

# Sentiment Analysis of Product on User Reviews using Big Data

<sup>#1</sup>Jyoti Jadhav, <sup>#2</sup>Abhishek Khandave, <sup>#3</sup>Monika Sancheti, <sup>#4</sup>Yogita Shinde,  
<sup>#5</sup>Prof. Rupali Gawande



<sup>1</sup> jyotijadhav2694@gmail.com,  
<sup>2</sup> abhishekkhandave@gmail.com,  
<sup>3</sup> sanchetim.16@gmail.com,  
<sup>4</sup> yogitashinde35@gmail.com  
<sup>5</sup> rupaligawande14@gmail.com

<sup>#1234</sup> B.E Student, IT Department  
<sup>#5</sup> Professor, IT Department

Marathwada Mitra Mandal's College Of Enginnering, Karvenagar,  
Pune, Savitribai Phule Pune University, India, 411048.

## ABSTRACT

Mining opinion targets and opinion words from online reviews are important tasks for fine-grained opinion mining, the key component of which involves detecting opinion relations among words. To this end, this paper proposes a novel approach based on the partially-supervised alignment model, which regards identifying opinion relations as an alignment process. Then, a graph-based co-ranking algorithm is exploited to estimate the confidence of each candidate. Finally, candidates with higher confidence are extracted as opinion targets or opinion words. Compared to previous methods based on the nearest-neighbor rules, our model captures opinion relations more precisely, especially for long-span relations. Compared to syntax-based methods, our word alignment model effectively alleviates the negative effects of parsing errors when dealing with informal online texts. In particular, compared to the traditional unsupervised alignment model, the proposed model obtains better precision because of the usage of partial supervision. In addition, when estimating candidate confidence, we penalize higher-degree vertices in our graph-based co-ranking algorithm to decrease the probability of error generation. Our experimental results on three corpora with different sizes and languages show that our approach effectively outperforms state-of-the-art methods. Identifying the sentiment of the text has recently gained a lot of popularity probably due to availability of huge datasets. Identifying the right sentiment from bulk of data becomes the real challenge. In this paper we suggest an approach to analyze the sentiment of the text available on social media.

**Keywords:** Text mining; sentiment analysis;

## ARTICLE INFO

### Article History

Received: 20<sup>th</sup> April 2017

Received in revised form :

20<sup>th</sup> April 2017

Accepted: 26<sup>th</sup> April 2017

**Published online :**

27<sup>th</sup> April 2017

## I. INTRODUCTION

As smart devices and wireless Internet have come into wide use, the amount of VOC (voice of customer) data is growing exponentially in the on-line world. Customer has been reflecting their opinion through Internet review sites and Social Networking Sites data and they become an important source for finding customer preference for products, services, an individual, or policies. Sentiment analysis (opinion mining) for online review sites, personal blogs, or Social Networking Sites data is becoming an active research issue in the business and IT areas. Sentiment analysis aims to determine the attitude of a customer with respect to some topic or the overall contextual polarity of a document.

Recommending Systems are new generation dynamic internet tool that help user for efficient product search via Information on the internet and receive information related to their preferences. To overcome the product overload of Internet shoppers, we introduce a semantic recommendation procedure which is more efficient. All product related data will be stored with it. Depending upon user's comments recommendations will be given to him. The suggested procedure recommends the semantic products to the customers and is originally based on semantic analysis and product classification. we will be developing a system to give recommendation about the good malls on the basis of shops and products which includes highest positive comment. Recommendation based on the analysis of the whole portal where the analysis of the Shopping Malls,

Products is being analyzed using hive in hadoop. Another important domain for sentiment analysis is the financial markets. There are numerous news items, articles, blogs, and tweets about each public company. A sentiment analysis system can use these various sources to find articles that discuss the companies and aggregate the sentiment about them as a single score that can be used by an automated trading system.

