Aspect Based Sentiment Analysis Approaches with Mining of Reviews: A Comparative Study

Neha Nandal, Jyoti Pruthi, Amit Choudhary

Abstract: *In the era of Technology, online marketing is becoming a new trend to ease things in the real world. Buying products online is now a facile task for people. Online Customers present their specific reviews on products they buy. These reviews project an impression on new customers of the product. Aspect based sentiment analysis concentrates on important aspects(or features) of the products which can be valuable for the customers while purchasing it online. To extract those features, the foremost work is to collect opinions on products. This paper presents a proposed work for extraction of aspects from opinions. Certain features play very important role while extracting opinions online and aspects from the data. Discussion on data collection and comparision of various methodologies to perform Aspect Level Sentiment analysis alogn with Machine learning methodologies has also been discussed here.*

Index Terms: *Review Mining, API, Crawler, Aspects, Sentiment Analysis.*

I. INTRODUCTION

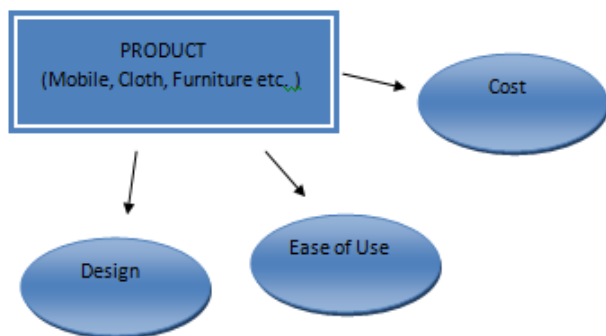Aspect Based Sentiment Analysis works with features in the data that can make impact on the domains.

Figure 1. Product and its Aspects

Figure 1 shows sample of aspects or features of any product. There can be certain types of aspects which can be valuable for the customer while purchasing.
For example , In a product like TV, features like Resolution, Picture quality, Cost, additional features etc. can be of focus for a customer.

It becomes very easy for customer to choose a product if they get reviews or opinions on particular feature of the product. There can different types of Aspect Level Sentiment Analysis like, Document Level , Sentence Level, Text Level having their own ways to extract Sentiments on the basis of Aspects. Document Level sentiment analysis considers complete document of data and extract features from that. Unlike Document Level, Sentence Level extracts features by taking in consideration the sentences. Text Level analysis focuses on particular text and provides the output in summarized manner.

A. Extraction of Reviews

To extract the reviews of customers on a website for any particular product, crawler can be used . Crawler visits the pages, crawls and data and collect it as a storage. There are several open source crawlers available like Scrapy, JSpider, WIRE, Pavuk etc. Crawlers are different in terms of various parameters like flexibility, Robustness, Distributed, Scalable, Efficient etc.
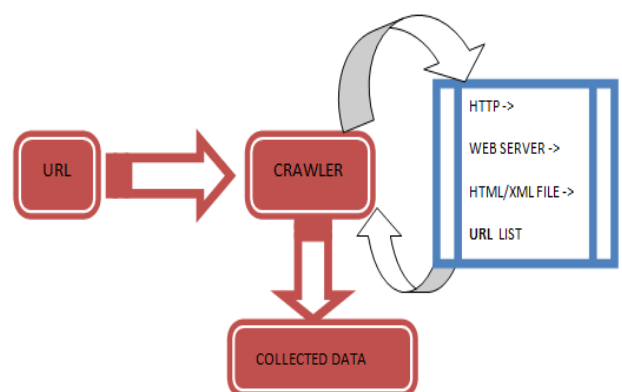

Figure 2. Flow of Crawler

Layout of the paper is as follows:

- Introduction to Mining of Reviews and Aspect Level Sentiment Analysis
- We presented how mining of reviews can be done for sentiment analysis.
- The Literature review that has been done for the work.

- Proposed Methodology of the work.
- Comparison of Methodologies for performing Sentiment Analysis
- Comparison of Machine Learning Approaches and Techniques.
- Conclusion and Future work.

II. LITERATURE REVIEW

The Research in the field of sentiment analysis, Aspect based sentiment analysis, opinion mining has been studied and discussed here. Bo Pang et.al.(2008)[1] presented a survey on sentiment analysis and opinion mining. They focused mainly on approaches and techniques that are directly related to opinion oriented systems. Discussion of datasets, evaluation techniques, resources are also provided in the paper.

