



Potential and limitations of the ISBSG dataset in enhancing software engineering research: A mapping review



Marta Fernández-Diego*, Fernando González-Ladrón-de-Guevara

Business and Economics Department, Universitat Politècnica de València, Camino de Vera, s/n, 46022 Valencia, Spain

ARTICLE INFO

Article history:

Received 31 July 2013

Received in revised form 7 January 2014

Accepted 8 January 2014

Available online 17 January 2014

Keywords:

Systematic mapping study

Research methods

Software engineering

ISBSG

Software effort estimation

Software cost prediction

ABSTRACT

Context: The International Software Benchmarking Standards Group (ISBSG) maintains a software development repository with over 6000 software projects. This dataset makes it possible to estimate a project's size, effort, duration, and cost.

Objective: The aim of this study was to determine how and to what extent, ISBSG has been used by researchers from 2000, when the first papers were published, until June of 2012.

Method: A systematic mapping review was used as the research method, which was applied to over 129 papers obtained after the filtering process.

Results: The papers were published in 19 journals and 40 conferences. Thirty-five percent of the papers published between years 2000 and 2011 have received at least one citation in journals and only five papers have received six or more citations. Effort variable is the focus of 70.5% of the papers, 22.5% center their research in a variable different from effort and 7% do not consider any target variable. Additionally, in as many as 70.5% of papers, effort estimation is the research topic, followed by dataset properties (36.4%). The more frequent methods are Regression (61.2%), Machine Learning (35.7%), and Estimation by Analogy (22.5%). ISBSG is used as the only support in 55% of the papers while the remaining papers use complementary datasets. The ISBSG release 10 is used most frequently with 32 references. Finally, some benefits and drawbacks of the usage of ISBSG have been highlighted.

Conclusion: This work presents a snapshot of the existing usage of ISBSG in software development research. ISBSG offers a wealth of information regarding practices from a wide range of organizations, applications, and development types, which constitutes its main potential. However, a data preparation process is required before any analysis. Lastly, the potential of ISBSG to develop new research is also outlined.

© 2014 Elsevier B.V. All rights reserved.

Contents

1. Introduction	528
2. Research methodology	528
2.1. Research questions	528
2.2. Search for primary studies strategy	529
2.3. Inclusion and exclusion criteria	529
2.4. Data collection	530
3. Results	530
3.1. RQ1	530
3.2. RQ2	531
3.3. RQ3	532
3.4. RQ4	534
3.5. RQ5	535

* Corresponding author. Tel.: +34 96 387 76 85; fax: +34 96 387 97 79.

E-mail addresses: marferdi@omp.upv.es (M. Fernández-Diego), fgonzal@omp.upv.es (F. González-Ladrón-de-Guevara).

3.5.1.	Reasons for the usage of complementary datasets.	536
3.5.2.	Advantages and limitations related to the usage of ISBSG.	536
4.	Discussion.	538
4.1.	Principal findings.	538
4.2.	Study limitations.	538
4.3.	Implications for research and practice.	539
5.	Conclusion and future work	540
	Appendix A	540
	References	544

1. Introduction

The International Software Benchmarking Standards Group [1] designed and currently maintains two international public repositories: Software Development & Enhancement with over 6000 software projects and Maintenance & Support with over 470 software applications. The repository contains more than 150 data fields collected from a wide range of countries, organizations, application types, and development types.

The goal of ISBSG is to help any type of organization (business, public, or non-profit) improve their IT resource management through the use of these datasets by performing their own analyses, estimations, comparisons, or benchmarking.

The ISBSG dataset for software development is organized by releases. The current version (since 2013) is release 12, which includes over 6000 projects distributed between years 1989 and 2013. Such a dataset makes it possible to estimate a project's size, effort, duration, and cost. Furthermore, ISBSG enables users to check project specification completeness, reduce project risk, control software development, plan infrastructure development, and benchmark performance. But ISBSG suffers from two major problems: the heterogeneity of data, i.e., the combining of data from heterogeneous sources [2], and most of the variables have a lot of missing values.

ISBSG implements two fields in its datasets, which can be used to filter out low quality cases from the analysis and help handle data validation and rating issues [S105]. Each project submitted to the ISBSG repository is validated against specific quality criteria and rated in four categories. As pointed out in Liebchen and Shepherd [3], the classification is primarily guided by the completeness of the software projects, which means that low quality data are interpreted as possessing high levels of missing values.

A mapping study can be considered as a secondary study that reviews articles related to a specific research topic. Such a study provides an overview of a research area to assess the existing evidence [4] and can identify gaps in the set of primary studies, where new or better primary studies are required. Mapping studies also pinpoint specific realms of knowledge where there may be an opportunity for more complete systematic literature reviews to be undertaken [5,6].

Ultimately, a mapping study aims at providing a classification, conducting a thematic analysis, or presenting publication channels [7]. Petticrew and Roberts [5] also suggest that this type of study "involves a search of the literature to determine what sorts of studies addressing the systematic review question have been carried out, where they are published, in what databases they have been indexed, what sorts of outcomes they have assessed, and in which populations." These studies require a rigorous searching process as well as detailed inclusion and exclusion criteria that are clearly defined in the research protocol and presented in the results report [8]. The main difference between a mapping study and a systematic literature review is the formulation of the research questions and the analysis of the available information [6].

The purpose of this study was to determine to what extent and how ISBSG has been the support for a group of researchers in

software engineering. Thereby, systematic mapping has been performed to "map out" papers that have used this dataset in an attempt to identify the topics, estimation methods, complementary datasets, and other issues that have been dealt with in research questions. In this way, a picture portraying the potential and limitations of ISBSG as a research facility has been obtained. Additionally, the investigation into the set of papers related to ISBSG and their classification are valuable results [9].

The rest of this paper will be organized as follows: Section 2 will describe the mapping process, Section 3 will report the mapping results, Section 4 will discuss the limitations of the study, and finally, Section 5 will outline the main conclusions obtained and future work lines.

2. Research methodology

Systematic mapping studies are a type of systematic literature review that aims to collect and classify research papers related to a specific topic [5–7,10,11].

This section provides an overview of the steps involved in the process of mapping review following Petersen et al. [7] including the formulation of the research questions, the search strategy for primary studies, the inclusion and exclusion criteria, and the data collection process.

2.1. Research questions

The primary goal addressed by this study is to analyze the use of the ISBSG dataset by the research community and to consider its potential to undertake the appropriate scientific research. A number of issues and their rationales were considered:

- Research question number 1 (RQ1): Which and how many journals and conferences include ISBSG-based research papers? The main purpose of this question is to provide a list of journals and conferences interested in such studies and to assess the relevance of papers based on ISBSG. To acknowledge the prestige associated with each journal and conference, their rankings are discussed. Furthermore, the temporal evolution of paper publications based on ISBSG as primary data was also taken into account for the most relevant journals and conferences.
- RQ2: To what extent are researchers working with the ISBSG dataset aware of other works also using the ISBSG dataset? By examining the citations between papers, it is possible to assess the awareness of researchers to the ISBSG body of knowledge and their ability to contribute to this body of knowledge. By scrutinizing the citations, it is possible to analyze the evolution of received citations over time.
- RQ3: What are the most investigated research topics and their trends over time? The motivation behind this question is to provide an update on the contents of ISBSG-related research. Furthermore, by analyzing the evolution of the topics, both emerging issues and obsolete issues can be identified.

Download English Version:

<https://daneshyari.com/en/article/550590>

Download Persian Version:

<https://daneshyari.com/article/550590>

[Daneshyari.com](https://daneshyari.com)