Available Online at www.ijcsmc.com

International Journal of Computer Science and Mobile Computing

A Monthly Journal of Computer Science and Information Technology

ISSN 2320–088X
IMPACT FACTOR: 6.017

IJCSMC, Vol. 6, Issue. 8, August 2017, pg.55 – 62

SENTIMENT ANALYSIS IN E-COMMERCE USING RECOMMENDATION SYSTEM

T. Karnan[1], G.Seenuvasan[2]

[1]Research Scholar, P.G and Research Department of Computer Science, Swami Vivekananda Arts and Science College, Orathur, Villupuram, Tamilnadu, India

[2]Assistant Professor, P.G and Research Department of Computer Science, Swami Vivekananda Arts and Science College, Orathur, Villupuram, Tamilnadu, India

Abstract- In the field of text mining the sentiment analysis is one of the current research topics. Sentiment analysis is the best solution for opinions and sentiment mining from natural language. Sentiment analysis gives significant information for decision making in various domains. The sentiments include ratings, reviews and emoticons. Several sentiment detection approaches are available which affect the quality of result. This paper finding the sentiments of people related to the services of E-shopping websites. The central goal is to recommend the products to users which are posted in E-shopping website and analyzing which one is the best product. The proposed system use stochastic learning algorithm which analyze various feedbacks related to the services and then the sentiments are classified as negative, positive and neutral. The pre-processing of the data is greatly affecting the quality of detected sentiments so that the fake reviews in the website are analyzed. Finally the classification technique can be used to analysis the products which are posted in E-shopping website. The proposed system find out fake reviews made by posting fake comments about a product by identifying the MAC address along with review posting patterns. In order to find out the review is fake or genuine the proposed system find out the MAC address of the user if the system observes fake review send by the same MAC address many a times it will inform the admin to remove that review from the system. The proposed system helps the user to find out correct review of the product.

Keywords: Text mining, Opinion Mining, E-shopping, fake reviews, MAC address, Recommendation, Emoticons.

I. INTRODUCTION

Many devices such as Laptops and Smart phones using internet have made online shopping very easy. Online shopping has become more and more popular because of its variety of types, lowest price and fast logistic systems. Now a days many people use online shopping to purchase the product through the Internet. The user comments become the most important information to judge the quality of the product. In order to improve the quality of the products the product manufacturers can obtain the comment about the product from the users. It is not easy to analyzing and concluding the large volume of data in the websites. So how to extract

© 2017, IJCSMC All Rights Reserved

useful information and build objective products' quality test system automatically to deal with the massive textual information is emerging in the related research field. Opinion Mining is technology based on text mining and natural language processing. It provides the approach to manage with the problem.

Today generating summary of the products has been attracting many researchers. Emotional orientation of each review is focused with Document-level sentiment analysis. It recognizes the opinion of the contents which authors express, mainly discusses the sentence-level opinion mining and treats the statements of the product' features for each viewpoint as analysis objects, then we can find authors' opinion inclinations. Therefore sentence-level sentiment analysis is the main task on opinion mining. The method can find the specific details of the comments and has a high confidential degree, but the operation is very complex. For example, if we take a type of laptop into consideration, we can divide the laptop's features into performance, price, appearance, endurance time, brand and so on. We consider each feature or attribute which each author expresses for each comment respectively, then do a comprehensive evaluation in order to avoid the overgeneralization. Feature-specific opinion mining attracts much attention. An object is an entity where as it can be a product, person, event, organization, topic or something else. It is associated with a hierarchy or taxonomy of components or a set of attributes. Meanwhile, each component can also have its own set of subcomponents or attributes. A feature is defined to show both components and attributes and it is the subject of a review. In fact, people follow the grammatical rules to organize sentences while writing articles. However, under informal circumstance the people usually abandonment it and there are so many spelling mistakes. This phenomenon is specifically prominent when people make comments after online shopping.

