A Recommendation System for Online Purchase Using Feature and Product Ranking

[1]Karthik.R.V, [2]Sannasi Ganapathy
School of Computing Science and Engineering
VIT University-Chennai Campus
Chennai-127, India
Karthikr.v2016@vitstudent.ac.in, sganapathy@vit.ac.in

Arputharaj Kannan
Dept. of Information Science and Technology
CEG Campus, Anna University
Chennai-25, India
kannan@annauniv.edu

***Abstract*—Social network occupies an important place and takes a considerable amount of time in people's daily life. It has become so popular that people are sharing a huge amount of data and opinion on social network/review sites, which in turn helps to find interesting insights for organizations / vendors or consumers. In this paper, we propose a new algorithm called Feature Based Product Ranking and Recommendation Algorithm (FBPRRA) for providing suggestions to the customers whose are interested in purchasing good quality products. The proposed algorithm analyzes online products and ranks them according to product reviews. Finally, it recommends the suitable product to the target customers. Experiments have been conducted using online reviews for evaluating the proposed algorithm and found that the proposed recommendation algorithm achieves better prediction accuracy than the exiting classifiers such as Naïve Bayes, Support Vector Machine, Random Forest, Decision Tree and K-NN.**

Keywords— Social Network; Sentiment Analysis; User Reviews; Recommendation; Opinion

I. INTRODUCTION

Social network and online shopping have become very popular these days. In Social networking sites, people share a lot of information in the form of images, opinions, ideas, etc. In many cases, they share their own experiences about a product or service and these online opinions are valuable as they play an important role in influencing a consumer buying decisions especially in online shopping. A lot of research is currently happening to extract opinion insights. The collective term for this is known as Sentimental analysis. This paper focuses on the importance of sentimental analysis in social networks and online shopping.

Feature selection is used to reduce the dimensionality of the vast data. It detects all the relevant features and discards the irrelevant ones. Feature Selection technique has several benefits: Improves the performance of machine learning, helps in data reduction, reduce computational burden and reduces the storage requirements and space and improves the accuracy of prediction. There are two types of feature selection methods namely filter based approach and wrapper based approach.

Classification algorithms are broadly categorized into two approaches namely lexicon based and machine learning based approach. Lexicon based approach assumes each word as independent, considers all the words with equal importance and ignore the context and meaning of the word in the sentence.

Machine learning approach overcomes this problem; it acts on the meaning or context of the word in the sentence. It contains two steps: 1] Extract the feature and storing it in the Feature vector and 2] Train the Feature vector by using classification algorithm. Machine learning approach is further classified into three categories: i) Supervised machine learning: deals with labeled data where machine is trained with training data and tested it by test data and unsupervised learning. ii) Unsupervised machine learning: deals with unlabeled data, machine is not trained with prior data iii) Semi supervised machine learning: this is introduced to deal with small-labeled sample problem where unlabeled data is much more when compare to label data.

In this paper, a new Feature Based Product Ranking and Recommendation Algorithm (FBPRRA) is proposed for recommending the good quality product to the customers by analyzing the online reviews. The rest of paper is organized as follows: In Section 2, Research works on sentimental analysis is provided. Section 3 explains the proposed work in online product recommendation using the sentimental analysis. Section 4 gives the results and discussions. Section 5 summarizes the work and suggests some future work.

II. LITERATURE SURVEY

There are many works that have been proposed in this direction by various researchers in the past [2-21]. Among them, Liu et al. [6] proposed a method to rank products using online reviews by intuitionistic fuzzy set theory. Their method uses two techniques: Sentiment analysis technique to identify the sentiment orientation for each product feature and ranking product technique which converts the sentiment orientation to fuzzy number. Wang et al. [17] proposed a new idea in the feature based ranking product method. In the proposed method, it identifies the implicit demographic information from acceptable volume of reviews and finally provides customer specific recommendation based on user category.

