



Evaluating the Performance of a Novel Rating System Based on Text Reviews

Vandurangi Chitti Babu & G.Sailaja

- #1 M.Tech Scholar, Department of Computer Science and Engineering,
Thandra Papraya Institute of Science and Technology,
Komatapally (Village), Bobbili (Mandalam)
Vizianagaram (District).
- #2 Assistant Professor, Department of Computer Science and Engineering,
Thandra Papraya Institute of Science and Technology,
Komatapally (Village), Bobbili (Mandalam)
Vizianagaram (District).

ABSTRACT

Now a day's sentiment analysis has become one of the fascinating domains in which each and every online shopping sites use this technique for gathering the product ratings or reviews based on various customers' sentiment. As we all know that in current days there were a lot of web sites evolved for providing feedback for the online products purchase. But there is no single accurate website which can provide the exact rating from the text reviews. For any web sites almost all the reviews are facing with overloading problem with continuous duplicate reviews that were posted by the same user. In this proposed paper, we mainly designed a novel sentiment-based rating prediction method (NRPS) to improve prediction accuracy in recommender

systems. As an extension we also included a concept like benchmark for the rated products based on text reviews. By conducting various experiments on our proposed model by taking a sample local website by taking some products into assumption, our simulation results show our NRPS greatly helps to improve the recommendation performance.

Key Words:

Rating, Recommendation,
Review System, Sentiment Based Rating.

1. INTRODUCTION

As we all know that, there is a lot of personal information that is available in online textual reviews. This information plays a very vital

role on decision making. For example, if we want to consider the best example for the textual reviews, the customer will decide what to buy if he or she sees valuable reviews posted by others, especially by the user's friends or friend of friends. Now a days we can also find out the rating stars along with text based reviews below the products that are available online. The customer will decide the product is good if it contains the highest rating stars along with number of good reviews are more, the same customer think that review is bad if the most number of text reviews are negative as well as less number of stars given for that appropriate product. We believe reviews and reviewers will do help to the rating prediction based on the idea that high-star ratings may greatly be attached with good reviews. Hence, how to mine reviews and the relation between reviewers in social networks has become an important issue in web mining, machine learning and natural language processing[1]-[3].

In addition to extracting user preferences, there is much work paying attention to the interpersonal interaction[4],[5]. Many approaches about the interpersonal influence in social networks have proved good performance in recommendation,

which can effectively solve the “cold start” problems. However, the existing approaches [6], [7], [8], [9], [10] mainly leverage product category information or tag information to study the interpersonal influence. These methods are all restricted on the structured data, which is not always available on some websites. However, user reviews can provide us ideas in mining interpersonal inference and user preferences. To address these problems, we propose a sentiment-based rating prediction method in the framework of matrix factorization. In our proposed work, we make use of social user's sentiment to infer ratings.

Figure. 1 is an example that illustrates our motivation. First, we extract product features from user reviews. Then, we find out the sentiment words, which are used to describe the product features. Besides, we leverage sentiment dictionaries to calculate sentiment of a specific user on an item/product. What is more, we combine social friend circle with sentiment to recommend. In Figure.1, the last user is interested in those product features, so based on the user reviews and the sentiment dictionaries, the last item will be recommended. Compared with previous work [8], [9], the main

difference is that: we use unstructured information to recommend instead of other structured social factors. Compared with [11], [12], [13], [14], the main difference is that: their work mainly focuses on classifying users into binary sentiment (i.e. positive or negative), and they do not go further in mining user's sentiment. In our paper, we not only mine social user's sentiment, but also explore interpersonal sentimental influence and item's reputation. Finally, we take all of them into the recommendation.

In that above figure we can clearly identify there are several products available in the site with a different choices like brand, quality and price. As we can see different professionals are available in the architecture, each and every professional will have a choice of choosing variety of products from the category, some will choose based on brand. Some will choose on color, price and quality with different types of reviews for each and every product what they purchase.

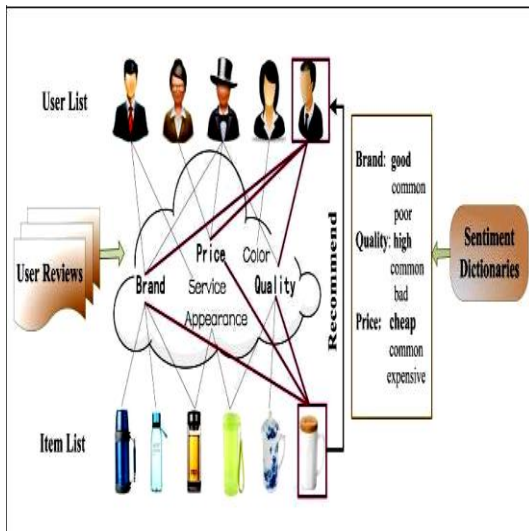


Figure 1. Represents the Working Flow of Sentiment Analysis System

From the figure 1, we can clearly identify the working flow of our proposed rating prediction system.

