



ScienceDirect

journal homepage: www.elsevier.com/pisc

Feature extraction and analysis of online reviews for the recommendation of books using opinion mining technique[☆]



Shahab Saquib Sohail^{a,*}, Jamshed Siddiqui^a, Rashid Ali^b

^a Department of Computer Sciences, Aligarh Muslim University, India

^b Department of Computer Engineering, Aligarh Muslim University, India

Received 15 February 2016; accepted 8 June 2016

Available online 18 July 2016

KEYWORDS

Book recommendation;
Customer review;
Opinion mining;
Feature extraction;
Online reviews;
Opinion mining

Summary The customer's review plays an important role in deciding the purchasing behaviour for online shopping as a customer prefers to get the opinion of other customers by observing their opinion through online products' reviews, blogs and social networking sites, etc. The customer's reviews reflect the customer's sentiments and have a substantial significance for the products being sold online including electronic gadgets, movies, house hold appliances and books. Hence, extracting the exact features of the products by analyzing the text of reviews requires a lot of efforts and human intelligence. In this paper we intend to analyze the online reviews available for books and extract book-features from the reviews using human intelligence. We have proposed a technique to categorize the features of books from the reviews of the customers. The extracted features may help in deciding the books to be recommended for readers. The ultimate goal of the work is to fulfil the requirement of the user and provide them their desired books. Thus, we have evaluated our categorization method by users themselves, and surveyed qualified persons for the concerned books. The survey results show high precision of the features categorized which clearly indicates that proposed method is very useful and appealing. The proposed technique may help in recommending the best books for concerned people and may also be generalized to recommend any product to the users.

© 2016 Published by Elsevier GmbH. This is an open access article under the CC BY-NC-ND license (<http://creativecommons.org/licenses/by-nc-nd/4.0/>).

Introduction

It is a common practice in the technologies being used for recommender systems to use customer's opinion and experiences for recommending desired products to the users (Aciar et al., 2006). The customers' opinions are generally available in an unstructured and unguided format which requires an extra effort in processing these reviews to get some

[☆] This article belongs to the special issue on Engineering and Material Sciences.

* Corresponding author.

E-mail addresses: shahabsaquibsohail@gmail.com (S.S. Sohail), jamshed_faiza@rediffmail.com (J. Siddiqui), rashidaliamu@rediffmail.com (R. Ali).

fruitful conclusion from it. Opinion mining has been used for the extraction of these reviews widespread. Opinion mining is basically concerned with extracting the opinions of the users and analyzing them to draw a meaningful conclusion from their respective ideas in terms of reviews and feedbacks (Fang et al., 2012). It has emerged significantly in past few years for various applications such as product recommendation, business analytics, information retrieval, etc.

The online reviews which are available at various online shopping sites are very helpful for the customers who seek for other's experiences and feedback about a particular product. A subjective conclusion from the user's review could help the customers to find a product of their choice or in deciding about a particular product whether to purchase it or leave it.

Books are very common item amongst other products which are being purchased frequently over the Internet using online shopping (Sohail et al., 2014a, 2015). In this paper we intend to present a method which extracts the features from online reviews and analyses those features to provide a platform for the users for purchasing online books. We have extracted the reviews of the customers and broadly categorized the book features into seven (7) categories. These features are based on the reviews collected from users, available at various online book stores like <http://www.amazon.com>, <http://www.goodreads.com/book/show/>, etc. The categorized features will help the users to find the suitable books for them.

The features are evaluated by the users explicitly, using the precision as an evaluation metric to validate the features extraction and analysis process. Each feature has a precision more than 75%; even one of the features has 100% precision. However, the average precision of all the categorized features is 90%.

Rest of the paper is organized as follows. In the section "Feature extraction and analysis", feature extraction process is explained with suitable figures and examples. The need of human intelligence and categorization of features are also discussed. The section "Performance evaluation" describes the performance evaluation parameter and performance of the feature extraction technique. In the section "Conclusion and future work", we summarize the conclusion and have given future directions.

Feature extraction and analysis

In this section we have described the steps of our procedure and the need of the human intelligence. Also, the features categorization process with the details of each feature is illustrated in subsequent sub section.

Human intelligence in feature extraction from reviews

Usually reviews from the customers contain some positive and negative words, and by analyzing the respective sense of a word researchers have concluded about a product quality and their features (Huang et al., 2008; Hu and Liu, 2004). Whereas in some cases, situation differ and a person says his words in a manner where negative words may be used in the favour of a product and not necessarily in a negative

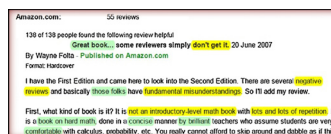


Figure 1 A sample customer review on books from amazon.com.

sense. We have shown a review example in Fig. 1, taken from biggest online book retailer, amazon.com.

The texts are highlighted in green and yellow colours for their positive and negative sense respectively. If we analyze, we find equal number of positive and negative words that compel to infer a neutral opinion about the particular book. However, this is not the case here. The words with a negative sense are used for the reviewers who do not review the book appropriately and not for the book. This is why we need human intelligence to infer correct results from customer's review as they are human emotions.