Zhang et al. [19] proposed a new method for ranking products through online reviews. In that, subjective sentences and comparative sentences [compare pair of product] are extracted from the online reviews, directed & weighted product graph is constructed using dynamic programming. Zhang et al. [20] further proposed subjective sentences against each feature are evaluated and finally overall ranking is provided based on each feature.

978-1-5386-6835-1/18/$31.00 ©2018 IEEE

Zhang et al [21] still improved the feature based ranking method by including the influence factors number of helpful votes got for the review and review posted age.

Fang et al [3] provided a multi model aspect opinion model (mmAOM) which accepts both the text based data and images as well. MmAOM find the correlation between the provided input text and the images. Ganapathy et al [4] conducted an effective survey about the roles and responsibilities of intelligent agents over the feature selection and classification processes. They have considered many key aspects for identifying the features which can contribute more for making effective decision in their survey. Priya et al [9] designed a new neuro fuzzy link based classification model for analyzing the user behaviors of the social networks. They have achieved better prediction accuracy by using their neuro-fuzzy model. In [10], the authors proposed a new behavior analysis model by using the existing clustering algorithm called genetic weighted fuzzy C-means clustering and the neural network based fuzzy classifier. They have used intelligent neuro-fuzzy rules for predicting the user's behaviors in social networks.

Sairamesh et al [14] introduced a new prediction system for predicting the user interests for providing the relevant information using relevance feedback and the users re-ranking. They have used a newly proposed re-ranking algorithm for categorizing the more relevant users by using the customer feedbacks. Sadhana et al [13] proposed a new recommendation system for identifying the opinions of various users by using target opinions from online reviews. In their system, they have used a semi-supervised word alignment technique for predicting the opinions of the customers.

III. PROPOSED APPROACH

This section describe in detail about the proposed algorithm called Feature Based Product Ranking and Recommendation Algorithm (FBPRRA) which is developed for recommending the suitable products based on the analysis report of the customer feedbacks.

Based on online reviews, Wang et al [17] have given an idea to recommend products for purchase by considering details about consumers who have purchased and to whom it was purchased. For example, consider this review by a consumer - "I bought an Optical mouse as a gift for my son, which is very good and useful." Based on the review, in later purchases, optical mouse can be recommended to anyone who is looking to purchase a gift for his/her son.

Here, the proposed work is to provide the recommendation by considering the 'occasion for purchase' (Wedding anniversary, birthday, any festival etc). This would be especially useful in a country like India where there is a large online shopping customer base and its people celebrate a large number of festivals.

Example reviews by customers:

"I bought this product for my friend's birthday"

"Bought it for my parents' wedding anniversary"

To do sentiment analysis on social media content and provide product feature based ranking, it is important to have a generic technique to satisfy all of the following factors.

Wang et al [17] found reasonable number of reviews containing demographic information while exploring leading sites, which contain about 15 million reviews.

Variant Inputs: Lots of research and study have been done on the text based sentimental analysis, but it is essential to have a generalized Sentiment analysis technique to handle all type of inputs, mixed languages, smileys, annotations, unstructured data from review blogs and social networking sites.

Identification and ignoring of sarcasms: In order to improve the accuracy, it is required to have an effective algorithm to identify and ignore sarcasms in the reviews from social media. Sarcasms might be in the form of simple text sentence. Sentiment lexicon plays an important role in classifying the feature into opinion polarity. It is required to have a suitable lexicon list to use it across domain.

Customer specific recommendations: On reading user profiles and previous buying history, user categories to predict and customer specific feature selection and weightage can be calculated automatically. Identify implicit features: There are still no effective methods to identify the implicit features of a product.

Proposed approach: In addition to demographic information, it identifies the age group of both purchaser and to whom it was purchased (target user), the technique can be further extended to consider occasions or festival related information from online reviews. By doing so, it is possible to recommend alternative products based on the time of occasion / festival.

In order to recommend a product, it is required to first understand age group of a customer. Based on existing techniques proposed by Zang et al [19], age group of purchaser and target user are identified from reviews. This is further improved in our approach by getting the age group information from user profiles also.

