Sentiment Analysis of Commit Comments in GitHub: An Empirical Study

Emitza Guzman, David Azócar, Yang Li
Technische Universität München
Faculty of Informatics
Garching, Germany
emitza.guzman@mytum.de, dazocar@gmail.com, liya@in.tum.de

ABSTRACT

Emotions have a high impact in productivity, task quality, creativity, group rapport and job satisfaction. In this work we use lexical sentiment analysis to study emotions expressed in commit comments of different open source projects and analyze their relationship with different factors such as used programming language, time and day of the week in which the commit was made, team distribution and project approval. Our results show that projects developed in Java tend to have more negative commit comments, and that projects that have more distributed teams tend to have a higher positive polarity in their emotional content. Additionally, we found that commit comments written on Mondays tend to a more negative emotion. While our results need to be confirmed by a more representative sample they are an initial step into the study of emotions and related factors in open source projects.

Categories and Subject Descriptors

D.2.9 [**Software Engineering**]: Management-Programming teams

General Terms

Human Factors

Keywords

Human Factors in Software Engineering, Sentiment Analysis

1. INTRODUCTION

Software development is a highly collaborative activity in which participants use mailing lists, forums, software code repositories and issue tracking tools, among others, to manage their work. These collaboration artifacts are communication channels where development teams can express their emotions concerning, for example, their satisfaction with the project or their difficulties while dealing with certain tasks. Previous research has shown that emotions affect task quality, productivity, creativity, group rapport and job satisfaction [1]. Developers and managers need to be aware of the emotions of the projects they are involved to take corrective actions when necessary and to have a better understanding of the social factors affecting the project. Determining emotions in open source projects is particularly challenging as traditional ways of collecting information through experience reports, interviews with managers or surveys are more difficult due to the non-hierarchical structure of open source projects, the geographical distribution of developers among different regions and the volunteer basis in which developers contribute. In this work we research the emotions expressed by developers who work in open source projects and collaborate through source code repositories. Additionally, we study the relationship between the expressed emotions and other factors, such as, used programming language, time and day of the week in which the collaboration occurred, team distribution and project approval. The research questions that guide our work are:

R1) Are emotions in commit comments related to the programming language in which a project is developed?
R2) Are emotions in commit comments related to the day of the week or time in which the commits were written?
R3) Are emotions in commit comments related to the team geographical distribution?
R4) Are emotions in commit comments related to project approval?

As done in previous research work [3], we propose to use lexical sentiment analysis on collaboration artifacts to measure expressed emotions and use statistical analysis to measure their interplay with other factors. For our study we analyze commit comments from different GitHub[1] projects.

2. RESEARCH METHODOLOGY

In the following sections we describe the data set in which we perform our analysis, as well as the lexical sentiment analysis approach we use.

2.1 Research Data

We extracted our analysis data from the dataset provided by Gousios [2], this dataset includes 90 of the 10-top starred software projects for the top programming languages on Github. We analyzed a total 60425 commit comments.

[1] https://github.com/

MSR'14, May 31 – June 1, 2014, Hyderabad, India

http://dx.doi.org/10.1145/2597073.2597118

Table 1: Example of SentiStrength scores in high emotional commit comments from the GitHub dataset.

Commit message	Word and sentence score	Total score
Sigh? It's fixed, man – rejoice!!	Sigh?[sentence: 1,-1] It's fixed ,man –rejoice[4]!![+1 punctuation emphasis] [sentence: 5,-1]	5
Wow amazing thread! even If I'm not a Rails developer!	Wow[3] amazing[3] [+1 consecutive positive words] thread![+1 punctuation emphasis] [sentence: 5,-1] even If I'm not a Rails developer ![+1 punctuation mood emphasis] [sentence: 2,-1]	5
MY PRECIOUSSS!!!	MY PRECIOUSSS[3] [+0.6 spelling emphasis] !!![+1 punctuation emphasis] [sentence: 5,-1]	5
If PHP code is producing errors with register_globals on you are terrible terrible programmer. If you are using magic_quotes you are simply stupid.	If PHP code is producing errors[-2] with register_globals on you are terrible[-4] terrible[-4] [-1 consecutive negative words] programmer.[sentence: 1,-5] If you are using magic_quotes you are simply stupid[-3].[sentence: 1,-3]	-5
But this commit message makes me sad :cry:	But this commit message makes me sad[-4] :cry[-4] [-1 consecutive negative words] :[sentence: 1,-5]	-5
This is really terrible - changing :private to :public without any deprecation warning? Not cool.	This is really terrible[-4] [-1 booster word] -changing :private to :public without any deprecation warning?[sentence: 1,-5] Not cool[2] [*-0.5 approx. negated multiplier] [sentence: 1,-5].	-5

