Twitter Sentiment Analysis for Electronic Goods Recommendation

Neha Vilas Jadhav[1] Pooja Jaydrath Kshirsagar[2] Nikhila Deepak Lodh[3] Dipti Gajanan Shilimkar[4]

Abstract— Twitter, Facebook and other social networking sites are the main source of data for analysis of product. Twitter is the most common micro blogging site which fetches the millions of tweets that are generated every minutes and every seconds. These tweets can be useful for sentiment analysis and analyzing users opinion about the particular product. And also enable the product manufacturing companies to gain the feedback about their product to improve the product quality. In this paper we are fetching the tweets about electronic goods e.g. Laptop, Mobiles etc. form the twitter account and by analyzing those tweets we are recommending the best product to the user.

Key words: Tweets, Hadoop, Sentiment Analysis, Twitter, Natural Language Processing

I. INTRODUCTION

Now a day's Social Media like Twitter is one of the best platform where many people express their opinion about anything like products, events, latest news etc. Due to high popularity, weblogs like tweets on twitter provides a very useful information which can be very helpful in analyzing the general public's sentiments and opinions of people. Analyzing this large amount of data on internet produces useful actionable knowledge that can be useful for vendors and other interested parties.

For fetching the tweets from twitter account, Twitter API is required. Currently there are three different versions of API's available [1].Those are REST API, Search API and Streaming API. The REST API allows developers to gather the user information and status data: the Search API is able to query specific twitter contents whereas Streaming API allows to collect twitter contents in real-time. The developers can mix those API's to create own applications. Hence sentiment analysis deals with millions of tweets generated every minute.

This paper mainly focuses on the tweets from twitter i.e. twitter data related to electronic goods. Twitter data has been used to address wide range of applications like movie reviews, political election prediction and product sales prediction. However, no studies have been conducted to recommend the best electronic product to the user. Such research is important to overcome the overhead of customer to analyze which product is best currently in the market or the manufacturing companies use this to improve the quality of product depending on the current scenario of the market. The result of above analysis can facilitate customer to buy the best product by comparing with the other available products in the market. Our approach is to fetch the tweets of particular #hashtag and perform the sentiment analysis i.e. finding the polarity of the tweets using hadoop framework. In order to handle so many tweets we are using apache hadoop framework.

II. LITERATURE SURVEY

In the last few years various evaluation dataset for tweeter sentiment analysis have been made publically available. This evaluation datasets consist of set of tweets. Sentiment analysis is nothing but the information extraction task and used natural processing language to analysis. Basically the aim of sentiment analysis is to express their feelings, emotions and opinion in positive, negative and many more.

In Nasukawa and Yi [2] explained how to extract sentiments associated with polarities of positive or negative for specific subjects from a text document, instead of classifying the complete document into positive or negative polarities. The main operation in this sentiment analysis are extraction of information and to express or indicates the positive (favorable) and negative (unfavorable) opinion for particular topic or subject. This is powerful functionality of this sentiment analysis.

In Ding [3], sentiment analysis illustrated to identifying semantic orientations of opinions expressed by users on specific products features.

Taylor et al.[4] presented sentiment analysis design of a opinion mining system for tourism. Aim of Taylor was to make this useful and easy to understand for opinion mining in many industries who does work for sentiment mining i.e. sentiment analysis. They also solve a specific problem in the Lake District tourism industry and used this proposal to implement the system which gives the best features extraction and the opinion mining.

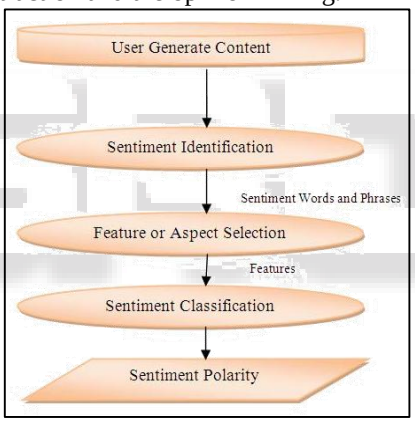

Fig. 1: Sentimental analysis process on user generated contents [5].

Haddi, Lui and Shi [6] implements the sentiment analysis for online movie reviews. In this approach they used different type of pre-processing methods and combined this methods to reduce the noise in the text documents and using chi-squared method they removed irrelevant features which do not affects on its orientation. Authors have gave extensive results that is achieving great accuracy by doing appropriate text pre-processing and it achieved on two data set that comparable to the sort of accuracy that can be achieved in subject categorization, make much easier problem.

In Moraes, Valiati & Neto [7], In this the focused part was comparing SVM and ANN considering the requirements to achieved good classification accuracies. In this, authors evaluated SVM and ANN as function (unigrams) approach.

III. ARCHITECTURE

The system architecture that is proposed for this paper is as shown in Figure.