2. RELATED WORK

In this section we will find the related background work that was analyzed and studied in order to implement this current thesis. This section will describe the work that is related to the various data mining techniques that are available in the literature and we can discuss about them in detail in this section.

Various Data Mining Techniques

Data mining processes have required an integration of techniques from multiple disciplines such as statistics, machine learning, database technology, pattern recognition, neural networks, information retrieval and

spatial data analysis. The process of data mining involves a keen observation with a set of algorithms to accomplish different tasks. All these algorithms attempt to fit a model to the given data set. All the data mining algorithms are mainly used to examine the input data and then determine a model that is very closest to the characteristics of the data being examined.

Generally the data mining algorithms can be classified into different ways based on the following three things like:

- a) **Model Based:** The purpose of the algorithm is to fit a model to the required data.
- b) **Preference Based:** This is based on the preferences or criteria that are being used.
- c) **Search Based:** In this method, we mainly find out the search techniques that are being used.

Of all the above three technique, our proposed application comes under preference based and search based technique. As in our proposed application we mainly try to find out the preference of finding

which product is having good privilege and which one is having less privilege. So based on these preference the user can search the best product from the list of several products. Here we used advanced search technique for finding the products in which search can be done either by category name or sub category name or product name.

3. A NOVEL PROPOSED RATING PREDICTION SYSTEM (RPS) USING TEXTUAL SENTIMENT APPROACH

In this section, we mainly describe the proposed novel **RPS Method Using Textual Sentiment Approach** for identifying the sentiment or rating for the products based on social sentiments.

MAIN MOTIVATION

The main motivation for proposing this current approach is carried out in order to propose the current RPS approach is to find the effective clues from text based reviews and try to predict social user's ratings. In this paper, we mainly extract product features from user review corpus, and then we introduce the method of identifying social user's

sentiment. In addition, we describe the three sentimental factors. At last we fuse all of them into our sentiment-based rating prediction method (RPS). The following sub-sections describe more details about our approach.

PRODUCT FEATURES EXTRACTION

In this section first we try to propose on the extraction of product features which mainly describe about the issues that are related to the products. In this thesis, we extract product features from textual reviews using LDA [11]. We mainly want to get the product features including some named entities and some product/item/service attributes. LDA is a Bayesian model, which is utilized to model the relationship of reviews, topics and words. In Figure.2, the shaded variables indicate the observed variables and the unshaded variables indicate the latent variables. The arrow indicates a conditional dependency between the variables and plates represented by the box. The definition of terminologies in LDA model is described as:

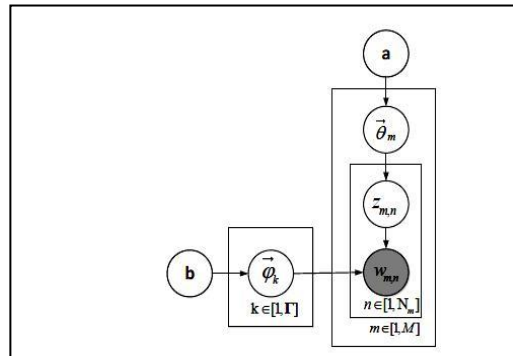


Figure 2. Shows the Graphical View of an LDA Method

From the above figure 2, the borders that are available mainly represents the replicates and the outer border clearly represents the user documents and the inner border that is available represent the repeated choice of words and topics that are available within the document.

- **V**: the vocabulary, it has Nd unique words. Each word is presented by the corresponding label $\{1, 2, \dots, Nd\}$.
- $w_i \in \{1, 2, \dots, Nd\}$: the word, each word of a review is mapped to **V** whose size is Nd through character matching.