Improve the weightage accuracy: If a review contains both positive and negative comments, it is required to weightage based on the sentiment severity and the emotion expressed.

In the existing approach, the consumer required to specify the interested feature and required to indicate the weight for those features. Most of the paper does not focus on the social network, where enormous volume of user generated reviews and comments are available. In proposed method, the user profile information and past purchase history information are retrieved. In the social media twitter/Facebook, reviews / twits for all the related products purchased by similar group are retrieved and weightage is calculated and stored in the lookup table.

In recent papers, the neutral reviews are considered, but it considers neutral when the review contains both positive and negative. This would reduce the accuracy of the recommendations. In all the existing works products with less than five reviews were ignored, which is not an ideal case.

All the newly launched products would have less number of reviews in nature. In this case, if most of the reviews have positive feedbacks, than the reviews can be considered and would give positive recommendation for the product. Since information from social media also considered, there will be more number of inputs/ feedback available in Social media for most recently launched product. Based on the feedback/ input in social media, weightage of the product is decided.

In the proposed algorithm, all these points are considered. By processing the User profile, past purchased product history and user posted comments/opinions in the social network, it is possible to find the user interested features and its weightage. Both feature and weightage are identified based on user profile, past product purchased history and review comments posted by the user. Alternative/ Similar products are identified by comparative sentences in reviews, crawling through related sites and comments posted in the social networking.

By applying sentiment analysis, the product ranking is calculated. If there are more positive reviews, then ranking will be high. Age group and target user are noted in this work based on demo graphic information. So, the top rated products are recommended based on the user age group. The proposed feature make use of benefits of sentiment analysis to know the good products based on online reviews and also through the demo graphic information approach it also gives the product recommendation based on the age group. For example: "Apple IPhone" is not good to recommended for all the users. For people of more than 60 years, it is useful to recommend the phone with more battery power.

Feature expressed in each review are identified and weightage is calculated. In proposed method a weightage scale is introduced if the review contains both positive and negative, it will not consider as neutral. This will improve the accuracy of the product recommendations. Using Demo graphic information and user profile information with Sentiment analysis is new approach, which is not yet tried as per our knowledge. This proposed algorithm helps to provide the product feature recommendation based on target consumer.

The proposed new algorithm consists of the following features:

- Occasion details are identified from the online reviews
- Actual end users are captured from the Review blog
- User-interested features and weightage calculation are automatic
- Providing product feature recommendation based on target consumer

The steps of the proposed algorithm are as follows:

Feature Based Product Ranking and Recommendation Algorithm (FBPRRA)

Input : Online review comments / feedbacks

Output : *Recommended product*

Fig.1 shows the proposed architecture.

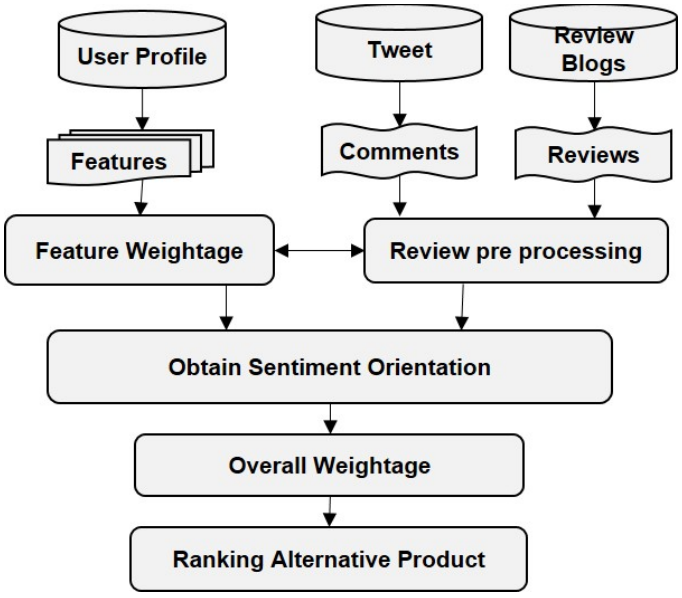

Fig. 1. Proposed architecture for feature based product ranking

For each nReview(i) new online product review

Do

(i) Identify the target consumer category

Uc= [u1, u2, u3, u4..] // Uc be the user category

(ii) Identify the opinion and calculate weightage

Begin

Let nReview be review of the product P

Wp= Avg ((Wipve *Wic) + (Wineu *Wic))

