International Journal of Emerging Technology in Computer Science & Electronics (IJETCSE) ISSN: 0976-1353 Volume 24 Issue 5 – APRIL 2017.

SENTIMENT EXTRACTION ON PRODUCTS RATINGS USING MATRIX FACTORIZATION AND COLLABORATIVE FILTERING

Asvini.P[1], Zayaraz.G[2], Swetha Sri.P[3]

[1]PG Student, [2]Professor, Department of Computer Science Engineering, Pondicherry Engineering College,Puducherry,India

[3]Assistant Professor, Department of Computer Science Engineering, CK College of Engineering & Technology,Cuddalore,India

Abstract: The advancements made in the social networks enable a phenomenal growth of information systems. The social data are created in the form of comments, blog posts and tweets. These kinds of social data assist to build the relationship between producers and consumers. Tracking the pulse of the social media contain, enables companies to gain feedback and insight in how to improve the market products better. Henceforth, new opportunities are developed and customer's satisfaction is predicted. In this paper, we have proposed an enhanced rating based prediction model which derive relevant knowledge from the pool of twitter data. Sentimental score is analyzed for each received data and make an interpersonal relationship between those data in both collaborative filtering and matrix factorization. Experimental results show the efficacy of our proposed work. Performance simulation is done in terms of precision, accuracy, RMSE and MAE values.

Keywords: Social networks, blog posts, tweets, customer satisfaction, knowledge derivation and twitter data.

I. INTRODUCTION

Due to its high popularity, Weblogs provide a wealth of information that can be very helpful in assessing the general public's sentiments and opinions [1]. It is therefore imperative to analyze them and distil useful knowledge that could be of economic values to vendors and other interested parties. Whereas marketing plays an important role for the newly released products, public opinion about the products might be crucial to determine their success in the long run. Analyzing the large volume of online reviews available would produce useful actionable knowledge that could be of economic values to vendors and other interested parties. Prediction of product performance is an extremely domain driven task, for which a deep understanding of a variety of aspects involved are important. Previous studies have confirmed that the sentiments expressed in the online reviews are strongly correlated with the sales performance of products [2].

From the recent studies regarding writing the reviews, online opinions, online comments, discussion forums, the most stakes is taken by film industry, which includes videos, songs, movies, television programs, etc [3]. It is very easy to get reviews about movies after or before its release from websites dedicated for movies and it was therefore decided to take up movie as a product in this study. If the prediction is focused on electronic goods, then it is required to consider different companies/brands, but here for movies it is possible to get exact amount of the box office revenue information. There has been previous research and comparing the results with previous results was also another motivation. Various economic functions have been utilized to examine the relationship between opinions discovered from product reviews and revenue growth, stock trading volume change, as well as the bidding price variation in commercial Websites, such as eBay [4].

Social media is increasingly being used by a large section of population in India. The content generated on social media websites have been largely untapped by businesses for gaining customer insights and predicting real outcomes. Microblogging services in recent times have been a popular communication tool among internet users [5]. It generates millions of daily messages for popular websites. Microblogging is online word of mouth branding like Twitter, is now serving as electronic word of mouth (eWOM), forming a eWOM branding which is based on social networking and trust. Twitter has been swamped with active users during the last years and much attention has been given in analyzing the social behavior and opinions of users. The wide-spread popularity of online social networks and the resulting availability of data have enabled the investigation of new research questions, such as the analysis and estimation of public opinion on various subjects.

The rest of the paper is organized as follows: Section II describes the related work; Section III presents the proposed work; Section IV presents the experimental analysis and atlast, concludes in Section V.

II. RELATED WORK

The author in [6] stated that customer comments articulated via the Internet are available to a large number of other customer's, and therefore can be expected to have a significant impact on the success of goods and services. This on consumer buying and communication behavior are tested in a large-scale study. The results illustrated that the consumers read online articulations mainly to save decision-making time and make better buying decisions. Structural equation modelings have shown that their motives for retrieving online articulations strongly influence their behavior.

