A Systematic Literature Review

of Combinatorial Testing

Ariffin Abdullah, Rohayanti Hassan and Zuraini Ali Shah

Universiti Teknologi Malaysia, 81310 UTM Skudai, MALAYSIA
e-mail: ariffin2@live.utm.my
rohayanti@utm.my
aszuraini@utm.my

Abstract

As a context of this research, combinatorial testing (CT) in simplifications of the parameter interactions strategy is to grade or rank the best algorithm in order to eliminated the parameter interaction on various crucial issue. These will become the most interesting research finding so as to enhance the efficiency of the test cases generations. The objective of this research is to identify and analyze existing combinatorial testing technique in the context of the formulated research questions. Search terms with relevant keywords were used to identify primary research that was related to combinatorial testing technique classified under journal articles, conference papers, workshops, symposiums, book chapters and IEEE bulletins. In the end of this research, we provided a primarily result of comparison on efficiency of the simplifications strategy in order to enhance the combinatorial testing technique.

Keywords: *Parameter interaction, Constraints, Null conditions, Condensed, Simplifications, Combinatorial testing.*

1   Introduction

Simplifications in the context of test cases (TC) generation are used to illustrate the challenges of software evaluation in combinatorial testing (CT) evolutionary (*t*-way) for dynamic web application that has thousand parameters input and interactions. The interaction consideration comes out with many solutions that are still complicated and not stratification in order to efficiently consume the cost and time. Combinatorial testing technique (CT) is one of the design principles that enable software under consideration for development to user interfaces (UI) function as expected. The combination of parameter value is based on the fact that

the failure in UI will be identified by a few interactions in parameter input of the applications [1]. We are facing a problem of a large scale application, for dynamic web-based to be acceptable by users or stakeholders, hence its entity and attributes must be well selected, executed and evaluated [2–4]. The nature and description of such constraints are highly domain specific, so for the practice in order to be essential in the account of the application, the parameter interaction has the highest demand. The approach in CT that fails to take account of constraints will produce many test cases that are either unachievable in practice or which yield expensively misleading results [4].

Interaction failures according to [1] have been divided into two factor [6, 7]. The first factor is, the interactions between variables after execution of the code will reveal the failure if the failure existing. The other factor will reveal the failure of the application if the arithmetic computation leads to the false result. The technique they use is required to identify the interactions faults. Later, the interaction is connected with their constraints and classification, to train the data to express the construction of it. There will be a lot possible value of combination in difference type like the example of integer or real value. Focusing in discovering the attribute is necessary to perform the combination. The minimal number of attributes can be example for simplification strategy in order to achieved researcher goal [2].

In this research, we were going to reach the goal to illustrate in terms of simplification in order to increase the efficiency of CT algorithm. We formulized two research questions (RQ1 and RQ2) and found the best solution in current studies to prove our goal. There are:

- RQ1: What are the existing techniques used for simplified parameter interaction in CT?
- RQ2: What are the descriptions and limitations of existing simplifications techniques?

We hope that the solution will have guided us in future to gain better algorithm for CT in terms of efficiency and effectiveness of strategy. The remainder of the article is structured as follows. Section 2 describes the research method used in this review. Section 3 discusses the research questions has been defined. Section 4 presents the results and discussions and Section 5 enumerates the research findings. Section 6 concludes the study.

2   Research Method

Our research method is based on technique proposed by [8] to execute this simplification research.

The review protocols consist of a few phases enumerated as follows: research questions, search strategy design and data extraction results. In the first phase, two sets of research questions had been chosen based on the aim of this study. In the second phase, search strategies were designed using excel table with the formulated research questions which consisted of the identification of search terms and the choices of literature resources. The third phase dealt with the collation of extracted data from paper journal and conference paper in many search resources. We find almost 15 papers that were related to our RQ answer and become proved evident for our goal.

2.1   Headings and format

The aim of this SLR is to understand and summarize the empirical proofs as regard to the state-of-the-art simplification techniques and identify areas for further research in order to complement the performances of existing techniques. To achieve this aim, two research questions (RQs) were formulated as presented below:

- RQ1: What are the existing techniques used for simplified parameter interaction in CT?
- RQ2: What are the descriptions and limitations of existing simplifications techniques?

These questions, which formed the basis for undertaking this research were intertwined and were simultaneously investigated.

