

# JVLC

**Journal of  
Visual Language and  
Computing**

**Volume 2019, Number 2**

Copyright © 2019 by KSI Research Inc.

All rights reserved. No part of this publication may be reproduced, stored in a retrieval system, or transmitted, in any form or by any means, electronic, mechanical, photocopying, recording, or otherwise, without the prior written consent of the publisher.

DOI: 10.18293/JVLC2019-N2

ISSN: 2573-7147 (online)

Journal preparation, editing and printing are sponsored by KSI Research Inc.

# **Journal of Visual Language and Computing**

## **Editor-in-Chief**

**Shi-Kuo Chang, University of Pittsburgh, USA**

## **Co-Editors-in-Chief**

**Gennaro Costagliola, University of Salerno, Italy**

**Paolo Nesi, University of Florence, Italy**

**Gem Stapleton, University of Brighton, UK**

**Franklyn Turbak, Wellesley College, USA**

**An Open Access Journal published by**

**KSI Research Inc.**

**156 Park Square Lane, Pittsburgh, PA 15238 USA**

## **JVLC Editorial Board**

Tim Arndt, Cleveland State University, USA  
Paolo Bottoni, University of Rome, Italy  
Francesco Colace, University of Salerno, Italy  
Maria Francesca Costabile, University of Bari, Italy  
Philip T. Cox, Dalhousie University, Canada  
Martin Erwig, Oregon State University, USA  
Andrew Fish, University of Brighton, United Kingdom  
Vittorio Fuccella, University of Salerno, Italy  
Angela Guercio, Kent State University, USA  
Erland Jungert, Swedish Defence Research Establishment, Sweden  
Kamen Kanev, Shizuoka University, Japan  
Robert Laurini, University of Lyon, France  
Jennifer Leopold, Missouri University of Science & Technology, USA  
Mark Minas, University of Munich, Germany  
Brad A. Myers, Carnegie Mellon University, USA  
Joseph J. Pfeiffer, Jr., New Mexico State University, USA  
Yong Qin, Beijing JiaoTung University, China  
Genny Tortora, University of Salerno, Italy  
Kang Zhang, University of Texas at Dallas, USA

## **Journal Production Associate Editors**

Jorge-Luis Pérez-Medina, Universidad de Las Américas, Ecuador  
Yang Zou, Hohai University, China



# Journal of Visual Language and Computing

Volume 2019, Number 2

November 2019

## Table of Contents

### Regular Papers

- Multilevel Visualisation of Topic Dependency Models for Assessment Design and Delivery: A Hypergraph Based Approach..... 69  
*Kendra Cooper and Hassan Khosravi*
- Metaphorical Parametric Model for Brand Mark Design: Towards a Universal Model of Computational Visual Communication Design . . . . . 83  
*Po Choy Ng and Clifford Sze-Tsan Choy*
- Graphical Animations of the Suzuki-Kasami Distributed Mutual Exclusion Protocol . . . . . 105  
*Dang Duy Bui and Kazuhiro Ogata*

### Research Notes

- An Empirical Analysis on Effectiveness of Source Code Metrics for Models Predicting Aging Related Bug. . . . . 117  
*Chinmay Hota, Lov Kumar and Lalita Bhanu Murthy Neti*
- Text Entry on Circular Smartwatches: The C-QWERTY Layout . . . . . 127  
*Gennaro Costagliola, Raffaele D'Arco, Sabato De Gregorio, Mattia De Rosa, Vittorio Fuccella and Daniele Lupo*
- Run-Time Adaptation of Modalities of Interaction and Context of Use in Web Mobile Apps..... 135  
*Danilo Camargo Bueno and Luciana A. M. Zaina*
- An Advanced Software Dashboard for Supporting the Development of Adaptable, Energy Efficient Applications..145  
*Vladimir Ivanov, Daria Larionova, Dragos Strugar, Giancarlo Succi and Eugene Zouev*



# Journal of Visual Language and Computing

journal homepage: [www.ksiresearch.org/jvlc](http://www.ksiresearch.org/jvlc)

## Multilevel Visualisation of Topic Dependency Models for Assessment Design and Delivery: A Hypergraph Based Approach

Kendra M. L. Cooper Ph.D.<sup>a,\*</sup><sup>1</sup>, Hassan Khosravi Ph.D.<sup>b</sup><sup>2</sup>

<sup>a</sup>Independent Scholar, Canada

<sup>b</sup>The University of Queensland, Australia

### ARTICLE INFO

#### Article History:

Submitted 4.14.2019

Revised 4.20.2019

Second Revision 10.23.2019

Accepted 10.25.2019

#### Keywords:

educational assessment

topic dependency model

hypergraph

visualisation

### ABSTRACT

The effective design and delivery of assessments in a wide variety of evolving educational environments remains a challenging problem. Proposals have included the use of learning dashboards, peer learning environments, and grading support systems; these embrace visualisations to summarise and communicate results. In an on-going project, the investigation of graph based visualisation models for assessment design and delivery has yielded promising results. Here, an alternative graph foundation, a two-weighted hypergraph, is considered to represent assessment results and their explicit mapping to one or more learning objective topics. The visualisation approach considers the hypergraph as a collection of levels; the content of these levels can be customized (i.e., filtered) and presented according to user preferences. A case study on generating hypergraph models using commonly available assessment data and a flexible visualisation approach using historical data from an introductory programming course is presented.

© 2019 KSI Research

## 1. Introduction


Assessment remains a core educational activity, even as environments continue to evolve beyond the traditional classroom. Blended, flipped, and massive open on-line courses are supported by a wide range of assessment tools and techniques [10]. Instructors have many options for assessing the required topics of a course (e.g., CS 100 Introduction to Computer Science) in a particular offering (e.g., CS 100 Term 1 2019 Section 001) [12]. More traditional material includes homework assignments and examinations; emerging material includes question repositories and games. This rich variety also introduces new challenges to educational stakeholders (e.g., students, instructors, administrators, ed-

ucation researchers) with respect to evaluating the coverage of assessment material and communicating achievements. Students may find it challenging to infer their strengths and weaknesses with respect to the topics and their relationships, which can impede their studies. Outside a classroom program administrators, course designers, course co-ordinators, and researchers also face challenges. Administrators find it challenging to compare the content and difficulty of formal assessments as well as students' outcomes across different offerings of a course. Course designers and co-ordinators find it challenging to ensure the required topics and their relationships (e.g. questions with a combination of topics  $T_1$ ,  $T_2$  and  $T_3$ ) have been assessed. Educational researchers need to compare the achievement results between control and experimental groups.

To enable different educational stakeholders to explore, understand, and communicate insights from educational data sets, a new field known as "Learning Dashboards" has emerged [20, 2, 18]. These dashboards employ data sets on students' engagement and performance that are commonly available in educational settings to provide interactive visualisation widgets and techniques in the educational context. A vari-

In this work, a topic dependency model for educational assessment is introduced. The model, based on a hypergraph foundation, is visualised with a multilevel approach that supports user selected filtering options for achievements and topic coverage.

\*Corresponding author

 [kendra.m.cooper@gmail.com](mailto:kendra.m.cooper@gmail.com) (K.M.L. Cooper);

[h.khosravi@uq.edu.au](mailto:h.khosravi@uq.edu.au) (H. Khosravi)

ORCID(s): 0000-0003-4771-1371 (K.M.L. Cooper)

<sup>1</sup>Independent Scholar, Canada.

<sup>2</sup>Senior Lecturer, The University of Queensland, Australia.

ety of traditional plots and charts have been adopted in these dashboards (e.g., pie, box, histogram, radial) to visualise the achievements of students. The research typically considers independent (stand-alone) topics. Additional discussion on the related work in this area is presented in Section 3.

As part of an on-going research project, the authors of this paper have explored a collection of topic dependency models (TDMs) for assessment in which the relationships among topics are considered [14, 8]. The TDMs use a two-weighted undirected graphs foundation to formally represent and visualise a wide variety of assessment data for one or more topics (i.e., topics and their dependency relationship) to meet the needs of diverse stakeholders. The collection consists of a course reference model (to establish the topics and dependencies covered in a course), in addition to classroom models, both static and dynamic. As the work presented in this article leverages these results, a background section on the original TDM collection is presented in Section 2.

In this work, an alternative graph foundation is explored for the TDM collection, a two weighted hypergraph, referred to as  $TDM_{MH}$ ; the current results focus on static classroom models. This graph can explicitly visualise the n-ary topic coverage and achievement inherent in questions (i.e., a question can assess 1:N topics). However, visualisations of their more general n-ary hyperedge relationships may be more difficult to understand in the broader community. Here, a method to generate  $TDM_{MH}$  using commonly available assessment data and a flexible approach to address the complexity of the graphs are presented. The visualisation is a multilevel filtering approach that considers the hypergraph as a collection of levels. The levels partition a graph based on the number of nodes associated with the hyperedges. The content of these levels can be customised (i.e., filtered) and presented according to user preferences. For example, instead of being presented with the complete hypergraph, the users can select topics, achievement or number of responses (e.g., maximum, minimum, range of values) to view part of it. In addition, the visualisations of the levels can be presented in either a cumulative or accumulative mode. The new models are presented using an illustrative example. In addition, a case study that illustrates how  $TDM_{MH}$  can be applied to provide insight for instructors is presented. This study uses historical data from a first-year undergraduate level offering of a course on programming and engineering design at The University of British Columbia. The case study includes a comparison of the use of the newly proposed  $TDM_{MH}$  with the original TDM approach.

## 2. Background: The Original TDM Collection

A collection of TDMs has been presented in the authors' previous work [14, 8] to address some of the challenges stakeholders face that relate to the design, delivery and analysis of assessments. An overview of the collection is presented in Figure 1. The TDM collection is revisited in this section as it provides a foundation for the current work.

### 2.1. Scenarios of Use

The wide variety of stakeholders interacting with these models (e.g., students, instructors, and so on) are shown at the top of the figure; they have roles inside and outside of the classroom.

The users can select the assessment data of interest. For example, they may need to explore data from formative (e.g. assignments) or summative (e.g., examinations) assessments; the data can be from one or more classes (current or previous). The course data includes a specification of the course topics that need to be assessed. The input assessment data are shown on the left of Figure 1.

A preliminary scenario analysis has identified 40 questions for the stakeholders. For example, instructors interact with the models to explore questions such as:

- What topics and their relationships do I need to assess?
- What topics are covered each of the assessments?
- How well are the students performing on the topics?
- How well are the students performing compared to other (current or previous) classes?
- Who in the class may be at risk of failing the course?
- How have all the topics and their relationships been assessed?
- How much has the class improved over time on the topics?

Here, scenarios of use include filtering by topic(s), or constraints involving achievement or coverage. Additional filters can be added in a straightforward way. Currently, the scenarios focus on using the static models; the reference and dynamic models for the  $TDM_{MH}$  are planned in future work. Two examples, derived from the scenario analysis in [14], are described below.

**Topics.** How are the students performing on a specific topic? One or more topics of interest can be selected in order to identify topics that may need additional attention. For example, an instructor can select one of many topics covered in a course; the levels are filtered to emphasize the topic(s) of interest.

**Achievements.** What topics have poor achievement? A constraint for visualising a particular achievement level can be selected. For example, an instructor can select to view the minimum achievement scores over all topics in order to identify those that need additional attention; the levels are filtered to emphasize the achievement scores of interest.

The users can choose to combine filters. In other words, they can filter the visualisations based on one or more constraints such as selecting both a topic and an achievement constraint, e.g., the highest performance involving topic  $T_1$ . The results are filtered on a level-by-level basis.

When visualising the filtered results, the user can also choose to view the results in either the cumulative or accumulative mode. In the accumulative mode, only the results for one level are presented. In the cumulative mode, the re-

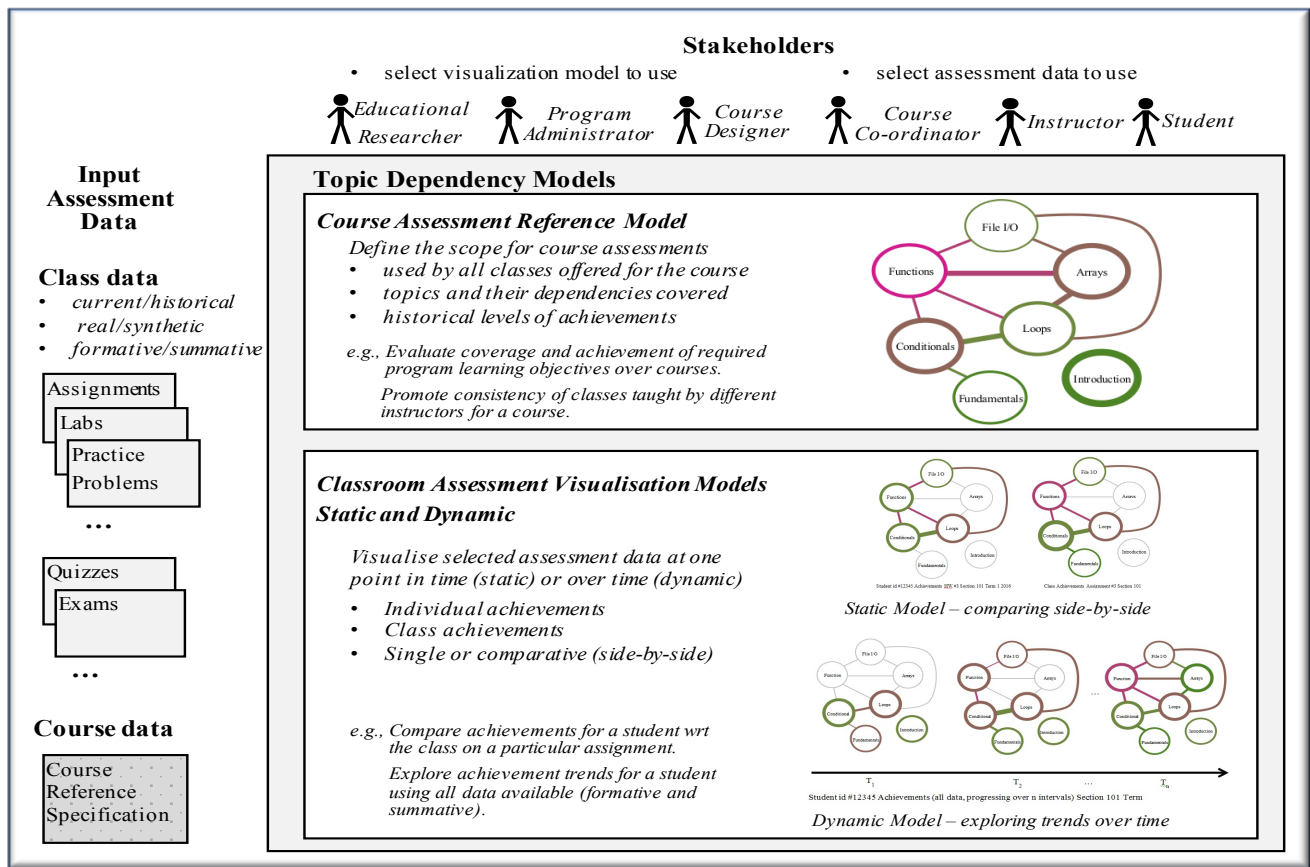


Figure 1: Original TDM model collection: two-weighted undirected graph foundation [14].

sults up to and including the current level of interest are presented.

## 2.2. Topic Dependency Models

Based on the requests from stakeholders, the input data are selected and transformed into visual models. As illustrated in the main block of Figure 1, the TDM collection consists of two kinds of models: Course Assessment Reference and Classroom Assessment Visualisation (static and dynamic). The Reference model establishes and communicates the required topics and their relationships for a course (e.g., a CS1 course is an introductory programming course covering variables, branching, loops, and so on). It provides a common foundation for all sections of a course offered over time (e.g., CS1 Section 001 Term 1 2019 Instructor A. Smith). The reference model helps to ensure the consistent coverage of topics by different instructors, clearly communicate the scope of the topics to students, and support administrative activities related to monitoring learning objective outcomes.

The Classroom Assessment Visualisation models present assessment data within a class and support comparisons of assessment data between classes at one point in time (static) and over time (dynamic). The assessment data are selectively visualised, e.g., for the whole class, individual students, specific topics, and so on. For example, the user can

choose to visualise static models, in a side-by-side comparison, for a specific assessment (e.g., Assignment 3) for all of the students in two classes. Alternatively, the user can choose to visualise dynamic models to explore the progression of a class over time.

The Reference and Classroom Assessment Visualisation models share a common graph foundation: a two-weighted undirected graph. The vertices represent topics in a course; the edges represent assessment material (e.g., questions) that address the topics the edges are related to. The weights are reflected in the visualisation using a colour palette (achievement) and width (coverage). As the graphs can only represent edges with up to two vertices, questions involving three or more topics must be redistributed in the models as combinations, in order to visualise them. For example, a question covering three topics (e.g., A, B, C) is redistributed with estimations into six relationships for visualisation: A,A; B,B; C,C, A,B; A,C; and B,C.

## 3. Related Work

The development of  $TDM_{MH}$  draws upon the literature from Learning Dashboards and hypergraph based methods. These are briefly presented in this section.

### 3.1. Learning Dashboards

The learning analytics community continues to actively investigate approaches that support the exploration of learning activities by different stakeholders. With the increase in the use of educational technologies and the advancements in the areas of learning analytics and educational data mining, a new field, commonly known as “Learning Dashboards” has emerged to help make sense of data sets in learning and education [20, 2, 18]. A variety of visualisations such as bar charts [9], pie charts [23], histograms [16], box plots [1], radar graphs [17], and skill meters [6] have been adopted to show the achievements of students for independent (stand-alone) topics. [20] presents a systematic literature review on the use of learning dashboards. Based on the findings of this literature review, the use of graph-based visualisations in learning dashboards has not received much attention. In addition, a number of studies provide strong evidence that opening the model to learners, leading to the notion of Open Learner Models (OLMs) [7], can be effective in helping students learn [3]. The OLMs commonly use a set of individual topics as their underlying structure for modeling learners’ knowledge state, which ignores relationships among topics. An emerging new field applied methods from process mining and sequential data mining to educational data [4] to facilitate better understanding of the educational process. Educational process mining methods predominantly use graphs, but their main focus is on using activity logs to visualise students’ learning process in terms of the time, place, path, pace of learning activities.

### 3.2. Hypergraph Based Methods

Hypergraphs have been adopted as a foundation for data analytics and visualisations in a wide variety of domains including data warehousing, communication network analysis, geospatial metadata, and cellular biology networks. To support interactive queries in data warehousing systems, a framework for developing dashboards called Dashboard-by-example has been proposed [13]. This framework adopts hypergraph-based techniques to transform dissimilar, heterogeneous data into a homogeneous knowledge space of clusters and partitions. The framework is flexible, as the hypergraph-guided data linkages support the exploration and aggregation of data from multiple perspectives. For the analysis of network traffic traces, available in massive communication logs, a hypergraph based visualisation is proposed in [11]. The network traffic visualisation approach utilises a frequent item set mining method to identify interesting traffic patterns in the large amount of data. The patterns are visualised as hypergraphs with explicit, multi-attribute relationships. A framework to comprehensively address all of the available geospatial metadata standards (i.e., documents) is proposed in [19]. Geospatial metadata describe geographic digital data resources such as earth imagery, geospatial databases and catalogues, and Geographic Information System files. The framework integrates hypergraphs and topic maps, representing the elements and their dependency relationships. The potential for adopting hypergraphs in the domain of cel-

lular biology is introduced in [15]; recently, for example, [21] proposes a framework that adopts hypergraphs and associated hypergraphs to describe, analyse, and identify metabolic network alignments at the full genome level. These alignments are used to discover important similarities and differences between distinct molecular networks: they reveal mappings between components (topological, biological functional) across different networks.

## 4. Methodology

This section presents the problem definition, graph foundation (hypergraph, level), analysis metrics, and an illustrative example. The graph generation and visualisation algorithm are presented in Section 5, using the example introduced in this section. A case study demonstrating real life examples of how  $TDM_{MH}$  can practically be used is presented in Section 6.

### 4.1. Problem Definition

Based on the related work analysis, a modeling and visualisation approach based on hypergraphs for educational assessment does not appear to be available at this time. The  $TDM_{MH}$  (static model) is introduced in this work to help address this gap. More specifically, the research questions are:

- RQ1. How can a two-weighted hypergraph be utilised to model topics, achievements, and coverage in the educational assessment problem domain?
- RQ2. How can the two-weighted hypergraph models representing assessment data be effectively visualised?
- RQ3. How can established graph metrics be used to quantitatively assess the visual complexity of the hypergraph based visualisations of the models?
- RQ4. What are the strengths and limitations of the  $TDM_{MH}$  in comparison to the original TDM?

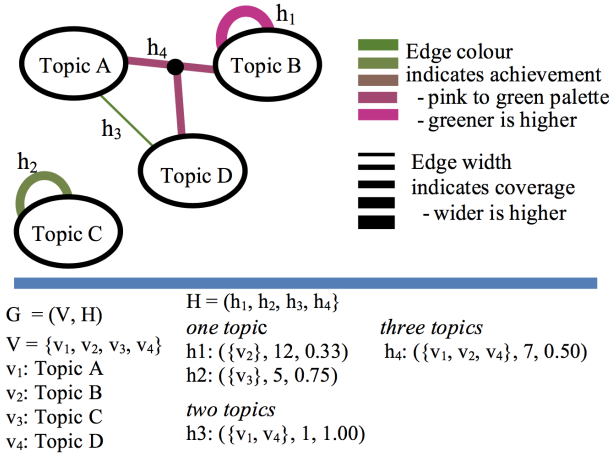
### 4.2. Hypergraph Foundation of the $TDM_{MH}$

In this section, the formal definition for the two-weighted hypergraph and the concept of a level are presented.

**Definition 1.** A two-weighted, undirected hypergraph  $G = (V, H)$ , where  $V$  is the set of vertices representing the topics, and  $H$  is the set of hyperedges. A hyperedge  $h \in H$  is represented as  $h = (C, c_1, c_2)$ , where  $C$  is the subset of the vertices being connected, and  $c_1$  and  $c_2$  represent the two weighted values for an edge. An edge involving only one node represents a self-loop.

In this work, the nodes are used represent topics; the hyperedges represent the assessment material that covers the topics. The weight  $c_1$  represents the number of learning objects that are tagged with the topics for the hyperedge; and  $c_2$  represents the represents the performance (e.g., achievement) on these learning objects. A simple example of a





**Figure 2:** Example of a two-weighted hypergraph: definition and visualisation.

$TDM_{MH}$  visualisation is illustrated in Figure 2. There are four topics in this example (A, B, C, and D); each topic is represented by a node. Topic B and Topic C both have self loops; Topic C is isolated which indicates the topic it represented is assessed on its own. For Topic B, the edge has a dark pink colour and a wide width, reflecting the poor average performance (33%) on the questions answered and the large number of responses. The hyperedge for Topic C is a medium green colour and has a medium width; this represents the good average performance (75%) and a moderate number of responses to questions. A 2-ary hyperedge exists between Topics A and D. This edge is dark green and thin, in representing the excellent performance (100%) on the related questions and a small number of responses. A 3-ary hyperedge exists between Topics A, B, and D. This edge is a brownish pink colour and has a moderate weight, which represents the moderate average performance (50%) and a moderate number of responses to questions on these three topics.

The concept of a level is used in this work to support the visualisation. For example, in the simple example given in Figure 2, there are three levels.  $Level_1$  is a subset of the hyperedges with one vertex (i.e.,  $h_1, h_2$ ),  $Level_2$  is a subset of the hyperedges with two vertices (i.e.,  $h_3$ ), and  $Level_3$  is a subset of the hyperedges with three vertices (i.e.,  $h_4$ ). More generally,  $Level_i$  is the subset of hyperedges involving  $i$  vertices.

### 4.3. Analysis Metrics

A basic collection of structural metrics has been selected, which are well-defined for hypergraphs (e.g., [5, 24]). As graphs are a special case of hypergraphs, the metrics can be applied to both hypergraphs and graphs. They are used in this work to help explore the complexity of the  $TDM_{MH}$  visualisations in Section 5. They are also applied in the comparative case study in Section 6 to compare and contrast the original TDM with the  $TDM_{MH}$ .

In the following definitions,  $G = (V, H)$  is an undirected hypergraph,  $v \in V$ , and  $h \in H$ . The order of the hypergraph

$G$  is the cardinality of  $V$ , i.e.,  $|V| = n$ . In other words, the order of the graph is the number of vertices it has. The size of the hypergraph  $G$  is the cardinality of  $H$ , i.e.,  $|H| = m$ . In other words, the size of the hypergraph is the number of hyperedges it has.

Degree metrics can be considered at both the individual node level and for a hypergraph as indicators of a graph's complexity. For an individual vertex the degree is the number of hyperedges (ingress or egress) the node is contained in. A node in a graph, denoted  $G(v)$ , can be contained in one or more hyperedges. For a hyperedge ( $h_j$ ),  $j \in J$  that contains the node  $v$ , the degree of that node is  $d(v) = |J|$ . In Figure 2, for example, the  $d(\text{Topic A})$  is two, as it is contained in two hyperedges  $h_2$  and  $h_3$ . The  $d(\text{Topic B})$  is three, as the hyperedge  $h_1$  contributes a count of two (egress and ingress) and  $h_4$  also contains the node.

Beyond the individual node degree computations, additional degree metrics are considered at the graph level. These include the total degree of  $G$ , denoted by  $d_t(G) = |H(v)|$ ; the average degree of a graph is the total divided by the order, i.e.,  $d_t(v)/|V|$ , or  $|H(v)|/|V|$ . In addition, the maximal degree of a hypergraph  $G$  is denoted by  $\Delta(G)$ ; the minimal degree is denoted by  $\delta(G)$ . For example, the average, maximal, and minimal degrees of Figure 2 are 2.25, 3, and 2.

### 4.4. Illustrative Example

An example based on six students, three formative assignments (five questions per assignment), and six topics has been defined by the authors. The following constraints are considered for creating the illustrated example:

- The questions must span a range of one to four topics.
- There must be a range in the individual student achievements in the responses (low to high scores).
- There must be a range in the number of responses to the questions.
- There must be a range in the average scores for the questions.
- Assessments on groups of topics demonstrate a range of achievements (very poor to very strong); one or more questions address the groups of topics.
- The multilevel visualisation algorithm with a list of filtering options (e.g., selected topic(s), achievement, cumulative/accumulative mode) can be clearly illustrated.

Table 1 presents a sample data set meeting these constraints. As input to the  $TDM_{MH}$  generation algorithm, the data are formatted into two CSV files: (1) A student-question-answer (SQA) file that contains the student identifier, question identifier, and score (correct/incorrect), and (2) a question-topic (QT) file that contains the question identifier and topics (i.e., tags) it addresses.

Q	Topic List	Student ID					
		$S_1$	$S_2$	$S_3$	$S_4$	$S_5$	$S_6$
<b>Assignment 1</b>							
$Q_1$	$T_1$	-	1	-	-	0	1
$Q_2$	$T_3$	1	1	1	1	1	1
$Q_3$	$T_4$	0	1	0	1	1	1
$Q_4$	$T_1, T_2$	1	1	0	1	1	0
$Q_5$	$T_1, T_4$	1	-	-	0	1	0
<b>Assignment 2</b>							
$Q_6$	$T_4$	0	0	1	0	0	1
$Q_7$	$T_1, T_4$	-	0	1	1	0	0
$Q_8$	$T_4, T_5$	0	1	0	0	0	1
$Q_9$	$T_1, T_4, T_5$	0	0	0	0	-	1
$Q_{10}$	$T_1, T_2, T_4, T_5$	0	0	0	1	0	1
<b>Assignment 3</b>							
$Q_{11}$	$T_1, T_4$	1	-	0	1	-	1
$Q_{12}$	$T_2, T_6$	0	0	0	0	0	0
$Q_{13}$	$T_1, T_2, T_6$	-	-	1	1	-	1
$Q_{14}$	$T_1, T_2, T_4, T_5$	-	0	1	0	0	1
$Q_{15}$	$T_2, T_4, T_5, T_6$	-	0	0	1	1	1

**Table 1**

Illustrative example: questions, topic lists, and student responses. The value 1 indicates the question is answered correctly; 0 indicates it is answered incorrectly; and - indicates it is not attempted.

## 5. $TDM_{MH}$ Static Model

In this section, the approach to generating and filtering a  $TDM_{MH}$  is presented. Section 5.1 presents methods for generating the graph and Section 5.2 explores filtering mechanisms to address the complexity of hypergraph visualisations. The results are briefly discussed in Section 5.3.

### 5.1. Generating a $TDM_{MH}$

This section demonstrates how commonly available input data (student achievements/grades for specific questions) and the mapping from the questions to the course topics are transformed into a  $TDM_{MH}$ . High-level code and notation are presented in Algorithm 1. The algorithm consists of three high level steps: create working dictionaries and matrices; define the  $TDM_{MH}$  graph elements (vertices and hyperedges); and visualise (i.e., plot) the  $TDM_{MH}$  graph.

The first six steps of Algorithm 1 transform the data in the SQA.csv and QT.csv files into working dictionaries and matrices. In Steps 1, 2, and 3 of the algorithm, three dictionaries are created,  $QDict$ ,  $SDict$ ,  $TDict$ , to map array and matrix indices to question, student, and topic identifiers. In Steps 4, 5, and 6, three working matrices are created:  $T$ ,  $A$ , and  $R$ . The information on topics assigned to each question is represented in matrix  $T$ , in which  $t_{ij} = 1$  indicates that question  $i$  is tagged with topic  $j$  and  $t_{ij} = 0$  indicates that question  $i$  is not tagged with topic  $j$ . The correctness of the answers provided by the users are represented in matrix  $A$ , where  $a_{ij} = 1$  indicates that user  $i$  has answered question  $i$  correctly,  $a_{ij} = 0$  indicates that user  $u$  has answered question  $i$  incorrectly, and  $a_{ij} = n$  indicates that user  $u$  has not attempted question  $i$ . Matrix  $R$  is used to keep track of attempted questions, where  $r_{ij} = 1$  if user  $i$  has attempted question  $i$  and zero otherwise. Figure 3 illustrates the data

### Algorithm 1 Generating a $TDM_{MH}$

**Require:**  $SQA.CSV$ ,  $QT.CSV$ ,  $filters$

**Create dictionaries and matrices for efficient indexing**

- 1:  $QDict \leftarrow CreateQDict(SQA.CSV)$
- 2:  $SDict \leftarrow CreateSDict(SQA.CSV)$
- 3:  $TDict \leftarrow CreateTDict(QT.CSV)$
- 4:  $T \leftarrow CreateT(QT.CSV, QDict, TDict)$
- 5:  $A \leftarrow CreateA(SQA.CSV, SDict, QDict)$
- 6:  $R \leftarrow CreateR(SQA.CSV, SDict, QDict)$

**Compute the Graph Elements: Vertices and Edges**

- 7:  $VList \leftarrow ComputeV(TDict)$
- 8:  $HList \leftarrow ComputeH(T, A, D, TDict)$

**Create and visualise the Graph**

- 9:  $TDMStatic \leftarrow CreateTDM(VList, HList)$
- 10:  $Visualise(TDMStatic, filters)$

stored in the dictionaries ( $QDict$ ,  $SDict$ ,  $TDict$ ) and the matrices ( $T$ ,  $A$ ,  $R$ ).

Steps 7 and 8 in Algorithm 1 establish the graph vertices and hyperedges.  $VList$  stores the list of the vertices of the  $TDM_{MH}$  graph and  $HList$  stores the list of its hyperedges. The coverage and competency associated with an edge are both computed using  $T$ ,  $A$  and  $R$  within the  $ComputeH$  function. The coverage ( $Cov$ ) associated with a hyperedge among a set of vertices  $V = v_j, \dots, v_k$  is computed based on the number of attempts that have been made on questions that are tagged with exactly topics that are included in  $V$ . It is computed using the following formula:

$$Cov(V) = \sum_{i \in QDict} \left( \prod_{j \in V} t_{ij} \prod_{j \notin V} (1 - t_{ij}) \right) \sum_{u \in SDict} r_{ui} \quad (1)$$

The outer summation loops through all of the questions. For a question  $i$ ,  $(\prod_{j \in V} t_{ij} \prod_{j \notin V} (1 - t_{ij}))$  results in one if question  $i$  is tagged with exactly the topics that are included in  $V$  and zero otherwise. Together, these two parts identify questions that are tagged with exactly topics that are included in  $V$ . The inner summation then counts the number of attempts that have been made on such a question. In the given example,  $Cov(\{T_1, T_4\})$  has coverage of 13, which consists of contributions of 4 from  $Q_5$ , 5 from  $Q_7$ , and 4 from  $Q_1$ .

The achievement ( $Achv$ ) associated with a hyperedge among a set of vertices  $V = v_j, \dots, v_k$  is computed as the success rate of answering questions that are tagged with exactly topics, which are included in  $V$ . It is computed using the following formula:

$$Achv(V) = \frac{\sum_{i \in QDict} \left( \prod_{j \in V} t_{ij} \prod_{j \notin V} (1 - t_{ij}) \right) \sum_{u \in SDict} a_{ui}}{Cov(V)} \quad (2)$$



**QDict:**

{'Q1':0, 'Q2':1, 'Q3':2, 'Q4':3, 'Q5':4, 'Q6':5, 'Q7':6, 'Q8':7, 'Q9':8, 'Q10':9, 'Q11':10, 'Q12':11, 'Q13':12, 'Q14':13, 'Q15':14}

**SDict:**

{'S1':0, 'S2':1, 'S3':2, 'S4':3, 'S5':4, 'S6':5}

**TDict:**

{'T1':0, 'T2':1, 'T3':2, 'T4':3, 'T5':4, 'T6':5}

Topics (6)						Questions (15)														
Questions (15)	Students (6)	$\begin{bmatrix} 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 \\ 1 & 1 & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 \\ 1 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 1 & 0 \\ 1 & 0 & 0 & 1 & 1 & 0 \\ 0 & 1 & 0 & 0 & 1 & 1 \\ 1 & 1 & 0 & 1 & 1 & 0 \\ 0 & 1 & 0 & 1 & 1 & 1 \end{bmatrix}$																		
		Matrix A																		
	Students (6)	$\begin{bmatrix} 0 & 1 & 1 & 1 & 1 & 1 & 0 & 1 & 1 & 1 & 1 & 1 & 0 & 0 & 0 \\ 1 & 1 & 1 & 1 & 0 & 1 & 1 & 1 & 1 & 1 & 0 & 1 & 0 & 1 & 1 \\ 0 & 1 & 1 & 1 & 0 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 \\ 0 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 0 & 1 & 0 & 1 & 0 & 1 & 1 \\ 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 \end{bmatrix}$																		
		Matrix R																		

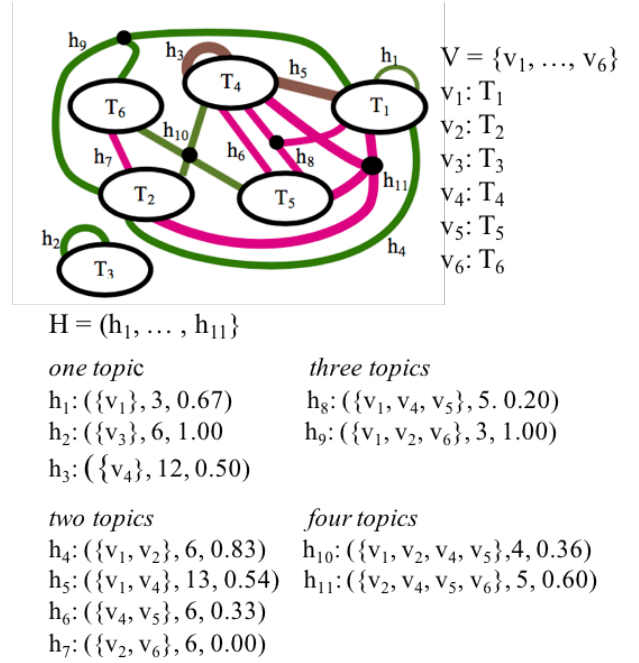
**Figure 3:** Illustrative Example Data Organization. The data represented in Table 1 are loaded and transformed into three dictionaries: *QDict*, *SDict* and *TDict* (Steps 1, 2, and 3 in Algorithm 1). The dictionaries and the input files are used to create three working matrices: T, A, and R (Steps 4, 5, and 6). These matrices are used to create the graph elements (Steps 7, 8). The  $TDM_{MH}$  model is created (Step 9) and subsequently visualised (Step 10) (refer to Figure 4).

The numerator of this formula count the number of correct answers given to questions that are tagged with exactly topics included in  $V$ . The formula is similar to how  $Con(V)$  is computed, but instead of using  $\sum_{u \in SDict} r_{ui}$  to count the number of attempts made on questions that are tagged with exactly topics included in  $V$ ,  $\sum_{u \in SDict} a_{ui}$  is used to compute only the number of correct answers. Dividing this number by  $Cov(V)$  produces the rate of correctly answering questions tagged with only topics in  $V$ . In the given example  $Achv(\{T_1, T_4\})$  is computed as  $\frac{2+2+3}{13} = \frac{7}{13} = 0.54$ . Note that  $Cov$  and  $Achv$  of a single topic  $j$  can be computed via the same two formulas by using  $V = v_j$ .

Step 9 of Algorithm 1 creates the graph model based on the  $VList$  and  $HList$  computed in steps 7 and 8 of the algorithm. The lower part of Figure 4 presents the  $VList$ ,  $HList$ , and two-weighted hypergraph model created for the illustrative example.

## 5.2. Visualising a $TDM_{MH}$

Step 10 of Algorithm 1 visualizes the model according to the user selected filtering options and display mode. The  $TDM_{MH}$  approach supports filtering combinations including one or more topics as well as one or more achievement conditions.



**Figure 4:** Complete  $TDM_{MH}$  for the data set provided in Table 1 using Algorithm 1 (no filtering, default cumulative mode). The example does not have questions that involve five or six topics.

Three straightforward visualisation examples are presented in this section for illustrative purposes: no filtering (default cumulative mode), filtering by one topic (accumulative mode), and filtering by an achievement condition (cumulative mode).

To accomplish the visualisation step, the levels in the graph are first identified based on the structure of the graph. The filters selected by the user are then applied to the levels. Lastly, the filtered levels are displayed in accordance with the mode (accumulative, cumulative).

The illustrative example (with six nodes) has six levels. Table 2 illustrates the levels before any filtering is done. For example,  $Level_1$  is the subset of hyperedges  $\{h_1, h_2, h_3\}$ .

### 5.2.1. Example 1 (no filtering)

This example illustrates the situation in which no filtering options are selected by the user and the default cumulative mode is used. In step 10 of Algorithm 1, the levels for the graph are first identified (refer to Table 2).

If the user does not select any filters, then all of the levels are displayed by default in the cumulative mode, including all topics, achievements, and coverage of the learning objects. In this situation, the filtered levels are the same as the original levels, i.e.,  $Level_i$  and  $Level'_i$  are the same (refer to Example 1 in Table 2).

The default cumulative mode is used, which results in the visualisation in Figure 4.

The metrics associated with this example are presented in Table 3. The row corresponding to the visualised model are emphasized with the lightgray colour; intermediate assessments are not highlighted. In the cumulative visuali-

Level ID	Level Membership	Example 1 Level' Membership (no filtering)	Example 2 Level' Membership: Topic is $T_1$	Example 3 Level' Membership: Achievement is $\leq 20\%$ OR $\geq 80\%$
$Level_1$	$\{h_1, h_2, h_3\}$	$\{h_1, h_2, h_3\}$	$\{h_1\}$	$\{h_2\}$
$Level_2$	$\{h_4, h_5, h_6, h_7\}$	$\{h_4, h_5, h_6, h_7\}$	$\{h_4, h_5\}$	$\{h_4, h_7\}$
$Level_3$	$\{h_8, h_9\}$	$\{h_8, h_9\}$	$\{h_8, h_9\}$	$\{h_8, h_9\}$
$Level_4$	$\{h_{10}, h_{11}\}$	$\{h_{10}, h_{11}\}$	$\{h_{10}\}$	$\emptyset$
$Level_5+$	$\emptyset$	$\emptyset$	$\emptyset$	$\emptyset$

**Table 2**

Illustrative Example: Levels before and after filtering (Examples 1, 2, and 3).

Example	$\delta(H)$ (min)	$\Delta(H)$ (max)	E (avg)	Order	Size
Example 1 (no filtering, default cumulative mode)					
Level 1'	2	2	2.00	3	3
Level 2'	1	2	1.60	4	5
Level 3'	1	2	1.20	2	5
Level 4'	1	2	1.20	2	5
Levels 1'	2	2	2.00	3	3
Levels 1'-2'	1	4	2.33	7	6
Levels 1'-3'	2	6	3.33	9	6
Levels 1'-4'	2	7	4.67	11	6
Example 2 (filter on topic, accumulative mode)					
Level 1'	2	2	2.00	1	1
Level 2'	2	2	1.33	2	3
Level 3'	2	2	1.20	2	5
Level 4'	1	1	1.00	1	4
Example 3 (filter on achievement, cumulative mode)					
Level 1'	2	2	1.00	1	1
Level 2'	1	2	2.00	2	3
Level 3'	1	2	1.20	2	5
Level 4' ( $\emptyset$ )	0	0	0	0	0
Levels 1'	2	2	2.00	1	1
Levels 1'-2'	1	2	1.50	3	4
Levels 1'-3'+	1	3	2.00	5	6

**Table 3**

Illustrative Example: Metric summaries (Examples 1, 2, and 3).

sation involving Levels 1'-4', the minimum and maximum degree node values occur with the isolated node  $T_3$ , 2, and node  $T_4$ , 6, respectively. The average degree node value is relatively high, 4.67. This indicates a more complex visualisation for the users. All six of the nodes and all 11 of the hyperedges are included in the visualisation.

### 5.2.2. Example 2 (filtering by topic)

This example illustrates the situation in which the user select the option to filter by one topic ( $T_1$ ) and the accumulative mode is used. In step 10 of Algorithm 1, the levels for the graph are first identified (refer to Table 2).

For example, if a user requests to filter with respect to topic  $T_1$ , the levels are filtered accordingly. In this case, for example, the filtered  $Level'_i$  is the subset of hyperedges  $\{h_1\}$ , as only  $h_1$  contains  $T_1$  (refer to Example 2 in Table 2).

After the levels are filtered, they are visualised. In the accumulative mode, only the hyperedges for one level are presented at a time. The four filtered levels are visualised in Figure 5.

The metrics associated with Example 2 are presented in

Table 3. The rows corresponding to the visualised model are emphasized with the lightgray colour; all four rows are highlighted in the table to indicate they are visualised. The minimum node degree values ([1..2]) show limited variation; the maximum degree node values ([1..2]) are also low. There are a smaller number of nodes ([1..2]); a small number of hyperedges ([1..2]) are involved in the filtered levels. This results in reduced average node degree values ([1.00..2.00]), reflecting the reduced visual complexity of the individual levels in comparison to Example 1.

### 5.2.3. Example 3 (filtering by achievement)

This example illustrates the situation in which the user select the option to filter by an achievement condition, which identifies the more extreme performance outcomes in the assessment data for the class ( $\leq 20\%$  OR  $\geq 80\%$ ); the cumulative mode is used. In step 10 of Algorithm 1, the levels for the graph are first identified (refer to Table 2).

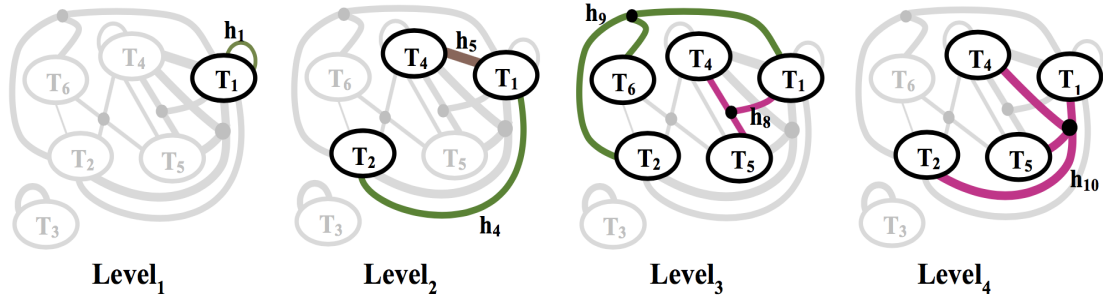
The levels are filtered according to the achievement condition selected by the user ( $\leq 20\%$  OR  $\geq 80\%$ ). There are five hyperedges that meet this achievement condition:  $h_2, h_4, h_7, h_8, h_9$  (refer to Example 3 in Table 2).

After the levels are filtered, they are ready to be visualised. In the cumulative mode, the presentation for a level includes the hyperedges for the current levels and the previous levels. The three (non-empty) filtered levels are visualised in Figure 6.

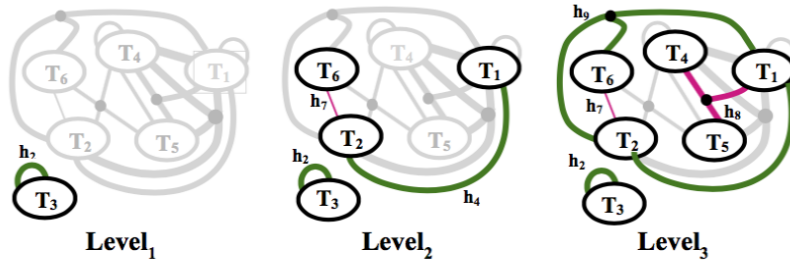
The metrics associated with this example are presented in Table 3. The rows corresponding to the visualised model are emphasized with the lightgray colour; rows containing intermediate assessments are not highlighted. The minimum node degree values ([1..2]), is low; the maximum degree node values ([2..3]) are also relatively low. There is a wider variation in the number of nodes involved in the cumulative visualisations ([1..6]); a relatively small number of hyperedges ([1..5]) are involved in the filtered levels. This results in reduced average node degree values [1.50..2.00] in the individual levels reflecting the reduced visual complexity of the individual levels in comparison to Example 1; the overall average in the cumulative visualisation involving the three non-empty levels is 2.00.

## 5.3. Discussion

The straightforward examples explore the use of the two-weighted hypergraph based  $TDM_{MH}$  model to generate (RQ1) and visualise (RQ2) assessment data. The representation of



**Figure 5:** Illustrative Example 2 multilevel visualisation: one topic is selected ( $T_1$ ) as a filter (accumulative mode).



**Figure 6:** Illustrative Example 3 multilevel visualisation: achievements at the more extreme ranges (very low  $\leq 20\%$ , very high  $\geq 80\%$ ) of the grades are selected using filter selections (cumulative mode).

the assessment data covering one or more topics is straightforward with the  $TDM_{MH}$  as it leverages the n-ary relationships of the hyperedges. The support for displaying all or part of the model of interest using the concepts of levels and filtering is flexible, allowing the user to focus on specific aspects of the assessment data. The preliminary set of metrics are applied, providing some insight on their relationship to the visual complexity of the models (RQ3). The average node index value appears to be useful as an indicator of the visual complexity. In the next section, the  $TDM_{MH}$  model is further explored using a historical case study in a self study and a comparative study with the original  $TDM$  model (RQ4).

## 6. Case Study

### 6.1. Historical Data

This case study is based on data collected from a first-year undergraduate level offering of a course on programming and engineering design at The University of British Columbia. This offering of the course had 377 students and was held during the Fall of 2016. The course covers many topics that are generally included in an introductory course on programming and engineering design in nine modules: number conversions, programming fundamentals, conditionals, loops, file I/O, functions, arrays, strings, and DAQ systems. Functions, strings and DAQ systems received two weeks of lecture time; all of the other modules received roughly one week of lecture time.