Sentiment analysis also known as opinion mining refers to the use of natural language processing and computational linguistics to extract subjective information from the given data and classify opinions. It is a wider concept and several tasks are involved in it. The most important are as follows:

(i) Sentiment Classification

It is also known as sentiment polarity or sentiment orientation. The opinion is classified into positive, negative or neutral. There are three levels of classification such as document level, sentence level and aspect level classification. The document level classification considers the whole document as an opinion and classifies accordingly. With this classification, it is possible to predict whether the review expresses a positive or negative opinion. The sentence level classification used to classifies the given sentence. If the given sentence is subjective, then it classifies it as positive or negative. Hence, the process is more related to the task of subjectivity classification that distinguishes sentences that express factual information from sentences that express subjective opinions. The aspect level classification classify the sentiment with respect to specific aspects of the entities. Therefore, this is the finest-grained and the most complex level as it is essential to extract many features and their relationships with high precision.

(ii) Subjectivity Classification

It comprises of evaluating whether a given sentence is subjective or not. It can be considered as the step that precedes sentiment classification. The accuracy of sentiment classification can be enhanced by employing an improved subjectivity classification.

(a) Opinion summarization

It is focused on extracting the main features of an entity shared with one or several documents and the sentiments regarding them. The task involves single document as well as multi-document summarization. Single document summarization analyzes the changes in the sentiment orientation throughout the document. In multi-document summarization, once the features have been detected, the system must group the different sentences which express sentiments related to those features.

(b) Opinion retrieval

The documents that express opinions or views are retrieved based on the given query. In this task, the documents are ranked as per the relevance score and the opinion score with respect to the query. The basic e-commerce reviews analysis is shown in figure 1.

© 2017, IJCSMC All Rights Reserved

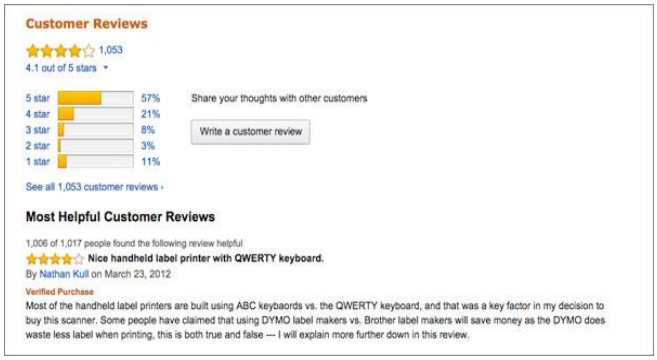

Figure 1: Analysis of Basic E-Commerce Reviews

II. LITERATURE REVIEW

Michael Jahrer et. al [1], proposed a system which help the users to find out data items within large web shops, to navigate through portals or to find friends with similar interests. The most interesting applications for recommender system have thousands of users which generate huge volume of data. For example, online shops collect purchase data and provide each user with a personalized shopping page on the login. The sources of information used for the recommender system can be common. Users generate actions like the purchase of a product, rating a product, creating a bookmark or clicking on a specific data item. Independently of the area of application or the type of information used, it is a major goal to increase the accuracy while retaining the capability of being able to use big datasets. Generating more accurate predictions is of general interest. For a subscription service like Netix, good recommendations are a key to customer loyalty. In the case of online stores better recommendations directly increase the revenue. The system provide a systematic empirical analysis of different blending methods on the Netix dataset. The Netix dataset is one of the largest available benchmark datasets for collaborative filtering algorithms today. It contains about 108 ratings, collected in a time period of 7 years. The system discuss and test several promising algorithms for blending, including neural network blending's, bagged gradient boosted decision trees, and kernel ridge regression. Our results show that linear blending is not optimal, and that it can be significantly outperformed by the presented methods. These methods are not limited to blending collaborative filtering predictors; they can be used for supervised regression problems in general.

Yu Zhang et. al [2], proposed a Collaborative filtering method which is an effective recommendation approach based on the intuitive idea that the preference of a user can be predicted by exploiting the information about other users which share similar interests. The Collaborative techniques exploit past activities of the users, such as their transaction history or product satisfaction expressed in ratings, to predict the future activities of the users. In recent years, collaborative filtering based recommendation systems have become increasingly popular because it is generally much easier to collect the past activities of users than their profiles, partially due to privacy considerations. Collaborative filtering is an effective recommendation approach in which the preference of a user on a data item is predicted based on the preferences of other users with similar interests. A big challenge in using collaborative filtering methods is the data sparsity problem which often arises because each user typically only rates very few items and hence the rating matrix is extremely sparse. In this paper, the author address the problem by considering multiple collaborative filtering tasks in different domains simultaneously and exploiting the relationships between domains. The main disadvantage of this method is a multi-domain collaborative filtering problem.