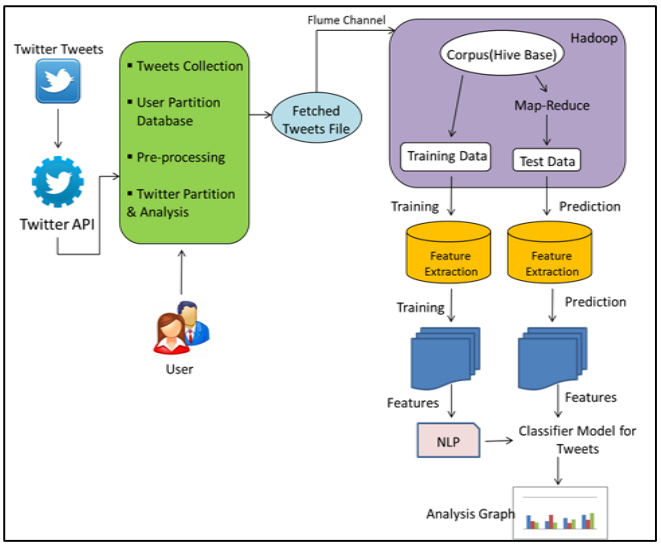

Fig. 2: Complete System Design

This paper uses both knowledge based methodologies and NLP toolkit for analysing the tweets. The process is as follows:

1) Tweets are fetched through a Twitter API. Then collected tweets are get preprocessed and then dumped on Hivebase (corpus) in Hadoop using flume channel and saved to text file.
2) Then corpus get divided into a 'training set' and a 'test set' and features are extracted from each respectively.
3) On the 'training set- feature extraction' NLP is applied. For NLP we have used Stanford CoreNLP 3.3.0 NLP toolkit which is responsible for analyzing the tweets and then classify them into the categories like positive, negative and neutral.
4) On the 'test set-feature extraction' the predictions are drived then these derived features from predictions are given to the Naïve Bayes classifier.
5) Then the ouput extracted using stanford coreNLP toolkit and the results generated by classifier get combined which will generate the analysis graph of showing the comparative view of electronic goods.

IV. Conclusion

We developed the system which fetch tweets from the twitter user account and which will make the comparative view of all electronic goods and will recommend the user which one is best to buy. Also the manufacturing company can use this analytical result to improve quality of product and tackle with other products in market.

V. Future Scope

In this we are only handling straight forward tweets and using Windows OS, so in future we can develop an application which can handle essay type tweets and make compatible with different platforms.

References

[1] Xing Fang, Justin Zhan,"Sentiment analysis using product review data", Journal of Big Data20152:5, DOI:10.1186/s40537-015-0015-2,pp.June 2015.
[2] T. Nasukawa, "Sentiment Analysis: Capturing Favorability Using Natural Language Processing Definition of Sentiment Expressions," pp. 70–77, 2003.
[3] X. Ding, S. M. Street, B. Liu, S. M. Street, P. S. Yu, and S. M. Street, "A Holistic Lexicon-Based Approach to Opinion Mining," pp. 231–239, 2008.
[4] E. Marrese-Taylor, J. D. Velasquez, and F. Bravo-Marquez, "Opinion Zoom: A Modular Tool to Explore Tourism Opinions on the Web," 2013 IEEE/WIC/ACM Int. Jt. Conf. Web Intell. Agent Technol., pp. 261–264, Nov. 2013.
[5] Shailesh Yadhav, "Sentiment Analysis and Classification: A Survey" Volume 3, isuue3, March 2015, International Journal of advance research in computer science and management studies,ISSN: 2327782
[6] E. Haddi, X. Liu, and Y. Shi, "The Role of Text Pre-processing in Sentiment Analysis," Procedia Computer. Sci., vol. 17, pp. 26–32, Jan. 2013.
[7] R. Moraes, J. F. Valiati, and W. P. Gavião Neto, "Document-level sentiment classification: An empirical comparison between SVM and ANN," Expert Syst. Appl., vol. 40, no. 2, pp. 621–633, Feb. 2013.
[8] R. Bell, Y. Koren, and C. Volinsky, "Modeling relationships at multiple scales to improve accuracy of large recommender systems" KDD '07: Proceedings of the 13th ACM SIGKDD international conference on Knowledge discovery and data mining, New York, NY, USA, 2007, ACM.
[9] O. Celma and P. Herrera, "A new approach to evaluating novel recommendations", RecSys '08: Proceedings of the 2008 ACM conference on Recommender systems, New York, NY, USA, 2008, ACM
[10] G. Adomavicius and A. Tuzhilin, "Toward the next generation of recommender systems: A survey of the state-of the-art and possible extensions," IEEE Trans. Knowl. Data Eng.
[11] Michael Hashler, "Recommender Lab: A Framework for Developing and Testing Recommendation Algorithms" Nov. 2011.
[12] G. Linden, B. Smith, and J. York, "Amazon recommendations: Itemto- item collaborative filtering," IEEE Internet Comput., Feb. 2003.