2.2 Sentiment Analysis

Sentiment analysis is the process of assigning a quantitative mood value (positive or negative) to a text snippet. For analyzing emotions in review comments we use SentiStrength[2], a lexical sentiment extraction tool specialized in dealing with short, low quality texts. Previous research has shown that SentiStrength has a good accuracy for short texts in Twitter and movie reviews [4]. A manual inspection of a small sample of commit comments in GitHub revealed that the messages were generally short and written in informal language, making SentiStrength a good candidate for analyzing the emotions in commit comments. SentiStrength assigns fixed scores to tokens in a dictionary where common emoticons are also included. Words with a negative emotion are given a value between [-5, -1] and words with a positive emotion are given a value in the [1, 5] range. The 1 and -1 values are used to give neutral scores to words, whereas 5 and -5 are used for words with a very positive and very negative emotion respectively. All the scores assigned by SentiStrength are integers. For example, "love" is assigned a score of $\{3, -1\}$ and "hate" a $\{1, -4\}$ score. Modifier words and symbols also alter the score. SentiStrength divides the text in commit comments into snippets of one or more sentences and assigns a positive and negative value to each of the sentences by taking the maximum and minimum scores among all the words in the sentence. The positive and negative score of a snippet is calculated by taking the maximum and minimum scores of the sentences forming it. We compute the emotion of an entire commit by calculating the positive and negative emotion score snippet average. For the case where both positive and negative snippet average emotion scores are in the [-1,1] range we assign the whole commit comment a score of 0. When the snippet average negative emotion score times 1.5 is less than the average positive score, we assign the commit emotion score to the negative snippet average score. When the opposite occurs the commit is assigned the positive snippet average emotion score. When comparing the snippet average emotions scores we multiply the snippet average negative scores by 1.5 because negativity is considered to be less frequent in human written texts[3]. Table 1 shows some examples of highly emotional commit comments from our dataset and their emotion scores. We consider a commit comment to be positive if its emotion score is in the (1,5] range, negative when it is in the (-1,-5] range and neutral when the emotion score is in the [-1,1] range.

[2]http://sentistrength.wlv.ac.uk/

3. RESULTS

In the following sections we present an analysis of the emotions expressed in commit comments relation between and their relation with additional factors such as programming language, time and day of the week, team distribution and user approval.

3.1 Emotions in Commit Comments

We analyzed the emotions present in the commit comments of all projects having more than 200 comments. In total we analyzed 29 projects. The average emotion score of the commit comments for each of the projects tended to neutrality (scores between 1 and -1). This is because many of the commit comments only describe technical aspects and no emotions are present or are only slightly positive or negative. Figure 1 shows the 6 projects with the highest number of commit comments and the emotion average score for each one of them.

While the average emotion score of the commit comments tended to neutrality, there were some highly emotional commit comments and other positive and negative commit comments whose emotional content was not reflected in the average, e.g., the commit comments shown in Table 1. In order to further examine this content we analyzed the proportion of positive, neutral and negative commit comments. Figure 2 shows this distribution for the 6 projects with the highest number of commit comments. The figure allows us to compare the distribution of emotions present in the commit comments among the different projects independently of the amount of commit comments in each project. It shows that the analyzed projects have an approximate equal distribu-

[3]As explained in the SentiStrength user manual: http://sentistrength.wlv.ac.uk/

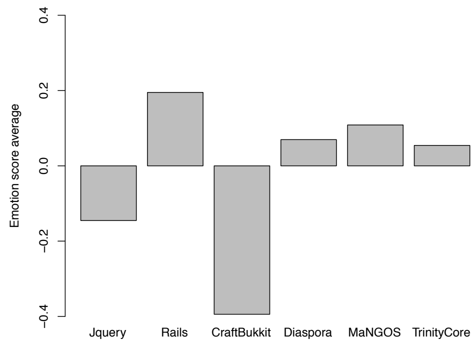


Figure 1: Emotion score average per project.

