



International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 4, Issue 7, July 2016

A Case Study on Sentiment Analysis from Social Big Data

Pooja Kawade, Nitin Pise, Pradnya Kulkarni

M.E Student, Dept. of Computer Engineering, MIT Pune, Maharashtra, India

Associate Professor, Dept. of Computer Engineering, MIT Pune, Maharashtra, India

Asst Professor, Dept. of Computer Engineering, MIT Pune, Maharashtra, India

ABSTRACT: With the fruition of web technology, there is a huge amount of data present in the web for the internet users. These users not only use the available resource, but also give their feedback. Thus additional useful information generated. A huge amount of data is generated by social media. More and more people express their opinion on social media such as facebook, twitter, youtube. Microblogging websites are rich sources of data for belief mining and sentiment analysis. It is important to explore, analyze and consolidate their views for better decision making. Sentiment analysis is a natural language processing and information extraction task that identifies the users views. It is used for extracting opinion about events, products and people and for understanding the current trends or state of the art.

KEYWORDS: Sentiment Analysis, Data Mining, Machine Learning Techniques, Social Media.

I. INTRODUCTION

Sentiment analysis is a type of natural language processing for tracking the mood of the public about a particular product. Sentiment analysis is also known as opinion mining. It collects and examines opinions about the particular product made in blog posts, comments, or tweets. Sentiment Analysis can track a particular topic, many companies use it to track or observe their products, status. For example, if someone is attacking on your brand on social media, sentiment analysis will score the post as immensely negative, and you can create alerts for posts with hyper-negative sentiment scores. A basic task in sentiment analysis is categorizing the polarity of a given text at the document, sentence whether the expressed opinion in a document, a sentence or an entity feature is positive, negative, or neutral. Sentiment classification looks at emotional states such as "angry," "sad," and "happy."

Sentiment Analysis Why Important in Social Media

Social media is more and more popular since mobile devices can access social network easily from anywhere. Therefore, Social media is becoming an important topic for research in many fields. As the number of people using social network are increasing day by day, to communicate with their peers so that they can share their personal feelings everyday and views. In today many companies have been using Social Media Marketing to advertise their products or brands, so it becomes essential for them that they can be able to calculate the success and usefulness of each product.

Sentiment Analysis	Subjectivity Analysis
Positive	Subjective
Negative	Subjective
Neutral	objective

Table 1. Classification of sentiment analysis and subjectivity analysis.

Table 1 shows the sentiment analysis and subjectivity analysis classification [5]. There are three main classification levels in SA: document-level, sentence-level, and aspect-level SA. Document-level SA aims to classify an opinion document as expressing a positive or negative opinion or sentiment.

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 4, Issue 7, July 2016

II. RELATED WORK

- In [2], G.Vinodhini-2012 in this paper Support Vector Machine (SVM), Naive-Bayes this machine learning techniques used for This are used classify reviews into either positive or negative orientation.
- In [5], Pak and Paroubek-2010, N-gram and POS-tags, multinomial Naive Bayes methods are used. In this paper a twitter corpus by using a Twitter API which automatically collected tweets from Twitter as well as annotating those using emoticons. Using that corpus, they built a sentiment classifier which used N-gram and POS-tags as features based on the multinomial Naive Bayes classifier.
- In [3], Machine Learning techniques, natural language processing this used. consider two dataset online review dataset and twitter dataset This paper shows that better accuracy in online review dataset. At last in accuracy term not giving the perfect accuracy for twitter dataset. Because of twitter as tweets consist of short hands as online review were written in more clear way as compared to tweets.
- In [4], Anuj Sharma-2013, used Hybrid sentiment classification model based on boosted SVM. This paper shows that SVM with bagging or boosting significantly outperforms a single SVM in terms of accuracy of sentiment based classification. the drawback is Learning phase of SVM is very time consuming for a large textual data and that some approximate algorithm can reduce the learning time of SVM.
- In [11], Balakrishnan Gokulakrishnan-2012, used SMO, BLR techniques. For Performance of algorithm for a neutral /polar /irrelevant classification it can give the best accuracy in term of SMO classifier. and for positive and negative it gives better accuracy in term of BLR Classifier.
- In [16], Anna Stavrianou-2013, Hybrid technologies, In this paper they used hybrid technology for better accuracy, shows that NLP alone do not capture the information wanted.
- In [15], Erik Boiy-2007 in this paper Symbolic techniques, Machine Learning techniques are used. and best result comes in machine learning technique approximate algorithm can reduce the learning time of SVM.

