International Journal of Emerging Technology in Computer Science & Electronics (IJETCSE)
ISSN: 0976-1353 Volume 13 Issue 4 –MARCH 2015.

Ranking Of Products Using Opinion Mining On Authentic Reviews

Barath Sriram R[1], Abhishek Balaji[2], Dhanya R[3],A.K. Mariappan[4]
1,2,3,4 Department of Information Technology, Easwari Engineering College

***Abstract*—E-Commerce is growing rapidly and more number of products are sold online. Many users are using the Internet to post the opinions about the products and its aspects. As there are large number of reviews, it is impossible to read all the reviews. So we propose a ranking system using opinion mining to rank the products based on features (aspects). Users can perform opinion search to rank products based on the feature. In this system, we consider the reviews of people who have purchased the product as those reviews are more likely to contain less spam and more utility. In this system, we perform sentence level sentiment classification of reviews to identify important product aspects and the opinions about those aspects.**

***Index Terms*- Aspect ranking, Opinion mining, Sentiment classification, Text classification.**

I. INTRODUCTION

Opinions and facts are the two major types of textual information in the web. Facts are objective expressions about entities, events and their properties. Opinions are usually subjective expressions that describe people's sentiments, appraisals or feelings toward entities, events and their properties. We generally focus on opinion expressions that convey people's positive or negative sentiments. Much of the existing research on textual information processing has been focused on mining and retrieval of factual information, e.g., information retrieval, Web search, text classification, text clustering and many other text mining and natural language processing tasks. Little work had been done on the processing of opinions until recent times. Yet, opinions are so important that whenever we need to make a decision, we hear others' opinions.

In this paper, a product ranking system using opinion mining techniques to find favorable products for users is presented. Some papers have explored opinion mining methods to identify the polarity or to improve accuracy. This is not just to rank the products based on any reviews present in the ecommerce website. We first check for the authenticity of the purchase when retrieving opinions. Then, classify the opinions into positive, negative and neutral reviews using sentiment analysis. The ranking of products is based on aspects and reviews as well. Every individual aspect such as touch screen, battery and camera resolution is taken into consideration while ranking the products. We check for the match between the products and aspect based opinion search query before we start. Eventually, our system would provide users to specify product features in a query, and send back the ranking results of all matched products.

The remainder of the paper is organized as follows. In Section II and III, we review the related works and current systems. Then the framework of a product ranking system is proposed in Section V. The system implementation and some product ranking results are described in Section V and VI.

II. RELATED WORKS

The product aspect ranking [1] is based on identifying the important aspects of a product using a shallow dependency parser and considers aspect frequency. It also uses a probabilistic aspect ranking algorithm to rank the aspects. The web products ranking [2] proposes a system to rank the products using opinion mining. It ranks the products specified by users using a score calculated using polarity of reviews, popularity weight and product release month. The product details and sentence polarity information are integrated into XML structure in this work. The aspect level opinion mining [3] proposes a Joint Aspect / Sentiment (JAS) model to jointly extract aspects and aspect dependent sentiment lexicons from online customer reviews. In this work, first the aspects of entities are identified and aspect dependent lexicons are identified. Aspect dependent sentiment lexicons can help in identifying target aspects where the aspects are not mentioned explicitly. This is called implicit aspect identification which is a challenging problem in aspect mining. A sample review such as "it is so delicious" reveals a positive opinion on the taste of food. Aspect aware sentiment lexicons can help infer the targeted aspects. In [4], the aspects and sentiments are derived from corpora. In [5], a novel method is proposed for retrieving product features from online reviews. It proposes a language modeling framework that combines a probabilistic model of opinion words and a stochastic mapping model between words to approximate a language model of features.

III. EXISTING SYSTEM

There are many papers on product ranking, based on the customer reviews, but there's no system for ranking of products based on its individual aspects. We generally rank the products based on opinion mining. In existing system, we extract the reviews through a crawler; classify the reviews based on sentiment analysis and natural language processing into positive, negative and neutral reviews. Ranking is done based on the paramount number of positive reviews a product has, in general. But there's no checking for authenticity of the customers in that particular site. The present systems focuses on ranking aspects of a particular product. We focus on ranking products of a particular kind based on a particular aspect.

