

Sentiment Analysis in Code Review Comments

Anne Veenendaal, Elliot Daly, Eddie Jones, Zhao Gang, Sumalini Vartak, Rahul S Patwardhan

Abstract— This paper proposed a novel method to predict the emotions and team dynamics by inspecting the code review comments. The code reviews from source control repository were extracted for development activity from the past three years for two open source projects. The comments were tokenized and features were extracted to detect sentiments such as encouragement, annoyance and fear. The emotion polarity of the code review comments were measure in terms of neutral vs rude (harshness). The technique used provided an accuracy of 67% in terms of recognition of harsh comments.

Keywords— Token, sentiment, review, code, polarity, SVM

I. INTRODUCTION

Many community-supported sites contain code review comments by volunteer programmers contributing to the code base. Sometimes the code review comments can sound harsh and may result in misunderstandings. This paper examines automatic sentiment analysis of code reviewer comments and predicts the emotional polarity of the comment as neutral vs rude (or harsh).

II. METHOD

The code review comments of 3 open source software code bases were downloaded. There were a total of 1536 comments in total. Each comment was manually annotated (expert human annotators) as neutral vs harsh. The annotation was done by 3 annotators to avoid inter annotator disagreement. The features defined for the text analysis were frequency of harsh words. The master list of harsh words used as reference was constructed using bag of words method applied on data sets for angry, rude, harsh sentiments. The number of syntax related comments, the number of comments on the code, the number of revisions and rollbacks were also tracked for each code change. The feature vector was used to train an SVM classifier for detecting class associated with rude emotion polarity.

III. CONCLUSIONS

The automatic classification of positive vs negative code review comments showed 67% accuracy. The accuracy degraded by 3% when the analysis was performed on the test data from another code review repository website that was not used during initial training. This was because the bag of words did not contain some of the words found in the test data. Additionally, the language used in some of the code reviews was technical syntax related and contained programming language specific jargon and abbreviations. This introduced a lot of misclassifications. There was also difference in grammar and positioning of the contradiction phrases. Future scope would include evaluation between international reviews so that subtle difference between American English and British English is captured.

REFERENCES

- [1] A.Veenendaal, Elliot Daly, Eddie Jones, Zhao Gang, Sumalini Vartak, Rahul S Patwardhan, Drunken Abnormal Human Gait Detection using Sensors, Computer Science and Emerging Research Journal, vol 1, 2013.
- [2] A.Veenendaal, Elliot Daly, Eddie Jones, Zhao Gang, Sumalini Vartak, Rahul S Patwardhan, Fear Detection with Background Subtraction from RGB-D data, Computer Science and Emerging Research Journal, vol 1, 2013.
- [3] A.Veenendaal, Elliot Daly, Eddie Jones, Zhao Gang, Sumalini Vartak, Rahul S Patwardhan, Code Definition Analysis for Call Graph Generation, Computer Science and Emerging Research Journal, vol 1, 2013.
- [4] A.Veenendaal, Elliot Daly, Eddie Jones, Zhao Gang, Sumalini Vartak, Rahul S Patwardhan, Multi-View Point Drowsiness and Fatigue Detection, Computer Science and Emerging Research Journal, vol 2, 2014.
- [5] A.Veenendaal, Elliot Daly, Eddie Jones, Zhao Gang, Sumalini Vartak, Rahul S Patwardhan, Group Emotion Detection using Edge Detection Mesh Analysis, Computer Science and Emerging Research Journal, vol 2, 2014.
- [6] A.Veenendaal, Elliot Daly, Eddie Jones, Zhao Gang, Sumalini Vartak, Rahul S Patwardhan, Polarity Analysis of Restaurant Review Comment Board, Computer Science and Emerging Research Journal, vol 2, 2014.
- [7] A.Veenendaal, Elliot Daly, Eddie Jones, Zhao Gang, Sumalini Vartak, Rahul S Patwardhan, Sentiment Analysis in Code Review Comments, Computer Science and Emerging Research Journal, vol 3, 2015.
- [8] A.Veenendaal, Elliot Daly, Eddie Jones, Zhao Gang, Sumalini Vartak, Rahul S Patwardhan, Temporal Analysis of News Feed Using Phrase Position, Computer Science and Emerging Research Journal, vol 3, 2015.
- [9] A.Veenendaal, Elliot Daly, Eddie Jones, Zhao Gang, Sumalini Vartak, Rahul S Patwardhan, Decision Rule Driven Human Activity Recognition, Computer Science and Emerging Research Journal, vol 3, 2015.
- [10] A.Veenendaal, Elliot Daly, Eddie Jones, Zhao Gang, Sumalini Vartak, Rahul S Patwardhan, Depression and Sadness Recognition in Closed Spaces, Computer Science and Emerging Research Journal, vol 4, 2016.
- [11] A.Veenendaal, Elliot Daly, Eddie Jones, Zhao Gang, Sumalini Vartak, Rahul S Patwardhan, Dynamic Probabilistic Network Based Human Action Recognition, Computer Science and Emerging Research Journal, vol 4, 2016.
- [12] A.Veenendaal, Elliot Daly, Eddie Jones, Zhao Gang, Sumalini Vartak, Rahul S Patwardhan, Fight and Aggression Recognition using Depth and Motion Data, Computer Science and Emerging Research Journal, vol 4, 2016.
- [13] A.Veenendaal, Elliot Daly, Eddie Jones, Zhao Gang, Sumalini Vartak, Rahul S Patwardhan, Sensor Tracked Points and HMM Based Classifier for Human Action Recognition, Computer Science and Emerging Research Journal, vol 5, 2016.
- [14] A. S. Patwardhan, 2016. "Structured Unit Testable Templated Code for Efficient Code Review Process", PeerJ Computer Science (in review), 2016.
- [15] A. S. Patwardhan, and R. S. Patwardhan, "XML Entity Architecture for Efficient Software Integration", International Journal for Research in Applied Science and Engineering Technology (IJRASET), vol. 4, no. 6, June 2016.
- [16] A. S. Patwardhan and G. M. Knapp, "Affect Intensity Estimation Using Multiple Modalities," Florida Artificial Intelligence Research Society Conference, May, 2014.
- [17] A. S. Patwardhan, R. S. Patwardhan, and S. S. Vartak, "Self-Contained Cross-Cutting Pipeline Software Architecture," International Research Journal of Engineering and Technology (IRJET), vol. 3, no. 5, May, 2016.
- [18] A. S. Patwardhan, "An Architecture for Adaptive Real Time Communication with Embedded Devices," LSU, 2006.

