A Comparative Study of Feature Extraction Methods from User Reviews for Recommender Systems

Pradnya Bhagat
Department of Computer Science and Technology, Goa University
Goa, India
pradnyabhagat91@gmail.com

Jyoti D. Pawar
Department of Computer Science and Technology, Goa University
Goa, India
jdp@unigoa.in

ABSTRACT

The Recommender systems technology is being massively exploited by e-commerce giants to enhance the shopping experience of their clients which in turn helps in improving the sales of the company. Most of the recommender systems in use today are based on Collaborative Filtering (CF) in which the known preferences of a group of users are used to make recommendations or predictions for the unknown preferences of other users. Although these ratings communicate about the quality of the product, they almost most of the times fail to express the reason behind people believing the product to be of a particular quality. This information can be inferred if analyze the information rich textual reviews written by the users.

In the current work, an attempt is made to study and implement various methods described in literature, to mine the product features from the user reviews associated with the product. A comparative study is presented at the end to appreciate the performance of the methods.

CCS CONCEPTS

• **Information systems → Recommender systems**; • **Computing methodologies → Natural language processing**;

KEYWORDS

Recommender Systems, Collaborative Filtering, User Reviews, Product Features, POS Tagging, Apriori, Latent Dirichlet Allocation

ACM Reference Format:
Pradnya Bhagat and Jyoti D. Pawar. 2018. A Comparative Study of Feature Extraction Methods from User Reviews for Recommender Systems. In *CoDS-COMAD '18: The ACM India Joint International Conference on Data Science & Management of Data, January 11–13, 2018, Goa, India.* ACM, New York, NY, USA, 4 pages. https://doi.org/10.1145/3152494.3167982

1 INTRODUCTION

The 1990s decade saw the evolution of web as a social networking platform where people could communicate with each other and express their opinions publicly on a global scale. Businesses and individuals started taking advantage from this technology by being able to connect to potential customers throughout the world. This has led to the emergence of e-commerce websites that have provided a platform to thousands of vendors to expand their business across the globe.

Although having many advantages, it is still not possible to get tangible experience for the products available on these websites. It may become a challenging task to validate the description and quality of the features through the descriptions made available by the vendor of the product. As a result, e-commerce websites have made it possible for the customers to share their experiences about the products with other customers to help them in making wise decisions. CF has proved to be the most successful technology in the development of recommender systems till date. Most of the research work on CF focuses on the explicit ratings specified by the users (Ex: 1- 5 stars), or implicit indications (purchases or click-throughs). Though these ratings indicate the quality of the product, they fail to quantify the reason for the product achieving that particular quality. In other words, there is no information available, presenting the different features of the product and the quality of these features.

To overcome these limitations, a better methodology can be developed if we look beyond the star ratings of the products and take into consideration the textual reviews written against every star rating by the customers. The reviews written by the customers are rich sources of information about the features of a product and their quality as perceived by the users.

In our work, an attempt is made to study the reviews written by the customers on an e-commerce website. With respect to this work, our scope is limited to study and implementation of different methods to automatically extract product features from the text reviews.

2 LITERATURE SURVEY

A significant amount of work has been done in recent years to extract product features from the textual reviews. [4] presents an approach to extract single noun and bi-gram features from user reviews using a combination of Natural Language Processing (NLP) and statistical methods. The approach assumes that the bigram topics can either be made up of Noun-Noun (NN) pairs or Adjective-Noun (AN) pairs. [1] presents an Apriori algorithm to find frequent itemsets from a transaction dataset. The approach has been adopted greatly in literature[2] to discover important topics (features) from text documents. The approach make a further assumption that if some words repeatedly occur together or in close proximity to one another in review sentences, that means they together form some important feature about that product. Hence, the algorithm is able to find multiword features without imposing any limitation on the

CoDS-COMAD '18, January 11–13, 2018, Goa, India

ACM ISBN 978-1-4503-6341-9/18/01…$15.00
https://doi.org/10.1145/3152494.3167982

number of words permitted in a feature. [3] uses an unsupervised technique of Latent Dirichlet Allocation (LDA) for topic extraction. The method is able to extract the main topics and the corresponding important words from the reviews. [7] presents a probabilistic approach for mining user preferences from reviews and mapping them onto numerical ratings based on Naive Bayes classifier. [11] attempts a statistical approach to identify polarity of nouns where no sentiment word is explicitly associated with the nouns. [8] proposes a domain independent approach to predict intensity of the sentiments expressed.