IV. PROPOSED SYSTEM

In the proposed product ranking system, products are ranked depending on its match with aspect based opinion search query. Also, aspect dependent sentiment lexicons are used to solve the implicit aspect identification problem. Better ranking is achieved as this problem is resolved. The ranking of products are done for every aspect of the product such as resolution, touch screen and so on. The reviews are then refined by checking for its authenticity as in whether it is a verified purchase or not. We use sentiment analysis to classify the reviews based on positive and negative comments using the words such as 'good', 'amazing', 'excellent' and 'poor', 'bad', 'disappointed' and so on in order to count the positive, negative and neutral reviews. Not only the lexicons, but also the stars given to every product are taken into account to rank them. Scores are given based on these criteria and ranking is done based on the polarity of the reviews .Thus we would achieve better ranking of products. The UI allows users to select the product category, feature and brand names.

SYSTEM ARCHITECTURE

In this system, we propose a framework where the user enters the specific product category and feature of the product in the user interface. At the back end, the reviews are extracted; refined and mining of opinions are done using NLP and sentiment analysis for every review and aspect of the products. Then the ranking is done and issued to the user. Also, while ranking the authenticity of reviews are considered.

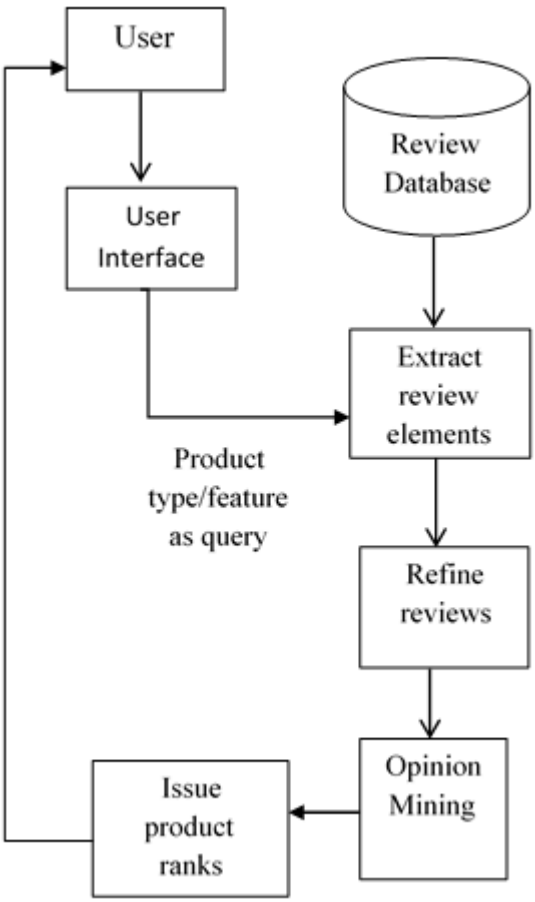

Fig. 1. System architecture

Figure 1 shows the proposed system architecture. The user reviews are stored in a database and sentiment analysis is carried out for the reviews. For the given query, all matched products are identified and ranked based on the chosen aspect and returned to the user. The user can give the query in the interface developed. The opinion mining is carried out based on a Naive Bayes classifier built using WEKA tool.

