Understanding the Personality Traits of Stack Overflow Users: Text Analysis with IBM Personality Insights

Clay Asato and Bruno C. da Silva
Department of Computer Science and Software Engineering
California Polytechnic State University
San Luis Obispo, California, United States of America
casato@calpoly.edu, bcdasilv@calpoly.edu

Abstract—The human aspect of programming has become a popular area of study in the Software Engineering community in recent years. Industry professionals and researchers alike are becoming more aware of sentiment, emotion, and personality traits in software engineering by exploring advancements in text analysis tools The Big Five personality model has been considered as a reliable model for mapping human personality traits. IBM Cloud Watson services provides an implementation of the Big 5 model for analyzing personality facets on a body of text. Therefore, this project aims to analyze the posts of Stack Overflow users to find a link between successful users and certain personality traits. Posts from Stack Overflow Users are concatenated and sent to the IBM Personality Insights API. Results are programmatically aggregated. Statistical analysis is run on the results. Trends in both trial groups were discovered. Both displayed high Openness and Neuroticism as well as low Agreeableness.

Index Terms—sentiment analysis, software engineering

I. INTRODUCTION

Sentiment Analysis is a relatively new area in Software Engineering research. The most common domain of Sentiment Analysis is consumer analytics and marketing [1] such as movie reviews [2]. On the surface, software does not appear to be connected to emotion or personality, but the human developers behind the software are emotional beings. Thus, impact of their feelings and humanity are more pronounced than one might expect. Recent study aims to understand the impact of this human aspect of Software Engineering.

Other research in Sentiment Analysis in Software Engineering has explored software code repositories like public GitHub repositories and forums like Stack Overflow [3]. Stack Overflow is a developer centered question and answer forum, with an emphasis on community interaction. Every submission, either question or answer, is rated by the community with votes, either +1 or -1. Answers can be edited by the community and accepted as the best answer [4]. While using source code as examples is helpful, the main interactions users have are in natural language, providing a unique opportunity for linguistic sentiment analysis. Previous work in Personality Analysis was performed in Github contributors. Rastogi and Nagappan

measured the differences in personality traits between different types of contributors on Github products and their varying roles in the projects. They also found that the personality traits exhibited by some users evolved overtime with the project they worked on [5] Related work in Stack Overflow Sentiment Analysis includes Bazelli et al. [3] use of Linguistic Inquiry and Word Count (LIWC) to find differences in personality traits of high, medium, and low reputed contributors as well as high and low scoring posts. Higher reputation and scoring users were found to have higher extroversion compared to users with lower reputation and scores.

Therefore our goal is to analyze the most successful authors on Stack Overflow to find a trend of personality types across these authors. To achieve this goal, we have defined the following research questions:

RQ1: What personality traits do the authors of the highest scoring Stack Overflow posts possess?

RQ2: What personality traits do the Stack Overflow users with the highest reputations possess?

To address the research questions, we have designed an empirical study analyze posts from the popular developers and measure their personality facets based on their posted text on the platform.

This report will begin with background information on the data collected and analyzed and the tools used to answer the research question. The results will be presented with analysis and conclusion at the end as well as opportunities for future study.

II. MEASURED CRITERIA

A. Post Score

Initially, we queried the authors of the highest scoring posts of all time and aggregated their highest scoring posts and analyzed that text. Post score is expressed by the number of up votes minus the number of down votes.

The initial and first primary run worked off of raw post score, but there were several issues with this metric. First, the score is just the net of votes, so a post with 1000 up votes and 950 down votes could have the same score as a post with

55 up votes and 5 down votes. The community's reaction to these two posts could be vastly different while the post score would be the same on the surface. Second, there were some instances of the highest scoring posts being only one of very few total posts by an author, so all of the posts by that author did not meet the 600 word minimum for a confident answer from IBM Personality Insights.

B. Reputation

Reputation scores are provided by Stack Overflow and follow a specific formula. Calefato et al. found that there was a positive correlation between successful answers on Stack Overflow and the reputation of the author [6]. Reputations are calculated by assigning point differentials to community interactions. It is usually a good indicator of a user who is active in the community and not only there get help for themselves. Contributions of help to other users like good answers to questions are weighted more than good questions or votes, so one can reasonably assume that those with higher reputation are the ones who contribute the most to the overall success of the site. The second main trial used reputation as the metric to find the most successful Stack Overflow users.

III. ENVIRONMENT AND METHODS USED

A. Data

Stack Overflow data was sourced by the SOTorrent project [7] from the Stack Overflow Dec. 2018 datadump. Data was provided in CSV and XML files and converted to a SQL database with provided SQL scripts. Queries were facilitated with MySQL and its Java connector.