The data have been collected from the final exam of the

course, which are captured via the Gradescope platform [22], a system for the on-line assessment of handwritten exams. The final exam of this offering consists of 17 independent sub-questions that formed a total of eight main questions. The questions are tagged using 20 concept-level topics, defined by the instructors, based on the eight modules that are covered in this course. For example, the Strings Module is further decomposed into three topics: String-Length, String-Copy and String-Compare, providing a finer level of granularity.

### 6.2. $TDM_{MH}$ Visualisation

Three visualisation cases are presented in this section: no filtering, filtering by topic, and filtering by an achievement condition. The cumulative mode is used in all three cases. In these visualisations, a self-loop on a node is represented by the width and colour associated with the circle representing that node. The metrics associated with these cases are presented in Table 4.

#### 6.2.1. Case 1 (no filtering)

Figure 7 shows the  $TDM_{MH}$  with no filters for the exam data described in Section 6.1.

It indicates that the exam adequately covers all of the modules of the course and that all modules except Conversions (including the Hex and Octal topics) are highly connected to one another. The exam includes questions that had only a single topic (e.g.,  $h_1$  on Hex), two topics (e.g.,  $h_{14}$  on While Loops and DAQ-Write), three topics (e.g.,  $h_9$  on Printing, File Output and 2D Arrays) and four topics (e.g.,

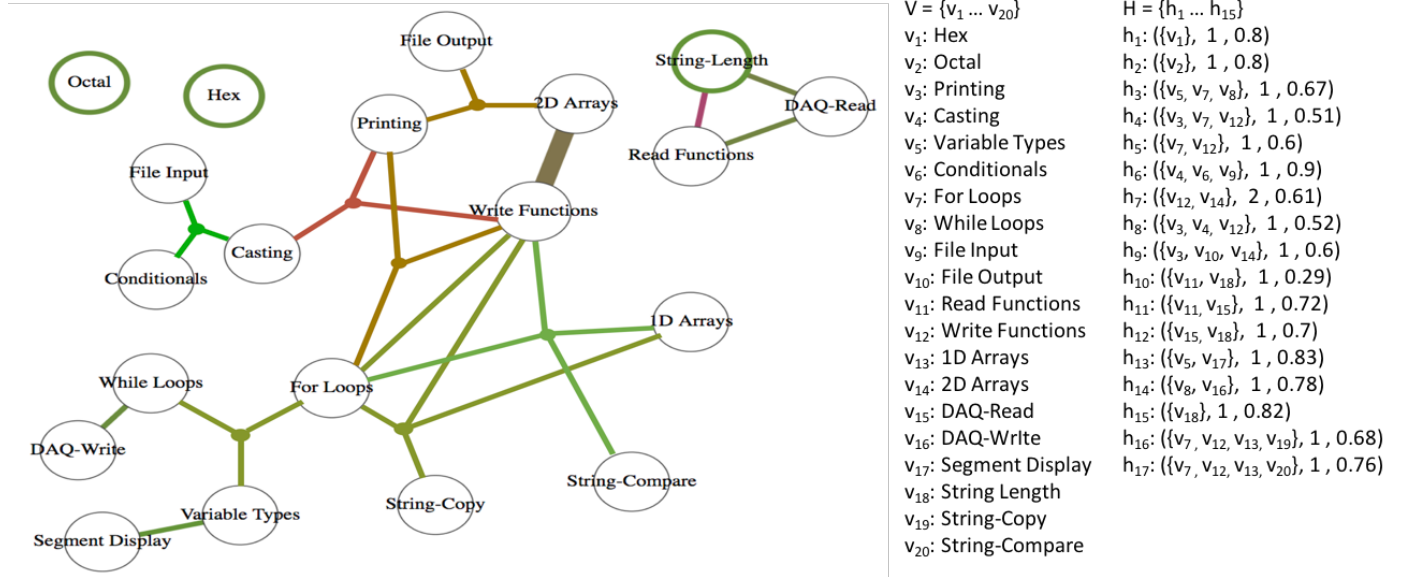


Figure 7:  $TDM_{MH}$  for the final exam of an offering of a first-year course on programming and engineering design

Case Study	$\delta(G)$ (min)	$\Delta(G)$ (max)	E (avg)	order	size
Case 1 (no filtering, default cumulative mode)					
Level 1'	2	2	2.00	3	3
Level 2'	1	2	1.40	7	10
Level 3'	1	3	1.0	5	10
Level 4'	1	2	0.80	2	5
All (Levels 1'-4')	1	6	2.15	17	20
Case 2 (filter on topic, Level 4, cumulative mode)					
Level 2'	1	2	1.33	2	3
Level 3'	1	2	1.50	2	4
Level 4'	1	2	1.60	2	5
Levels 1'-4'	1	6	2.25	6	8
Case 3 (filter on achievement, Level 3, cumulative mode)					
Level 2'	1	1	1.00	2	4
Level 3'	1	3	0.67	3	6
Levels 1'-3'	1	3	1.62	5	8

Table 4

Case Study:  $TDM_{MH}$  Metric Summaries for Cases 1, 2, and 3.

$h_{17}$  on For Loops, 1D Arrays, Write Functions, and String-Compare). The further decomposition to concept-level topics provides insights which might have not been possible to gauge using module-level tags. For example it shows that students have done well on questions on String-Copy and String-Compare but not so well on questions on String-Length. As another example, further decomposition of items from the Arrays Module shows that students are able to do well in 1D-Arrays, but perform quite poorly on questions on 2D-Arrays.

For the first case study, its associated metrics are presented in the top part of Table 4. The row corresponding to the visualised model is emphasised with the lightgray colour; the rows with intermediate assessment values are not highlighted. The minimum degree node value is one; this occurs, for example, with the nodes representing File Input (i.e.,  $v_9$ ),

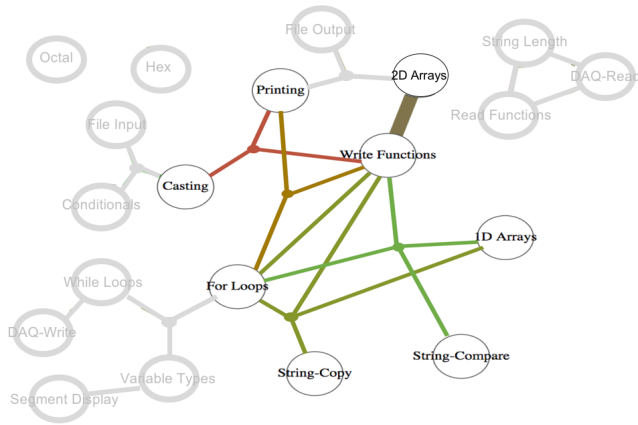
File Output (i.e.,  $c_10$ ), and others. The maximum degree node value is six which occurs with a node representing the Write Functions topic ( $v_{12}$ ). The average degree node value is 2.15 for the visualised model; this study includes all of the hyperedges and nodes, 17 and 20 respectively.

### 6.2.2. Case 2 (filtering by topic)

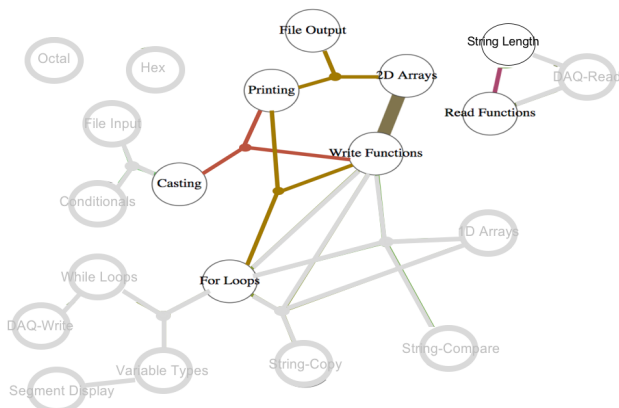
Figure 8 applies the cumulative topic-based  $level_4$  filter on the “Write Functions” concept. This filter will enable instructors to determine the coverage and performance of students on a particular concept. The  $TDM_{MH}$  in the given example demonstrates that the Write Function has been extensively covered in the exam. The performance of students on questions that covered write functions in combination with 1D arrays, String-copy, and String-compare have been quite good; however, their performance on questions that cover Write functions in combination with 2D Arrays, Casting, and Printing has not been as good.

The metrics associated with this case are presented in Table 4. The row corresponding to the visualised model is emphasised with the lightgray colour; the rows with intermediate assessments are not highlighted. The minimum degree node value is one (e.g., 2D Arrays ( $v_{14}$ )). The maximum degree node value is six; this occurs with Write Functions ( $v_{12}$ ). The average degree node value is 2.25 for the visualised model, which is higher than the average value for Case 1 (no filtering). The Write Functions is the topic that has the highest number of appearances in the hyperedges. Filtering on a topic results in a higher average degree value in this study, as there are more connections among the nodes selected; a subset of six hyperedges and eight nodes are considered.





**Figure 8:** Cumulative topic-based  $level_4$  filter on "Write Functions" is selected.



**Figure 9:** Cumulative achievement  $level_3$  filter with " $\leq 60$ " is selected.

### 6.2.3. Case 3 (filtering by achievement)

Figure 9 applies the cumulative achievement  $level_3$  filter with " $\leq 60$ ". This filter enables instructors to determine gaps in students' knowledge. The  $TDM_{MH}$  in the given example demonstrates questions concept combinations that the students have performed poorly on in this exam. Interestingly, all of the hyperedges that include Printing are selected in this filter. This may suggest that there is a general misunderstanding or misconception about how print statements are used.

The metrics associated with this case are presented in Table 4. The row corresponding to the visualised model is emphasised with the lightgray colour; the rows with intermediate assessments are not highlighted. The minimum degree node value is one (e.g., String-Length ( $v_{18}$ )). The maximum degree node value is three (e.g., Write Functions ( $v_{12}$ )). The average degree node value, 1.62, is quite low for the visualised model. Here, the filtering selects a subset of five hyperedges and eight nodes for consideration.

### 6.2.4. Discussion

The three presented cases explore the use of the two-weighted hypergraph based  $TDM_{MH}$  model to generate (RQ1)

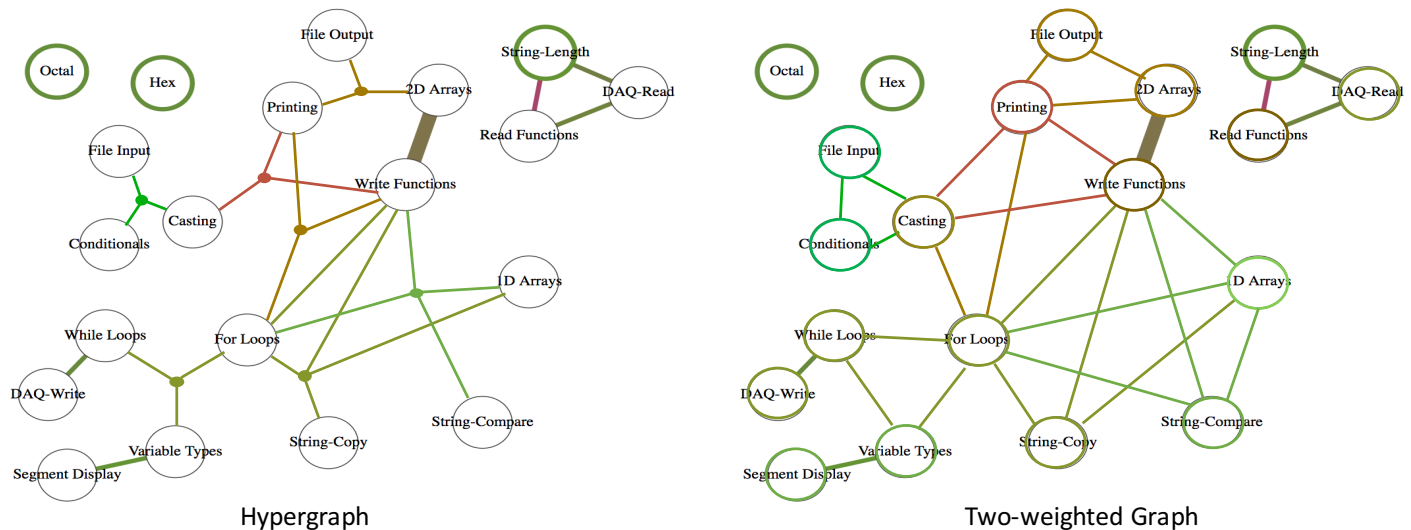
and visualise (RQ2) assessment data from a historical data set. The support for visualising all or part of the model of interest using the concepts of levels and filtering allows instructors to gain additional insights. The preliminary set of metrics are applied to quantify the complexity of the models (RQ3). The average node index value appears to be useful as an indicator of the visual complexity of the model. As demonstrated by Case 2 and Case 3, applying filters may increase or decrease the average node index of the model compared to Case 1 which had no filters. Therefore, the number of nodes in combination with the average node index value may be a more useful indicator of the visual complexity of the model.

## 6.3. Comparing the TDM and $TDM_{MH}$ Approaches (no filtering)

This section compares and contrasts the use of the  $TDM_{MH}$  with the previously proposed TDM approach in this study, Case 4. The discussion is organised with respect to the four RQs. Figure 10 visualises the final exam results for the data set described in Section 6.1; this study extends Case 1 in 6.2.1 (i.e., no filtering) to include the TDM approach.

The two approaches can both be used to represent the assessment data. Essentially, the nodes represent assessment topics and the two-weighted edges/hyperedges represent the coverage and achievements in the assessment data. A colour palette and line widths can visualise the weights in both the  $TDM$  and  $TDM_{MH}$ . The common layout of the nodes has been performed manually, with the goal of reducing the number of crossing edges. This layout, however, does not strongly reflect the organisation of topics within a course. For example, three DAQ topics (DAQ-Read, DAQ-Write, Segment Display) are taught within one module of the course; these topics are scattered in the visualisations.

The fundamental difference between the two models is based on their interpretation of how a question  $i$  that is tagged with a set of topics  $V = v_j, \dots, v_k$  contributes to the coverage and achievement of the models. In the algorithm used for creating the  $TDM_{MH}$  hypergraphs, question  $i$  only contributes to the coverage and achievement of the hyperedge that is associated with  $V$  ( $Achv(V)$ ) (see Section 5.1 for details). In contrast, in the algorithm used for creating the  $TDM$  two-weighted graphs, question  $i$  contributes to the coverage and achievement for all of the edges that include an element of  $V$  (see [14] for details). This approach ensures that a question that is tagged with more than two topics (i.e.,  $|V| > 2$ ), where a particular edge is not associated with  $V$ , contributes to the computation of achievements within the TDM. Let's consider an example using the models that are visualised in Figure 10. In the  $TDM_{MH}$  approach, the question tagged with the three topics of File Input, Conditionals and Casting only contribute to the coverage and achievement of one hyperedge; this is the one associated with these three topics. In the two-weighted TDM, this question contributes to the coverage and achievement of six separate edges: self loops on each of the File Input, Conditionals and Casting topics as well as edges between File Input and Casting, File



**Figure 10:** Comparison of the proposed two-weighted hypergraph ( $TDM_{MH}$ ) and the original two-weighted graph ( $TDM$ ) for visualising the final exam results of an offering of a first-year course on programming and engineering design

input and Conditionals, and Casting and Conditionals.

Based on the difference in the interpretation of how a question contributes to the coverage and achievement of edges, the two models differ quite significantly. A few of these similarities and differences with reference the metrics provided in Table 5 are presented below.

- With respect to the number of nodes, both visualisations are the same, as this is not impacted by the different graph foundations: 20 topics need to be visualised. Filtering options that impact the number of nodes (e.g., multi-level visualisations in the hypergraphs) are not part of the Case 4 study.
- The  $TDM_{MH}$  approach provides explicit information about questions that are tagged with more than two topics, whereas this information is implicitly provided in the two-weighted TDM. For example, as discussed previously, the contribution of a question that is tagged with three topics (e.g., File Input, Conditionals and Casting) is explicitly visualised in the hypergraph approach; however, the contribution of this question is implicitly distributed among six edges in the two-weighted graph approach.
- The  $TDM_{MH}$  can have up to  $2^n$  hyperedges where as the two-weighted graph TDM can only have  $\frac{n(n-1)}{2}$  edges for a graph with  $n$  vertices. This enables  $TDM_{MH}$  to model more complex domains with more than one edge between two vertices. For example there are a total of three edges between For Loops and Write Functions. One based on a question that involved only these two topics; on hyperedge based on a question that involved these two topics as well as 1D arrays and Strong Compare; and a third hyperedge based on a question that involved these two topics as well as 1D Arrays and String Copy.
- The  $TDM_{MH}$  approach is generally sparser compared to the two-weighted TDM approach. This is because in the  $TDM_{MH}$  approach each question can only contribute to one edge. In contrast, in the two-weighted graph approach one question can contribute to multiple edges. For example, in Figure 10, the hypergraph TDM has a total of 17 hyperedges whereas the two-weighted TDM has a total of 49 edges. The average node degree values of the  $TDM$  and  $TDM_{MH}$  are 2.80 and 2.15, respectively, indicate this measure may provide some insight into the visual complexity of the models.
- The main strength of  $TDM_{MH}$  is that it accurately represents the underlying data set. Therefore, the model can be used to determine precisely the achievement and coverage on a set of nodes  $V$ . The main limitation of  $TDM_{MH}$  is in its inference power as it lacks the ability to infer achievements or coverage on the non-complete subsets of  $V$ . As an example, referring to Figure 10,  $TDM_{MH}$  provides accurate information about how For Loops have been covered in combination with other topics, but it cannot infer achievement on For Loops by itself.
- The main limitation of the original TDM approach is that it cannot accurately represent the underlying data set; therefore, the model does not have the capacity to determine precisely the achievement and coverage on a set of nodes  $V$ . The main strength of this model is in its inference power as it can infer achievements or coverage of subset of nodes of size 1 (self loops) or size 2 (general edges). For the same example given before, the original  $TDM$  can be used to infer the achievement and coverage on the topic of For Loops by itself; however, it cannot be used to provide highly

Model	$\delta(G)$ (min)	$\Delta(G)$ (max)	E (avg)	order	size
<i>TDM</i>	2	10	2.80	49	20
<i>TDM<sub>MH</sub></i>	1	6	2.15	17	20

**Table 5**

Case Study: *TDM* and *TDM<sub>MH</sub>* Metric Summaries (no filtering, default cumulative mode)

accurate information about how the topic of For Loops has been covered in combination with other topics.

## 7. Conclusions and Future Work

The preliminary results of modelling and visualising topic-level achievements and coverage in the educational assessment problem using a two-weighted, multilevel, hypergraph-based model (called *TDM<sub>MH</sub>*) are presented. The generation of *TDM<sub>MH</sub>* utilises matrix computations, which makes it scalable and efficient. The visualisation helps to address the complexity of a hypergraph through a flexible, multilevel approach. A set of metrics are employed to quantitatively assess the visual complexity of the hypergraph based visualisations of the models. A case study that illustrates how *TDM<sub>MH</sub>* can be applied to provide insight in the context of a large university course is presented. The case study also compares and contrasts the proposed approach with the original *TDM* model that uses a graph instead of a hypergraph.

There are several limitations in the current work which restrict the generalisability of the results. One of the significant limitations of this study is that *TDM<sub>MH</sub>*s have only been applied to one course from the computer science domain. It is important to acknowledge that these may be hard to comprehend for stakeholders without formal training in algorithmic literacy. A second limitation of the current work is that the case studies only explore the benefits of using *TDM<sub>MH</sub>*s from an instructor's perspective. A comprehensive stakeholder and scenario analysis study is planned for future work, which would investigate the usability of the model across a range of disciplines.

A number of other directions are also planned for the next steps in this research. (1) Adding layout algorithms that optimise the model visualisation; the algorithm may use a range of options such as reducing the number of crossing edges in the graph or grouping topics with respect to instructional design constraints (e.g., groups of related topics covered in a course). (2) Exploring the use of dynamic models to present a series of *TDM<sub>MH</sub>*s that illustrate changes on achievements and coverage over time. (3) Exploring the use of 3D and hierarchical models for visualising a *TDM<sub>MH</sub>*. (4) Exploring additional metrics to improve the quantitative assessment of the *TDM* and *TDM<sub>MH</sub>* models' complexity.

## References

- [1] Albert, D., Nussbaumer, A., Steiner, C., 2010. Towards generic visualisation tools and techniques for adaptive e-learning, in: Proc of the 18th International Conference in Computers in Education, pp. 61–65.
- [2] Argumentation, V., 2003. Software tools for collaborative and educational sense-making. Paul A. Kirschner, Simon J. Buckingham Shum and Chad S. Carr (Eds.) Springer-Verlag: London, 117–135.
- [3] Bodily, R., Kay, J., Alevy, V., Jivet, I., Davis, D., Xhakaj, F., Verbert, K., 2018. Open learner models and learning analytics dashboards: a systematic review, in: Proc of the 8th International Conference on Learning Analytics and Knowledge, pp. 41–50.
- [4] Bogarín, A., Cerezo, R., Romero, C., 2018. A survey on educational process mining. Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery 8.
- [5] Bretto, A., 2013. Hypergraph Theory. Springer.
- [6] Bull, S., Ginon, B., Boscolo, C., Johnson, M., 2016. Introduction of learning visualisations and metacognitive support in a persuadable open learner model, in: Proc of the Sixth International Conference on Learning Analytics & Knowledge, pp. 30–39.
- [7] Bull, S., Kay, J., 2010. Open learner models, in: Advances in Intelligent Tutoring Systems. Springer, pp. 301–322.
- [8] Cooper, K., Khosravi, H., 2018. Graph-based visual topic dependency models: Supporting assessment design and delivery at scale, in: Proc of the 8th International Conference on Learning Analytics and Knowledge, pp. 11–15.
- [9] Dawson, S., Macfadyen, L., Risko, E.F., Foulsham, T., Kingstone, A., 2012. Using technology to encourage self-directed learning: The collaborative lecture annotation system, in: Proc of the conference of the Australian Society for Computers in Tertiary Education.
- [10] Englund, C., Olofsson, A.D., Price, L., 2017. Teaching with technology in higher education: understanding conceptual change and development in practice. Higher Education Research & Development 36, 73–87.
- [11] Glatz, E., Mavromatidis, S., Ager, B., Dimitropoulos, X., 2014. Visualizing big network traffic data using frequent pattern mining and hypergraphs. Computing 96, 27–38.
- [12] Griffin, P., Care, E., 2015. Assessment and Teaching of 21st Century Skills. Springer.
- [13] Hoang, D., Nguyen, B., Tjoa, M., 2012. Dashboard by-example: A hypergraph-based approach to on-demand data warehousing systems, in: Proc of the 2012 IEEE International Conference on Systems, Man, and Cybernetics, pp. 1853–1858.
- [14] Khosravi, H., Cooper, K., 2018. Topic dependency models: Graph-based visual analytics for communicating assessment data. Journal of Learning Analytics 5.
- [15] Klamt, S., Haus, U.U., Theis, F., 2009. Hypergraphs and cellular networks. PLoS computational biology 5, e1000385.
- [16] Leony, D., Pardo, A., de la Fuente Valentin, L., de Castro, D.S., Kloos, C.D., 2012. Glass: a learning analytics visualization tool, in: Proc of the 2nd International Conference on Learning Analytics and Knowledge, pp. 162–163.
- [17] May, M., George, S., Prevot, P., 2011. Travis to enhance online tutoring and learning activities: Real-time visualization of students tracking data. Interactive Technology and Smart Education 8, 52–69.
- [18] Park, Y., Jo, I.H., 2015. Development of the learning analytics dashboard to support students' learning performance. Journal of Universal Computer Science 21, 110.
- [19] Rajaram, G., Karnatak, H.C., Venkatraman, S., Manjula, K.R., Krithivasan, K., 2018. A novel computational knowledge-base framework for visualization and quantification of geospatial metadata in spatial data infrastructures. GeoInformatica 22, 269–305.
- [20] Schwendimann, B.A., Rodriguez-Triana, M.J., Vozniuk, A., Prieto, L.P., Boroujeni, M.S., Holzer, A., Gillet, D., Dillenbourg, P., 2016. Perceiving learning at a glance: A systematic literature review of learning dashboard research. IEEE Transactions on Learning Technologies 10, 30–41.
- [21] Shen, T., Zhang, Z., Chen, Z., Gu, D., Liang, S., Xu, Y., Li, R., Wei, Y., Liu, Z., Yi, Y., Xie, X., 2018. A genome-scale metabolic network alignment method within a hypergraph-based framework using a rotational tensor-vector product. Scientific Reports 8.
- [22] Singh, A., Karayev, S., Gutowski, K., Abbeel, P., 2017. Gradescope: a fast, flexible, and fair system for scalable assessment of handwritten

- work, in: Proc of the 4th Conference on Learning@ Scale, pp. 81–88.
- [23] Tervakari, A.M., Silius, K., Koro, J., Paukkeri, J., Pirttila, O., 2014. Usefulness of information visualizations based on educational data, in: Global Engineering Education Conference, pp. 142–151.
- [24] Voloshin, V., 2009. Introduction to Graph and Hypergraph Theory. Nova Science Publishers.



# Journal of Visual Language and Computing

journal homepage: [www.ksiresearch.org/jvlc/](http://www.ksiresearch.org/jvlc/)

## Metaphorical Parametric Model for Brand Mark Design

Towards a Universal Model of Computational Visual Communication Design

Po Choy Ng<sup>a</sup> and Clifford Sze-Tsan Choy<sup>a</sup>

<sup>a</sup>*School of Design, Hong Kong Polytechnic University, Hong Kong*

### ARTICLE INFO

#### Article History:

Submitted 4.14.2019

Revised 5.20.2019

Second Revision 8.15.2019

Accepted 9.20.2019

#### Keywords:

Visual language

Visual communication design

Conceptual metaphor

Metaphorical expression

Image schema

Visual rhetoric

### ABSTRACT

Visual communication design involves two aspects of visual representation, namely, information and expression. The goal of building a computational model for visual communication design has existed for many years. However, this objective is only partially fulfilled as most of the solutions are focused on the aspect of information and seldom touch upon the expression. Lakoff and Johnson claim that "...metaphor is not just a matter of language, ..., human thought processes are largely metaphorical." Thus, based on the theory of conceptual metaphor rather than language, a computational model for visual communication is proposed. As brand mark design is a challenging problem that demands the representation of both information and expression, the first attempt of the development of this new model is targeted to brand mark design. The assembling of MPM is based on the Metaphorical Relation (MR), which manages the spatial arrangement of its components to signify the intended meanings. All components and relations of the MPM are parametrically controllable for generating infinite visual forms to represent different kinds of information and expression.

© 2019 KSI Research

## 1. Introduction

The idea of formalising visual communication design as a model of visual language has been carried out by art and design practitioners for many years. In general, there are two major groups of practitioners—the informationists Neurath, Bertin, Richards and Engelhardt [1, 2, 3, 4, 5] and the expressionists Kandinsky and Klee [6, 7, 8]. Brinton [1] stated that: "The principles for a grammar of graphics presentation are so simple that a remarkably small number of rules would be sufficient to give a universal language." With the advent of computing, his idea of making a universal visual language for graphic representation has been partially fulfilled. However, most of the previous research works are limited to the presentation of quantitative information such as variable data printing [9], automatic layout [10], font matching [11] or data visualisation [12]. Although there are some attempts to create algorithms for showing expressions in generative art [13,

14] or generative design [15, 16], they are mostly focusing on the synthesis of visual forms but not generate works based on the creative concepts or the semantic meanings. Yet, there is no well-established universal model to represent both information and expression.

As an attempt to create a universal computational model which can generate designs to present information as well as expression, this paper proposes the Metaphorical Parametric Model (MPM) for generating visual communication design. Since brand mark design is related to both information and expression and has been considered as "one of the most difficult to perfect [17]", it is chosen as the topic for an initial attempt.

This paper is organised as follows: Section 2 outlines the theoretical framework of the MPM. Section 3 reports the corpus analysis of 8000 brand marks and explains how the findings are utilised for the development of the model proposed. Section 4 presents the model of Metaphorical Parametric Model (MPM) and discusses how to represent visual objects for making metaphorical expressions. Section 5 introduces the concepts of Metaphorical Relations (MR) and Metaphorical Expressions (ME).

<sup>a</sup>Corresponding author

Email address: [ng.pochoy@connect.polyu.hk](mailto:ng.pochoy@connect.polyu.hk)

ORCID: 0000-0003-2863-469X

Section 6 demonstrates how to use the MPM to recreate an existing brand mark. It also discusses how to change the MR or the components of the MR to generate new designs. Section 7 further discusses the concept of MPM and related issues. Section 8 concludes this paper and suggests future developments.

## 2. Background

In the following paragraphs, the theoretical framework of the MPM is introduced and described. Five principal theories which are essential to the construction of this model are introduced, including the conceptual metaphor, image schema, primary metaphor, visual rhetoric, and symmetric object alignment.

### 2.1 Conceptual Metaphor—the Foundation of Language

In their seminal book “Metaphors We Live By”, Lakoff and Johnson state that, “...metaphor is not just a matter of language, ... human thought processes are largely metaphorical.” [18, p.3] Based on such premise and many examples elucidate in the book, they further confirm that “Our ordinary conceptual system, in terms of which we both think and act, is fundamentally metaphorical in nature [18, p. 153].” They define that conceptual metaphors are different from the linguistic ones but specific cognitive concepts embedded in our bodily experience. Based on the nature of the source domain of the mapping, they differentiate three types of conceptual metaphors, namely, ontological, structural, and orientational metaphors.

1. **Structural metaphor:** This type of metaphor is grounded in systematic correlation within our experience in which the concept is expressed with another different structured, clearly defined or well familiar concept. For example, the same structure of war can be used as the metaphorical expressions “ARGUMENT IS WAR”. Therefore, the concepts such as attack, indefensible, strategy, win, gain, ground and related concepts will be used to form a systematic way of talking about the battling aspects of arguing.
2. **Oriental metaphors:** In this type of metaphor, the mapping is focused on adopting the structure of one concept rather than a whole system of concepts of the corresponding object. Most of these are related to spatial orientations such as up-down, in-out, front-back, on-off, deep-shallow, or central-peripheral related to the physical environment. Hence, the orientational metaphor represents a spatial orientation concept such as HAPPY IS UP and results as an expression like “I’m feeling up today.”
3. **Ontological metaphors:** The concept of this type of metaphor is based on the mapping of a source in

the physical world to the target which can be an activity, emotion or idea. Take the “INFLATION IS AN ENTITY” as an example, by using “inflation” as an entity, it allows us to quantify, identify it, take it as a cause and act upon it as the “Inflation is lowering our standard of living” or “Inflation is hacking us into a corner.”

In summary, the conceptual metaphor theory provides at least two fundamental concepts of the MPM proposed here. First, it states that our conceptual systems are based on conceptual metaphors rather than languages. Second, metaphorical expressions are constructed with a system through the mapping of meanings from the source concepts to the target concepts.

### 2.2 Image Schemas — the Recurring Patterns of Embodied Experience that Structure Our Concepts.

The term “image-schema” is only briefly mentioned in “Metaphors We Live By” [18] and the more in-depth discussions are found in the later publications [19, 20, 21, 22, 23]. Johnson [20] explains that the idea of a close connection between perceptual and semantic structure is not new and have been discussed by scholars such as Dewey, Merleau-Ponty and Churchland. By using the formation of the CENTER-PERIPHERY image schema as an example, he describes that once a baby is born, it will develop its visual focusing abilities gradually. It learns to focus on an area and highlight a figure against a background that fades out into an indefinite perceptual horizon on its periphery. In order to survive, animals must acquire such focusing skill and know that the centre part is more important than its periphery. Such perceptual experience always reveals the same recurring schematic structure which has a focal centre surrounded by a horizon that fades off. The CENTER-PERIPHERY structure, as well as many other image schemas, is not restricted to vision but relevant to all our senses. When our focus is confined to the voice on the telephone, our attention will be diverted away from the humming noise from the computer nearby. On the other hand, we can shift the focus and take the voice as periphery and make the humming noise as the centre.

The CENTER-PERIPHERY schema is critical to our fundamental preconceptual understanding of our world. It is a continuous structure or pattern through which we organise and unify our perceptions, motor programs, spatial orientation temporal sequences and so forth. An image schema is not an abstract, finite proposition, concept or concrete image but an evolving pattern of our imaginatively structured experience. It is a recurrent pattern which comes from our bodily (sensorimotor) experience of the world. Through different types of image schemas, we can connect different perceptual phenomenon

and our understanding of more abstract domains through imaginative metaphoric and metonymic projections.

### 2.3 Spatial Primitives—the Prelinguistic Conceptual Building Blocks

Through a long period of study of the development of the conceptual system of baby, Mandler [24, 25, 26, 27] makes noteworthy interpretation of the principle of image schema from the perspectives of developmental psychologist rather than those of the cognitive linguistics. She argues that concept formation is top-down in nature and organised as a hierarchical object conceptual system. The early conceptual categorization of objects are not refined and it is difficult for infants or even adults to define conceptual categorization of objects such as dog or vehicle. Thus, she proposes the attentional processes for the formation of the first concepts, by finding patterns in perceptual data and redescribing (reformatting) them. She states that “Initially, redescription is done via the conceptual primitives that are the vocabulary of the mechanism (called Perceptual Meaning Analysis) that does the reformatting. This mechanism is an attentive process that extracts spatial information from perceptual displays and while retaining its analogue character recodes it into a skeletal (somewhat topological-like) form. For example, the infant attends to an apple being put into a bowl, but Perceptual Meaning Analysis outputs something like thing into a container. Redescriptions like this enable the concept formation that makes conscious thought possible.” [24, p. 212]

Mandler further highlights that the early global concepts are mainly related to spatial information, especially movement in space which she calls “conceptual primitives”. This spatial information is primarily related to concepts of both objects and events which is different from the other non-perceptual schema such as force schemas. So, the “container” primitive has a bounded space with an inside and an outside. The “linked paths” primitives consist of objects moving together collectively. For Mandler, these primitives are not merely copies of sensory impressions but require descriptive processing before the perceptual encoding is transformed as the conceptual descriptions. Based on the principle just mentioned, she considers that spatial image-schemas should be differentiated from the other conceptual representations.

Though the term “prelinguistic primitive” [25] has been used in the early article in her discussion of the human conceptual system and image schema, in the later one, Mandler [27, pp. 510–511] suggests that “Spatial primitives are the first conceptual building blocks, image schemas are simple spatial stories built from them, and schematic integrations use the first two types to build concepts that include non-spatial elements. These three kinds of structure and some others as well have often come under the umbrella term of ‘image schemas’.”

She explains that spatial and motion primitives are essential to the first image schema. It lets the infant simplify the situations and find a way to understand and remember the events; so, they can form new conceptual structures by unifying different sensations and feelings of force. However, Mandler suggests that the term “image schema” is different from these new structure and should be limited to imageable information, which constitutes the foundations of the conceptual system. The term ‘schematic integrations’ should be used to represent the other structures that include internal feelings of force, the other emotions and sensory information. So, starting with spatial primitives, infants can construct image schema to run mental simulations of spatial events. Then, based on these simulations, they can connect disparate experience and integrate them into new conceptual structures and create schematic integrations which eventually composed with more non-spatial elements. When the non-spatial, non-imageable information is incorporated into the conceptual system, she considers that these structures are no longer image schemas anymore but are blends that integrate non-spatial components into spatial events. They also allow the infants, for the first time, to conceptualise non-spatial experiences as meaningful aspects of organised spatial stories.

Finally, Mandler concludes that “All three structures of the schematisation process (spatial primitives, image schemas, and schematic integrations) are powerful tools for reducing the immense variety of perceptions and sensations to discrete kinds of event that the human mind can easily represent.” [26, p. 19]

### 2.4 Metaphor and Image Schema in Information Visualization

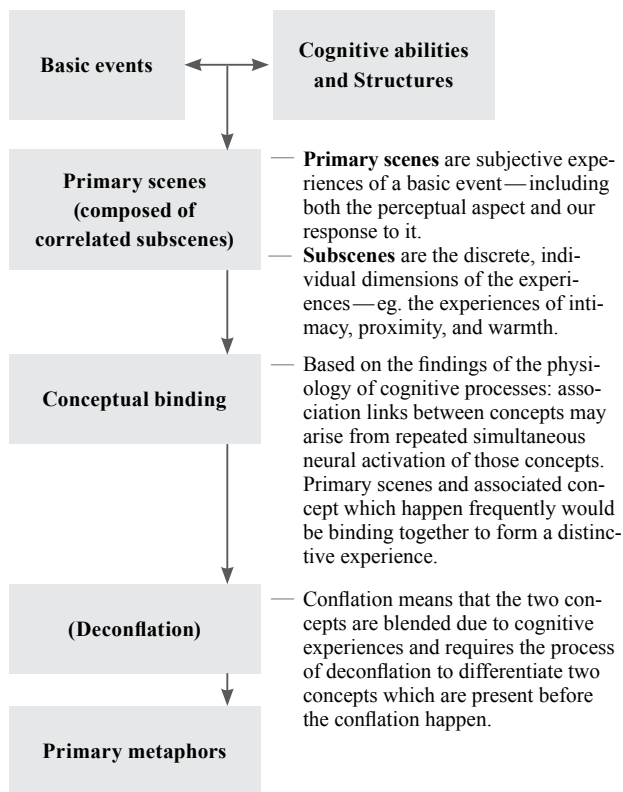
Image schema is now well accepted and the concept has been confirmed by research in cognitive sciences and developmental psychology [21, 22, 23]. Risch [28] lists as many as sixty-six image schemas and explains that many information graphics are based on the principle of image schema whereas people’s early perceptual experience of spatial relations such as “containment”, “path-following” and “object dynamics” become generalised and formalised subconsciously. He makes a comprehensive discussion on the use of image schema and metaphor for the visualisation of information. He describes that there are two distinctive types of graphics namely analogical graphics and metaphorical graphics:

1. **Analogical graphics:** It depicts the inherent spatial relations from the source to the target of the same domain such as the mapping of the length of line to a dimension.
2. **Metaphorical graphics:** It expresses non-spatial concepts in spatial terms through the functional alignment with image schemas.

He proposes [28, p.2] that ‘...the standard inventory of image schemas derived from linguistic and cognitive studies can serve as the basis for developing a kind of visualization “grammar.” Such a grammar would employ graphical analogues of image schematic patterns as syntactic elements that relate concepts expressed using conventional signifiers such as text, colour, and symbology.’

## 2.5 Primary Metaphor—the Basic Components of Complex Metaphor

Grady [29] suspects that if a metaphor is a conceptual rather than a linguistic phenomenon, it should be revealed in the other cognitive behaviours which have been supported by various researches [30, 31]. To investigate the related questions, he proposes a new approach to conceptual metaphor as Figure 1. For Grady, primary metaphors refer to bindings of this type of distinct concepts emerged from primary scenes as described in the figure whereas the conceptual mapping of this type constitutes the patterns of metaphoric language.



**Figure 1: From basic events to primary metaphors with descriptions [29]**

Grady argues that primary source concepts should be universal with basic properties, relations and activities came from our experience of the world. A typical characteristic of the primary source concepts is that they have

image content represented as a mental image which is multimodal and not limited to vision. Contrarily, primary target concepts lack image content or sensory content; they are not the direct perceptions of the world but responses. Target concepts like “similarity”, “happen”, and “difficult” are merely universal cognitive operations and structures.

Therefore, based on the principle mentioned above, Grady specifies the primary metaphor as follows:

Metaphor:	SIMILARITY IS PROXIMITY
Motivation:	The tendency for similar objects to be clustered together. And/or the tendency for adjacent objects to appear similar because of they are in similar conditions.
Examples:	These fabrics aren’t quite the same but they’re close. His singing style is miles away from mine.

Furthermore, he also compiles a list of primary metaphor and categorised as five groups as shown in Table 1.

**Table 1: Categories of Primary Metaphor**

Category	Count
Atemporal relations	19
Quantity and degree	4
Time, action and event structure	28
Affect, evaluation and social relations	28
Thought and consciousness	21

Once Grady has proposed the concept of primary metaphor, he argues that there are many other more complicated conceptual structures which are based on primary metaphors and can be constructed through the process of conceptual binding. For example, the complex metaphor “VIALE COMPLEX ENTITY IS ERECT PHYSICAL STRUCTURE” can be deconstructed as “VIABILITY IS ERECTNESS,” grounded in our experiences with objects where erectness correlates with functionality and health. Furthermore, “ORGANIZATION IS PHYSICAL STRUCTURE” is grounded in the manipulation of complex objects, which correlates with the formation of mental representation of logical relationships among parts. Both of these basic metaphors come from our experience with physical objects in the world and each of the two primary metaphors focuses to a specific part of our experience with objects while they could happen concurrently. Thus, a viable entity stands up firmly as a building which is an erect physical structure and constitutes the compound metaphor “VIALE COMPLEX ENTITY IS A BUILDING.”



Grady states that the theory of primary metaphor simplifies the construction of metaphor by distributing the process to a set of fundamental modules rather than dealing with many complicated matters at once. By deconstructing a complex metaphor into primary metaphors, it becomes easier to understand the underlying concept of each metaphor. At the same token, it provides a set of basic components which could facilitate the construction of complex metaphor for different purposes. Such a principle not only valid for the linguistic metaphor but also a useful strategy to keep visual metaphor universal if the underlying principles could be utilised by users of different backgrounds.

## 2.6 Integrated Theory of Primary Metaphor

After more scholars have investigated on the studies related to conceptual metaphor, Lakoff and Johnson [19] propose the Integrated Theory of Primary Metaphor which provides some complementary principles to support the understanding and utilisation of the theories of conceptual metaphor and primary metaphor. This integrated theory consists of four parts, namely, Johnson's theory of conflation in the course of learning, Grady's theory of primary metaphor, Narayanan's neural theory of metaphor, and Fauconnier and Turner's theory of conceptual blending and each of these theories are described as follows:

### 1. Johnson's theory of conflation in the course of learning:

Young children may not be able to distinguish the subjective (non-sensorimotor) experiences and judgments and sensorimotor experiences and regularly conflated them if they occur together at once for a time.

An infant may make a connection between the subjective experience of affection with the sensory experience of warmth during childhood. The association of two different domains are automatically and repeatedly connected and eventually leads to conflation. However, the children can separate the domain during a period of differentiation while the cross-domain associations are still maintained. This persisted mapping of conceptual metaphor leads the same infant to associate affection with warmth and making the metaphorical expression "a warm smile" or the sense of being help closely as "a close friend" later in life.

### 2. Grady's theory of primary metaphor:

According to Grady, primary metaphors are the "atomic" components which can be used to construct complex metaphors like "molecular" structures. Through the recurring process of conflation which makes cross-domain association occurred and gives rise to primary metaphor, universal early experience of conflation eventually lead to conventional conceptual metaphor.

### 3. Narayanan's neural theory of metaphor:

During the process of conflation, the associations are realised

neurally in simultaneous activation. A neural network that defines the conceptual domains will result in permanent neural connections within the brain.

### 4. Fauconnier and Turner's theory of conceptual blending:

In certain cases, different conceptual domains can be activated together and form cross-domain connections. Such "conceptual blends" can be original or conventional while two or more primary metaphors can be combined to make larger complex metaphors through the mechanism of conventional blends.

Lakoff and Johnson summarise that there are hundreds of primary metaphors come from our experiences repeatedly. Through conflation and deconflation, people formulate a rich inferential structure and shape the neural network they have. Narayanan's neural theory of metaphor explains how primary metaphors are formed and operates due to our embodied experience of the world. They make many sensorimotor experiences become the resources for metaphorical extension and functions as the components for building complex metaphor.

## 2.7 Different Opinions on the Conceptual Metaphor Theory

Throughout many years of its development, the Conceptual Metaphor Theory (CMT) has received different criticism and gone through different levels of refinements. Szwedek [32] criticised that structural and orientational metaphors are based on existing objects, thus metaphors are ontological. All non-physical phenomena such as social, political phenomenon as well as events, actions, activities and states have to be objectified first before they can be assigned structure and orientation. On the other hand, Hernández and Pérez, [33] argued that metaphor is not fundamentally ontological as explained by Lakoff and Johnson [18], image schema and primary metaphor are more fundamental.

In a special issue of the retrospective of conceptual metaphor, Fusaroli and Morgagni [34] mentioned that some authors such as Leezenberg [35] suggested that embodied concepts do not come from an individual's experience only but largely shaped by the social, cultural and linguistic practice. Regarding the influence from society and culture, Brandt [36], Deignan and Cameron [37] pointed out that the conceptual metaphor is not all dependent on bodily experience but conditions affected by the context. All these lead to the direction that the CMT should concern with the conceptual patterns affected by cultural practices and contextual issues. Based on studies of different researchers, plenty of evidence [38, 39, 40] reveals that people understand certain domains through the other domains more often than through the language. People talk and think about time in term of space and motion but not from the opposite direction.

## 2.8 Visual Rhetoric—the Means to Manipulate Visual Presentation

Marcus [41] states that: “Visual rhetoric refers to those devices which modify or affect a viewer’s conceptions and attitudes toward a visual presentation.” Among the early studies of visual rhetoric, Bonsiepe [42] and Durand [43, 44] have provided concrete guidelines and trigger the awareness of visual rhetoric among the practitioners in the design and advertisement industries. Through the examinations of different taxonomies of visual rhetoric [45] proposed by different researches, the framework offered by Durand has been identified as the most applicable one for the objective of the current study. The matrix of “Classification of Figure” presented by Durand as shown in Table 2 provides a rather structural approach to construct different types of visual rhetorical figures. This matrix is organised with the “rhetorical operations” assigned to the horizontal axis and the semantic relations assigned to the vertical axis. Then the cell corresponds to the intersection between the horizontal row and vertical column indicates the rhetorical figure resulted from the interaction of the rhetorical operations and the semantic relations. The concepts of the rhetorical operation listed on the horizontal axis are as follows:

- Addition: one or more elements are used in the composition
- Suppression: one or more elements are deleted from the composition
- Substitution: it can be formed by a suppression followed by an addition. So part of the components are removed and the other components will be added to the composition.
- Exchange: it consists of two reciprocal substitutions in which the two substitutions are inversely related.

Regarding the relations between objects, Durand has distinguished the objects by the nature of either “similarity” and “dissimilarity”. Between two elements, he indicates different relations between form and content in Table 2. The definitions of the label in the matrix are as follows:

- Identity: only relations of identity
- Similarity: at least one relation of identity and one of difference
- Opposition: at least one relation of opposition
- Difference: only relations of difference
- Paradox and double meaning are mix relations of content and relations of form. The relations of content are first seen as homologous to the relations of forms; however, comprehensive examination reverses this interpretation.

Based on the interactions of different relations, Durand proposed the complete set of rhetorical operations as shown in Table 2. This schematic approach provides an effective way to construct visual rhetoric that seems to match with the traditional rhetorical figures as listed in the matrix. Such an approach also provides a useful exemplar to construct a system of visual rhetorical operations to work with the computational model proposed in this study. However, the relations of forms and relations of elements listed in Table 3 are much more complicated than the descriptions in his paper as a form could refer to either the physical form or the linguistic form. Furthermore, a form or element has different attributes such as material, colour, texture, size, and shape which can affect the meanings as well. After all, when the verbal rhetoric is transferred to visual rhetoric, the limitation of the current understanding of visual rhetoric becomes obvious and reflects that additional knowledge beyond those of verbal rhetoric is required.