2.2   Search Strategy

The detailed description of the search strategies utilized in this research consisted of search terms, literature resources and search process as explained. We used only excel table to summarize the current technique and approaches. We classed them into 10 columns and there were:

- Search based
- Year
- Technique or Approaches
- Input parameter
- Interaction strategy (in terms of CT)
- Combination logic

- Similarity
- Simplification evolved
- Researcher

We also used other method such as mind maker or other data analysis application depended on how easy the planning to collect the data was. No matter what the importing things we can see through these papers to lookup our answers.

2.3   Data Extractions Result

At this point, we struggled on detail through paper journal and paper conference on step-by-step. That was from abstract to the last chapter of the paper. There was an impact on challenges of approaches we found from it but also how we try to merge their knowledge and understandings to answer our RQ.

In our research, we found 2,398 prospective studies that were realized. Next, the titles of these studies were used to scrutinize and collate relevant studies. This task was necessary to eliminate duplicate and irrelevant studies. Consequently, 94 relevant studies were selected. Thereafter, the references of each selected study were perused to identify the important studies that might have been missed out during the initial search process. For this research we only selected a few paper from 94 relevant studies to complete the effort for these research. At the end of the exercise, 10 studies were selected and deemed capable of providing answers to the formulated research questions.
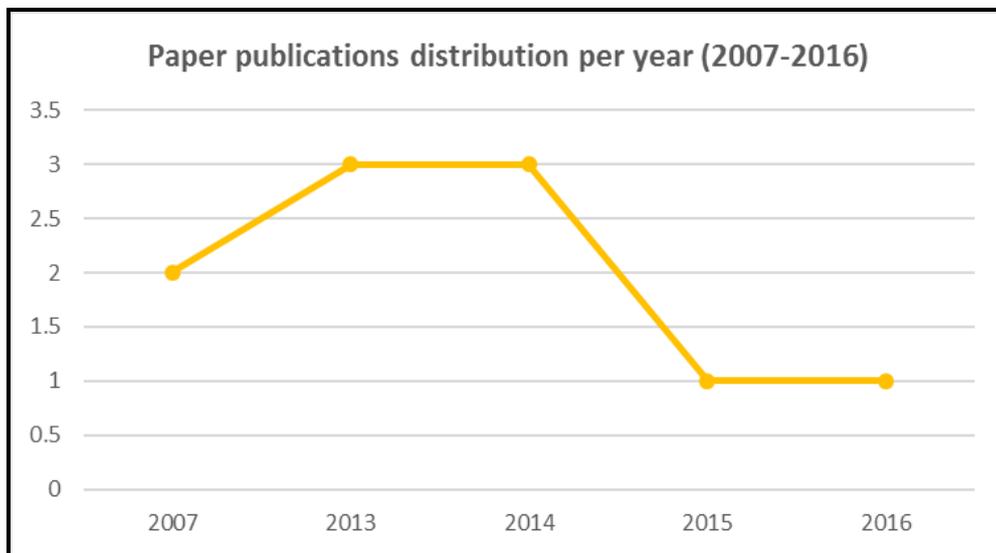
Fig. 1 Number of papers by year of publication
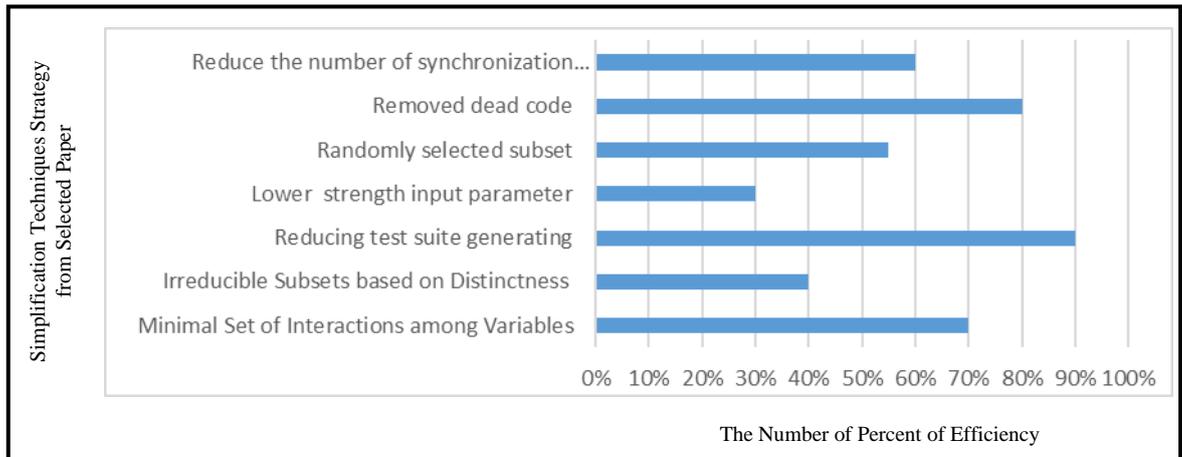
Fig. 2 Existing Simplification Strategy

