

IMPROVED MACHINE LEARNING APPROACH FOR SENTIMENT POLARITY ESTIMATION FOR OPINION MINING

Hettikankanama H.K.S.K.^{1*}, Vasanthapriyan S.¹ and Rathnayake R.M.K.T.²

¹Department of Computing and Information Systems, Faculty of Applied Sciences, Sabaragamuwa University of Sri Lanka, Sri Lanka

²Department of Physical Sciences and Technology, Faculty of Applied Sciences, Sabaragamuwa University of Sri Lanka, Sri Lanka

*keshalasanduni@gmail.com

Opinions of others may be essential in making decisions or selecting from variety of alternatives. Review of customer feedback helps to improve sales and eventually benefits the company. Goal of this study is to compare performance measurements of several machine learning algorithms and propose a novel product ranking model considering user reviews which enable multi-user recommendation. Objectives of this study are to identify different types of methods used for sentiment analysis and sentiment polarity estimation, propose more effective algorithm for estimating sentiment polarity of review text and facilitating customers by providing recommendations based on user given textual reviews. In methodology, first, relevant data was gathered and they were pre-processed to strain the valuable words to estimate polarity. After the sentiment analysis, some supervised learning methods were applied on the data set and their performance measurement results are recorded. In the next step results are compared and best accurate algorithm was proposed for opinion mining. Next step is to predict sentiment polarity values using proposed algorithm and finally a system was implemented to rank products considering opinions given for them. As results K-Nearest-Neighbours algorithm was proposed as best accurate algorithm as it had 92% of accuracy. It had highest precision, recall, cross validation precision, F1-measure and also least Mean error. Other algorithms that used are decision tree, random forest, Ada-boost and ensemble algorithm. A system which ranks products in online shopping sites by considering opinions given by users is resulted. Conclusion highlights that all three objectives are fulfilled by literature review where current available models of opinion mining taken to consider, propose k neighbours algorithm as the best accurate classifier finally and final resulted system which ranked products in shopping sites according to average sentiment polarity respectively. Implementing a system to identify opinions in Sinhala, Singlish or other language using same methodology, consider more variables to rank products like review age and user credibility and combining this system with current available recommendation mechanism are recommended as future work.

Keywords: Machine learning, Opinion mining, Product recommendation, Sentiment analysis