3 EXPERIMENTAL AND COMPUTATIONAL DETAILS

The experimental work carried out consists of studying various methods described in the literature to extract product features from a corpus of text reviews and making a comparison based on the features retrieved. The dataset [5] used in the experiments is sourced from Amazon.com which is limited to the category Mobile Cell Phones and Other Related Accessories.

3.1 Feature Extraction using Noun Occurrence (FENO)

An analysis of the dataset unveiled that most of the product features occur as nouns in the review dataset. Hence, as a preliminary method we consider all the nouns as the features in the dataset. The challenge here is, along with the product *feature nouns*, there are many other nouns that occur in the product reviews which in no way represent product features *(non-feature nouns)*. For example, if we consider the following sentence:

Example 1: *My family loved the look of the cellphone.*

After Parts of Speech (POS) [9] [10] tagging we get:

My_PRP family_NN loved_VBD the_DT look_NN of_IN the_DT cellphone_NN.

We can observe that, there are three nouns in the given sentence; *family*, *look* and *cellphone*. While we want to retain *look* and *cellphone* in our result set, since they form the features of a particular item belonging to Mobile phones and accessories category, the noun *family* clearly does not constitute as a product feature in the given dataset. Further observation of the dataset revealed that frequency of occurrence of *feature nouns* is considerably more than the frequency of occurrence of *non-feature nouns*. Hence, we eliminate the *non-feature nouns* by retaining only those nouns which appear more number of times than a specified threshold. As the final result set, we extract the top 25 nouns from the generated nouns based on the occurrence frequency.

3.2 Single Word Feature Extraction Using Occurrence Patterns (SW-FEOP)

It is observed that most of the *feature nouns* occur in close proximity to sentiment words [4]. This pattern is not followed by non-feature nouns. For example, consider a sentence:

Example 2: *My friend suggested me to buy this awesome phone because it has an excellent camera.*

After POS tagging, we get:

My_PRPS friend_NN suggested_VBD me_PRP to_TO buy_VB this_DT awesome_JJ phone_NN because_IN it_PRP has_VBZ an_DT excellent_JJ camera_NN.

The nouns occurring in the above sentence are; *friend*, *phone* and *camera*. As can be seen, *phone* has an adjective awesome associated with it and *camera* has an adjective excellent associated with it. Since the reviewer wants to describe the phone, he will use some sentiment words to express his opinions about the features of the product. On the other hand, if words like *friend*, *neighbor* or *relative* occur also in the review, there will hardly be any sentiment words associated with them. This information can be utilized to differentiate *feature nouns* from *non-feature nouns*. Hence, in this method only the nouns which are associated with sentiment words in the reviews are extracted. The top 25 most occurring nouns from the results generated are considered in the final result set.

3.3 Bi-gram Feature Extraction Using Occurrence Patterns (B-FEOP)

This method tries to extract bi-word features from the review set [4]. Consider the following two sentences as example:

Example 3: *The camera mode of the mobile is good.*
The front camera of the mobile is good.

POS tagging of these sentences gives us:

The_DT camera_NN mode_NN of_IN the_DT mobile_NN is_VBZ good_JJ.
The_DT front_JJ camera_NN of_IN the_DT mobile_NN is_VBZ good_JJ.

As can be seen in the first sentence, *camera_NN mode_NN* forms one topic, which is formed by a noun followed by a noun. In the second sentence, the words *front_JJ camera_NN* form one topic which is formed by an adjective and a noun occurring consecutively. Hence, to produce a set of bi-gram topics, all bi-grams from the global review set which conform to one of following basic POS co-location patterns are extracted:

(1) A noun followed by a noun (NN) such as *camera mode.*
(2) An adjective followed by a noun (AN) such as *front camera.*

There are candidate topics that need to be filtered out to avoid including ANs that are actually opinionated single-noun topics; for example, *excellent camera* also forms an adjective-noun pair, but is a single-noun topic (*camera*) and not a bi-gram topic. To achieve this, the bi-grams whose adjective is found to be a sentiment word (*e.g. excellent, good, great, lovely, terrible, horrible etc.*) are excluded using an English opinion lexicon [6].