**Table 2:** Classification of Figure [43]

Relations between Elements	Rhetorical Operations			
	<i>Addition</i>	<i>Suppression</i>	<i>Substitution</i>	<i>Exchange</i>
<i>Identity</i>	Repetition	Ellipsis	Hyperbole	Inversion
<i>Similarity</i> <i>of form</i> <i>of content</i>	Rhyme		Allusion	Hendiadys
	Simile	Circumlocution	Metaphor	Homology
<i>Difference</i>	Accumulation	Suspension	Metonymy	Asyndeton
<i>Opposition</i> <i>of form</i> <i>of content</i>	Zeugma	Dubitation	Periphrasis	Anacoluthon
	Antithesis	Reticence	Euphemism	Chiasmus
<i>False homologies</i> <i>Ambiguity</i> <i>Paradox</i>	Antanaclasis	Tautology	Pun	Antimetabole
	Paradox	Preterition	Antiphrasis	Antilogy

**Table 3:** Interaction of Relations of Form and Content [43]

Relations between Elements	Relations of form		
	<i>Identity</i>	<i>Difference</i>	<i>Opposition</i>
<i>Identity</i>	identity	similarity of content	paradox
<i>Difference</i>	similarity of form	difference	opposition of form
<i>Opposition</i>	double meaning	opposition of content	homologous opposition

When Durand [43, 44] first introduced the theory of visual rhetoric in 1970, the theory of conceptual metaphor had not emerged yet. So during that period, the use of visual rhetorical figures was mainly based on the intuition of advertising people who select objects to symbolise the meaning that they would intend to express. Such know-how based on intuition and common practice prevail in the advertising and design industries for many years. Most of the time, the practitioners may consider the making of visual rhetoric as a kind of creative act emerging from their imaginations without noticing the underlying theories. In the current study, through the integration of the theory of Durand and the other theories introduced in this section, it reflects that conceptual metaphor and rhetoric are crucial for communication. However, visual rhetoric is different from verbal rhetoric and a different theory should be developed for visual communication and the related issues will be discussed in Section 7.







## 2.9 From Conceptual Metaphor Theory to Visual Rhetoric and Symmetric Object Alignment

By the time when the researchers in the marketing and linguistic studies notice the connection between the visual rhetoric and conceptual metaphor, the component which binds all the theories mentioned above finally emerged. Schilperoord, Maes, and Ferdinandusse [46] describe that people tend to organise objects as cohesive configurations according to the century-old gestalt theory [47]. Hence, designers often deliberately utilise these grouping principles for their aesthetic and expressive purposes. The making of visual expressions through the manipulation of metaphorical relations and mapping of concepts between visual objects is defined as symmetric object alignment (SOA).

Through the incorporation of the theory of primary metaphor and SOA in visual representations, Ortiz [48] explains that the formal configuration of visual metaphor in pictorial advertising could be distinguished as three types as shown in Table 4. The first type can be regarded as a fusion in which different objects are blended together

and appear as a single hybrid image. The second type appears as an omission in which only one single object is shown while the other is hidden and suggested by the context. The third one can be considered as a juxtaposition which shows at least two objects at once.

**Table 4:** Three Basic Types of Visual Metaphors

Type	Logo Sample	
Fusion: Blending of different objects in a hybrid image		
Omission: Only one object appears and the other is suggested through the context	 Food Company	 Child, Infant's Product Company
Juxtaposition: Both objects appear simultaneously		

Concerning the nature between the objects chosen for the alignment, they could be categorised as three different types including the alignment of different objects, the alignment of incompatible objects, and the alignment of identical objects.

Based on various combinations of the formal structures and the selected objects, Ortiz has analyzed a series of advertisement and identified five types of primary metaphors generally found in advertisements which include:

- “SIMILARITY IS ALIGNMENT”
- “SIMILARITY IS PROXIMITY”
- “THE NATURE OF AN ENTITY IS ITS SHAPE”, “CONSTITUENTS ARE CONTENTS”
- “CATEGORIES/SETS ARE BOUNDED SPATIAL REGIONS”

Ortiz summaries that the nature of SOA could be either metaphorical or literal. The SOA is metaphorical when the object conveys the idea of similarity and belongs to the same category; also, when the alignment, proximity, shape, inclusion, and space indicate the presence of primary metaphors which are grounded by experiences. The SOA is literal if the element presented does not seem similar but merely shows objects different from the others.

The examples provided by Ortiz show that SOA can be utilised in various kinds of visual rhetoric to construct visual metaphor for making metaphorical expressions. Although her study focuses on the discussion of

advertisings and photographic images, the same principle of SOA could be found in brand mark design as reflected in the results of the corpus analysis described in next section. However, as the structure of visuals is different from text, different approaches of using SOA are necessary for the development of visual languages and the related issues will be further discussed in the later sections.

### 3. Corpus Analysis of Brand Marks

Pragglejaz Group [49] proposes the “metaphor identification procedure” (MIP) for identifying the metaphorically used words in discourse. They describe that different researchers will handle the material and process differently based on their purpose whereas some may do a large corpus while the experimental psycholinguists may only verify or identify the target words or discourse in small sources.

For this study, due to the nature of the corpus is visual form rather than text, the corpus analysis demands the judgement of human other than the processing of computers. Thus, under the constraint of time and resource, a flexible and pragmatic approach is adopted. The process focuses on two key issues including:

- The identification of the basic components and structures of brand mark designs.
- The identification of the use of the conceptual metaphor and visual rhetoric in the brand mark design.

Two sets of books [50, 51] with collections of approximately 8000 brand marks were chosen. Though one of the publishers of these books is from Japan while the other is from South Korea, the samples in the collections are selected from worldwide. In general, these type of books are used by designers for reference, most of the samples in the collections were well selected by editor with design expertise. Hence, the samples in these collections consist of the choice of two different groups of experts who would screen for professional-grade design samples. Those digital files distributed with the book is another crucial factor to consider the choice of these collections. Thus, all the brand marks can be loaded into the database for the corpus analysis without extra preparation work beforehand.

Two fresh graduate design assistants had been assigned to identify the shapes and objects found in each sample. All the shapes found in the samples were deconstructed and identified as one or more of the shapes listed in the codebook which was compiled according to literature reviews. During the process of the corpus analysis, if any shape extracted from the mark could not be classified as a shape listed in the codebook, the new item would be added to the codebook until all the unique shapes were recorded. The objective of this task was to identify the essential set of shapes required for brand mark design.

Other than the basic geometric shapes, the assistants had to identify and record the objects represented by these basic shapes.

Here, the basic shapes are defined as the simple, primitive forms found in brand marks, they may or may not resemble any object found in reality. However, these forms mostly appear as geometric shapes or distinctive shapes which can be represented by equations. These objects can be either physical entities or graphic representations; also, they can be either natural or manmade objects. Thus, two small circles on a circle might signify a human face as Figure 2 when they are reading as a single object. On the other hand, the same configuration can be considered as two small circles overlay on the front layer of a larger circle if they are perceived as separated shapes. In both of these cases, both the circle and human face will be an entry as shape and object respectively.

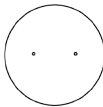


Figure 2: Two dots enclosed in a circle, forming a human face

From the result of the analysis, certain distinctive patterns are noticeable and part of the findings are listed and discussed as follows:

Table 5: Summary of the Brand Mark Analysis

Item	Count
Total number of brand marks analysed	8095
Total Object count	23738
Unique object identified	1538
Geometric Shapes	11784 (49.65%)
Organic Shapes	1195 (50.35%)

As Table 5 shows above, nearly half of the shapes in the corpus are abstract forms and mostly can be considered as geometric shapes. These brand marks may consist of one or more components which may appear as abstract forms, iconic objects or a combination of both. The findings reflect that both geometric and iconic shapes are essential for brand mark design. Since geometric shapes do not depict specific objects but are usually perceived as abstract forms, they can be used as components of design to represent a wide range of objects or topics. In many cases, it is the composition and arrangement of these objects that constitute the meaning of the brand marks. The figure listed above indicates that without using complex forms or many diversified objects, it is possible to generate a large variety of brand marks in the collections with



a small set of objects using geometric shapes.

Since the top hundred of the most used objects count for 67% of the items in the corpus, the corpus analysis has focused on this set instead of processing thousands of objects for efficiency. In the second phase, these marks have gone through further analysis for identifying their synthetic and semantic structures. As the meanings of visual objects are affected by the conditions, contexts and cultures, the annotations done by the assistants might not be precise and comprehensive due to their experience and background. However, because of this analysis is not aimed for making an exhaustive study of the corpus but rather a strategy to identify how meanings are signified in brand marks regarding the theory of conceptual metaphor and visual rhetoric. So, even though this corpus analysis has its limitations, it provides the idea and knowledge to build computational models for making visual expressions.

The result of the corpus analysis indicates that the objects used in the brand mark designs consist of either iconic or abstract forms. These designs may appear as a single or more items in which several basic components could be combined as an abstract or iconic object. Thus, based on the background theories described in the earlier sections and the result of the analysis just mentioned, the current study proposes a new model to generate brand mark designs.

#### 4. Introduction of the Metaphorical Parametric Model (MPM)

Metaphorical Parametric Model (MPM) is a type of two-dimensional model designed for making visual metaphorical expressions based on conceptual metaphor theory and visual rhetoric. Instead of representing the synthetic structure of a specific object, it represents the semantic structure and the semantic relations in between. Moreover, the synthetic process of MPM is guided by its semantic components. Thus, through different combinations of parametric settings of the MPM, different forms could be generated to signify diversified meanings ranged from literal representation to metaphorical expression.

##### 4.1 MPM Types

MPM consists of two different types of basic entities namely element and compound. An element is a single, simple and distinctive visual object while a compound is composed of two or more components which could be either an element or another compound.

##### 4.2 Elements and Metaphorical Properties







Based on the analysis of the samples, approximately sixty primary components are encoded as the “elements”

of the MPM currently. An element is the most common and essential form found in a specific domain such as the landmark in the current study. It is a simple, abstract, and geometric-like two-dimensional object generated by certain geometric logic and equations while each form is controlled by one or more parameters. Elements can be classified as two types, namely line and shape. The major difference between them is that a line is defined as an open path without an enclosed area. However, both types could suggest different expressions according to the art and design theories introduced by Kandinsky and Klee [6, 7, 8].

Among all graphic forms, both lines and shapes could trigger cognitive responses. Through experiments with the imaging from fMRI, Larson, Aronoff, Sarinopoulos and Zhu [52] recognise that the data support the notion that visual threat can be triggered by a simple downward V-shape with no reference of the other contextual or affective cues. Besides, the tests of the association of meaning between the downward-pointing and upward-pointing triangles also support the hypothesis that simple geometric forms could convey emotions. Thus, the results support that our brains could detect the stimulus of the geometric shapes with much less stimulus information than the earlier studies of these authors [53].

Regarding the semantics of lines and shapes, findings done by researchers such as the studies of lines by Ibáñez and Delgado-Mata [54] and the studies on shape by Blazhenkova and Kumar [55] could provide guiding principles to set the parametric settings of the MPM for signifying the intended expressions. Thus, as indicated in Table 6, with different parameters, a single element such as the Archimedean Spiral or Trapezium could suggest a multitude of expressions as listed.

**Table 6: Generated Forms of Various MPM and Their Possible Metaphorical Expressions**

Archimedean Spiral					
Regular		Offsetted		Outward	
					
					
Possible Metaphorical Expressions (Respectively)					
regular	harmony	graceful	living	springy	dynamic
hypnotic	gentle	organic	biological	cheerful	changing
cryptic	intricate	nature	structural	extroverted	unexpected
mysterious	infinity	exquisite	dynamic	excited	energetic

Trapezium					
Asymmetric		Large width/ high ratio		Complex	
Possible Metaphorical Expressions (Respectively)					
free-spirited	distinct	structured	precise	opposing	focused
unrestricted	individual	geometric	absolute	conflicting	against
open	intentional	dignified	strong	complicated	difficult
independent	anarchic	historic	bold	contradict	agitated

### 4.3 Compounds

A “compound” of MPM is a two-dimensional visual object formed by two or more components assembled with specific relations. The component of a compound can be either an element or another compound. A compound only provides two or more semantic placeholders for holding objects; it has no definite synthetic form other than connected together by a set of conceptual relations controllable by parameters. In general, a set of default parametric settings which matches with the general knowledge of most of the people will be assigned to the initial form of the compound.

When different parameters or objects are assigned to the compound, the results could be quite distinct from the default settings. Figure 3 is the conceptual diagram of the compound “sun”. and the top right version in Figure 4 is the “sun” generated with the default settings. With different parameters used in the other versions, even without changing the objects assigned to the placeholder, the other variants in Figure 4 become quite departed from the default version.

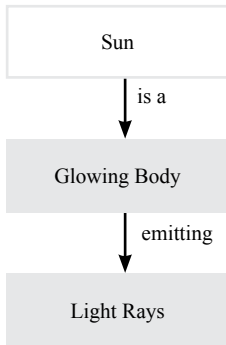


Figure 3: The conceptual structure of compound “Sun”

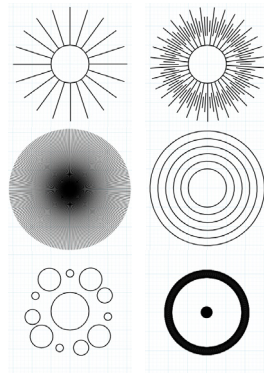


Figure 4: Sun generated with different parameters

As the components and the parameters in the compound “sun” can be switched according to the intended semantic meaning, it can be used to construct different design as shown in Figure 5 and 6 in which the concept of the sun is still retained but the visual forms and expressions are different.



Figure 5: Existing brand marks recreated using compounds “sun” on the right, original design on the left



Figure 6: The other existing brand marks that could be represented by the compound “sun”

### 4.4 Deconstruction and Construction of Compounds

To create a compound, the target object should be deconstructed as the relevant units and reconstructed as a conceptual structure matched with the general knowledge of most of the people or the target audience. Depending on the background, cultures or shared experiences, different people may have different concepts of the structure of the object. In additions, as the focuses or viewing angles towards the same object could be coming from diverse perspectives, an object could be deconstructed and reconstructed in many different ways. For example, a bird is a simple object, but it is unlikely to use one 2D model for all as a bird can stand still or fly; it could be viewed from the bottom, the side or from the top. Unlike 3D modelling, it is necessary to build various 2D models to anticipate different conceptions or situations. Carlier, Leonard, Hahmann, Morin, and Collins [56] conducted a study to collect data about how people decompose the structure of shapes. Their investigation covers 1200 shapes in 70 classes; it has at least 24 annotations per shape and a total of 41,953 annotations provided by 2861 participants. The result reflects that there are many ways to decompose shapes into diverse configurations according to different perceptions. Therefore, many compounds can be created to represent different perceptual models of the same object.

## 5. Metaphorical Relations, Relational Operations and Metaphorical Expressions

Unlike the other models, the MPM only provides placeholders for assembling objects according to conceptual relations which are eventually represented as spatial relations among the components of the model. Although many researchers have made tremendous efforts on the study of mereotopological relations, the focuses are mostly on the studies of physical conditions or spatial relations for engineering, mathematical or scientific purpose. To build a model to signify different metaphorical expressions, the metaphorical relation (MR) is proposed. For constructing the specific spatial relations to signify the MR, several Relational Operations (RO) are encoded in the MPM. The detailed discussion of the features and functionality of MR and RO will be discussed in the following sections.

### 5.1 Relational Operation (RO)

Relational operation (RO) is the operation which arranges the position of objects according to the specific spatial relations. It could arrange the objects according to the concepts of different MRs and then signifying different spatial relationships or metaphorical expressions.

Currently, fourteen ROs which are essential for the arrangement of the components are defined in the MPM. These relations include: alignTo, topOf, bottomOf, rightOf, leftOf, rotateAroundPoint, parallelTo, perpendicularTo, scaleBy, matchOneSide, matchExactSize, inside, meetAt and tipOf.

### 5.2 Metaphorical Relations and Metaphorical Expressions

A large proportion of conceptual metaphors belong to the orientational metaphor type are associated with spatial relations. While mathematicians and researchers have investigated the related theories for many years [57, 58], their efforts seldom come across the notice of visual communication designers. Nonetheless, the diagrammatic renditions of their mathematic theories provide great resources to signify the spatial relations between objects for making innumerable metaphorical expressions in different circumstances. These spatial relations, which could be considered as different kinds of image schemas, have been used to signify different meanings in the pictogram of Chinese oracle-bone inscriptions in ancient. Two distinctive examples in Figure 7 show the principle of image schemas by positioning the shorter strokes in two different relative positions and suggest the opposite meanings.

In the early stage of this study, the purpose of RO of the MPM was mainly for the synthesis of the visual



**Figure 7:** Ancient Chinese characters based on pictogram for “up” (left group) and “down” (right group) and their modern counterparts

composition; it is aimed at managing the spatial relations between different components in the MPM through geometric transformations. However, in the later stage, it has been found that these spatial relations could be constructed according to the principle of image schema to signify specific information and expression metaphorically. Therefore, other than organizing the composition of designs, the RO is also utilised to construct Metaphorical Relation (MR) to signify an additional layer of meanings.

The development of the MR has also gone through different stages of evolution. At the initial stage, the principle of MR is more or less conformed to the conventional theories of two-dimensional design taught in design schools or practised by the professional designers. In this stage, the specifications of relational position among the objects are based on some rather straightforward and common patterns. For example, the concept of “Top of” mainly concerns about the relative position between two objects whereas one object would be placed at a position above the other one. This rather simple approach works in design practice as the human agent will fill in the rest of the missing gap intuitively without noticing that many subtle judgments have been made during the design process. Nonetheless, it will encounter different problems if the process is carried out by the computational method as many subtle conditional data and instructions required to create the generative design are missing. For example, all the conditions in Figure 9 can be considered as a “Top of” relations, however, each variation required additional information to differentiate the subtle variations in each case. In these cases, they all have a “Top of” relations in terms of the relative position between two objects, yet they have differences in the other spatial relations such as the distance between the objects or their “Length” relationship. Therefore, though all these compositions carry a “Top of” relations, each suggests different semantics according to the combined effect constituted by different types of spatial relations operated in each composition.













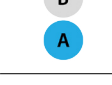














After a further study of the theories of image schema and incorporating the spatial primitives proposed by Mandler, a more elaborate schematic study of the interaction between different types of MR has been carried out. Through the deconstruction and reconstruction of our concept of the MR, it has been restructured and divided into six different dimensions namely, length, direction, distance, intersection, container and parallelism as listed in Table 7 below.

**Table 7: MR Dimensions**

Dimension Name	Description
Length	Comparison between comparable measurements
Direction	Relative positioning between two objects
Distance	How far apart objects are, relative to their size and context
Intersection	Count of intersecting points on object boundaries
Container	Description of the topological part-whole relationship between the two objects
Parallelism	Comparison of the orientation of various measurements (tension vectors, width-height ratio)













The MR proposed above using six dimensions to identify different types of spatial relations. Users can use RO to specify either one or more dimensions to define a specific MR between objects. Therefore, both single or multiple spatial dimensions can be applied according to the requirements. In Table 8 different examples of brand mark design demonstrate the MR as applied in real-world situations. In this table, the classification may only indicate the typical dimensions of spatial relation showing in a design. In many cases, more than one dimension could be operated in a single MR. In the first example of “Longer than”, some strips can be identified in the conditions of “Longer than” compared to the others. At once, some stripes are on the “Top of” the others while the others are on the “Right of” and “Bottom of” the other objects. So the possible meaning suggested by one single design can be quite rich as multiple MRs are operated in the design while different components interact with one another.

**Table 8: Spatial Relations between Two Objects and their Possible Metaphorical Expressions**

Relation	Description	Logo Examples	Possible Metaphorical Expressions
<b>Length</b>			
Longer than	Measurement (e.g width) of A is longer than B	  	authoritative dominant established expending full generous influential plentiful powerful
Equal to	Measurement of A is visually identical to B	  	accompany balance equality even matching partner symmetric competitor fair
Shorter than	Measurement of A is shorter than B	  	delicate fine gentle humble insignificant minor modest progeny subdominant
<b>Directional</b>			
Top of	A has a greater y-position (upwards) than B	  	authority climax dominant high rank important hierarchy superior good influential
Bottom of	A has a smaller y-position (downwards) than B	  	foundation earthy low rank base basic support endorse subordinate inferior
Left of	A has a smaller x-position (towards left) than B	  	accompany assist backward extend lead liberal member past peer
Right of	A has a greater x-position (towards right) than B	  	active advance assist extend forward follow future companion lateral
Front of	A has a greater z-index (forward) than B	  	apparent credible dominant exposed forward important inviting noticeable open
Back of	A has a smaller z-index (backward) than B	  	backing back up concealed emerge hidden introverted obscure dependant disguised



Relation	Description	Logo Examples			Possible Metaphorical Expressions		
Distance							
Far	Size of A is larger than the size of B					distant remote unrelated	unfamiliar strange reserved unknown foreign alienated
Near	Size of A is smaller than the size of B					closely related intimate friendly	familiar intrusive dependent approachable intimidating recognizable
Superimpose	A and B are of same size					closeness intimacy familiar	sameness wholeness interchangeable uniformity equality monotony
Intersection							
Disjoint	Boundaries of A and B do not have 0 intersection points					separation disconnection dissociation	division avoidance correlate comparison standalone independent
Meet	Boundaries of A and B have exactly 1 intersection point					contact interactive minimal	focus fragile brief instant approaching association
Multi-joint	Boundaries of A and B have more than 1 intersection points					connection joining dividing	portion meeting similarity agreement bond common
Shared Boundary	A and B shares the same partial boundary					sharing association agreement	peer acquaintance dependent togetherness union connection
Container							
Inside	A is within the interior of B					bounded protected secure	sheltered surrounded attention cherished child comforting
Across	A consists of the interior and exterior of B concurrently					affiliated component associated	consensus equality friendly linked share protection
Outside	A is at the exterior of B					accompany balance equality	foreign partnering symmetric unrelated disconnection competitor
Enclosed	A is within the hollow region of B					confined still passive	imprisoned protection shelter restricted secure certainty
Semibounded region	A is within the exterior of B's concave region					comfort sheltered bounded	guarding safe compassion inviting supportive oppressive
Conceptual boundary	A is between the exterior and the concave region of B					breakthrough progression active	improvement movement advancement possibility aggressive opportunity

Relation	Description	Logo Examples	Possible Metaphorical Expressions
<b>Parallelism</b>			
Parallel	The orientation of A and B are identical or have a 180 degrees difference	  	coordinated peer uniform peaceful polite structured obedient equal organised
Perpendicular	The orientation of A and B have a 90 degrees difference	  	established respected anarchic different resistance opposition obedience structure organised
Divergent	The direction of the force of A is parting from that of B's	  	spreading loose creative different irregular organic parting dissonant varying
Convergent	The direction of the force of A is heading towards that of B's	  	focused precision intense connecting approaching imminent acceleration inevitable agreement

What makes spatial relations even more complicated is when objects of different synthetic and semantic properties interact with one and the other. For example, despite that the definition of the MR “Shared Boundary” is simply defined as: “Parts of the boundary of A are the same as parts of the boundary of B”, the actual application of this relations could appear as different forms due to the differences between the form of object A and object B. Since a MR can be realised as so many possible compositions, it is necessary to identify the conditions that might constitute to these variations as well as the deviations in semantics. Currently, many scenarios have been tested and recorded in our knowledge base for analysis and further development. In general, these scenarios are identified either based on corpus analysis as mentioned in Section 3 or using a conventional design process. When using a conventional design process, a designer will use different design methods to seek solutions similar to a common design assignment. Some preliminary ideas based on either lateral thinking or rational process will be applied at the early stage. If a design can be matched with the MR presented in the exiting design, such scenarios will be kept in the knowledge base for further usage.

Then the preliminary ideas remained will go through another round of design process to seek for the other possibilities. As design is considered as a kind of ill-defined problem and new possibilities are always expected. So, this task is targeted for identifying a rich set of scenarios to facilitate the current development rather than making an exhaustive set. Once a collection of scenarios is available, it can be used as the reference for coding or annotating. In Table 9, a collection of scenarios that represents

the MR “Shared Boundary” is shown. In each one, the criteria of different dimensions of MR are listed except “Parallelism” which has no effect to “Shared Boundary”. With these additional criteria, many kinds of compositions with “Shared Boundary” can be generated to signify even more specific metaphorical expressions as listed in Table 8 since the conditions of the MR will become more confined to certain aspects. In the conceptual diagram of Table 9, only rectangles are used to illustrate the concepts of the different scenarios of “Shared Boundary”. In real-world applications, different kinds of shapes will be used as shown in the brand mark samples. Therefore, another set of scenarios will be required if the boundary is shared by a line and a shape. Thus, to assure that a MR can work with different kinds of shapes or objects, more tables and specifications similar to Table 9 are required for handling the other scenarios.

In Table 8, a number of spatial relations are listed with brand mark examples to illustrate how specific arrangement of visual objects can potentially signify different expressions. Table 8 only indicates one-to-one relations; however, these relations could be applied to one-to-many or many-to-many situations as shown in Figure 8.

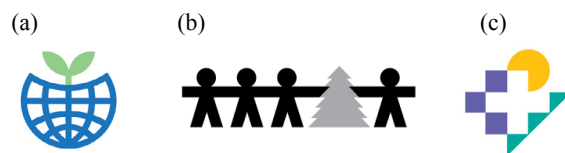






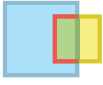











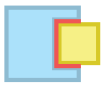





Figure 8: (a) one-to-one, (b) one-to-many, (c) many-to-many relations

**Table 9: Examples of Shared-Boundary Interacting with Different MR Dimensions**

No.	Conceptual diagram (yellow = main object)	Logo examples	MR Dimension				Possible Metaphorical Expressions	
			Length	Direction	Distance	Container		
(a)			Equal to	*	Near / Far	Outside	precision allied dependent	organised constant reliable
(b)			*	*	Near / Far	Outside	comparable similarity resemblance	supportive commonality coexistence
(c)			*	*	Near / Far	Across	supportive hinting sub-dominant	assisting complementary back-up
(d)			*	*	Near / Far	Across	attached dependant occupying	engaging distinction speciality
(e)			Equal To	Front of	Near / Far	Inside	sharing proportional component	extension composing integral
(f)			*	Front of	Near / Far	Inside	bounded travelling into opening	internal composing elemental
(g)			*	*	Near / Far	Enclosed	bounded constrained reserved	static focal point secure
(i)			*	*	Near / Far	Semi-bounded region	embrace valued important	protected secure immunity
(j)			*	*	Near / Far	Semi-bounded region	protected contained completion	within surrounded safe
(k)			*	*	Near / Far	Conceptual Boundary	interconnected dependent coincide	agreement blockade defensive
(l)			*	*	Near / Far	Conceptual Boundary	communal sharing mutual	correspond accompany mixing

\* This composition works for any type of this MR dimension

Through constructing the MR in a MPM with RO, different metaphorical expressions (ME) could be signified based on the principle of symmetrical object alignment (SOA). By choosing a source object with specific properties and combining with the relevant MR, the properties could be mapped to the target object and suggest the associated meaning. Thus, not only the literal meanings but also the metaphorical expressions could be delivered to the receivers from the design generated with the MPM.

As one of the main concerns of the MPM is for handling semantics, especially those related to expression and emotion, the principle of semantic differential could be utilised to control the parameters. By mapping the value of a semantic differential scale relevant to a specific parameter which controls the metaphorical relation, the form of the model could be modified and express such meaning respectively. For example, the weight of “authority” could be mapped to the relation “Top of”

whereas the distance between the top and bottom objects indicates the level of the “authority”. The position of the top object could be placed at either a higher or lower position to signify different degrees of authority as shown in Figure 9. At the same token, the effect could be enhanced by concatenating the “Larger than” relation to the size of an object.

Other than passing the semantic values to the responding parameters according to the principles of symmetrical object alignment and semantic differential, machine learning could be applied to learn and drive the model if the proper interface is available. Nevertheless, the experimentation of these ideas are out of the scope of this paper and further development should be done in future.

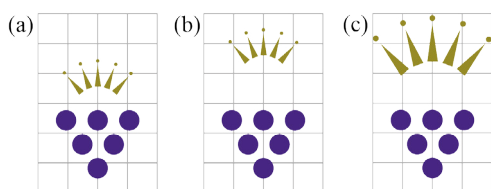


Figure 9: (a) Original brand mark, (b) increase value of parameter “yOffset” in TopOf and (c) increase value of parameter “xScale” of ScaleBy

### 5.3 Resolving problems of contradicting relations

As some of the MRs such as “Top of/ Bottom of”, “Left of/Right of” shown in Table 7 are contrary pairs, some relations will signify divergent meanings depend on which object is under attention. For example, the object “A” may carry meanings of “authority”, “high rank” or “good” while the object “B” may show the ideas of “support”, “low rank” or “inferior”. To a certain extent, the expressions of “authority” vs. “support”, “high rank” vs. “low rank” are complementary pairs. On the other hand, the expression of “good” and “bad” could be contradictory when a unified brand personality is preferable. In general, this type of conflict could be avoided if the MPM is assembling according to the theory of visual rhetoric as only the appropriate object will be selected. Although a detailed explanation of visual rhetoric is out of the scope of the current discussion, a brief description could elucidate how the MPM would work with semantic issues. In general, if a unified concept is required, the choice of the nature of both object “A” and “B” should complement each other as the samples show in the top row of Table 8. Hence the expression of the complementary pair such as “authority/ support”, “high rank/ low rank”, and “good” will be dominant while the contradicted meaning “bad” will be submerged. After all, the metaphorical expressions of a design are affected by the nature of the objects

as well as the relations between the objects found in the design.

If an object of a brand mark is a noticeable sign which carries distinctive meanings, the metaphorical expressions suggested by the metaphorical relations between objects could be overwhelmed by the semantics of such objects. For example, the construction of the brand mark in Figure 10, is based on the principle of the rhetorical figure “Antithesis”. The object “A” and “B”, which carry the “opposition of content” as defined by Durand [43, 44], are combined as a single compound. The composition of this design can be constructed by at least two different MR namely, “Left of / Right of ” and “Cover / Cover by”. In this case, metaphorical relations suggested by the “Left of / Right of ” or “Cover / Cover by” will become minor. Here, the contradictive expressions suggested by the “happy face” object and “sad face” objects on both sides are showed in parallel and suggest the expression of “Anthesis” for representing the theatre group.



Figure 10: Brand mark based on the principle of rhetorical figure “Antithesis”

## 6. An Example of the Procedure of Making Brand Mark Design with MPM

A formal procedure of making brand mark design with MPM is described in this section so that users can acquire a better understanding of how to utilise the model in real-world applications.

- Before working on a design, the designer or user should get the design specifications with information such as the nature of the business, the target group, the brand personalities or specific requirement collected in advance. Here, an existing brand mark of Swiss Airline is chosen and recreated as an MPM for the purpose of demonstration and explanation.
- In general, visual communication designers will search for visual metaphors to represent the design concept. The choice of the metaphor could come from analytical processes such as a visual inventory of the other existing work created for a similar business or lateral thinking such as brainstorming. As this demonstration is based on an existing design, the first step to do is analysing and identifying the metaphor in the original design. Through the deconstruction of the original brand mark, two visual metaphors representing the concepts “Swiss” and “airline” are identified as Figure 11.



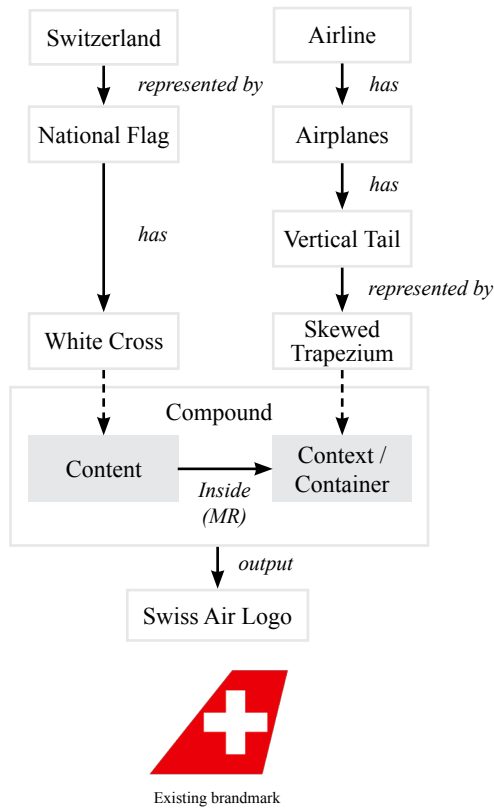


Figure 11: Swiss Airline logo analysis

- Depending on the design requirements, a brand mark design can be created by either an element, a compound or different combinations of both. As the Swiss Airline brand mark consists of two metaphorical concepts, creating a compound with two components will be a logical approach to initiate the design. To represent these two concepts, a compound consists of one placeholder to contain the concept “Swiss” and another one for the concept “airline” as shown in Figure 11 is created.
- Then these two placeholders are connected by one MR “inside” which will place the object “Swiss” enclosed within the boundary of the object “airline”. According to the principle of image schema, this MR may suggest the metaphorical expressions such as category, material made, inside, locate within, affordances, and boundary.
- Then the metaphorical concepts are encoded as a compound as shown in Table 10. Following the original design, a trapezium and a cross are assigned to the placeholders of the compound first. By setting the relevant parameters to the MR, a replicate is created.
- By changing the parameters or the elements of the compound, design alternatives can be generated as shown in Table 11. With a different element on the

back, the meaning of the “airline” can be transformed. Similarly, changing the front element can signify the other meaning as well.

Table 10: Coded Simulation of Existing Brand Mark Design

## Swiss Air Brand Mark Example

## Simulated Result



## Code Sample (Javascript)

```
//create trapezium to signify airline
var tail = Trapezium.createComposed({
  length: 190, height: 170,
  angleLeft: 60,
  angleRight: 90.5 })
tail.gt("skew", "x", -13, "CM", "g")
tail.style("fill", "red")






//create cross to signify Switzerland
var cross = CrossShape.createComposed({
  width:200, height:200,
  thicknessX: 60, thicknessY: 60 })
cross.style("fill", "white")

//establish MR between the tail and the cross
cross.addRelation("alignTo",tail,"CR",
  "CR",[15,0])
cross.addRelation("inside", tail, 0.85)

//plot resulting brand mark onto screen
var logo = Compo.createComposed({})
logo.addItem(tail, cross)
logo.plot()
```

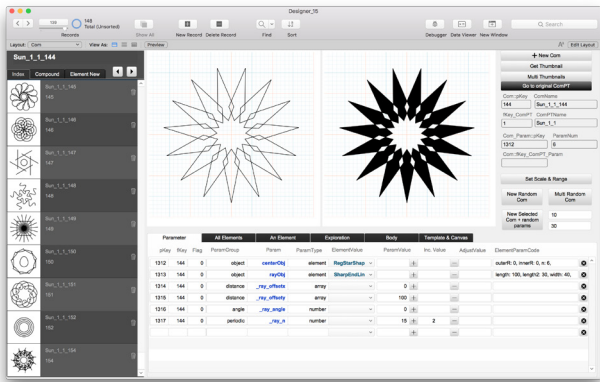
Table 11: Variations of the Simulated Result

Simulated Result	Description
<b>Change of Object</b>	
<p>Generated designs</p>	By changing the shape of the back element, different designs are created. These shapes are not the visual metaphors to represent the concept “airline”; however, they could represent the other brands from the Swiss.
<p>Existing brandmark</p>	

Simulated Result	Description
<b>Change of Object</b>	
	By changing the shape in the front element, another group of design are created. As the visual metaphor of “Swiss” is replaced, this design could signify an airline associated with different places or institutions.
	
	
<b>Change of Relation</b>	
	By changing the relations between the elements using Right Of, Part Of, Overlap and Top Of instead of Inside. The semantic meaning will change and signify different metaphorical expressions.
	

By using the RO to create different MRs between the elements of the original composition, new metaphorical expressions can be suggested as well. By referring to the relations shown in Table 8, each of these new relations can signify a manifold of meanings diverged from the original meaning.

Figure 12: Screenshot of Prototype with GUI Interface



As the MPM and its related operations are still under development currently, the coding library has not released to public usage yet. However, a prototype built with FileMaker Pro, which is a relational database supporting HTML and Javascript for rapid prototyping, is available for designers with no programming experience

to do design or experimentation. Through the GUI (Figure 12), users can assemble elements as compounds or using compounds to construct other compounds. By placing different objects into the placeholders and modifying the parameters of the objects or the MRs, users can explore a large number of designs in a short period.

## 7. Discussion



In order to create a universal model to facilitate visual communication design, building a model with the capacity for showing expression is just as important as representing information. The MPM proposed in this paper is an initial attempt to develop a new computational model for visual communication design based on the conceptual metaphor theory and visual rhetoric. The model is a representation of the semantic and metaphorical structure rather than any specific physical structure of a target object. Unlike some conventional architectural models which represent the physical structures of architectural forms, it only represents the most dominant conceptual structure perceived by the target groups. The structure of this model is not based on a definite ontological structure but a combination of conceptual and metaphorical structures. These structures are not only affected by our embodied experience but also our general knowledge prevailing in a specific group, society, or culture, not to mention the rules and regulations enforced by institutions or the other matters.

Based on the theory of symmetrical object alignment (SOA), we can map the metaphorical properties from a source object to the target object and signifying the intended metaphorical expressions. However, choosing the relevant components to construct a computational model requires specific domain knowledge. So, other than the knowledge of computing and coding, defining an MPM for making professional-grade brand mark design requires the understanding of visual communication design as well as the culture of the target audience.

When a visual form is displayed, the visual rhetorical treatments on graphic attributes are also crucial. For the same object, the choice of colour, line quality, texture or particular style could affect the expressions of the outcome. This kind of treatments is the same as the “metaphorical graphics”, which is a type of metaphor for information visualization, described by Risch [28]. The visual rhetorical operations applied to graphic attributes always carry another layer of metaphorical properties which could affect the expressions signified in the design.

A detail discussion of the visual rhetorical effect of graphic attribute is out of the scope of this paper, however, the current MPM has already incorporated the concept that all graphics attributes can be used to signify metaphorical expressions as the examples shown in Table 12.

**Table 12: Variations of metaphorical expressions resulted from the changes of graphic attributes**

Simulated Result	Description
<i>Change of Attribute</i>	
	Changes in actual vector point location
	Changes in styling methods (Vector points are not changed)

It is obvious that the current MPM created with vector graphics can not represent photographic or raster images in the design. However, with the relevant adjustments, the MPM can incorporate raster image processing functions to generate both vector and raster graphics at once. So such limitation is mainly the technical issues related to the rendering of graphics and should not hinder the potentials of the MPM.

This paper mainly focuses on how the conceptual metaphor theory and image schemas are employed in establishing the structures of MPM. It explains the assembling of visual objects as conceptual metaphors for making metaphorical expressions through the manipulations of metaphorical relations. Though it has explained the concept of symmetric object alignment, there is no discussion on how to choose which object as the source object and how it is mapped to the target object. Also, there is only a brief description of how to arrange the objects to compose the final design based on the theory of visual rhetoric. In order to elucidate the design process such as how to select objects and control their parameters, it will require a more detail introduction and in-depth descriptions of visual rhetorical operation (VRO) which is out of the scope of this paper.

Currently, a few of VROs based on the seminal paper of Durand [43, 44] have been implemented to facilitate the construction of the MPM. By using the VRO together with the MPM, plenty of diversified designs could be generated and the results are promising. The idea of mapping different conventional rhetorical figures into an interactive matrix might help Durand to know how to apply verbal rhetorical figures to advertisings. However, in the later stage of the current research, after trying the implementation of some of the rhetorical operations suggested by Durand, the limitations of his studies become noticeable.

Since the structure of verbal and visual representation are quite different, the current study reveals that, for the development of visual language, the research of visual rhetoric should be based on its nature while verbal

rhetoric should be taken for reference only. For example, the concept of inversion in the text is quite limited due to its linear arrangement. In graphics or painting, inversions [59] could be applied to more than one dimension which may include vertical and horizontal in spatial relations and the hue, lightness, and saturation in colour. Also, some symmetry patterns created with “repetition” can only be created in a two-dimensional plane but not the linear text. For example, the figure and ground phenomenon are gestalts unique to visual perception. As shown in many works of M.C. Escher, the metaphorical expressions revealed from the figure and ground gestalt in the symmetry pattern could not be recreated by the verbal rhetorical concept or in the linguistic form. Baruchello [60] has made a comprehensive study to identify the gestalt principles found in a large compilation of conventional rhetorical figures. The finding indicated that gestalt patterns could signify different rhetorical effects. However, the study also revealed that there is no schematic or distinctive correlation between the mapping of gestalt principles and rhetorical figures. Hence, it is obvious that the mapping between the gestalt and verbal rhetoric are not bilateral and it is quite reasonable to expect that some of the rhetorical effects that created with gestalt principles are unlikely to be found in the verbal rhetoric.

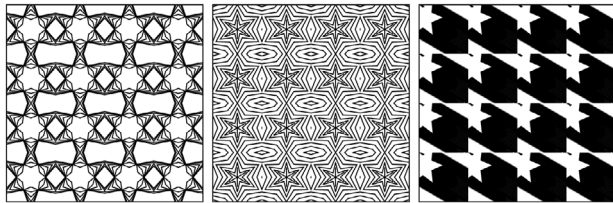
## 8. Conclusion and Future Development

The MPM proposed in this paper is only the first attempt towards a new type of model for visual communication design. Although the current study of the MPM is focused on brand mark design, obviously the same model could repurpose for signage design or icon design without the need of much modification of the core structure of MPM if the domain knowledge of these design types is provided. With the integration of the knowledge system of page layout and information graphics with the model, the application of the MPM can be further extended to the area of advertising or poster designs.

The usage of MPM is more versatile than expected; based on some experiments conducted previously, the same model can be used to simulate various types of pattern design or generative art as shown in Figure 13. This model is flexible and the users can modify the components and relations based on the theory of conceptual metaphor or visual rhetoric for an instant. Furthermore, though this model is initially developed for visual communication designers dealing with two-dimensional works, it could be adapted in different design disciplines immediately. For example, this model could be an ideal creative supporting tool (CST) for architects who need to design floor plans or facade layouts with symbolic meanings. As all the objects and relations in the MPM are encoded to signify the conceptual metaphor and visual rhetoric, this model

will allow the architects to create new designs in an environment which is difficult to achieve with conventional CAD systems.

Other than extending the application of MPM just mentioned, the study of visual rhetoric should never be ignored. When the MPM is applied to advertising, the utilisation and integration of visual rhetoric will demand different concerns as the visuals and objects used in advertising are more complicated and diverse. Furthermore, the structures and the components involved in a brand mark and advertising are quite dissimilar; hence, both require different kinds of visual rhetoric. So, other than meeting the demand that the investigation of visual metaphor should be based on the nature of the visual language and inherent to our visual perceptions as pointed out in the last section, the study of visual rhetoric should also be based on the nature of the types of visual communication. To fulfil these two demands, a new paradigm of the study of visual rhetoric should be established so that it can be grounded on its own and released from the shadow of verbal rhetoric and verbal language.



**Figure 13: Three different patterns based on the same element “Regular Star Shape”**

Last but not least, as Lakoff and Johnson [18] described a common conceptual metaphor found in human beings—ARGUMENT IS WAR. With such a concept, people use the structure of war to see an argument between different parties. People use the word and ideas such as fight, indefensible, weak point, target, demolished, shoot, strategy, opponent to deal with an argument. They suggested that if there is a culture which views argument as a dance, participants as performers, and the objective is to give a pleasing performance, then the result should become more constructive. In other words, the choice of metaphor could affect the thoughts and behaviour of people to a large extent beyond our awareness.

If metaphor is more fundamental than language, there should be dictionaries to archive the meanings of metaphors as well. As MPM is inherently a model created for representing visual metaphors, it would be an ideal model to contain visual metaphor and formalised as visual vocabularies. If a large number of visual vocabularies encoded as MPM are stored in a knowledge base, powerful resources would be available to facilitate different activities related to communication and thinking. As metaphor

is part of our living as suggested by Lakoff and Johnson, the formation of metaphor is a continuous process and new meanings always emerge under different social and cultural activities. The currency and meaning of a metaphor can be changed from time to time. Therefore, it would be beneficial if there is a knowledge base for archiving the visual vocabulary in different space-time rather than one time for a single culture only.

Eventually, whether a user is from engineering, architecture, fashion or communication design background, or even a musician or poet, all could find some approaches to use the model to support one’s works—by switching the mental image through the transformations of visual metaphors. Hence, through the building of MPM knowledge base with the metaphorical concepts of different cultures collected, we may be able to create a universal visual language to facilitate better communication and creativity for all in future.

## References

- [1] W. C. Brinton, *Graphic presentation*. New York: Brinton Associates, 1939.
- [2] O. Neurath, *International picture language: the first rules of isotype*. London: Kegan Paul, Trench, Trubner & Co. Ltd., 1936.
- [3] J. Bertin, *Semiology of graphics: diagrams, networks, maps*. Madison, WI: University of Wisconsin Press, 1983.
- [4] C. J. Richards, “Diagrammatics: an investigation aimed at providing a theoretical framework for studying diagrams and for establishing a taxonomy of their fundamental modes of graphic organization,” London: Royal College of Art, 1984.
- [5] Y. Engelhardt, “Syntactic structures in graphics,” *Computational Visualistics and Picture Morphology*, vol. 5, no. 1, 2007, pp. 23–36.
- [6] W. Kandinsky, *Point and line to plane*, NY: Solomon R. Guggenheim Foundation, 1947.
- [7] P. Klee, *Paul Klee notebooks Volume1 the thinking eye*, 1st ed., vol. 1. London: Lund Humphries Publishers Limited, 1973.
- [8] P. Klee, *Paul Klee notebooks volume 2 the nature of nature*, 1st ed., vol. 2. London: Lund Humphries Publishers Limited, 1973.
- [9] X. Lin, “Active layout engine: Algorithms and applications in variable data printing,” *CAD Comput. Aided Des.*, vol. 38, no. 5, 2006, pp. 444–456.
- [10] A. Jahanian, J. Liu, Q. Lin, D. R. Tretter, E. O’Brien-Strain, S. Lee, N. Lyons, and J. P. Allebach, “Recommendation system for automatic design of magazine covers,” in *International Conference on Intelligent User Interfaces, Proceedings IUI*, 2013, pp. 95–105.
- [11] T. J. Morris, “A Software Automation Framework For Image-Typeface Matching In Graphic Design,” M.A. thesis, Massachusetts Institute of Technology, USA, 2011.
- [12] B. J. Fry, “Computational Information Design,” Ph.D. thesis, Massachusetts Institute of Technology, USA, 2004.
- [13] L. Xiong and K. Zhang, “Generation of Miro’s surrealism,” in *Proc. VINCI*, 2016, pp. 130–137.
- [14] W. Shi, “A generative approach to Chinese shanshui painting,” *IEEE Comput. Graph. Appl.*, vol. 37, no. 1, pp. 15–19, 2017.
- [15] M. T. Chi, C. C. Hu, and Y. J. Jhan, “A sketch-based generation system for oriental cloud pattern design” in *Proc. SG*, 2014, pp. 27–38.



