The Journal of Systems and Software 000 (2016) 1–20

Contents lists available at ScienceDirect

The Journal of Systems and Software

journal homepage: www.elsevier.com/locate/jss

Exploring quality measures for the evaluation of feature models: a case study

Carla I.M. Bezerra[a,b,c,*], Rossana M.C. Andrade[a,c,1], Jose Maria Monteiro[c]

[a] Group of Computer Networks, Software Engineering and Systems (GREat), Campus do Pici - Bloco 942-A - ZIP: 60455-760 - Fortaleza - CE, Brazil
[b] Campus Quixadá, Federal University of Ceará (UFC), Quixadá, Brazil
[c] Graduate Program in Computer Science, Federal University of Ceará (UFC), Fortaleza, Brazil

ARTICLE INFO

Article history:
Received 15 June 2015
Revised 7 July 2016
Accepted 25 July 2016
Available online xxx

Keywords:
Software product line
Quality evaluation
Measures
Feature models

ABSTRACT

Evaluating the quality of a feature model is essential to ensure that errors in the early stages do not spread throughout the Software Product Line (SPL). One way to evaluate the feature model is to use measures that could be associated with the feature model quality characteristics and their quality attributes. In this paper, we aim at investigating how measures can be applied to the quality assessment of SPL feature models. We performed an exploratory case study using the COfFEE maintainability measures catalog and the S.P.L.O.T. feature models repository. In order to support this case study, we built a dataset (denoted by MAcchiATO) containing the values of 32 measures from COfFEE for 218 software feature models, extracted from S.P.L.O.T. This research approach allowed us to explore three different data analysis techniques. First, we applied the Spearman's rank correlation coefficient in order to identify relationships between the measures. This analysis showed that not all 32 measures in COfFEE are necessary to reveal the quality of a feature model and just 15 measures could be used. Next, the 32 measures in COfFEE were grouped by applying the Principal Component Analysis and a set of 9 new grouped measures were defined. Finally, we used the Tolerance Interval technique to define statistical thresholds for these 9 new grouped measures. So, our findings suggest that measures can be effectively used to support the quality evaluation of SPL feature models.

© 2016 Elsevier Inc. All rights reserved.

1. Introduction

Among the techniques for software reuse, one that has gained relevance is Software Product Line (SPL). Clements and Northrop (2002) defined SPL as a collection of software intensive systems using and sharing a group of common characteristics managed to meet the needs of a particular segment of the market or mission and developed from a common set of core assets and a predetermined shape.

Quality evaluation is essential in the SPL context justified by the fact that an error or inconsistency in an SPL artifact can be propagated to all its products. It is important to notice that quality assessment in SPL presents more complexity than in traditional software development due to two aspects: i) different products can be derived from the same SPL; and ii) different products in the same SPL may require different levels of quality Etxeberria and Sagardui (2008b). According to Etxeberria and Sagardui (2008b), the quality evaluation of all artifacts and software products of a given SPL proves to be impractical, both for economic reasons and the effort needed.

One of the most important assets of an SPL is the feature model. As believed by Kang et al. (1990), this artifact captures the common features and differences among end products resulting from the same SPL. In particular, it models all possible products of an SPL in a given context Benavides et al. (2010). The structure of the feature model is composed by features. As stated by Böckle et al. (2005), features describe the functional as well as the quality characteristics of the system under consideration. The feature model is a relevant intermediate product. This means that evaluating the quality of a feature model is critical to ensure that errors in the early stages do not spread throughout the SPL.

It is difficult to control what you cannot measure, therefore software measures have become a concern in the field of software engineering. They play an important role in understanding,

* Corresponding author.
E-mail address: carlabezerra@great.ufc.br, carlailane@gmail.com (C.I.M. Bezerra).
[1] Researcher Scholarship - DT Level 2, sponsored by CNPq.

http://dx.doi.org/10.1016/j.jss.2016.07.040
0164-1212/© 2016 Elsevier Inc. All rights reserved.

controlling and improving software quality. Thus, to evaluate the quality of an artifact, a popular strategy is the use of measures.

Hence, a lot of measures have been proposed to evaluate the quality of feature models Bagheri and Gasevic (2011); Montagud et al. (2012); Berger and Guo (2014); Bezerra et al. (2015). In our previous work, Bezerra et al. (2015) presented a measures catalog (denoted by COfFEE - CatalOg of measures for Feature modEl quality Evaluation) able to be used to support the quality evaluation of a feature model. In order to identify these measures, firstly a systematic mapping was conducted. Then, to evaluate the use of the proposed catalog, we applied the measures in six feature models in the several domains.

Still, according to Montagud and Abrahão (2009), several work have been proposed in order to ensure the quality in SPLs. The majority of them focus on the evaluation of quality attributes at the architecture level (e.g. Thiel (2002); Matinlassi et al. (2002); Olumofin and Mišić (2007); Benavides et al. (2007); Kim et al. (2008); Junior et al. (2013)) and few methods focus on the evaluation of the relevant domain attributes Etxeberria and Sagardui (2008a,b).