The author in [7] studied that both a movie's box office revenue and WOM valence significantly influence WOM volume. WOM volume in turn leads to higher retrieve other customer's online articulations from web based consumer opinion platforms. This positive feedback mechanism highlighted the importance of WOM in generating and sustaining retail revenue. The author in [8] hypothesized that buyers suspect that many reviewers are authors or other biased parties. They found marginal (negative) impact of 1-star reviews is greater than the (positive) impact of 5-star reviews. The results suggested that new forms of customer communication on the Internet have an important impact on customer behavior.

The author in [9] evaluated the impact of eWOM attributes and factors on e-commerce sales using real-world data from a multiproduct retail e-commerce firm. The researchers validated a conceptual model of eWOM and its impact on sales. Their research showed that the interactions among eWOM postings, product category, volume of postings, and product were statistically significant in explaining changes in product sales. The author in [10] examined the extent to which people were willing to accept and adopt online consumer reviews and the factors that encouraged adoption. The research findings reported comprehensiveness and relevance to be the most effective components of online postings.

The author in [11] reported negative eWOM had a greater effect than positive eWOM. Related to eWOM communication is sentiment analysis or opinion mining, the author in [12] stated that opinion mining required the retrieval of relevant documents and then ranking those documents according to expressed opinions about a query topic. Certainly, though, one could be interested in aspects other than a ranked list. The author in [13] developed an application for analyzing and comparing consumer opinions for a set of competing products. The author in [14] leveraged the PageRank algorithm to measure movies based on user reviews. Their results compared favourably with the actual box office rankings. The author in [15] presented a survey of the various techniques for opinion mining.

III. PROPOSED WORK

The enhanced rating based prediction model is explained using Quality Data miner techniques. The following are the modules which are described:

a) Product managing systems:

This is the first module that manages the products based on its domain. It also ensures add/ removal of any products.

b) Review and command management systems:

This phase represents the command given for each product. It is done by using Opinion word feature mining which comprised of unigram and bigram formation and classification of opinion word and Opinion word polarity classification which comprised of polarity classification and orientation identification of the sentence.

c) POS tagging:

POS tagging is also called as grammatical tagging or word. Category disambiguation is the process of making different context to its definition. It also assists to make different adjacent and related words in phrase like nouns, verbs, adverbs etc.

d) Aspect Ranking:

Aspect ranking framework consists of three components, aspect identification, sentiment classification based on aspects and probabilistic aspect ranking. First, the aspects of the reviews are calculated for each sentiment classifier. And similarly, probabilistic based aspect ranking is estimated for each opinion. From these results, sentimental words are classified into positive or negative words.

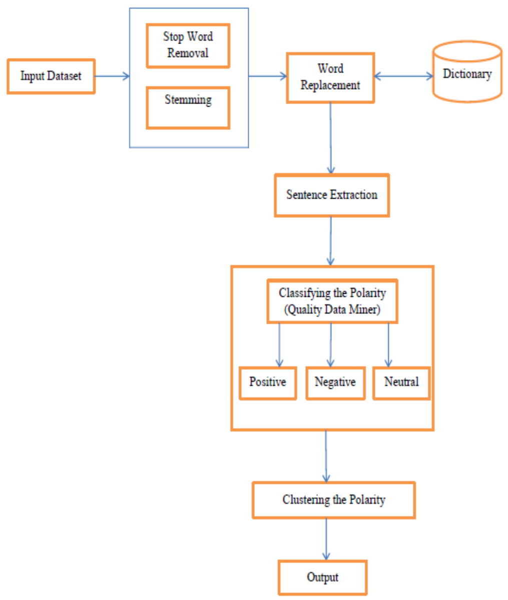

Fig.1. Proposed architecture using QD miner

IV. EXPERIMENTAL RESULTS

This section depicts the experimental analysis of the proposed rating based prediction model. The performance analysis is done in terms of

a) *Precision*

Precision is the fraction of retrieved instances that are relevant and it is given as:

Precision = (Positive precisions +Negative precisions)/ (Total reviews) * 100

b) *Accuracy*

Accuracy is the estimation of the proposition of correctly identified keywords. It is given as:

Accuracy= (TP+TN)/ (TP+TN+FP+FN)