- [16] Y. Li, K. Zhang, and D. Li, “Rule-based automatic generation of logo designs,” *Leonardo*, vol. 50, no. 2, pp. 177–181, 2017.
- [17] J. Zhang, J. Yu, K. Zhang, X. S. Zheng, and J. Zhang, “Computational aesthetic evaluation of logos,” *ACM Trans. Appl. Percept.*, vol. 14, no. 3, pp. 1–21, 2017.
- [18] G. Lakoff and M. Johnson, *Metaphors we live by*. Chicago and London: The University of Chicago Press, 1980.
- [19] G. Lakoff and M. Johnson, *Philosophy in the flesh: The embodied mind and its challenge to western thought*. 1999.
- [20] M. Johnson, “Image-schematic bases of meaning,” *Semiotic inquiry*, vol. 9, pp. 109–118, 1989.
- [21] R. Gibbs, “The psychological status of image schemas,” *Cognitive linguistics research*, vol. 29, pp. 113–136, 2005.
- [22] R. Gibbs, H. Colston, “The cognitive psychological reality of image schemas and their transformation,” *Cognitive Linguistics*, vol. 6, no. 4, pp. 347–378, 1995.
- [23] T. Rohrer, “Image Schemata in the Brain,” *Cognitive linguistics research*, vol. 29, pp. 165–196, 2005.
- [24] J. M. Mandler, “On the birth and growth of concepts,” *Philos. Psychol.*, vol. 21, no. 2, pp. 207–230, 2008.
- [25] J. M. Mandler, “Prelinguistic Primitives,” in *Proceedings of the Seventeenth Annual Meeting of the Berkeley Linguistics Society: General Session and Parasession on The Grammar of Event Structure*, 1991, pp. 414–425.
- [26] J. M. Mandler, “The spatial foundations of the conceptual system,” *Lang. Cogn.*, vol. 2, no. 1, pp. 21–44, 2010.
- [27] J. M. Mandler and C. P. Cánovas, “On defining image schemas,” *Lang. Cogn.*, vol. 6, no. 04, pp. 510–532, 2014.
- [28] Risch, J.S. On the role of metaphor in information visualization. ArXiv, abs/0809.0884, 2008. Available: <https://arxiv.org/abs/0809.0884> [Accessed Aug.13, 2019].
- [29] J. Grady, “Foundations of meaning: Primary metaphors and primary scenes,” Ph.D. dissertation, University of California, Berkeley, USA, 1997.
- [30] D. Gentner, & D.R. Gentner, “Flowing waters or teeming crowds: mental models of electricity,” In D. Gentner & A.L. Stevens (Eds.), *Mental Models*. Hillsdale, NJ: Erlbaum.
- [31] R. Gibbs, “The poetics of mind: figurative thought, language, and understanding,” Cambridge: Cambridge University Press, 1994.
- [32] A. Szwedek, “An alternative theory of metaphorisation,” *Language and meaning: Cognitive and functional perspectives*, pp. 312–327, 2007.
- [33] F. J. R. de M. Hernández and I. L. Pérez, “The Contemporary Theory of Metaphor: Myths, Developments and Challenges,” *Metaphor Symb.*, vol. 26, no. 3, pp. 161–185, Jul. 2011.
- [34] R. Fusaroli and S. Morgagni, “Conceptual metaphor theory: thirty years after,” *Cognitive Semiotics*, vol. 5, no. 1–2, pp. 1–12, 2013.
- [35] M. Leezenberg, “From cognitive linguistics to social science: thirty years after *Metaphors We Live By*,” *Journal of Cognitive Semiotics*, vol. 5, no. 1–2, pp. 140–152, 2013.
- [36] L. Brandt, “Metaphor and the communicative mind,” *Journal of Cognitive Semiotics*, vol. 5, no. 1–2, pp. 37–72, 2013.
- [37] A. Deignan and L. Cameron, “A re-examination of UNDERSTANDING IS SEEING,” *Journal of Cognitive Semiotics*, vol. 5 no. 1–2, pp. 220–243, 2013.
- [38] R.W. Gibbs, “Making good psychology out of blending theory,” *Cognitive Linguistics*, vol. 11, no. 3–4, pp. 347–358, 2000.
- [39] R.W. Gibbs, “Prototypes in dynamic meaning construal,” *Cognitive Poetics in Practice*, pp. 27–40, 2003.
- [40] R.W. Gibbs and M. Tendahl, “Cognitive effort and effects in metaphor comprehension: Relevance theory and psycholinguistics,” *Mind & Language*, vol. 21, no. 3, pp. 379–403, 2006.
- [41] A. Marcus, *Visual Rhetoric in a Pictographoc-Ideographic Narrative*. Semiotics Unfolding, 3, 1501–1503, 1983.
- [42] G. Bonsiepe, “Visual/verbal rhetoric,” *Ulm J. Ulm Sch. Des.* 14/15, pp. 37–42, 1965.
- [43] J. Durand, “Rhétorique et image publicitaire,” *Communications*, vol. 15, no. 1, pp. 70–95, 1970. [Online]. Available: [https://www.persee.fr/doc/comm\\_0588-8018\\_1970\\_num\\_15\\_1\\_1215](https://www.persee.fr/doc/comm_0588-8018_1970_num_15_1_1215).
- [44] J. Durand and T. Van Leeuwen, “Rhetoric and the advertising image,” 1983. [Online]. Available: [http://jacques.durand.pages-perso-orange.fr/Site/anglais/texte\\_2\\_ang.htm](http://jacques.durand.pages-perso-orange.fr/Site/anglais/texte_2_ang.htm).
- [45] B. A. Huhmann and P. A. Albinsson, “Assessing the Usefulness of Taxonomies of Visual Rhetorical Figures,” *J. Curr. Issues Res. Advert.*, pp. 1–25, 2018.
- [46] J. Schilperoord, A. Maes, and H. Ferdinandusse, “Perceptual and conceptual visual rhetoric: The case of symmetric object alignment,” *Metaphor Symbol*, vol. 24, no. 3, pp. 155–173, Jul. 2009.
- [47] R. Arnheim, *Art and Visual Perception: A Psychology of the Creative Eye* (New version, expanded and rev. ed.). Berkeley, CA: University of California Press, 1974.
- [48] M. Ortiz, “Visual Rhetoric: Primary Metaphors and Symmetric Object Alignment,” *Metaphor Symb.*, vol. 25, no. 3, pp. 162–180, Jul. 2010.
- [49] Pragglegaz Group, “MIP: A Method for Identifying Metaphorically Used Words in Discourse,” *Metaphor Symb.*, vol. 22, no. 1, pp. 1–39, Jan. 2007.
- [50] Y. S. Jung, *The Best in World Trademarks: Corporate Identity, Brand Identity, Volume 1–2*, Korea: Millim, 2005.
- [51] *New logo: a collection of corporate identities*. Singapore: Page One, 2004.
- [52] C. L. Larson, J. Aronoff, I. C. Sarinopoulos, and D. C. Zhu, “Recognizing threat: A simple geometric shape activates neural circuitry for threat detection,” *J. Cogn. Neurosci.*, vol. 21, no. 8, pp. 1523–1535, 2009.
- [53] C. L. Larson, J. Aronoff, and J. J. Stearns, “The Shape of Threat: Simple Geometric Forms Evoke Rapid and Sustained Capture of Attention,” *Emotion*, vol. 7, no. 3, pp. 526–534, 2007.
- [54] J. Ibáñez and C. Delgado-Mata, “Emotional line: Showing emotions through the sharpness, height, width and speed of a series of concatenated digital shifting curves,” in *Proc. SG*, 2014, vol. 8698 LNCS, pp. 98–111.
- [55] O. Blazhenkova and M. M. Kumar, “Angular versus curved shapes: correspondences and emotional processing,” *Perception*, vol. 47, no. 1, pp. 67–89, 2018.
- [56] A. Carlier, K. Leonard, S. Hahmann, G. Morin, and M. Collins, “The 2D shape structure dataset: A user annotated open access database,” *Comput. Graph.*, vol. 58, pp. 23–30, 2016.
- [57] E. J. Golin and S. P. Reiss, “The specification of visual language syntax,” *J. Vis. Lang. Comput.*, vol. 1, no. 2, pp. 141–157, 1990.
- [58] M. J. Egenhofer and M. P. Dube, “Topological relations from metric refinements,” in *Proc. 17th ACM SIGSPATIAL Int. Conf. Adv. Geogr. Inf. Syst. –GIS*, 2009, p. 158.
- [59] S. Kim, *Inversions*, Peterborough, NH: BYTE Books, 1981.
- [60] G. Baruchello, “A classification of classics. Gestalt psychology and the tropes of rhetoric,” *New Ideas Psychol.*, vol. 36, pp. 10–24, 2015.





# Journal of Visual Language and Computing

journal homepage: [www.ksiresearch.org/jvlc](http://www.ksiresearch.org/jvlc)

## Graphical Animations of the Suzuki-Kasami Distributed Mutual Exclusion Protocol<sup>\*,\*\*</sup>

Dang Duy Bui<sup>a</sup>, Kazuhiro Ogata<sup>a,\*</sup>

<sup>a</sup>*School of Information Science, Japan Advanced Institute of Science and Technology (JAIST), 1-1 Asahidai, Nomi, Ishikawa 923-1292, Japan*

### ARTICLE INFO

#### Article History:

Submitted 3.1.2019

Revised 6.1.2019

Second Revision 8.1.2019

Accepted 10.10.2019

#### Keywords:

distributed protocols

graphical animations

Maude

model checking

SMGA

state machines

### ABSTRACT

A state machine that formalizes a distributed mutual exclusion protocol called the Suzuki-Kasami protocol is graphically animated. The messages that have been just put into (or sent) and deleted from (or received) the network are crucial information and then visually explicitly displayed on the designated places in a state picture. The protocol uses some pieces of information that are seemingly owned by each node but actually shared by all nodes. The pieces of information are visually explicitly displayed on two designated places. One main purpose of graphically animating state machines is to make it possible for humans to visually perceive characteristics or properties of the state machines. We demonstrate that carefully observing graphical animations makes it possible for human users to perceive some characteristics or properties of the state machine formalizing the Suzuki-Kasami protocol and the properties are confirmed by model checking. To make it more likely for human users to be able to perceive such properties, it is necessary to design good state pictures. We summarize some tips on how to design good state pictures for mutual exclusion protocols.

© 2019 KSI Research

## 1. Introduction

Many kinds of information and communication technology (ICT) systems can be formalized as state machines. Some ICT systems, such as the Internet, have become important societal infrastructures, they must be highly reliable. It is crucial to comprehend ICT systems better so as to make them highly reliable. Because ICT systems can be formalized as state machines, one possible way to comprehend ICT systems is to understand state machines that formalize the ICT systems. There may be multiple possible ways to understand state machines, and one promising way to do so would be to rely on human visual perception. Therefore, Nguyen and Ogata [8] have developed a tool called SMGA that generates graphical animations of state machines. Some shared-memory mutual exclusion protocols and some communica-

tion protocols have been tackled with SMGA so far. But, ICT systems that are societal infrastructures are often in the form of distributed systems. In this paper, thus, a distributed mutual exclusion protocol called the Suzuki-Kasami protocol [14] is tackled with SMGA.


As all other distributed systems, the Suzuki-Kasami protocol uses a network with which messages are exchanged among nodes. The message that has been just put into the network (just sent by a node) or just deleted from the network (just received by a node) is crucial information and then we prepare one place dedicated to the one that has been just put into the network and one place dedicated to the one that has been just deleted from the network. There may be too many messages in the network to display all of them on the designated place. If that is the case, a limited number of messages is displayed and the others are suppressed. Instead, “...” is displayed. The protocol uses some pieces of information that are seemingly owned by each node but actually shared by all nodes. The pieces of information are visually explicitly displayed on the designated places in a state picture.

Understanding a state machine is to know properties the state machine enjoys. The more state machine properties we know, the better we understand the state machine. One

<sup>\*</sup>This work was partially supported by JSPS KAKENHI Grant Number JP26240008 & JP19H04082.

<sup>\*\*</sup>The present paper is an extended and revised version of the paper [4] presented at DMSVIVA 2019.

<sup>\*</sup>Corresponding author

 [bddang@jaist.ac.jp](mailto:bddang@jaist.ac.jp) (D.D. Bui); [ogata@jaist.ac.jp](mailto:ogata@jaist.ac.jp) (K. Ogata)

ORCID(s): 0000-0002-2700-1762 (D.D. Bui); 0000-0002-4441-3259 (K. Ogata)

DOI reference number: 10-18293/JVLC2019-N2-012

main purpose of SMGA is to make it possible for humans to perceive characteristics or properties of a state machine by observing graphical animations of the state machine. We guess some properties of the state machine formalizing the Suzuki-Kasami protocol by observing its graphical animations and confirm the properties by model checking. We use Maude [5], a rewriting logic-based computer language, as a specification language for state machines and a model checker. We mention some SMGA functionalities, such as playing forward and backward frame-by-frame playback and finding out states that satisfy some conditions from a state sequence used in an input file. We also summarize how to design state pictures for mutual exclusion protocols.

The rest of the paper is organized as follows. Sect. 2 mentions some preliminaries needed to comprehend the technical contents of the paper. Sect. 3 describes state machine graphical animation (SMGA). Sect. 4 describes the Suzuki-Kasami protocol. Sect. 5 describes how to specify the Suzuki-Kasami protocol in Maude. Sect. 6 describes how to revise SMGA so that the Suzuki-Kasami protocol can be tackled and reports on the case study in which the protocol has been graphically animated with the revised version of SMGA. Sect. 7 mentions some functionalities of SMGA. Sect. 8 mentions on how to confirm guessed properties of the protocol with model checking. Sect. 9 summarizes some tips on how to design state pictures for mutual exclusion protocols. Sect. 10 mentions some related work. Sect. 11 concludes the paper and mentions some future directions.

## 2. Preliminaries

Let us consider as an example a test&set mutual exclusion (or spin-lock) protocol whose pseudo-code is as follows:

### Loop

“Remainder Section”

rm : **repeat while** test&set(*locked*);

“Critical Section”

cs : *locked* := false;

Multiple processes participate in the protocol. Each process is in Remainder Section, working on some tasks that do not require any shared resources. When a process needs some shared resources, it is supposed to move to Critical Section where it uses the shared resources. After that, it goes back to Remainder Section. We suppose that each process is located at either rs or cs. A process is located at rs and cs if and only if it is in Remainder Section and Critical Section, respectively. *locked* is a Boolean variable shared with all processes participating in the protocol. test&set(*locked*) atomically performs the following: it sets *locked* to true and returns the old value stored in *locked*. When a process wants to enter Critical Section, it repeatedly conducts test&set(*locked*) until false is returned and then goes to cs. When it leaves Critical Section, it sets *locked* to false and goes back to rs.

The protocol is formalized as a state machine  $M \triangleq \langle S, I, T \rangle$  that consists of a set  $S$  of states, where some states

$I \subseteq$  are initial states, and a binary relation  $T$  over states, where each element  $(s, s') \in T$  is a state transition, saying that state  $s$  can go to state  $s'$ . We use Maude [5], a programming/specification language based on rewriting logic, to specify state machines. Maude makes it possible to specify complex systems flexibly and is also equipped with model checking facilities, such as a reachability analyzer (or a search command).

When there are three processes  $p_1, p_2$  and  $p_3$  participating in the protocol, let  $M_{TS} \triangleq \langle S_{TS}, I_{TS}, T_{TS} \rangle$  be the state machine formalizing the protocol. A state in  $S_{TS}$  can be expressed as follows:  $\{(pc[p_1]: l_1)(pc[p_2]: l_2)(pc[p_3]: l_3)(locked: b)\}$ , where  $li$  (for  $i = 1, 2, 3$ ) is either rs or cs and  $b$  is either true or false.  $(pc[p_i]: li)$  and  $(locked: b)$  are called observable components that are name-value pairs, meaning that process  $p_i$  is located at  $li$  and the value stored in variable *locked* is  $b$ , respectively.  $pc[p_i]$  and *locked* are names, while  $li$  and  $b$  are values.  $(pc[p_i]: li)$  is also called a pc or  $pc[p_i]$  observable component and  $(locked: b)$  is also called a locked observable component. Observable components are glued as members of an associative-commutative collection that is called a soup. Therefore, a state is expressed as a braced soup of observable components.  $I_{TS}$  consists of one state that is expressed as follows:  $\{(pc[p_1]: rs)(pc[p_2]: rs)(pc[p_3]: rs)(locked: false)\}$ , which will be referred as *ic*.

$T_{TS}$  is specified by the following two rewrite rules:

```
r1 [enter] : {(locked: false) (pc[I]: rs) 0Cs}
=> {(locked: true) (pc[I]: cs) 0Cs} .
```

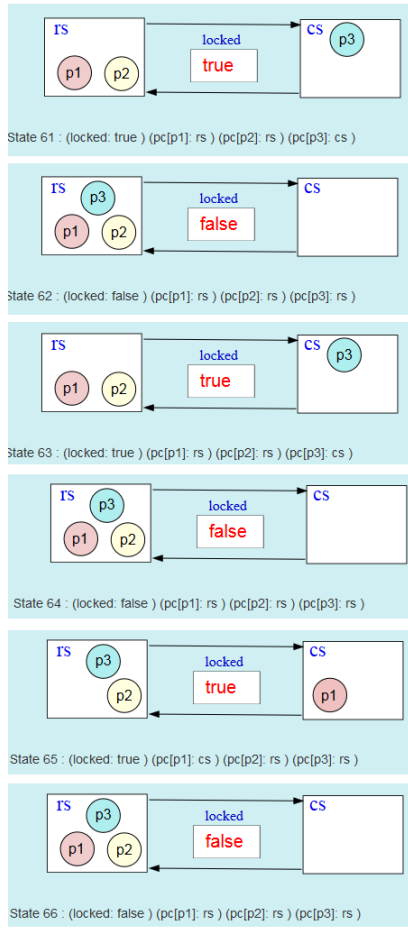
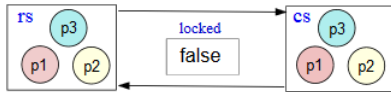
```
r1 [exit] : {(locked: B) (pc[I]: cs) 0Cs}
=> {(locked: false) (pc[I]: rs) 0Cs} .
```

Rewrite rules are defined with  $r_1$ , while conditional ones are defined with  $cr_1$  and their conditions are written after *if*. *enter* and *exit* are labels (or names) given to the rules, respectively.  $I$  is a Maude variable of process identifications,  $B$  is a Maude variable of Boolean values and  $0Cs$  is a variable of observable component soups.

By substituting  $B, I$  and  $0Cs$  with *true*,  $p_3$  and  $(pc[p_1]: rs) verb!(pc[p_2]: rs)!$ , the left-hand side of rule *exit* comes to equal State 61 shown in Fig. 1 because soups are associative and commutative and then the right-hand side comes to equal State 61. By substituting  $B, I$  and  $ICs$  with *false*,  $p_1$  and  $(pc[p_2]: rs) (pc[p_3]: rs)$ , the left-hand side of rule *enter* comes to equal State 64 shown in Fig. 1 and then the right-hand side comes to equal State 65. These are how rewrite rules describe state transitions.

## 3. State Machine Graphical Animation (SMGA)

SMGA [8] has been implemented with Draw SVG ([www.drawsvg.org](http://www.drawsvg.org)). It basically takes a finite computation (or a finite state sequence) of a state machine and generates a graphical animation of the state machine. For each state, a picture is designed and then we get a series of pictures from a finite computation. Such a series of pictures is regarded as a movie

Figure 1: A sequence of pictures for  $M_{TS}$ Figure 2: A picture of states in  $S_{TS}$ 

film. This is how SMGA generates a graphical animation of a state machine.

We could design a picture of states in  $S_{TS}$  as shown in Fig. 2 when there are three processes. An input file to SMGA consists of three parts: `###keys`, `###textDisplay` and `###states`. Fig. 3 shows a snip of an input file to SMGA for  $M_{TS}$ . In the `###keys` part, the names (or keys) used in the observable components that constitute each state are enumerated. For  $M_{TS}$ , `locked`, `pc[p1]`, `pc[p2]` and `pc[p3]` are enumerated. There are two different ways to display the value stored in an observable component: (1) the value is displayed at a designated place and (2) one of the diagram objects associated with the value appears, while the other diagram objects disappear. For (1), the value is treated as a text and displayed horizontally from left to right by default. Some data structures, such as stacks, should be displayed in a different way.

```
###keys
locked pc[p1] pc[p2] pc[p3]

###textDisplay

###states
(locked: false) (pc[p1]: rs) (pc[p2]: rs) (pc[p3]: rs) ||
(locked: true) (pc[p1]: cs) (pc[p2]: rs) (pc[p3]: rs) ||
(locked: false) (pc[p1]: rs) (pc[p2]: rs) (pc[p3]: rs) ||
(locked: true) (pc[p1]: rs) (pc[p2]: rs) (pc[p3]: cs) ||
```

Figure 3: A snip of an input file to SMGA for  $S_{TS}$ 

For example, a stack should be displayed vertically from top to bottom. In the `###textDisplay` part, we can specify how values treated as texts are displayed for (1). For  $M_{TS}$ , the value stored in the `locked` observable component is displayed as (1). Because it is displayed by default, nothing is specified in the `###textDisplay` part. The values stored in the `pc[p1]`, `pc[p2]` and `pc[p3]` observable components are displayed as (2). For example, if the value stored in the `pc[p1]` observable component is `rs`, the circle on which 1 is written appears in the `rs` region and otherwise the circle appears in the `cs` region. In the `###states` part, a sequence of states is written, where states are separated by `||`.

Fig. 1 shows a sequence of pictures for  $M_{TS}$  generated by SMGA. In State 61, the value stored in the `locked` observable component is `true`, the value stored in the `pc[p1]` observable component is `rs`, the value stored in the `pc[p2]` observable component is `rs`, the value stored in the `pc[p3]` observable component is `cs`. Hence, `true` is displayed by default at the designated place, the circle on which 1 is written appears in the `rs` region, the circle on which 2 is written appears in the `rs` region and the circle on which 3 is written appears in the `cs` region in State 61. In State 62, the value stored in the `locked` observable component changes to `false`, the value stored in the `pc[p3]` observable component changes to `rs` and the other values do not change. Hence, `false` is displayed by default at the designated place, the circle on which 3 is written appears in the `rs` region and the other two circles remain in the `rs` region in State 62.

Observing such a sequence of pictures or a graphical animation of a state machine, we could guess some properties of the state machine [9]. Observing the sequence of pictures shown in Fig. 1, for example, we could guess the property that `locked` is `false` if and only if there is no process in the critical section, or equivalently `locked` is `true` if and only if there exists a process in the critical section. We use the Maude reachability analyzer (or the search command) to confirm that the guessed property is invariant with respect to  $M_{TS}$ . The following command can be used to confirm it:

```
search [1] in TS : ic
=>* {(locked: false) (pc[I]: cs) 0Cs} .
```

Maude exhaustively searches the reachable states from the initial state `ic` for a state that can match `{(locked: false) (pc[I]: cs) 0Cs}`. If there is such a state, in which a process `I` is located at `cs` and `locked` is `false`, then the property is

try( <i>i</i> )	↔	rem	<b>procedure</b> P1
setReq( <i>i</i> )	↔	11	<i>requesting</i> := true;
chkPrv( <i>i</i> )	↔	12	<b>if</b> $\neg have\_privilege$ <b>then</b>
incRN( <i>i</i> )	↔	13	<i>rn</i> [ <i>i</i> ] := <i>rn</i> [ <i>i</i> ] + 1;
sndReq( <i>i</i> )	↔	14	<b>for all</b> <i>j</i> ∈ {1, ..., <i>N</i> } − { <i>i</i> } <b>do</b> send request( <i>i</i> , <i>rn</i> [ <i>i</i> ]) to node <i>j</i> ; <b>endfor</b>
wtPrv( <i>i</i> )	↔	15	<b>wait until</b> privilege( <i>queue</i> , <i>ln</i> ) is received; <i>have_privilege</i> := true; <b>endif</b>
exit( <i>i</i> )	↔	cs	Critical Section;
cmpReq( <i>i</i> )	↔	16	<i>ln</i> [ <i>i</i> ] := <i>rn</i> [ <i>i</i> ]; <b>for all</b> <i>j</i> ∈ {1, ..., <i>N</i> } − { <i>i</i> } <b>do</b> <b>if</b> ( <i>j</i> ∉ <i>queue</i> ) ∧ ( <i>rn</i> [ <i>j</i> ] = <i>ln</i> [ <i>j</i> ] + 1) <b>then</b> <i>queue</i> := enq( <i>queue</i> , <i>j</i> ); <b>endif</b> <b>endfor</b>
updQ( <i>i</i> )	↔	17	<b>if</b> <i>queue</i> ≠ empty <b>then</b> <i>have_privilege</i> := false; send privilege(deq( <i>queue</i> ), <i>ln</i> ) to node top( <i>queue</i> ); <b>endif</b>
chkQ( <i>i</i> )	↔	18	<i>request</i> := false;
trsPrv( <i>i</i> )	↔	19	<b>endproc</b>
rstReq( <i>i</i> )	↔	110	
			// request( <i>j</i> , <i>n</i> ) is received; P2 is indivisible.
			<b>procedure</b> P2
			<i>rn</i> [ <i>j</i> ] := max( <i>rn</i> [ <i>j</i> ], <i>n</i> );
			<b>if</b> <i>have_privilege</i> ∧ $\neg request$ ∧ ( <i>rn</i> [ <i>j</i> ] = <i>ln</i> [ <i>j</i> ] + 1) <b>then</b> <i>have_privilege</i> := false; send privilege( <i>queue</i> , <i>ln</i> ) to node <i>j</i> ; <b>endif</b> <b>endproc</b>
recReq( <i>i</i> )	↔		

**Figure 4:** Suzuki-Kasami distributed mutual exclusion protocol in an Algol-like language

violated. No such a state is found by Maude and then the guessed property is invariant with respect to  $M_{TS}$  when there are three processes.

#### 4. Suzuki-Kasami Protocol

The Suzuki-Kasami protocol is a distributed mutual exclusion protocol [14]. The basic idea is that a node that has a privilege is only allowed to enter its critical section, and there exists one and only one privilege in the system. The privilege is owned by a node, or is in the network being transferred by a node to another. We suppose that  $N$  nodes participate in the protocol and the natural numbers  $1, \dots, N$  are used as their identifications. Let Node be  $\{1, \dots, N\}$ . The  $N$  nodes have no memory in common and can communicate only by exchanging messages. The communication delay is totally unpredictable, namely that although messages eventually arrive at their destinations, they are not guaranteed to be delivered in the same order in which they are sent. There are two kinds of messages used in the Suzuki-Kasami protocol: request and privilege messages. A request message is in the form request(*j*, *n*), where *j* is the identification of the node that has sent the message and *n* is a request number. A privilege message is in the form privilege(*q*, *a*), where *q* is a queue of node identifications and *a* is a natural number array of size  $N$ .

The Suzuki-Kasami protocol consists of two procedures P1 and P2 for each node  $i \in \text{Node}$ . The procedures for node  $i$  are shown in Fig. 4.

*request* and *have\_privilege* are Boolean variables. *request* indicates whether or not node  $i$  wants to enter its critical section, and *have\_privilege* indicates whether or not node  $i$  owns the privilege. *queue* is a queue of Node. It contains the identifications of nodes that wait to enter their critical sections. *ln* and *rn* are natural number arrays of size  $N$ . *ln*[*j*] for each node  $j \in \text{Node}$  is the sequence number of node  $j$ 's request granted most recently. *rn* records the largest request number ever received from each of the other nodes. Node  $i$  uses *rn*[*i*] to generate the sequence numbers of its own requests. For each node  $i \in \text{Node}$ , its *rn* is always meaningful, while its *queue* and *ln* are meaningful only when node  $i$  owns the privilege. When the privilege is in the network, *queue* and *ln* contained in the privilege message are meaningful. For each node  $i \in \text{Node}$ , initially, *request* is false, *have\_privilege* is true if  $i = 1$  and false otherwise, *queue* is empty, and *ln*[*j*] and *rn*[*j*] for each  $j \in \text{Node}$  are 0.

If node  $i$  wants to enter its critical section, it first calls its own procedure P1, which sets *request* to true. If it happens to own the privilege, it immediately enters the critical section. Otherwise, it generates the next sequence number, namely, incrementing *rn*[*i*], and sends the request message request(*i*, *rn*[*i*]) to all other nodes. When it receives a privilege message privilege(*queue*, *ln*), it enters the critical section. When it leaves the critical section, it sets *ln*[*i*] to its current sequence number *rn*[*i*], meaning that the current request has been granted, and updates *queue* such that if there are nodes that want to enter their critical sections and whose identifications are not yet in the queue, their identifications are added to the queue. After that, if *queue* is not empty, node  $i$  sets *have\_privilege* to false and sends the privilege message privilege(deq(*queue*), *ln*) to the node found in the front of the queue. Otherwise, node  $i$  keeps the privilege. Finally, node  $i$  sets *request* to false and leaves procedure P1.

Whenever request(*j*, *n*) is delivered to node  $i$ , node  $i$  executes its own procedure P2. However, procedure P2 has to be atomically executed. When node  $i$  executes procedure P2, it sets *rn*[*j*] to *n* if *n* is greater than *rn*[*j*]. Then, if node  $i$  owns the privilege, does not want to enter its critical section, and the *n*th request of node  $j$  has not been granted, that is, *rn*[*j*] = *ln*[*j*] + 1, then it sets *have\_privilege* to false and sends the privilege message privilege(*queue*, *ln*) to node  $j$ .

#### 5. Specification of Suzuki-Kasami Protocol

Let Nat, Bool, Loc, NodeQueue, and NatNArray be the set of all natural numbers, the set of the Boolean values (true and false), the set of all locations, such as rem and ll, the set of all queues of Node, and the set of all natural number arrays whose sizes are  $N$ , respectively. A request message addressed to node  $i \in \text{Node}$  by node  $j \in \text{Node}$  is expressed as msg(*i*, req(*j*, *k*)), where  $k \in \text{Nat}$ , msg is used as the constructor of messages and req is used as the constructor of requests. A privilege message addressed to node  $i \in \text{Node}$  is expressed as msg(*i*, priv(*q*, *a*)), where  $q \in \text{NodeQueue}$ ,  $a \in \text{NatNArray}$  and priv is used as the constructor of privileges. Let Req and Priv be  $\{\text{req}(i, n) \mid i \in \text{Node}, n \in \text{Nat}\}$



```

r1 [sndReq] : {(pc[I]: 14) (idx[I]: K) (rn[I]: RN) (nw: NW) (tran: T) OCs}
=> {(pc[I]: if K == N then 15 else 14 fi) (idx[I]: if K == N then 1 else K + 1 fi) (rn[I]: RN)
(nw: (if K == I then NW else msg(K, req(I, RN[I])) ; NW fi) (tran: sndReq(I)) OCs} .

r1 [wtPrv] : {(pc[I]: 15) (havePriv[I]: F) (ln[I]: LN') (queue[I]: Q')
(nw: (msg(I, priv(Q, LN)) ; NW)) (tran: T) OCs}
=> {(pc[I]: cs) (havePriv[I]: true) (ln[I]: LN) (queue[I]: Q) (nw: NW)
(tran: wtPrv(I)) OCs} .

r1 [trsPrv] : {(pc[I]: 19) (havePriv[I]: F) (ln[I]: LN) (queue[I]: Q) (nw: NW)
(tran: T) OCs}
=> {(pc[I]: 110) (havePriv[I]: false) (ln[I]: LN) (queue[I]: Q)
(nw: (msg(top(Q), priv(get(Q), LN)) ; NW)) (tran: trsPrv(I)) OCs} .

crl [recReq] : {(pc[I]: L) (request[I]: F) (havePriv[I]: F') (rn[I]: RN) (ln[I]: LN)
(queue[I]: Q) (nw: (msg(I, req(J, X)) ; NW)) (tran: T) OCs}
=> {(pc[I]: L) (request[I]: F) (havePriv[I]: if C then false else F' fi)
(rn[I]: RN[J] := Max) (ln[I]: LN) (queue[I]: Q)
(nw: if C then (msg(J, priv(Q, LN)) ; NW) else NW fi) (tran: recReq(I)) OCs}
if I /= J /\ L /= 110 /\ L /= 18 /\ L /= 17 /\
Max := if (RN[J]) < X then X else RN[J] fi /\ C := F' and not(F) and Max == (LN[J]) + 1 .

```

Figure 5: Rewrite rules specifying  $T_{SK}$ 

and  $\{\text{priv}(q, a) \mid q \in \text{NodeQueue}, a \in \text{NatNArray}\}$ , respectively. The network is formalized as a soup of messages that are request and privilege messages. The set  $\text{MsgSoup}$  of all soups of messages is inductively defined as follows:  $\text{void} \in \text{MsgSoup}$ , for each  $i \in \text{Node}$ ,  $r \in \text{Req}$ , and  $p \in \text{Priv}$ ,  $\text{msg}(i, r) \in \text{MsgSoup}$  and  $\text{msg}(i, p) \in \text{MsgSoup}$ , and for each  $ms_1, ms_2 \in \text{MsgSoup}$ ,  $ms_1 ; ms_2 \in \text{MsgSoup}$ . A semicolon  $;$  is used as the constructor of soups of messages.  $\text{void}$  denotes the empty soup of messages and is the identity of  $;$ , namely that  $ms ; \text{void} = \text{void}$ ;  $ms = ms$  for each  $ms \in \text{MsgSoup}$ . Each message is also treated as the singleton soup that only consists of the message.

The Suzuki-Kasami protocol is formalized as a state machine  $M_{SK} \triangleq \langle S_{SK}, I_{SK}, T_{SK} \rangle$ , which is specified in Maude. P1 is divided into 12 regions shown in Fig. 4. The 12 regions are referred as the 12 locations, such as  $\text{rem}$  and  $\text{l1}$ . We suppose that each node is at one of those 12 locations. Procedure P2 is regarded as one region and then there are totally 13 regions in the Suzuki-Kasami protocol. The 13 regions are given names, such as  $\text{try}(i)$  and  $\text{setReq}(i)$ , shown at the left-most column in Fig. 4. For each node  $i$ , there are 13 kinds of transitions that corresponds to the 13 regions. The 13 region names are used to refer to the 13 kinds of transitions. Let us note that  $\text{sndReq}(i)$  and  $\text{updQ}(i)$  correspond to each iteration of the loops at labels  $\text{l4}$  and  $\text{l7}$ , respectively.

When there are three nodes, a state in  $M_{SK}$  is expressed as follows:  $\{n(1) \ n(2) \ n(3) \ (nw: ms) \ (tran: t)\}$ , where  $n(i)$  is as follows:

```

(#req[i]: n) (pc[i]: l) (request[i]: b1)
(havePriv[i]: b2) (rn[i]: a1) (ln[i]: a2)
(queue[i]: q) (idx[i]: j)

```

where  $ms$  is a soup of messages in the network,  $t$  is the transition that has been just taken,  $n$  is the number of requests made by node  $i$ ,  $l$  is the location where node  $i$  is,  $b1$  is the value of the node  $i$ 's request,  $b2$  is the value of the node  $i$ 's have\_privilege,  $q$  is the value of the node  $i$ 's queue and  $j$  is the value of the node  $i$ 's  $j$ , a loop variable. The state expression defines  $S_{SK}$ .

When there are three nodes,  $I_{SK}$  consists of one state that is expressed as follows:  $\{n(1) \ n(2) \ n(3) \ (nw: \text{void}) \ (tran: \text{notran})\}$ , where  $\text{notran}$  means that no transition has been taken and  $n(i)$  is as follows:

```

(#req[i]: 0) (pc[i]: rem) (request[i]: false)
(havePriv[i]: (I == 1)) (rn[i]: ia)
(ln[i]: ia) (queue[i]: empty) (idx[i]: 1)

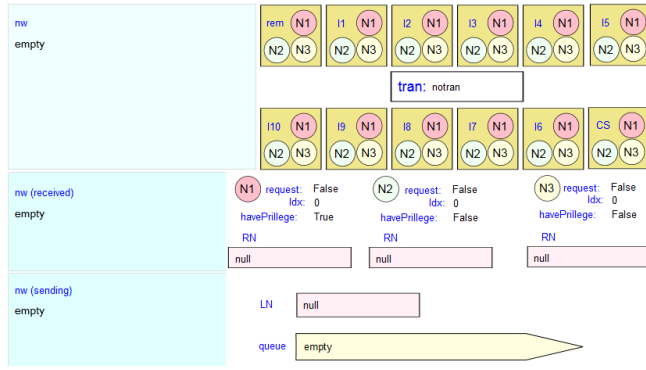
```

where  $ia$  denotes the natural number array such that each slot is 0,  $I == 1$  is true if  $I$  is 1 and false otherwise, and  $\text{empty}$  denotes the empty queue.

$T_{SK}$  is specified in terms of (conditional) rewrite rules. There are 13 rules, among which four rules are shown in Fig. 5. The words starting with a capital letter, such as  $X$ ,  $I$ ,  $\tau$  and  $OCs$ , are Maude variables. Their types (or sorts) could be understood from the context. For example,  $X$ ,  $I$ ,  $\tau$  and  $OCs$  are variables of  $\text{Nat}$ ,  $\text{Loc}$ , transition names and observable component soups, respectively.  $RN[J] := \text{Max}$  is the array assignment at index  $J$ .  $\text{top}(Q)$  is the top element of  $Q$ .  $\text{get}(Q)$  denotes the queue obtained by deleting the top from  $Q$ . What is called a matching equation<sup>1</sup>  $V := T$ , where  $V$  is a fresh variable and  $T$  is a term, can be used in rule conditions and is like **let** expressions in functional programming languages. The rules will be described later by observing some state pictures generated by SMGA. Let us, however, mention one thing about rule  $\text{recReq}$  that formalizes procedure P2. The rule has  $L \neq 110 \wedge L \neq 18 \wedge L \neq 17$  as part of the condition, saying that node  $I$  does not receive any requests messages if it is located at  $\text{l10}$ ,  $\text{l8}$  or  $\text{l7}$ . The condition is not explicitly mentioned in the original paper [14] of the Suzuki-Kasami protocol. If we do not use the condition, the protocol may cause lockout that a node that wants to enter its critical section never be there if a node does not try to enter its critical section unboundedly many times [10].

<sup>1</sup>In general, a matching equation is in the form  $T1 := T2$ , where  $T1$  and  $T2$  are terms.



Figure 6: A picture of states in  $S_{SK}$ 

```

###keys
nw tran #req[1] #req[2] #req[3] pc[1] pc[2] pc[3] requesting[1]
requesting[2] requesting[3] havePriv[1] havePriv[2] havePriv[3]
rn[1] rn[2] rn[3] ln[1] ln[2] ln[3] queue[1] queue[2] queue[3]
idx[1] idx[2] idx[3]

###textDisplay

###conditionDisplay
ln queue
ln[1] ln[2] ln[3] ***priv(, ln)
ln[1]++++havePriv[1]==true
ln[2]++++havePriv[2]==true
ln[3]++++havePriv[3]==true
***priv(, ln)++++nw
queue[1] queue[2] queue[3] ***priv(queue, _)
queue[1]++++havePriv[1]==true
queue[2]++++havePriv[2]==true
queue[3]++++havePriv[3]==true
***priv(queue, _)++++nw

###states
(nw: void tran: notran (#req[1]: 0) (#req[2]: 0) (#req[3]: 0)
(pc[1]: rem) (pc[2]: rem) (pc[3]: rem) (requesting[1]: false)
(requesting[2]: false) (requesting[3]: false) (havePriv[1]: true)
(havePriv[2]: false) (havePriv[3]: false) (rn[1]: ia) (rn[2]: ia)
(rn[3]: ia) (ln[1]: ia) (ln[2]: ia) (ln[3]: ia) (queue[1]: empty)
(queue[2]: empty) (queue[3]: empty) (idx[1]: 1) (idx[2]: 1)
idx[3]: 1) ||
(nw: void tran: try(2) (#req[1]: 0) (#req[2]: 1) (#req[3]: 0)
(nc[1]: rem) (nc[2]: 1) (nc[3]: rem) (requesting[1]: false)

```

Figure 7: A snip of an input file to SMGA for  $S_{SK}$ 

## 6. Graphical Animations of Suzuki-Kasami Protocol

Fig. 6 shows a picture of states in  $S_{SK}$  when there are three nodes participating in the Suzuki-Kasami protocol. There is a pane (called the nw pane) located in the left upper corner where the messages in the network are displayed. Under the nw pane, there is a pane (called the nw (received) pane) where the message that has been just received by a node (or just deleted from the network) is displayed. Under the nw (received) pane, there is a pane (called nw (sending) pane) where the message that has been just sent by a node (or just put into the network) is displayed. For each node  $i = 1, 2, 3$ , there are places to display the node  $i$ 's *request*,  $j$  (or *idx*), *have\_privilege* and *rn*. There is always exactly one *queue* that is meaningful and then there is one place to display the meaningful *queue*. There is always exactly one *ln* that is meaningful and then there is one

place to display the meaningful *ln*. If there is a node whose *have\_privilege* is true, its *queue* and *ln* are displayed there. If there is a privilege message in the network, namely that there is no node whose *have\_privilege* is true, then nothing is displayed there because you can see the meaningful *queue* and *ln* in the privilege message in the network. There are 12 panes that correspond to the 12 locations, such as *rs* and *l1*. There is one more pane in the picture where the transition that has been just taken is displayed.

Fig. 7 shows a snip of an input file to SMGA for  $M_{SK}$ . We have added one more part called the `###conditionDisplay` part to an input file to SMGA. Although each node has its own variables *ln* and *queue*, they have meaningful values only if the node owns the privilege. Hence, we would like to only display the meaningful *ln* and *queue* on the two places on a state picture, respectively. Just below the line where `###conditionDisplay` is written in the input file shown in Fig. 7, we write the two names *ln* and *queue* because at most one *ln* and at most one *queue* are displayed. The next line `ln[1] ln[2] ln[3] ***priv(, ln)` specifies that there are four possible places where the meaningful *ln* is stored: the *ln[1]* observable component, the *ln[2]* observable component, the *ln[3]* observable component and the privilege message in the network. The next line `ln[1]++++havePriv[1]==true` says that if the *havePriv[1]* observable component stores true, the value stored in the *ln[1]* observable component is displayed on the place for *ln*. The following two lines can be interpreted likewise. The next line `***priv(, ln)++++nw` says that if there exists a privilege message in the network, nothing is displayed on the place for *ln*. The following five lines specify how to deal with *queue* and can be interpreted in the same way as the five lines for *ln*.

The *nw* observable component consists of the messages that have been sent and have not been received yet. Those messages are displayed on the *nw* pane. It may be impossible to display all messages on the *nw* pane. If that is the case, the messages that can fit to the *nw* pane are only displayed, which are followed by "...". If there exists a message that has been just put (or sent) into the network, the message is displayed on the *nw* (sending) pane. If there exists a message that has been just deleted (or received) from the network, the message is displayed on the *nw* (received) pane.

Fig. 8 shows a sequence of four pictures for  $M_{SK}$ . The four pictures correspond to four consecutive states State 242, State 243, State 244 and State 245 in a finite computation of  $M_{SK}$ . State 242 goes to State 243 by rewrite rule *sndReq* (or *sendReq(2)*), State 243 goes to State 244 by rewrite rule *trsPrv* (or *transferPriv(1)*) and State 244 goes to State 245 by rewrite rule *wtPrv* (or *waitPriv(3)*). Taking a look at the first picture (of State 242) immediately makes us recognize that node 1 is located at I9, node 2 is located at I4, node 3 is located at I5, node 1 owns the privilege, there is one message denoted *msg(1, req(2, 5))*, the rule *sendReq(2)* (or *sndReq(2)*) has been just taken and so on. What is displayed as the content of *ln* is (1 : 4), (2 : 4), 3 : 3, which says that *ln[1]* is 4, *ln[2]* is 4 and *ln[3]* is 3, meaning that the node 1's fourth request has been granted, the node 2's fourth re-

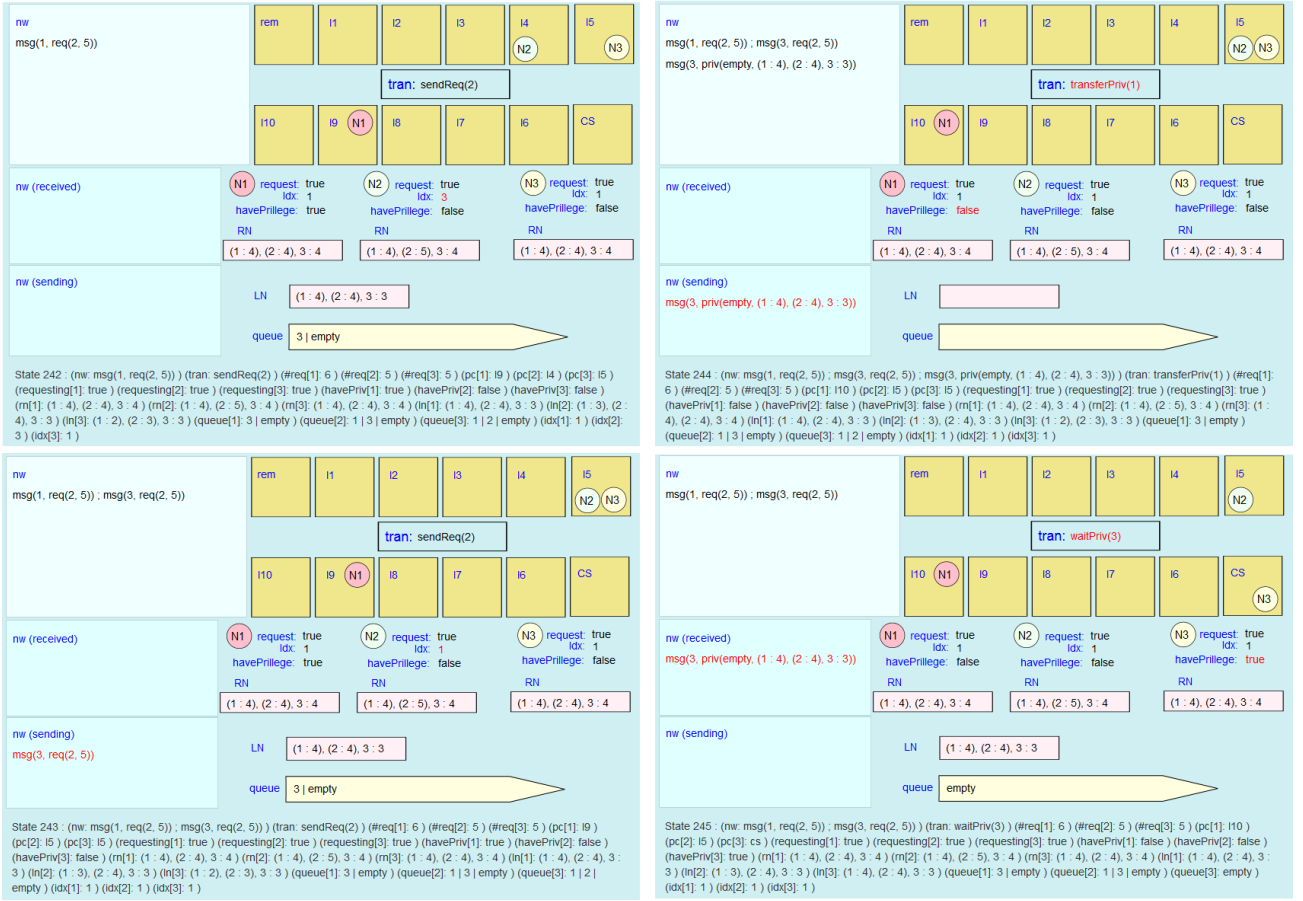


Figure 8: A sequence of pictures for  $M_{SK}$  (1)

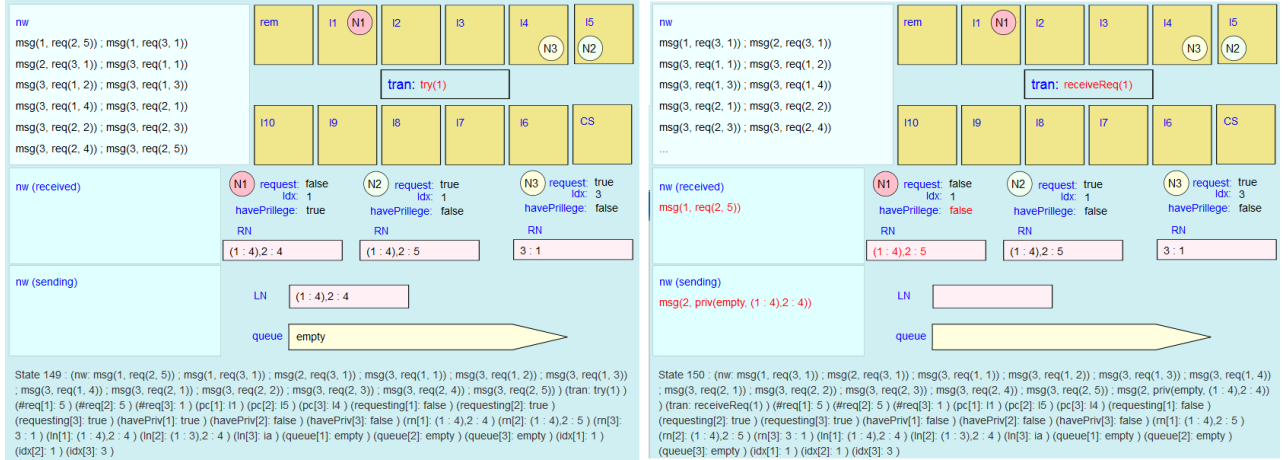


Figure 9: A sequence of pictures for  $M_{SK}$  (2)

quest has been granted and the node 3's third request has been granted. What is displayed as the content of *queue* is 3 | empty, which says that there is one element in *queue* and the element is 3, meaning that node 3 has been waiting to enter its critical section.

