Product aspect ranking and its application

Thirunagari Lakshmi Narasimha Vedavyas
vedavyasthirunagari@gmail.com
SRM Institute of Science and Technology, Chennai, Tamil Nadu

Vanam Udayram
udayramvanam@gmail.com
SRM Institute of Science and Technology, Chennai, Tamil Nadu

Kateeb Shaik Sharukh Mastan
shaik.sharukhmastan@gmail.com
SRM Institute of Science and Technology, Chennai, Tamil Nadu

V. Ethirajulu
Ethirajulu.v@vdp.srmuniv.ac.in
SRM Institute of Science and Technology, Chennai, Tamil Nadu

ABSTRACT

E-commerce is a transaction of buying or selling something online. E-commerce allows the customers to overcome the barriers of geographical and also allows them to purchase anytime and from anywhere and also consumers having the privilege to review positively or negatively on any product over the online. The consumer review is very important in knowing the product's aspect and feature and it also very useful for both other consumers and firm. So in the way of finding the product aspect ranking we have proposed the methodologies in which it extracts the reviews and preprocessing, finding the aspect identification of the product, classifying the positive, negative and neutral reviews of the product by the sentiment classifier and also proposing the ranking algorithm used for the product ranking. In the data preprocessing there are methods available in which it initially differentiates the meaning and meaningless words and also it removes the postfix from each word and then tokenize each sentence by removing the emotion icons and also space. In aspect identification, we will identify the aspect from the numerous review which is given by the consumer whether it is positive or negative and on its basis of high or low score we will give a ranking. The main aims of sentiment classifier are to classify the review. The concurrent consideration of the aspect frequency and the pressure of consumer's opinion given to each aspect on their overall opinions in the products aspects ranking and in its application.

Keywords—Product aspects, Aspect ranking, Aspect identification

1. INTRODUCTION

The world is digital today. We do a lot of activities online including buying stuff we need. E-commerce has grown into a major industry throughout the world. Unlike the mouth-to-mouth publicity and reviews used for products bought offline, online websites and platforms depend on reviews to promote the products and to market their quality.

Online reviews have been proven to increase the likelihood of buying a product. Reviews, today, are not just describing the positive or negative aspect of the entire product but contact details of the product including the most favorable and least favorable minute attribute of the product. For example, battery life could be the only positive feature of a certain smartphone besides other features. This concept is called feature in the literature that has produced research outcomes on reviews.

There are exclusive websites and platforms for reviews such as cnet.com and pricegrabber.com where millions of users have written reviews on several million products. The focus of this research is extracting the valuable knowledge from the review which is generally complex in nature unless the hidden meaning is extracted. The reviews could be broadly classified into positive, negative, and neutral. However, the classification of the reviews into these categories is not as easy for any mechanism.

2. LITERATURE SURVEY

Internet has provided unlimited power to the common user to describe the usefulness of a product online. The exponentially increasing number of reviews have only made it difficult to search for the right attribute of the product a consumer is looking for in a product. It is not an easy task to get clear idea of the screen resolution quality of a recently released iPhone before you search a million reviews and hundreds of pages.

The increasing use of indirect references, emojis, sarcasm, contextual meaning, and other styles of review writing has made it difficult for ordinary mechanisms to extract the hidden meaning in the message. The first line of features extracted by the proposed mechanism in the literature are the spelling errors, subjectivity, and also the readability aspect of a given review. The other features include subjectivity usefulness, identity, and the history of the user as per the reviews posted before. These features in the research proposed in the literature offer a comprehensive approach to deal and analyze the review. This

mechanism deals into the depth of a review to analyze its hidden details. [1]

Lexicon based methods of sentiment analysis were the first of its kind to analyze the sentiment of a text. They relied on the literal meaning of the words used pin the text. Reviews were classified positive, negative or neutral based on the words used. However, these primitive methods failed to understand the hidden meaning. Another advanced method of semi-supervised sentiment analysis was proposed in the literature that used both lexicon based and the contextual meaning to extract the meaning of the review. Biparte graph representation method was used to determine the meaning and classify the sentiment of the review. The semi-supervised method was found to be effective in comparison with the purely supervised method from the literature [2].