Where True Positive (TP) - Keyword correctly identified as a Keyword

True Negative (TN) – Non- Keyword correctly identified as non-keyword

False Positive (FP) – Non-Keyword incorrectly identified as a keyword

False Negative (FN) – Keyword incorrectly identified as non-Keyword

c) *Root Mean Square Error (RMSE)*

Root Mean Square Error (RMSE) is the difference between predicted keywords and the observed keywords. It is given as:

$$RMSE = \sqrt{\sum_{i\in\Re_{test}} \left(\hat{R}_{u,i} - R_{u,i}\right)^2 / |\Re_{test}|}$$

d) *Mean Absolute Error (MAE)*

Mean Absolute Error (MAE) is the quantity estimation of obtained from its outcomes. It is given as:

$$MAE = \sum_{i\in\Re_{test}} \left|\hat{R}_{u,i} - R_{u,i}\right| / |\Re_{test}|$$

Where Ru, is the real rating value of user u to item i, Ru, is the predicted rating value. $|\Re test|$ denotes the number of user-item pairs in the test set.

Techniques	Positive reviews	Negative reviews	Neutral reviews
Collaborating filtering	120	43	38
Matrix factorization	80	20	10

Table 4.1 Overall input data used in techniques

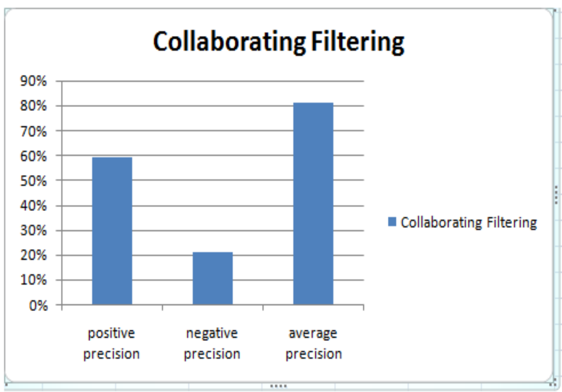

Fig. 2. Precision achieved using collaborative filtering

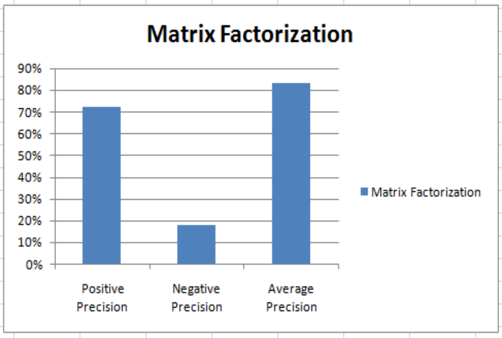

Fig.3. Precision achieved for matrix factorization

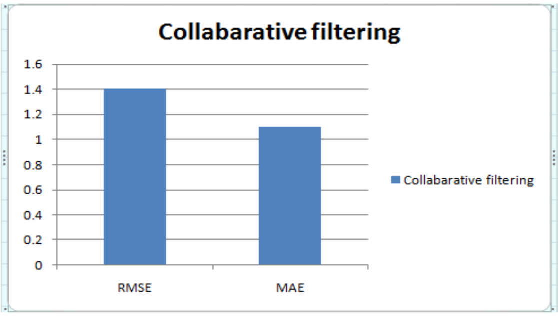

Fig.4. RMSE and MAE value achieved for collaborative filtering

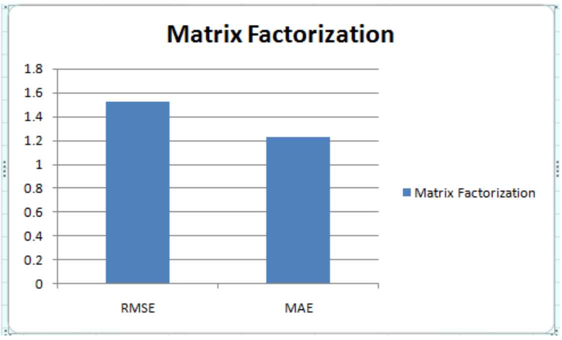

Fig.5. RMSE and MAE value achieved for matrix factorization

V. CONCLUSION

Recently, the growth of Social media enables different organizations to understand the behaviors of the customers in order to enhance the markets. Thus, the sentiment analysis concepts have been introduced to predict the behaviors of the customers toward the products. In this paper, we have proposed rating based prediction model that deals using social media data. Sentiment analysis on Online review are done by forming dictionary which shows that it is easier to build dictionary on phrases but complex in case of Twitter as tweets. So, hidden relationship between different keywords and a dictionary of the words on the basis of different categories of comments & tweets is formed. Future work include to determine their features for the movie in detail i.e. make polarity check on different features such as actors, directors, scripts, music etc. and make the dictionary for them.

REFERENCES

[1] Xiaojiang Lei et al, "Rating Prediction based on Social Sentiment from Textual Reviews", IEEE transactions on Multimedia, 18(9), 2016.