V. IMPLEMENTATION

EXPERIMENTAL DATA

The dataset was collected by extracting reviews from e-commerce sites such as Amazon, Flipkart for mobile phones and cameras. Totally, 27214 reviews were collected consisting of 92231 sentences. First, the Amazon and Flipkart product pages were downloaded using Python and the required data were extracted using HTML parser library called BeautifulSoup4 in Python. Each row contains information such as category, product name, stars, review title, review text, helpfulness, and verified purchase. After all the data was collected, only the verified purchase reviews were considered since the reviews given by buyers are more likely to contain less spam and more utility.

```
    def closeLinks(self):
        self.File.close()

    def getproductlinks(self):
        return self.productlinks

f = Flip(category="Mobile Phones",url="")
f.getLinks("flipkart samsung.txt")
f.createWriter("flipkart-reviews/Samsung.csv")
links = f.getproductlinks()
for l in links:
    f.crawl(f.changelink(l),0,False)
f.closefile()
f.closeLinks()
```

```
('Mobile Phones', u' Samsung S6812 - Galaxy Fame ', u'4 stars', u'Perfect in small size', u'I purchased this phone and using it last 4 months. I was
('Mobile Phones', u' Samsung S6812 - Galaxy Fame ', u'4 stars', u'Good Performer but problematic', u"This phone is very good in performance and speed
('Mobile Phones', u' Samsung S6812 - Galaxy Fame ', u'3 stars', u'Economic Phone within 8K range....', u'Good phone... i think this phone with all th
('Mobile Phones', u' Samsung S6812 - Galaxy Fame ', u'5 stars', u'Nice Phone - Delivery is so fast within 48hrs', u"I just bought this phone, its a n
('Mobile Phones', u' Samsung S6812 - Galaxy Fame ', u'3 stars', u'fame is awsome', u'samsung fame awsome mobile . but problem flipkart says to 4gb in
('Mobile Phones', u' Samsung S6812 - Galaxy Fame ', u'3 stars', u'Cheap price but low in performance', u"Got the phone well packed and timely deliver
('Mobile Phones', u' Samsung S6812 - Galaxy Fame ', u'5 stars', u'Excellent android phone with all features', u'It is excellent phone with all featur
('Mobile Phones', u' Samsung S6812 - Galaxy Fame ', u'5 stars', u'BEST PHONE UNDER THIS RANGE', u"I using this phone for more than 15 days.....i'm f
('Mobile Phones', u' Samsung S6812 - Galaxy Fame ', u'3 stars', u'Samsung Fame S 6812', u'i bought this Samsung product. I wanted to have a small and
('Mobile Phones', u' Samsung S6812 - Galaxy Fame ', u'5 stars', u'Wonderful phone', u'I am vey glad to be an owner of this outstanding phone.It inclu
```

Fig. 2. Data extracted from Amazon website

REVIEW PREPROCESSING

I. Splitting Reviews into sentences

All the reviews are split into individual sentences to perform sentence level sentiment analysis. It is essential to split into sentences to identify the aspects as they are given anywhere in the review text.

II. Removing Stopwords

Stopwords are insignificant words in the review text. We used the stopwords corpus in NLTK (Natural Language Toolkit) library in Python. We also removed additional stopwords like the website names, opinion about delivery as they focus on the merchants and not on the products. As the review "Touch should improve" conveys a negative meaning it should be classified as negative. So words such as "should", "could have" are not removed.

III. Correcting Spelling Mistakes

The list numbers in the reviews are removed and the spellings of the words are corrected so that the exact meanings of the words are interpreted for allotting scores for every review.

IV. Removing spam content in reviews

We removed grammatically meaningless words such as "ossssummmmm" and other language words which do not contribute to the review scores.

V. POS Tagging

POS tagging is done to read texts in the reviews and assign parts of speech to every word. In order to identify the aspects in all the reviews extracted, we do POS tagging for every wordin a given sentence and thereby classify the words as noun, verb, adjective etc.
Figure 3 shows the reviews data after preprocessing. This data is now used as input to the sentiment classifier.

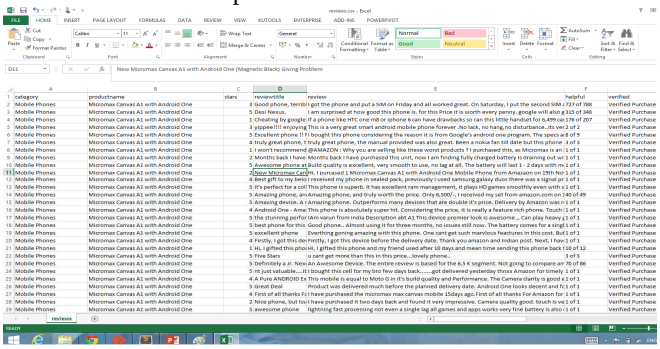

Fig. 3.Preprocessed reviews

ASPECTS IDENTIFICATION

We do static aspect extraction from the reviews and then find the sentences that have aspects such as touch, battery, memory and camera mega pixels, and store those sentences. We use the aspects corpus from MPQC corpus to develop an aspect dependent sentiment lexicons to solve implicit aspect identification problem.

