

Polarity Analysis of Restaurant Review Comment Board

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

Abstract— This paper proposed bag of words and dictionary based polarity estimation technique from text. The restaurant review comment board data was obtained from popular restaurant review web sites. The bag of words and dictionary of positive and negative words was constructed. Additionally, the polarity of reviewed emotions was estimated using a bag of words and dictionary representing positive and negative emotions. The process was evaluated using the test data, by classifying the tokenized test text into positive and negative reviews and positive and negative emotions and then was compared with manual annotation. The estimation results indicated 81% accuracy for positive reviews and 83% accuracy for negative reviews.

Keywords— Review polarity, emotion recognition, positive and negative reviews, positive and negative emotions, bag of words, dictionary, semantic analysis, sentiment analysis.

I. INTRODUCTION

There has been meteoric rise in the number of blogs, reviews, comments, message boards, social network portals and web sites in general on the internet. It is important to analyze this text data for several useful applications such as determining the success of a business, customer trends, marketing, advertisement campaigns and spam filtering. This research focuses on two aspects of sentiment analysis on a restaurant review website. First aim is to estimate the polarity of the comments (positive review vs negative review). The second aim is to estimate the polarity of emotions of the reviewer between negative emotions (angry, frustrated, disappointed, dissatisfied and Positive emotions (happy and content).

II. METHOD

Data Collection: We collected 788 comments from various restaurant review sites. A panel of 2 annotators was created to manually annotate each comment as positive and negative review. Additionally, each comment was also annotated with positive and negative emotion. In case of conflicting annotation, a third annotator was used to resolve the conflict.

Feature Definition: A bag of words representing positive review was created. For instance, the collection of positive review words was {great, delicious, fantastic, fresh, best, awesome, tasty, good, friendly, healthy, nice}. Similarly, a collection of words representing negative reviews was created {horrible, terrible, bad, smell, awful, nasty, bad, slow, ridiculous, sloppy, degrade, sucks, yucky, disgusting, crap, mean, rude, dirty}. Each review was then analyzed for the presence of these words and the frequency of the words was recorded. The second feature was constructed by counting the presence of contradictory phrases such as {even though, but, although, on the contrary, nevertheless, even then, even if, in spite of}. The third feature was constructed by counting the frequency of distance between the words representing the positive and negative words within the same review. The same method was used to create feature vectors for emotion polarity recognition.

Training: The feature vector was used to train an SVM classifier using radial basis function as the kernel. The slack variable was optimized and set to 0.2. The data was divided into 60% training and 40% test data. The classifier was trained using 2-fold cross validation. The classifier was also evaluated against a test data set from another restaurant review website that was not used in the initial training. The same method was applied to train the classifier for the emotion analysis.

III. RESULTS

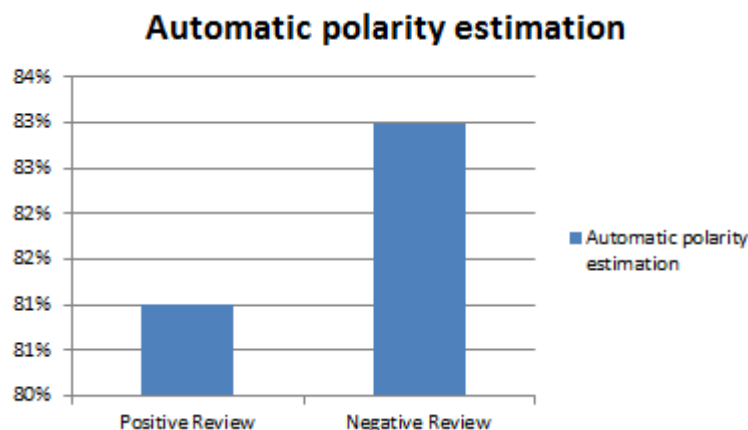


Fig. 1. Review polarity recognition accuracy

The positive review recognition accuracy was lower compared to the negative review recognition accuracy. This was because the number of negative words was higher and had better discriminatory power compared to the collection of positive review words.

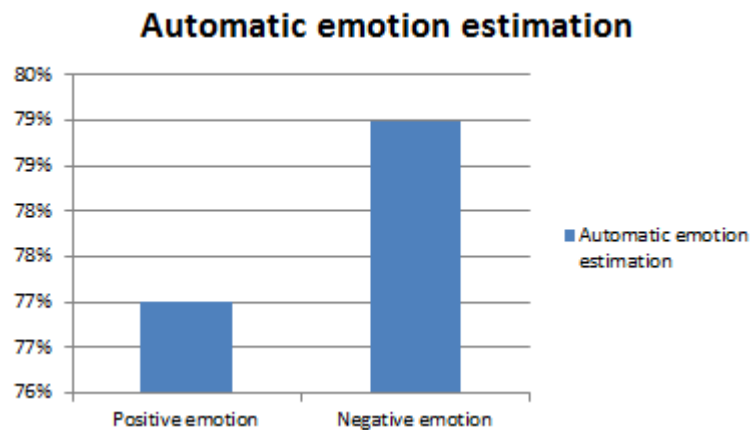


Fig. 2. Emotion polarity recognition accuracy

The negative emotion recognition accuracy was 79% and positive emotion recognition accuracy was 77%.