Online auctioning has gained potential in the recent times and an aggregator mechanism was proposed in the literature that collects the data from several auction websites and then shortlist the hot items in the website. A list of analyzed reviews and the auctions from several websites offer valuable data for the seller to promote and buyer to decide the right product. This method works in two steps. All the features of the product available from the reviews are to be extracted in the first step. An HMM based training method is used to train the system so that the information is extracted in detail. The second step discovers the knowledge and then summarized the list of items to be recommended for the user to buy. A semi-supervised method and graphs minute algorithm is used to extract the information. Frequency and position of the review are the two factors considered in the algorithm. Several experiments revealed the effectiveness of the method proposed in the literature [3].

Helpfulness is another important feature of reviews. Extracting the helpfulness of a review based on different factors such as timeliness and expertise of the reviewer is also discussed and effective solutions proposed online. Writing style to extract the tone of the reviewer was also considered in the design and the effectiveness of considering these factors for analyzing the sentiment of a review was found to be very effective using an advanced method from the literature. Dataset of IMBD was considered and the results were proven to be effective [4].

Wrapper based method from the literature was analyzed for its efficiency. Information extraction was possible using the learning attributes from the previous wrappers and characteristics from the same. The knowledge of the previous information was used on a new website to accurately analyze the review. Headers and standard segments analyzed from the previous websites were easily ignored and the actual text was analyzed making this method very fast and accurate. Generative model was used to analyze the text fragments. Format of the data in the website was also a parameter designed in the analysis. Bayesian approach of learning was used by the researchers and the results of the wrapper method offered good results. Increase in the number of reviews in the training data was also found to be increasing the efficiency. EM techniques were used with Bayesian methods to improve the accuracy [5].

The review of literature has offered insights about the current state of research on this topic. There is sufficient research done in this regard. However, the growing number of reviews and the difficulty in shortlisting and analyzing the reviews is a difficult task. The proposed design of sentiment analysis is based on the analysis of several research works from the literature.

3. SYSTEM DESIGN

The number of products available online on several platforms amounts to several millions. This number is rapidly growing every day. A simple statistic reveal the following numbers – Bing – 5 million products, Amazon – 36 million products, Shopper – 5 million products, etc. It is a common practice that all the websites recommend, in fact encourage, the users to write the reviews for their products. Different features of the products are to be reviewed for better design of the system. Larger the data, accurate is the analysis. Amazon has recently introduced a concept listing the most positive and negative review along with features to like the review.

Aspect ranking framework is the contribution of this paper to the research literature. Important aspects of the products are automatically identified based on the large number of customer reviews available on a particular product. The frequency of the important features is often higher for any product. Such frequently mentioned features have been proven to increase the likelihood of buying any product. Frequency-based solution is a method proposed in this research. Different aspects of a product and their specific opinions are the strong foundation for the overall result of the analysis. This assumption is used in the design. The overall ratings, their consistency, and the features are the core aspects of the proposed design. For example – battery could be the most negative aspect of the iPhone whereas the overall rating will still be high. Display/screen quality may be more important than the battery feature but may not be mentioned as frequently by the reviewers. Thus, the frequency based method is not the only way to analyze the reviews and review the product.
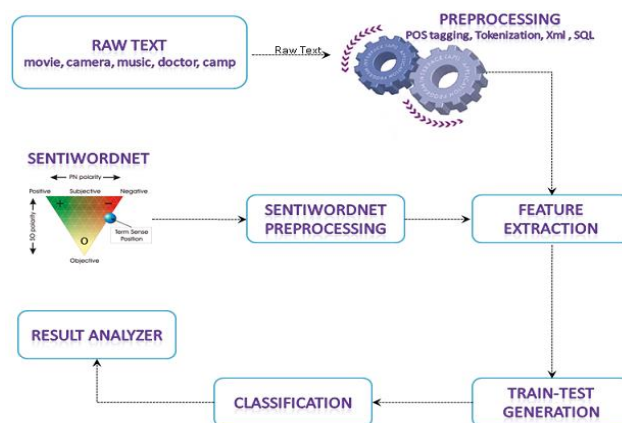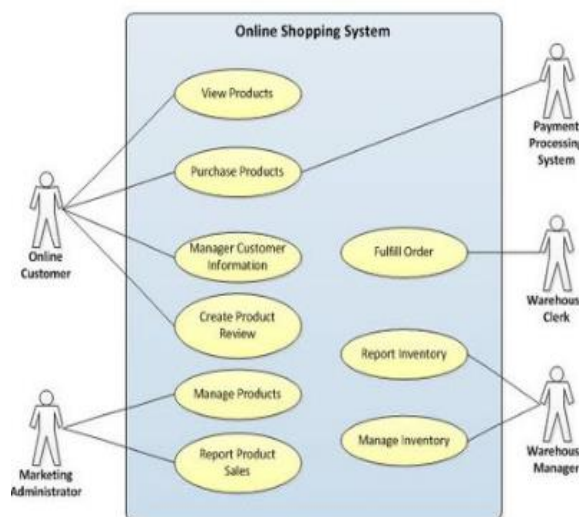