<p>➤ d_m : the document/review of a user, it corresponds to a word set of the review. A user with only one document. All documents denote as $D = \{d_1, d_2, \dots, d_M\}$.</p>
<p>➤ Γ: the number of topics (const scalar).</p>
<p>➤ $\theta^{\rightarrow m}$: the multinomial distribution of topics specific to the document m. One proportion for each document, $\Theta = \{\theta^m\}_{m=1}^M$ ($M \times \Gamma$ matrix)</p>
<p>➤ $\phi^{\rightarrow k}$: the component for each topic, $\Phi = \{\phi^k\}_{k=1}^{\Gamma}$ ($\Gamma \times k$ matrix)</p>
<p>➤ z_m, n: the topic associated with the n-th token in the document m.</p>
<p>➤ a, b: Dirichlet priors to the multinomial distribution θ^m and ϕ^k.</p>

THE GENERATIVE PROCESS OF LDA

In this section we mainly discuss about the generative process of the LDA Approach. The input of LDA model is all users' document sets D , and we assign the number of topic Γ (we set 50 empirically). The output is the topic preference distribution for each user and a topic list, which contains at least 10 feature words under each topic. The generative process of LDA consists of three steps:

For each document d_j , we choose a dimensional Dirichlet random variable $\theta^m \sim \text{Dirichlet}(a)$.

➤ For each topic z_k , where $k \in [1, \Gamma]$, we choose $\phi^k \sim \text{Dirichlet}(b)$.

➤ For each topic z_k , the inference scheme is based upon the observation that:

$p(\theta, \Phi | D_{train}, a, b) = \sum_z p(\theta, \Phi | z, D_{train}, a, b) P(z, | D_{train}, a, b)$ (3). We obtain an approximate posterior on θ and Φ by using a Gibbs sampler to compute the sum over z .

- Repeating the process above and eventually we get the output of LDA.

4. IMPLEMENTATION MODULES

Implementation is the stage where the theoretical design is converted into programmatically manner. In this stage we will divide the application into a number of modules and then coded for deployment. We have implemented the proposed concept on Java programming language with JEE as the chosen language in order to show the performance this proposed NRPS for finding the sentiment of any products based on social sentiment generated by textual reviews. The front end of the application takes JSP, HTML and Java Beans and as a Back-End Data base we took My SQL data base along with a local product data set with some set of products which we use in general world.

The application is divided mainly into following 2 modules, in which each and every module has a set of various sub modules. They are as follows:

1. Admin Module
2. User Module

1. ADMIN MODULE

In this module, the admin has to login by using valid user name and password. After login successful he can do some operations such as add categories, add posts, list of all posts, list of all recommended posts, list of all

reviewed posts, list of users, list of all search history, update posts, lists of bad reviews by date wise, list of good reviews by date wise, list of all neutral reviews date wise, all friend request, all product reviews and rating results. Here the admin is the one who is maintaining this product data base by inserting a set of items inside the database.

SUB MODULES

The following are the sub-modules that are available by the admin. Now let us discuss about them in detail as follows:

i) ADD PRODUCTS POSTS

In this module, the admin can add the post by including product name, price, description and corresponding product image. Here the products can be shopping, general, stationary or any house holds products. Here the admin maintain a validation steps for developing this sub-module like all the fields should be filled during the insertion of products. If the admin miss any field during the insertion, the product will not be added successfully. Here the validation is written for inserting all fields properly during product insertion and then only the



product is added successfully inside the product database.

ii) VIEW ALL POSTS

In this module, the admin can view the post by searching keyword and can get all the information about the product like product name, price, description and corresponding product image. Here the admin after inserting the products based on category ,he can view the set of products that are inserted in the database. All the products are displayed based on category wise and sub-category wise and the user can search the product either with product names or category name or by sub-category wise.

iii) INVENT SENTIMENT ANALYSIS

The admin can analyze the sentiment based on products from positive sentiment words, products from negative sentiment words, products from neutral sentiment words and View Products Rating based on sentiment words. Here the sentiment analysis is nothing but classifying the different products based on the type of sentiment given by the various users. Actually for each and every product the sentiment /feed back is given by a set of users who try to view the post ,based on the

keywords they enter in the feedback, the sentiment analysis is done.

2. USER MODULE

In this module, there are n numbers of users are present. User should register before doing some operations. After registration successful he has to login by using authorized user name and password. Login successful he will do some operations like view user details, search for products posts, view my search history, view recommended, search for top N posts and logout.