Taking a look at the second picture (of State 243) makes

us recognize that the rule  $sendReq(2)$  (or  $sndReq(2)$ ) has been just taken, the message  $msg(1, req(2, 5))$  has been just put into the network and node 2 has just moved to  $I5$  from  $I6$ . The state transition from State 242 to State 243 visually describes an instance of what rule  $sndReq$  (or  $sendReq(2)$  or  $sndReq(2)$ ) in Fig. 5 does.

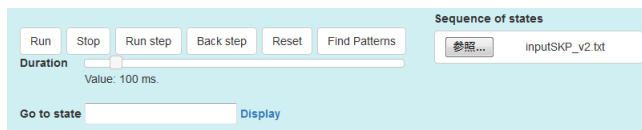
Taking a look at the third picture (of State 244) makes us recognize that the rule `transferPriv(1)` (or `trsPrv(1)`) has been just taken, the message `msg(3,priv(empty, (1 : 4),(2 : 4),3 : 4))` has been just put into the network and node 1 has just moved to 110 from 19. The state transition from State 243 to State 244 visually describes an instance of what rule `trsPrv` (or `transferPriv(1)` or `trsPrv(1)`) in Fig. 5 does.

Taking a look at the fourth picture (of State 245) makes us recognize that the rule `waitPriv(3)` (or `wtPrv(3)`) has been just taken, the privilege message has been just received by node 3 (or just deleted from the network) and node 3 has just moved to `cs` from 15. The state transition from State 244 to State 245 visually describes an instance of what rule `wtPrv` (or `transferPriv(1)` or `trsPrv(1)`) in Fig. 5 does.

Fig. 9 shows another sequence of two pictures for  $M_{SK}$ . The two pictures correspond to two consecutive states State 149 and State 150 in another finite computation of  $M_{SK}$ . State 149 goes to State 150 by rewrite rule `recReq` (or `receiveReq(1)`). If it is possible to display all messages in the network on the nw pane, SMGA does so. Otherwise, a limited number of messages are displayed on the nw pane and the others are depressed. In the first picture (of State 149), all messages in the network are displayed on the nw pane. `msg(1,req(2,5))` is received by node 1 and `msg(2,priv(empty, (1 : 4),2 : 4))` is put into the network by node 1. Then, it is impossible to display all messages in the network on the nw pane. Therefore, 10 messages out of 12 ones are displayed on the nw pane and the two messages `msg(3,req(2,5))` and `msg(2,priv(empty, (1 : 4),2 : 4))` are depressed in the second picture (of State 150). Instead of displaying the two messages, “...” is displayed on the nw pane in addition to the 10 messages.

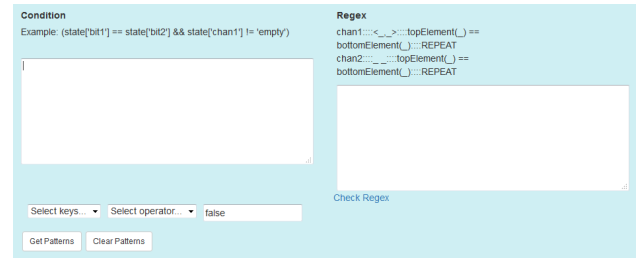
## 7. Some Functionalities of SMGA

In addition to screening state machine movies, SMGA allows us to play forward and backward frame-by-frame playback, etc. Under the state sequence movie (or state picture) displayed by SMGA, there are several buttons shown as follows:



The “Run step” button plays forward frame-by-frame playback and the “Back step” does backward one. The “Duration” adjuster can change how fast a movie is played. When you write a number  $N$  in the box just right-hand side of “Go to state” and click “Display,” the picture of state  $N$  is displayed.

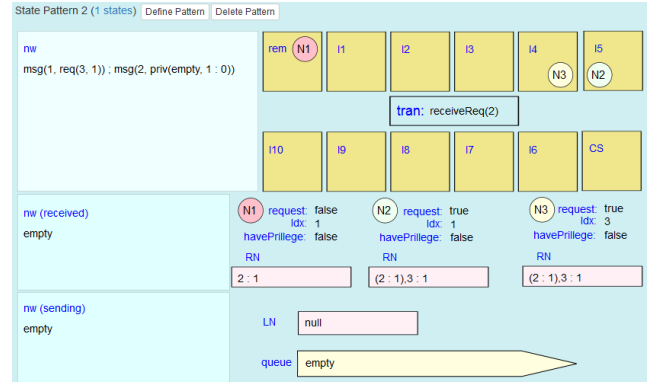
The “Find Patterns” button makes it possible to find states that satisfy some conditions. When the button is clicked, the following appears:



Writing conditions in the box under “Condition” and regular expressions in the box under “Regex”, the “Get Patterns” button finds all states appearing in the state sequence of the input file that satisfy the conditions and regular expressions. When the following condition is written in the condition box,

```
(state['havePriv[1]'] == 'false' ) &&
(state['havePriv[2]'] == 'false' ) &&
(state['havePriv[3]'] == 'false' )
```

SMGA finds out 7 states among 1000 states that satisfy the condition in the first input file. One state found is as follows:



We notice that there exists a privilege message in the network in the state.

When the following condition is written in the condition box,

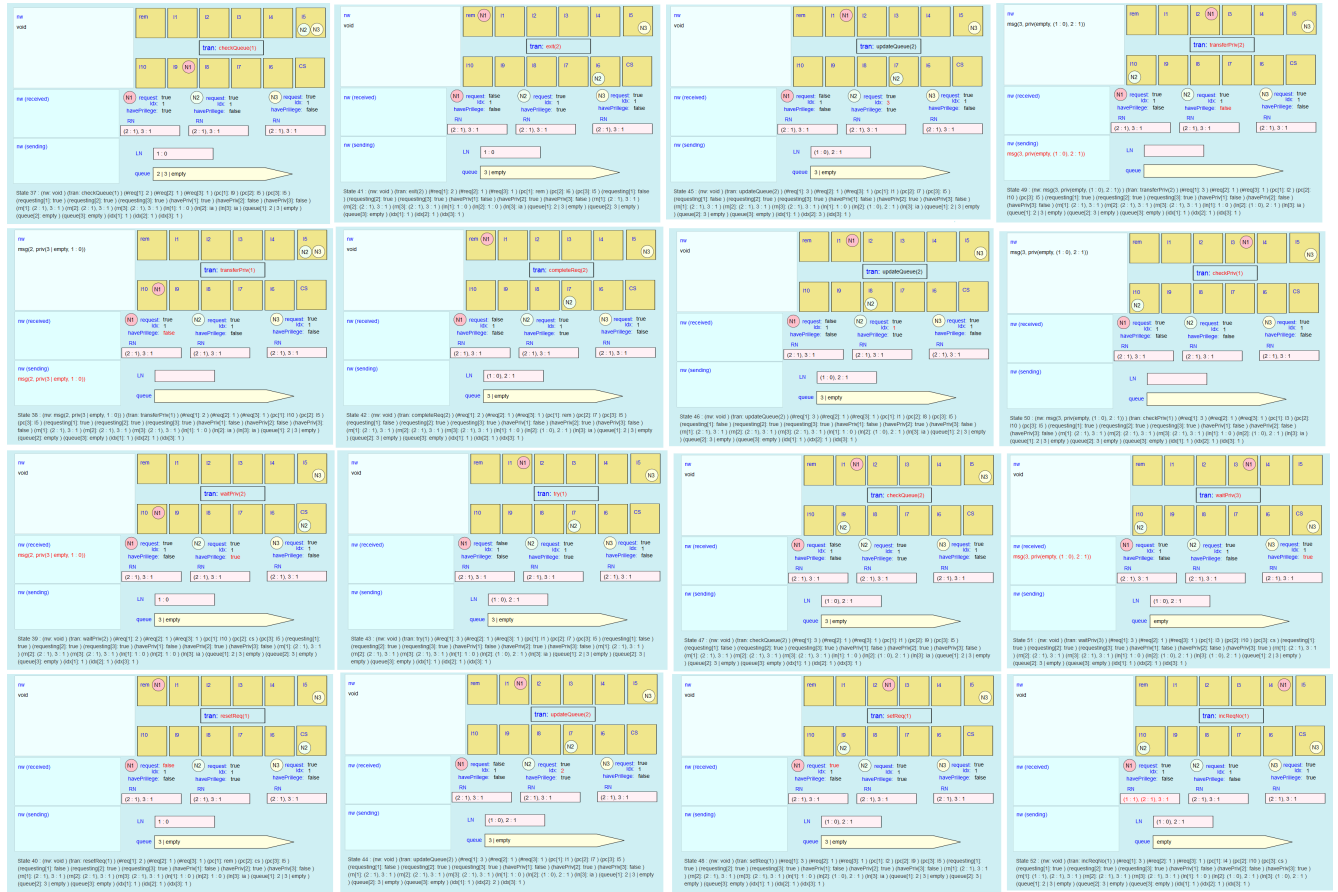
```
(state['pc[1]'] == 'cs') &&
(state['pc[2]'] == 'cs')
```

SMGA does not find out any states among 1000 states that satisfy the condition in the first input file. From it, we may guess that two or more nodes can not be in their critical sections at the same time. For the following condition,

```
(state['havePriv[1]'] == 'true') &&
(state['havePriv[2]'] == 'true')
```

SMGA does find out any states, either. From it, we may guess that two or more node *have\_privilege*'s cannot be true at the same time.

The SMGA functionality that finds out states that match regular expression for the Suzuki-Kasami protocol has not been implemented. We would like to express some regular expression that says that there exists a privilege message in the network. It is one piece of our future work to implement the SMGA functionality with regular expressions.

Figure 10: A sequence of pictures for  $M_{SK}$  (3)

## 8. Confirmation of Guessed Properties with Model Checking

Observing some graphical animations of  $M_{SK}$  could help us visually perceive some characteristics or properties of  $M_{SK}$ . Fig. 10 shows yet another sequence of states for  $M_{SK}$ . Carefully observing such graphical animations of  $M_{SK}$ , we notice that there is always at most one node located at cs, l6, l7, l8 or l9 at any given moment. The guessed property can be confirmed by model checking with Maude as follows:

```
search [1] in SK : ic
=>* (pc[I]: l1) (pc[J]: l2) OCs .
```

for  $l1, l2 \in \{cs, l6, l7, l8, l9\}$ . The search command checks if there is a state reachable from ic such that two nodes I & J are located at cs, l6, l7, l8 or l9 at the same time. Because the command does not find any such state, the guessed property is true when there are three nodes.

Carefully observing some graphical animations of  $M_{SK}$ , we also guess that there is a privilege message in the network if and only if there is no node located at cs, l6, l7, l8 or l9 at any given moment. This is true in all pictures shown in Fig. 8, Fig. 9 and Fig. 10. The guessed property can be confirmed by model checking with Maude as follows:

```
search [1] in SK : ic
=>* (nw: (msg(I,priv(Q,A)) ; NW))
(pc[J]: l1) OCs .
```

for  $l1 \in \{cs, l6, l7, l8, l9\}$ . The search command checks if there is a state reachable from ic such that a privilege message is in the network and a node is located at cs, l6, l7, l8 or l9 at the same time. Because the command does not find any such state, the guessed property is true when there are three nodes.

Carefully observing some graphical animations of  $S_{SK}$ , we also guess that there exists a node whose *have\_privilege* is true if and only if there does not exist any other node at cs, l6, l7, l8 or l9 at any given moment. This is also true in all pictures shown in Fig. 8, Fig. 9 and Fig. 10. The guessed property can be confirmed by model checking with Maude as follows:

```
search [1] in SK : ic
=>* (pc[I]: l1) (havePriv[J]: true) OCs
such that I = J .
```

for  $l1 \in \{cs, l6, l7, l8, l9\}$ . The search command checks if there is a state reachable from ic such that a node I is located at cs, l6, l7, l8 or l9 and a node j owns the privilege at the same time, where  $I \neq J$ . Because the command does not find

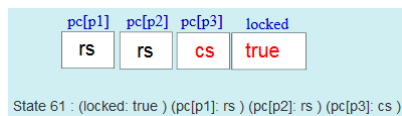


any such state, the guessed property is true when there are three nodes.

Note that model checking only guarantees that the three guessed properties are true when there are three nodes. We need to use theorem proving so as to guarantee that the properties are true when there are an arbitrary number of nodes [2].

## 9. Some Tips on How to Design State Pictures

SMGA does not automatically produce visual representations or pictures of states but human users are supposed to design state pictures. Any state pictures are not good. For example, one possible state picture of the test&set protocol is as follows:



Although state pictures like this could be automatically generated, we do not think that this state picture is very good for the test&set protocol because the state picture is almost the same as the text representation shown below the state picture. The state picture for the test&set protocol shown in Fig. 1 and Fig. 2 is better than the one shown above. This is because (1) we can immediately realize how many sections there are totally, (2) how many processes there are in each section and (3) the relations among the sections such that a process located at *cs* will move to *rs*. We can also visually perceive that (4) variable *locked* is shared by all processes and (5) what section each process is located at.

We summarize some tips on how to design state pictures for mutual exclusion protocols based on our experiences.

- To recognize what sections there are at which each process or node is located, allocate the pane (or place) for each section such that the relations among the sections are visually perceived and display some diagram, such as a circle on which a process or node ID is written, on the designated pane;
- To recognize what pieces of information, such as the network for the Suzuki-Kasami protocol and variable *locked* for the test&set protocol, are shared by all processes or nodes, allocate the pane (or place) for each such piece of information such that we can visually perceive they are shared by all processes and nodes and display them on the designated panes adequately;
- To recognize whether there are some that are more crucial than the others among the shared resources, such as the messages that have been just put into and deleted from the network, prepare the panes (or places) for them and display them there adequately;
- To recognize what pieces of information are owned by each process or node, allocate the panes (or places) for them to make it possible to visually perceive what pieces of information are owned by what processes or

nodes and display them on the designated panes adequately.

There may be some pieces of information that are seemingly stored in each process or node variables but actually the pieces of information are shared by all processes or nodes. *queue* and *In* used in the Suzuki-Kasami protocol are such pieces of information. To realize it, we need to comprehend the Suzuki-Kasami protocol to some extent. One important lesson learned from our experiences is that it is necessary to comprehend mutual exclusion protocols well to some extent so as to design reasonably good state pictures.

## 10. Related Work

Alloy [7] is a relational-logic based specification (or modeling) language. Its environment is also called Alloy. Alloy is equipped with a SAT-based bounded model checker. When it finds a counterexample, it automatically visualizes states. PAT (Process Analysis Toolkit) [13] is an enhanced simulator, model checker and refinement checker for concurrent and real-time systems. Its simulator can automatically visualize states or state sequences, such as counterexamples. Automatic state visualization could help human users comprehend states or state sequences, such as counterexamples, better to some extent. Each application or system, however, has its own characteristics and then must have a good picture that cannot be automatically generated but should be designed by human users especially if the main purpose of graphically animating state machines is to find out non-trivial properties that could be used as lemmas for theorem proving. A graphical user interface for Maude-NPA has been developed [11]. Maude-NPA is a high-level security protocol analysis language and system implemented on the top of Maude. The graphical user interface is dedicated to Maude-NPA and then cannot be used for our main purpose.

Visualization of formal specifications have been attempted. Hoxha, et al. [6] have proposed how to visualize real-time temporal logic formulas because it is a error prone task to specify desired requirements in such logic for conventional engineers. Tikhonova, et al. [15] have proposed how to visualize formal specifications in a DSL implemented on the top of Event-B and Arcainiet al. [1] have proposed a visual notation for Abstract State Machines (ASMs). Unlike these studies, SMGA does not aim at visualizing formal specifications.

Computer networks have been grown and intricate. Social networking service (SNS) has been used by many people over the world and then networks constituted of those SNS users have become very complex. Visualization is one promising way to comprehend such complex networks and then network visualization has been intensively studied. Tools, such as Gephi [3] and Cytoscape [12], have been developed. Some visualization techniques used in those tools could be used to implement some functionalities given by SMGA. Because state machines cannot be necessarily expressed as networks only, however, those network visualization tools cannot be directly used to graphically animate state machines.



## 11. Conclusion

We have described graphical animations of the Suzuki-Kasami protocol with a revised version of SMGA. Observing them has made us guess some properties of the state machine formalizing the Suzuki-Kasami protocol. We have used the Maude reachability analyzer (the search command) to confirm that the guessed properties are invariant with respect to the state machine when there are three nodes and each node enters its critical section once. The case study demonstrates that state machine graphical animations could make humans perceive state machine properties. We have summarized tips on how to design good state pictures for mutual exclusion protocols.

One piece of our future work is to graphically animate state machines that formalize other protocols than mutual exclusion protocols with SMGA or a further revised version of SMGA, although Alternating Bit Protocol (ABP) has been tackled with SMGA [8]. Another piece of our future work is to formally verify that the Suzuki-Kasami protocol enjoys the three properties by theorem proving as were done in [2]. The current implementation of SMGA is completely independent from Maude, which is pros and cons. The pros is that SMGA can play graphical animations of state sequences from another tool, such as Java Pathfinder, provided that state sequence formats conforms to SMGA. The cons is that state sequences needs to be generated in advance. It would be preferable to integrate SMGA with Maude so that state sequences generated by Maude can be visualized on-the-fly, which is yet another future work of ours.

## Acknowledgment

The authors would like to thank the anonymous reviewers who carefully read an earlier version of the paper and gave them valuable comments without which they were not able to complete the present paper.

## References

- [1] Arcaini, P., Bonfanti, S., Gargantini, A., Riccobene, E., 2016. Visual notation and patterns for abstract state machines, in: STAF 2016 Collocated Workshops, Revised Selected Papers, Springer. pp. 163–178. doi:[10.1007/978-3-319-50230-4\\_12](https://doi.org/10.1007/978-3-319-50230-4_12).
- [2] Aung, M.T., Nguyen, T.T.T., Ogata, K., 2018. Guessing, model checking and theorem proving of state machine properties – a case study on Qlock. *Intl J. Softw. Eng. & Comput. Sys.* 4, 1–18. doi:[doi:10.15282/ijsecs.4.2.2018.1.0045](https://doi.org/10.15282/ijsecs.4.2.2018.1.0045).
- [3] Bastian, M., Heymann, S., Jacomy, M., 2009. Gephi: An open source software for exploring and manipulating networks, in: 3rd AAAI ICWSM, pp. 361–362.
- [4] Bui, D.D., Ogata, K., 2019. Graphical animations of the suzuki-kasami distributed mutual exclusion protocol, in: DMSVIVA 2019, KSI Research Inc., pp. 125–134. doi:[10.18293/DMSVIVA2019-012](https://doi.org/10.18293/DMSVIVA2019-012).
- [5] Clavel, M., Durán, F., Eker, S., Lincoln, P., Martí-Oliet, N., Meseguer, J., Talcott, C., 2007. All About Maude. volume 4350 of *LNCS*. Springer.
- [6] Hoxha, B., Bach, H., Abbas, H., Dokhanchi, A., Kobayashi, Y., 2014. Towards formal specification visualization for testing and monitoring of cyber-physical systems, in: DIFTS14.
- [7] Jackson, D., 2012. Software Abstraction. The MIT Press.
- [8] Nguyen, T.T.T., Ogata, K., 2017a. Graphical animations of state machines, in: 15th IEEE DASC, pp. 604–611. doi:[10.1109/DASC-PICom-DataCom-CyberSciTec.2017.107](https://doi.org/10.1109/DASC-PICom-DataCom-CyberSciTec.2017.107).
- [9] Nguyen, T.T.T., Ogata, K., 2017b. Graphically perceiving characteristics of the MCS lock and model checking them, in: 7th Intl Workshop SOFL+MSVL, Springer. pp. 3–23. doi:[10.1007/978-3-319-90104-6\\_1](https://doi.org/10.1007/978-3-319-90104-6_1).
- [10] Ogata, K., Futatsugi, K., 2007. Comparison of Maude and SAL by conducting case studies model checking a distributed algorithm. *IEICE Trans.* 90-A, 1690–1703. doi:[10.1093/ietfec/e90-a.8.1690](https://doi.org/10.1093/ietfec/e90-a.8.1690).
- [11] Santiago, S., Talcott, C.L., Escobar, S., Meadows, C.A., Meseguer, J., 2009. A graphical user interface for Maude-NPA, in: 9th Spanish Conf. Prog. & Lang., pp. 3–20. doi:[10.1016/j.entcs.2009.12.002](https://doi.org/10.1016/j.entcs.2009.12.002).
- [12] Shannon, P., Markiel, A., Ozier, O., Baliga, N.S., Wang, J.T., Ramage, D., Amin, N., Schwikowski, B., Ideker, T., 2003. Cytoscape: A software environment for integrated models of biomolecular interaction networks. *Genome Res.* 13, 2498–2504. doi:[10.1101/gr.1239303](https://doi.org/10.1101/gr.1239303).
- [13] Sun, J., Liu, Y., Dong, J.S., Pang, J., 2009. PAT: Towards flexible verification under fairness, in: 21st CAV, Springer. pp. 709–714.
- [14] Suzuki, I., Kasami, T., 1985. A distributed mutual exclusion algorithm. *ACM TOCS* 3, 344–349. doi:[10.1145/6110.214406](https://doi.org/10.1145/6110.214406).
- [15] Tikhonova, U., Manders, M., Boudewijns, R., 2016. Visualization of formal specifications for understanding and debugging an industrial DSL, in: STAF 2016 Collocated Workshops, Revised Selected Papers, Springer. pp. 179–195. doi:[10.1007/978-3-319-50230-4\\_13](https://doi.org/10.1007/978-3-319-50230-4_13).



# Journal of Visual Language and Computing

journal homepage: [www.ksiresearch.org/jvlc](http://www.ksiresearch.org/jvlc)

## An Empirical Analysis on Effectiveness of Source Code Metrics for Models Predicting Aging Related Bug<sup>\*,\*\*</sup>

Chinmay Hota<sup>a,1</sup>, Lov Kumar<sup>a,\*,2</sup> and Lalita Bhanu Murthy Neti<sup>a,\*\*</sup>

<sup>a</sup>BITS Pilani Hyderabad Campus, India

### ARTICLE INFO

#### Article History:

Submitted 4.30.2019

Revised 5.19.2019

Second Revision 7.29.2019

Accepted 10.10.2019

#### Keywords:

Aging

Software Engineering

Source Code Metrics

Machine Learning

Feature selection

### ABSTRACT

Reliability of any software plays a very important role to assess the quality of the software systems. Reliability analysis and evaluation during early phases of software development life cycle (SDLC) significantly help developer and analysts in the proper allocation of limited resources during testing and maintenance process. The goal of this work is to develop one model using the measurement of the internal structure of the software system i.e., source code metrics for predicting the reason of failure in software during continuously running for a certain time i.e., software aging. Aging-Related Bugs are related with failure during the execution of the software, that leads to degradation in quality, system crashing, misuse of resources, etc.. In this paper, seven different sets of software metrics, seven model training algorithms, one data sampling technique have been empirically investigated and evaluated for predicting aging-related bugs classes. The trained aging-related bugs prediction models are validated using 5-fold cross-validation techniques. The final observation of this experiment work is assessed over seven open source application software systems. The high-value of performance parameters confirm the predicting capability of data sampling, sets of metrics, and training algorithms to predict aging-related bugs classes.

## 1. Introduction

Today's scenario, the majority of the software corporation has it in mind to apply the Object-Oriented (OO) tools for construction the contemporary software system due to its efficient blueprint characteristic such as reusability (extending the code use again) and reducing vulnerability, which facilitates more rapid product development [3][5]. With the help of these features, we can achieve the OO properties like-Inheritance, data abstraction, polymorphism, etc.. Software development or solutions of business requirement using OO programming expect high-quality such reliability, maintainability, reusability, portability, etc.. Measuring quality of software and understanding their relationship with software metrics can help tester, architect, and developer control and estimate quality of software by analyzing the source code. The experimental study presented in the paper is motivated


by the need to correlate and investigate aging-related bugs work and source code metrics. Aging-related bugs are related with failure during the execution of the software, that leads to degradation in quality, system crashing, misuse of resources, etc.. The effort and cost of fixing aging-related bugs increase exponentially if these bugs are not encountered before the deployment of the software [4][10]. The aging-related bugs prediction models can be applied in the early phase of SDLC. Further, their use can improve the quality of the software as well as reduces the maintenance and testing cost. The experimental study presented in this paper is to develop one model for automatic detection of aging-related bugs work. In particular, our motivation is to investigate the application of different feature selection techniques, classification techniques, and data imbalanced techniques for the task of predicting run time failure i.e., aging-related bugs. Thus, the following research questions are addressed in this work:

- **RQ1: What is the capability of different classification techniques to train the models for predicting aging-related bugs?** In this research question, the 5-fold cross-validation technique has been considered to val-

<sup>\*\*</sup> Aging Related Bug Prediction.

<sup>\*</sup>Corresponding author

<sup>\*\*</sup>Principal corresponding author

 f20160069@hyderabad.bits-pilani.ac.in (C. Hota);

lovkumar@hyderabad.bits-pilani.ac.in (L. Kumar);

bhanu@hyderabad.bits-pilani.ac.in (L.B.M. Neti)

ORCID(s):

**Table 1**  
Experimental Dataset Description

IDName	Total classes	non-aging classes	% non-aging classes	aging classes	% aging classes
P1Linux driver net	2292	2283	99.61	9	0.39
P2Linux ext3	29	24	82.76	5	17.24
P3Linux driver scsi	962	958	99.58	4	0.42
P4Linux ipv4	117	115	98.29	2	1.71
P5MySQL optimizer	36	33	91.67	3	8.33
P6MySQL replication	32	28	87.5	4	12.5
P7MySQL innodb	402	370	92.04	32	7.96

idate the models developed using seven different classification techniques. The capability and significance of all considered classification techniques are assessed based on the seven different datasets using rank-sum test and AUC values.

- **RQ2: What is the effect of different sets of metrics for training the models using classification techniques?**  
In this research question, seven different sets of features extracted from source code metrics i.e., software metrics are used as an input to train the aging-related bug prediction models. The capability and significance of these sets of features are computed and compared using AUC values on the seven different datasets.
- **RQ3: What is the capability of data sampling technique to train the models for predicting aging-related bugs?**

In this research question, aging-related bug prediction models are trained using balance data and imbalance data. The capability and significance of the data sampling techniques are computed and compared using AUC values on the seven different datasets.

## 2. Related Work

Grottke et al. [4] investigated that aging-related bugs caused software failures have been observed in numerous events and operating conditions, where it has caused major losses with respect to hazards, the human being as well as money.

Junjun Zheng et al. [10] proposed six number of rejuvenation rules, which are particularly meet with the time-based constraints. They demonstrated a systematic study to measure those six software rejuvenation rules numerically and illustrated that the planned rejuvenation rules often termed as wait-time rules are better to the other Markovian arrival processes (MAPs). In practice eliminating entire bugs is even impractical that eventually makes a system probable of being aged after some time. Therefore, efforts must have been made to detect aging-related bugs from software systems to enable swift aging-related bugs detection and to avoid future system-failures. Both bug identification and prediction are the challenging tasks of the developer.

Padhy et al. [8] developed a framework called re-usability optimization based aging-related bugs prediction system, which

will estimate the re-usability level. Their focus was on re-usability risk management, where they argued that excessive reused system enters into aging-related bugs. Rejuvenation is one of the practical approaches to avoid failures, which causes the software aging-related bugs. Generally, those aging-related bugs are hard to eliminate completely in the software development as well as testing phase. These types of bugs are often known as a memory leak. The process of handling the software aging-related bugs play the key role and proactive management is required to handle such kind issues.

## 3. RESEARCH BACKGROUND

In this section, the details regarding dataset, sampling technique, techniques used for selecting significant uncorrelated predictors, ELM with three kernels have been summarized.

### 3.1. Experimental Dataset

In this experiment, the empirical results are computed over seven large-scale software applications. These software are mainly used for the development of mission-critical, distributed applications, near real-time applications. This software come in the category of large-scale software used in real-world applications, including safety-critical and business contexts. The data has been collected using github Repository<sup>1</sup>. The github Repository website is a very well-known data repository consisting of various research datasets related to software engineering on defects, code analysis, effort, bugs, maintainability, refactoring, and test generation. The primary objective of using these datasets to replicated our experiment finding, so that these experimental findings are used by other researchers for benchmarking and comparison. In this paper, experiments are conducting on a dataset containing information of aging-related bugs. The dataset contains various types of metrics such as program size related metrics, Halstead metrics, McCabe's Cyclomatic complexity, and aging-related metrics [7]. The characteristics of these datasets are given in Table 1. The data has been collected for seven different projects (Four projects: Linux sub-systems and three projects: MySQL sub-systems). From Table 1, it can be clearly seen that Linux driver net is the largest project containing 2292 classes and Linux ext3 is the

<sup>1</sup><https://github.com/lov505/Aging-data>

**Table 2**  
Software Metrics

Types of Metrics Sets	Metrics List
Program size metrics	AltAvgLineCode, AltAvgLineComment, AltCountLineCode, AvgCyclomatic, AvgCyclomaticModified, AvgCyclomaticStrict, AvgEssential, AvgLineComment, CountClassCoupled, CountDeclClassMethod, CountDeclClassVariable, CountDeclFunction, CountDeclInstanceMethod, CountDeclInstanceVariablePrivate, CountDeclMethod, CountDeclMethodPrivate, CountDeclMethodProtected, CountDeclMethodPublic, CountInput, CountLine, CountLineCode, CountLineCodeDecl, CountLineCodeExe, CountLineComment, CountLineInactive, CountLinePreprocessor, CountSemicolon, CountStmt, CountStmtDecl, CountStmtEmpty, AltAvgLineBlank, AltCountLineBlank, AltCountLineComment, AvgLine, AvgLineBlank, AvgLineCode, CountClassBase, CountClassDerived, CountDeclClass, CountDeclInstanceVariable, CountDeclInstanceVariableProtected, CountDeclInstanceVariablePublic, CountDeclMethodAll, CountDeclMethodConst, CountDeclMethodFriend, CountLineBlank, CountOutput, CountPath, CountStmtExe
MCC metrics	CyclomaticStrict, Essential, Knots, MaxCyclomatic, MaxCyclomaticModified, MaxEssentialKnots, MaxInheritanceTree, MaxNesting, MinEssentialKnots, SumCyclomatic, SumCyclomaticModified, SumCyclomaticStrict, Cyclomatic, CyclomaticModified, MaxCyclomaticStrict, PercentLackOfCohesion, RatioCommentToCode, SumEssential
Halstead metrics	Number of distinct operands, Program length, Difficulty, Number of distinct operators, Effort, Total number of operands, Volume, Program vocabulary, Total number of operators
Aging metrics	DerefUse, UniqueDerefSet, DeallocOps, AllocOps, DerefSet, UniqueDerefUse

smallest project containing 29 classes among all projects. It is also observed from Table 1 that are considered data are highly imbalanced because % of aging-related classes varies from 0.39% to 17.24%.

### 3.2. Software Metrics

The features extracted from source code called software metrics used to measure the size and complexity of source code. These extracted features are considered as an input to train the models for predicting aging-related bugs. In this work, we have used 49 metrics related to program size, 18 metrics related to McCabe's Cyclomatic complexity, 9 metrics related to Halstead metrics, and 6 metrics related to aging-related bugs as shown in Table 2 for measuring size and complexity of source code. The abbreviation and details explanation of these metrics sets are explained with practical applications in [1][6][2]<sup>2</sup>.

### 3.3. Feature Selection Techniques

The next step of this work is to apply feature selection techniques on feature extracted from source code metrics for finding the right sets of uncorrelated significant features. In this step, some important features are selected from all considered features and these selected features are considered by training algorithms to train the models for predicting aging-related bugs. In this paper, we have used univariate rank-sum test and logistic regression to remove insignificant predictor for aging-related bugs. Ramsum test is used to test null hypothesis i.e., no statistical significance between the values of a feature of the class having aging-related bugs and the values of a feature of the class having non-aging-related bugs. The considered null hypothesis has been tested at 0.05 p-value i.e.,  $p\text{-value} \leq 0.05$ : hypothesis rejected means metrics is relevant for aging-related bugs prediction and  $p\text{-value} > 0.05$ : hypothesis accepted means metrics is irrelevant for aging-related bugs prediction. After removing

all insignificant predictors, we are using cross-correlation analysis to remove highly correlated features and finally select significant uncorrelated features. The primary objective of the above step is to select independent significant features having a high correlation with aging-related bugs.

### 3.4. Classification Algorithms:

In this research, we are using extreme learning machine (ELM) algorithm with different kernels functions to train the aging-related bugs prediction models by considering features extracted from source code metrics as input. ELM is based on the concept of randomly generated hidden neurons with an objective to reduce the learning time of the models [9]. ELM algorithm overcomes various problems such as improper learning rate and overfitting, a problem due to local minima, backpropagation. Finally, the trained models using ELM for predicting aging-related bugs are compared with widely used classifiers using AUC value and accuracy.

### 3.5. Data Sampling Technique to handle Imbalanced Data:

Models trained using imbalanced data are the common problem in many domains related to data mining and machine learning such as medical diagnosis, defect prediction, fraud detection etc.. A data is called imbalanced if it has not equal distributions of classes. These types of data sets may be affected the actual performance of models. It can be seen from Table 1 that the percentage of aging-related bugs classes present is considered datasets varies from 0.39% to 17.24 %. Since, the percentage of aging-related bugs classes are not equal to percentage of non-aging-related bugs classes, so models developed using these datasets suffered from class imbalance problem. So, in this work, upscale sampling technique has been considered to deal with class imbalanced class problem. The performance of the developed models using upscale data are computed using accuracy and AUC values and compared with original data.

<sup>2</sup>[https://github.com/sealuzh/user\\_quality/wiki/Code-Quality-Metrics](https://github.com/sealuzh/user_quality/wiki/Code-Quality-Metrics)



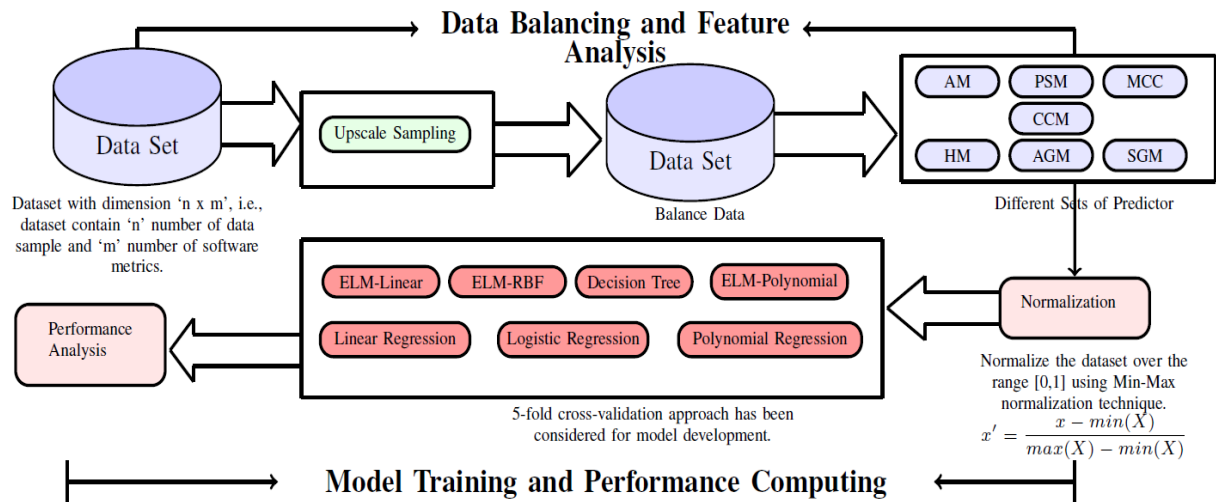


Figure 1: Framework of proposed work

### 3.6. Performance Parameter

In order to measure the quality of aging-related bugs prediction models, various performance metrics need to be examined which indicated the effectiveness of the trained models. In this work, we have considered two different performance metrics such as Accuracy (%) and area under curve value (AUC) for evaluating the quality of aging-related bugs prediction models. A high value of these metrics indicates that the built prediction model is good.

## 4. Research Framework and Experimental Results

This section presents the experimental framework used for developing aging-related bugs prediction models using different sets of source code metrics. Figure 1 shows the empirical framework consisting various steps for development of aging-related bugs prediction model.

From Figure 1, it can be seen that five major steps are used in the proposed framework. The five steps are: (1) Data balancing using upscale sampling technique (2) Categorization of all considered metrics into different sets and also selection of significant and uncorrelated right sets of metrics (3) Data normalization using Min-Max normalization technique (4) Predictive model trained using different classification techniques (5) Significance and capability evaluation using different performance parameters. The above mentioned steps are applied on different packages of two long-living, complex, and large software systems i.e., MySQL and Linux kernel. In this work, eighty two different metrics such as Halstead metrics, McCabe's cyclomatic complexity, size related metrics, and aging-related bugs metrics have been considered as predictors for aging-related bugs prediction models. From Table 1, we can observed that the considered datasets are suffering from class imbalance problem. So, we have considered upscale sampling technique to address the issue of class imbalance problem. The sec-

ond step consists of Categorization of all considered metrics into different sets i.e., Halstead metrics (HM), McCabe's cyclomatic complexity (MCC), size related metrics (PSM), aging-related bugs metrics (AGM), and selection of right set of significant predictor (SGM) and uncorrelated significant predictor (CCM) using rank-sum test and cross correlation analysis. Then, Min-Max Normalization has applied before training the model to normalize the all selected features in the same range of 0 to 1. Further, seven different classification techniques have been used to train the aging-related bugs prediction model. These models are validated using 5-fold cross-validation approach. Finally, accuracy, and AUC values are computed for each model to evaluate capability and significance of above considered classification techniques, data sampling technique, and different sets of metrics.

### 4.1. Feature Selection Techniques:

In this section, we have summarized the results obtained after rank-sum test on original data and upscale sampling data for different projects. In this study, file having at least one aging-related bugs bugs considered as aging-related bugs classes else non-aging-related bugs classes. Table 3 provides the relevant metrics related program size metrics after rank-sum test for two projects on without sampling data i.e., original data and upscale data. Similarly, the relevant features for other metrics sets are computed using rank-sum test. In Table 3, only 0 and 1 are used to represent relevant features and irrelevant metrics i.e., cell containing 1 denotes the relevant metrics for project using original data or upscale data and cell containing 0 denotes the irrelevant metrics for project using original data or upscale data. It can be observed from Table 3 that AltAvgLineBlank is relevant for all cases except scsi, ext3, and ipv4 projects. Similarly, It can also observed from Table 3 that CountClassBase, CountClassCoupled, CountClassDerived, CountDeclClassMethod, CountDeclClassVariable etc. are irrelevant for aging-related bugs

**Table 3**  
Rank-sum Test for Program Size Metrics

	Original Data							Upscale Sampling						
	Linux				MySQL			Linux				MySQL		
	net	scsi	ext3	ipv4	innodb	optimizer	replication	net	scsi	ext3	ipv4	innodb	optimizer	replication
AltAvgLineBlank	1	0	0	0	1	1	1	1	1	1	1	1	1	1
AltAvgLineCode	1	0	0	0	1	1	0	1	1	1	1	1	1	1
AltAvgLineComment	1	1	0	0	1	1	1	1	1	1	1	1	1	1
AltCountLineBlank	1	0	1	0	1	1	1	1	1	1	1	1	1	1
AltCountLineCode	1	0	1	0	1	1	1	1	1	1	1	1	1	1
AltCountLineComment	1	0	1	0	1	1	1	1	1	1	1	1	1	1
AvgCyclomatic	1	1	0	0	1	1	0	1	1	1	0	1	1	1
AvgCyclomaticModified	1	1	0	0	1	1	1	1	1	1	0	1	1	1
AvgCyclomaticStrict	1	1	0	0	1	1	0	1	1	1	0	1	1	1
AvgEssential	0	1	0	0	1	1	0	1	1	1	0	1	1	1
AvgLine	1	1	0	0	1	1	0	1	1	1	1	1	1	1
AvgLineBlank	1	0	0	0	1	1	1	1	1	1	1	1	1	1
AvgLineCode	1	0	0	0	1	1	0	1	1	1	1	1	1	1
AvgLineComment	1	1	0	0	1	1	1	1	1	1	1	1	1	1
CountClassBase	0	0	0	0	0	0	0	0	0	0	0	0	0	0
CountClassCoupled	0	0	0	0	0	0	0	0	0	0	0	0	0	0
CountClassDerived	0	0	0	0	0	0	0	0	0	0	0	0	0	0
CountDeclClass	0	0	0	0	1	0	0	0	0	0	0	1	1	1
CountDeclClassMethod	0	0	0	0	0	0	0	0	0	0	0	0	0	0
CountDeclClassVariable	0	0	0	0	0	0	0	0	0	0	0	0	0	0
CountDeclFunction	1	0	1	0	1	0	1	1	1	1	1	1	0	1
CountDeclInstanceMethod	0	0	0	0	0	0	0	0	0	0	0	0	0	0
CountDeclInstanceVariable	0	0	0	0	0	0	0	0	0	0	0	0	0	0
CountDeclInstanceVariablePrivate	0	0	0	0	0	0	0	0	0	0	0	0	0	0
CountDeclInstanceVariableProtected	0	0	0	0	0	0	0	0	0	0	0	0	0	0
CountDeclInstanceVariablePublic	0	0	0	0	0	0	0	0	0	0	0	0	0	0
CountDeclMethod	0	0	0	0	0	0	0	0	0	0	0	0	0	0
CountDeclMethodAll	0	0	0	0	0	0	0	0	0	0	0	0	0	0
CountDeclMethodConst	0	0	0	0	0	0	0	0	0	0	0	0	0	0
CountDeclMethodFriend	0	0	0	0	0	0	0	0	0	0	0	0	0	0
CountDeclMethodPrivate	0	0	0	0	0	0	0	0	0	0	0	0	0	0
CountDeclMethodProtected	0	0	0	0	0	0	0	0	0	0	0	0	0	0
CountDeclMethodPublic	0	0	0	0	0	0	0	0	0	0	0	0	0	0
CountInput	0	0	0	0	0	0	0	0	0	0	0	0	0	0
CountLine	1	0	1	0	1	1	1	1	1	1	1	1	1	1
CountLineBlank	1	0	1	0	1	1	1	1	1	1	1	1	1	1
CountLineCode	1	0	1	0	1	1	1	1	1	1	1	1	1	1
CountLineCodeDecl	1	0	1	0	1	0	1	1	1	1	1	1	1	1
CountLineCodeExe	1	0	1	0	1	1	1	1	1	1	1	1	1	1
CountLineComment	1	1	1	0	1	1	1	1	1	1	1	1	1	1
CountLineInactive	1	0	0	0	1	0	1	1	1	1	1	1	1	1
CountLinePreprocessor	1	0	0	0	1	0	1	1	1	1	1	1	1	1
CountOutput	0	0	0	0	0	0	0	0	0	0	0	0	0	0
CountPath	0	0	0	0	0	0	0	0	0	0	0	0	0	0
CountSemicolon	1	0	1	0	1	1	1	1	1	1	1	1	1	1
CountStmt	1	0	1	0	1	1	1	1	1	1	1	1	1	1
CountStmtDecl	1	0	1	0	1	0	1	1	1	1	1	1	1	1
CountStmtEmpty	1	1	0	1	1	1	1	1	1	1	1	1	1	1
CountStmtExe	1	1	1	0	1	1	1	1	1	1	1	1	1	1

prediction i.e., these metrics are not considered as input to train the aging-related bugs prediction models.

## 4.2. Results and analysis

We have considered seven sets of source code metrics as input to train the model using seven different classification techniques to predict aging-related bugs. Five fold cross validation method has been used to validate the trained models

**Table 4**  
Accuracy and AUC values for Linux net and Scsi

	Without Sampling (net)													
	Accuracy							AUC						
	LR	PR	LOR	DT	ELMLIN	ELMPY	ELMRBF	LR	PR	LOR	DT	ELMLIN	ELMPY	ELMRBF
Fold1	98.69	94.32	98.25	99.78	99.78	93.45	99.78	0.49	0.47	0.49	0.50	0.50	0.47	0.50
Fold2	98.47	95.64	98.47	99.56	99.35	76.47	99.56	0.49	0.73	0.49	0.50	0.50	0.63	0.50
Fold3	97.82	89.11	98.91	98.69	98.91	88.24	99.56	0.74	0.70	0.50	0.50	0.50	0.69	0.50
Fold4	98.69	93.23	98.47	99.56	99.56	89.52	99.56	0.99	0.97	0.74	0.50	0.50	0.45	0.50
Fold5	97.16	92.14	97.82	99.13	99.13	96.94	99.56	0.74	0.96	0.49	0.50	0.50	0.49	0.50
Upscale Sampling (net)														
	LR	PR	LOR	DT	ELMLIN	ELMPY	ELMRBF	LR	PR	LOR	DT	ELMLIN	ELMPY	ELMRBF
Fold1	50.00	50.00	98.47	99.78	97.16	64.11	50.00	0.50	0.50	0.98	1.00	0.97	0.64	0.50
Fold2	49.95	49.95	98.47	99.56	95.84	54.87	49.95	0.50	0.50	0.98	1.00	0.96	0.55	0.50
Fold3	50.00	50.00	99.02	99.34	96.72	67.72	50.00	0.50	0.50	0.99	0.99	0.97	0.68	0.50
Fold4	50.00	50.00	98.57	99.23	95.18	69.30	50.00	0.50	0.50	0.99	0.99	0.95	0.69	0.50
Fold5	50.05	50.05	98.69	99.67	95.51	68.02	50.05	0.50	0.50	0.99	1.00	0.96	0.68	0.50
Without Sampling (scsi)														
	LR	PR	LOR	DT	ELMLIN	ELMPY	ELMRBF	LR	PR	LOR	DT	ELMLIN	ELMPY	ELMRBF
Fold1	95.31	76.04	97.40	100.00	99.48	82.81	100.00	0.50	0.50	0.50	0.50	0.50	0.50	0.50
Fold2	92.75	76.68	97.93	98.96	98.96	37.31	99.48	0.47	0.88	0.49	0.99	0.50	0.68	0.50
Fold3	93.75	86.98	97.40	99.48	98.44	86.46	99.48	0.47	0.44	0.49	0.50	0.49	0.93	0.50
Fold4	95.85	79.79	98.96	99.48	99.48	95.85	99.48	0.48	0.40	0.50	0.50	0.50	0.48	0.50
Fold5	96.35	93.23	96.88	98.96	97.92	76.04	99.48	0.48	0.47	0.49	0.50	0.49	0.88	0.50
Upscale Sampling (scsi)														
	LR	PR	LOR	DT	ELMLIN	ELMPY	ELMRBF	LR	PR	LOR	DT	ELMLIN	ELMPY	ELMRBF
Fold1	50.00	50.00	94.27	99.22	93.49	56.51	50.00	0.50	0.50	0.94	0.99	0.93	0.57	0.50
Fold2	50.00	50.00	92.93	99.74	95.03	63.35	50.00	0.50	0.50	0.93	1.00	0.95	0.63	0.50
Fold3	50.00	50.00	95.83	99.48	93.49	75.26	50.00	0.50	0.50	0.96	0.99	0.93	0.75	0.50
Fold4	49.87	49.87	93.99	98.69	93.47	53.26	49.87	0.50	0.50	0.94	0.99	0.93	0.53	0.50
Fold5	50.13	50.13	96.34	98.69	92.43	68.41	50.13	0.50	0.50	0.96	0.99	0.92	0.68	0.50

and performance of these trained models are express in terms of area under curve value (AUC) and accuracy. The value of Accuracy and AUC of trained model using different techniques for Linux net project on each fold shown in Table 4. Similarly, the Accuracy and AUC values for other projects are computed. It can be seen from Table 4 that:

- Aging-related bugs prediction models trained using Upscale sampling data have relatively better AUC values as compare to original data.
- Aging-related bugs prediction models trained using logistic regression and decision tree have relatively better AUC values as compare to other techniques.
- Aging-related bugs prediction models trained using ELM with linear kernel function have relatively better AUC values as compare to other kernel functions.