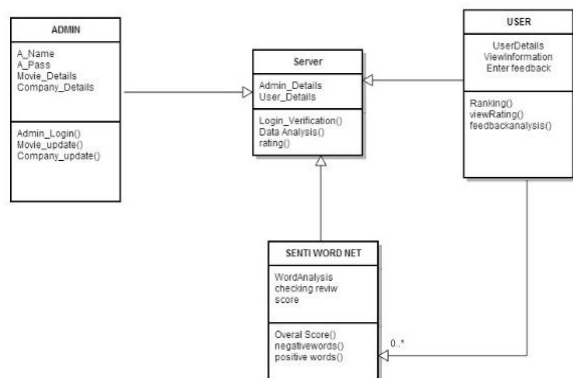**Fig. 1: System architecture**

Fig. 2: UML use case diagram
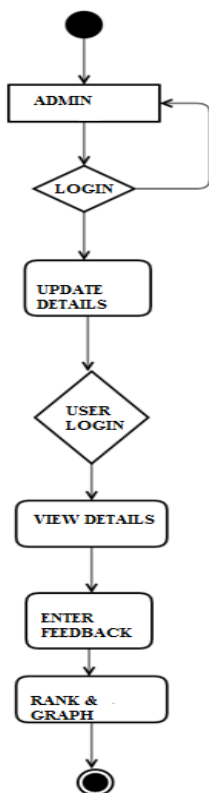
Fig. 3: Class diagram
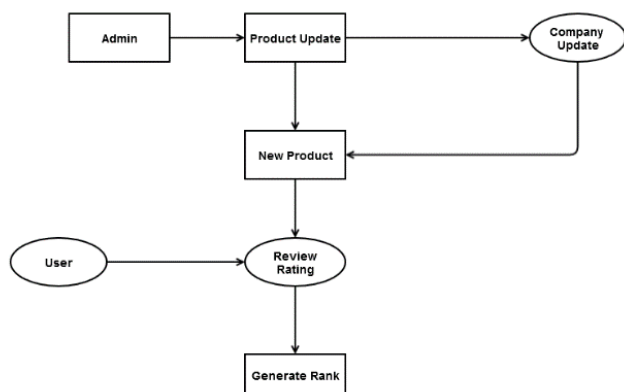

Fig. 4: Activity diagram

Fig. 5: Overall architecture diagram

4. RESULTS AND DISCUSSION
The following are the system module of the proposed system.
- Product Aspect Identification.
- Product Aspect Ranking.
- Probabilistic Aspect Ranking.
- Extractive Review.
- Sentiment Classification.
- Consumer review

There is no standard format for a review. Each website or platform employs their own method of reviews. Some of the popular websites such as CNet have a detailed format for reviews. A consumer is to give an overall review for the product and then continue to review the positive and negative features of the product in a detailed manner. Also, a paragraph of text has to be accompanied with ratings so that the review is detailed. Reevoo, another website recommends the users to leave a review with overall rating and also positive and negatives of the reviews with free text. Pros and cons of any reviews can be extracted in the first step of the proposed design. Identification of frequent noun terms uses lexicon methods. This method is effective in detecting the noun terms whereas the free text can be analyzed based on aspect identification feature of the proposed method.

Identification of the feature is the first step, then the consumer opinions are used to analyze the meaning of the review (sentiment analysis). The final step is the implementation of the probabilistic aspect ranking algorithm. Two features of the algorithm are the most important. Aspect frequency and the importance of the aspect that is an aggregate of all the previous reviews of the reviewer – the overall rating. A customer is expected to review several aspects of the product. It is possible that a list of features could be given for the customers to review in a particular product along with *others* option for any missing feature that the reviewer needs to mention. Each feature requires a minimum and maximum rating to ensure the analysis is standard throughout the reviews. The maximum and minimum rating differs across the websites. It is possible to standardize the ratings for analysis using statistical methods.