IV. CONCLUSIONS

The automatic classification of positive vs negative restaurant review comments showed more than 80% accuracy. The accuracy degraded by 2% when the analysis was performed on the test data for polarity of positive and negative emotions. This was because the bag of words did not contain some of the words found in the test data. In addition, the language used in some of the restaurant reviews was colloquial and abbreviations. This introduced many 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 US English and UK English is captured.

REFERENCES

- [1] Fan, R.-E., Chang, K.-W., Hsieh, C.-J., Wang, X.-R., & Lin, C.-J. (2008, June). Liblinear: A library for large linear classification. *J. Mach. Learn. Res.*, 9, 1871–1874.
- [2] Golder, S. A., & Macy, M. W. (2011). Diurnal and seasonal mood vary with work, sleep, and daylength across diverse cultures. *Science*, 333(6051), 1878–1881.
- [3] Huang, E. H., Socher, R., Manning, C. D., & Ng, A. Y. (2012). Improving Word Representations via Global Context and Multiple Word Prototypes. In Annual meeting of the association for computational linguistics (acl).
- [4] Klein, D., & Manning, C. (2003). Accurate unlexicalized parsing. In Proceedings of the 41st annual meeting on association for computational linguistics-volume 1 (pp. 423–430).
- [5] Pang, B., & Lee, L. (2005). Seeing stars: Exploiting class relationships for sentiment categorization with respect to rating scales. In Annual meeting-association for computational linguistics (Vol. 43, p. 115).
- [6] Wang, S., & Manning, C. (2012). Baselines and Bigrams: Simple, Good Sentiment and Topic Classification. In Proceedings of the 50th Annual Meeting of the Association for Computational Linguistics.
- [7] C. Zirn, M. Niepert, H. Stuckenschmidt, M. Strube, Fine-Grained Sentiment Analysis with Structural Features, in: 5th International Joint Conference on Natural Language Processing (IJCNLP 2011), Asian Federation of Natural Language Processing, 2011, pp. 336–344.
- [8] S. Somasundaran, G. Namata, J. Wiebe, L. Getoor, Supervised and Unsupervised Methods in Employing Discourse Relations for Improving Opinion Polarity Classification, in: 2009 Conference on Empirical Methods in Natural Language Processing (EMNLP 2009), Association for Computational Linguistics, 2009, pp. 170–179.
- [9] B. Chardon, F. Benamara, Y. Mathieu, V. Popescu, N. Asher, Measuring the Effect of Discourse Structure on Sentiment Analysis, in: 14th International Conference on Intelligent Text Processing and Computational Linguistics (CICling 2013), volume 7817 of Lecture Notes in Computer Science, Springer, 2013, pp. 25–37.
- [10] W. Mann, S. Thompson, Rhetorical Structure Theory: Toward a Functional Theory of Text Organization, *Text* 8 (1988) 243–281.
- [11] J. Chenlo, D. Losada, Effective and Efficient Polarity Estimation in Blogs Based on Sentence-Level Evidence, in: 20th ACM Conference on Information and Knowledge Management (CIKM 2011), Association for Computing Machinery, 2011, pp. 365–374.
- [12] S. Gerani, M. Carman, F. Crestani, Proximity-Based Opinion Retrieval, in: 33rd International ACM SIGIR Conference on Research and Development in Information Retrieval (SIGIR 2010), Association for Computing Machinery, 2010, pp. 403–410.
- [13] R. Santos, B. He, C. Macdonald, I. Ounis, Integrating Proximity to Subjective Sentences for Blog Opinion Retrieval, in: 31st European Conference on Information Retrieval (ECIR 2009), Springer, 2009, pp. 325–336.
- [14] B. He, C. Macdonald, J. He, I. Ounis, An Effective Statistical Approach to Blog Post Opinion Retrieval, in: 17th ACM Conference on Information and Knowledge Management (CIKM 2008), Association for Computing Machinery, 2008, pp. 1063–1072.
- [15] B. Pang, L. Lee, A Sentimental Education: Sentiment Analysis using Subjectivity Summarization based on Minimum Cuts, in: 42nd Annual Meeting of the Association for Computational Linguistics (ACL 2004), Association for Computational Linguistics, 2004, pp. 271–280.
- [16] Anne 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.
- [17] Anne 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.