Silvana Aciar(2010)[2] focused on context based information in customer reviews and developed a method to identify the same. The dataset that has been used in the paper is of www.tripadvisor.com. The identification of review's sentences has been done using classification text mining. Presented work has been explained using a case study.

A. Gural Vural et.al.(2012)[3] focused on web crawling. Researchers presented a work on crawling the web which is focused on sentiments. Various strategies and tools to perform the same also discussed in the paper. They developed the framework which collects URLs of web pages and crawls them to collect the data. They presented comparion of several strategies to perform sentiment based crawling and showed that their framework provides efficient results in crawling number of pages.

Zhang Hailong et.al(2014) [4] presented a survey on methods of sentiment analysis. They also discussed cross domain and cross lingual approaches in the work. The results they presented showed that techniques like SVM and Naïve Bayes shows higher precision than lexicon based approaches.

Abhishek Kaushik et.al.(2015)[5] presented a study on sentiment analysis tools and methods. They showed basic methodology of sentiment analysis and classification of polarities. Discussion of approaches, tools have been shown in paper using tabular format. They discussed how sentiment analysis can help in several domains.

Joseph Mei et.al.(2015)[6] developed self-guiding web crawler and collected data for four sentiment based classes i.e. extremist webpages, news discussion extremist, anti-extremists and webpages having no discussion of extremists.They used decision trees to calculate the results and work showed success rate of 80% for differentiating four classes.

Kumar Ravi et.al.(2015)[7] presented a survey on sentiment analysis and opinion mining. They presented various techniques, approaches for perfoming opinion mining and sentiment analysis. The paper covered views presented in 100+ articles. The literature from 2002-2015 covered in the paper and presented in summarized way with tabular format.

Anurag P. Jain et.al(2015)[8] presents work for analyzing user sentiments. They have used data mining classifiers. Comparison of various classifiers also discussed in the paper.

The results presented in the work shows that K-NN provides high accuracy in prediction.

Alessia D'Andrea et.al.(2015)[9] Researchers in this work presented applications of sentiment analysis. They presented tables for approaches of Sentiment Analysis, tools available for sentiment analysis, Different application area for sentiment analysis.

Rushlene Kaur Bakshi et.al.(2016)[10] discussed sentiment analysis and opinion mining. They presented a steps for checking the affects of tweets over stock market. They followed steps like data cleaning and finally sentiment analysis on the cleaned data. The algorithm developed for the work shows effective results in terms of Positive, Negative and Neutral Sentiments.

Devika M D et.al. (2016)[11] presented a comparative analysis of different techniques and approaches. They covered approach of sentiment analysis and machine learning and presented them in tabular format on the basis of parameters like scope, accuracy etc.

Fatemeh Hemmatian et.al. (2017)[12] discussed about Opinion mining, the approaches, techniques and methods of it. They presented available challenges in the area and compared the advantages and disadvantages of approaches. Discussion of methods for aspect identification, extraction, classification and evaluation have also been done in the paper.

Sharma, Jyoti. (2017)[13] presented a review paper on mining of text and sentimental analysis. In paper several techniques of mining text and performing sentimental analysis are discussed.

Shubham Goyal(2017)[14] presented a work on sentiment analysis. They used text mining for collection of data and for classification purpose they utilized hybrid approach of KNN algorithm and Naïve Bayes algorithm. The data they have collected was of Tweets.

Shahid Shayaa et.al. (2018)[15] presented a survey on Opinion Mining and Sentiment analysis. They discussed several methodologies, techniques and approaches to perform sentiment analysis and opinion mining. Challenges in terms of Technical and Non-Technical sentiment analysis have been discussed in the paper.

Mr. S.M. Vohra et.al(2018)[16] presented a study on sentiment analysis. In this work they showed several techniques and approaches which can be utilized for performing sentiment analysis. Several tools for sentiment analysis are also discussed. The comparison they have made is presented in tabular format.

Lucia Martin-Gomez et.al.(2018)[17] presented a very new approach to identify sentiments of human towards music. They identified first the parameters of music which can make impact. Then techniques like KNN, Apriori are applied for the evaluation.