3   Threats to Validity

The publications biasness and inaccurate extraction of data were considered to be the major threats militating against this review protocol. The studies were chosen based on the search strategy described previously which include (a) various literature databases, (b) selection criteria, and (c) quality criteria. The index terms corresponding or relating to the specified research questions were used to detect relevant studies that were utilized in this review. However, there existed the possibility of missing important studies because, not all studies can be extracted using the terms that were related to the research questions in their titles, abstracts or keywords. To curb this threat, a manual scrutiny of the references of all the extracted studies was patiently carried out to identify those studies that were missed out during the initial search.

Additionally, a precise definition of the selection criteria that complied with the research questions was enforced to avoid incorrect exclusion of desired studies. The studies were selected via a meticulous application of the quality assessment criteria and where discrepancies existed; inconsistencies were immediately resolved. This way, a wide range of further studies were detected. However, the second category of envisaged threat is known as publication biasness. This is a situation where positive results on simplifications techniques are more likely to be reported than negative results, or scholars claiming that, their technique outpaced others. Consequently, this can lead to an overestimation of the performances of existing techniques. To avert this threat, publications that dealt with comparisons of various existing techniques were searched for and included in the selected studies in order to obtain an objective evaluation result across the various techniques. This is because, in most cases; these comparative studies present unbiased reports. Finally, to minimize the threat associated with inaccurate

extraction of data, all the selected studies were re-evaluated to identify the true positives, a situation where the title of a study could connote relevance but the contents do not contain answers to any of the research questions.

4   Result and Analysis

This section presents and analyzes the findings of this review. We start by presenting an overview of the selected studies. Secondly, we present a detailed description of the findings of this review in line with the research questions in separate sub-sections. The review results are also interpreted in this section.

4.1   Overview of Selected Studies

10 studies were selected for this research. Among them, eight papers were published in journals and two papers appeared in conference proceedings. The respective number of papers by year of publication is depicted in Figure 1. However, the detailed descriptions of the selected studies are shown in the Table A1 of the last section.

4.2   Simplifications Parameter Interaction Techniques (RQ1)

From the respective number of paper, we have found 7 several simplifications strategy from existing technique. Figure. 2 shows the plot of the various techniques and their performances to enhance their approached. There are some efforts that have been invested in simplification research. This is evidenced with the number of techniques available in the literature. These techniques are used to determine the strategy with greater value to business successes. The highest percentages until the low percentages effort techniques are detailed as follows: With the highest percentages of 90% excellence efforts, Dale *et. al*, has done a great job by proposing new algorithm named as Interaction-based Test Suite Minimization (ITSM) to enhance the CT design for an effective test planning. ITSM is used to cover the target from different form or explicit set of value combinations to be covered. The set of tests that the CTD algorithm may choose from is limited to the input tests. Exact implementation details depend on the CTD algorithm used, but common to all is the need to represent the set of possible tests. In the execution, there found more improved ways to reduce the constant factor in the algorithm. One of the ways is to avoid unnecessary calculations by the number of times that uncovered target counted. As large data sets, the improved algorithm is about 32 times faster than the straight forward solution [3].

Second highest effort by implementing the Tansuo technique [9], Wenhua *et. al*, evaluated the effectiveness of Tansuo's exploration process guided by CT coverage, for *t = 1, 2, 3*, with respect to code coverage, and find that the

navigation structure exploration by Tansuo, in general, resulted in high code coverage, it eliminated the dead of code to increase the statement coverage on subject applications and it also did the comparison between two other strategies and Tansuo proved are the most effective. Sangeeta and Manuj had claimed by proposing their approach namely as Directed Statement Flow Graph (DSFG) [1]. Where each node represents a statement and each edge represents the directed flow of control of statements within the program. Data flow analysis techniques focus on the flow of data through the program. The flow of data is analyzed with the help of the definition and usage of data. The two usages defined are *c-use* (computation use) and *p-use* (predicate use). A variable is considered *c-used* when it occurs in an output statement, as a parameter within a function, in a subscript expression or as a part of an assignment statement [10]. A variable is considered *p-used* when it is used as a condition in a branch statement [10]. The results indicate that the proposed approach is able to achieve a considerable reduction in the number of interactions to be tested.