End

Where,

Wp is the weightage of the review

Wipve is positive opinions in review

Wineu is neutral opinions in review

Wic is the weightage of the opinion

(iii) Identify the target user / occasions expressed in the review

Begin

Let nReview be review of the product P

Identify target users (tUser)

a) Filter out the target users in tUser if present

eg … Bought for …. My son / gift to present it to wife ….

b) Filter out the Occasion details occD**…**

eg...Bought for …. Festival

End

(iv) Include all the extracted data in reference database

Begin

insertIntoReferencetable (P, Pc,pUg, Wp, tUg, OccData)

End

Where,

P is the product purchased

Pc is the Category of the product

pUg is the purchaser user group

Wp is the weightage of the product

tUg is the target / actual end

occData is the Occasion

The proposed algorithm works with four major activities: First, it identifies the target consumer category for each online product review which is received as input. Second, the algorithm identifies the opinion and calculates the weights for each review of the particular product. Third, it identifies the target user expressed in the review in each input product review. Finally, it includes all the extracted data in reference database for the each input product review which is collected in online from users.

IV. Result And Discussion

This work considered a movie review dataset and the RapidMiner tool is used for evaluating the proposed classification algorithm and also compared with the existing classification techniques such as Naive Bayes, Super Vector Machine (SVM), Decision tree, Neural Network, KNN and Random Forest.

Before start the classification, preprocess steps: tokenize, remove stop words and stemming are applied to improve the performance.

Using the pattern extraction method, customer review is checked whether any demo graphic information target user age group / any occasion information is present. This additional information is stored in the database with the product name. Purchaser age group is also stored if user profile information is available, this information is later used for product recommendation.

User interested feature is also extracted by previous user comments and review's. If there are no previous comments, then feature from similar age group is selected. Based on the sentiment analysis outcome, ranking is calculated.

Fig. 2 shows the preprocessing steps for the movie review dataset.

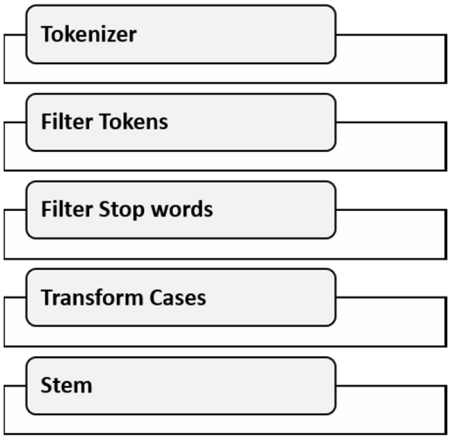

Fig. 2. Preprocessing the input movie review dataset

Table I shows the comparative analysis between the proposed classification and the existing classification techniques which are available in the literature according to the classification accuracy.

TABLE I. Comparison Of Accuracy Level Of Different Classification Techniques

S.No.	Learning Technique	Accuracy
1	Support Vector Machine	58.00%
2	Decision Tree	58.00%
3	Naive Bayes	50.00%
4	KNN	59.00%
5	Random Forest	54.00%
6	Proposed algorithm	81.23%

From Table I, it can be seen that the performance of the proposed algorithm is higher than other existing classification algorithms such as Support Vector Machine, Decision Tree, Naive Bayes, KNN and Random Forest. This is due to the fact that the use of demographic information for decision making.

