Abstract

In this era, the data of the user on social media is generating in every millisecond. The importance of data can be noted or observed as these are the reviews, emotions, and opinions of the human being in the form of text. The customer-generated data can be related to events, food, products, etc. This information is a “key to success” for those who do business or are in government and other individuals. As the data is in the form of bulk so, it is evident that to analyze a content that is generated by a user must be complicated to manage as well as it must be time-consuming. For this, we need an intelligent system that helps us to figure out whether the content is positive, negative or neutral in the form of categories. This smart system commonly named sentiment analysis (SA), opinion mining (OM), subjectivity mining, etc. Opinion mining is the systematized mining approach. Through this, we can classify, thoughts and emotions from the text, speech, and databank sources through Natural Language Processing (NLP). The actual purpose of writing this paper is to determine the idea of human emotions with the help of BERT model, where we took a dataset of IMDB movie reviews, which
are generated by a users’ data. Our experimental methodology is adequate and robust, which in turn describes the quality of sentiment analysis.

References

1. Abbasi, A., France, S., Zhang, Z. and Chen, H., 2010. Selecting attributes for sentiment classification using feature relation networks. IEEE Transactions on Knowledge and Data Engineering, 23(3), pp.447-462.
2. Tetlock, P.C., Saar-Tsechansky, M. and Macskassy, S., 2008. More than words: Quantifying language to measure firms’ fundamentals. The Journal of Finance, 63(3), pp.1437-1467.
3. Wilson, T., Wiebe, J. and Hoffmann, P., 2005. Recognizing contextual polarity in phrase-level sentiment analysis. In Proceedings of Human Language Technology Conference and Conference on Empirical Methods in Natural Language Processing.
4. Yu, H. and Hatzivassiloglou, V., 2003, July. Towards answering opinion questions: Separating facts from opinions and identifying the polarity of opinion sentences. In Proceedings of the 2003 conference on Empirical methods in natural language processing (pp. 129-136). Association for Computational Linguistics.
5. Tan, L.K.W., Na, J.C., Theng, Y.L. and Chang, K., 2011, October. Sentence-level sentiment polarity classification using a linguistic approach. In International Conference on Asian Digital Libraries (pp. 77-87). Springer, Berlin, Heidelberg.
6. Das, S.R., 2011. News analytics: Framework, techniques and metrics. In The Handbook of News Analytics in Finance (Vol. 2). John Wiley & Sons Chichester.
7. Pang, B., Lee, L. and Vaithyanathan, S., 2002, July. Thumbs up?: sentiment classification using machine learning techniques. In Proceedings of the ACL-02 conference on Empirical methods in natural language processing-Volume 10 (pp. 79-86). Association for Computational Linguistics.
8. Melville, P., Gryc, W. and Lawrence, R.D., 2009, June. Sentiment analysis of blogs by combining lexical knowledge with text classification. In Proceedings of the 15th ACM SIGKDD international conference on Knowledge discovery and data mining (pp. 1275-1284). ACM.
9. an, C., Lee, L., Tang, J., Jiang, L., Zhou, M. and Li, P., 2011, August. User-level sentiment analysis incorporating social networks. In Proceedings of the 17th ACM SIGKDD international conference on Knowledge discovery and data mining (pp. 1397-1405). ACM.
10. Mishne, G. and Glance, N.S., 2006, March. Predicting movie sales from blogger sentiment. In AAAI spring symposium: computational approaches to analyzing weblogs (pp. 155-158).
11. Maas, A.L., Daly, R.E., Pham, P.T., Huang, D., Ng, A.Y. and Potts, C., 2011, June. Learning word vectors for sentiment analysis. In Proceedings of the 49th annual meeting of the association for computational linguistics: Human language technologies-volume 1 (pp. 142-150). Association for Computational Linguistics.
12. Denecke, K., 2008, April. Using sentiwordnet for multilingual sentiment analysis. In 2008 IEEE 24th International Conference on Data Engineering Workshop (pp. 507-512). IEEE.
13. Radford, A., Jozefowicz, R. and Sutskever, I., 2017. Learning to generate reviews and discovering sentiment. arXiv preprint arXiv:1704.01444.
14. Wei, J.W. and Zou, K., 2019. Eda: Easy data augmentation techniques for boosting performance on text classification tasks. arXiv preprint arXiv:1901.11196.
15. Radford, A., Narasimhan, K., Salimans, T. and Sutskever, I., 2018. Improving language understanding with unsupervised learning. Technical report, OpenAI.
16. Howard, J. and Ruder, S., 2018. Universal language model fine-tuning for text classification. arXiv preprint arXiv:1801.06146.
17. Devlin, J., Chang, M.W., Lee, K. and Toutanova, K., 2018. Bert: Pre-training of deep bidirectional transformers for language understanding. arXiv preprint arXiv:1810.04805.
18. Ramos, J., 2003, December. Using tf-idf to determine word relevance in document queries. In Proceedings of the first instructional conference on machine learning (Vol. 242, pp. 133-142).
19. Adelia, R., Suyanto, S. and Wisesty, U.N., 2019. Indonesian Abstractive Text Summarization Using Bidirectional Gated Recurrent Unit. Procedia Computer Science, 157, pp.581-588.
20. Yazdavar, A.H., Ebrahimi, M. and Salim, N., 2017. Fuzzy based implicit sentiment analysis on quantitative sentences. arXiv preprint arXiv:1701.00798.
21. Dolnicar, S., Grün, B. and Yanamandram, V., 2013. Dynamic, interactive survey questions can increase survey data quality. Journal of Travel & Tourism Marketing, 30(7), pp.690-699.
22. Menner, T., Höpken, W., Fuchs, M. and Lexhagen, M., 2016. Topic detection: identifying relevant topics in tourism reviews. In Information and Communication Technologies in Tourism 2016 (pp. 411-423). Springer, Cham.
23. Schmunk, S., Höpken, W., Fuchs, M. and Lexhagen, M., 2013. Sentiment analysis: Extracting decision-relevant knowledge from UGC. In Information and Communication Technologies in Tourism 2014 (pp. 253-265). Springer, Cham.
24. Alaei, A.R., Becken, S. and Stantic, B., 2019. Sentiment analysis in tourism: capitalizing on big data. Journal of Travel Research, 58(2), pp.175-191.
25. García-Pablos, A., Duca, A.L., Cuadros, M., Linaza, M.T. and Marchetti, A., 2016. Correlating languages and sentiment analysis on the basis of text-based reviews. In Information and Communication Technologies in Tourism 2016 (pp. 565-577). Springer, Cham.
26. Wang, P., Hu, J., Zeng, H.J. and Chen, Z., 2009. Using Wikipedia knowledge to improve text classification. Knowledge and Information Systems, 19(3), pp.265-281.
27. Hu, X., Zhang, X., Lu, C., Park, E.K. and Zhou, X., 2009, June. Exploiting Wikipedia as external knowledge for document clustering. In Proceedings of the 15th ACM SIGKDD international conference on Knowledge discovery and data mining (pp. 389-396). ACM.
28. Boubacar, A. and Niu, Z., 2014. Conceptual clustering. In Future Information Technology (pp. 1-8). Springer, Berlin, Heidelberg.

Index Terms

Computer Science Information Sciences

Keywords
Using BERT for Checking the Polarity of Movie Reviews

Text classification, sentiment analysis, natural language processing, Bidirectional Encoder Representations from Transformers.