Categorization of features

In Sohail et al. (2013) authors have used the technique for categorization of features in book recommendation. But the author did not apply any validation of the features. We have modified the features and analyzed it again and categorized seven features from a different way to provide a better understanding for the book customers. The features are namely, (1) Frequency of Occurrence in Search Engine Results Page (SERP), (2) Useful Content, (3) Extraneous Content, (4) Sufficient Material, (5) Physical Attributes, (6) Market Availability and (7) Price.

The first feature deals with frequency of occurrence of a book in SERP. A good and most visited book should eventually come top in search engine ranking and it shows its importance. The second, third and fourth features i.e. useful content, extraneous content and sufficient material are concerned with the content of the books. Usually a review from user contains about a book whether it is useful or it has less examples though written well, etc. these features are grouped separately in respective features. Physical attributes are concerned with the quality of books cover, hard copy page quality, etc. Market availability is the availability of a book online, availability of its review and there availability in the market as mentioned in user's reviews. The last feature categorized is price. Whether price of a book, costly or cheaper, has a significance in the eye of a customer or not?

Performance evaluation

Various metrics have been used for evaluation of features extraction and user feedback (Shani and Gunawardana, 2011). Authors in Sohail et al. (2014b, 2014c) have used precision as a major evaluation metric. We define precision as follows:

precision

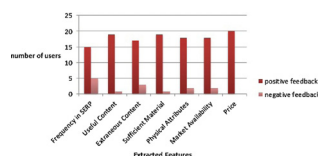
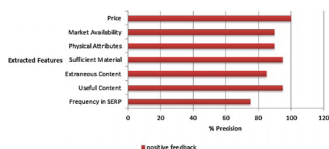
$$= \frac{\text{number of users with positive feedback for categorized feature}}{\text{total number of users involved in the feedback}} \quad (1)$$

Table 1 User feedback table for categorized features.

	Frequency in SERP	Useful content	Extraneous content	Sufficient material	Physical attributes	Market availability	Price
Number of positive feedback	15	19	17	19	18	18	20
Number of negative feedback	5	1	3	1	2	2	0

Table 2 Extracted feature and corresponding percentage-precision.

	Frequency in SERP	Useful content	Extraneous content	Sufficient material	Physical attributes	Market availability	Price
% precision	75	95	85	95	90	90	100

**Figure 2** User positive–negative feedback for extracted features.**Figure 3** % precision of extracted features.

We have presented all the seven features to 20 postgraduate students and academic personalities. The users were asked to give their feedback for the categorized features. We have obtained positive feedback for all the features from majority of the users. Features 1 and 3 were rated positive by 15 and 17 students respectively. Features 2 and 4 were rated positive by 19 users; feature 5 and 6 were rated positive by 18 participants. The seventh feature, ‘price’ were rated positive by all 20 users. Hence, by using equation 1, we can easily observe that the average precision for categorized feature is 90%. The total number of positive and negative feedback is tabulated and shown in Table 1. A pictorial representation of the value is depicted in Fig. 2. The extracted features – precision value is given in Table 2 and a corresponding graph is shown in Fig. 3.

Conclusion and future work

In this paper we have presented a feature extraction and analysis method. We have used human intelligence and categorized book-features from online reviews which may help users in finding the books of their choice. The method has been evaluated by using precision as an evaluation metric. The average precision of all the categorized features is 90%. Further, no feature has value of precision less than .75 i.e. at least 75% persons have positive feedback for each feature individually. Also, one of the features has 100% precision. The high value of the precision indicates the accuracy of the method adopted.

In future, we can enlarge the number of users. Also, there should be some automatic method that can recommend

books based on these features. One can use some mining technique or soft computing methods over these features to recommend books for users.

Conflict of interest

None.

Funding

None.

References

- Aciar, S., et al.,2006. Recommender system based on consumer product reviews. In: Proceedings of the 2006 IEEE/WIC/ACM International Conference on Web Intelligence. IEEE Computer Society.
- Fang, Y., et al.,2012. Mining contrastive opinions on political texts using cross-perspective topic model. In: Proceedings of the Fifth ACM International Conference on Web Search and Data Mining. ACM.
- Hu, M., Liu, B., 2004. Mining and summarizing customer reviews. In: KDD'04.
- Huang, W., et al., 2008. Analysis of the user behavior and opinion classification based on the BBS. Appl. Math. Comput. 205 (2), 668–676.
- Shani, G., Gunawardana, A., 2011. Evaluating Recommendation Systems. Recommender Systems Handbook. Springer US, pp. 257–297.
- Sohail, S.S., Siddiqui, J., Ali, R.,2013. Book recommendation system using opinion mining technique. In: 2013 International Conference on Advances in Computing, Communications and Informatics (ICACCI). IEEE.
- Sohail, S.S., Siddiqui, J., Ali, R.,2014a. Ordered ranked weighted aggregation based book recommendation technique: a link mining approach. In: 2014 14th International Conference on Hybrid Intelligent Systems (HIS). IEEE.
- Sohail, S.S., Siddiqui, J., Ali, R., 2014b. User feedback scoring and evaluation of a product recommendation system. In: Proceeding of IC3, pp. 525–530.
- Sohail, S.S., Siddiqui, J., Ali, R., 2014c. User feedback based evaluation of a product recommendation system using rank aggregation method. Advances in Intelligent Informatics, Advances in Intelligent Systems and Computing, vol. 320., pp. 349–358.
- Sohail, S.S., Siddiqui, J., Ali, R., 2015. OWA based book recommendation technique. Procedia Comput. Sci. 62, 126–133.