The goal of this work is to investigate the way measures can be applied to the quality assessment of SPL feature models. For this, we performed an exploratory case study using the COfFEE catalog (proposed in Bezerra et al. (2015)) and the S.P.L.O.T. feature models repository Mendonça et al. (2009). This case study was influenced by Jedlitschka and Pfahl (2005); Kitchenham et al. (2008); Robson and McCartan (2016) and based on the guidelines defined in Runeson and Höst (2009). This new paper extends our previous work Bezerra et al. (2015) by:

- **Building a dataset:** we built a dataset (denoted by MAcchiATO - MeAsures dATaset for feaTure mOdel) containing the values of 32 measures from COfFEE for 218 software feature models, extracted from the S.P.L.O.T. repository;
- **Analyzing the correlation:** we analyzed the correlation among the 32 measures in COfFEE applying the Spearman's rank correlation coefficient in order to indicate the statistical correlation;
- **Grouping measures:** we grouped the measures in COfFEE and defined 9 new grouped measures applying the Principal Components Analysis technique (PCA); and
- **Defining thresholds:** we defined statistical thresholds for these 9 new grouped measures using two different strategies: "three-sigma rule" and tolerance interval.

It's important to emphasize that using the MAcchiATO dataset, three data analysis techniques were applied in order to better understand the existing relationships between the 32 measures in COfFEE. Firstly, we have used the Spearman's rank correlation coefficient to describe the statistical correlation between each pair of measures. Our analysis showed that correlations among the measures in COfFEE exist. Thus, not all of the 32 measures are necessary to reveal the quality of a feature model. Some can be used interchangeably and just 15 measures could be used. Besides, most measures are defined at the level of individual aspects. So, to summarize the measure at a high level there is a need for grouping methods. As a feature model evaluation requires the use of different measures, possibly, with widely varied output ranges, there is a need to combine these measures into a unified quality assessment. To deal with this problem we used Principal Component Analysis (PCA). PCA is a mathematical procedure that transforms a number of (possibly) correlated variables into a (smaller) number of uncorrelated variables called principal components. With PCA, the large set of 32 measures in COfFEE was reduced to a small set of 9 new grouped measures that still contains most of the information in the large set. We also defined statistical thresholds for the 9 new measures, based on two different approaches: "three-sigma rule" and tolerance interval. The thresholds can be used by industry and academia as parameters for quality evaluation of feature models.

In order to validate the three different data analysis techniques used in this exploratory case study (Spearman's rank correlation coefficient, Principal Component Analysis and Tolerance Interval), we performed a cross-validation test, with 10 rounds. The results of this test showed that these techniques will generalize to an independent dataset.

The remainder of this paper is organized as follows. Section 2 addresses the related work. Section 3 presents the key concepts involved in this work. It formalizes the concept of feature model and quality models besides presenting a systematic mapping about the quality evaluation of feature models and the COfFEE catalog. Section 4 discusses the study case design, highlighting the research questions, case and subjects selection, data collection procedures and data analysis procedures. The results and the proposed validation are presented in Section 5. In Section 6 our results and implications for researchers and practitioners are discussed. Section 7 presents the threats to validity, and at last, Section 8 concludes this paper and points out directions for future work.

2. Related work

In the last years, several approaches to ensure quality in SPL have been proposed. However, most of these works has focused on the development and validation of product configurations. Just a few of them have been investigating quality assurance aspects, such as the definition of quality measures or the internal and external evaluation of quality attributes. Some work identified measures for the quality evaluation of feature models. Other work evaluated the statistical correlation between quality measures. However, we did not find studies that have produced grouped measures for quality evaluation of feature models or have defined thresholds for these measures.

Bagheri and Gasevic (2011) proposed a number of structural measures to assess the quality of feature models in SPLs. They validated their measures using measurement theoretic principles. A controlled experimentation was performed in order to analyze whether these structural measures can be good predictors of the three maintainability sub-characteristics: analysability, changeability and understandability. However, the measures presented in their study cover just three maintainability sub-characteristics, while this work includes seven subcharacteristics. Other measures of maintainability mentioned in our work were not covered by Bagheri and Gasevic. In addition, this paper used only 14 feature models for collecting and analyzing measures, and most used features models are not related to software. Our study evaluated a larger set of measures (32 measures), and performed a correlation analysis between these measures using a set of 218 feature models. The fact that we use a lot of data for correlation analysis makes our study most reliable.

Berger and Guo (2014) performed a correlation analysis with code measures and feature model measures in SPLs. Some features models measures were identified in literature and others measures were proposed by the authors. They used a set of 8 real SPLs to perform their analysis, which for a correlation study is considered a small set.

Unlike Berger and Guo (2014), our work used a set of 218 feature models, all of them related to software, in the correlation analysis. Besides, our work proposed a set of grouped measures by applying PCA and used two different strategies to define thresholds for the grouped measures.

Montagud et al. (2012) published a systematic review aiming to identify studies that have quality attributes and/or measures for SPL. These attributes and measures were classified using a set of

Download English Version:

https://daneshyari.com/en/article/4956420

Download Persian Version:

https://daneshyari.com/article/4956420

Daneshyari.com