POLARITY CLASSIFICATION

All the sentences are classified as positive or negative or neutral using a sentiment classifier. We then find the average polarity of every individual review. The aspect score for a product is calculated as follows:

$$\text{Aspect score for a product} = \sum_{i=1}^{n} (s_i)_{pos}$$

Where s_i is the sentence in a review R for a product, $(s_i)_{pos}$ is the positive polarity score for the sentence s_i.. The aspect score for a product is calculated by adding all the positive scores. We used WEKA for classifying each sentence. Using Naïve Bayes classifier with two class "good", "bad" and "neu" as class attributes to denote the polarity of the review. We manually trained 1000 reviews and then rest as test set for the classifier. The table 1 shows the accuracy of the classifier we built.

TABLE 1: Detailed accuracy of the classifier

TP Rate	FP Rate	Precision	Recall	F-Measure	ROC Area	Class
1	0.034	0.967	1	0.983	0.95	Good
1	0	1	1	1	0.964	Bad
0.9	0	1	0.9	0.947	0.979	Neu

RANKING OF PRODUCTS BASED ON REVIEW SCORES AND ASPECTS:
After the identification of aspects and polarity of every sentence, we rank the products based on review scores and we perform ranking for every aspect of the products considered.

USER INTERFACE
Selecting Products: After choosing the product name(s) and product category, the system searches for products with the given query and retrieves the matching products. The user can choose multiple brands and also specify a limit to limit the number of products returned for the search query.
Features: The system allows user to select features for the specified product category. Multiple features can also be given in the search query.
Finally, pressing button "Submit" would retrieve the products along with their aspect scores for the specified aspects and rank them according to the scores.

VI. RESULTS

The results obtained contains the classification of reviews and GUI which gives the ranking of products based on reviews and aspects for every brand and also gives the number of individual sentences of the reviews obtained.
Figure 4 shows the scores accredited to reviews.

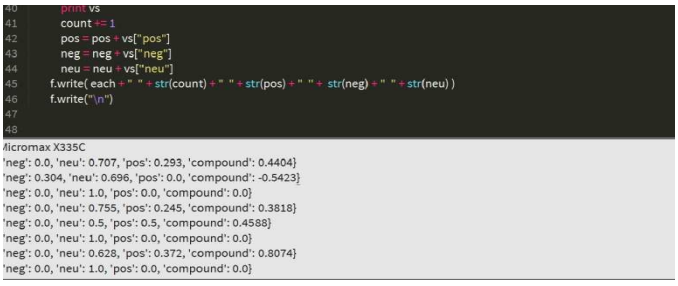

Fig. 4. Accrediting scores to reviews

Figure 5 shows the products ranked based on review scores for the "Mobile phones" category and the brands "Nokia" and "Micromax".

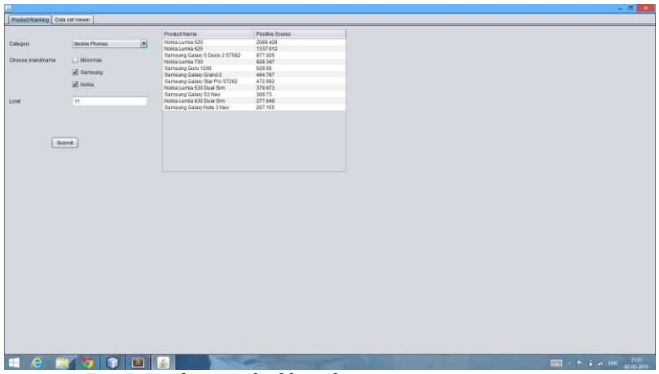

Fig. 5. Products ranked based on review scores

Figure 6 shows the products ranked based on aspects for the given product category and brands. The figure also shows the products ranked based on review scores for comparison.

