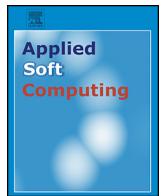




Contents lists available at [ScienceDirect](#)

Applied Soft Computing

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



Extracting features from online software reviews to aid requirements reuse

Noor Hasrina Bakar^a, Zarinah M. Kasirun^{a,*}, Norsaremah Salleh^b, Hamid A. Jalab^a

^a Department of Software Engineering, Faculty of Computer Science & IT, University of Malaya, 50603 Kuala Lumpur, Malaysia

^b Department of Computer Science, Kulliyah of ICT, International Islamic University Malaysia, P.O. Box 10, 50728, Kuala Lumpur, Malaysia

ARTICLE INFO

Article history:

Received 31 December 2015
Received in revised form 24 June 2016
Accepted 25 July 2016
Available online xxx

Keywords:

Requirements reuse
Software engineering
Natural language processing
Unsupervised learning
Latent semantic analysis

ABSTRACT

Sets of common features are essential assets to be reused in fulfilling specific needs in software product line methodology. In Requirements Reuse (RR), the extraction of software features from Software Requirement Specifications (SRS) is viable only to practitioners who have access to these software artefacts. Due to organisational privacy, SRS are always kept confidential and not easily available to the public. As alternatives, researchers opted to use the publicly available software descriptions such as product brochures and online software descriptions to identify potential software features to initiate the RR process. The aim of this paper is to propose a semi-automated approach, known as Feature Extraction for Reuse of Natural Language requirements (FENL), to extract phrases that can represent software features from software reviews in the absence of SRS as a way to initiate the RR process. FENL is composed of four stages, which depend on keyword occurrences from several combinations of nouns, verbs, and/or adjectives. In the experiment conducted, phrases that could reflect software features, which reside within online software reviews were extracted by utilising the techniques from information retrieval (IR) area. As a way to demonstrate the feature groupings phase, a semi-automated approach to group the extracted features were then conducted with the assistance of a modified word overlap algorithm. As for the evaluation, the proposed extraction approach is evaluated through experiments against the truth data set created manually. The performance results obtained from the feature extraction phase indicates that the proposed approach performed comparably with related works in terms of recall, precision, and F-Measure.

© 2016 Elsevier B.V. All rights reserved.

1. Introduction

Requirements for any existing system can be extracted and reused for production of a new similar system [8]. However, reuse of software features extracted from Software Requirement Specifications (SRS) is only viable to practitioners who have access to these software artefacts. SRS usually reside within company databases that are kept confidential and therefore makes it hard for external researchers to access and further explore its reuse potential. In related research, software features were extracted from various other forms of Natural Language Requirements (high-level requirements) when SRS are not easily accessible, for example product brochures were used by Ferrari et al. [11], online product listings were used by Davril et al. [9], and the use of multiple web reposi-

tories were reported by Yu et al. [29]. Compared to related works, this research is focusing on extracting software features that can represent the functionality of software being reviewed. Few works in the requirements engineering (RE) area used the mobile app reviews to either extract the new feature requests or for the purpose of redesigning the existing functionalities for the mobile apps. However, to the best of current knowledge, none of the works has reported the use of online software reviews purposely to initiate the Requirements Reuse (RR) process.

Various information can be obtained from software reviews that are available on the Internet. These can include user opinions or sentiments towards certain products, user complaints, new feature requests, and also statements about existing software functionalities. Software reviews are accessible artefacts that not only contain beneficial information for new users before making software purchasing decision, but these reviews also contain important information for developers, in which it encompasses the evaluations made by customers including ideas for improvements, new feature requests, and existing functionalities of software being reviewed. Information obtained from the software reviews can be

* Corresponding author at: Department of Software Engineering, Fac of Computer Science and Information Technology, University of Malaya, 50603 Kuala Lumpur, Malaysia.

E-mail address: zarinahmk@um.edu.my (Z.M. Kasirun).

Table 1
Research problems and proposed solutions.

#	Problems	The Proposed Solution
1	Manual and ad hoc reuse of requirements can be very arduous, time-consuming, and labour-intensive on the results.	Propose a less complex semi-automated process to <i>extract software features</i> that can reduce manual extraction effort.
2	Existing approaches provide little guidelines for practitioners' adoptions.	Perform experiment to <i>demonstrate the proposed feature extraction process</i> .
3	Very limited works in this area that provide empirical evaluations towards their results.	<i>Provide results comparison</i> and empirical evaluations towards the proposed approach as compared to manual approach.

valuable for developers who have no access to the technical documentation of the software product such as SRS, but have the intention to reuse the information (RR process) for the development of similar products. In manual and ad hoc reuse, domain analysts need to firstly read the whole requirement documents to get an overview on what the product is about. Second, to extract related features, domain analysts need to manually select sentences or phrases that indicate the software feature, in which can be accurate but require more processing time and error-prone when dealing with large data.