## 5. Comparison

**RQ1: RQ1: what is the overall predictive capability of model trained using different classification techniques to predict aging-related bugs?** In this research question, 5-fold cross validation technique has been considered to validate the models developed using seven different classification techniques. The significance and capability of all considered classification techniques are assessed based on the

seven different datasets using Boxplots, Descriptive Statistics, and Statistical tests analysis.

**Boxplots and Descriptive Statistics: Classification Techniques** In this paper, boxplot diagram has been considered for visual comparison showing distribution, variability, outliers, degree of dispersion of different classification techniques as depicted in Figure 2. The first, second, and third sub-figures of Figure 2 shows the boxplot diagram of AUC value for classification techniques trained using original data, upscale sampling data, and overall. The median value of AUC value for each classification technique has been presented using red line. The descriptive statistics i.e., **Min**, **25%**, **Mean**, **Median**, **75%**, **Max** for each box are presented in Table 5. It can be seen from Table 5 and Figure 2 that aging-related bugs prediction models trained using decision tree (DT) have relatively better AUC values as compare to other techniques. It can be also seen from Table 5 and Figure 2 that ELM with liner kernel on upscale sampling data and over all provided good results as compared to other polynomial kernel and RBF kernels.

**Comparison of different Classification techniques using Statistical tests:** In this work, we have also considered pairwise rank sum test with Bonferroni correction to compare the performance of developed aging-related bugs prediction models using different classification techniques. We have considered seven sets of metrics, and two different datasets for each seven project. Hence, each classification technique

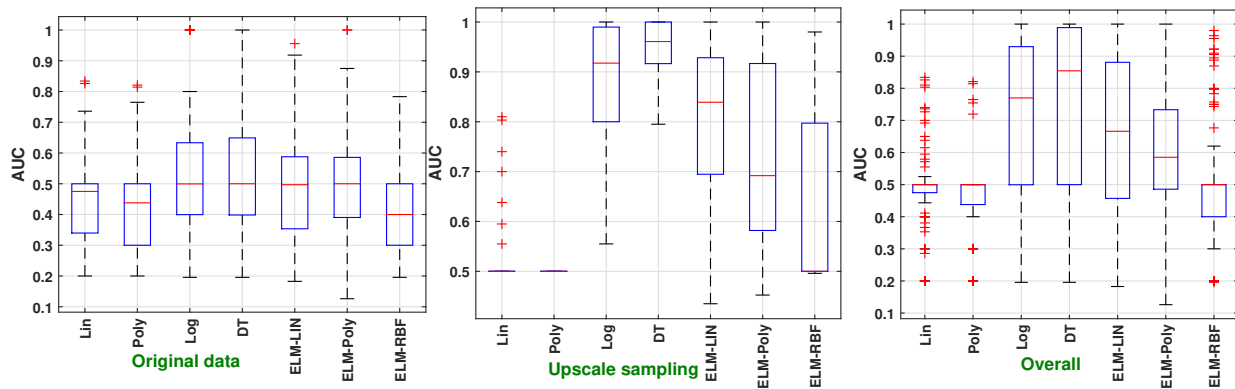


Figure 2: Boxplots for AUC for Classification Techniques

has 490 data points or AUC values [7 sets of features \* (1 original data + 1 upscale sampling data) \* 7 projects \* 5 folds]. The null hypothesis of the rank sum test is that *aging-related bugs prediction models developed using different classification techniques are not significantly different*. The rank sum test with Bonferroni correction adjust the threshold p-

**Table 5**  
Statistical Measures on AUC: Classification Techniques

Original Data						
	Min	Max	Mean	Median	25%	75%
LR	0.20	0.83	0.45	0.48	0.34	0.50
PR	0.20	0.82	0.43	0.44	0.30	0.50
LOR	0.20	1.00	0.54	0.50	0.40	0.63
DT	0.20	1.00	0.55	0.50	0.40	0.65
ELMLIN	0.18	0.96	0.49	0.50	0.35	0.59
ELMPLOY	0.13	1.00	0.50	0.50	0.39	0.59
ELMRBF	0.20	0.78	0.41	0.40	0.30	0.50
Upscale Sampling						
	Min	Max	Mean	Median	25%	75%
LR	0.50	0.81	0.53	0.50	0.50	0.50
PR	0.50	0.50	0.50	0.50	0.50	0.50
LOR	0.56	1.00	0.88	0.92	0.80	0.99
DT	0.80	1.00	0.95	0.96	0.92	1.00
ELMLIN	0.44	1.00	0.81	0.84	0.69	0.93
ELMPLOY	0.45	1.00	0.73	0.69	0.58	0.92
ELMRBF	0.50	0.98	0.63	0.50	0.50	0.80
Overall						
	Min	Max	Mean	Median	25%	75%
LR	0.20	0.83	0.49	0.50	0.48	0.50
PR	0.20	0.82	0.47	0.50	0.44	0.50
LOR	0.20	1.00	0.71	0.77	0.50	0.93
DT	0.20	1.00	0.75	0.85	0.50	0.99
ELMLIN	0.18	1.00	0.65	0.67	0.46	0.88
ELMPLOY	0.13	1.00	0.62	0.59	0.49	0.73
ELMRBF	0.20	0.98	0.52	0.50	0.40	0.50

value at  $\frac{0.05}{\text{number of different pairs}}$ . Since, we have considered seven different classification techniques, so the number of different pairs is  $7c_2 = 7 * 6/2 = 21$ . Hence, null hypothesis is accepted only if the p-value is less than  $\frac{0.05}{21} = 0.0024$ . The results of rank sum test with Bonferroni correction for different classification techniques are summarized in Table 6. We observe from Table 6 that the p-value of 13 out of 21 pairs are less than 0.0024 i.e., hypothesis rejected means model trained using different classification techniques are significantly different.

**Table 6**  
P-value: Classification Techniques

	LR	PR	LOR	DT	ELMLIN	ELMPLOY	ELMRBF
LR	1.00	0.17	0.00	0.00	0.00	0.00	0.98
PR	0.17	1.00	0.00	0.00	0.00	0.00	0.36
LOR	0.00	0.00	1.00	0.23	0.08	0.01	0.00
DT	0.00	0.00	0.23	1.00	0.00	0.00	0.00
ELMLIN	0.00	0.00	0.08	0.00	1.00	0.23	0.00
ELMPLOY	0.00	0.00	0.01	0.00	0.23	1.00	0.00
ELMRBF	0.98	0.36	0.00	0.00	0.00	0.00	1.00

**RQ2: what is the overall predictive capability of model developed using different sets of metrics as input to predict aging-related bugs?** In this paper, seven different sets of metrics are considered as input to developed a model for predicting aging-related bugs. The significance and capability of all considered sets of metrics are assessed based on the seven different datasets using Boxplots, Descriptive Statistics, and Statistical tests analysis.

**Comparison of different sets of metrics using Boxplots and Descriptive Statistics:** In this paper, boxplot diagram has been considered for visual comparison showing distribution, variability, outliers, degree of dispersion of different sets of metrics as depicted in Figure 3. The first, second, and third sub-figures of Figure 3 shows the boxplot diagram of AUC value for classification techniques trained using original data, upscale sampling data, and overall. The median value of AUC value for each classification technique

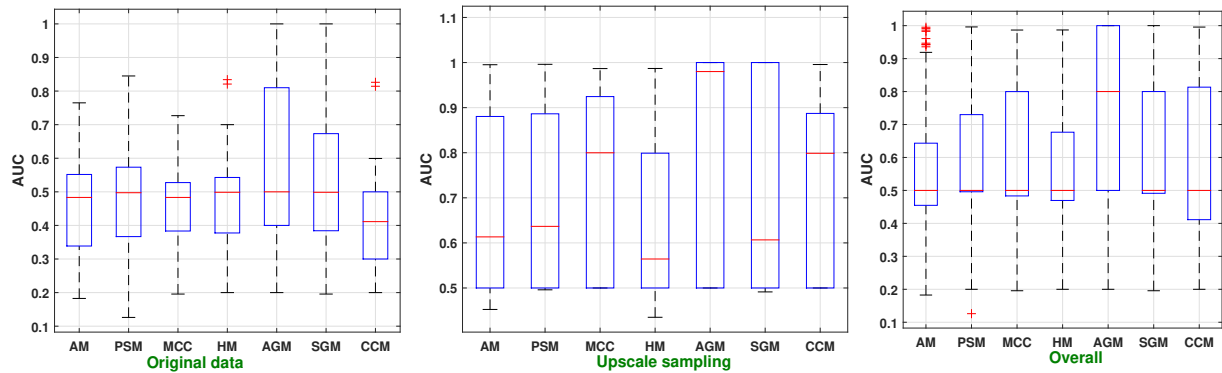


Figure 3: Boxplots for AUC for Sampling Technique

has been presented using red line. The descriptive statistics i.e., **Min**, **25%**, **Mean**, **Median**, **75%**, **Max** for each box are presented in Table 7. It can be seen from Table 7 and Figure 3 that aging-related bugs prediction models developed using aging-related bugs (AGM) metrics as input have relatively better AUC values as compare to other sets of metrics. It can be also seen from Table 7 and Figure 3 that significant sets of features (SGM) provided results as compared to other sets of metrics except AGM.

*Comparison of different sets of metrics using Statistical tests:* In this work, we have also considered pairwise rank sum test with Bonferroni correction to compare the performance of developed aging-related bugs prediction models using different sets of metrics. We have considered seven different classification techniques, and two different datasets for each seven project. Hence, each each set of metrics has 490 data points or AUC values [7 classification techniques \* (1 original data + 1 upscale sampling data) \* 7 projects \* 5 folds]. The null hypothesis of the rank sum test is that *aging-related bugs prediction models developed using different sets of metrics are not significantly different*. The rank sum test with Bonferroni correction adjust the threshold p-value at  $\frac{0.05}{\text{number of different pairs}}$ . Since, we have consid-

ered seven sets of metrics, so the number of different pairs is  $7c_2 = 7 * 6/2 = 21$ . Hence, null hypothesis is accepted only if the p-value is less then  $\frac{0.05}{21} = 0.0024$ . The results of rank sum test with Bonferroni correction for sets of metrics are summarized in Table 8. We observe from Table 8 that the p-value of the models trained using aging-related bugs metrics are less than 0.0024 i.e., hypothesis rejected means model trained using aging-related bugs metrics and other sets of metrics are significantly different.

**RQ3: what is the overall predictive capability of model developed using balance and imbalance data to predict aging-related bugs?**

In this work, aging-related bugs prediction models are trained using balance data and imbalance data. The significance and capability of the data sampling used for balancing data is assessed based on the seven different datasets using rank-sum test and AUC values. .

Table 7

Statistical Measures on AUC: Sets of Metrics.

	Original Data					
	Min	Max	Mean	Median	25%	75%
AM	0.18	0.76	0.45	0.48	0.34	0.55
PSM	0.13	0.85	0.46	0.50	0.37	0.57
MCC	0.20	0.73	0.44	0.48	0.38	0.53
HM	0.20	0.83	0.46	0.50	0.38	0.54
AGM	0.20	1.00	0.61	0.50	0.40	0.81
SGM	0.20	1.00	0.52	0.50	0.38	0.67
CCM	0.20	0.83	0.42	0.41	0.30	0.50
	Upscale Sampling					
	Min	Max	Mean	Median	25%	75%
AM	0.45	1.00	0.67	0.61	0.50	0.88
PSM	0.50	1.00	0.69	0.64	0.50	0.89
MCC	0.50	0.99	0.75	0.80	0.50	0.92
HM	0.44	0.99	0.64	0.56	0.50	0.80
AGM	0.50	1.00	0.83	0.98	0.50	1.00
SGM	0.49	1.00	0.70	0.61	0.50	1.00
CCM	0.50	1.00	0.74	0.80	0.50	0.89
	Overall					
	Min	Max	Mean	Median	25%	75%
AM	0.18	1.00	0.56	0.50	0.45	0.64
PSM	0.13	1.00	0.58	0.50	0.50	0.73
MCC	0.20	0.99	0.60	0.50	0.48	0.80
HM	0.20	0.99	0.55	0.50	0.47	0.68
AGM	0.20	1.00	0.72	0.80	0.50	1.00
SGM	0.20	1.00	0.61	0.50	0.49	0.80
CCM	0.20	1.00	0.58	0.50	0.41	0.81

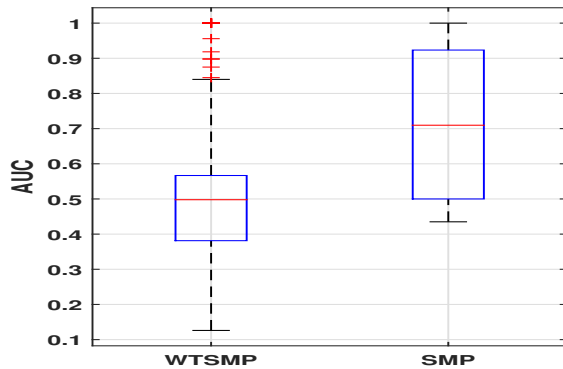
*Comparison of different sampling techniques using Box-plots and Descriptive Statistics:* In this paper, boxplot diagram has been considered for visual comparison showing distribution, variability, outliers, degree of dispersion of original data and upscale sampling data as depicted in Figure 4. The median value of AUC value for each data set has been presented using red line. The descriptive statistics i.e., **Min**, **25%**, **Mean**, **Median**, **75%**, **Max** for each box are presented



**Table 8**

P-value: Sets of Metrics

	AM	PSM	MCC	HM	AGM	SGM	CCM
AM	1.00	0.67	0.41	0.87	0.00	0.17	0.73
PSM	0.67	1.00	0.65	0.52	0.00	0.33	0.96
MCC	0.41	0.65	1.00	0.26	0.00	0.51	0.59
HM	0.87	0.52	0.26	1.00	0.00	0.15	0.59
AGM	0.00	0.00	0.00	0.00	1.00	0.01	0.00
SGM	0.17	0.33	0.51	0.15	0.01	1.00	0.31
CCM	0.73	0.96	0.59	0.59	0.00	0.31	1.00

**Figure 4:** Boxplots for AUC

in Table 9. It can be seen from Table 9 and Figure 4 that aging-related bugs prediction models developed using sampling data i.e., balance data have relatively better AUC values as compare to original data. *Comparison of different*

**Table 9**

Statistical Measures on AUC: Original data and Upscale Data

	Min	Max	Mean	Median	25%	75%
WTSMP	0.13	1.00	0.48	0.50	0.38	0.57
SMP	0.44	1.00	0.72	0.71	0.50	0.92

*sampling techniques using Statistical tests:* In this work, we have also considered pairwise rank sum test with Bonferroni correction to compare the performance of developed aging-related bugs prediction models using original data and sampled data. We have considered seven different classification techniques, seven sets of metrics, and seven project. Hence, each data set has 1715 data points or AUC values [7 classification techniques \* 7 sets of features \* 7 projects \* 5 folds]. The null hypothesis of the rank sum test is that *aging-related bugs prediction models developed using original data and sampled data are not significantly different*. The rank sum test with Bonferroni correction adjust the threshold p-value at  $\frac{0.05}{\text{number of different pairs}}$ . Since, we have considered two datasets, so the number of different pairs is  $2c_2 = 1$ . Hence, null hypothesis is accepted only if the p-value is less than  $\frac{0.05}{1} = 0.05$ . The results of rank sum test with Bonferroni correction for sets of metrics are summarized in Table 10.

We observe from Table 10 that the p-value of the models trained using original data and sampled data are significantly different

**Table 10**

P-value: Original data and Upscale Data

	WTSMP	SMP
WTSMP	1.00	0.00
SMP	0.00	1.00

## 6. Conclusion

The primary goal of this research is to find the effect of internal structure of software on aging-related bugs. In this experiment, we also empirically evaluated, analyzed, and compared the performance of developed aging-related bugs prediction models using different classification techniques, sets of metrics, selected set of metrics, and data sampling technique. The major conclusions of this study are summarized as follows.

- The aging-related bugs prediction models trained using decision tree (DT) have relatively better AUC values as compare to other techniques.
- The aging-related bugs prediction models developed using aging-related bugs (AGM) metrics as input have relatively better AUC values as compare to other sets of metrics.
- Upscale sampling technique provides relatively better performance in detecting aging-related bugs compared to the models built by original data.

## References

- [1] Aggarwal, K., Singh, Y., Kaur, A., Malhotra, R., . Empirical analysis for investigating the effect of object-oriented metrics on fault proneness: a replicated case study. *Software process: Improvement and practice* .
- [2] Chidamber, S.R., Darcy, D.P., Kemerer, C.F., 1998. Managerial use of metrics for Object-Oriented software: An exploratory analysis. *IEEE Transactions on Software Engineering* 24, 629–639.
- [3] Chidamber, S.R., Kemerer, C.F., . Towards a metrics suite for Object-Oriented design. volume 26. *ACM*.
- [4] Grottke, M., Li, L., Vaidyanathan, K., Trivedi, K.S., 2006. Analysis of software aging in a web server. *IEEE Transactions on reliability* 55, 411–420.
- [5] Malhotra, R., Chug, A., 2014. Application of group method of data handling model for software maintainability prediction using object oriented systems. *International Journal of System Assurance Engineering and Management* 5, 165–173.
- [6] Malhotra, R., Singh, Y., . On the applicability of machine learning techniques for object oriented software fault prediction. *Software Engineering: An International Journal* .
- [7] Natella, R., 2017. Aging-related bugs and software complexity metrics Aging-related bugs and software complexity metrics. URL: <https://doi.org/10.5281/zenodo.581659>, doi:10.5281/zenodo.581659.
- [8] Padhy, N., Singh, R., Satapathy, S.C., 2017. Enhanced evolutionary computing based artificial intelligence model for web-solutions software reusability estimation. *Cluster Computing* , 1–18.

- [9] Sun, C., Wang, B., Liu, Y., Sun, D., Chen, D., 2016. Numerical simulations for near-field acoustic holographic data extrapolation based on the neural network elm method, in: *Information Science and Electronic Engineering: Proceedings of the 3rd International Conference of Electronic Engineering and Information Science (ICEEIS 2016)*, January 4-5, 2016, Harbin, China, CRC Press. p. 87.
- [10] Zheng, J., Okamura, H., Li, L., Dohi, T., 2017. A comprehensive evaluation of software rejuvenation policies for transaction systems with markovian arrivals. *IEEE Transactions on Reliability* 66, 1157–1177.

# Journal of Visual Language and Computing

journal homepage: [www.ksiresearch.org/jvlc](http://www.ksiresearch.org/jvlc)

## Text entry on circular smartwatches: the C-QWERTY layout

Gennaro Costagliola<sup>a,\*</sup>, Raffaele D'Arco<sup>b</sup>, Sabato De Gregorio<sup>b</sup>, Mattia De Rosa<sup>a,\*\*</sup>, Vittorio Fuccella<sup>a,\*</sup> and Daniele Lupo<sup>b</sup>

<sup>a</sup>Dipartimento di Informatica, University of Salerno, Via Giovanni Paolo II, 84084 Fisciano (SA), Italy

<sup>b</sup>University of Salerno, Via Giovanni Paolo II, 84084 Fisciano (SA), Italy

### ARTICLE INFO

#### Article History:

Submitted 4.18.2019

Revised 5.20.2019

Second Revision 8.1.2019

Accepted 10.9.2019

#### Keywords:

Text entry

Smartwatch

Gestures

Touchscreen

### Abstract

We present a study aimed at comparing different circular layouts for entering text on smartwatches. In particular, we measured the extent through which the use of a QWERTY layout increases user performance in the earliest sessions of use. To this aim, we designed C-QWERTY, a soft keyboard designed for circular smartwatches in which the keys are arranged along the edge of the screen in a circular layout. In order to make the keyboard more familiar to users, the order of the keys is similar to the one of traditional rectangular QWERTY keyboards. The method supports two interaction modes: tapping and gesture, in which a whole word can be written with a single gesture. As an evaluation, we compared the C-QWERTY layout with Cirrin, another circular layout in which the order of the characters has been optimized to minimize the distance between successive keys during writing. The experimental results showed that the C-QWERTY layout outperformed the Cirrin one with respect to the text entry speed with an average increase of 27 percent.

© 2019 KSI Research

## 1. Introduction

Smartwatches have recently emerged as new wearable computing devices, but the small size of the screen may however, in some cases, make it difficult to interact with them. One challenge in this area is text entry, and researchers are introducing new text entry techniques for smartwatches in order to increase efficiency [1, 6].

Interestingly, many of them are adopting the QWERTY layout. This is probably done to exploit users' previous knowledge, as most of them use the QWERTY keyboard habitually, and to avoid that users have to invest time to learn a new layout. However, when the full QWERTY layout is fitted on a smartwatch screen the keys become very small. To solve this problem various techniques have been developed, as described in the next section.

In any case, a rectangular layout is more suitable for screens of the same type, while in a circular display, as in


smartwatches, it can waste a lot of screen space. In this case a circular layout could allow a more efficient use of the screen space, but it is obviously very different from the QWERTY layout already known by users.

The purpose of this paper is to present a study on circular layouts for text entry on smartwatches. For this purpose we have decided to measure how much a QWERTY-like circular layout can improve user performance, especially at the first sessions of use. For this reason we have designed C-QWERTY, a soft keyboard for circular smartwatches in which the keys are arranged along the edge of the screen in a circular layout, and the keys have similar order of the ones of a traditional rectangular QWERTY keyboards. C-QWERTY also supports two types of interaction, tapping and gesture, in which a whole word can be written with a single gesture. To evaluate the performance of the C-QWERTY layout we therefore decided to compare it, through a user study, with another circular layout called Cirrin, in which the order of the characters has been optimized to minimize the distance between successive keys during writing.

This paper is an extended version of work published in [5]. We extend our previous work by adding an optional spell checker to the method and by expanding the empirical eval-

\*Corresponding author

\*\*Principal corresponding author

 gencos@unisa.it (G. Costagliola); matderosa@unisa.it (M. De Rosa); vfuccella@unisa.it (V. Fuccella)

ORCID(s): 0000-0003-3816-7765 (G. Costagliola);  
0000-0002-6922-5529 (M. De Rosa); 0000-0003-3244-3650 (V. Fuccella)

uation. In particular we have doubled the number of participants and compared the method performances both with and without the spell checker.

The paper is organized follows: Section 2 describes previous work on text entry on smartwatches and on circular layouts; Section 3 describes our layout and text entry method, Section 4 shows its experimental evaluation and Section 5 the evaluation results. Finally, Sections 6 concludes the paper with a discussion on future work.

## 2. Related Work

As in the case of smartphones and tablets [14, 26, 20, 8, 7], numerous text entry techniques have been proposed to facilitate text entry on smartwatches. To solve the problem caused by the small screen of smartwatches, a common approach is to introduce a further interaction step on soft keyboards. In SplitBoard [13] only a part of the keyboard is shown at any given time and a flick is used to change the displayed part. In ZoomBoard [19], instead, a zooming step is used to increase key sizes. In ZShift [15] a callout is used to show a zoomed copy of the screen under the user's finger. Such interaction steps, however, increase the time required to select each key. Another approach to the problem is to use gestures to select keys. SwipeBoard [4] is a eye-free text entry method in which two swipes are used to enter each character. DualKey [11] uses keys with two letters and tapping gestures with two different fingers to choose between them.

Other techniques instead use a word-based (e.g. WatchWriter [10]) or a sentence-based (e.g. VelociTap [25]) text entry approach.

It can be noted that all of above mentioned techniques are based on rectangular (QWERTY) keyboards. A rectangular layout, however, is naturally more suitable for screens of the same type. On a circular screen, instead, this layout can cause a waste of screen space. In this case a circular layout could allow a more efficient use of the screen space. In fact, in addition to the common rectangular layouts, there are also some keyboards with circular layouts, that are used to exploit such layout advantages in scenarios such as rotational interfaces [21] or personal area on the tabletop [2].

One of the first circular layout to be proposed is Cirrin, a text entry method for pen input devices based on a soft keyboard proposed by Mankoff and Abowd in 1998 [18]. Here, the keys are arranged in a circle, as shown in Figure 1, and the layout of the letters in the circle is such as to minimize the average distance traveled by the pen to write a word. This is because the user, to insert a word, simply draws a path, starting from the inside of the circle, that crosses the circumference in points corresponding to the characters of the desired word, in the right order. A space character is automatically inserted when the pen is raised. Cirrin has shown to have the potential to be faster than the classic QWERTY virtual keyboards and is particularly suitable for pen based devices.

Some other techniques use characters in alphabetical order. Shoemaker et al. [22] proposed a circular keyboard for

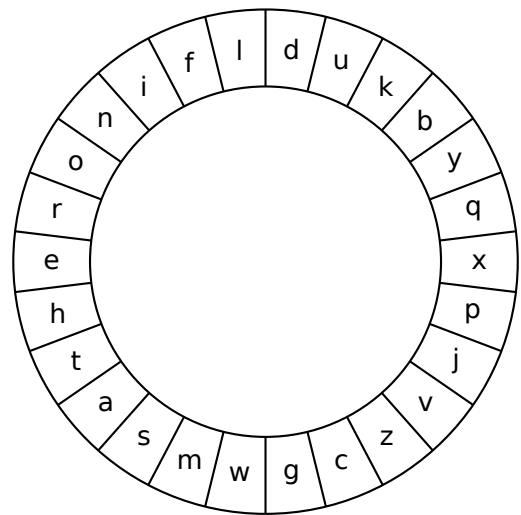


Figure 1: Cirrin Layout.

entering text on large wall displays. The selection is performed by moving the pointer over the desired character and by pressing a button to enter it. BubbleCircle [12] uses a ring arrangement of alphabetic characters on a tabletop display and magnifies the next most probable characters based on the user's input. TUP [21] is a text entry method for touch sensitive wheels. The characters are positioned on fixed positions on the wheel and the user presses a select key to enter the highlighted character.

## 3. C-QWERTY

Our layout, C-QWERTY, is a keyboard layout for circular smartwatches in which the alphanumeric keys are arranged along the edge of the screen in a circular layout. In order to make the keyboard more familiar to users, the keys are arranged in a similar way to that of traditional QWERTY keyboards, as shown in Figure 2.

In particular the keys corresponding to the first row of the QWERTY keyboard have been positioned at the top of the screen (from right to left: q, w, e, r, t, y, u, i, o, p), while those corresponding to the third keyboard row have been positioned at the bottom of the screen (from right to the left: z, x, c, v, b, n, m). In the remaining space on the sides, in the central part of the screen, the keys corresponding to the second keyboard row have been positioned partly on the left (from top to bottom: a, s, d, f) and partly on the right (from top to bottom: g, h, j, k, l).

In order to make the best use of the available screen space, our C-QWERTY soft keyboard implementation has on the center of the screen a text field that shows the transcribed text. Above this field there are the space and backspace keys, while below it there are up to 4 suggestion words in order to complete the current written word. This functionality uses a 20K words dictionary. An optional spell checker that uses the same dictionary is also available. When it is active, it tries to correct any misspelled word as the user types; its language model is also used in the suggestion fea-



Figure 2: C-QWERTY Layout.

ture to allow the system to propose words that do not necessarily start with the characters written by the user.

The keyboard has two interaction modes for entering text: tapping and gesture. In the first mode, by tapping on a key it is possible to enter the corresponding character. In gesture mode, in addition to tapping, it is also possible to enter an entire word with a single gesture. This is accomplished by dragging the finger in sequence over the alphanumeric keys corresponding to the characters of the desired word. As soon as the finger passes over a key, the system enters the corresponding character, and also automatically adds a space when the user raises his finger at the end of the gesture.

The space and backspace keys and the suggestion words can be used in both modes only by tapping.

## 4. Evaluation

We carried out a user-study aimed at evaluating the speed and accuracy of the C-QWERTY layout. To this aim we decided to compare it with the Cirrin layout described in Section 2 and shown in Figure 1.

In the experiment we asked the participants to transcribe some sentences using both layouts (C-QWERTY and Cirrin) and both interactions modes (tapping and gesture). Furthermore, the participants were divided into two groups, depending on whether the spell checker was enabled or not.

### 4.1. Participants

For the experiment, 24 participants (4 female) were recruited. They were all university students between 20 and 30 years old ( $M = 23.8$ ,  $SD = 2.6$ ) and chose to participate for free. All were usual users of computers and smartphones, while most of them had little experience with smartwatches.



Figure 3: A participant during the experiment.

Most of them also declared a good English knowledge. Figure 3 shows a participant during the experiment.

### 4.2. Apparatus

The experiment was conducted on a Ticwatch Pro equipped with a Snapdragon Wear 2100 Quad Core 1.2 Ghz processor and running the Wear OS operating system. The device weighs 58.5 grams and has a circular display with a 1.39" diagonal and a resolution of 400×400 pixels.

The experimental software is a Wear OS application that implements the two layouts (C-QWERTY and Cirrin), the two interaction modes (tapping and gesture) and the optional spell checker.

At startup the application asks the user to choose the desired layout, interaction mode and whether to enable the spell checker. After this selection the application shows for a few seconds the sentence to be transcribed, after which the keyboard that the participant must use to transcribe the sentence is shown. During this phase the sentence is also shown on a computer screen, in order to allow the participant to re-look at it if necessary. After writing the sentence the participant can confirm it by performing a long press over the text field at the center of the screen, after which the system shows the next sentence (or asks the user to transcribe again the same sentence, if he has exceeded the 15% of non corrected errors).

A personal computer with a 19" monitor was used to show the participant the current sentence.

### 4.3. Procedure

Before starting the experiment, the participants filled out a questionnaire with the following information: personal data (age, gender), dominant hand, the hand with which they actually performed the experiment, previous experiences with smartwatches and with text entry on smartwatches, level of proficiency with the English language.

The experiment was conducted in a well-lit laboratory. The participants were given the watch and told to wear it on the wrist where it was most convenient for them and possibly



rest their arm on a desk. Moreover, during the entire experiment they remained seated. Then, they had a short practice session in which the text entry method was explained and tested (fixed example sentences were shown on the PC monitor). The participants were given all the recommendations related to the experiment, and in particular to:

- read and memorize the sentence before starting to copy it;
- balance speed and accuracy when writing;
- correct mistakes made while entering text. Since the only way to correct errors is by using the backspace key, they were also told to avoid correcting errors noticed only after having already written other words.

The measured tasks started after the participants understood the procedure. The task was to transcribe short text sentences. Each participant had to enter six sentences in each of the four test conditions (the first sentence was for training and therefore not measured). The sentences were randomly chosen from the MacKenzie and Soukoreff set [17] that do not include punctuation or numbers. The participants were allowed a rest period of a few minutes at the end of each test condition.

At the end of the experiment the participants were asked to complete a System Usability Scale (SUS) [3] questionnaire for each layout (C-QWERTY e Cirrin). SUS includes ten statements, to which respondents had to specify their level of agreement using a five-point Likert scale. The questions alternate between positive and negative (since they are in a rather standard form we do not include them here). Each SUS questionnaire has a score between 0 and 100, of which we then calculated the averages on all participants.

In addition to SUS, we also asked for the preferred layout (C-QWERTY or Cirrin), the preferred interaction method (tapping or gesture) and the reasons for these choices. We also collected further feedback through an open form and verbal interaction.

#### 4.4. Design

The experiment was a three-way with two within-subjects factors and one between-subjects factor design. The two within-subjects factors were the Layout and the Interaction mode. The Layout included two levels: C-QWERTY and Cirrin, while the Interaction mode included two levels: tapping and gesture. The one between-subjects factor was the Group with two levels: NSC (No Spell Checker used) and SC (Spell Checker used).

As dependent variables we included:

- Speed: text entry speed measured in words per minute (wpm) as specified in [16].
- Accuracy: the text entry accuracy using both the Total Error Rate (TER) and the Non Corrected Error Rate (NCER), calculated as specified in [24].

**Table 1**

Counterbalancing used during the experiment.

Participants	Group	Orderings			
1, 5, 9	NSC	C-QWERTY TAPPING	C-QWERTY GESTURE	CIRRIN TAPPING	CIRRIN GESTURE
2, 6, 10	NSC	CIRRIN TAPPING	CIRRIN GESTURE	C-QWERTY TAPPING	C-QWERTY GESTURE
3, 7, 11	NSC	C-QWERTY GESTURE	C-QWERTY TAPPING	CIRRIN GESTURE	CIRRIN TAPPING
4, 8, 12	NSC	CIRRIN GESTURE	CIRRIN TAPPING	C-QWERTY GESTURE	C-QWERTY TAPPING
13, 17, 21	SC	C-QWERTY TAPPING	C-QWERTY GESTURE	CIRRIN TAPPING	CIRRIN GESTURE
14, 18, 22	SC	CIRRIN TAPPING	CIRRIN GESTURE	C-QWERTY TAPPING	C-QWERTY GESTURE
15, 19, 23	SC	C-QWERTY GESTURE	C-QWERTY TAPPING	CIRRIN GESTURE	CIRRIN TAPPING
16, 20, 24	SC	CIRRIN GESTURE	CIRRIN TAPPING	C-QWERTY GESTURE	C-QWERTY TAPPING

- GPC (gestures per character): the average number of gestures needed to enter a character (each interaction with the touchscreen is counted as a gesture).

We counterbalanced the two factors, as shown in Table 1.

## 5. Results

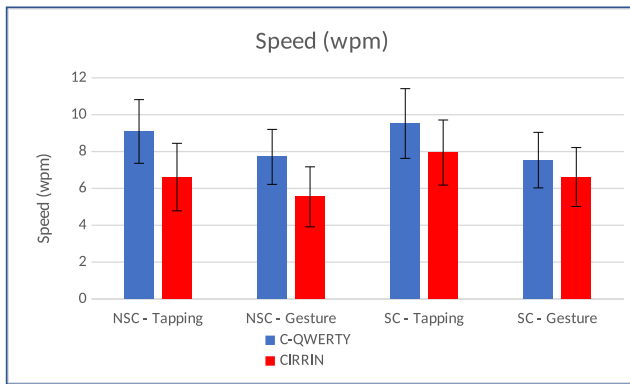
All participants completed the experiment. One participant of the *NSC group* was unable to successfully use the gesture interaction and basically used tapping when performing the two gesture tasks. For each participant the experiment lasted about 30 minutes. We tested significance using an analysis of repeated variance measures (ANOVA) [9].

### 5.1. Speed

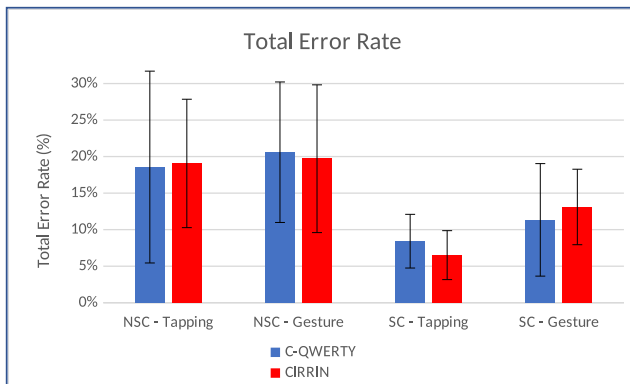
The text entry speeds are shown in Figure 4. The grand mean for it was 7.6 wpm. *C-QWERTY* was the fastest layout with an average of 8.5 wpm, outperforming *Cirrin* at 6.7 wpm. Regarding the interaction mode, *tapping* was fastest with an average of 8.3 wpm, outperforming *gesture* at 6.8 wpm. Regarding the groups, the *SC group* was fastest with an average of 7.9 wpm, outperforming the *NSC group* at 7.2 wpm.

Without distinguishing between the two groups, the highest speeds were obtained by *C-QWERTY/tapping* at 9.3 wpm and *C-QWERTY/gesture* at 7.6 wpm, while the slowest speeds were obtained by *Cirrin/tapping* at 7.3 wpm and *Cirrin/gesture* at 6.1 wpm. As expected, the highest speed was obtained by *C-QWERTY/tapping/SC* at 9.5 wpm, while the slowest speed was obtained by *Cirrin/gesture/NSC* at 5.5 wpm.

This is probably because in the C-QWERTY mode, the participants could immediately find the position of the letters, given the similarity with the well known QWERTY layout. Moreover the gesture mode is slower than the tapping mode, probably due the fact that some keys are hidden by the hand when performing gestures. The SC group has recorded better speeds on average, although not in a statistically significant manner. This however gives us a clue that the use of a language model can cause an increase in speed



**Figure 4:** Speeds (in wpm). Error bars show the standard deviation.



**Figure 5:** Total error rates. Error bars show the standard deviation.

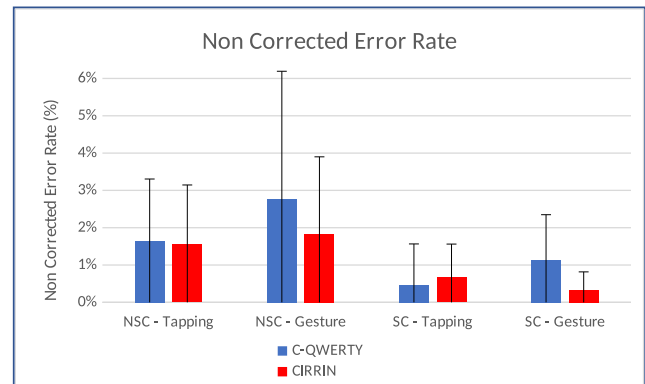
for the C-QWERTY and Cirrin circular layouts. In particular, from the chart in Figure 4, it can be seen that, with respect to C-QWERTY, for Cirrin there is a more marked difference between the NSC and SC groups. This is probably caused by the fact that with Cirrin lower typing speeds, the use of a suggestion to complete a word allows for a greater speed increase, while for C-QWERTY faster typing speed, the use of a suggestion does not give the same advantage.

From the ANOVA resulted that the effect of the layout on the speed was highly significant ( $F_{1,22} = 47.098$ ,  $p < .0001$ ). There was also a highly significant effect for interaction mode ( $F_{1,22} = 42.953$ ,  $p < .0001$ ). There was not a statistically significant difference for the group ( $F_{1,22} = 1.371$ ,  $ns$ ) and for the interactions between: layout and group ( $F_{1,22} = 4.267$ ,  $ns$ ), interaction mode and group ( $F_{1,22} = 0.964$ ,  $ns$ ), layout and interaction mode ( $F_{1,22} = 1.642$ ,  $ns$ ), layout, interaction mode and group ( $F_{1,22} = 0.221$ ,  $ns$ ).

## 5.2. Accuracy

Average values for TER and NCER are shown in Figure 5 and Figure 6.

The grand mean for TER was 14.7%. There was little difference between layouts, with C-QWERTY at 14.7% and Cirrin at 14.6%. Regarding the interaction mode, *tapping* was more accurate with an average of 13.1%, while



**Figure 6:** Non corrected error rates. Error bars show the standard deviation.

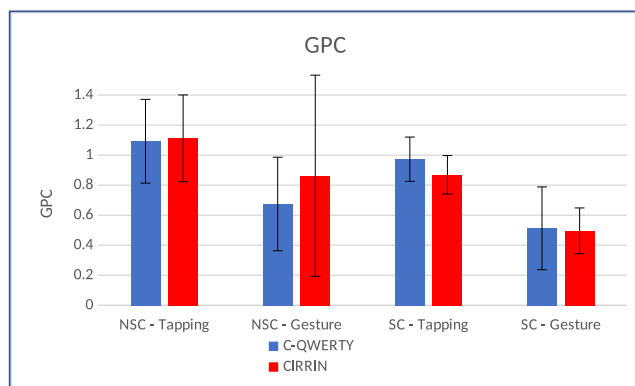
*gesture* reached 16.2%. Regarding the groups, the SC group was more accurate with an average of 9.8%, while the NSC group reached 19.5%. The lowest TER was obtained by Cirrin/tapping/SC at 6.5%, while the highest by C-QWERTY/gesture/NSC at 20.6%. From the ANOVA resulted that the effect of the group was highly significant ( $F_{1,22} = 13.528$ ,  $p = .0013$ ). However, there was not a statistically significant difference for the layout ( $F_{1,22} = 0.014$ ,  $ns$ ), the interaction mode ( $F_{1,22} = 3.929$ ,  $ns$ ), and for the interactions between: layout and group ( $F_{1,22} = 0.003$ ,  $ns$ ), interaction mode and group ( $F_{1,22} = 1.235$ ,  $ns$ ), layout and interaction mode ( $F_{1,22} = 0.329$ ,  $ns$ ), layout, interaction mode and group ( $F_{1,22} = 1.606$ ,  $ns$ ).

The grand mean for NCER was 1.3%. There was little difference between layouts, with C-QWERTY at 1.5% and Cirrin at 1.1%. There was also little difference between interaction modes, with *tapping* at 1.1% and *gesture* at 1.5%. Regarding the group, the SC group was more accurate with an average of 0.7%, while the NSC group reached 1.9%. The lowest NCER was obtained by Cirrin/gesture/SC at 0.3%, while the highest by C-QWERTY/gesture/NSC at 2.8%. From the ANOVA resulted that the effect of the group was significant ( $F_{1,22} = 6.190$ ,  $p = .0209$ ). However, there was not a statistically significant difference for the layout ( $F_{1,22} = 2.947$ ,  $ns$ ), the interaction mode ( $F_{1,22} = 2.695$ ,  $ns$ ), and for the interactions between: layout and group ( $F_{1,22} = 0.202$ ,  $ns$ ), interaction mode and group ( $F_{1,22} = 0.969$ ,  $ns$ ), layout and interaction mode ( $F_{1,22} = 1.673$ ,  $ns$ ), layout, interaction mode and group ( $F_{1,22} = 0.012$ ,  $ns$ ).

## 5.3. Gestures per Character

The GPCs are shown in Figure 7. The grand mean for it was 0.82. The C-QWERTY layout had the lower value at 0.81, with Cirrin at 0.83. Regarding the interaction mode, as expected *gesture* had the lower value at 0.64, with *tapping* at 1.01. Regarding the group, the SC group had the lower value at 0.71, with the NSC group at 0.94. The lowest value was obtained by Cirrin/gesture/SC at 0.50, with the highest value obtained by Cirrin/tapping/NSC at 1.11.

From the ANOVA resulted that the effect of the interaction mode on the KSPC was highly significant ( $F_{1,22} =$



**Figure 7:** Gestures per Character. Error bars show the standard deviation.

24.658,  $p = .0001$ ). There was also a significant effect for group ( $F_{1,22} = 8.792$ ,  $p < .0071$ ). However, there was not a statistically significant difference for the layout ( $F_{1,22} = 0.164$ ,  $ns$ ), and for the interactions between: layout and group ( $F_{1,22} = 2.385$ ,  $ns$ ), interaction mode and group ( $F_{1,22} = 0.304$ ,  $ns$ ), layout and interaction mode ( $F_{1,22} = 1.196$ ,  $ns$ ), layout, interaction mode and group ( $F_{1,22} = 0.122$ ,  $ns$ ).

#### 5.4. User Satisfaction and Free-form Comments

The average SUS score was 70.3 ( $SD = 15.1$ ) for C-QWERTY and 59.2 ( $SD = 18.1$ ) for Cirrin. A Wilcoxon matched-pairs signed-ranks test [23] performed on SUS scores revealed a statistical significance between the two techniques ( $Z = 2.85$ ,  $p < .005$ ).

The trend on such scores was confirmed by the choice of the preferred layout, with 20 participants that preferred C-QWERTY and only 4 that preferred Cirrin. Most of them motivated their choice by declaring that the C-QWERTY layout was more intuitive and familiar.

Regarding the interaction mode 22 participants declared their preference for tapping, with only 2 that preferred gesture. Most of them stated that in gesture mode it is too difficult to see the keys and the suggestions during writing.

#### 5.5. Discussion

During the experiment, even if most participants had little experience with smartwatches, most participants immediately learned how the method works and showed a fast learning process. To this, it may have contributed the fact that participants were all usual users of smartphones and computers. Moreover, all of them were in the 20-30 age group and university students (some of them even in computer science). This may have influenced the obtained speeds in absolute terms, with respect to the general population. However, the relative difference between the different layouts should not have been affected by this.

Most participants preferred the C-QWERTY layout, stating its familiarity as main motivation. This confirms that the QWERTY layout remains recognizable even after its adaptation to a circular layout. All the participants com-

plained about the gesture mode, stating that it is too difficult to see the keys and press them correctly. Some participants also complained about the Cirrin layout stating that the position of the characters confused them. Some participants also complained about the touchscreen accuracy, stating that sometimes it is difficult to click on a key or a suggestion.

Most of them also appreciated the word suggestion functionality, stating that it was a fundamental help, since it made possible to transcribe the sentences more quickly.

Many of the NSC group have instead complained of the lack of a spell checker and of the fact that the entry of a single wrong character is enough to ensure that the desired word will never be shown as suggestion. No complaints about the spell checker were made by the SC group.

In this study we decided to compare the two layouts both with no spell checker and a simple suggestion feature, and with a spell checker and a more advanced suggestion feature. We have been able to observe that in the second case the performances have increased on average by 9%, but not in a statistically significant way. This, anyway, gives us a clue that the use of a language model can cause an increase in speed. As expected, instead, a significant improvement in accuracy for both TER and NCER has been recorded.

## 6. Conclusions and further works

In this paper we presented a study comparing different circular layouts for entering text on smartwatches in order to measure the extent through which the use of a circular QWERTY-like layout increases user performance. In particular, we compared our proposed layout C-QWERTY to Cirrin, a layout in which the order of the characters has been optimized to minimize the distance between successive keys during writing. The experimental results show that the C-QWERTY layout significantly outperforms the Cirrin layout in the first sessions of use.

Future work will focus on measuring the performance when the user experience increases and on comparing C-QWERTY with other text entry methods.