III. ARCHITECTURE OF SENTIMENT ANALYSIS



Fig.1. Architecture of sentiment analysis.

•Data Collection

Consumers express their sentiments on public forums like the blogs, discussion boards, product as well as on their private logs Social network sites like Facebook and Twitter. Feelings are expressed in different way, context of writing, usage of short forms and slang, making the data huge. Manual analysis of sentiment data is virtually impossible. Therefore, special programming languages like R are used to procedure and analyze the data.

• Text Preparation

Text preparation means filtering the mined data before analysis. Text preparation is nothing but data pre-processing.

• Sentiment Detection

At this stage, each sentence of the opinion is examined for subjectivity. Sentences with subjective information are retained and that which conveys objective expressions are discarded.

• Sentiment Classification

Sentiments can be generally classified into two groups, positive and negative. At this stage of sentiment analysis method, each subjective sentence detected is ordered into groups-positive, negative, good, bad, like, dislike.

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 4, Issue 7, July 2016

- Presentation of Output

The idea of sentiment analysis is to change unstructured text into meaningful data. After the completion of analysis, the text results are displayed on graphs like pie chart, bar chart. Carrying out sentiment analysis is the main task for all the product and service providers today.

IV. TECHNIQUES OF SENTIMENT ANALYSIS

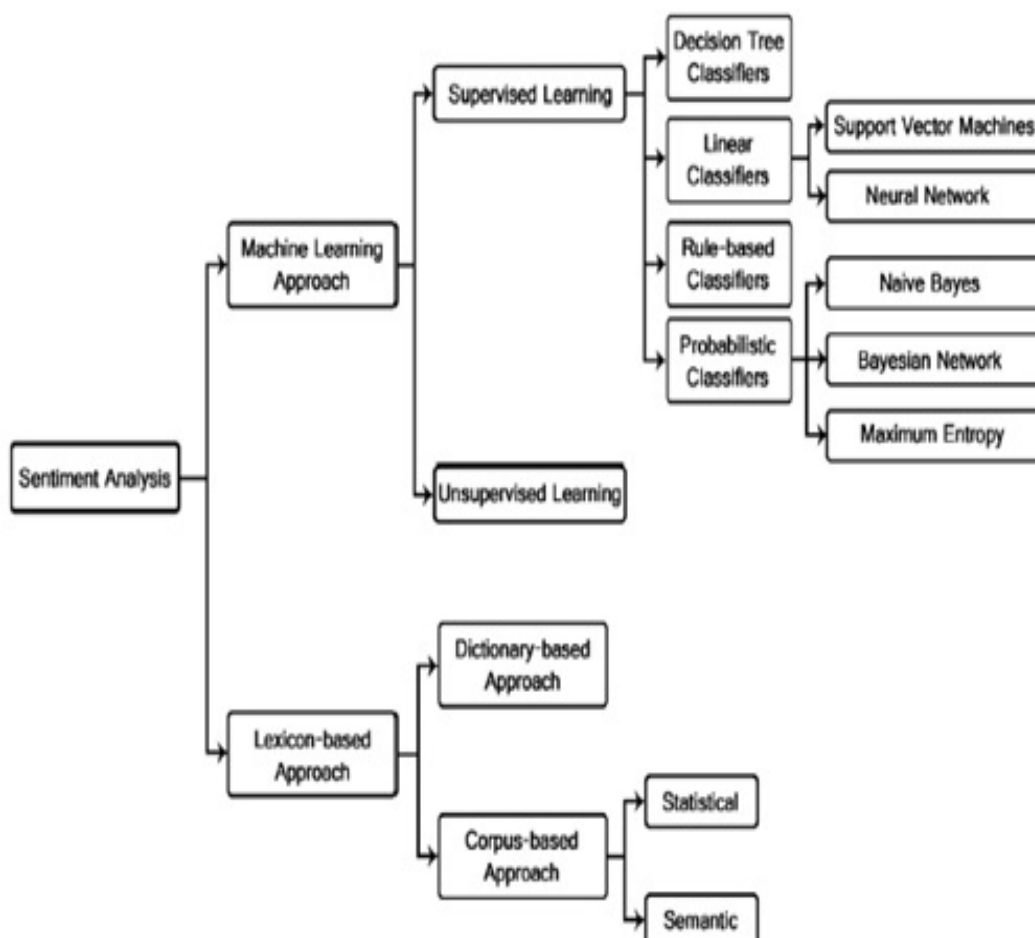


Fig.2. techniques of sentiment analysis

Machine learning techniques classified into two basic techniques as defined below[2].

- Supervised Machine Learning Techniques: In supervised learning, we are given a data set and already know what our correct output should look like, have the idea that there is a relationship between the input and the output. Supervised machine learning techniques are used for classified document or sentences into finite set of class that is into positive, negative and neutral. Training data set is available for all kind of classes. An optimal scenario will agree for the algorithm to correctly determine the class labels for unseen instances. This requires the learning algorithm to simplify from the training data to unseen situations in a "reasonable" way. We are using Support Vector Machine (SVM), Naive-Bayes, K-nearest neighbor (KNN), Logistic regression for classification purpose. SVM efficiently classifies news articles, Blogs into positive, negative or neutral set. Naive-Bayes efficiently classifies tweets or small piece of sentences called Crunches. KNN give good result for sentence level sentiment analysis.