3.4 Feature Extraction using Frequent Itemset Generation (FEFIG)

The method attempts to extract topics using Apriori frequent itemset generation algorithm [1]. The algorithm works in two steps:

(1) In the first step, it finds all frequent itemsets from the transactions that satisfy a user specified minimum support.
(2) In the second step, it generates rules from the discovered frequent itemsets.

To generate topics from the reviews, we break down the reviews into sentences and consider every sentence as one transaction. Next, after applying the pre-processing steps, we keep only the nouns and the adjectives from every sentence, since nouns and adjectives are observed to be the terms representing features of the product. The assumption is such that, the features from all the nouns and adjectives will be the most frequently occurring terms (items) and hence will have higher support. So, to find the frequently occurring terms, we need only the first step of the Apriori algorithm, i.e. to find the frequent itemsets which are candidate features. The results of applying the Apriori algorithms on our dataset is given below. The support threshold applied is 2% while generating the results and the top 25 item sets are extracted as the final result set.

3.5 Feature Extraction using Latent Dirichlet Allocation (FELDA)

In this method the statistical model Latent Dirichlet Allocation [3] is used in which a document is considered to have a set of topics. The model represents documents as mixtures of topics that are made up of words with certain probabilities.

As it is seen in the above methods, adjectives and nouns are the only parts of speech that constitute to product features in the dataset. Hence, to get better results, only the adjectives and nouns from reviews are retained and LDA is applied. Since in our dataset we already know the topic and are only interested in finding important words constituting that topic, we set the number of topics parameter to one.

4 RESULTS AND DISCUSSIONS

Table 1 displays the results of experimented methods on the stated dataset[5]. Only the most frequently occurring top 25 features obtained using the experimented methods are considered for the comparative study.

Manual evaluation is be used to evaluate the success rate of the methods. The features obtained using the experimented methods are reviewed manually and are divided into two categories; features that constitute to product features and features obtained that do not contribute to the product features in the Cell Phones and Accessories category. The success rate of the methods is explained with the help of a graph in Figure 1.

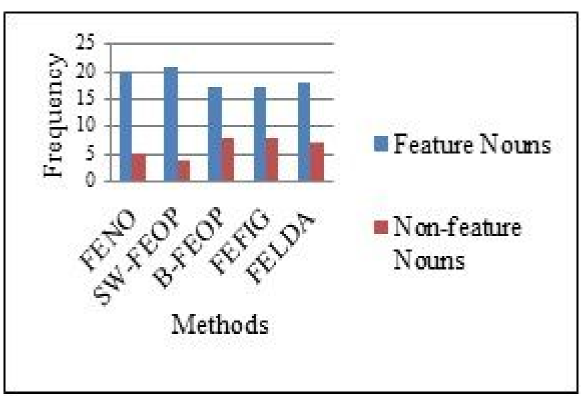

Figure 1: Performance of methods experimented

Table 1: Features obtained using experimented methods

	FENO	SW-FEOP	B-FEOP	FEFIG	FELDA
1	battery	battery	battery life	battery	battery
2	button	bit	battery pack	button	cable
3	cable	case	belt clip	cable	car
4	car	charge	car charger	case	case
5	case	charger	cell phone	case phone	charge
6	charge	cover	customer service	charger	charger
7	charger	deal	external battery	device	device
8	color	design	few days	easy	easy
9	cover	device	galaxy note	good	good
10	day	feature	galaxy s3	great	great
11	device	fit	galaxy s4	iphone	iphone
12	headset	job	home button	little	little
13	iphone	part	iphone 4s	nice	nice
14	phone	phone	new trent	other	phone
15	port	plastic	only thing	phone	port
16	power	price	phone case	power	power
17	price	product	power bank	price	price
18	problem	protector	power button	product	product
19	product	quality	same time	protector	protector
20	protector	review	samsung galaxy	quality	quality
21	quality	screen	screen protector	screen	review
22	review	side	sound quality	screen protector	screen
23	screen	thing	usb cable	thing	speaker
24	thing	time	usb port	time	time
25	time	way	wall charger	usb	usb

