Exploring Sentiment Analysis to understand Software Crowdsourcing Challenges

Ricardo Rodrigo M. Melo, Leticia S. Machado, Cleidson R.B. de Souza[1]

Universidade Federal do Pará (UFPA)
{ricardorodrigomm, leticia.smachado}@gmail.com,
cleidson.desouza@acm.org

Abstract. *Companies are increasingly using crowdsourcing to accomplish specific software development tasks. This paper describes the initial results from an exploratory study using sentiment analysis in a software crowdsourcing context. In this case, we classify the polarity of the messages exchanged in online forums associated with crowdsourcing challenges. Our ultimate goal is to understand whether the crowd is talking good or bad things, and explore if negative messages can be a factor that influence the decrease in motivation among crowd members and, finally, if positive sentiments can help to encourage more contributions to challenges.*

1. Introduction

Software Crowdsourcing (SW CS) is a software development strategy where a large number of online users may be engaged to contribute in several software development activities. Such strategy, based on the crowd, has been used for companies who are seeking to increase the speed of their software development efforts [Stol, 2014] [Yang, 2015]. This strategy is usually structured around platforms that allow a requester submit a task to be performed and connect with the crowd that chooses to work in this task and provide a solution for it. These platforms usually explore a competitive approach: members of the crowd independently create a solution while competing against each other by monetary rewards for task completion.

Sentiment analysis, also called opinion mining, is the field of study of data mining which has the objective to analyze, understand, process, and extract the textual data, which may include opinions, sentiments, evaluation, assessment, attitudes, and emotions to entities such as products, services, organizations, individuals, and specific topics [Alamsyah, 2017].

In this paper, we argue that sentiment analysis can be used to explore the opinions of the crowd about a particular task in a software crowdsourcing context. We use mining techniques and natural language processing to understand how good (or bad) the crowd is talking about a particular task. Understanding how the crowd feels during a challenge, or at least part of it, can serve as a starting point to understand and mitigate existing communication and collaboration issues in SW CS platforms. In fact, it is important to identify factors that support the crowd motivation during SW CS challenges so that they can submit a solution for tasks [Stol, 2014]. Therefore, in this paper we investigate whether sentiment analysis might contribute to the SW CS context leveraging the diversity of the crowd and the communication among the crowd members.

2. Background

2.1. Software Crowdsourcing

Software Crowdsourcing, or simply SW CS, is a particular way of designing and creating software through the engagement of a pool of online members who can be tapped on-demand to contribute to various types of software development tasks.

SW CS is usually structured around software platforms. These are marketplaces that allow requesters to seek crowd members to perform their tasks and, at the same time, support crowd members in finding tasks to work on. Examples of SW CS platforms include TopCoder, uTest, and Passbrain.

Challenges in SW CS include motivation, remuneration, coordination and communication, and task decomposition. In particular, motivation is a topic that has received considerable attention in the literature, given that it is reported to be a major factor in SW CS project success [Stol, 2014].

2.2. Sentiment Analysis in forums

The term sentiment analysis has been used to describe different tasks and problems [Araujo, 2016]. Besides, there is the concept of Opinion Mining (OM), also known as sentiment classification, a recent subdiscipline at the crossroads of information retrieval and computational linguistics which is concerned not with the topic a text is about, but with the opinion it expresses [Esuli, 2006].

There are multiple existing sentiment analysis methods that explore different techniques, usually relying on lexical resources or learning approaches [Araujo, 2016]. In order to aid the extraction of opinions from text, recent research has tried to automatically determine the "PN-polarity" of subjective terms, i.e. identify whether a term that is a marker of opinionated content has a positive or a negative connotation [Esuli, 2006].

The importance of sentiment analysis is reported in the literature in many ways. In particular, for the context of communication forums, advantages of this approach include: (i) provide feedback on proposed solutions [Bhatia, 2012]; (ii) separate topics compared to other types of social media, providing a specific and exhaustive discussion [Alamsyah, 2017]; and (iii) provide useful information that is quickly exposed, thus benefiting members in the decision-making process [Nirmala, 2012] [Li, 2010].