## References

- [1] Arif, A.S., Mazalek, A., 2016. A survey of text entry techniques for smartwatches, in: Proceedings, Part II, of the 18th International Conference on Human-Computer Interaction. Interaction Platforms and Techniques - Volume 9732, Springer-Verlag, Berlin, Heidelberg. pp. 255–267. doi:[10.1007/978-3-319-39516-6\\_24](https://doi.org/10.1007/978-3-319-39516-6_24).
- [2] Blanch, R., Ortega, M., 2009. Rake cursor: Improving pointing performance with concurrent input channels, in: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, ACM, New York, NY, USA. pp. 1415–1418. doi:[10.1145/1518701.1518914](https://doi.org/10.1145/1518701.1518914).
- [3] Brooke, J., et al., 1996. Sus-a quick and dirty usability scale. Usability evaluation in industry 189, 4–7.
- [4] Chen, X.A., Grossman, T., Fitzmaurice, G., 2014. Swipeboard: A text entry technique for ultra-small interfaces that supports novice to expert transitions, in: Proceedings of the 27th Annual ACM Symposium on User Interface Software and Technology, ACM, New York, NY, USA. pp. 615–620. doi:[10.1145/2642918.2647354](https://doi.org/10.1145/2642918.2647354).
- [5] Costagliola, G., De Rosa, M., D'Arco, R., De Gregorio, S., Fucella, V., Lupo, D., 2019. C-QWERTY: a text entry method for circular smartwatches, in: The 25th International DMS Conference

- on Visualization and Visual Languages, pp. 51–57. doi:[10.18293/DMSVIVA2019-014](https://doi.org/10.18293/DMSVIVA2019-014).
- [6] Costagliola, G., De Rosa, M., Fuccella, V., 2017. Handwriting on smartwatches: An empirical investigation. *IEEE Transactions on Human-Machine Systems* 47, 1100–1109. doi:[10.1109/THMS.2017.2754938](https://doi.org/10.1109/THMS.2017.2754938).
- [7] Costagliola, G., De Rosa, M., Fuccella, V., 2018. A technique for improving text editing on touchscreen devices. *Journal of Visual Languages & Computing* 47, 1 – 8. URL: <http://www.sciencedirect.com/science/article/pii/S1045926X17302422>, doi:[10.1016/j.jvlc.2018.04.002](https://doi.org/10.1016/j.jvlc.2018.04.002).
- [8] Fuccella, V., De Rosa, M., Costagliola, G., 2014. Novice and expert performance of keystretch: A gesture-based text entry method for touch-screens. *IEEE Transactions on Human-Machine Systems* 44, 511–523. doi:[10.1109/THMS.2014.2314447](https://doi.org/10.1109/THMS.2014.2314447).
- [9] Girden, E., 1992. ANOVA: Repeated Measures. Number No. 84 in ANOVA: Repeated Measures, SAGE Publications. URL: <https://books.google.it/books?id=pCw3jwEACAAJ>.
- [10] Gordon, M., Ouyang, T., Zhai, S., 2016. Watchwriter: Tap and gesture typing on a smartwatch miniature keyboard with statistical decoding, in: Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems, ACM, New York, NY, USA. pp. 3817–3821. doi:[10.1145/2858036.2858242](https://doi.org/10.1145/2858036.2858242).
- [11] Gupta, A., Balakrishnan, R., 2016. Dualkey: Miniature screen text entry via finger identification, in: Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems, ACM, New York, NY, USA. pp. 59–70. doi:[10.1145/2858036.2858052](https://doi.org/10.1145/2858036.2858052).
- [12] Hinrichs, U., Schmidt, H., Isenberg, T., Hancock, M.S., Carpendale, S., 2008. BubbleType: Enabling text entry within a walk-up tabletop installation. Technical Report. University of Calgary.
- [13] Hong, J., Heo, S., Isokoski, P., Lee, G., 2015. Splitboard: A simple split soft keyboard for wristwatch-sized touch screens, in: Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems, ACM, New York, NY, USA. pp. 1233–1236. doi:[10.1145/2702123.2702273](https://doi.org/10.1145/2702123.2702273).
- [14] Isokoski, P., 2004. Performance of menu-augmented soft keyboards, in: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, ACM, New York, NY, USA. pp. 423–430. doi:[10.1145/985692.985746](https://doi.org/10.1145/985692.985746).
- [15] Leiva, L.A., Sahami, A., Catala, A., Henze, N., Schmidt, A., 2015. Text entry on tiny qwerty soft keyboards, in: Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems, ACM, New York, NY, USA. pp. 669–678. doi:[10.1145/2702123.2702388](https://doi.org/10.1145/2702123.2702388).
- [16] MacKenzie, I.S., 2015. A note on calculating text entry speed. URL: <http://www.yorku.ca/mack/RN-TextEntrySpeed.html>.
- [17] MacKenzie, I.S., Soukoreff, R.W., 2003. Phrase sets for evaluating text entry techniques, in: CHI '03 Extended Abstracts on Human Factors in Computing Systems, ACM, New York, NY, USA. pp. 754–755. doi:[10.1145/765891.765971](https://doi.org/10.1145/765891.765971).
- [18] Mankoff, J., Abowd, G.D., 1998. Cirrin: A word-level unistroke keyboard for pen input, in: Proceedings of the 11th Annual ACM Symposium on User Interface Software and Technology, ACM, New York, NY, USA. pp. 213–214. doi:[10.1145/288392.288611](https://doi.org/10.1145/288392.288611).
- [19] Oney, S., Harrison, C., Ogan, A., Wiese, J., 2013. Zoomboard: A diminutive qwerty soft keyboard using iterative zooming for ultra-small devices, in: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, ACM, New York, NY, USA. pp. 2799–2802. doi:[10.1145/2470654.2481387](https://doi.org/10.1145/2470654.2481387).
- [20] Oulasvirta, A., Reichel, A., Li, W., Zhang, Y., Bachynskyi, M., Vertanen, K., Kristensson, P.O., 2013. Improving two-thumb text entry on touchscreen devices, in: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, ACM, New York, NY, USA. pp. 2765–2774. doi:[10.1145/2470654.2481383](https://doi.org/10.1145/2470654.2481383).
- [21] Proschowsky, M., Schultz, N., Jacobsen, N.E., 2006. An intuitive text input method for touch wheels, in: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, ACM, New York, NY, USA. pp. 467–470. doi:[10.1145/1124772.1124842](https://doi.org/10.1145/1124772.1124842).
- [22] Shoemaker, G., Findlater, L., Dawson, J.Q., Booth, K.S., 2009. Mid-air text input techniques for very large wall displays, in: Proceedings of Graphics Interface 2009, Canadian Information Processing Society, Toronto, Ont., Canada, Canada. pp. 231–238. URL: <http://dl.acm.org/citation.cfm?id=1555880.1555931>.
- [23] Siegel, S., 1956. Nonparametric statistics for the behavioral sciences. McGraw-Hill, New York, NY, US.
- [24] Soukoreff, R.W., MacKenzie, I.S., 2003. Metrics for text entry research: An evaluation of msd and kspc, and a new unified error metric, in: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, ACM, New York, NY, USA. pp. 113–120. doi:[10.1145/642611.642632](https://doi.org/10.1145/642611.642632).
- [25] Vertanen, K., Memmi, H., Emge, J., Reyat, S., Kristensson, P.O., 2015. Velocitap: Investigating fast mobile text entry using sentence-based decoding of touchscreen keyboard input, in: Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems, ACM, New York, NY, USA. pp. 659–668. doi:[10.1145/2702123.2702135](https://doi.org/10.1145/2702123.2702135).
- [26] Zhai, S., Kristensson, P.O., 2012. The word-gesture keyboard: Reimagining keyboard interaction. *Commun. ACM* 55, 91–101. doi:[10.1145/2330667.2330689](https://doi.org/10.1145/2330667.2330689).





# Journal of Visual Language and Computing

journal homepage: [www.ksiresearch.org/jvlc](http://www.ksiresearch.org/jvlc)

## Run-time adaptation of modalities of interaction and context of use in web mobile apps

Danilo Camargo Bueno, Luciana A. M. Zaina

<sup>a</sup>Department of Computer Science, Federal University of São Carlos - UFSCAR, Sorocaba (SP), BRA

### ARTICLE INFO

#### Article History:

Submitted 04.29.2019

Revised 08.01.2019

Second Revision 10.09.2019

Accepted 10.09.2019

#### Keywords:

Dynamic Adaptation

Mobile web app

Context of use

Modalities of interaction

Front-end Framework

### ABSTRACT

Several approaches have emerged to support the development of mobile web applications (apps). Front-end frameworks (FeF) have emerged to support developers in the construction of responsive mobile web apps. However, these frameworks do not provide resources to handle easily variables of the context of use and to deal with different modalities of interaction. Considering this gap, we proposed the HyMobWeb, an approach that assists developers in working with these aspects. By grasping the popularity of the FeFs, HyMobWeb proposes a flexible and reusable approach based on FeF structure. It works with a hybrid approach that treats the adaptation in two phases. The static one, performed in the development time, allows developers to implement the resources of adaptation. The dynamic one performs the adaptation during the run-time. In this article, we present the HyMobWeb dynamic adaptation approach. As end-users are the receivers of the dynamic adaptation our evaluation was carried out in the perspective of this audience. The results showed that the adaptations regarding the context of use were the ones that presented more impact on the user experience. We concluded that the HyMobWeb dynamic adaptation provides ways to enhance user interaction in mobile web apps when compared with RWD resources.

© 2019 KSI Research

## 1. Introduction


Recently, developers have been facing issues on how to build web applications that are suitable to interact in small mobile devices [15, 16]. Some of these issues come up from the limitations that some browsers presented in supporting applications (apps) of different mobile platforms [1, 15]. Frequently, the development of Web applications require an extra effort of developers. Considering a base web version of the app, they have to customize the web versions to use resources (e.g. sensors, screen size) of the mobile devices [18].

The interaction in mobile devices create a new demand for the developers. They are forced to deal with the context in which the application will be used [7, 15]. Context of use encompasses a set of elements as such as users preferences, device capabilities and environmental. These elements together provide the information about the context of use [12, 7].

Besides of questions about the user experience, by treating the aspects relevant to the context of user interaction the developers can enhance user performance on the apps [10, 15, 25]. Non-adaptive user interface for mobile devices can cause some frustration that directly affects the performance and usability of the users. The context of use can also affect the user preferences for different modalities of interaction [13].

The adaptation to different modalities of interaction can become suitable in situations in which the users have some discomfort caused by the luminosity, background noises, and other environmental issues. By providing different modalities of interaction developers can improve the usability in mobile apps [10].

Although the adaptation of mobile apps is not a new topic, there are some gaps to be explored regarding the tools and techniques to support the adaptations [1, 15, 20, 13]. Responsive Web Design (RWD) is an approach proposed by Ethan Marcotte that is widely used by Web developers. Its main principles are flexible layouts with relative dimensions, flexible images and videos content through dynamic

 danilocbueno@ifsp.edu.br (D.C. Bueno); lzaina@ufscar.br (L.A.M. Zaina)

ORCID(s): 0000-0002-7359-9937 (D.C. Bueno); 0000-0002-1736-544X (L.A.M. Zaina)

DOI reference number: 10-18293/JVLC2019-N2-011

resizing; and adaptation based on media queries<sup>1</sup> [16, 11].

Although RWD is widely used, it presents some limitations regarding the context of use. The detection in the context changes are restricted to the adaptation considering the screen resolution and device orientation. Additionally, the support to code the adaptations is merely provided by showing/hiding elements and changing some visual attributes [16, 17, 12].

From the increasing of popularity of RWD many Front-end Frameworks (FeF) came up to aid developers in the construction of web mobile applications. FeF combines the concepts of RWD to the User Interface Design Patterns (UIDP). UIDPs are descriptions of best practices or some sort of heuristics to fulfill design problems [23]. They are reusable solutions (e.g. menus, buttons, lists, and so on) that can be applied in different domains. A set of components built in HTML (Hypertext Markup Language), CSS and the JavaScript programming language composes the core of FeFs.

Bootstrap, Zurb Foundation, Pure, Materialize, KickStart<sup>2</sup> are examples of FeFs. Among them, Bootstrap has become the most popular one to support developers and designers work. Its popularity is a consequence of the framework maturity (precursor among the other FeFs) [19, 22].

However, these FeFs concerns strictly in handling features linked with RWD (e.g. different screen sizes) [12]. FeFs do not provide resources to implement adaptations of different input modalities or even different contexts of use [17]. As a consequence, developers have to employ more effort on the implementation of such features. Additionally, there is no pattern to support the designing of different interaction modalities as well as to treat the context of use.

Based on the RWD and FeF restrictions, we proposed the HyMobWeb approach. It provides resources to develop a hybrid adaptation on user interfaces of web mobile apps. A hybrid approach works with static and dynamic strategies of adaptation. HyMobWeb enables developers to indicates in code time (i.e. static strategy of adaptation) which parts of the web mobile user interface can be adapted during the run-time (i.e. dynamic strategy of adaptation). Static and dynamic adaptations are implemented from an extension of FeF. This approach has the advantage that it can be applied to different FeFs as they have similar structures. From HyMobWeb approach, developers can define different modalities of interaction considering the aspects of context of use. The static adaptation of HyMobWeb was previously presented in [5].

In this article, our goal is to present the HyMobWeb dynamic adaptation. To explore the HyMobWeb proposal, we extended the Bootstrap FeF and applied it in the development of a mobile web app. With the participation of 14 volunteers, we evaluated the dynamic adaptation from the perspective of end-users. To explore the potential of the proposal, we developed two versions of apps and conducted a comparison

of them. The first one was implemented by using the basic RWD resources of Bootstrap. In the second, we evolved the first one by applying the resources available in the extension. We highlight that dynamic adaptation approach, and its evaluation was not presented in other work.

This article is organized as follow: related work is presented and discussed in Section 2; as both adaptations are interconnected, we briefly give an overview of HyMobWeb approach and of static adaptation in Sections 3 and 4, respectively; Section 5 presents the dynamic adaptation approach in details and Section 6 its evaluation; finally the discussion about the proposal and the final remarks are discussed in Sections 7 and 8, respectively.

## 2. Related Work

Our related work concerns on discussing of mobile web adaptations in different perspectives as follow.

Semantic Transformer [21] promotes the adaptation of user interface from web desktop to mobile web approach. The adaptation mechanism modifies the user interface structure to get one with low cost to be performed in a mobile device. Small Screen Device (SSD) Browser [1] also presents a desktop-mobile Web adaptation method based on visual and structural arrangement of interface elements. It identifies the position of elements and the relationship among the elements of the document and thus provides the mobile web user interface. Besides, the method allows users to set their own adaptations. However, a particular browser has to be adopted to use the proposal.

Tree Adapt [2] provides the adaptation by rearranging the interface elements to keep only the content which can fit into mobile devices screen. As a result users can visualize different contents by clicking on a title which triggers the block's expansion. W3Touch [18] takes the injection of JavaScript code to track touchscreen events and collected data from the user interaction (interaction tracking). By combining the data of user interaction to adaptation rules W3Touch automatically detects and suggests web page components. These adaptation rules are defined in accordance to the mobile device features. Yigitbas et al. [25] presents the CoBAUI, a component-based development framework for building adaptive interfaces. The work is based on the Angular - framework for the development of advanced Web applications. The developer create the component using the angular architecture and write the code that will perform the adaptation according to the previously defined rules.

A Web interface adaptation approach by combining the modalities of voice and graphic is proposed by [15]. The approach follows a set of rules to select the suitable adaptation to a given device. During the application development phase, the interface elements can be marked by a CSS class. Considering these marks and the adaptation rules, a multimodal interface is automatically generated. However, this approach demands for a browser extension which are not available on mobile browsers. To make the proposal feasibility, the authors developed a new mobile browser which has to

<sup>1</sup>media query is a CSS (Cascading Style Sheets) resource from which user interface can be adapted according to the features of devices.

<sup>2</sup>Links: [getbootstrap.com](http://getbootstrap.com), [foundation.zurb.com](http://foundation.zurb.com), [purecss.io](http://purecss.io), [materializecss.com](http://materializecss.com), [getkickstart.com](http://getkickstart.com)

be installed in the devices. AdaM approach [20] adapt interfaces in real-time collaborative environments promoting the rearrangement of the interface elements. It works based on the analysis of information from the context of use, such as the user's device, functions, preferences, and access rights.

Although the works above discuss the topics of use interfaces adaptation, they present some gaps. Semantic Transformer [21], Tree Adapt [2], W3Touch [18] and AdaM approach [20] present different ways of manipulating a Web page's content so that it can adapt to the size of the mobile device. However, they concentrate on working in the content adaptation and not in the interface elements adaptation to different modalities of interaction.

SSD Browser and Manca proposal [1, 15] introduce barriers to their adoption so these proposals demand for the installation of new mobile browsers. CoBAUI [25] handles only variable of context of use, not including different modalities of interaction. Besides, as far as we know there were no works that explore the adaptation by applying the features of FeFs. And also we could not found works which provide resources to deal with the adaptation to different modalities of interaction and to the context of use.

### 3. HyMobWeb approach

HyMobWeb is a hybrid approach which supports the adaptation of the web mobile apps concerning on variables of the context of use in combination with the different modalities of interaction. The hybrid strategy blends static and dynamic approaches for interface adaptation. On the one hand, static adaptation is implemented by developers in coding time. On another hand, in dynamic strategy the adaptation in the code happens during the run-time [7]. By taking the advantages of user interface design patterns available in FeFs, HyMobWeb interconnects both strategies of adaptation without requesting the use of new web browsers. Additionally, HyMobWeb extends the structure of FeF to provide developers with resources to work on the adaptation of modalities of interaction and context of use. By using the same structure of the FeFs HyMobWeb approach becomes ease to be implemented in different FeFs.

Figure 1 illustrates the approach. The core of HyMobWeb is a Domain Specific Language (DSL). This DSL provides a set of codes which aids developers, during code time, to define what will be adapt considering the context of use and the modalities of interaction. The step (A) in Figure 1 shows how the static adaptation works. Based on the application constructed from a FeF (1a), the developers use the HyMobWeb DSL (2.1a) to indicate in which interface elements the adaptations will undergo (3a) during the run-time. The markings are introduced in the code of the interface elements.

Through an adaptation engine (Figure 1 - step B), HyMobWeb performs the dynamic adaptation by parsing the source code of the web mobile apps and thus modifying the elements which were previously marked in the development time (step A). In the dynamic adaptation, the engine checks each marking that was inserted during the coding time (i.e.

from static adaptation (1b)), interprets such codes and delegates the execution to multimodal (2.1b) and context (2.2b) handlers. The *Multimodal Handler* manages the adaptations of the interaction modalities of the elements. The *Context Handler* analyzes continuously the changes that occur in the variables of user context, and then performs the adaptation. Finally, the adaptation engine returns the code adapted to the application (4b).

As mentioned in the Introduction section, the static adaptation of HyMobWeb and its evaluation were already presented in [5]. As the dynamic adaptation are interconnected to static, we will discuss briefly the static one in next section. Sequentially, we will present the dynamic one which is the focus of this work.

### 4. Static adaptation

The main component of static adaptation is the HyMobWeb DSL. It allows developers to set which elements of context of use and modalities of interaction will be take into account in a run-time adaptation. The approach considers three categories of variables of context of use based on the proposal of [21]: *user* (i.e. preferences, goals and tasks, physical position), *technology* (i.e. screen resolution, connectivity, browser, battery), and *environment* (location, luminosity, noise level). Our approach proposed two distinct ways for the implementation of the code of the adaptation.

In the first one, the developer uses CSS language following the grammar presented in Definition 1. The `@context` points to the *Context Handler* that a context variable should be treated; the `<aspect>` describes which variables of the context will be considered; the `<state>` informs which is the value that will triggers the adaptation; and the `{...}` defines the CSS code that will be applied to one or more elements to provide the adaptation. Some examples are `@context (user-activity: walking){ ... }`, `@context (battery-level: low){ ... }`, `@context (noise-level: high){ ... }`. They are variables of context of use that when changes to that state will cause the adaptation. They represent the conditions of the user walking, the device's battery at a low level and a high noise level in the environment, respectively.

**Definition 1.** *grammar to use in CSS code*

$$\text{@context ( <aspect> : <state> ) } \{ \dots \} \quad (1)$$

The second way allows developers to configure which interface elements will suffers the adaptation. To do this, developers make a linking between HTML elements and CSS class following the grammar presented in Definition 2. The `<object-behavior>` represents the behavior that the object should take; the aspect reports the available context aspects (battery-level, luminosity, user-activity); the state describes the values that trigger the execution of the behavior. This proposal can be applied to a variety of contexts and states. HyMobWeb provides a list of pre-defined behaviors as such as *visible*, *hidden*, *bigger*, *darken*. For instance, some potential classes can be *.visible-on-user-activity-walking*, *.increase-on-luminosity-low* and *.hidden-on-noise-level-low*.

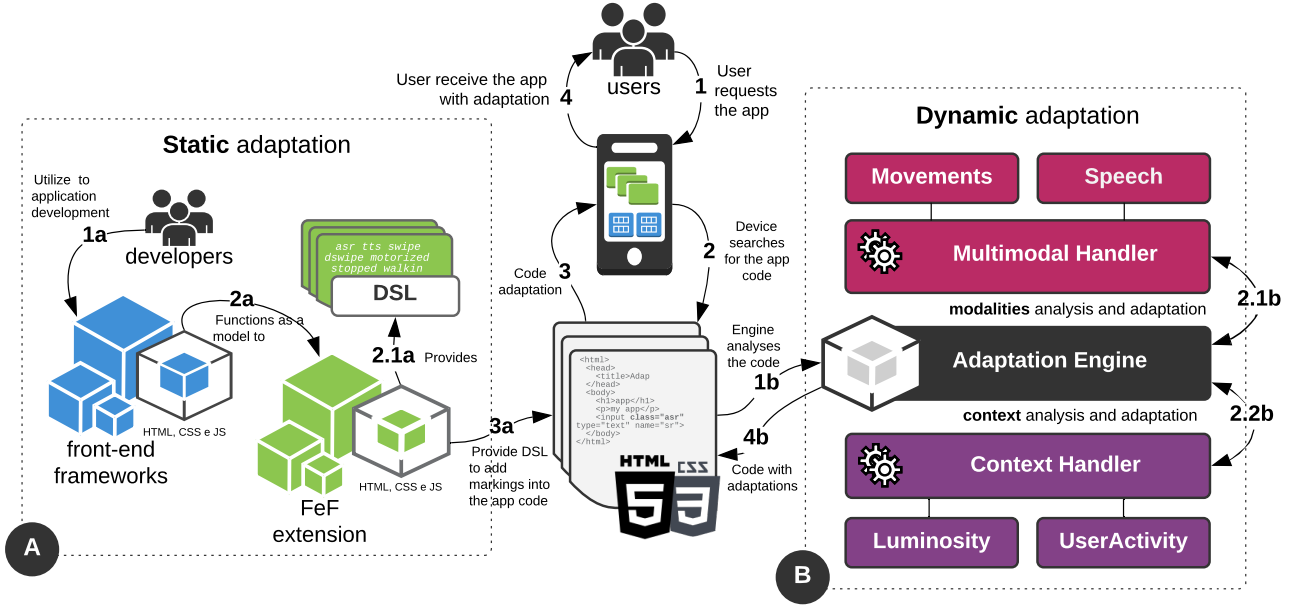


Figure 1: HyMobWeb approach.

**Definition 2.** grammar to use in HTML code

$$.<object-behavior>-on-<aspect>-<state> \quad (2)$$

Following the same structure of grammar above, HyMobWeb offers the possibility of using new interaction modalities. In the `<aspect>`, developers report on which aspect the modalities of interaction will be available (i.e. interface element, screen, movements, device). Some examples are `.listen-on-element-focus` (activate automatic speech recognition when the element receive focus), `.speak-on-user-activity-walking` (activate speech synthesizer when the user is walking), `.hidden-on-movements-swipeleft` (hide the element when the swipe to the left movement is recognized), and `.vibrate-on-luminosity-low` (vibrate the device when the ambient light level is low).

HyMobWeb DSL is generic enough to be applied to different aspects. However, technological features of devices and browsers' features have to be taken into account during the implementation of the grammar.

To implement an instance of our approach, we extended the *Bootstrap* FeF. We added the modalities of *Speech* and *Movements* providing to developers ways to work with automatic speech recognition, text-to-speech, swipe movements, double swipe, double tap and pinch. Regarding the context of use, the extension supported developers on the analysis of characteristics of user and environment using the contexts of *User Activity* and *Luminosity*. These consider the user's physical state (i.e. stopped, walking or motorized) or the level of luminosity (i.e. low, medium or high), respectively.

As mentioned previously, we already carried out the evaluation of static strategy of HyMobWeb approach with 19 developers. The findings showed that the developers were able to use the static adaptation of HyMobWeb and also demon-

strated the developers could perceive the utility of the approach. The details of static adaptation implementation and evaluation can be found in [5].

## 5. Dynamic Adaptation

Dynamic adaptation (see Figure 1-B) is performed based on the settings of variables of context and modalities of interaction. These settings are done during the coding phase (i.e. static adaptation). The core of HyMobWeb dynamic strategy is the *Adaptation Engine*. The adaptation engine is composed by a set of HTML, CSS and Javascript codes that are installed on the web mobile application. They execute the parsing on the code and perform the adaptations. *Multimodal* and *Context Handlers* (see Figure 1 - 2.1b and 2.2b respectively) are responsible to the treatment of the variables and to the changings in the code in accordance with the results of code analysis. HyMobWeb adaptation engine (see Figure 2) is conceived as a based-component architecture which makes easy its extension. The operation of the *Context Handler* and *Modality Handler* components will be explained individually.

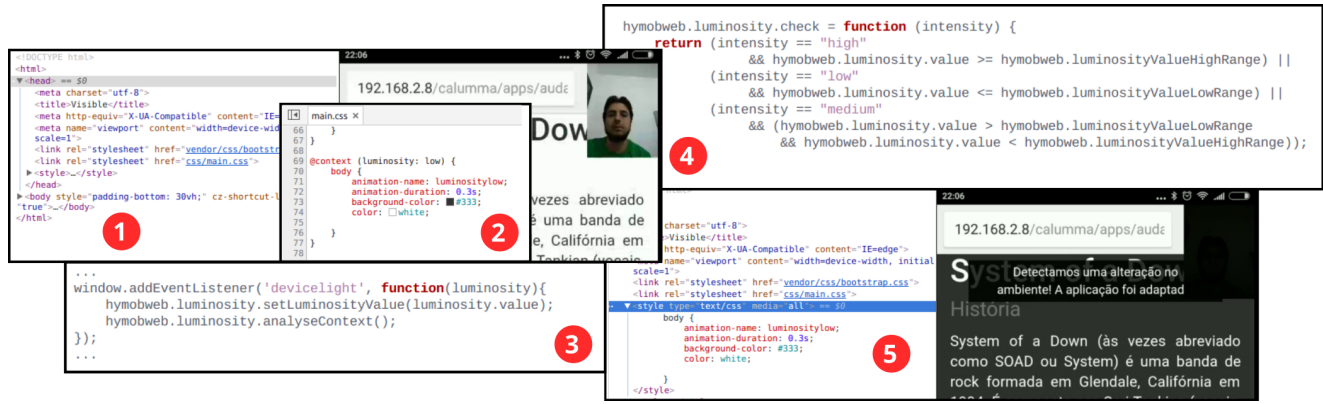
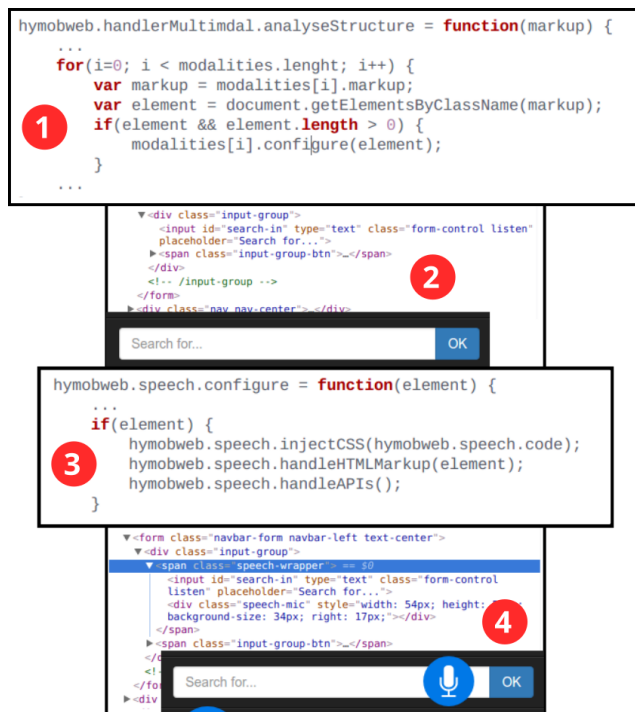
### 5.1. Context Handler

First, the *Context Handler* calls `loadContexts()` method that loads a list of all context aspects which are available (i.e. context variables that are implemented). We implemented these aspects by using regular expression. After loading the context aspects, the `analyzePresentation()` method performs the parsing into the CSS code of the application to search for matchings of code chunks and the regular expressions. Figure 3 shows two examples of regular expressions (read from the left to the right). The first one refers to the checking of







Figure 4: Example of dynamic adaptation of *Luminosity* aspect.Figure 5: Example of dynamic adaptation of *Speech* aspect.

pect it is not necessary to insert code in the application, because the marking and the visual of the element in question are not changed. The modality, in this case, will only be responsible to modifying the behavior of the element. The implementation was accomplished using the Touch Events API that is supported in Mozilla Firefox, Google Chrome, and Microsoft Edge browsers[4, 9].

## 6. Evaluation of the dynamic adaptation

As the dynamic adaptation has a direct impact on the use of the apps, our evaluation focused on exploring the HyMobWeb approach in the perspective of end-users. The data and results presented in this article have not published before. The evaluation in the perspective of developers (i.e. evalua-

tion of static adaptation) can be found in [5].

To examine the dynamic adaptation, we carried out a controlled experiment with the objective of verifying end-users feedback. This evaluation is relevant, since end-users are those who will use the resources provided by dynamic adaptation. We followed the Wohlin et al [24] guidelines to organize the planning, execution and analysis phases of this study.

### 6.1. Planning

Initially, we defined the research question (RQ): *How is the experience of end-users regarding the features provided from HyMobWeb dynamic adaptation?*. To answer our RQ, we took a comparison method. We developed two versions of a music player mobile web app. Both versions of the app provided features to browse among the albums available in the library, visualize the album's songs, play and pause a song, and turn the volume up or down.

The two versions were built by a developer with two years of experience in web development. First, the developer implemented a music player version only using the RWD resources available in *Bootstrap* FeF. We named it of *Base* version. This version had restrictions about the use of different modalities of interaction and adaptation to the context of use. For the second version, the developer extended the *Base* one by applying the resources of the HyMobWeb DSL (i.e. resources available in FeF extension - see Figure 2-A). The details about DSL can be found in Section 4. The developer added the features of adaptation modalities of interaction and context of use by applying the static adaptation. This second version was named *HyMobWeb*.

We elaborated four tasks (see Table 6.1) to guide the participants interaction during the study. For each task, we had different treatments to the two versions (i.e. the *Base* and *HyMobWeb* apps). A tutorial was developed to assist the end-users to accomplish the tasks.

Self-Assessment Manikin (SAM) instrument was adopted to collect the end-users feedback. SAM is a pictograph evaluation method to measure emotional responses of users after some sort of stimulus. Three dimensions are considered by this technique: pleasure (if the participant had a positive

**Table 1**

List of tasks - aspects of modalities of interaction (a) and context of use (b)

Task	Aspect	Base	HyMobWeb
(T1) Browse through the available albums until found a specific one	Movements (a)	Touch	Swipe
(T2) Search for a particular album through a keyword	Speech (a)	Typing	Speak
(T3) Read a text in the mobile app in a low-light environment	Luminosity (b)	Not available	Available
(T4) Select a song while walking	UserActivity (b)	Not available	Available

or negative reaction), arousal (body stimulation level from an event or object) and dominance (feeling in control of the situation or controlled by it). The user chooses a value on a scale of one to nine on each dimension, using images, to represent their emotions after each task [3]. Besides of SAM, a post-experiment questionnaire was applied to collect the participants' profile that includes questions about their age, education level, the frequency of Internet access and use of device resources, and their perceptions about the *Base* and *HyMobWeb* versions.

A pilot test was running with two individuals. This test aimed at checking whether the study instructions and procedures were sufficiently clear to the participants. We observed some difficulties of the individuals during the pilot test, and hence, we refined the tutorial.

## 6.2. Execution

A total of 14 participants took part in the study voluntarily and were selected by convenience. They were students, teachers and collaborators from the Federal Institute of São Paulo (IFSP), Itapetininga, Brazil. The study was carried out in the Informatics lab. All the volunteers signed the consent form in which the participants agreed about the use of their data for academic purposes.

All the participants were used mobile devices (smartphones) daily. Most of them aged 16 to 18 years old (75%), attended high school (91.7%), and accessed Internet through mobile devices frequently (95.4% access daily). To avoid problems during the experiment execution, two smartphones were provided to the participants during the study (one ASUS ZenFone 5 and one Xiaomi Redmi 4).

An initial explanation about the study conduction, its goal and evaluation process were done. The tasks were presented in a textual way on the mobile device screen. All the interactions of the participants were recorded by using the Ace Screen Recorder application<sup>7</sup> that was installed into the smartphones.

<sup>7</sup><https://play.google.com/store/apps/details?id=com.dev47apps.screenrecorder>

**Table 2**

Results: participants' interaction (ID), which version the participant has interacted first (APP) - Base (BA) or HyMobWeb (HW), and SAM indexes values - pleasure index (PI), arousal index (AI) and dominance index (DI).

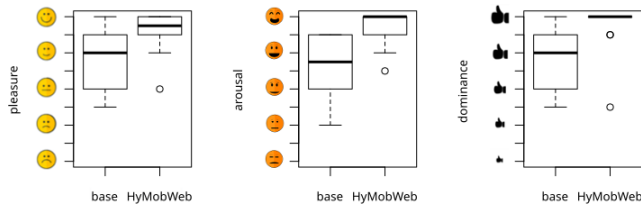
		(T1) Movements						(T2) Speech					
		Base			HyMobWeb			Base			HyMobWeb		
ID	APP	PI	AI	DI	PI	AI	DI	PI	AI	DI	PI	AI	DI
1	BA	4	4	4	8	9	8	3	4	4	9	9	9
2	HW	7	6	7	9	9	9	6	7	8	8	9	8
3	BA	5	3	5	8	8	9	8	6	8	9	8	8
4	BA	8	8	8	9	8	9	4	5	6	9	5	9
5	BA	8	8	8	9	9	9	6	6	5	9	9	9
6	BA	8	7	7	9	9	9	7	5	6	9	7	7
7	HW	8	8	9	9	9	9	9	7	9	9	9	7
8	HW	5	5	5	7	7	9	5	5	6	8	8	8
9	HW	8	8	9	5	6	4	7	8	7	9	9	9
10	BA	6	6	7	8	9	9	5	7	6	8	9	8
11	HW	9	7	9	9	9	9	9	9	9	7	9	6
12	BA	5	5	6	9	9	9	5	6	5	9	9	9
13	HW	7	8	5	8	8	9	9	9	8	6	7	5
14	BA	7	6	6	8	8	8	9	9	8	5	2	6
		(T3) Luminosity						(T4) UserActivity					
		Base			HyMobWeb			Base			HyMobWeb		
ID	APP	PI	AI	DI	PI	AI	DI	PI	AI	DI	PI	AI	DI
1	BA	5	4	5	8	9	9	3	2	5	7	8	9
2	HW	6	7	8	9	9	9	5	6	6	8	8	8
3	BA	5	4	8	9	8	8	5	6	4	9	8	9
4	BA	5	5	5	9	9	9	6	8	9	9	9	9
5	BA	6	6	5	9	9	9	8	8	8	9	9	9
6	BA	7	5	6	9	7	7	3	3	2	9	8	7
7	HW	7	7	9	9	9	7	7	7	9	9	9	8
8	HW	5	5	6	8	8	8	6	6	6	9	9	9
9	HW	7	8	7	9	9	9	9	8	8	6	8	6
10	BA	5	7	6	8	9	8	6	7	7	9	8	7
11	HW	7	7	9	9	9	9	8	6	9	9	9	9
12	BA	4	5	5	9	9	9	7	7	7	5	5	6
13	HW	5	2	4	9	9	9	7	6	3	9	7	9
14	BA	5	6	5	9	9	8	3	5	5	9	9	9

All participants used both app versions (i.e. *Base* and *HyMobWeb*). However, we performed a balance in the order in which the versions were used. From this balance, we avoid the introduction of biases in the results which could be caused by the order the versions were used. Hence some participants interacted first with the *Base* version and others with the *HyMobWeb* one.

After finishing each task, the participant answered the SAM questionnaire. At the end of all the tasks, the users responded the post-questionnaire composed of seven questions about their frequency of use of mobile devices and four non-mandatory questions about their perception about the both app versions.

## 6.3. Analysis

The SAM and post-questionnaire answers and the individual recordings of participants interaction were analyzed. To answer our RQ (*How is the experience of end-users regarding the features provided from HyMobWeb dynamic adaptation?*), we analyzed the videos contained the participants' interactions tasks and the SAM answers for each one. Additionally, our results were complemented by the responses of open-questions found in the post-questionnaire. Table 2 presents the results per participants for each task. We also associated the aspect (see Table 6.1) observed for each task.



**Figure 6:** Boxplots for the pleasure, arousal and dominance of the Movements aspect (T1).

#### 6.4. Threats of Validity

The validity of our study was treated by taking into account some strategies considering the levels of validity treatment: internal (i), external (ii), construction (iii) and completion (iv) [24].

Internal threat (i) refers to the tiredness of the participants. To mitigate this, we prepared a set of tasks that could be accomplished in 20 minutes. External threat (ii) refers to the sample. The participants of our study represented typical users of mobile devices. For them, *smartphones* are devices that take part in their daily so they could naturally perform the tasks.

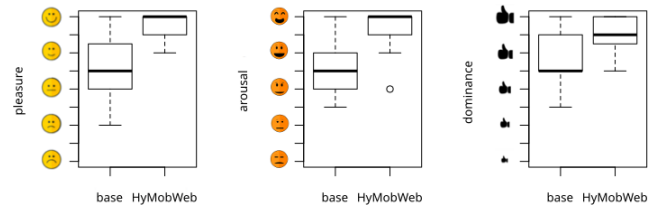
Construct threat was mitigated in two ways (iii). First, the participants' interaction was randomly divided into two groups. While one group had their first interaction by using the *Base* version the other group started its interaction with *HyMobWeb*. Afterwards, the groups changed the order of interaction. Besides, the tasks were shuffled, so the album requested to be located in the *Base* version was in a different position than the same one on *HyMobWeb* version.

We handled the threat of conclusion (iv) by crossing the data collected from the perspective of the participants (i.e. from using SAM) and from the analysis of the videos. This proceeds avoided our conclusions were driven by only one source of data.

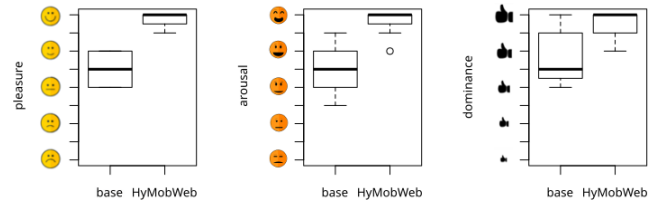
#### 6.5. Findings

Our results will be discussed for each aspect of modality of interaction or context of use related to the tasks (see Table 6.1). We took the descriptive statistics to support our considerations about the findings. Additionally, the results showed in Table 2 give the background to our discussions.

First, considering the *Movements* aspect (see Figure 6) related to the first task (T1 - see Table 6.1). The results show that SAM indexes have their values concentrated at the highest levels for the *HyMobWeb* version. The three dimensions (i.e. PI, AI and CI) revealed similar results for the *Base* version. This version presents a measure of central tendency equal to or less than 7. On the other hand, *HyMobWeb* version presents values of 8.5 for the indexes of pleasure (PI), and 9 for arousal (AI) and dominance (DI). In dominance index (DI), the concentration of the values is even more perceptible. In this dimension, only two users stand outside the maximum SAM scale. This result suggests the participants had a higher level of control while using the swipe mode than by using the touch interaction. However, some outliers was found (see Figure 6). In the post-questionnaire, we could



**Figure 7:** Boxplots for the pleasure, arousal and dominance for the Speech aspect (T2).



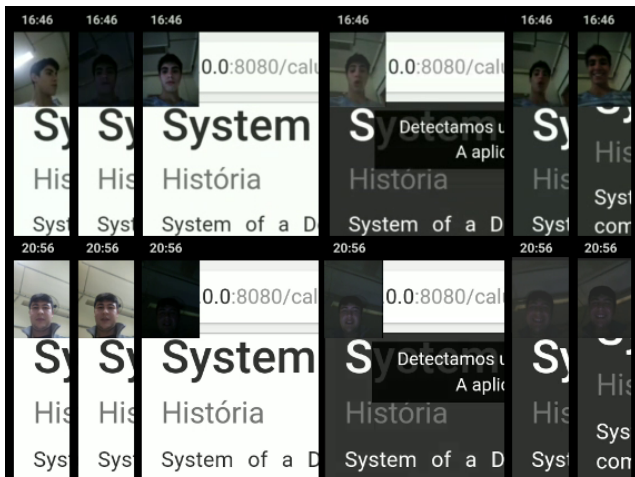
**Figure 8:** Boxplots for the pleasure, arousal and dominance indexes for the Luminosity aspect (T3).

found the explanation from a user statement: “*I was afraid to move in the albums on too fast and end up missing that one I wanted*”. By examining the videos, we could see the participants struggled in performing some touch interaction. As the *Base* version used the RWD resources, some icons were automatically reduced in their size. Additionally, we noticed that the users attempted to perform the swipe movement in the album gallery even when the instructions explicitly oriented to take the touch interaction.

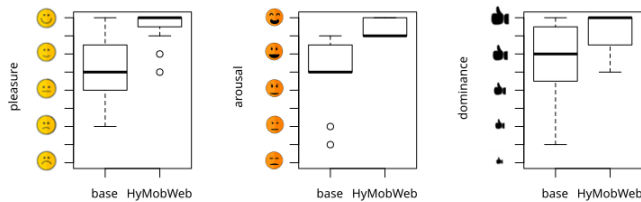
Figure 7 shows the results for the *Speech* aspect related to the second task (T2 - see Table 6.1). We see that the pleasure index values (PI) are greater in the *HyMobWeb* version than in the *Base* one. For the voice recognition resource to work, the internet connection should be available. Consequently, the problems caused by some interruptions in the Internet signal that happened may have impacted on the results of the arousal index (AI). Nevertheless, besides the speech mode, the participants had the typing mode available even while Internet access fails. Taking into account the post-questionnaire answers, we see that most of the participants stated they liked to keep the interaction by speech mode for being more practical. Some participants reported that they choose voice mode only in the situation that typing mode is not suitable, for instance, when they are driving a car.

The highest levels in all the SAM indexes were obtained by *Luminosity* aspect, which is linked to the third task (T3 - see Table 6.1). In Figure 8, we see that whereas the median remains at the highest level of the scale (i.e. 9) in the three indexes for the *HyMobWeb* version, for the *Base* one the value is 6. Some participants reported that “*...when the luminosity changes automatically the reading becomes more comfortable...*”; “*...this change avoids I squint my eyes to see better..*”. Besides, by observing the videos that contained the users' interaction, we could notice a positive facial expression as soon as the luminosity automatically changed (see Figure 9).

Finally, in the *UserActivity* aspect linked to the last task



**Figure 9:** Positive facial expression - reaction when the Luminosity aspect worked.



**Figure 10:** Boxplots for the pleasure, arousal and dominance for the UserActivity aspect (T4).

(T4 - see Table 6.1), we see again that the *HyMobWeb* version got the highest levels in the SAM indexes (see Figure 8). Google API<sup>8</sup> provided the resources to catch the user's movements (e.g. if they are standing, walking or moving). However, this API takes a few seconds for the recognition of the user activity. Consequently, the delays provoked by the API can be the cause of some outliers (see Figure 8). Although this affected the participants' feedback regarding the SAM indexes negatively, we could get some positive responses from the post-questionnaire as such: *"I liked this resource...while I am walking or running I could see the most important information in the same position on the screen...the things do not move from one position to another..."*; *"...the recognition of my moving improves my ability for interacting with the app..."*.

Retaking the RQ (*How is the experience of end-users regarding the features provided from HyMobWeb dynamic adaptation?*), we can respond that the resources of *HyMobWeb* dynamic adaptation could provide a good experience to end-users.

## 7. Discussion

Our findings revealed that the three SAM dimensions (i.e. pleasure, arousal and dominance) presented high values for the *HyMobWeb* version. We could notice that the use of different modalities of interaction made the user in-

teraction easier in some tasks. However, the aspects of the context of use were the ones that presented more impact on the user experience. Taking into account all the aspects we considered in this study, we can state that the Luminosity and User Activity were those that most have influence into the pleasure index values.

Different solutions for working with dynamic adaptation in the context of mobile web apps can be found in the literature [21, 1, 18, 2, 15, 20]. However, our approach presents new contributions. First, our approach concerns on providing resources for performing the app adaptation based on an extension of FeFs features and RWD techniques. As FeFs and RWD are widely known by software developers [17, 18] they will spend less time in learning how to apply the adaptation resources. Consequently, their learning curve for the implementation of the adaption resources will be reduced. Besides, *HyMobWeb* proposes a DSL and a structure to implement the apps adaptation that is flexible and easily extensible to different FeFs [13]. Given the flexibility of *HyMobWeb*, new handlers can be inserted into the approach in addition to Context and Modalities handlers. This flexibility allows that other resources could be implemented or customized to attend the developers' demands.

One of the advantages of *HyMobWeb* dynamic adaptation is that the values of context of use could be caught without using intelligent agents or external plugins. Differently, the works of [1] and [15] that request the use of additional resources. User actions can be monitored only using resources of HTML, CSS and Javascript that are native lightweight technologies for web development. Although such mechanisms have already been used in other approaches [2], in *HyMobWeb* proposal, these consider the standard structure of FeF. By following these structures, our approach becomes reusable for different apps.

Additionally, *HyMobWeb* provides ways to promote two different adaptations in the same app. One focused on the modality of interaction and another on variables of the context of use. The adaptation of modes of interaction and context of use had already been addressed in other [15, 18, 25]. However, as far as we know, our approach is the first that brings these two possibilities together.

## 8. Final remarks and future works

*HyMobWeb* is an approach that provides resources to perform adaptation on web mobile application. The core of the proposal is to allow developers to dealing with different modalities of interaction and with context of use through hybrid adaptation. During the coding time (i.e. static adaptation), developers can identify in the code which parts can be adapted during the run-time (i.e. dynamic adaptation). By extending the *Bootstrap* FeF, we provided concrete ways to use the *HyMobWeb* proposal. The evaluation of static adaptation was already done in another work.

In this work, we focused on presenting the dynamic adaptation of the approach and its evaluation. The dynamic adaptation was built from an extensible architecture based on han-

<sup>8</sup><https://developers.google.com/android/reference/packages>



dler components which give the flexibility to adding other kinds of user interaction variables. The implementation of variables of context of use could be used in combination with different modalities of interaction.

The evaluation from the end-users perspective showed that the adaptations provided by HyMobWeb can potentially improve the user experience in mobile web apps. Some limitations of our proposal are: HyMobWeb does not provide ways of end-users set their preferences and we did not explore the complexity of handling various contextual aspects and different modalities of interaction together.

As future works, we intend to explore other categories and elements of interaction to identify their problems and hence to propose solutions in the scope of HyMobWeb. We also intend to run testings to observe the performance of our approach in different settings of browsers and Internet conditions.