There are certain aspects that are considered important in reviews of a particular product. Frequently commented feature, opinion of the consumer that influence the overall opinion for a buyer. The aggregation is not straight forward throughout the features. Ranking and weights are to be used in the calculations to even the features.

Important, unimportant, unique, special are some of the levels of weights that could be used in designing the algorithm. Higher number of levels may be used for accuracy. However, it is important not to make the algorithm complex given the large number of reviews it has to process in short time.

The weighted sum of opinions on a particular product is calculated. The overall rating or and the rating *r* are calculated before the analysis is initiated. The matrix representation of these terms indicate the importance of the weight on each row and column. Numbers of the larger indices represent higher weights.

Unorganized reviews and the large number of reviews add to the complexity of the analysis algorithm. The large number of reviews available online are more distracting and misleading to any customer. Although the current method of reviews is working, it is important that every word of the reviews is analyzed to extract all the potential of that review to influence the buying habit of the customer.

The automated summarization and analysis of the reviews is the contribution of this research. The summarization of the existing reviews is classified into abstractive and extractive. The abstractive method analyzes the review to extract all the hidden meaning using contextual and summarization methods. The review is converted into a natural language using linguistic methods of analysis. The most important information of the review is shortlisted and summarized.

Aspect level sentiment analysis and classification is an advanced method. Lexicon-based methods are the first and strong methods used. The parts of the text is analyzed using different methods and a list of words, phrases that determine the sentimental orientation of the text is analyzed. The quality of the lexicon used in the sentiment influences the quality of the results. Supervised methods usually train the algorithm with the previous data. The classifier is then used to analyze and predict the positive/negative nature of the sentiment in the text.

Support Vector Machine (SVM), Naive Bayes, and Maximum Entropy (ME) are the most commonly used classification models. Large amount of training data is required for the supervised algorithm for higher accuracy. Processing the training data for any classification is a complex task. The classified data is considered as the input for training data.

It is possible to extend this method to a document level of sentiment analysis and classification as the goal of the method is to determine the sentiment irrespective of the number of words or amount of data. The aspects of reviews, features reviewed, and weight of each feature vary for each product. It is not possible to standardize the features for all the products.

A typical example of smartphone is considered for understanding. IPhone may be highly reviewed for its ease of use and touch screen but negatively reviewed for battery backup. However, the overall ratings for the product is high in spite of a significant number of negative reviews on a particular aspect. This is a similar scenario for a large number of products. Also, it is not easy to predict the type of reviews for similar products either. Thus, the aspect ranking feature proposed in this paper offers a solution for the existing problem in terms of reviewing the specific features of a product. The sentiment classification proposed in this paper is an advanced version of the existing methods of sentiment analysis in which the aspects of a particular product are ranked depending on various factors such as weight and importance. The sentiment classification has been tested for over 100 random reviews along with document level classification. Different training and testing data revealed that the design is effective.

The following are the screenshots of the forms for various modules.

Fig. 6: Registration

Fig. 7: Product category

Fig. 8: Product rating

Fig. 9: Product description

Fig. 10: Owner question

Fig. 11: Analytical graph

5. CONCLUSION

In this research article, a novel version of aspect ranking framework is proposed. The highlight of this method is that it identifies the important aspects of a particular product with the data of the large number of customer reviews. The three important features of the proposed method are – aspect identification, sentiment classification (aspect), and ranking of the aspect. These three steps ensure that the aspect ranking method accurately analyzes the review and ranks it based on the aspects analyzed. Free text reviews were first analyzed in this paper to implement the aspect identification and classification of the sentiment. A probabilistic aspect ranking method (algorithm) was developed that considered the various aspects of the product to rank the aspects after analysis. Each and every aspect is important in this analysis, thus offering accuracy for the reviews analyzed. The aspects of each product are then ranked with importance scores. 97,560 consumer reviews were analyzed as part of implementation for 21 popular products in different (8) domains. Document-level sentiment analysis and extractive review summarization are the contributions of this research. The corpus is available publicly on request and the results prove that

the solution is effective in sentiment identification. The solution could be used for various applications to improve the accuracy.

6. REFERENCES

[1] Anindya Ghose, and Panagiotis G. Ipeirotis, "Estimating the Helpfulness and Economic Impact of Product Reviews: Mining Text and Reviewer Characteristics", IEEE Transactions on Knowledge and Data Engineering, Vol. 23, Issue 10, pp. 1498-1512, Oct. 2011.

