

Summarizing, Classifying and Diversifying User Feedback

Emitza Guzman
Technische Universität München
Department of Applied Software Engineering
Boltzmannstr. 3, 85748 Garching
emitza.guzman@mytum.de

Abstract: Current software users can give feedback about a software application through diverse online mediums such as blogs, forums, instant messaging and product review sites. This produces a large amount of textual information, making it difficult to take user feedback into consideration in the software evolution process. We propose to automatically summarize, classify and diversify textual user feedback in order to reduce information overload and give users a stronger voice in software evolution. In this paper we describe the research questions, the envisioned approach and the achieved progress related to the presented thesis.

1 Introduction

In the early days of the digital age, software users were a few engineers or scientists with specific technical requirements. However, with the evolution of computing power, the appearance of more affordable personal computers, the Internet and smartphones, the definition of user has extended to include more heterogeneous groups of people with a wide variety of needs and expectations [Pag13]. Current software engineering research has pointed out the importance of taking these needs and expectations into account in order to make useful and usable software, and reduce maintenance efforts [Kuj03, KKLK05, VMSC02]. In order to keep software useful and relevant through its evolution it is necessary that user needs and expectations are considered in the post-deployment phase [BD04, KLF⁺11]. With the growing trend of Internet use, more users are writing feedback about software applications, requesting features and reporting bugs through social media, online forums, specialized user feedback platforms or directly in the application distribution platforms (e.g. GooglePlay¹ or the AppStore²) by means of their integrated review system. This tendency produces a large amount of textual data which can be burdensome to manually analyze and process. Because of this, it is difficult for developers to remain aware of feature requests and bug reports that users write, as well as the general opinion that users have about an application and its features. This thesis researches possible ways to overcome this problem. More concretely, it analyzes summarization, classification and diversification techniques commonly used in the natural processing language and information retrieval communities and proposes to apply them to feedback given by users of software

¹<https://play.google.com>

²<http://www.apple.com/iphone-5s/app-store/>

applications. The summarization of user feedback will allow analysts and developers to have a quicker overview of the aggregated user feedback content, whereas classification will separate user feedback into categories allowing developers and analysts to concentrate on the categories that are relevant to the evolution task they are performing. Furthermore, diversification will retrieve user feedback that varies in its content, allowing developers to obtain user feedback with conflicting opinions and with a varied set of topics.

2 Research Questions

This thesis aims to reduce developers' and analysts' information overload when analyzing and processing user feedback by automatically summarizing, classifying and diversifying it. The research questions that guide this work are the following:

Summarization

RQ1: How can user feedback be automatically summarized?

RQ2: Do analysts and developers benefit from summarized user feedback?

Classification

RQ3: What are the characteristics of the different types of user feedback (e.g. feature requests, bug reports, reviews about existing features)? What are the characteristics of useful feedback in the perspective of developers and analysts?

RQ4: How can the different types of feedback be automatically classified? How can useful feedback be automatically classified?

Diversification

RQ5: How can we retrieve diverse feedback in terms of the mentioned software features and the sentiments associated to them?

RQ6: What are the benefits of retrieving diverse feedback in the software evolution process?

3 Envisioned Approach

We will apply the approach to feedback given by users through application distribution platforms, such as GooglePlay and the AppStore, and through product review platforms, such as Epinions³ and Ohloh⁴. The following paragraphs describe the approach that will be followed to answer the research questions mentioned in Section 2.

Summarization We plan to analyze different summarization techniques previously used to summarize content written in social media and product reviews [HL04, MLW⁺07, PE05].

³<http://www.epinions.com>

⁴<http://www.ohloh.net/>

We will focus on summarizing software features and the sentiments associated to these features. For this initial step we will apply topic modeling algorithms, frequent item set mining and sentiment analysis. We will evaluate our approach against manually generated summaries. Additionally, we will do a controlled experiment to measure the effort differences when developers and analysts deal with summarized and non-summarized user feedback.

Classification We will apply content analysis to a random sample of user feedback in order to find the characteristics that define feature requests, bug reports and reviews about existing software features. Furthermore, we will also use content analysis to define the characteristics of useful and useless feedback in the developers' and analysts' perspective. We will combine these results with data mining algorithms for the automatic classification of feedback. To evaluate the approach we will compare our results against manually labeled user feedback.

Diversification We will focus on retrieving feedback that mentions different sets of software features and sentiments. Our assumption is that retrieving feedback which mentions a wide spectrum of software features and sentiments can benefit the decision process of developers and analysts in software evolution concerning which features to implement, remove or improve. For extracting diverse user feedback we will use information retrieval algorithms employed to extract diverse product reviews [KD11, TNT11] and adapt them to the peculiarities of software feedback if needed. We will use metrics for diversification commonly used in information retrieval [CKC⁺08] for the evaluation of our approach.

