

Survey on Sentiment Analysis

¹Neelam Soni, ²Dr. Nitesh Dubey

¹ Research Scholar Dept of CSE

² Professor, Dept of CSE

^{1,2} Global Nature Care Sangathan Group of Institution,
Jabalpur, Madhya Pradesh, India.

Abstract- Sentiment Analysis (SA) is an ongoing field of research in text mining field. SA is the computational treatment of opinions, sentiments and subjectivity of text. This survey paper tackles a comprehensive overview of the last update in this field. Many recently proposed algorithms' enhancements and various SA applications are investigated and presented briefly in this survey. These articles are categorized according to their contributions in the various SA techniques. The related fields to SA (transfer learning, emotion detection, and building resources) that attracted researchers recently are discussed. The main target of this survey is to give nearly full image of SA techniques and the related fields with brief details. The main contributions of this paper include the sophisticated categorizations of a large number of recent articles and the illustration of the recent trend of research in the sentiment analysis and its related areas.

Keywords- Data Analytics, sentiment Analysis, Decision making..

I. INTRODUCTION

Analytics is an art of processing raw data to extract some reasonable information. Data Analytics is widely used in many industries and organization to make a better Business decision. By applying analytics to the structured and unstructured data the enterprises brings a great change in their way of planning and decision making. Data analysis is the process of verifying, cleaning, and transforming in order to retrieve useful information from the data. This information will be more helpful in suggesting business conclusions and decisions making. Data Analysis has a variety of angles and methods that combines many techniques in order to provide better accuracy. One of the most popular methods of data analysis technique is data mining that mainly concentrates on modelling and discovery of knowledge for prediction process rather than descriptive purposes. Predictive analytics is mainly used for predicting forecasting/classification where as text analytics make use of statistical, linguistic and structural techniques in order to retrieve information from text sources [1]. This text sources are mostly in the form of unstructured data.

Sentiment analysis (or) opinion mining plays a significant role in our daily decision making process [2]. These decisions may range from purchasing a product such as mobile phone to reviewing the movie to making investments---all the decisions will have a great impact on the daily life. In ancient days before buying a product / service people will seek opinion from their friends, neighbors, etc. But in internet era it is easy to seek opinion from different people around the world. Now-a- days people before buying any product/service will make a glance on review sites (e.g. CNET), e-Commerce sites (e.g. Amazon, eBay) and social media (e.g. twitter) to get a feedback about the specific product (or) service in market[3].

Sentiment Analysis makes use of 3 terms in order to fetch the sentiment .That is object and feature, opinion holder, opinion and orientation. Sentiment Analysis deals with several technical challenges such as object identification, opinion orientation classification, and feature extraction. Usually sentiment analysis can be performed using supervised and unsupervised learning such as naive Bayes, Neural Networks, Support Vector Machine [4]. Among these three techniques SVM is considered to be more suitable for sentiment Analysis. Sentiment classification can be performed in 3 stages such as

- Document level
- Sentence level
- Feature level

In document and sentence level the sentiment analysis make use of only a single object and extracts only a single opinion from the single opinion holder. But this type of assumptions is not suitable for many situations. Extracting sentiment for entire document/blog will not be efficient as extracting sentiment by considering aspects of each subject in the particular sentence.

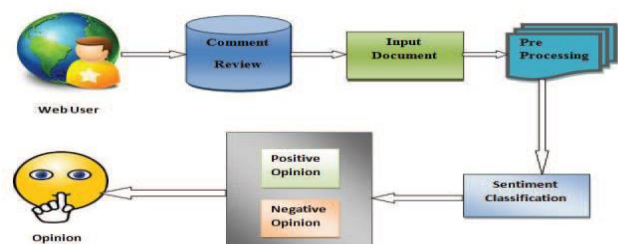


Fig 1.1: Sentiment Analysis concepts.

The Fig 1.1 shows the overall process of sentiment analysis which starts from preprocessing of review dataset and continuous the sentiment classification or opinion mining through the various machine algorithms or some other dictionary based techniques.

The data sets used in SA are an important issue in this field. The main sources of data are from the product reviews. These reviews are important to the business holders as they can take business decisions according to the analysis results of users’ opinions about their products. The reviews sources are mainly review sites. SA is not only applied on product reviews but can also be applied on stock markets [5], news articles, [6] or political debates [7]. In political debates for example, we could figure out people’s opinions on a certain election candidates or political parties. The election results can also be predicted from political posts. The social network sites and micro-blogging sites are considered a very good source of information because people share and discuss their opinions about a certain topic freely. They are also used as data sources in the SA process. There are many applications and enhancements on SA algorithms that were proposed in the last few years [8]. This survey aims to give a closer look on these enhancements and to summarize and categorize some articles presented in this field according to the various SA techniques.

1.2 Different Approaches to Sentiment Analysis

The Sentiment Classification (SC) techniques, as shown in Fig. 1.2, are generally classified on the basis of following Techniques:

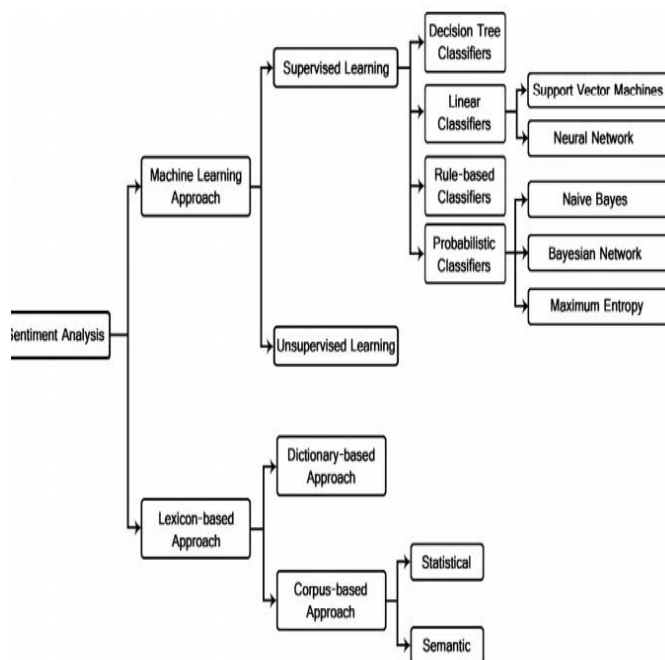


Fig. 1.2 Sentiment classification techniques [9].