[2] Vikas Sindhwani, and Prem Melville, "Document-Word Co-regularization for Semi-supervised Sentiment Analysis", 2008 Eighth IEEE International Conference on Data Mining, 15-19 Dec 2008.

[3] Tak-Lam Wong, and Wai Lam, "Hot Item Mining and Summarization from Multiple Auction Web Site", Fifth IEEE International Conference on Data Mining (ICDM'05), 27-30 Nov 2005.

[4] Yang Liu, Xiangji Huang, Aijun An, and Xiaohui Yu, "Modeling and Predicting the Helpfulness of Online Reviews", 2008 Eighth IEEE International Conference on Data Mining, 15-19 Dec 2008.

[5] Tak-Lam Wong, and Wai Lam, "A Probabilistic Approach for Adapting Information Extraction Wrappers and Discovering New Attributes", Fourth IEEE International Conference on Data Mining, 2004.

[6] M. Taniguchi, N. Masuhara, and K. Burnett, "Water, energy, and food security in the Asia Pacific region," Journal of Hydrology: Regional Studies, vol. 11, pp. 9–19, Dec. 2015.

[7] V. Jalali, S. A. Kapourchal, and M. Homaee, "Evaluating performance of macroscopic water uptake models at productive growth stages of durum wheat under salineconditions," Agricultural Water Management, vol. 180, pp. 13–21, Jan. 2017.

[8] Y. E. Duan, "Design of intelligent agriculture management information system based on IoT," IEEE Computer Society, vol. 1, pp. 1045–1049, Mar. 2011.

[9] S. Suttles, "Agricultural energy use and the proposed clean power plan," Evaluation & the Health Professions, vol. 34, no. 3, pp. 362–370, Sep. 2015.

[10] D. J. Rodriguez, A. Delgado, and P. Dayquil, "Thirsty Energy," pp. 1–72, Jun. 2013.

[11] Y. Xiang, J. Liu, and F. Li, "Optimal active distribution network planning: a review," Electric Power Components and Systems, vol. 44, no. 10, pp. 1075–1094, May.2016.

[12] dQ. Liu, L. I. Mei-Hong, and M. F. Zhu, "Application of solar energy in intelligent ecological agriculture," Journal of Beihua University, vol. 13, no. 3, pp. 344–347, Mar. 2012.

[13] P. J. Sonneveld, G. L. A. M. Swinkels, and J. Campen, "Performance results of a solar greenhouse combining electrical and thermal energy production," Biosystems Engineering, vol. 106, no.1, pp. 48–57, May. 2010.

[14] P. J. Sonneveld, G. L. A. M. Swinkels, and G. P. A. Bot, "Feasibility study for combining cooling and high grade energy production in a solar greenhouse," Biosystems Engineering, vol. 105, no. 1, pp. 51–58, Jan. 2010.

[15] A. Yano, K. Tsuchiya, and K. Nishi, "Development of a greenhouse side-ventilation controller driven by photovoltaic energy," Biosystems Engineering, vol. 96, no. 4, pp. 633–641, Apr. 2007

[16] J. Y. Zan, Z. M. Liu, and H. Liao, "Simulation research on photovoltaic greenhouse temperature," Journal of Yunnan Normal University, vol. 2, no. 34, pp. 42–47, Feb. 2014.

[17] F. X. Zhang and Y. L. Wang, "National golden solar engineering photovoltaic eco-agriculture demonstration project," Agricultural Knowledge, vol. 23, pp. 38–38, Oct. 2013.

[18] 18 Q. Sun, T. Gao, and R. Yang, "Design and realization of smart micro grid for household in rural area," Transactions of the Chinese Society of Agricultural Engineering, vol. 29, no. 13, pp .150–157, Jul. 2013.

[19] M. Nasir, H. A. Khan, and A. Hussain, "Solar PV based scalable DC microgrid for rural electrification in developing regions," IEEE Transactions on Sustainable Energy, vol. 9, no. 1, pp. 390–399, Aug. 2017.

[20] M. J. Sarker, B. Asare-Bediako, and J. G. Slootweg, "DC micro-grid with distributed generation for rural electrification," in Proceedings International Universities Power Engineering Conference. IEEE, pp. 1–6, Sep. 2012.