Many companies such as Facebook, Twitter and Uber use sentiment analysis to monitor online conversations and understand real user reviews, complaints and suggestions about their product, service or brand.

In this study, we conduct sentiment analysis of the messages exchanged in the communication forums of SW CS projects. Our goal is to understand the coordination and collaboration issues among the crowd during these competitive challenges.

3. Methods

We chose the TopCoder SW CS platform because it is regarded, arguably, as the largest and most successful one [Stol, 2014]. A web scrapping process was used to collect data about communication forums hosted on TopCoder. From several development challenges (SW CS tasks) hosted on Topcoder, we selected a sample of 25 challenges for analysis. The selection was based on the criteria described in the Table 1.

Table 1. Summary of criteria

Category	Metrics	Description
(i) Active period of analysis	Between July and August 2017	Tasks available and open for registration in the period.
(ii) Number of registrants	> 15	Number of crowd workers who applied for a task.
(iii) Financial reward	> $500	Task budge. The value the task requester is willing to pay
(iv) Task phase	Registration and submission	Indicates the task status (registration, submission, review)
(v) Task challenge	Development Challenges	Indicates the challenge area (design, development, and data science)
(vi) Task category	Code	Captures the different task categories (conceptualization, design, coding, etc.)
Total	25 Challenges	Total of challenges during the period (July - 9 tasks, and August - 16 tasks)

The period of data collection is from July 2017 until August 2017, the busiest Topcoder months [Dubey et al., 2016]. A total of 1,184 messages were collected. However, after the criteria adopted for data analysis (Table 1), only 1,053 messages were used. More specifically, 496 messages were sent by co-pilots (mediators) and 557 messages by crowd members[1]. Messages were sent by 120 different people (11 co-pilots and 109 crowd members) distributed among 216 threads. We analyzed only the messages from the crowd members, i.e., 557 messages.

Figure 1 - Workflow of sentiment analysis performed

A data preprocessing step was used to separate the sentence in words (Tokenization), throw away words that are not important or less significant (Stopwords). The workflow of sentiment analysis performed is shown in Figure 1.The preprocessing step aims to reduce the volume of data before starting the execution of the analysis steps. The iFeel[2] sentiment analysis framework was used to this goal. iFeel is a Web Application that allows one to detect sentiments in any form of text including unstructured social media data. It is free and gives access to 18 sentiment analysis methods, namely: AFINN, Emolex, Emoticons, EmoticonDS, Happiness Index, OpinionFinder, NRCHashtag, Opinion Lexicon, Panast, SANN, SASA, Sentiment140, Sentistrength, SentiWordNet, SOCAL, StanfordDeep Learning, Umigon, and Vader. These methods, implemented in the iFeel tool [Messias, 2016] are regarded as the “state-of-the-practice” sentiment analysis methods for English [Araujo, 2016].

[1] Co-pilots are *platform moderators* who are "a special person to answer questions from the crowd", i.e., these moderators’ goal is to alleviate the communication difficulties between customers and crowd members [Ågerfalk et al, 2015].

[2] http://blackbird.dcc.ufmg.br:1210/

4. Results

As mentioned, 557 messages from crowd members were submitted to the iFeel application in the English language. Upon completion of the analysis, the results of each of the 18 methods used by the application were stored in spreadsheets. Polarity Results for 18 Sentiment Analysis Methods are show in Figure 2.

Method	Positive	Negative	Neutral
1. AFINN	44,7%	14,8%	40,5%
2. Emolex	43,3%	15,0%	41,7%
3. Emoticons	0,0%	0,0%	100,0%
4. EmoticonDS	98,6%	0,0%	1,4%
5. Happiness Index	28,7%	4,4%	66,9%
6. MPQA	17,6%	8,4%	74,1%
7. NRCHashtag	21,0%	63,3%	15,8%
8. Opinion Lexicon	36,1%	16,8%	47,1%
9. Panast	0,0%	0,2%	99,8%
10. SANN	63,9%	11,6%	24,6%
11. SASA	24,4%	7,8%	67,9%
12. Sentiment140	42,3%	40,9%	16,8%
13. Sentistrength	34,3%	16,6%	49,1%
14. SentiWordNet	70,5%	23,4%	6,2%
15. SOCAL	36,1%	19,2%	44,7%
16. StanfordDeep Learning	8,6%	59,5%	31,9%
17. Umigon	19,0%	20,2%	60,9%
18. Vader	22,0%	2,4%	75,6%