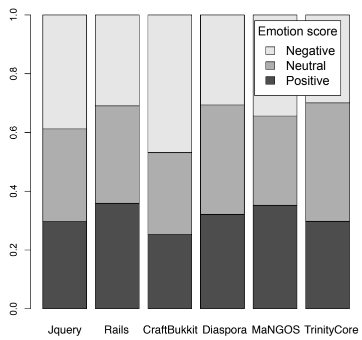


Figure 2: Proportion of positive, neutral and negative commit comments per project.

tion between the amount of positive, neutral and negative messages, despite having an average that tends to neutrality.

3.2 Emotions and Programming Language

The projects in our dataset were implemented in 14 different programming languages. Table 2 shows the average emotion score of five programming languages. In the table we observe that Java projects tend to have a slightly more negative score than projects implemented in other languages. We performed a Wilcoxon rank sum test which confirmed that commit comments from projects written in Java are more negative than projects implemented in six other languages (p-value<=0.002), namely C, C++, JavaScript, PHP, Python and Ruby. Statistical tests on the emotion scores of the commits of the remaining programming languages did not yield significant results.

3.3 Emotions, Day and Time of the Week

We grouped all of the commit comments into different categories representing the day of the week they had been written. Most commit comments, 78%, were written during the week, while 22% were written during weekends. Table 3(b) shows an overview of the average emotion score of commit comments grouped by weekday. While the emotion scores are generally neutral, the emotion of commit comments on Mondays tends to a more negative emotion. A Wilcoxon rank sum test of Monday against each of the other days confirmed that commit comments were more negative on Monday than on Sunday, Tuesday, and Wednesday (p-value<=0.015). No other significant relations were found between the emotion scores of the commit comments of the other week days.

Table 2: Emotion score average grouped by programming language.

Language	Commits	Mean	Stand. Dev.
C	6257	0.023	1.716
C++	16930	0.017	1.725
Java	4713	-0.144	1.736
Python	2128	-0.018	1.711
Ruby	15257	0.002	1.714

Table 3: Emotion score average of commit comments grouped by weekday.

Weekday	Commits	Mean	Stand. Dev.
Monday	9517	-0.043	1.732
Tuesday	9319	0.005	1.712
Wednesday	9730	0.008	1.716
Thursday	9538	0.001	1.728
Friday	9076	-0.016	1.739
Saturday	6701	-0.027	1.688
Sunday	6544	0.022	1.717

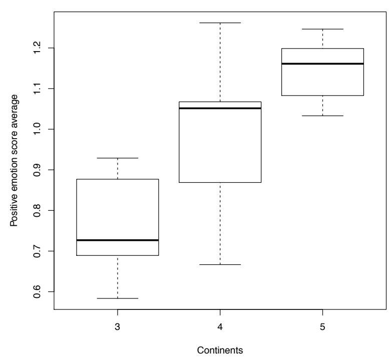
Table 4: Emotion score average of commit comments grouped by time of the day.

Time of Day	Commits	Mean	Stand. Dev.
Morning	12714	0.001	1.730
Afternoon	19809	0.004	1.717
Evening	16584	-0.023	1.721
Night	11318	-0.016	1.713

To analyze the relationship between the time of the day in which a commit comment was written and the emotion in the commit message we analyzed the timestamp and emotion score of each commit comment. We separated the commit comments into comments written during the morning [6:00-12:00), afternoon [12:00-18:00), evening [18:00-23:00) and night [23:00-6:00). Commit comments written in the afternoon were the most common, 33%, while 27% were written in the evening, 21% in the morning and 19% at night. We performed a Wilcoxon rank sum test against the commit comments of each part of the day and did not find any significant difference between their emotion scores, with exception of the emotion scores of afternoon and evening (p-value = 0.066), where the emotion scores of the afternoon comments were significantly more positive than those of the evening.