Raghunandan H. Keshavan et. al [3], studied a low complexity algorithm, based on a combination of spectral techniques and manifold optimization. The system that proves performance guarantees that are order-optimal in a number of circumstances. Collaborative filtering was studied from a graphical models perspective which introduced an approach to prediction based on Restricted Boltzmann Machines (RBM). Exact learning of the model parameters is intractable for such models, but the authors studied the performances of a contrastive divergence, which computes an approximate gradient of the likelihood function, and uses it to optimize the likelihood locally. Based on empirical evidence, it was argued that RBM's have several advantages over spectral methods for collaborative filtering. An objective function analogous to the one used in the present paper was considered early on in Srebro and Jaakkola, which uses gradient descent in the factors to minimize a weighted sum of square residuals.

© 2017, IJCSMC All Rights Reserved

Morgan Harvey et. al [4], implemented content filtering systems, based on techniques from information retrieval, are designed to assist in this process by narrowing down the number of items a user has to look through in order to fulfil a particular information need. These systems rely on textual descriptions of items and seek to match these descriptions with a user's profile in order to suggest useful items. One significant issue with this content-based filtering is that for some types of items it can be extremely difficult to choose suitable descriptive terms to search for. Another, more accurate, approach to discovering items of interest is provided by ratings-based collaborative filtering systems, which use past ratings to predict items the user may like. Such systems predict which items a given user will be interested in based on the information provided in their user profile. These profiles consist of votes or ratings for items in the system that the user has already viewed and evaluated.

HaoMa et. al[5], provided the process of trust generation is a unilateral action that does not require user to confirm the relationship. This also indicates that user does not need to even know user in the real life. "Social friendships" refer to the cooperative and mutual relationships that surround us, such as classmates, colleagues, or relatives, etc. Lots of social networking Web sites, like Facebook and Orkut, are designed for online users to interact and connect with their friends in the real life. From the definition, it can see that trust-aware recommender systems cannot represent the concept of "social recommendation", since the idea of "social recommendation" anticipates to improve recommender systems by incorporating a social friend network. Secondly, trust-aware recommender systems are based on the assumption that users have similar tastes with other users they trust. This hypothesis may not always be true in social recommender systems since the tastes of one user's friends may vary significantly. Some friends may share similar favors with this user while other friends may have totally different tastes. Hence, trust-aware recommendation algorithms cannot be directly applied to generate recommendations in social recommender systems. Thirdly, due to the rapid growth of Web 2.0 applications, online users spend more and more time on social network related applications since interacting with real friends is the most attractive activity on the Web. On the contrary, only few online systems, like Opinions, have implementations of trust mechanism. Thus, in order to provide more proactive and personalized recommendation results to online users, they should pay more attention to the research of social recommendation, in addition to the existing research of trust aware recommendation.

III. EXISTING METHODOLOGIES

Recommender Systems are necessary to provide personalized services on the Web. Recommending items which match a user's preference has been researched for a long time, and there exist a lot of useful approaches.

Collaborative Filtering With Explicit Feedbacks

First, the system discuss existing Collaborative Filtering methods with explicit feedbacks. Collaborative Filtering with explicit feedbacks that both positive and negative feedbacks are observed in the dataset. The Collaborative Filtering methods can be divided into the memory-based method, the model based method and the combination of the two. The memory-based method comprises the Neighborhood method, which calculates the similarity of the users or items. The model-based method contains the Matrix Factorization model, the Probabilistic model and Cluster based model. This approach is based on the idea that there are latent factors which represent the user-item preference relationships between users and items, and unknown preferences can be predicted using latent factors and the relationship between users-latent factors and items-latent factors. The biggest problem in Collaborative Filtering is the sparseness of observed values. It means feedbacks are observed in very small portion of all possible user-item pairs. However the Matrix Factorization model is known to work better than other models even if the data is sparse.