SUB MODULE:

i) SEARCHES FOR GOOD REVIEWS AND BAD REVIEW

In this module, user searches for reviews for the post and can get the following information like product name, price, description and corresponding product image. The user can recommend the product and can give review using sentiment words (such as good or bad product like that) based on brand, Quality, Price.

ii) SEARCHES FOR PRODUCTS POSTS



In this module, user searches for the products or posts after they get login into their account. He will try to enter the search the products based on any field like product name, product category name or product sub category name and so on. Once if the product is matched with any of the field then only the product information is displayed on their login account. If none of the products are matched for the search keyword, then he will get an error like products not matched.

iii) USER RECOMMEND PRODUCTS

In this module, user searches for the products or posts after they get login into their account. Once after searching any product in his login, he may want to recommend the same for others who registered in the same account. For recommendation he can either recommend individually in his friends list or he can recommend at a time for all friends. Along with recommendation he can also give some valuable comments for the product which will give some more detail information about the product to the end users

iv) USER SEARCH FOR TOP-N POSTS

In this module, user searches for the products or posts which are almost top rated by the various users. Here if he specify the top K value from his login, he will get the top rated and accepted products from the database. The top rated will be decided for any product based on the number of positive reviews given by various users for that appropriate product. This function is applied for all the products that are inserted or available in the database, this is not limited for a individual product that is available in the database.

5. CONCLUSION

In this paper, we for the first time designed and developed a novel rating prediction system (RPS) based on user social sentiment on textual reviews. Here we took a sample product data set as input and try to propose the proposed system and we observed the performance of our current approach based on user sentiment reviews. Also as an extension we also implemented a concept like bench mark based on the percentage values that are collected from each and every individual category. By conducting various experiments on our proposed method, we finally came to an conclusion that our proposed approach

is best in providing a accurate and efficient feedback for the products under various input parameters.

6. REFERENCES

[1] M. Bellare, C. Namprepre, and G. Neven. Security proofs for identity-based identification and signature schemes. *J. Cryptology*,22(1):1–61, 2009.

[2] W. Sun, S. Yu, W. Lou, Y. T. Hou, and H. Li, “Protecting your right: Attribute-based keyword search with fine-grained owner-enforced search authorization in the cloud,” in *Proc. IEEE Conf. Comput. Commun.*, 2014, pp. 226–234.

[3] S. Yu, C. Wang, K. Ren, and W. Lou, “Achieving secure, scalable, and fine-grained data access control in cloud computing,” in *Proc. IEEE Conf. Comput. Commun.*, 2010, pp. 1–9.

[4] M. Li, S. Yu, Y. Zheng, K. Ren, and W. Lou, “Scalable and secure sharing of personal health records in cloud computing using attribute-based encryption,” *IEEE Trans. Parallel Distrib. Syst.*,vol. 24, no. 1, pp. 131–143, Jan. 2013.

[5] S. Kamara and K. Lauter, “Cryptographic cloud storage,” in *Proc.14th Int. Conf. Financial*

Cryptography, Data Security, 2010, pp. 136–149.

[6] H. Feng, and X. Qian, “Recommendation via user’s personality and social contextual,” in *Proc. 22nd ACM international conference on information & knowledge management*. 2013, pp. 1521-1524.

[7] K.H. L. Tso-Sutter, L. B. Marinho, L. Schmidt-Thieme, “Tag-aware recommender systems by fusion of collaborative filtering algorithms,” in *Proceedings of the 2008 ACM symposium on Applied computing*, 2008, pp. 1995-1999.

[8] F. Li, N. Liu, H. Jin, K. Zhao, Q. Yang, X. Zhu, “Incorporating reviewer and product information for review rating prediction,” in *Proceedings of the Twenty-Second international joint conference on Artificial Intelligence*, 2011, pp. 1820-1825.

[9] 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.

[10] 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.

[11] X. Qian, H. Feng, G. Zhao, and T. Mei, “Personalized recommendation combining user interest and social circle,” IEEE Trans. Knowledge and data engineering. 2014, pp. 1763-1777.

[12] 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.

7. ABOUT THE AUTHORS



VANDRANGI CHITTI BABU is currently pursuing his 2 years M.Tech in Department of Computer Science and Engineering at Thandra Papraya Institute of Science and Technology, Komatapally (Village), Bobbili (Mandalam), Vizianagaram (District).

His area of interest includes in Mining, Network, Web Technologies, Computer Networks, HCI.



G.SAILAJA,

M.Tech, B.Tech, received her B.Tech from Rajah AL-AMAN College of Engineering affiliated to J.N.T.U Kakinada and Received the M.Tech Computer Science and Engineering with Specialization in Computer Science and Technology from Andhra University, Visakhapatnam. Currently working as Assistant professor in Thandra Paparaya Institute of Science & Technology (TPIST) Komatapalli Bobbili, Vizianagaram Dist, Andhra Pradesh from Dec 2014 onwards. Her Area Of interest including C, C++, and JAVA are web technologies, Computer networks, compiler design, flat, HCI, mc, and CNS.