III. METHODOLOGY

Our approach can be summarized as follows:

B. **Proposed Methodology**

The methodology has been developed to perform Aspect Level Sentiment Analysis. This paper compares the methodologies that can be utilized in the work for performing Aspect Based Sentiment Analysis.
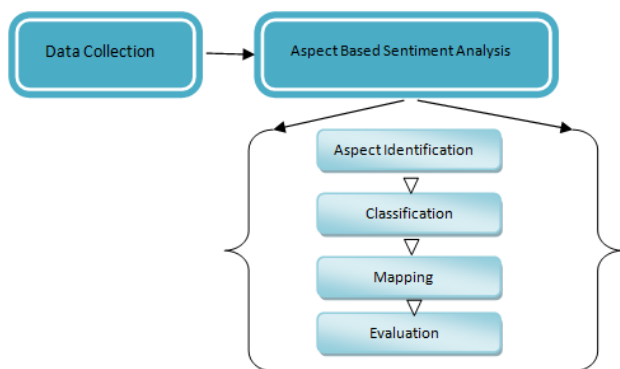

Figure 3. Proposed Methodology

For Data Collection, API Crawler will be used to collect reviews from Amazon.com. Products in which customers are most interested will be considered for data collection. API Crawler is able to crawl webpages of reviews and collect data in structured form.

After Collection of data, Aspects will be identified on which classification and mapping will be performed. Finally evaluation of results in terms of Precision, Recall, Accuracy will be provided.

Comparison of results with the existing approaches will be done in the work.

C. **DataSet Collection**

An API crawler has been developed to collect data from Amazon Products. The Crawler is very fast. The products chosen for collection are based on the number of reviews on the data. Products like Mobile phones of Samsung, Redmi, OnePlus, Earphones.
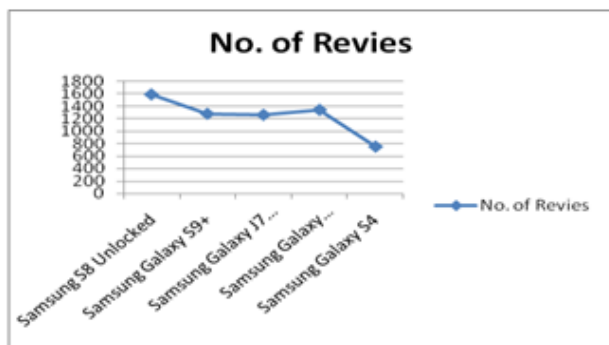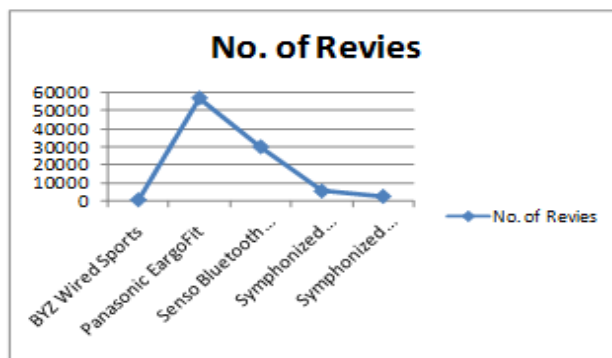

Figure 4. No. of Reviews on Mobiles

Figure 5. No. of Reviews on Earphones

Figure 4 and 5 are samples of number of reviews on products of amazon. Crawler is fast enough to collect data. The collected data is in the structured form.

IV. SENTIMENT ANALYSIS TECHNIQUE AND CLASSIFIERS

Comparison of Methodologies

There are several methodologies to perform Sentiment Analysis. Different types of classifiers in machine learning can be utilized for performing classification. This section discusses techniques for performing sentiment analysis and different machine learning classifiers.

A Sentiment Analysis Techniques

There are mainly two ways to perform sentiment analysis i.e. Lexicon Based Approach and Machine Learning Approach. Table 1 comprises of Sentiment Analysis Techniques with their important points to make comparison.

A.**1 Machine Learning Based Technique:**

This technique is usually based on Supervised Learning. It is very fast and accurate. In this technique we have to divide our data into two parts as Training Data and Testing Data. Classifier will get trained with the help of training data and it will test on testing data in terms of Accuracy. Some of the techniques the machine learning uses are Frequency of terms, Negations, POS tagging, opinion words.

A.**2 Lexicon Based Technique:**

This technique is basically is unsupervised technique. It can utilize techniques like Corpora based, DictionaryBased and Manual. In this technique, the classification has been done using comparison of aspects with the lexicons which are developed earlier. The steps this techniques used to perform classification are first preprocessing, then tokenization, sentiment score and classification.