[2] [2] D. Tang, Q. Bing, T. Liu, "Learning semantic representations of users and products for document level sentiment classification," in Proc. 53th Annual Meeting of the Association for Computational Linguistics and the 7 th International Joint Conference on Natural Language Processing, Beijing, China, July 26-31, 2015, pp. 1014–1023.

[3] Y. Zhang, G. Lai, M. Zhang, Y. Zhang, Y. Liu, S. Ma, "Explicit factor models for explainable recommendation based on phrase-level sentiment analysis," in proceedings of the 37th international ACM SIGIR conference on Research & development in information retrieval, 2014.

[4] W. Zhang, G. Ding, L. Chen, C. Li , and C. Zhang, " Generating virtual ratings from Chinese reviews to augment online recommendations," ACM TIST, vol.4, no.1. 2013, pp. 1-17.

[5] X. Lei, and X. Qian, "Rating prediction via exploring service reputation," 2015 IEEE 17th International Workshop on Multimedia.

[6] X. Yang, H. Steck, and Y. Liu, "Circle-based recommendation in online social networks, " in Proc. 18th ACM SIGKDD Int. Conf. KDD, New York, NY, USA, Aug. 2012, pp. 1267–1275.

[7] M. Jiang, P. Cui, R. Liu, Q. Yang, F. Wang, W. Zhu, and S. Yang, "Social contextual recommendation," in proc. 21st ACM Int. CIKM, 2012, pp. 45-54.

[8] Z. Fu, X. Sun, Q. Liu, et al., "Achieving Efficient Cloud Search Services: Multi-Keyword Ranked Search over Encrypted Cloud Data Supporting Parallel Computing," IEICE Transactions on Communications, 2015, 98(1):190-200.

[9] Y. Ren, J. Shen, J. Wang, J. Han, and S. Lee, "Mutual Verifiable Provable Data Auditing in Public

Cloud Storage," Journal of Internet Technology, vol. 16, no. 2, 2015, pp. 317-323.

[10] W. Luo, F. Zhuang, X. Cheng, Q. H, Z. Shi, "Ratable aspects over sentiments: predicting ratings for unrated reviews," IEEE International Conference on Data Mining (ICDM), 2014, pp. 380-389.

[11] T. Nakagawa, K. Inui, and S. Kurohashi, "Dependency tree-based sentiment classification using CRFs with Hidden Variables," NAACL, 2010, pp.786-794.

[12] Xiaojiang Lei, Xueming Qian, Member, IEEE, and Guoshuai Zhao, "Rating Prediction based on Social Sentiment from Textual Reviews," IEEE Transactions On Multimedia, MANUSCRIPT ID: MM-006446

[13] Mrs. R.Nithya, Dr. D.Maheshwari. 2014 Sentiment Analysis on Unstructured Review, International Conference on Intelligent Computing Application, IEEE.

[14] Ms. K. Mouthami, Ms. K.Nirmala Devi, Dr. V.Murali Bhaskaran. 2010 Sentiment Analysis and Classification based on Textual Reviews, Dept of CSE, Tamil Nadu, IEEE.

[15] Nargiza Bekmamedova, Graeme Shanks 2013 Social Media Analytics and Business Value: A Theoretical Framework and Case Study, Department of Computing and Information Systems, University of Melbourne.