B. Analysis

1) IBM Personality Insights: IBM Personality Insights is part of the IBM Cloud service and uses the Watson AI. The API takes in a string of natural language text from a single author and returns percentile ratings for multiple personality facets, Big Five, Needs, and Values [8]. The Big Five personality facets, Openness, Conscientiousness, Extraversion, Agreeableness, and Neuroticism, were used for this study. IBM counts a "high" score to be that above 0.75 and a "low" score to be below 0.25.

The IBM Personality Insights Java API was used to analyze the post body text after cleansing of source code and HTML tags. Results were returned as JSON objects and saved to disk.

C. Collection

Query and data collection was facilitated by a Java program. Text was concatenated and cleansed of HTML tags and source code and sent to IBM Personality Insights. Results as JSON files were accessed for their Big-5 Personality metrics and output to comma separated value files and converted to Microsoft Excel format for quick analysis.

IV. TRIALS

A. Initial Trial

The first trial involved selecting the users with the most popular posts of all time by Score value and selecting each of their highest scoring posts, cleansing and appending them together, and sending them to IBM Personality Insights.

```
SELECT OwnerUserId FROM Posts ORDER BY
    Score DESC;
```

Of the top 100 posts, only authors with enough posts to meet the minimum of 600 words had their posts analyzed. The ID of the author was parameterized in the following SQL query.

```
String query = "SELECT␣Body␣FROM␣Posts␣
    WHERE␣OwnerUserId='"
            + Integer.toString(userId)
            + "'␣ORDER␣BY␣Score␣DESC;"
```
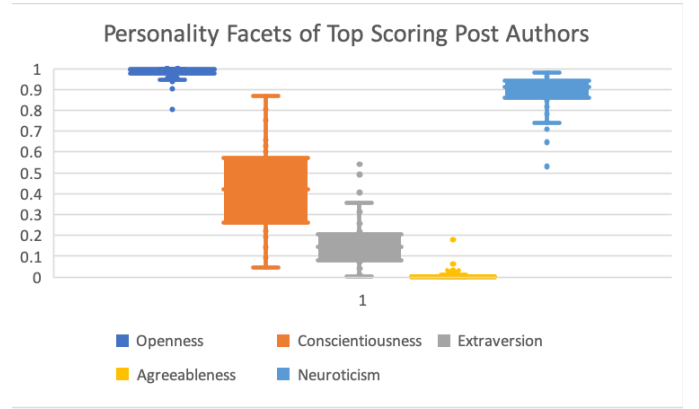

Fig. 1. Personality Insights results by top scoring posts

1) Results: Initial results pointed to interesting trends in the top scoring post authors on Stack Overflow Figure 1 expresses each of the character facets of the selected users. The trial of 82 authors returned trends in high Neuroticism and Openness with mixed Extraversion and Conscientiousness, and low Agreeableness.

2) Lessons Learned: The initital trial run was meant to take in 100 users, but the minimum word count for IBM Personality Insights restricted some entries from analysis because the authors did not have enough posts overall to meet the minimum word count for Personality Insights analysis. This was addressed in the second trial run.

B. Second Trial

With the first trial laying out the groundwork for the experiment, the second trial attempted to focus on the author of the posts more than the posts themselves. As such, the precalculated Reputation score was used as the criteria for successful Stack Overflow users. Users with high reputations could reasonably be assumed to engage with the rest of the

community in a positive and helpful manner. In addition, whereas some authors of the top scoring posts did not have enough raw post text to analyze or could potentially have only one high scoring post, users with high reputation would also be more likely to contribute consistently good content to the site.

The top 150 users with the highest reputation scores as of the December 2018 SOTorrent dataset were selected

```
SELECT Id FROM Users ORDER BY Reputation
    DESC LIMIT 150;
```

Each user's top scoring posts up to the minimum word count of 600 were analyzed with IBM Personality Insights using the same structured query as in the first trial.
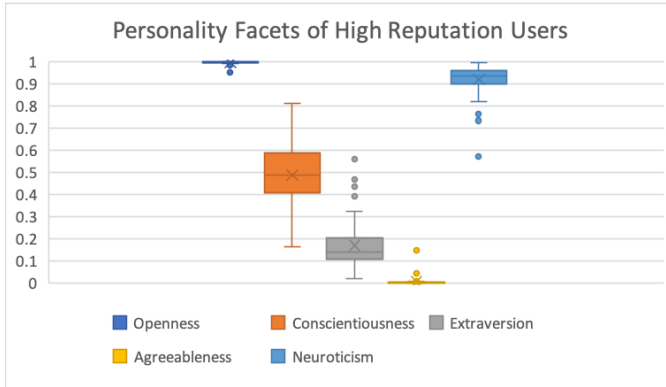

Fig. 2. Personality Insights results by top reputation users

1) Results: The resulting trial set was 107 users. Similar to the first trial, the users with the highest reputations all posessed high Neuroticism and Openness and low Agreeableness with mixed Extraversion and Conscientiousness as seen in Figure 2.