5 CONCLUSION

The work compares performance of five feature extraction methods on a real world dataset [5]. As can be concluded, FENO being the simplest detects only single word features. Being a basic method, it does not guarantee high quality results. The second method SW-FEOP, tries to improve upon the first method by adding an additional constraint of a sentiment word being associated with the noun. As can be seen from Figure 1, this method delivers better performance compared to all other methods. The third method B-FEOP tries to find multi-word features by considering NN and AN pairs. It identifies that a product feature is mostly preceded by an adjective which may or may not be a sentiment word. FEFIG is based on Apriori algorithm and is able to find multiword features without any length limitation. Since Apriori algorithm makes multiple passes on the data to find the itemsets, this method is considerably slower than all other methods. FELDA tries to find topics using LDA. Since it considers every word as an independent entity, we get only single word topics using this method.

The future work plan consists of improving upon the existing feature extraction methods. The sentiments corresponding to features and the intensity of the sentiments associated also needs to be studied. Based on this, we propose to improve upon the existing recommender system algorithms by adding the dimension of context to the recommendation algorithms.

REFERENCES

[1] Rakesh Agrawal and Ramakrishnan Srikant. 1994. Fast Algorithms for Mining Association Rules. *Proceedings of the 20th VLDB Conference Santiago, Chile* (1994).

[2] Ruihai Dong, Kevin McCarthy, Michael P. O'Mahony, Markus Schaal, and Barry Smyth. 2012. Towards an Intelligent Reviewer's Assistant: Recommending Topics to Help Users to Write Better Product Reviews. *IUI'12, Lisbon, Portugal* (2012).

[3] Ruihai Dong, Markus Schaal, Kevin McCarthy Michael P. O'Mahony, and Barry Smyth. 2012. Unsupervised Topic Extraction for the Reviewer's Assistant. *Springer-Verlag London* (2012).

[4] Ruihai Dong, Markus Schaal, Michael P. O'Mahony, and Barry Smyth. 2013. Topic Extraction from Online Reviews for Classification and Recommendation. *Proceedings of the Twenty-Third International Joint Conference on Artificial Intelligence* (2013).

[5] Ruining He and Julian McAuley. 2016. Ups and downs: Modeling the visual evolution of fashion trends with oneclass collaborative filtering. *WWW* (2016).

[6] Minqing Hu and Bing Liu. 2004. Mining and Summarizing Customer Reviews. *Proceedings of the ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD-2004)* (2004).

[7] Cane Wing ki Leung, Stephen Chi-fai, Chan Fu-lai, and Chung Grace Ngai. 2011. A probabilistic rating inference framework for mining user preferences from reviews. *World Wide Web* (2011).

[8] Raksha Sharma, Mohit Gupta, Astha Agarwal, and Pushpak Bhattacharyya. 2015. Adjective Intensity and Sentiment Analysis. *Proceedings of the 2015 Conference on Empirical Methods in Natural Language Processing* (2015).

[9] Ann Taylor, Mitchell Marcus, and Beatrice Santorini. 2003. The Penn Treebank: An Overview. *In: Abeille A. (eds) Treebanks. Text, Speech and Language Technology, vol 20. Springer, Dordrecht* (2003).

[10] Kristina Toutanova, Dan Klein, Christopher Manning, , and Yoram Singer. 2003. Feature-Rich Part-of-Speech Tagging with a Cyclic Dependency Network. *In Proceedings of HLTNAACL* (2003).

[11] Lei Zhang and Bing Liu. 2011. Identifying Noun Product Features that Imply Opinions. *Proceedings of the 49th Annual Meeting of the Association for Computational Linguistics:shortpapers* (2011).