	A	B	C
1	products	battery	touch
2	Nokia Lumia 520	Nokia Lumia 730	Nokia Lumia 520
3	Nokia Lumia 625	Nokia Lumia 520	Nokia Lumia 625
4	Nokia Lumia 730	Nokia Lumia 625	Nokia Lumia 730
5	Nokia Lumia 530 Dual Sim	Nokia Lumia 530 Dual Sim	Nokia Lumia 530 Dual Sim
6	Nokia Lumia 630 Dual Sim	Nokia Lumia 630 Dual Sim	Nokia Lumia 630 Dual Sim
7	Nokia XL	Nokia XL	Nokia XL
8	Nokia Lumia 630 Single SIM	Nokia Lumia 630 Single SIM	Nokia Lumia 630 Single SIM
9	Nokia X2	Nokia X2	Nokia Lumia 638
10	Nokia Lumia 638	Nokia Lumia 638	Nokia Lumia 830
11	Nokia Lumia 830	Nokia X Plus	Nokia X2

Fig. 6. Products ranked based on aspects: battery, touch screen.

CONCLUSION

In this paper, we presented a product ranking system based on aspects of the product. We performed opinion mining at aspect level. We find that querying products by aspects will retrieve effective results and also relevant products based on the user's needs.

REFERENCES

[1] Zheng-Jun Zha, Jianxing Yu, Jinhui Tang Meng Wang, Member, IEEE, and Tat-Seng Chua, " Product Aspect Ranking And its Applications", IEEE Transactions On Knowledge And Data Engineering, Vol. 26, No. 5, May 2014

[2] Yin-Fu Huang and Heng Lin, "Web Product Ranking Using Opinion Mining", IEEE Symposium on Knowledge and Data Intelligence, 2013

[3] XU Xueke, CHENG Xueqi, TAN Songbo, LIU Yue, SHEN Huawei,"Aspect-Level Opinion Mining of Online Customer Reviews", China Communications, March 2013

[4] M. Ganapathibhotla and B. Liu, "Mining opinions in comparative sentences," Proc. the 22nd International Conference on Computational Linguistics, pages. 241-248, 2008.

[5] Lizhen, Liu; Wei, Song; Hanshi, Wang; Chuchu, Li; Jingli, Lu, "A novel feature-based method for sentiment analysis of Chinese product reviews", China Communications, March 2014, Vol. 11, Issue: 3, Page 154 – 164.

[6] R. Xu and C. Kit, "Corse-fine opinion mining," Proc. International Conference on Machine Learning and Cybernetics, pages. 3469-3474,2009.

[7] Hu M., Liu B., Mining and summarizing customer reviews. In Proceedings of the tenth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, pages 168–177. ACM, 2004

[8] Z. Zhang, Y. Li, Q. Ye, and R. Law, "Sentiment classification for Chinese product reviews using an unsupervised Internet-based method,"

[9] L. R. Bahl and R. L. Mercer, "Part-of-speech assignment by a statistical decision algorithm," Proc. the IEEE International Symposium on Information Theory, pages. 88-89, 1976.

[10] H. Schmid, "Part-of-speech tagging with neural networks," Proc. the15th International Conference on Computational Linguistics, pages.172-176, 1994.

[11] P. D. Turney, "Thumbs up or thumbs down? semantic orientation applied to unsupervised classification of reviews," Proc. the 40th ACL Annual Meeting on Association for Computational Linguistics, pages.417-424, 2002.

[12] WordNet 2.1 for Windows, http://wordnet.princeton.edu/wordnet/

[13] P. D. Turney and M. L. Littman, "Measuring praise and criticism: inference of semantic orientation from association," ACM Transactions on Information Systems, Vol. 21, pages. 315-346, 2003.

[14] Sasha Blair-Goldensohn, Kerry Hannan, Ryan McDonald,Tyler Neylon,George A. Reis,Jeff Reynar," Building a Sentiment Summarizer for Local Service Reviews**.",** NLPIX2008, April 22, 2008, Beijing, China.

[15] . Yu, J. Ma, S. Tsuchiya, and F. Ren, "Opinion mining: a study on semantic orientation analysis for online document," *Proc. the 7th WorldConference on Intelligent Control and Automation*, pages. 4548-4552,2008