Mainly, this work contributes to the existing body of knowledge within the feature extraction from NL requirements, which employ the NL processing and data mining techniques in Software Engineering area. Secondly, the application of Soft Computing techniques from the Support Vector Machine area is applied in this real world Software Engineering problem of requirements reuse. The application of Latent Semantic Analysis, LSA (within the scope of Support Vector Machine area) was demonstrated in this paper as a way to identify similar requirements documents, in which researchers interested in exploring natural language artefacts can benefit from the implementation steps outlined in this paper. In addition to LSA, this paper present the use of clustering algorithm such as K-Means, Fuzzy C-Means and Self Organizing Map towards data obtained from LSA. This paper presents the possibilities of clustering algorithm application in software engineering area. Furthermore the quality of documents clustering results produced are then evaluated by measuring its silhouette values, and comparisons between silhouette analysis and clustering accuracy are thoroughly presented. Thirdly, a part from applying the soft computing techniques, this research provides a dataset from software reviews extracted from the internet, which can be later used by other researchers to further explore on improving the feature extraction approach, or for suggesting a new approach for requirements reuse. As to date this paper is written, to the best knowledge of the authors, there is no dataset that is made freely available for requirements reuse research that are based on software reviews.

1.1. Problem statement

Inspired by the related works, this research is aiming to explore the semi-automated process to extract software features from public data, which can assist the process to detect reusable software features from natural language (NL) requirements. The definition of features is referred to as a prominent or distinctive user visible characteristic of a software product [21]. Previous works in the RR area within the Software Product Line Engineering (SPLE) community such as Ferrari et al. [11] and Hariri et al. [15] have looked at the publicly available data as an input to the RR process, but as to date this paper is written, and to the best of the author's knowledge, none of the related works has reported the use of online software reviews to initiate the RR process. Table 1 lists out the research

problems and the proposed solutions that will be further explored in this paper:

Two issues that motivate this research are algorithm complexity and missing guidelines for future practitioners' adoption. The purpose of this paper is to present a less complex approach for automated feature extraction for RR and to provide insights for future practitioners' adoption.

In this work, a semi-automated process is presented to extract the phrases from software reviews. Collection of these phrases typically functions as "action" and "object". To represent a system function, it is believed that a requirement statement must at least encompasses of an "action" (usually represented by verbs), followed by an "object" (represented by verbs, nouns, and adjectives). The representation of "actions" and "objects" that exist in user reviews is believed to expose the software functionalities. This is important so that it can be reused in the production of similar system in the near future. The NL processing approaches for the extraction of the software features from the user reviews is employed in this work. The main aim is to provide lists of early software features which can provide input to domain analysts in the RR process.

The remainder of this paper is organised as follows: Section II describes the related works; Section III presents the proposed approach in detail; Section IV discusses the experiment results and presents threats to validity, and finally, a conclusion and description about future research are provided at the end of the paper.

2. Related works

The related works are divided into three parts: reviews that relate to RE and software product lines, works that are related to feature extraction from NL requirements in software product lines, and lastly the related works in the area of RE that utilise user reviews.

2.1. Related review in the area of RE and SPLE

A systematic literature review was conducted by Alves et al. [1] related to RE in SPLE. Important findings from their review revealed that there is a need to improve the overall quality of the reviewed studies in terms of empirical validations. Additionally, they have reported that most of the studies do not provide sufficient guidelines for practitioners to adopt the proposed approach and there are very limited commercial or open source tools that are currently accessible, which hinders the practitioners' adoption of the proposed approach. As for the research trend, a growth in the number of approaches to handle NL requirements in a more automated way is anticipated in the future.

2.2. Related works for feature extraction from NL requirements and SPLE

Pertaining to feature extractions from NL requirements process within the SPLE context, few authors use online product listings [15], product brochures [11], or web repositories [29] for deriving reusable product lines features. For instance, data mining approaches were used by Hariri et al. [15] to discover common features across products and relationships among features by taking online product listing as their input. Incremental diffusive clustering algorithm and association mining were then used in their study to group common software features. Similar effort appeared in Yu et al. [29], which used project profiles from Sourceforge.net and Softpedia.com as the input to the proposed feature recommendation process. To achieve that, they have introduced an incremental diffusive clustering algorithm for feature discovery and used association mining to augment initial feature profile followed by the

Download English Version:

<https://daneshyari.com/en/article/4963636>

Download Persian Version:

<https://daneshyari.com/article/4963636>

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