The rest of the technique also proposed their techniques in order to enhance the CT by consideration of the best input and value of parameter interactions for improving the efficiency of cost consuming and time effort.

4.3   Descriptions and Limitations of Existing Simplification Strategy

The descriptions and limitations of existing simplifications strategy are enumerated in Table A2 of the Appendix section. These limitations serve as the basis for any improvement hoped for. The descriptions of these techniques are necessary to proffer a glimpse of how each simplifications technique function. While frantic efforts have been invested in developing and improving simplification techniques, some limitations still exist which require urgent research attention. These limitations were gained from the selected studies.

4.4   Discussion from Finding Literature

Table A2 shows the description and limitation of existing simplifications strategy based on their own method to reach the goals. W. Wang *et. al*., built the navigation graph using Tansuo's techniques by selecting appropriate input values for parameters and systematically combined them to reach new pages. Meanwhile, Sangeeta and Manuj identified the interaction based on value strength that coverage as minimal set of interaction. For large scale application, this approaches still need to be expended due to strength of minimal set of interaction is difficult to define. V. Angelova *et. al.*, proposed their method as they located the minimal subset of attributes that relevant to all parameter in application. It is hard to do the pairing between parameter with large scale in attributes as it will take more time

and work in order to simplify the techniques. J. Petke et. al., searched the hard and soft constraints in applications to perform test cases, then chosen the soft constraint in order to reduce process but they had an issue that went to solve the problem of the strength of large parameter and a lot of fault found in action. D. Blue *et. al.*, tried to reduce the test suite and selected the minimal set of interaction to generate the result but it still need a hard work to support the large parameter in big application that hard to find the minimal set of interaction. Based on above discussions, we have explained *how* and *why* each method is different and reflect on their result in Figure 2.

The benefit of above approach is that it uses advanced algorithms derived by contradiction and optimized for speed in numerous competitions. For specific examples, they try to eliminated the irrelevant test suite or test cases of the task to polynomial (by looking for a specific value parameter with only one true literal and by excluding the reason for conflict). In our web-based application, each field in user interface is a variable and we create one test cases per field of the table. Each field is a set of test cases of the variables for which the corresponding column has many value. Finally, we add a cardinality null conditions that restrict the simplifications strategy solution. As we find solutions, we locate irrelevant test cases to disable the already found test cases and their extensions.

5   Conclusion

The aim of this research was to identify and examine the status quo of combinatorial testing simplifications techniques. The method utilized in this research was a systematic literature review. With this method, some pertinent research questions were formulated based on the aim of the study to identify and ascertain existing simplifications techniques, their limitations and processes. This was carried out through identifying, assessing and interpreting relevant studies. The essence of this research was to identify areas for possible improvement or enhancement via systematic evaluation of relevant and current studies in simplification strategy techniques as reported in the literature. Data were extracted from these primary studies and the extracted data were synchronized. The research objectives were considered to have been met and the formulated research questions were considered to have been addressed. Some studies that dealt with the method of enhancing existing simplification techniques were identified. It was discovered that, although a lot of simplification techniques exist; improvements are still required. Some of these improvements border on scalability, computational complexities, ease of use, reliability of results, validation of techniques in industrial settings and requirements resolvability and dependency issues. Therefore, further studies on simplification techniques can dwell on the aforementioned limitations. Some threats to validity were eminent in this study; most of which have been identified and curbed at the early stage of this research.