V. THREATS TO VALIDITY

The Personality Insights API is mainly meant for businesses to use for marketing purposes. It is very powerful in its analysis, but results could be skewed by technical keywords used in Stack Overflow [1]. Stack Overflow as a platform exists in an intermediate state between purely technical and purely casual interaction, and as such might not be suited for sentiment analysis tools made strictly for Software Engineering.

VI. CONCLUSIONS

Both trials yielded similar results. The highest scoring post authors and highest reputation users exhibited high Openness and Neuroticism along with low Agreeableness according to Personality Insights results. These facets had surprisingly low variance among them and we were surprised at the low agreeableness exhibited in both the most popular posts and the highest reputation users.

Although the results warrant analysis by experts in Psychology, some conclusions can be drawn at this time. High

Openness suggests that the authors have an exploratory or curious tendency [8]. They are seeking knowledge, which is to be expected of a platform like Stack Overflow. A good question on Stack Overflow is one that is common to other programmers, but not widely known online. The high Neuroticism score, or tendency towards self-consciousness or nervousness, likely reflects stress that students or professionals might have if they are looking for a solution to their issue with a deadline on the horizon [8]. In addition, the low consistently low agreeableness score points to the nature of Stack Overflow itself and what constitutes a good answer that users are likely to vote up. Agreeableness is ones tendency to identify with others emotionally their willingness collaborate. A good answer on Stack Overflow is expected to be clear and detailed; it will also be confident and authoritative. Agreeableness is a good trait to have in interpersonal communication, but for technical help, a clear-cut answer is preferable.

These results warrant further work, but we still believe they can be useful in industry. Professional communication often favors low openness as they favor a straightforward dialog over flattery and fluff. Ultimately, the ideal communicator will possess the ability to display both high and low openness at the appropriate time.

VII. FUTURE WORK

The Personality Insights API offers many more data points in addition to the Big Five, so there is more analysis that can be done on the Personality Insights data. Other metrics measured by Stack Overflow can also offer more ways to measure the success of a user.

Other sentiment analysis tools can be used to infer user's emotions. IBM's Natural Language Understanding API is an example of this. The minimum word count for a confident response is also smaller, so analysis could potentially be performed on a per-post basis rather than per-user [9].

The most surprising result of this study was the consistently low agreeableness scores of the trials. This seems to be counter-intuitive as one would expect popular answers to be agreeable. This warrants further investigation into the factors that contributed to the low agreeableness scores of the analyzed users.

REFERENCES

[1] B. Lin, F. Zampetti, G. Bavota, M. Di Penta, M. Lanza, and R. Oliveto, "Sentiment analysis for software engineering: How far can we go?" in *2018 IEEE/ACM 40th International Conference on Software Engineering (ICSE)*, May 2018, pp. 94–104.

[2] R. Socher, A. Perelygin, J. Wu, J. Chuang, C. D. Manning, A. Ng, and C. Potts, "Recursive deep models for semantic compositionality over a sentiment treebank," in *Proceedings of the 2013 Conference on Empirical Methods in Natural Language Processing*. Seattle, Washington, USA: Association for Computational Linguistics, Oct. 2013, pp. 1631–1642. [Online]. Available: https://www.aclweb.org/anthology/D13-1170

[3] B. Bazelli, A. Hindle, and E. Stroulia, "On the personality traits of stackoverflow users," in *2013 IEEE International Conference on Software Maintenance*, Sep. 2013, pp. 460–463.

[4] "What is reputation? how do i earn (and lose) it? - help center," 2019. [Online]. Available: https://stackoverflow.com/help/whats-reputation

[5] A. Rastogi and N. Nagappan, "On the personality traits of github contributors," in *2016 IEEE 27th International Symposium on Software Reliability Engineering (ISSRE)*, Oct 2016, pp. 77–86.

[6] F. Calefato, F. Lanubile, M. C. Marasciulo, and N. Novielli, "Mining successful answers in stack overflow," in *Proceedings of the 12th Working Conference on Mining Software Repositories*, ser. MSR '15. Piscataway, NJ, USA: IEEE Press, 2015, pp. 430–433. [Online]. Available: http://dl.acm.org/citation.cfm?id=2820518.2820579

[7] S. Baltes, L. Dumani, C. Treude, and S. Diehl, "Sotorrent: Reconstructing and analyzing the evolution of stack overflow posts," in *Proceedings of the 15th International Conference on Mining Software Repositories*, ser. MSR '18. New York, NY, USA: ACM, 2018, pp. 319–330. [Online]. Available: http://doi.acm.org/10.1145/3196398.3196430

[8] "Personality models," Mar 2019. [Online]. Available: https://cloud.ibm.com/docs/services/personality-insights?topic=personality-insights-models

[9] "Natural language understanding: About," Feb 2019. [Online]. Available: https://cloud.ibm.com/docs/services/natural-language-understanding?topic=natural-language-understanding-about