Collaborative Filtering With Implicit Feedbacks

Here, the system discuss existing Collaborative Filtering methods with implicit feedbacks. Basically, a dataset with implicit feedbacks contains of user-item pairs where the user provided feedbacks to the data item. Often timestamps are also provided. Existing works for Collaborative Filtering with implicit feedbacks assume that implicit feedbacks are observed as one-class positive feedbacks and missing values do not indicate the negative feedbacks. Thus, existing Collaborative Filtering methods with explicit feedbacks cannot be directly applied to the dataset with implicit feedbacks because they require both positive and negative feedbacks in the dataset. To address this problem, many existing approaches try to find possible negative values hidden in missing values and employed a weighted Matrix Factorization model. They initially filled all missing values with negative values, and assign weights to discount the relative contribution of each value to prediction. The weights are determined based on the number of items to which a user provided feedback, or the number of users who gave feedbacks for data item or given uniformly. This approach has the problem of running time. Basically, the Matrix Factorization model can save computation for missing values, but the weighted Matrix Factorization model must cope with all elements even when the original dataset is given very sparse and also proposed a sampling based method. This approach samples only a part of the missing values and replaces them with negative values. Three kinds of sampling methods are proposed: User-oriented sampling, Item-oriented sampling and Uniform sampling.

© 2017, IJCSMC All Rights Reserved

User-oriented sampling assumes that the number of negative values hidden in missing values and the amount of past feedbacks given by a user are related. Item-oriented sampling assumes that the number of negative values hidden in missing values and the amount of past feedbacks given for an item are related. In this work, this method is combined with our proposals. Also, proposed an extension of the weighted Matrix Factorization model. They incorporate similarity matrices over items and users to the weighted Matrix Factorization model. Finally proposed the joint model of the Matrix Factorization model and non-negative Matrix Factorization model which classify missing values into positive and negative feedbacks. Packet and Koenigstein proposed a bayesian generative model which predicts the probability with which missing values are converted to negative feedbacks. There exist other approaches that use the auxiliary data to treat missing values. Limitations of existing approaches as follows:

- Only analyzed ratings from user reviews
- Fake reviews can't be analyzed by existing work
- Unable to identify genuine reviews
- Handle only limited number of product reviews

IV. PROPOSED FRAMEWORK

A recommendation system has been implemented based on hybrid approach of stochastic learning and context based engine. The proposed framework have tried to combine the existing algorithms for recommendation to come up with a hybrid one. It improves the performance by overcoming the drawbacks of traditional recommendation systems. Recommender systems being a part of information filtering system are used to forecast the bias or ratings the user tends to give for an item. Among different kinds of recommendation approaches, collaborative filtering technique has a very high popularity because of their effectiveness.

These traditional collaborative filtering systems can even work very effectively and can produce standard recommendations, even for wide ranging problems. For item based on their neighbor's preferences entropy based technique creates better suggestions than others. Whereas other techniques like content based suffers from poor accuracy, scalability, data sparsity and big-error prediction. To find these possibilities the proposed system have used user-based collaborative filtering approach. In this Item based collaborative filtering technique first examine the User item rating matrix and identify the relationships among various items, and then use these relationships in order to compute the recommendations for the user. Then using cosine similarity which is a similarity weight is going to play an important role in the collaborative item based filtering approach and hence in order to maintain or select the trustable users from the given set of user. Hence they give us a method to increase or decrease the significance of a particular user or item. The present methodology using adjusted similarity for computation of similar weights of items.

Sentiment analysis of natural language texts is a large and growing field. Sentiment analysis or Opinion Mining is the computational treatment of opinions and subjectivity of text. Sentiment analysis is an Information Extraction task that intends to acquire writer's feelings expressed in positive or negative comments, after analyzing his documents. The term 'Presence' is more important to sentiment analysis then term 'Frequency' which was earlier used for traditional information retrieval. It has also been reported that unigrams surpass bigrams for classifying movie reviews by sentiment polarity. The proposed framework can implement stochastic learning algorithm to analyze reviews, ratings, and emoticons. The proposed work is described as follows:

Framework construction

E-commerce framework is used to buy the products in online to easy retrieval the mobile products. This module is used to create android and web site for recommending best mobiles in specific area. Admin is the responsibility for maintaining the all details in server and server can be design in server. There are two accounts such as admin and user account. Admin can login to the system and post item details with expiry dates. User can login to the mobile to choose the language and area. Then view the products with specified filter. This module is used to create web site buy or post products for users. Admin can login to the system and post products with features. User can login to the system to view product details.