3.4 Emotions and Team Distribution

To study the relationship between the emotions expressed in commit comments and the geographical distribution of the projects' developers we analyzed the data of the 29 projects with more than 200 commit comments. We manually grouped the locations of all developers associated to the 29 projects into countries. Additionally, we also grouped the different countries into their corresponding continents. The projects were on average distributed in 21.17 different countries and 76% of the projects were distributed on at least 4 continents. We found no significant correlation between the emotion score of the commit comments and the number of countries or continents in which each project was distributed. However, the Spearman correlation between the average of all positive commit comments (those with an emotion score larger than 1) and the number of countries in which a project is distributed is a moderate positive correlation of 0.418 . This means that the number of country locations in a project is positively correlated to the amount of emotion (emotion polarity) present in positive commit comments. To gain more insight of how strongly positive emotion scores differ when the geographical distribution increases we performed an additional statistical test. For this we separated all projects into 5 groups, differentiating them by the number of continents they are distributed in. Because our data was not normally distributed we applied a Mann-Whitney test, which revealed that for the positive average score the projects distributed in 3, 4 and 5 continents have significantly different medians. This shows us that as the number of continents involved in a project increases, the amount of emotion in positive commit comments also increases (see Figure 3).



Continents	Mean	Stand. Dev.
2	1.031	–
3	0.761	0.141
4	0.996	0.157
5	1.148	0.078

Figure 3: Relationship between continent distribution and positive emotion score average (only one project is distributed in 2 continents).

Figure 4 visualizes the relationship between the average emotion score of positive commit comments, the country and continent distribution and the user starring. In the Figure we see that projects with a higher distribution, either country or continent wise, tend to have a higher amount of positive emotions in their positive commit comments. We discuss the influence of user starring in the next section.

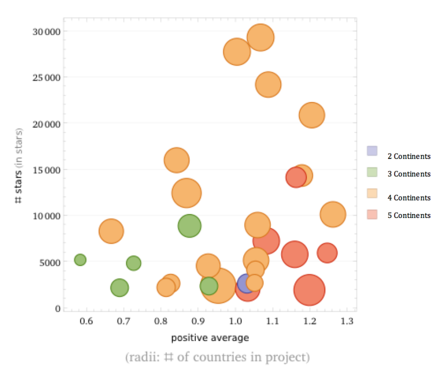


Figure 4: Positive emotion score average, country and continent distribution, and star relationship.

3.5 Emotions and Project Approval

GitHub allows its users to *star* its hosted projects, as a way of showing appreciation. Starring in GitHub can be seen as the equivalent of *liking* in other social media platforms. We tested the correlation between the average emotion score of each projects and its number of stars but found no correlation. The Spearman correlation between the average of all positive commit comments belonging to a project and its number of stars was of 0.316 indicating a positive weak correlation between star rating and the amount of positive emotion present in positive commit comments.

4. CONCLUSIONS

We presented lexical sentiment analysis as a mechanism for extracting emotions expressed in commit comments as quantitative values. Additionally, we analyzed the relationship between these extracted emotions and different factors such as, used programming language, time and day of the week in which the commit comments were written, team distribution and project approval. While the sample is not representative, our approach presents possible research directions that could be explored in future work to gain a deeper understanding of the factors that affect the emotions of developers working in open source projects, as well as how these emotions have an effect on developers' outcome. Our studies confirm the importance of not only considering the average emotion score of a whole document, in this case commits, when studying emotions with the aid of sentiment analysis. The positive and negative average emotion score, as well as the distribution of positive, negative and neutral documents should also be taken into consideration to get a deeper understanding of the emotional content, as averages tend to opaque this information. The definition of further metrics concerning emotions for open source projects will be the subject of future work.

5. ACKNOWLEDGMENTS

This work was partially funded by a scholarship from the Mexican National Council of Science and Technology.

6. REFERENCES

[1] M. De Choudhury and S. Counts. Understanding affect in the workplace via social media. In *Proc. of the 2013 conference on Computer supported cooperative work - CSCW '13*, pages 303–316, Feb. 2013.

[2] G. Gousios. The GHTorrent dataset and tool suite. In *Proceedings of the 10th Working Conference on Mining Software Repositories*, MSR'13, pages 233–236, 2013.

[3] E. Guzman and B. Bruegge. Towards Emotional Awareness in Software Development Teams. In *Foundations of Software Engineering - FSE '13*, pages 671–674, 2013.

[4] M. Thelwall, K. Buckley, and G. Paltoglou. Sentiment strength detection for the social web. *Journal of the American Society for Information Science and Technology*, 63(1):163–173, Jan. 2012.