One more dataset "Amazon ecommerce dataset" is taken for evaluation. Dataset contains product name, manufacturer, price, number of reviews, number of answered questions, product description and customer reviews.

Fig. 3 shows the presence of demographic information in the Amazon dataset taken. Out of 100000 product items, 30% of reviews contain target user or an occasion detail, which in turn helps to recommend products for similar age groups or during the time of festival. Inputs from the Social media like Twitter would also help to find the feedback of the product. Result helps to recommend the product based on the age group of the purchaser and also based on the time of the purchase.

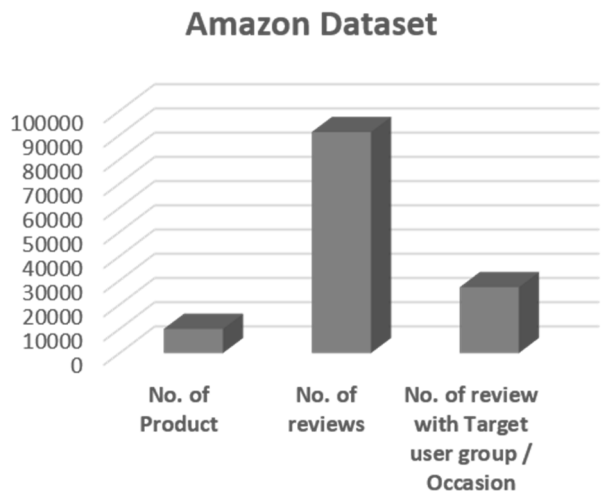

Fig. 3. Product reviews with target analysis on amazon dataset

Table II shows few reviews in the data set that contain target user/occasion details, these review details would be very useful for recommending products later based on the purchaser / to whom he/she has made the purchase.

TABLE II. AMAZON DATASET DETAILS

S.No	Reviews with Target User Group / Occasion	Information Retrieved
1	Recommend : Birthday present // 5.0	Occasion: Birthday
2	My 2 year old loves chuggington so we decided to get her this set	Age group: Target User: 2 yrs. child
3	Very good train set and very popular with the young infants in my family	Age group : Target User : 2- 5 yrs. child
4	19 Feb. 2015 // My daughter loves it | Good in parts	Age group : Child , Purchaser : Young parent

V. CONCLUSION AND FUTURE WORK

In this paper, a new online product recommendation system has been proposed and implemented for recommending online products efficiently. It is based on target audience's demographic information (age group of purchaser and to whom product was purchased) and other influencing factors like occasion, festival period while calculating the weightage. The proposed method - Feature based product ranking and recommendation system takes the reviews/feedbacks from the review sites and also from Social media. Social media sites contain more feedbacks and reviews when compared to the review sites. It also identifies interested feature details of users automatically by considering his/her past comments / reviews. This avoids an existing problem where user has to provide his/her interested feature details and weightage manually, at worst case, entire feature has to be considered. In the proposed idea, the most important factor considered is festival / occasion time. The main purpose of this work is to handle huge volumes of review data from different datasets and to improve the accuracy of the predictive online recommendation system.

The proposed approach can be improved further by including Fuzzy logic. By introducing Fuzzy logic, influencing factors like time and location, the accuracy of product recommendation system can be further improved.

REFERENCES

[1] C. Catal and M. Nangir, "Sentiment classification model based on multiple classifier ", Applied Soft Computing, vol. 50, pp. 135–141, November 2016.

[2] F. Corea,"Can Twitter Proxy the Investors' Sentiment? The Case for the Technology Sector", Big Data Research, vol. 4, pp. 70–74, 2016.

[3] Q. Fang, C. Xu, J. Sang, M.S. Hossain and G. Muhummad, "Word of mouth understanding: entity centric multimodal aspect opining mining in Social media", IEEE Transactions on Multimedia, vol.17, pp. 411-427, July 2016.