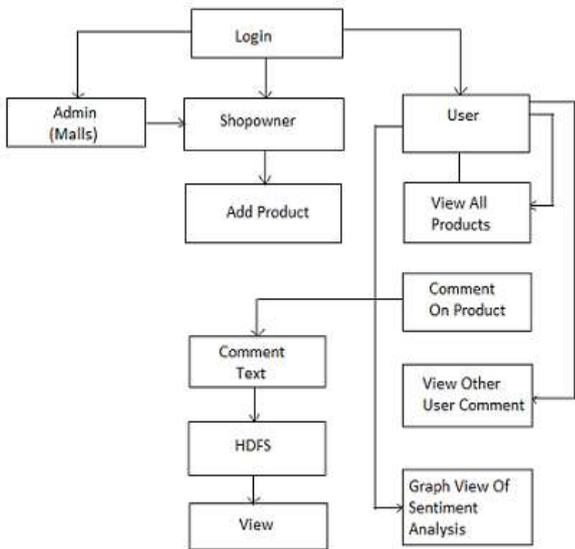


Fig 1: Architecture

**II. PROJECT METHODOLOGY**

There will be 3 user roles such as a Admin, Shop owner, user.

- 1) Admin role: Admin is able to add the Mall and Shops.

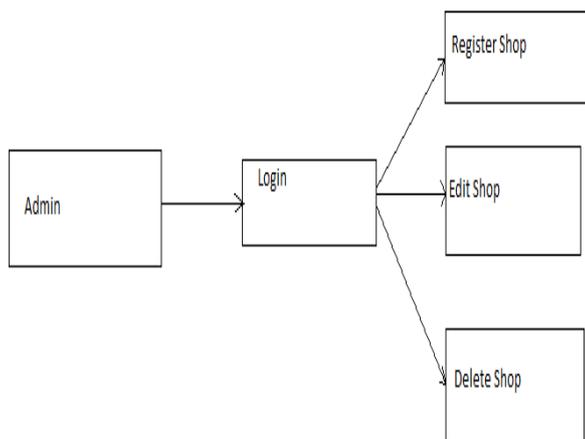


Fig 2: Admin role

- 2) Shop owner role: Shop Owner is able to add the products on his/her shop.

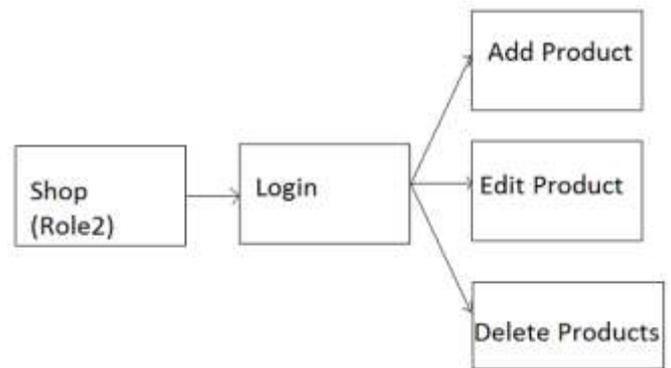


Fig 3: Shop owner role

- 3) User role:

- User is able to view all the Mall, Shops, and the particular products added in the particular shop.
- User is able to comment on individual shop and the products User is also able to view the comments commented by other Users.
- User is able to view the graphical Analysis of the Comments on shop & Products.
- All the data that will be generated after adding malls, shops and products, huge data will be collated at the back end which will be stored in textual format and analysis will be done using Hive where local system data will be stored into Hadoop and with HQL sentiment analysis will be predicted.

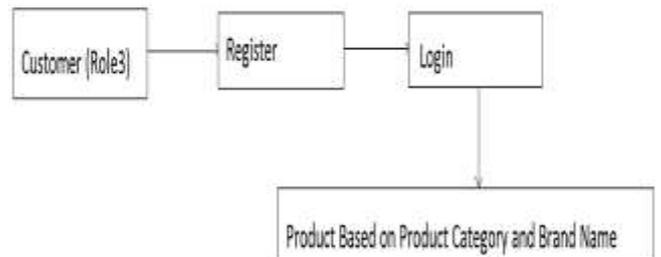


Fig 4: User role

**III. RELATED WORK**

- 1. Analysis of Sentiments using Unsupervised Learning Techniques[3]

In this paper proposes the sales and distribution integration data analysis solution based on big data analysis.