- [18] Anne 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.
- [19] Anne 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.
- [20] Anne 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.
- [21] Anne 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.
- [22] Anne 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.
- [23] Anne 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.
- [24] Anne 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.
- [25] Anne 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.
- [26] Anne 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.
- [27] Anne 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.
- [28] Anne 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.
- [29] A. S. Patwardhan, 2016. "Structured Unit Testable Templated Code for Efficient Code Review Process", *PeerJ Computer Science* (in review), 2016.
- [30] 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.
- [31] A. S. Patwardhan and G. M. Knapp, "Affect Intensity Estimation Using Multiple Modalities," *Florida Artificial Intelligence Research Society Conference*, May. 2014.
- [32] 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.
- [33] A. S. Patwardhan, "An Architecture for Adaptive Real Time Communication with Embedded Devices," *LSU*, 2006.
- [34] A. S. Patwardhan and G. M. Knapp, "Multimodal Affect Analysis for Product Feedback Assessment," *IIE Annual Conference. Proceedings. Institute of Industrial Engineers-Publisher*, 2013.
- [35] 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).
- [36] A. S. Patwardhan and G. M. Knapp, "EmoFit: Affect Monitoring System for Sedentary Jobs," preprint, *arXiv.org*, 2016.
- [37] A. S. Patwardhan, J. Kidd, T. Urena and A. Rajagopalan, "Embracing Agile methodology during DevOps Developer Internship Program", *IEEE Software* (in review), 2016.
- [38] A. S. Patwardhan, "Analysis of Software Delivery Process Shortcomings and Architectural Pitfalls", *PeerJ Computer Science* (in review), 2016.
- [39] A. S. Patwardhan, "Multimodal Affect Recognition using Kinect", *ACM TIST* (in review), 2016.
- [40] A. S. Patwardhan, "Augmenting Supervised Emotion Recognition with Rule-Based Decision Model", *IEEE TAC* (in review), 2016.
- [41] A. S. Patwardhan, Jacob Badeaux, Siavash, G. M. Knapp, "Automated Prediction of Temporal Relations", *Technical Report*. 2014.
- [42] R. Santos, C. Macdonald, R. McCreadie, I. Ounis, I. Soboroff, *Information Retrieval on the Blogosphere, Foundations and Trends in Information Retrieval* 6 (2012) 1-125.
- [43] I. Ounis, C. Macdonald, I. Soboroff, Overview of the TREC 2008 Blog Track, in: *17th Text Retrieval Conference (TREC 2008)*, National Institute of Standards and Technology, 2008.
- [44] A. S. Patwardhan, "Edge Based Grid Super-Imposition for Crowd Emotion Recognition", *International Research Journal of Engineering and Technology (IRJET)*, May. 2010.
- [45] A. S. Patwardhan, "Human Activity Recognition Using Temporal Frame Decision Rule Extraction", *International Research Journal of Engineering and Technology (IRJET)*, May. 2010.
- [46] A. S. Patwardhan, "Low Morale, Depressed and Sad State Recognition in Confined Spaces", *International Research Journal of Engineering and Technology (IRJET)*, May. 2011.
- [47] A. S. Patwardhan, "View Independent Drowsy Behavior and Tiredness Detection", *International Research Journal of Engineering and Technology (IRJET)*, May. 2011.
- [48] A. S. Patwardhan, "Sensor Based Human Gait Recognition for Drunk State", *International Research Journal of Engineering and Technology (IRJET)*, May. 2012.
- [49] A. S. Patwardhan, "Background Removal Using RGB-D data for Fright Recognition", *International Research Journal of Engineering and Technology (IRJET)*, May. 2012.
- [50] A. S. Patwardhan, "Depth and Movement Data Analysis for Fight Detection", *International Research Journal of Engineering and Technology (IRJET)*, May. 2013.
- [51] A. S. Patwardhan, "Human Action Recognition Classification using HMM and Movement Tracking", *International Research Journal of Engineering and Technology (IRJET)*, May. 2013.
- [52] A. S. Patwardhan, "Feedback and Emotion Polarity Extraction from Online Reviewer sites", *International Research Journal of Engineering and Technology (IRJET)*, May. 2014.
- [53] A. S. Patwardhan, "Call Tree Detection Using Source Code Syntax Analysis", *International Research Journal of Engineering and Technology (IRJET)*, May. 2014.
- [54] A. S. Patwardhan, "Walking, Lifting, Standing Activity Recognition using Probabilistic Networks", *International Research Journal of Engineering and Technology (IRJET)*, May. 2015.
- [55] A. S. Patwardhan, "Online News Article Temporal Phrase Extraction for Causal Linking", *International Research Journal of Engineering and Technology (IRJET)*, May. 2015.
- [56] A. S. Patwardhan, "Online Comment Processing for Sentiment Extraction", *International Research Journal of Engineering and Technology (IRJET)*, May. 2016.