[4] S. Ganapathy, K. Kulothungan, S. Muthurajkumar, M. Vijayalakshmi, P. Yogesh and A. Kannan, "Intelligent feature selection and classification techniques for intrusion detection in networks: a survey", EURASIP Journal on Wireless Communications and Networking, Vol.271, pp. 1-16, 2013.

[5] T.L. Griffiths, M. Steyvers, D.M. Blei and J.B. Tenenbaum, "Integrating topics and syntax", In advance in neural information processing systems, MIT press, vol. 17, pp. 537-544, 2005.

[6] Y. Liu, J. Wu and Z. Fan, "Ranking products through online reviews: A method based on sentiment analysis technique and intuitionistic fuzzy set theory", Information Fusion, pp. 149–161

[7] M. Meire, M. Ballings and D. Poel, "The added Value of auxiliary data in Sentiment analysis of Facebook posts", Decision Support Systems, vol. 89, pp. 98-113, 2016.

[8] S.A. Mulay, J. Shrijeet, M.R. Shaha, H.V. Vibhute and P. Mahesh, "Sentimental analysis and opinion mining with social networking - Predicting Box office collection of Movie", International journal of emerging research in management and technology, vol.5, No.1, pp. 74-79, January 2016.

[9] P.I. Priya, G.D. Kumar, K. Arputharaj and S. Ganapathy, "Neuro Fuzzy Link based Classifier for the Analysis of Behavior Models in Social Networks", Journal of Computer Science, vol.10, No.4, pp. 578-584, 2014.

[10] P.I. Priya, D.K. Ghosh, A. Kannan and S. Ganapathy, "Behaviour Analysis Model for Social Networks using Genetic Weighted Fuzzy C-Means Clustering and Neuro-Fuzzy Classifier", International Journal of Soft Computing, vol. 9, No.3, pp. 138-142, 2014.

[11] S. Qian and T. Zing, "Multi modal event topic model for social event analysis", IEEE Transactions on Multimedia, vol. 18, No. 2, pp. 233 - 246, Feb 2016.

[12] S.D. Roy and S.K. Chakraborty, "Impact of Social Media / Social Networks on Education and life", vol. I, No. I, pp. 141-147, February 2015.

[13] S.A. Sadhana, L. Sairamesh, S. Sabena, S. Ganapathy and A. Kannan, "Mining Target Opinions from Online Reviews Using Semi-supervised Word Alignment Model", Second International Conference on Recent Trends and Challenges in Computational Models (ICRTCCM), pp. 196-200, 2017.

[14] L. Sairamesh, S. Ganapathy, R. Bhuvaneshwari, K. Kulothungan, V. Pandiyaraju and A. Kannan, "Prediction of user interests for providing relevant information using relevance feedback and re-ranking", International Journal of Intelligent Information Technologies (IJIIT), vol.11, No.4, pp. 55-71, 2015.

[15] P. Teisseyre, "Feature Ranking for multi label classification using Markov Networks", Neuro computing, vol. 205, pp.439-454, 2016.

[16] B.V. Kumar and M.B. Jadav, "Analysis of Various Sentiment Classification Techniques", International journal of Computer applications, vol. 140, pp. 22-27, April 2016.

[17] J. Wang, W. Zhao, Y. He and X. Li, "Leveraging product adapter information from Online reviews for product recommendation", AAAI Conference on web and social media, pp. 464-472, 2015.

[18] R. Xia, "Dual Sentiment Analysis", Information Fusion, 2015.

[19] K. Zhang, R. Narayanan and A. Chaudhary, "Mining Online Customer Reviews For Ranking Products", EECS Department, Northwestern University, Technical Report, 2000.

[20] K. Zhang, R. Narayanan and A. Chaudhary, "Voice of the customers: mining online customer reviews for product feature-based ranking", WOSN'10, Boston, USA, 2010.

[21] K. Zhang, Y. Cheng, W.K. Liao and A. Chaudhary, "Mining millions of reviews: a technique to rank products based on importance of reviews", Proceedings of the 13th International Conference on Electronic Commerce, ACM, New York, USA, Article No.12, 2011.