- [19] A. S. Patwardhan and G. M. Knapp, "Multimodal Affect Analysis for Product Feedback Assessment," IIE Annual Conference. Proceedings. Institute of Industrial Engineers-Publisher, 2013.
- [20] A. S. Patwardhan and G. M. Knapp, "Aggressive Action and Anger Detection from Multiple Modalities using Kinect", submitted to ACM Transactions on Intelligent Systems and Technology (ACM TIST) (in review).
- [21] A. S. Patwardhan and G. M. Knapp, "EmoFit: Affect Monitoring System for Sedentary Jobs," preprint, arXiv.org, 2016.
- [22] A. S. Patwardhan, "Embracing Agile methodology during DevOps Developer Internship Program", IEEE Software (in review), 2016.
- [23] A. S. Patwardhan, "Edge Based Grid Super-Imposition for Crowd Emotion Recognition", International Research Journal of Engineering and Technology (IRJET), May. 2010.
- [24] A. S. Patwardhan, "Human Activity Recognition Using Temporal Frame Decision Rule Extraction", International Research Journal of Engineering and Technology (IRJET), May. 2010.
- [25] A. S. Patwardhan, "Low Morale, Depressed and Sad State Recognition in Confined Spaces", International Research Journal of Engineering and Technology (IRJET), May. 2011.
- [26] A. S. Patwardhan, "View Independent Drowsy Behavior and Tiredness Detection", International Research Journal of Engineering and Technology (IRJET), May. 2011.
- [27] A. S. Patwardhan, "Sensor Based Human Gait Recognition for Drunk State", International Research Journal of Engineering and Technology (IRJET), May. 2012.
- [28] A. S. Patwardhan, "Background Removal Using RGB-D data for Fright Recognition", International Research Journal of Engineering and Technology (IRJET), May. 2012.
- [29] A. S. Patwardhan, "Depth and Movement Data Analysis for Fight Detection", International Research Journal of Engineering and Technology (IRJET), May. 2013.
- [30] A. S. Patwardhan, "Human Action Recognition Classification using HMM and Movement Tracking", International Research Journal of Engineering and Technology (IRJET), May. 2013.
- [31] A. S. Patwardhan, "Feedback and Emotion Polarity Extraction from Online Reviewer sites", International Research Journal of Engineering and Technology (IRJET), May. 2014.
- [32] A. S. Patwardhan, "Call Tree Detection Using Source Code Syntax Analysis", International Research Journal of Engineering and Technology (IRJET), May. 2014.
- [33] A. S. Patwardhan, "Walking, Lifting, Standing Activity Recognition using Probabilistic Networks", International Research Journal of Engineering and Technology (IRJET), May. 2015.
- [34] A. S. Patwardhan, "Online News Article Temporal Phrase Extraction for Causal Linking", International Research Journal of Engineering and Technology (IRJET), May. 2015.
- [35] A. S. Patwardhan, "Online Comment Processing for Sentiment Extraction", International Research Journal of Engineering and Technology (IRJET), May. 2016.
- [36] A. S. Patwardhan, "Analysis of Software Delivery Process Shortcomings and Architectural Pitfalls", PeerJ Computer Science (in review), 2016.
- [37] A. S. Patwardhan, "Multimodal Affect Recognition using Kinect", ACM TIST (in review), 2016.
- [38] A. S. Patwardhan, "Augmenting Supervised Emotion Recognition with Rule-Based Decision Model", IEEE TAC (in review), 2016.
- [39] A. S. Patwardhan, Jacob Badeaux, Siavash, G. M. Knapp, "Automated Prediction of Temporal Relations", Technical Report. 2014.