

ECE1785 Project Proposal: Emotion Mining in Coding

Jiaming Xu 1007698831¹, Teng Yue 1007826792¹, and Wenrui Xu 1008313228¹

¹Group 8, Department of Electrical & Computer Engineering, University of Toronto

February 14, 2022

1 Background

Nowadays, with the increasing demand of the Internet applications, software development engineers are playing more and more important roles in the industry. Meanwhile, the coding efficiency of them, which contributes to companies' competitiveness in the industry, is of vital significance to dot-com companies. In order to determine what factors have great impacts on the coding efficiency, many companies have done surveys in this area, and they have made some adjustments to promote engineers' coding efficiency according to the results. As is widely acknowledged, a positive emotion can be a catalyst to work efficiency. However, as far as we are concerned, the previous surveys paid too much attention to the external factors like gender, age, family status, instead of engineers' internal factors, such as emotions. Very few researches connected the relationship between coding efficiency and emotions. Therefore, we expect to mine out the relationship between them, in a gesture to attract more attention to the psychological health of software development engineers.

2 Research Questions

How do we define and categorize emotions?

How do we know engineers' emotions from their codes?

How do we measure the efficiency of coding?

Does the emotion change have impact on the efficiency of coding?

3 Methodology

As to the categories of emotions [3], we plan to use Paul's basic emotion framework [1]. This is very convincing and easy enough to operate on.

Some research plans considered: 1. Find out some open-source code repositories, get the commit logs of them and analyze the emotions contained in the commit messages. Calculate the number of lines each contributor committed in a certain period of time. Compare the efficiency in different emotions. We found some similar emotional mining works on the website which is written by Ramiro [2]. But what he did is just mine out the emotions without connecting it with the coding efficiency.

2. Do some interviews among peers with regard to emotions. Ask the participants to make records of their moods every time they start to work, and calculate their coding efficiency.

We think these methods are appropriate in three ways:

1. As a student, we don't have enough time and money to do interviews on a large amount of people. These methods can help us get the original data quickly and easily.

2. The existing emotional mining and NLP technology is well developed, we think we can mine out enough information from the commit messages.

3. The working efficiency is hard to define, so we can only count on the coding lines to measure it.

4 Expected Results

We expect to mine out the relationship between emotions and the coding efficiency. We should quantitatively analyze the results to find out the difference of the coding efficiency between different emotions. We will contribute to the industry after we make it clear that the psychological health of programmers can influence the coding efficiency greatly.

References

- [1] P. Ekman. Basic emotions. *Handbook of cognition and emotion*, 98(45-60):16, 1999.
- [2] R. Gómez. Exploring Expressions of Emotions in Github Commit Messages. Website, 2012. <https://geeksta.net/geeklog/exploring-expressions-emotions-github-commit-messages>.
- [3] Wikipedia. Emotion-wikipedia. Website, 2022. <https://en.wikipedia.org/wiki/Emotion>.