4 Current State

We have designed and implemented an initial approach to summarize the content and sentiments present in user feedback [GB13] and are currently evaluating the approach on user reviews from the AppStore and GooglePlay. The approach summarizes the content present in user feedback as sets of co-occurring words and assigns each feedback a quantitative value which expresses the sentiment present in the feedback. It uses a topic modeling algorithm [BNJ03] for content summarization and lexical sentiment analysis [TBP⁺10] for extracting sentiments from text. We plan to improve the current approach by including other summarization techniques, and by evaluating more extensively against manually generated summaries and, as mentioned in Section 3, by measuring the effect that the summaries have on developers' and analysts' effort through controlled experiments.

References

- [BD04] Bernd Bruegge and Allen H Dutoit. *Object-Oriented Software Engineering Using UML, Patterns and Java-(Required)*. Prentice Hall, 2004.
- [BNJ03] David M. Blei, Andrew Y. Ng, and Michael I. Jordan. Latent Dirichlet Allocation. *The Journal of Machine Learning Research*, 3:993–1022, 2003.

- [CKC⁺08] Charles L A Clarke, Maheedhar Kolla, Gordon V Cormack, Olga Vechtomova, Azin Ashkan, Stefan Büttcher, and Ian MacKinnon. Novelty and diversity in information retrieval evaluation. In *Proceedings of the 31st International Conference on Research and Development in Information Retrieval*, pages 659–666. ACM, 2008.
- [GB13] Emitza Guzman and Bernd Bruegge. Towards Emotional Awareness in Software Development Teams. In *Symposium on the Foundations of Software Engineering - FSE '13*, 2013.
- [HL04] Mingqing Hu and Bing Liu. Mining opinion features in customer reviews. In *Proceedings of the International Conference on Knowledge Discovery and Data Mining - KDD '04*, pages 755–760. AAAI Press, 2004.
- [KD11] Ralf Krestel and Nima Dokoohaki. Diversifying product review rankings: Getting the full picture. In *Web Intelligence and Intelligent Agent Technology - WI-IAT '11*, volume 1, pages 138–145. IEEE, 2011.
- [KKLK05] Sari Kujala, Marjo Kauppinen, Laura Lehtola, and Tero Kojo. The role of user involvement in requirements quality and project success. In *Proceedings. 13th International Conference on Requirements Engineering, 2005.*, pages 75–84. IEEE, 2005.
- [KLF⁺11] Andrew J Ko, Michael J Lee, Valentina Ferrari, Steven Ip, and Charlie Tran. A case study of post-deployment user feedback triage. In *Proceedings of the 4th International Workshop on Cooperative and Human Aspects of Software Engineering*, pages 1–8. ACM, 2011.
- [Kuj03] Sari Kujala. User involvement: a review of the benefits and challenges. *Behaviour & information technology*, 22(1):1–16, 2003.
- [MLW⁺07] Qiaozhu Mei, Xu Ling, Matthew Wondra, Hang Su, and ChengXiang Zhai. Topic sentiment mixture: modeling facets and opinions in Weblogs. In *Proc. of the 16th international conference on World Wide Web - WWW '07*, pages 171–180, 2007.
- [Pag13] Dennis Pagano. *Portneuf - A Framework for Continuous User Involvement*. Doctoral thesis, Technische Universität München, 2013.
- [PE05] Ana-Maria Popescu and Oren Etzioni. Extracting product features and opinions from reviews. In *Proceedings of the Conference on Human Language Technology and Empirical Methods in Natural Language Processing - HLT '05*, pages 339–346. Association for Computational Linguistics, 2005.
- [TBP⁺10] Mike Thelwall, Kevan Buckley, Georgios Paltoglou, Di Cai, and Arvid Kappas. Sentiment strength detection in short informal text. *Journal of the American Society for Information Science and Technology*, 61(12):2544–2558, 2010.
- [TNT11] Panayiotis Tsaparas, Alexandros Ntoulas, and Evimaria Terzi. Selecting a comprehensive set of reviews. In *Proceedings of the 17th International Conference on Knowledge Discovery and Data Mining*, pages 168–176. ACM, 2011.
- [VMSC02] Karel Vredenburg, Ji-Ye Mao, Paul W Smith, and Tom Carey. A survey of user-centered design practice. In *Proceedings of the Conference on Human factors in Computing Systems: Changing our world, changing ourselves*, pages 471–478. ACM, 2002.