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 4, Issue 7, July 2016

• Unsupervised Machine Learning Techniques: Unsupervised machine learning techniques not use training data set for classification. Clustering algorithms like K-means clustering, Hierarchical clustering used to classify data into categories.

V. APPLICATION OF SENTIMENT ANALYSIS

Following are the major applications of sentiment analysis in real world scenarios.

- Reputation Monitoring - Twitter and Facebook are a central point of many sentiment analysis applications. The most common application is maintain the reputation of a particular brand on Twitter and/or Facebook.
- Result prediction - By analyzing sentiments from related sources, one can predict the probable outcome of a particular event.
- Decision making - Another important application is that sentiment analysis can be used as an important aspect supporting the decision making systems. For instance, in the financial markets investment. There are numerous news items, articles, blogs, and tweets about every public company.
- Product and service Review- The most common application of sentiment analysis is in the area of reviews of customer products and services.

The main sources of data are from the product reviews. These reviews are important to the industry holders as they can take business decisions according to the analysis results of users.

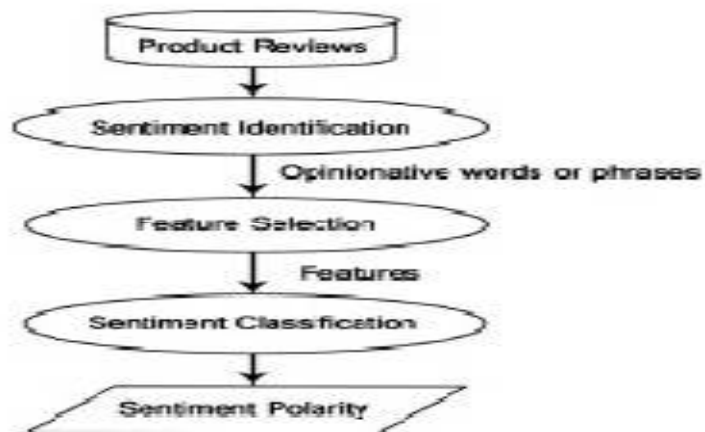


Fig. 3. Sentiment analysis on product review

The social network sites and micro-blogging sites are considered a very good source of information because people share and discuss their feelings about a certain topic freely. They are also used as data sources in the SA process. These fields contain Emotion Detection(ED), Building Resources (BR) and Transfer Learning (TL). Emotion detection aims to extract and analyze sentiments, while the emotions could be explicit or implicit in the sentences.

VI. CHALLENGES IN SENTIMENT ANALYSIS

There are several challenges in Sentiment analysis. The first is a opinion word that is measured to be positive in one situation may be measured negative in another situation. A second challenge is that people don't always express opinions in a same way.