Reviews collection

Admin collect reviews and have various types of reviews. Reviews may be rating reviews, text reviews and smileys reviews. All reviews are stored in database for future evaluation. Ratings, reviews and emoticons are stored in database. Rating, Reviews and Emoticons are the evaluation or assessment of something, in terms of quality (as with a critic rating a novel), quantity or some combination of both.

© 2017, IJCSMC All Rights Reserved

Sentiment analysis

Sentiment analysis refers to the use of natural language processing, text analysis, computational linguistics to systematically identify, extract, quantify, and study affective states and subjective information. Sentiment analysis is widely applied to voice of the customer materials such as reviews and ratings for applications that range from marketing to customer service to buy the products efficiently. Admin can analyze whether the product is positive or negative. In star rating, we can calculate star count values. In text reviews, extract keywords and matched with database. Then smileys reviews are calculated based positive and negative symbols.

Recommendation system

Recommender systems are a subclass of information filtering system that seek to predict the "rating" or "preference" that a user would give to an item. User can search the product in search bar and view the list of products based on price and review details. Implement the stochastic learning algorithm to classify the products such as positive or negative. Positive products are display in recommendation panel based on ratings and reviews. If the product has negative review means, automatically the positive products in recommendation panel.

Fake reviews monitoring

In this module, fake reviews are analyzed by admin. A media access control address (MAC address) of a computer is a unique identifier assigned to network interfaces for communications at the data link layer of a network segment. Admin can get user account details, Mobile address and Order id details. So one user can post one reviews that will be genuine reviews. The proposed framework is shown in figure 2.

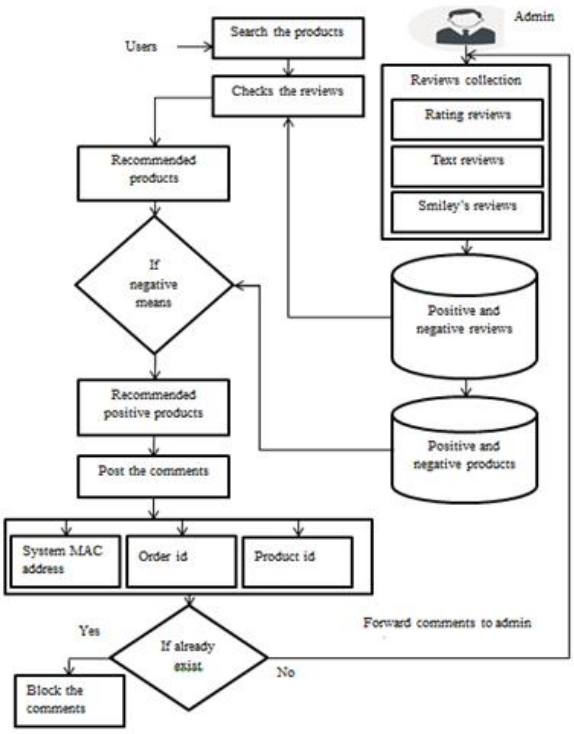

Figure 2: A **Proposed Framework for Sentiment Analysis in E-Commerce**

V. EXPERIMENTAL RESULTS

There are a number of evaluation matrices are used to evaluate the retrieval performance. For matric evaluation, the proposed framework using precision and recall. Where, precision measure the availability of relevant products from product database based on feedbacks and recall measure the availability of relevant products from the overall database

© 2017, IJCSMC All Rights Reserved

$$\text{Precision} = \frac{\text{No of relevant products extracted}}{\text{Total no of products extracted}}$$

$$\text{Recall} = \frac{\text{No of relevant products extracted}}{\text{Total no of products in database}}$$

To analyses the recommendation system, various types of recommendation algorithms are used. The proposed system took some products in e-commerce based sentiments after that it calculate the average precision and average recall for every sentiment. The following figure 3 depicts the proposed framework provided the better result in comparison of framework.