Figure 2 - Polarity Results for 18 Sentiment Analysis Methods

For each method used, the cell (positive, negative or neutral) representing the most expressive polarity was highlighted. Of the 18 methods presented, only 2 obtained results with greater negative expressiveness among the messages sent by the crowd. Six out of the 18 methods obtained a more expressive result as positive and, finally, 10 methods classified the messages as mostly neutral[3]. In summary, methods that the most expressive polarity was negative represented 11% of the total methods used in this study. 33% represented positive polarity and most methods used (56%) provided a neutral classification (See Figure 3).

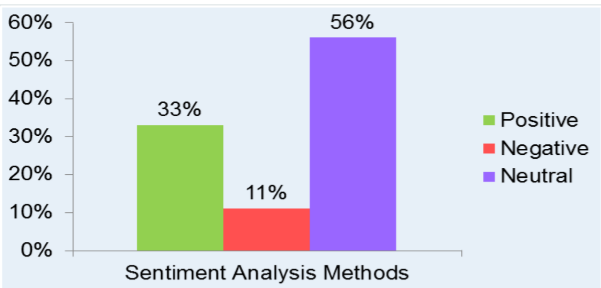

Figure 3 – Summary of Sentimental Analysis Methods

5. Discussion

Keeping high motivation to perform and submit a solution for tasks is one of the main concerns for software crowdsourcing platforms [Stol, 2014]. Meanwhile, "bad messages", i.e., messages expressing negative sentiments, might influence the crowd, especially its motivation. If the crowd is mostly expressing a negative sentiment, that might suggest a problem with the task or the platform or other aspect. In any case, it is

[3] It is important to mention that when unsure about the results, most methods classify the messages as neutral.

essential to identify and to change the situation to maintain the crowd motivated. In other words, if good words (positive sentiments) can help encourage crowd contributions to the challenges, it is important to monitor what is going in the forums and act when necessary. We speculate that this approach might even increase the number of solutions submitted in crowdsourcing challenges. Previous studies have looked into sentiment analysis of messages in news reports and blogs containing large volume of public opinion information [Bhatia, 2012] [Alamsyah, 2017] [Nirmala, 2012] [Li, 2010]. We are not aware of any study using sentiment analysis to study crowdsourcing challenges.

We believe that classifying the polarity of the messages exchanged in challenge forums to understand how positive or negative is the crowd sentiment might serve as a starting point to investigate what aspect of the SW CS task a crowd member is discussing. For example, a member would want to separate messages related to: unclear task documentation, reduced time to solve the task, unfair and uncareful issues among crowd competitors. In other words, the crowd might be expressing different feelings about different topics. For instance, a positive feeling about the platform, while at the same time, a negative feeling about the task. Currently, we are not able to make this distinction, but we believe this is something important to be explored.

Finally, sentiment analysis is a subjective information and might help to understand the social sentiment of the crowd while monitoring online conversations. Even though several methods have been presented in this study to classify the polarity for the messages sent by the crowd, it is necessary to better understand the goals and contexts of each method, because there are different metrics implemented in these methods. In short, it is necessary to evaluate which methods are more adequate to represent polarity in the context of software crowdsourcing challenges.

6. Conclusions and Future Work

In this exploratory paper, we used sentiment analysis in the software crowdsourcing context. We analyzed 557 messages sent by crowd members in 25 TopCoder challenges.

After classifying the polarity of the messages exchanged in the forums, we concluded that most of the analyzed sentiments are classified as neutral. However, some classification methods presented a significant number of positive polarities. The main goal of this paper was to understand how much the crowd is talking about positive or negative aspects in challenge forums, i.e., are there more “good” or “bad” messages?

The limitations of this paper include the low number of challenges and messages analyzed. In this way, as future work, we plan to broaden the analysis of forums expanding analysis for data collected in 2018 as well: 62 challenge forums, 2.471 messages, of these 1.176 messages sent by co-pilots and 1.295 messages posted by the crowd only.