International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 4, Issue 7, July 2016

VII. TOOLS FOR SENTIMENT ANALYSIS

TOOLS FOR SENTIMENT ANALYSIS	TECNIQUES USED BY TOOLS
EMOTICONS	Emoticons contained in the text
LIWC	Dictionary and sentiment classified categories
SentiStrength	LIEC dictionary with new features to strength and weak sentiments
SentiWordNet	Lexical dictionary and scores obtained by semi-machine learning approaches
Sentiment140	API that allows classifying tweets to polarity classes positive, negative and neutral.
RapidMiner	For machine learning, data mining, text mining, predictive analytics and business analytics
Weka	Waikato Environment for Knowledge Analysis (Weka) for data preprocessing, clustering classification, regression, visualization.

VIII. CONCLUSION

In this paper we discussed architecture techniques and applications of sentiment analysis and we have seen different tools that are useful to find out different classification and analysis and we also study the challenges of sentiment analysis.

REFERENCES

1. G.Vinodhini, RM.Chandrasekaran, "Sentiment Analysis and Opinion Mining: A Survey", IJARCSSE Volume 2, Issue 6, June 2012.
2. Jalaj S. Modha, Gayatri S. Pandi, Sandip J. Modha, "Automatic Sentiment Analysis for Unstructured" ,Data Volume 3, International Journal of Advanced Research in Computer Science and Software Engineering Issue 12, December 2013.
3. Rajni Singh, Rajdeep Kaur, "Sentiment Analysis on Social Media and Online Review" International Journal of Computer Applications (0975 8887) Volume 121 No.20, July 2015.
4. Anuj Sharma, Shubhamoy Dey, "A Boosted SVM based Sentiment Analysis Approach for Online Opinionated Text", RACS13, October 14, 2013, Montreal, QC, Canada. ACM 978-1-4503-2348-2/13/10.
5. Ms. Gaurangi Patil, "Sentiment Analysis Using Support Vector Machine", International Journal of Innovative Research in Computer and Communication Engineering, Vol. 2, Issue 1, January 2014.
6. Walaa Medhat, Ahmed Hassan, Hoda Korashy, "Sentiment analysis algorithms and applications: A survey", Ain Shams University Ain Shams Engineering Journal 2014.
7. Krushikanth R. Apala, Merin Jose, Supreme Motnam, "Prediction of Movies BoxOffice Performance Using Social Media", 2013 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining.
8. Neethu M S, Rajasree R, "Sentiment Analysis in Twitter using Machine Learning Techniques", IEEE 31661, 4th ICCNT 2013 July 4 - 6, 2013, Tiruchengode, India.
9. Onur Kucukunc, "A Large-Scale Sentiment Analysis for Yahoo! Answers", WSDM12, February 812, 2012, Seattle, Washington, USA. ACM 978-1-4503-0747-5/12/02
10. Rutuja Tikait, "Product Aspect Ranking Techniques: A Survey", International Journal of Innovative Research in Computer and Communication Engineering, Vol. 2, Issue 11, November 2014.
11. BalakrishnanGokulakrishnan, "Opinion Mining and Sentiment Analysis on a Twitter Data Stream", The International Conference on Advances in ICT for Emerging Regions - ICTer 2012 : 182-188.
12. Tetsuya Nasukawa, "Sentiment Analysis: Capturing Favorability Using Natural Language Processing" K-CAP03, October 2325, 2003, Sanibel Island, Florida, USA. ACM 1-58113-583-1/03/0010.
13. Alessia DAndrea Approches, "Tools and Applications for Sentiment Analysis Implementation", International Journal of Computer Applications (0975 8887) Volume 125 No.3, September 2015.
14. Anis Zarrad, "The Evaluation of The Public Opinion", 2014 IEEE/ACM 7th International Conference on Utility and Cloud Computing.
15. Erik Boiy; Pieter Hens; Koen Deschacht; Marie-Francine Moens, "Automatic Sentiment Analysis in On-line Text", Proceedings ELPUB2007 Conference on Electronic Publishing Vienna, Austria June 2007.
16. Anna Stavrianou, "NLP-based Feature Extraction for Automated Tweet Classification", 2013 IEEE International Conference on Advances in Social Networks Analysis and Mining.