- 2. An approach towards comprehensive sentimental data analysis and opinion mining[4]

This paper is the outcome of our research in gathering opinion and review data from popular portals, e-commerce websites, forums or social networks; and processing the data using the rules of natural language and grammar to find out what exactly was being talked about in the user's review and the sentiments that people are expressing.

- 3. The Solution for Performance Improvement of Electric Distribution Network Line Loss Based on Hadoop Big Data Technology[5]

In this paper proposes the sales and distribution integration data analysis solution based on big data analysis. It explains the key technologies such as sales and distribution data integration, line loss data processing and line loss data feedback.

#### IV. PROPOSED SYSTEM

In propose system we develop Recommending Systems for user that help for efficient product search via Information on the internet and receive information related to their preferences. It is also helpful for user because user view the comment on the product and that basis user have to purchases the product. To overcome the product overload of Internet shoppers.

#### V. OBJECTIVE

To overcome the product overload of Internet shoppers. A semantic recommendation procedure which is more efficient. The suggested procedure recommends the semantic products to the customers. Social media Monitoring has been growing very rapidly so there is a need for various organizations to analyze customer behavior or attitude of particular product. Text analytics and sentiment analysis can help organization to derive valuable business insights. Attitude can be calculated based on polarity check. Sentiment analysis on Online review are done by forming dictionary which shows that it is easier to build dictionary on phrases.

#### VI. CONCLUSION

First dataset cleaning is done. Stop words, spaces are removed then keywords are obtained. Exact matching keywords are found out from the dataset. We have formed the rating dictionary and have given rating values from -1 to +1. Sentiment Analysis is used for calculation. User is able to view the comments of all the user that has been commented. Graphical Analysis of the comments as per the positive, negative or neutral. Large amount of data is managed on it Our main contribution is focused on detecting opinion relations between opinion targets and opinion words. So we have to make updations in the rating dictionary as passive user's reviews changes. So Recommendation is dynamic and more realistic. We are using Map-Reduce in java to reduce number of same keywords into one in the long. Finally we will run this project on Hadoop. It is designed to scale up from single servers to thousands of commodity machines, each offering local computation and storage.

#### REFERENCES

- [1] Neha R. Kasture, Poonam B. Bhilare, An Approach for Sentiment Analysis on Social Networking Sites, 2015.
- [2] Kang Liu, Liheng Xu, Jun Zhao, Co-Extracting Opinion Targets and Opinion Words from Online Reviews Based on the Word Alignment Model, 2015.

- [3] M. S. Usha, Dr. M. Indra Devi, Analysis of Sentiments using Unsupervised Learning Techniques, 2013.
- [4] Pooja Kherwa, Arjit Sachdeva, Dhruv Mahajan, Nishtha Pande, Prashast Kumar, An approach towards comprehensive sentimental data analysis and opinion mining, 2014.
- [5] Lihua Sun, Mu Hu, Qingqiang Meng, Feng Lin, Yakang Qian, Yunbo She, Zheng Ma, Xuan Pei, Shuting Song, Nanjing NARI, The Solution for Performance Improvement of Electric Distribution Network Line Loss Based on Hadoop Big Data Technology, 2015.
- [6] Laila M. Qaisi, Ibrahim Aljarah, A twitter sentiment analysis for cloud providers: A case study of Azure vs. AWS, 2016.
- [7] Mehadi Hasan, Noor Hussian Shaon, Kamrul Hasan "Friend recommendation framework for social networking sites using user's online behavior" 2016.
- [8] Kushboo.r. Shrote, A.V Deorankar. "Review based service recommendation for big data", 2016.