Acknowledgment

We thank PROPESP/UFPA and CNPq for the financial support (#420801/2016-2 and #311256/2018-0). The first and second authors also thank CAPES for financial support.

References

Ågerfalk, P. J., Fitzgerald, B., and Stol, K. J. (2015). Software Sourcing in the Age of Open: Leveraging the Unknown Workforce. Springer, 71p.

Alamsyah, A., Nuruz, F. (2017). "Measuring Public Sentiment Towards Service Level in Online Forum using Naïve Bayes Classifier Method and Word Cloud". In CRS-ForMIND International Conference and Workshop 2017.

Araujo, M.; Reis, J.; Pereira, A.; Benevenuto, F. "An Evaluation of Machine Translation for Multilingual Sentence-level Sentiment Analysis". In: ACM Symposium on Applied Computing (SAC), 2016, ACM Press, Pisa, Italy.

Bhatia, S., Biyani, P., &Mitra, P. (2012). "Classifying user messages for managing web forum data". In Fifteenth International Workshop on the Web and Databases, WebDB pp. pp.13-18. Scottsdale, AZ: ACM.

Dubey, A.; Abhinav, K.; Taneja, S.; Virdi, G.; Dwarakanath, A.; Kass, A.; Kuriakose, M.S. Dynamics of software development crowdsourcing. In: ICGSE, 2016, pp. 49-58.

Esuli, A., & Sebastiani, F. (2006). SENTIWORDNET: A Publicly Available Lexical Resource for Opinion Mining. LREC.

Gray, L. Suri, S., Ali, S., Kulkarni, D. (2016) "The crowd is a collaborative network". In: Proceedings of the 19th CSCW Conference. ACM. p. 134-147.

Karim, M.R.; Saremi, R.; Ruhe, G. "Who should take this task? Dynamic decision support for crowd workers". In: ESEM ACM/IEEE, 2016, p. 8.

LaToza, T.D.; Chen, M.; Jiang, L.; Zhao, M.; Van Der Hoek, A. "Borrowing from the crowd: A study of recombination in software design competitions". In: 37th International Conference on Software Engineering, v1, 2015, pp. 551-562.

Li, N., and Wu, D.D. (2010). "Using text mining and sentiment analysis for online forums hotspot detection and forecast". Decision Support Systems, 48, 354-368.

Machado, L. S.; Zanatta, A. L.; Marczak, S.; Prikladnicki, R. "The Good, the Bad and the Ugly: An Onboard Journey in Software Crowdsourcing Competitive Model". In: 4th International Workshop on CrowdSourcing in Software Engineering (CSI-SE). Collocated with the 39th ICSE, 2017, pp. 2-8.

Messias, J; Diniz, João P.; Soares, E.; Ferreira, M.; Araújo, M.; Bastos, L.; Miranda, M.; Benevenuto, F. (2016). "Towards sentiment analysis for mobile devices". In: IEEE. Advances in Social Networks Analysis and Mining, 2016. p. 1390–1391.

Nag, S.; Heffan, I.; Saenz-Otero, A.; Lydon, M. "SPHERES Zero Robotics software development: Lessons on crowdsourcing and collaborative competition". In: Aerospace Conference, 2012 IEEE, pp. 1-17.

Nirmala, K., and Bhaskarn,V., (2012). "Online Forums Hotspot Prediction Based on Sentiment Analysis". Journal of Computer Science.

Peng, X.; Ali Babar, M.; Ebert, C. "Collaborative Software Development Platforms for Crowdsourcing". IEEE Software, v. 31, n. 2, 2014, pp. p. 30–36.

Stol, K.-J.; Fitzgerald, B. "Two's company, three's a crowd: a case study of crowdsourcing software development". In: ICSE, ACM Press, 2014, p.p.187.

Tausczik, Y.; Wang. P. "To Share, or Not to Share? Community-Level Collaboration in Open Innovation Contests". CSCW, 2017.

Yang, Y.; Saremi, R. "Award vs. worker behaviors in competitive crowdsourcing tasks". In: ESEM, IEEE, 2015, pp. 1-10.