Table 1. Comparison of Sentiment Analysis Techniques

Technique	Key Points
Lexicon-Based	• **Comparatively provides less performance than Machine Learning** • **Getting data is easier as compare to machine learning.** • **This technique is based on stored words and if the words are less in number then results can suffer.** • **Recall parameter can suffer if stored data is not sufficient and the words stored in corpora are mostly context dependant which can affect accuracy, precision and recall.**
Machine Learning	• **Better than Lexicon Based Method in terms of Performance** • **Big Training labelled data is required** • **It may not provide correct results if it is less of training data.**
Hybrid	• **This Technique is combinaition of both Lexicon and Machine Learning.** • **It provides highly accurate results** • **The results that are generated using Hybrid Technique can be more stable.** • **It is difficult to handle data with unwanted noise.**

B. Classifiers

Classification is a techniques to classify the data in defined classed. The data is divided into two parts as Training data and Testing data. Classifier trains itself using training data checks it's accuracy on testing data. There are different types of classifiers which can be utilized in classification. Performance of various classifiers depends on application and dataset.
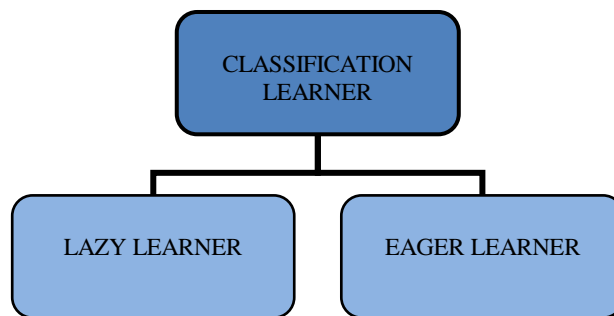

Figure 6. Types of Learner

Lazy Learners are those learners which takes more time in predicting results and less time in getting trained.
Eager Learners are those learners which takes less time in prediction of results but more time in getting trained.
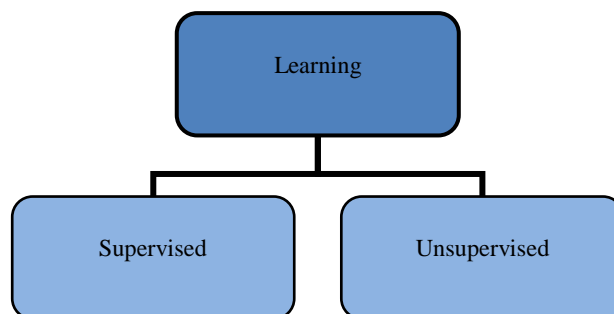Figure 6 shows types of Learner.

Figure 7. Types of Learning

Supervised Learning uses a supervisor or trainer. It uses a labeled data to train the classifier
Unsupervised Learning doesnot use any supervisor or trainer. It uses data which is unlabelled and allows the algorithm to give results without any supervision.
Figure 7 shows types of learning

Table 2 presents several types of Classifier.

Table 1. Classifiers

Classifier	Learner Type	Learning Type
Decision Trees	Eager, Discriminative	Supervised Learning
Naïve Bayes	Eager, Generative	Supervised Learning
K-Nearest Neighbour	Lazy, Discriminative	Supervised Learning
Support Vector Machines	Discriminative	Supervised Learning

V. CONCLUSION AND FUTURE SCOPE

Sentiment Analysis is a technique to identify sentiments in terms on polarities. Aspect Level Sentiment Analysis is a technique which focuses on particular important features in the data. Our work is to perform Aspect Level Sentiment Analysis for which the proposed methodology shown in the paper. Data Collection for the same has been done using API Crawler. The data is collected in the form of reviews of products from amazon.com. The discussion of varioussentiment analysis techniques have been done in the paper. The classifier which can be utilized for performing classification are also been discussed and shown in tabular format. Our Future plan is to implement the proposed methodology on the collected data to perform Aspect Level Sentiment Analysis.