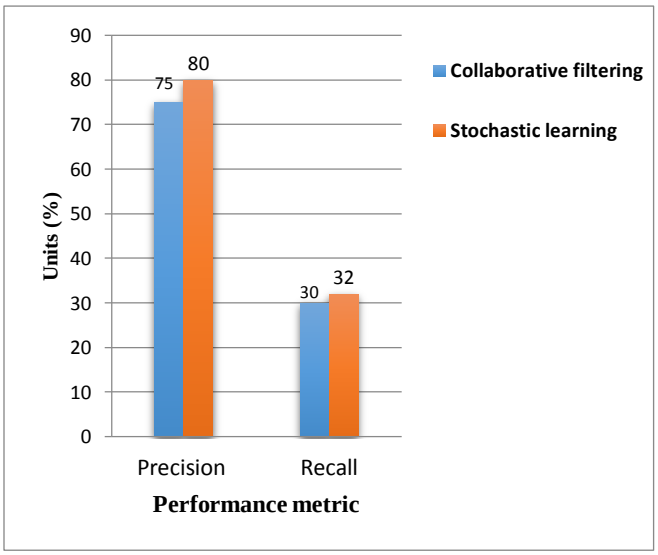

Figure 3: Performance graph

VI. CONCLUSION

In this paper, the proposed framework presented a novel implementation of a product recommendation system based on hybrid recommendation algorithm. The main advantages of this framework is to provide a visual organization of the data based on the underlying structure and a significant reduction in the size of the search space per result output. This framework also provide a simple method to search the products anywhere and anytime. Ratings, reviews and emoticons are analyzed and categorized as positive and negative sentiments. Search the products based on price based filtering and reviews based filtering. MAC based filtering approach can be used to avoid fake reviews. Supermarket can benefits because easy buying, easy transactions and to get more customers. Our method was evaluated against real user data collected through an online website, by using a subset of the movies liked by each user as input to the system. The current results are notably better than random approach. Hybrid Recommendations is one of the main modules of the system which helps overcome the drawbacks of the traditional Collaborative and Content Based Recommendations.

REFERENCES

[1] M. Jahrer, A. Toscher and R. Legenstein, "*Combining predictions for accurate recommender systems*," in KDD'10, pp. 693-702, 2010.

[2] Y. Zhang, B. Caoand D. Y. Yeung, "Multi-domain collaborative filtering," in Proc.UAI, pp. 725-732, 2010.

[3] R. Keshavan, A. Montanari, and S. Oh, "Matrix completion from noisy entries," Journal of Machine Learning Research, vol. 11, pp. 2057-2078, 2010.

© 2017, IJCSMC All Rights Reserved

[4] M. Harvey, M. Carman, I. Ruthven, and F. Crestani, "Bayesian latent variable models for collaborative item rating prediction," in CIKM'11, pp. 699-708, 2011.
[5] H. Ma, D. Zhou, C. Liu, M. Lyu, and I. King, "Recommender Systems with Social Regularization," in ACM WSDM, pp. 287-296, 2011.
[6] P. Bedi, H. Kaur, and S. Marwaha, "Trust based recommender system for semantic web," in IJCAI'07, pp. 2677-2682, 2007.
[7] P. Cui, F. Wang, S. Liu, M. Ou, S. Yang, and L. Sun, "Who should share what? item-level social influence prediction for users and posts ranking," in SIGIR, pp. 185-194, 2011.
[8] S. Scellato, A. Noulas, and C. Mascolo, "Exploiting Place Features in Link Prediction on Location-based Social Networks," in KDD'11, pp. 1046-1054, 2011.
[9] Z. Wang, L. Sun, W. Zhu, S. Yang, H. Li, and D. Wu, "Joint Social and Content Recommendation for User Generated Videos in Online Social Network," IEEE Trans. Multimedia, vol. 15, no. 3, pp. 698-709, 2013.
[10] Y. Chen, A. Cheng, and W. H. Hsu, "Travel Recommendation by Mining People Attributes and Travel Group Types from Community-Contributed Photos," IEEE Trans. Multimedia, vol. 15, no. 6, pp. 1283-1295, 2013.

© 2017, IJCSMC All Rights Reserved