Appendix

Table A1: Overview of selected studies with respective reference numbers

Publication type	Publication year	Publication name	Refs.
Journals	2016	Software Testing, Verification and Reliability	[9]
Journals	2014	SIGSOFT Software Engineering Notes	[1]
Journals	2014	Proceedings of the 15th International Conference on Computer Systems and Technologies	[2]
Journals	2014	Computer	[14]
Journals	2013	Proceedings of the 2013 9th Joint Meeting on Foundations of Software Engineering	[4]
Journals	2013	Proceedings of the 2013 International Conference on Software Engineering	[3]
Journals	2007	SIGSOFT Software Engineering Notes	[5]
Journals	2007	Software Testing, Verification and Reliability	[11]
Conference	2015	Software Testing, Verification and Validation Workshops (ICSTW), 2015 IEEE Eighth International Conference on	[12]
Conference	2013	Instrumentation, Measurement, Computer, Communication and Control (IMCCC), 2013 Third International Conference on	[13]

Table A2. Descriptions and Limitations of Existing Simplifications Strategy

No.	Name of Techniques	Description	Limitations/Sources
1.	Combinatorial *Tansuo's* exploration process	Techniques are required to select appropriate input values for parameters and systematically combine them to reach new pages.	The number of submission tests are required to achieve a certain level of combinatorial coverage.
2.	Data flow analysis techniques	The technique is required that can identify the optimal value of the strength of coverage or the minimal set of interactions to be tested so as to minimise the testing costs.	The strength and large scale interaction.
3.	Classification Algorithms with SAT Solver Tool Support	The element by attributes is large and some attributes are redundant. It is interesting to find a minimal subset of attributes, which discriminates between all elements.	Coverage on paring parameter interaction and to simplified the solutions.
4.	Separation considerations of value parameter	Real systems are typically constrained. Hard constraints must be accounted for to avoid the generation of inapplicable or misleading test cases, but soft constraints also can contribute to reducing the test space. Real testers are also required.	The relationship between interaction strength and faults found
5.	Interaction-based Test-suite Minimization	Reduces an existing test suite and preserving its interaction coverage	For large scale system requires generating all the corresponding data and is highly laborious task

References

[1] S. Sabharwal, M. Aggarwal, Identifying Interactions for Combinatorial Testing using Data Flow Techniques, ACM SIGSOFT Soft. Eng. Notes, Nov. 2014 Vol. 39 No. 6.

[2] V. Angelova, Early Concept of Classification Test and Contemporary Approach to Diagnostic Test and Combinatorial Test, Inter. Conf. on Com. Sys. and Tech. - CompSysTech'14, June 2014.

[3] D. Blue *et. al*, Interaction-Based Test-Suite Minimization, ICSE 2013, San Francisco, CA, USA, 2013.

[4] J. Petke *et. al*, Efficiency and Early Fault Detection with Lower and Higher Strength Combinatorial Interaction Testing, ESEC/FSE'13, Aug. 18–26, 2013.

[5] R. Krishnan *et. al*, Combinatorial Testing: Learnings from our Experience, ACM SIGSOFT Soft. Eng. Notes, Vol. 32 Num. 3, May 2007.

[6] Mathur, A.P. 2008. Foundations of Software Testing. Pearson Education.

[7] Kuhn, D.R., Kacker, R.N., and Lei, Y. 2010. Practical Combinatorial Testing. NIST Special Publication. 800-142.

[8] B.A. Kitchenham, S. Charters, Guidelines for Performing Systematic Literature Reviews in Software Engineering, Tech. Rep. EBSE-2007-01, Keele University and University of Durham, 2007.

[9] Wang, W. Sampath, Sreedevi, Lei, Y. Kacker, *et. al*, Using Combinatorial Testing to Build Navigation Graphs for Dynamic Web Applications, Soft. Test., Ver. and Rel., 2016.

[10] Kuhn, D.R., Kacker, R.N., and Lei, Y. Practical Combinatorial Testing. NIST Special Publication. 800-142, 2010.

[11] Lei, Yu, Carver, *et. al*, A combinatorial testing strategy for concurrent programs Soft. Testing, Veri. and Reli. 2007.

[12] I. Segall, R. Tzoref-Brill, A. Zlotnick, Combining minimization and generation for combinatorial testing Software Testing, Veri. and Valid. Works. (ICSTW), 2015 IEEE Eighth Inter. Conf. on 2015.

[13] W. Jianfeng, W. Changan, J. Shouda, Test Data Generation Algorithm of Combinatorial Testing Based on Differential Evolution, Instr. Mea. Comp. Com. and Cont. (IMCCC), 2013 Third Inter. Conf. on 2013.

[14] C. Yilmaz, S. Fouch, et. al, Moving Forward with Combinatorial Interaction
      Testing, Computer 2014.