## 9. Acknowledgement

The authors thank the grant 2013/25572-7, São Paulo Research Foundation (FAPESP).

## References

- [1] Ahmadi, H., Kong, J., 2012. User-centric adaptation of web information for small screens. *J. Vis. Lang. Comput.* 23, 13–28. URL: <http://dx.doi.org/10.1016/j.jvlc.2011.09.002>, doi:10.1016/j.jvlc.2011.09.002.
- [2] Anam, R., Ho, C.K., Lim, T.Y., 2014. Tree Adapt: Web Content Adaptation for Mobile Devices. *International Journal of Information Technology and Computer Science* 6, 1–13. URL: <http://www.mecs-press.org/ijitcs/ijitcs-v6-n9/v6n9-1.html>, doi:10.5815/ijitcs.2014.09.01.
- [3] Bradley, M.M., Lang, P.J., 1994. Measuring emotion: The self-assessment manikin and the semantic differential. *Journal of Behavior Therapy and Experimental Psychiatry* 25, 49–59. URL: <http://www.sciencedirect.com/science/article/pii/0005791694900639>, doi:https://doi.org/10.1016/0005-7916(94)90063-9.
- [4] Brubeck, M., Moon, S., Barstow, A., Schepers, D., 2013 (accessed 13 March 2016). Touch Events. W3C Recommendation. W3C. <http://www.w3.org/TR/2013/REC-touch-events-20131010/>.
- [5] Bueno, D.C., Zaina, L.M., 2017. Hymobweb: A hybrid adaptation of context-sensitive web interfaces with multimodality support in mobile devices. *SBC Journal on Interactive Systems* 8, 20–34.
- [6] Burnett, D., Shuang, Z.W., 2010 (accessed 13 April 2016). Speech Synthesis Markup Language (SSML) Version 1.1. W3C Recommendation. W3C. <http://www.w3.org/TR/2010/REC-speech-synthesis11-20100907/>.
- [7] Cirilo, C.E., Prado, A.F., de Souza, W.L., Zaina, L.M., 2010. A hybrid approach for adapting web graphical user interfaces to multiple devices using information retrieved from context., in: *The 16th International Conference on Distributed Multimedia Systems*, pp. 168–173.
- [8] Desruelle, H., Blomme, D., Gielen, F., 2011. Adaptive Mobile Web Applications Through Fine-Grained Progressive Enhancement. *The Third International Conference on Adaptive and Self-Adaptive Systems and Applications*, 51–56 URL: [http://www.thinkmind.org/index.php?view=article{&articleid=adaptive}\\_{2011}\\_{3}\\_{20}\\_{50030}](http://www.thinkmind.org/index.php?view=article{&articleid=adaptive}_{2011}_{3}_{20}_{50030}).
- [9] Deveria, A., 2017 (accessed 01 April 2017). Can I use... Support tables for HTML5, CSS3, etc. URL: <https://caniuse.com/>.
- [10] Dumas, B., Solórzano, M., Signer, B., 2013. Design guidelines for adaptive multimodal mobile input solutions, in: *Proceedings of the 15th International Conference on Human-computer Interaction with Mobile Devices and Services*, ACM, New York, NY, USA. pp. 285–294. URL: <http://doi.acm.org/10.1145/2493190.2493227>, doi:10.1145/2493190.2493227.
- [11] Fabri, D.W., Krempser, T., Filgueiras, L.V.L., 2013. Estudo de responsive web design aplicado a um sistema de pesquisa de opinião na área médica, in: *Proceedings of the 12th Brazilian Symposium on Human Factors in Computing Systems*, Brazilian Computer Society, Porto Alegre, Brazil, Brazil. pp. 264–267. URL: <http://dl.acm.org/citation.cfm?id=2577101.2577159>.
- [12] Ghiani, G., Manca, M., Paternò, F., Porta, C., 2014. Beyond responsive design: Context-dependent multimodal augmentation of web applications, in: *Proceedings of the 11th International Conference on Mobile Web Information Systems*, Springer International Publishing, Cham. pp. 71–85. doi:10.1007/978-3-319-10359-4\_6.
- [13] Iqbal, M.W., Ahmad, N., Shahzad, S.K., Feroz, I., Mian, N.A., 2018. Towards adaptive user interfaces for mobile-phone in smart world. *International Journal of Advanced Computer Science and Applications* 9. URL: <http://dx.doi.org/10.14569/IJACSA.2018.091177>, doi:10.14569/IJACSA.2018.091177.
- [14] Kostiaainen, A., 2016 (accessed 13 March 2016). Ambient light sensor. <https://www.w3.org/TR/2016/WD-ambient-light-20160830/>.
- [15] Manca, M., Paternò, F., Santoro, C., Spano, L.D., 2013. Generation of multi-device adaptive multimodal web applications, in: *Proceedings of the 10th International Conference on Mobile Web Information Systems - Volume 8093*, Springer International Publishing, New York, NY, USA. pp. 218–232. URL: [http://dx.doi.org/10.1007/978-3-642-40276-0\\_17](http://dx.doi.org/10.1007/978-3-642-40276-0_17), doi:10.1007/978-3-642-40276-0\_17.
- [16] Marcotte, E., 2017. Responsive web design: A book apart n4. A Book Apart, Eyrolles.
- [17] Nebeling, M., Norrie, M.C., 2014. Beyond responsive design: Adaptation to touch and multitouch, in: *Web Engineering*, Springer International Publishing, Cham. pp. 380–389. doi:10.1007/978-3-319-08245-5\_23.
- [18] Nebeling, M., Speicher, M., Norrie, M., 2013. W3touch: Metrics-based web page adaptation for touch, in: *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, ACM, New York, NY, USA. pp. 2311–2320. URL: <http://doi.acm.org/10.1145/2470654.2481319>, doi:10.1145/2470654.2481319.
- [19] Otto, M., Thornton, J., et al., 2015 (accessed 20 April 2015). Bootstrap: the world's most popular mobile-first and responsive front-end framework. URL: <https://getbootstrap.com/>.
- [20] Park, S., Gebhardt, C., Rädle, R., Feit, A.M., Vrzakova, H., Dayama, N.R., Yeo, H.S., Klokmoose, C.N., Quigley, A., Oulasvirta, A., Hilliges, O., 2018. Adam: Adapting multi-user interfaces for collaborative environments in real-time, in: *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems*, ACM, New York, NY, USA. pp. 184:1–184:14. URL: <http://doi.acm.org/10.1145/3173574.3173758>, doi:10.1145/3173574.3173758.
- [21] Paternò, F., Zichittella, G., 2010. Desktop-to-mobile web adaptation through customizable two-dimensional semantic redesign, in: *Proceedings of the Third International Conference on Human-centred Software Engineering*, Springer-Verlag, Berlin, Heidelberg. pp. 79–94. URL: <http://dl.acm.org/citation.cfm?id=1939212.1939221>.
- [22] Spurlock, J., 2013. Bootstrap: Responsive Web Development. O'Reilly Media.
- [23] Toxboe, A., 2016 (accessed 15 April 2016). UI-Patterns. URL: <http://ui-patterns.com/>.
- [24] Wohlin, C., Runeson, P., Hst, M., Ohlsson, M.C., Regnell, B., Wessln, A., 2012. Experimentation in Software Engineering. Springer Publishing Company, Incorporated.
- [25] Yigitbas, E., Josifovska, K., Jovanovikj, I., Kalinci, F., Anjorin, A., Engels, G., 2019. Component-based development of adaptive user interfaces, in: *Proceedings of the ACM SIGCHI Symposium on Engineering Interactive Computing Systems*, ACM, New York, NY, USA. pp. 13:1–13:7. URL: <http://doi.acm.org/10.1145/3319499.3328229>, doi:10.1145/3319499.3328229.



# Journal of Visual Language and Computing

journal homepage: [www.ksiresearch.org/jvlc](http://www.ksiresearch.org/jvlc)

## An Advanced Software Dashboard for Supporting the Development of Adaptable, Energy Efficient Applications<sup>★</sup>

Vladimir Ivanov<sup>a</sup>, Dragos Strugar<sup>a</sup>, Giancarlo Succi<sup>a</sup> and Eugene Zouev<sup>a</sup>

### ARTICLE INFO

#### Article History:

Submitted 4.8.2019

Revised 4.30.2019

Second Revision 5.20.2019

Accepted 8.15.2019

#### Keywords:

energy efficiency  
dashboards

### ABSTRACT

Developing features based solely on requirement documents and specifications has been a traditional way of building software. This paper provides a different approach by combining the notions from Artificial Intelligence (AI) - Evolutionary Algorithms (EA) and Complexity Theory. It represents the software to be build - a dashboard - as a Complex System, and metrics inside of them as agents that are interacting in a non-linear fashion. Furthermore, the paper suggests different visualization techniques required to represent such complex relationships.

© 2019 KSI Research

## 1. Introduction

The goal of our ongoing research was to come up with a self-adjusting metric analysis system that would allow managers as well as engineers to make more informed decisions in the development of Adaptable, Energy Efficient software systems. Systems that pervade our everyday life are inherently dynamic since they need to operate in a continuously changing environment and must be able to quickly react and adapt to different types of changes, even unanticipated while guaranteeing the efficient use of the available resources.

This scenario poses to software engineering a new challenge that concerns the development of software for Adaptable, Energy Efficient systems, i.e. systems that explicitly characterize the resources under control, that dynamically evolve to maintain an acceptable consumption of resources.

In the paper we discuss solutions that help defining and validating a quantitative framework to guide the development and the evolution of software systems using a variety of metrics collected throughout the life-cycle of software systems, from the initial concept to the deployment, execution, and maintenance, optimizing the performances of the sys-

tems under a variety of nowadays relevant factors, including quality, productivity, efficient use of resources.

Developing such software products first requires an understanding of the problem that is expected to be solved. Therefore, developing of adaptable software systems requires a customizable dashboard, which is a well-proven solution for this purpose. Thus, we have created a web application to demonstrate our vision of the dashboard best-suited for this use case. The purpose of the global data collection system, to which the dashboard belongs to, is to present the obtained data in a specific structure and form, as well as to provide the users with an overview of the collected information.

Numerous academic papers emphasize the importance of the good design for the dashboard. In [8], Few points out that currently the entire purpose of dashboards is not only to display all the necessary information, but rather to provide a medium for communication and team collaboration.

The literature review section is primary based on the works of Zorin [11], Pishulin [10], whose preceding studies grounded the system that we are currently developing. It is also based on the work of Sarikaya, Correll, Bartram, Tory, and Fisher [20], who performed a systematic literature review on designing dashboards for different domains. These authors split the dashboards into types by different criteria, and explained the common patterns and features for each of them. Finally, it is based on the work of Yigitbasioglu, and Velcu [26], who also reviewed many sources for designing the dashboard, and suggested the mechanism for making decisions in their func-

<sup>★</sup>This document is the results of the research project funded by the Russian Science Foundation, grant # 19-19-00623.

✉ [v.ivanov@innopolis.ru](mailto:v.ivanov@innopolis.ru) (V. Ivanov); [d.strugar@innopolis.ru](mailto:d.strugar@innopolis.ru) (D. Strugar); [g.succi@innopolis.ru](mailto:g.succi@innopolis.ru) (G. Succi); [e.zouev@innopolis.ru](mailto:e.zouev@innopolis.ru) (E. Zouev)

ORCID(s):

<sup>1</sup>Innopolis University, Russian Federation

tionality.

Section 2 presents in greater detail the common problems which developers of dashboards are likely to encounter. Section 3 summarizes fundamental goals and metrics of the GQM (Goal Question Metric) model, as required by software engineers for the evaluation of their performance. Section 4 covers possible features of the future dashboard. Section 5 lists some common visualization patterns. Section 6 shows the overall design of the working solution. Then, sections 7 and 8 talk about the Complex Adaptive Systems and how they map to our use case. And finally, section 9 gives the final thoughts and reflections.

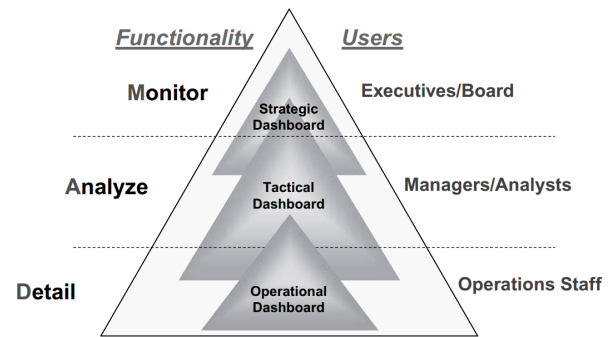
## 2. Problems Designation

One of the main challenges in dashboard design is selecting appropriate metrics and displaying them in a meaningful and structured way. This choice has great influence on the understanding of those metrics and it is particularly important in agile development environments [14], especially in the context of mobile development [21, 5, 4, 3], to track open source projects [23, 15, 18, 19, 6], in domain-centered development [24, 25], or in safety critical situations. The dashboard should “provide intuitive, actionable, flexible, and programmable visualization to support effective decision making.” In addition, effective representation of selected metrics is also a great challenge. Solving this problem would allow users to easily detect and address the issues that may occur during the development process.

To solve these problems and choose more appropriate visualization techniques for more effective representation, Brath and Peters [1] suggest answering the following three questions:

- What metrics does the user need to see?
- What context does each metric require to make it meaningful?
- What is the visual representation that best communicates the metric?

In [22], the authors suggest to focus on the important metrics, with concise visualization. Such an approach that reduces redundancy by focusing on the goals obtained during the requirement collection and analysis phase is very useful. Moreover, it helps to cope with making the wrong decisions in certain situations [26]. A similar view on data visualization is presented in [12], where the authors insist that the dashboard should be designed to be useful. This does not mean that it should contain all possible information that can be visualized, but only the necessary and sufficient data. The authors support this idea by suggesting the concepts of the “right” data and “right” visualization technique. Therefore, it is first necessary to determine which metrics should be displayed and explain the rationale of their choice. Second, choose the way to present them to the user, which can consequently minimize the time required for understanding them.



**Figure 1:** Correspondence of the dashboard types and functionality. (Taken from [7, p. 103])

In addition, as noticed in [20], the design of the dashboard and its functionality greatly depends on its type. Few [8] and Eckerson [7] introduced three types of dashboards. Eckerson represents them in the form of a pyramid, with the operational dashboard type at the base, the tactical in the middle layer, and the strategic at the top (as shown in Figure 1). The three dashboard types differ in purpose and level of abstraction and interaction. According to this model, the strategic view is intended for monitoring the current situation. This allows minimum interaction with the user, contains an overview and consists of the most meaningful data. This is because, as mentioned in [20], the people often make screenshots and put them onto slides for showing a general picture. The purpose of the tactical view is to enable the analysis. This should contain more detailed information from the tactical view. The operational view should present a thorough form of the collected data and metrics. That would allow the user to summarize and also would help to find the reasons which led to this state, and come to possible solutions. In addition, according to the results of the surveys by Zorin, the operational becomes more effective than the tactical and the strategic dashboard type.

In [26], Yigitbasioglu and Velcu suggest a path for making design decisions, that requires determining the four characteristics of a dashboard under development:

1. its purpose is enabling
  - consistency,
  - monitoring,
  - planning, and,
  - communication,
2. users and their
  - tasks,
  - knowledge and
  - cognitive styles,
3. design features (functional or visual) and
4. outcomes, that is performance progress reached by this dashboard (i.e. improved speed, consistency etc.).

Taking into account that respondents of Zorin's surveys are potential users, we can presume the variance in their education levels, age, company size and the position they are working on. Users prioritized their needs in the dashboard in the following order from the most to the least demanded:

1. performance monitoring;
2. planning;
3. communication;
4. measurement consistency.

Therefore, these needs along with their priorities should be taken into account when designing a new dashboard from scratch. The next couple of sections provide a more comprehensive description of how these can be formed in the GQM model.

### 3. The main goals and metrics

For selecting the necessary metrics in [12], the authors suggest using GQM+Strategies. This concept is a traditional Goal-Question-Metric approach supplemented by the links between different layers of organizational goals, i.e., high-level and measurement goals [17]. According to this approach, the authors ensure that a good dashboard should meet business goals. Hence, the development of such a dashboard requires the participation of all stakeholders.

As part of a study, Zorin conducted surveys with the representatives of software engineering companies, and extracted six common goals they wanted to achieve:

- Improving effort estimation efficacy;
- Using resources in a more efficient way;
- Executing testing activities in a more efficient and systematic way;
- Improving the quality of the development process;
- Completing projects successfully;
- Completing projects phases successfully.

Later in [11], Zorin et al. summarize initial results, and extract three main goals:

- more effective effort estimation;
- more efficient use of resources;
- better software quality and development process.

The metrics needed for evaluating the achievement of these goals were divided into five groups, displaying the:

- progress status of the project;
- speed of the work performed;
- status of testing;
- status of software quality;

- effectiveness of effort estimation.

In addition, Zorin distinguishes the most frequent metrics:

1. iteration Burndown chart;
2. team velocity;
3. code coverage;
4. effort estimation accuracy.

The summary of the results in the GQM model, shown in Figure 2.

In [10], Pishulin et al. validated the results of the surveys by Zorin [11]. The authors investigated three goals highlighted in those studies, and determined the most suitable metrics for measuring them. For assessing the effectiveness of effort estimation, they identified the following key metrics:

1. iteration Burndown;
2. effort Estimation Accuracy;
3. team Velocity.

For the evaluation of software quality and development process, the metrics are:

1. passed tests;
2. code coverage;
3. unresolved defects;
4. class / method length;
5. iteration burndown;
6. defect removal;
7. defected density.

For the goal of more efficient use of resources, the author did not provide any information due to the lack of obtained information.

For describing the available manipulations, one can split metrics to three types:

- raw,
- composite
- expression.

The raw metrics are those extracted directly from the collected user activities and source code (primary set of metrics described above). The composite metrics are the ones constructed from two raw or composite metrics by applying:

- Simple arithmetic operations:
  - Addition;
  - Subtraction;
  - Multiplication;
  - Division;
- Simple mathematical functions:
  - Average;
  - Maximum;
  - Minimum.

The expression metrics are obtained from one or more Raw or Composite type, and can be aggregated by some user-defined arithmetic expression.

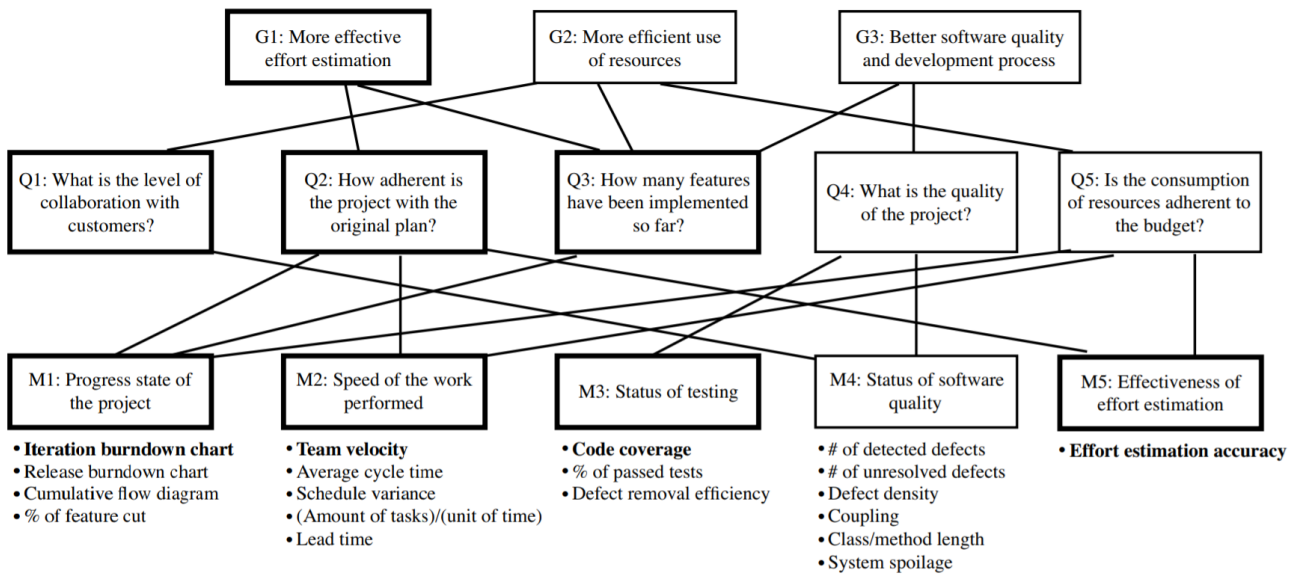


Figure 2: Summary of the GQM with the most common aspects in bold (Taken from [11])

#### 4. Functionality

The main mission of the dashboard is allowing the user to monitor key metrics in terms of completion of predefined goals. Furthermore, the dashboard should provide various tools for tracking the development progress and the product quality, and executing comparison of the current values to the perfect (or expected) values. Despite the amount of functionality, it is substantial to fit the dashboard to a single computer screen without compromising the content [8, 26].

In [12], the authors point out that understanding the displayed data should require the minimum effort from the user. The most relevant information should be provided with a “push” strategy. But at the same time, support interactivity with the ability to switch to “pull” mode. Moreover, the authors notice that the most relevant information should attract user’s attention. However, at the same time, it is very important to find a balance and avoid a motley design. The authors suggest displaying the same data every time in the same place in order to make users become accustomed to the design.

One of the main features that the dashboard should implement is alerting users about the deviations of measurements from normal values. For example, red could be used for going out of predefined by the user predefined range, green for those in the range and any other for neutral or default. [12, 20] covered an idea of using arrows for showing the current tendency. The method in [7] can be useful for color-blind people to help them with correct interpretation of results. At the same time, it is significant not to overuse colors, visual structures and other catchy elements of design, in order not to overload the view [26].

According to [20], it is significantly important to make the dashboard customizable and adaptable to different users and situations. This objective can be obtained by allowing

the users to set the goals, by themselves select the metrics to be displayed and set their own admissible and critical borders. Furthermore, the author draws our attention to the fact that filters [20] and comparison support [20] are able to supplement the dashboard with more flexibility.

Another frequently mentioned concept lies in presenting different views on collected data and metrics, which can be reached by two ways. The first one is to introduce roles with specific goals. Therefore, according to the goals, different analysis and visualization techniques should be selected with different points of view about the data [2]. In addition, roles can help to solve the problem of data visibility and privacy [20].

The second one is providing different levels of abstractions and different degree of details presented to the user. For introducing the general view about the situation, there should be a way to combine many metrics to several numbers, and avoiding both detailed and redundant information [2]. More complete and thorough views can be reached by drill-down navigation or an “exploration mode” [8, 16, 20]. It means that for “exploring” the nature of a certain result, it should be possible to switch between a generic and a more detailed view of the same metrics down to the raw data view. Using this technique would provide users with convenience in accuracy of data analysis, performed automatically [20].

According to [10], respondents want to have the ability to watch individual team members’ metrics that can be satisfied by introducing hierarchical views to collected data on the part of individual developers, teams and even the whole company. Furthermore, interviewees would like the dashboard to display possible reasons of metrics changes and recommendations for recovering and improvement of current situation. Thus, the what-if simulation can take place [20].



## 5. Visualization

Considering the structure of the dashboard we are relying on solutions from [12, 13] suggests the design based on Andon board with tiles, where each tile shows a numeric metric. This representation can speed up the understanding of the data and decision making as well as providing a unified way of metrics visualization in order to maintain the system scalability. However, there pointed a need in specific views for some metrics.

In [26], the authors make a investigation on data representation formats. Some researchers prefer graphical forms to tabular ones, whereas some place tables over graphs. Others do not exalt any of them believing that the choice of representation format highly depends on the task it is intended to solve. Graphs are more useful for tasks implicating comparisons and studying relationships of data, while the tabular form is more suitable for obtaining particular information. Thereby, the authors admit possibility of switching to the displaying format more preferable by the user.

[9] suggests using different treemaps, texture, and bump mapping; animated zooming and panning for visualization of metrics. However, in [7] the author underlines that sometimes even at first glance simple features can bring additional complexity. The ease of use becomes one of the most important characteristic of any product. As noticed in [20], some authors even suggest reducing the interactivity for simplifying the system. Therefore, finding a balance in flexibility and customization is crucial.

## 6. Results

This section focuses on establishing a strong relationship between the metrics and the goal. We are also going to present screenshots and various UI components we created. Results are presented in a form of a web page built with state-of-the-art front end web development library React.js.

As we noted earlier, the combination of flexibility and customization is a major factor in the User Experience (UX). That is precisely why we chose to create numerous widgets that users have control over. These widgets encapsulate graphs, charts, percentages, numerical values and time management tools alongside other key functionalities. They are designed to be easily reusable and re-sizable. This also contributes to the fact that the solution is fully responsive and works on all screen sizes. Some of the examples you can see on Figures 4, 5, and 6.

Figure 3 showcases all the parts of the dashboard. The page contains 14 metrics selected for the study. Each of these metrics is placed in one of the widgets, or so called tiles. There is a total of 3 bar graphs, 2 line charts, 4 percentage-based values, 4 numerical values and 1 date value.

This prototype assumes that the software development team is using the Scrum methodology, with 7 iterations completed. Each iteration is two weeks long. The task of this prototype is to find out if the relationship between the metrics and the goal is obvious to the engineers. In addition, it

is also very important to immediately see the most representative metrics for the Metric-Goal relationship.

This is precisely why we chose to make the connection with Adaptable systems. Not only would applying the notions from the Complexity Theory allow the dashboard to present the most important metrics for certain situations, but it would also prevent it from being susceptible to change.

## 7. Complex Adaptive Systems

Realizing that the dashboard needs to adapt to rapid change in today's software development life-cycle, we suggest that modern dashboards should be a part of any complex software system, and more specifically, the Complex Adaptive Systems. We represent each of the software development metrics, depicted as widgets on the 3, as agents within the system. Some of them are dependent on one another and some of them are not. The ones which are, are not strictly interacting in a linear, predictable way.

By embracing adaptation these agents have the potential to synchronize their internal states with the other agents in the system. Additionally, the system should be able to recognize these changes and self-adjust with the emergence of globally coherent patterns of adjustment developing.

Then, this Complex Adaptive System, the dashboard, should be able to feed back this information to micro-level agents. In order to preserve the relevance of certain metrics, the system makes the natural selection based on their fitness criteria to the environment. In our case, the fitness function is broadly defined as follows: An agent has a higher contribution/fitness to the overall system if and only if a slight change in that specific metric would yield a significant change in the overall system fitness, the difference between the expected value and the actual value is above average, or a metric answers some custom questions that dashboard users may have.

Applied and contextualized to our use case, the metrics are more relevant if the current value greatly differs from the expected one, or a slight change in a specific metric may yield a substantial increase in the well-being of the entire system. Such a system embraces the complex dynamic between the micro-level components (metrics) and the overall system (the dashboard). The interaction between the differentiation of micro- and macro-level agents with different goals and agendas creates the core dynamic of complexity in our system.

Intrinsic to the Adaptive Systems is the notion of innovation; i.e. coming up with novel outcomes that we could not have predicted ahead of time. This is the crucial aspect of designing and developing a dashboard engineered to dynamically present the most useful metrics to the person using the dashboard. For example, a correlation between some two metrics may yield a significant improvement in the overall system functioning. However, that relationship has not been made beforehand. It is only by feeding the data to the system that it is able to come up with the correlations which may drastically change the overall system fitness.



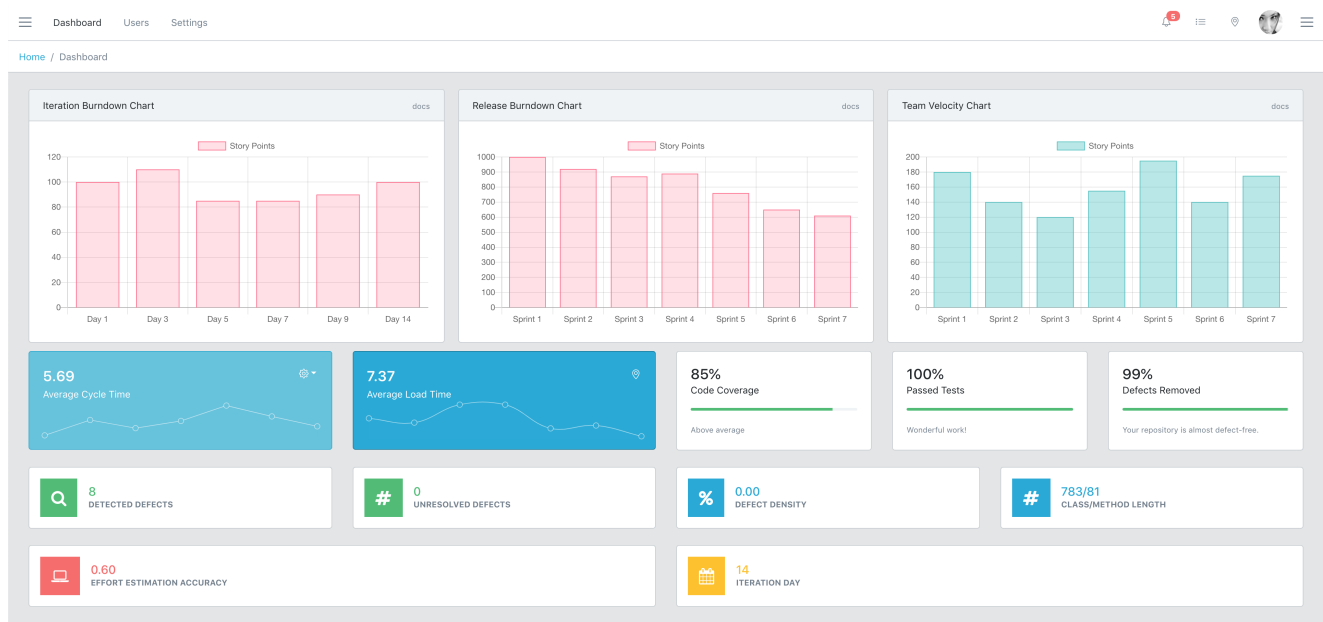


Figure 3: HTML Web Interface Results (InnoMetrics)

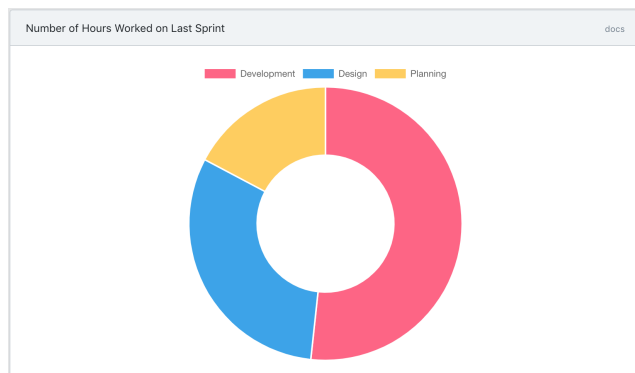


Figure 4: Doughnut Graph Representation

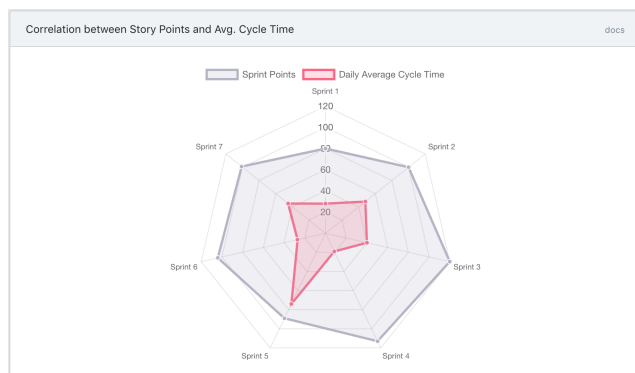


Figure 5: Polar Graph Representation

## 7.1. Non-linear Dynamics

We realize that common assumptions that managers and stakeholders may have when creating and maintaining the dashboards for their products are very often not close to the

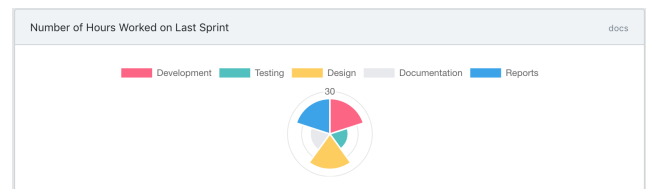


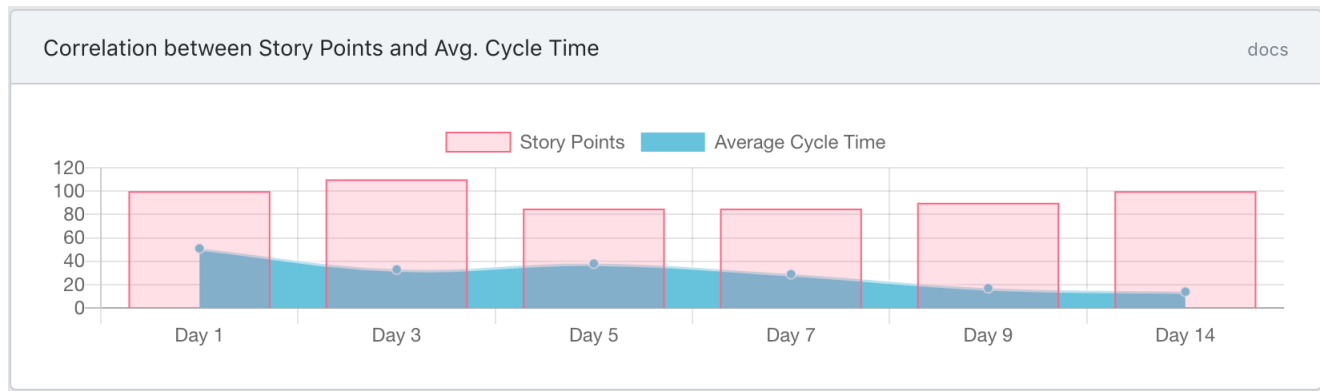
Figure 6: Radial Graph Representation

real-world outcome. For example, they may think that one metric may be very important, and that turns out not to be true. And over time that results in a waste of their time looking at the metric and trying to improve it, while not focusing on the metrics that are actually significant at the moment. Put differently, the common assumptions that stakeholders may have about predicting the outcome based only on the initial input do not often work out in Complex Adaptive Systems due to their emergent complexity. Non-linear dynamics in adaptive systems is necessary to constantly change the internal states of the agents within the system, resulting in the change the entire system's state.

## 7.2. Dynamic Equilibrium

Utilizing the concept of the Dynamic Equilibrium allowed us to embrace our GQM model to come up with the state which has the following characteristics:

- the current state is never completely stable, which results in the full stagnation;
- the current state is never in complete chaos, where there is nothing to bind individual actors together;
- the current state is always in a so called Dynamic Equilibrium where all actors are loosely bound to each other



**Figure 7:** Correlation between Story Points and Average Cycle Time

with the plethora of room to innovate and improve.

## 8. Applying the Complex System techniques to our use case

As previously noted, the agents in the Complex System are represented as widgets. Each of these widgets encapsulates one metric that is used for maintaining and monitoring the software development process. Different visual representation apply to each of these metrics, of course. Some of them are more useful and easier to think about when depicted as continuous graphs, some of them as plan numbers, as seen on Figure 3.

It is then challenging to decide how these metrics can bring the variety, and the non-linear dynamics that Complex Systems impose. There is a total of 14 metrics. It is only by combining them that we will be able to achieve what we have suggested above. We needed to assume a synergy between multiple metrics to see how they correlate to one another and how the increase in a specific factor or a group of factors can influence the whole system.

That is specifically why we decided to create custom widgets that represent the correlation between two or more metrics in the system. Even though such functionally complex widgets may contain a lot of data, as they are a relevant component to look after, they play a crucial role in the self-sustainability and continuous evolution of the dashboard system.

An example of such a combination is shown in Figure 7. One can immediately notice a very significant property. It is that the graph is now a complex structure consisting of two sub-graphs: a bar graph and a colored line graph. When the main 14 metrics we first presented, none of the structures had more than one metric encapsulated within the widget. Now widgets containing these complex forms with not only two but potentially more sub-graphs indicate a special correlation between them. Let us take a closer look at the Figure 7 and determine why this graph was chosen to be present to the manager who is responsible for monitoring the dashboard.

To start off, it is important to note that this graph spans the time period of exactly 14 days, the duration of one Sprint.

Therefore, the X axis represents the corresponding stage of the Sprint. Next, the Y axis is responsible for showing the number of Story Points that a team has managed to obtain on the specific day. It also shows the Average Cycle Time of a task in the sprint (measured in minutes).

From the chart it is also easy to notice that the cycle time is fairly low, indicating that the tasks are well segmented and there are a lot of them, as the number of story points is not low. However, the reason this graph is particularly interesting is that the Average Cycle Time graph is trending downwards. That means that it took more time for developers and designers to complete the tasks at the beginning of the Sprint rather than on the end of the Sprint.

There are numerous possible reasons for that, but the manager can be certain that his/her team did not try to hurry up at the Sprint end to finish most of the tasks because the number of Story Points is evenly distributed throughout the week. The manager might be misled to thinking that that is the case only by looking at the Average Cycle Time. The combination of the two graphs frees him/her from that suspicion, and now he/she can focus on other, less critical reasons why that is the case.

Here are some of the possible reasons:

- the scope/size of the tasks is not uniformly distributed throughout the Sprint;
- the team decided to first finish the complex tasks before moving on to easier ones.

As one can see, neither of the two reasons are as critical as the team doing most of the work at the Sprint's end. Thereafter, the widget managed to effectively show that information without the manager wasting his/her precious time on tracking down the issue that did not exist in the first place. This is an example of an effective combination of the two metrics that is far more useful when analyzed together rather than separately. It was determined to be important due to the fact that the trend was noticed in the Average Daily Cycle Time, and that such regularity did not cause any fluctuation in the number of Daily Story Points in the Sprint. Rather than looking at the raw data and realizing that his/her team is

trying to do most of the work at the end of the Sprint, manager focused his/her attention to other problems and ways the workflow can be improved. The dashboard served as a helper tool to get the job done by enhancing workers' expertise and not getting in the way of already productive existing workflow pipelines.

## 9. Conclusion

We have shown our previous work in the field of Software Metrics and suggested a possible way of improving it, as well as to give rise to the importance of designing and developing compelling dashboards. We have previously decided on 14 key metrics for the software development process. Now, by continuous improvement and natural selection of the most relevant agents the dashboard should be able to achieve the dynamic equilibrium using non-linear dynamics. The dashboard would also be able to combine several metrics that have a higher chance of indicating a possible flaw in the workflow. That does not necessarily mean that it is going to achieve the optimal state right away, nor does it mean that just by applying the techniques presented here will we truly solve the problem of inadequate dashboards in general. But rather this all means that our dashboard will try to get there over time, by facilitating adaptation, collaboration and expertise while avoiding chaos.

## References

- [1] Brath, R., Peters, M., 2004/10. Dashboard design: Why design is important. DM Direct 85. URL: [http://cs.furman.edu/~pbatchelor/csc105/articles/TUN\\_DM\\_ONLINE.pdf](http://cs.furman.edu/~pbatchelor/csc105/articles/TUN_DM_ONLINE.pdf).
- [2] Coman, I.D., Sillitti, A., Succi, G., 2009. A case-study on using an automated in-process software engineering measurement and analysis system in an industrial environment, in: Proceedings of the 31st International Conference on Software Engineering, IEEE Computer Society, Washington, DC, USA. pp. 89–99. URL: <http://dx.doi.org/10.1109/ICSE.2009.5070511>, doi:10.1109/ICSE.2009.5070511.
- [3] Corral, L., Georgiev, A.B., Sillitti, A., Succi, G., 2013a. A method for characterizing energy consumption in Android smartphones, in: Green and Sustainable Software (GREENS 2013), 2nd International Workshop on, IEEE. pp. 38–45. URL: <http://dx.doi.org/10.1109/GREENS.2013.6606420>, doi:10.1109/GREENS.2013.6606420.
- [4] Corral, L., Sillitti, A., Succi, G., 2013b. Software development processes for mobile systems: Is agile really taking over the business?, in: Engineering of Mobile-Enabled Systems (MOBS), 2013 1st International Workshop on the, pp. 19–24. doi:10.1109/MOBS.2013.6614218.
- [5] Corral, L., Sillitti, A., Succi, G., Garibbo, A., Ramella, P., 2011. Evolution of Mobile Software Development from Platform-Specific to Web-Based Multiplatform Paradigm, in: Proceedings of the 10th SIGPLAN Symposium on New Ideas, New Paradigms, and Reflections on Programming and Software, ACM, New York, NY, USA. pp. 181–183. URL: <http://doi.acm.org/10.1145/2048237.2157457>, doi:10.1145/2048237.2157457.
- [6] Di Bella, E., Sillitti, A., Succi, G., 2013. A multivariate classification of open source developers. Information Sciences 221, 72–83.
- [7] Eckerson, W.W., 2010. Performance Dashboards: Measuring, Monitoring, and Managing Your Business. Wiley. URL: <https://www.amazon.com/Performance-Dashboards-Measuring-Monitoring-Managing/dp/0470589833?SubscriptionId=AKIAIOBINVZYXZQZ2U3A&tag=chimbori05-20&linkCode=xm2&camp=2025&creative=165953&creativeASIN=0470589833>.
- [8] Few, S., 2006. Information Dashboard Design: The Effective Visual Communication of Data. O'Reilly Media. URL: <https://www.amazon.com/Information-Dashboard-Design-Effective-Communication/dp/0596100167?SubscriptionId=AKIAIOBINVZYXZQZ2U3A&tag=chimbori05-20&linkCode=xm2&camp=2025&creative=165953&creativeASIN=0596100167>.
- [9] Holten, D., Vliegen, R., van Wijk, J., 2005. Visual realism for the visualization of software metrics, in: 3rd IEEE International Workshop on Visualizing Software for Understanding and Analysis, IEEE. URL: <https://doi.org/10.1109/vissof.2005.1684299>, doi:10.1109/vissof.2005.1684299.
- [10] Ivanov, V., Pischulin, V., Rogers, A., Succi, G., Yi, J., Zorin, V., 2018a. Design and validation of precooked developer dashboards, in: Proceedings of the 2018 26th ACM Joint Meeting on European Software Engineering Conference and Symposium on the Foundations of Software Engineering, ACM, New York, NY, USA. pp. 821–826. URL: <http://doi.acm.org/10.1145/3236024.3275530>, doi:10.1145/3236024.3275530.
- [11] Ivanov, V., Rogers, A., Succi, G., Yi, J., Zorin, V., 2018b. Pre-cooked developer dashboards: What to show and how to use, in: Proceedings of the 40th International Conference on Software Engineering: Companion Proceedings, ACM, New York, NY, USA. pp. 402–403. URL: <http://doi.acm.org/10.1145/3183440.3195028>, doi:10.1145/3183440.3195028.
- [12] Janes, A., Sillitti, A., Succi, G., 2013. Effective dashboard design. Cutter IT Journal 26, 17–24.
- [13] Janes, A., Succi, G., 2014. Lean Software Development in Action. Springer Berlin Heidelberg. URL: <https://doi.org/10.1007/978-3-642-00503-9>, doi:10.1007/978-3-642-00503-9.
- [14] Kivi, J., Haydon, D., Hayes, J., Schneider, R., Succi, G., 2000. Extreme programming: a university team design experience, in: 2000 Canadian Conference on Electrical and Computer Engineering. Conference Proceedings. Navigating to a New Era (Cat. No.00TH8492), pp. 816–820 vol.2. doi:10.1109/CCECE.2000.849579.
- [15] Kovács, G.L., Drozdik, S., Zuliani, P., Succi, G., 2004. Open Source Software for the Public Administration, in: Proceedings of the 6th International Workshop on Computer Science and Information Technologies.
- [16] López, L., Martínez-Fernández, S., Gómez, C., Choraś, M., Kozik, R., Guzmán, L., Vollmer, A.M., Franch, X., Jedlitschka, A., 2018. Q-rapids tool prototype: Supporting decision-makers in managing quality in rapid software development, in: Lecture Notes in Business Information Processing. Springer International Publishing. pp. 200–208. URL: [https://doi.org/10.1007/978-3-319-92901-9\\_17](https://doi.org/10.1007/978-3-319-92901-9_17), doi:10.1007/978-3-319-92901-9\_17.
- [17] Pedrycz, W., Russo, B., Succi, G., 2012. Knowledge transfer in system modeling and its realization through an optimal allocation of information granularity. Appl. Soft Comput. 12, 1985–1995. URL: <http://dx.doi.org/10.1016/j.asoc.2012.02.004>, doi:10.1016/j.asoc.2012.02.004.
- [18] Petrinja, E., Sillitti, A., Succi, G., 2010. Comparing OpenBRR, QSOS, and OMM assessment models, in: Open Source Software: New Horizons - Proceedings of the 6th International IFIP WG 2.13 Conference on Open Source Systems, OSS 2010, Springer, Heidelberg, Notre Dame, IN, USA. pp. 224–238. URL: [http://dx.doi.org/10.1007/978-3-642-13244-5\\_18](http://dx.doi.org/10.1007/978-3-642-13244-5_18), doi:10.1007/978-3-642-13244-5\_18.
- [19] Rossi, B., Russo, B., Succi, G., 2012. Adoption of free/libre open source software in public organizations: factors of impact. Information Technology & People 25, 156–187. URL: <http://dx.doi.org/10.1108/09593841211232677>, doi:10.1108/09593841211232677.
- [20] Sarikaya, A., Correll, M., Bartram, L., Tory, M., Fisher, D., 2018. What do we talk about when we talk about dashboards? IEEE Transactions on Visualization and Computer Graphics, 1–1 URL: <https://doi.org/10.1109/tvcg.2018.2864903>, doi:10.1109/tvcg.2018.2864903.
- [21] Sillitti, A., Janes, A., Succi, G., Vernazza, T., 2004. Measures for mobile users: an architecture. Journal of Systems Architecture 50, 393–405. URL: <http://dx.doi.org/10.1016/j.sysarc.2003.09.005>, doi:10.1016/j.sysarc.2003.09.005.

- doi:10.1016/j.sysarc.2003.09.005.
- [22] Staron, M., Meding, W., Hansson, J., Höglund, C., Niesel, K., Bergmann, V., 2014. Dashboards for continuous monitoring of quality for software product under development, in: *Relating System Quality and Software Architecture*. Elsevier, pp. 209–229. URL: <https://doi.org/10.1016/b978-0-12-417009-4.00008-9>, doi:10.1016/b978-0-12-417009-4.00008-9.
  - [23] Succi, G., Paulson, J., Eberlein, A., 2001. Preliminary results from an empirical study on the growth of open source and commercial software products, in: *EDSER-3 Workshop*, pp. 14–15.
  - [24] Valerio, A., Succi, G., Fenaroli, M., 1997. Domain analysis and framework-based software development. *SIGAPP Appl. Comput. Rev.* 5, 4–15. URL: <http://doi.acm.org/10.1145/297075.297081>, doi:10.1145/297075.297081.
  - [25] Vernazza, T., Granatella, G., Succi, G., Benedicenti, L., Mintchev, M., 2000. Defining Metrics for Software Components, in: *Proceedings of the World Multiconference on Systemics, Cybernetics and Informatics*, pp. 16–23.
  - [26] Yigitbasioglu, O.M., Velcu, O., 2012. A review of dashboards in performance management: Implications for design and research. *International Journal of Accounting Information Systems* 13, 41–59. URL: <https://doi.org/10.1016/j.accinf.2011.08.002>, doi:10.1016/j.accinf.2011.08.002.

# **Journal of Visual Language and Computing**

**Volume 2019, Number 2**