REFERENCES

1. Bo Pang and Lillian Lee(2008), Opinion Mining and Sentiment Analysis, Foundations and Trends in Information Retrieval, Vol. 2, No 1-2(2008) 1-135.
2. Silvana Aciar(2010), Mining Context Information from Consumer'sReviews,http://ids.csom.umn.edu/faculty/gedas/cars2010/Aciar-CARS-2010.pdf
3. A. Gural Vural, B. Barla Cambazoglu, Pinar Senkul (2012),Sentiment Focused Web Crawling, CIKM'12, October 29–November 2, 2012, Maui, HI, USA, ACM Transactions on the web, DOI: 10.1145/2396761.2398564
4. Zhang Hailong, Gan Wenyan, Jiang Bo(2014), Machine Learning and Lexicon Based Methods for Sentiment Analysis: A Survey, Published in: 2014 11th Web Information System and Application Conference, DOI: 10.1109/WISA.2014.55
5. Abhishek Kaushik, Anchal Kaushik, Sudhanshu Naitani(2015), A Study on Sentiment Analysis:Methods and Tools, International Journal of Science and Research (IJSR), Volume 4 Issue 12, December 2015
6. Joseph Mei, Richard Frank (2015), Sentiment Crawling: Extremist Content Collection through a Sentiment Analysis Guided Web-Crawler, IEEE/ACM International Conference,2015, DOI: 10.1145/2808797.2809373
7. Kumar Ravi, Vadlamani Ravi (2015),A Survey on Opinion Mining and Sentiment Analysis:Tasks, Approaches and Applications, Knowledge Based Systems, Volume 89, November 2015,Pages 14-46.
8. Anurag P. Jain, Vijay D. Katkar(2015) Sentiment Analysis of Twitter Data using data mining, 2015 International Conference on Information Processing (ICIP), 10.1109/INFOP.2015.7489492
9. Alessia D'Andrea, Fernando Ferri, Patrizia Grifoni, Tiziana Guzzo(2015), Approaches, Tools and Applications for Sentiment Analysis Implementation, International Journal of Computer Applications (0975 – 8887), Volume 125 – No.3, September 2015
10. Rushlene Kaur Bakshi ; Navneet Kaur ; Ravneet Kaur ; Gurpreet Kaur (2016), Opinion Mining and Sentiment Analysis, 2016 3rd International Conference on Computing for Sustainable Global Development (INDIACom), 31 October 2016
11. Devika M D, Sunitha C, Amal Ganesh (2016), Sentiment Analysis: A Comparative Study On Different Approaches, Fourth International Conference on Recent Trends in Computer Science & Engineering, Chennai, Tamil Nadu, India, Procedia Computer Science 87(2016) 44-49.
12. Fatemeh Hemmantian, Mohammad Karim Sohrabi (2017), A Survey on Classification Techniques for Opinion and Sentiment Analysis, https://doi.org/10.1007/s10462-017-9599-6
13. Sharma, Jyoti. (2017). Review Paper on Text mining and Sentimental analysis. International Journal of Engineering and Computer Science. 6. 10.18535/ijecs/v6i11.05.
14. Shubham Goyal (2017), Review Paper on Sentiment Analysis of Twitter Data Using Text Mining and Hybrid Classification Approach, © 2017 IJEDR | Volume 5, Issue 2 | ISSN: 2321-9939
15. Shahid Shayaa, Noor Ismawati Jaafar, Shamshul Bahri, Ainin Sulaiman, (2018), Sentiment Analysis of Big Data: Methods, Applications, and Open Challenges, IEEE Access, Volume 6, 2018, DOI: 10.1109/ACCESS.2018.2851311
16. Mr. S. M. Vohra, J.B. Teraiya(2018), A Comparative Study of Sentiment Analysis Techniques, Journal of Information, Knowledge and Research in Computer Engineering., ISSN: 0975 – 6760| NOV 12 TO OCT 13 | VOLUME – 02, ISSUE – 02
17. Lucia Martin-Gomez, Maria Navarro Caceres(2018), Applying Data Mining for Sentiment Analysis, Trends in Cyber-Physical Multi-Agent Systems. The PAAMS Collection - 15th International Conference, PAAMS 2017 (pp.198-205)

II. AUTHORS PROFILE

Neha Nandal (Pursuing Ph.D, M.Tech) is currently pursuing her PhD from Manav Rachna University, Faridabad. Her Area of Interest is Pattern Recognition and Machine Learning.

Dr. Jyoti Pruthi is currently working as Associate Professor in Manav Rachna Univerisity, Faridabad. Her research area includes Search Engine, Hidden Web.

Dr. Amit Choudhary is currently working as Associate Professor and HOD of Computer Department in Maharaja Surajmal Institute, Delhi. His area of research is Pattern Recognition and Soft Computing.