A. Machine Learning

Machine learning algorithms can be addressed as a combination of methods to automatically detect the available pattern in the given set of data. It makes use of undiscovered patterns to forecast the future data (or) to implement the decision making under uncertainty. Machine learning can be performed in 2 ways such as supervised and unsupervised.

Supervised learning is performed by considering the target value (i.e. label) and unsupervised learning is conducted by not considering the target value (i.e. label). There are various types of algorithms for supervised learning such as classification (Decision tree, Naive bayes etc) and unsupervised learning algorithm such as clustering (SOM, Neural network) [11].

B. Dictionary Based

Dictionary oriented approach is considered as easiest way for performing sentiment analysis. Dictionaries such as WordNet, SentiWordNet are publically available to perform SA. Dictionary can be build by providing a set of sentiment words Iteration of Algorithm ends when there are no new words found to add in dictionary.

C. Ontology Based

Ontology is a word mostly used in philosophy and it is used in many different areas which mean that “theory of Existence/ Nature of Being”. Ontology is related to the conceptualization which captures a structure of specific domain. Ontology clarifies the concepts in the domain and also the relationship between those concepts. Different types of ontologies are available such as upper ontology, domain ontology and hybrid ontology. Among this domain ontology will be very suitable for sentiment analysis. Researchers mostly prefer ontology in feature extraction phase of sentiment analysis. One of the important thing to be addressed in ontology is taxonomy (i.e.) Ontology will be constructed in the hierarchical form. Ontology can be developed by using several ontology languages such as OWL and RDF. Most commonly ontology was developed by using the tool protégé. Ontology includes several components such as Individuals, classes and properties.

II. SENTIMENT ANALYSIS CHALLENGES

Now-a-days Sentiment Analysis is dealing with various issues such as:

A) Polarity Shift

Polarity Shift is a most important issue to be addressed in Sentiment Analysis. Polarity Shift means that Polarity

(Sentiment) of the sentence is calculated in different way from the polarity actually expressed in the Sentence. This problem is mainly arises due to polarity shifters such as negation (e.g. "I don't like this car") and contrast (e.g. "good, but not in my own style"). In the above mentioned example the sentence "I don't like this car" is very similar to "I like t his car". Here the polarity shifter is "Don't".

B) Binary Classification

Binary Classification is another important problem to be addressed in which the given review's Polarity is classified only by using "Positive", "Negative" by ignoring the "Neutral". This type of problem mainly arises when the sentiment classification is purely based on machine learning algorithms. Opinion mining that only considers positive and Negative will not have good accuracy. Now-a-days the classification is extended by considering 5 possibilities such "Positive", "Strong Positive", "Negative", "Strong Negative" and "Neutral". By increasing the classification category it is possible to improve the accuracy of the opinion mining.

C) Data Sparsity problem

Third issue to be addressed is Data Sparsity problem which is caused due to the imposed character limit in micro blog/social media websites. For instance the maximum character limit in twitter is 140. Due to this limitation people will not express their opinion in clear manner. All these three issues are closely related to the accuracy of the sentiment analysis.

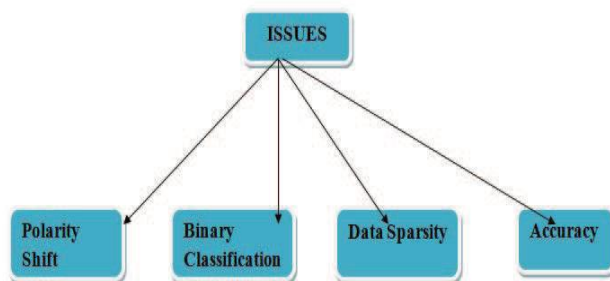


Fig 1.3: Issues in Sentiment Analysis.

III. CONCLUSION

Sentiment analysis (or) opinion mining plays a significant role in business decision making. Many of the organization and enterprises will take their business decision only based on their customer review. There are several techniques for performing sentiment analysis. This paper specifies the sentiment analysis under 3 categories such as Machine learning, dictionary based, Ontology based and gives the clear knowledge about various approaches. This survey gives the knowledge about the sentiment analysis issues such as

Polarity shift problem, data sparsity, binary classification briefly and how they are handled in different domains.

REFERENCES

- [1] RuiXia,FengXu,JianfeiYu," Polarity shift detection, elimination and ensemble: A three stage model for document-level sentiment analysis" Information Processing and Management 52 (2016) 36– 45.
- [2] Wanxiang Che, Yanyan Zhao, Honglei Guo, Zhong Su, and Ting Liu," Sentence Compression for spect-Based Sentiment Analysis" IEEE/ACM TRANSACTIONS ON AUDIO, SPEECH, AND LANGUAGE PROCESSING, VOL. 23, NO. 12, DECEMBER 2015
- [3] Fangzhao Wu, Yongfeng Huang, Yangqiu Song, Shixia Liu," Towards building a high quality micro blog-specific Chinese sentiment lexicon", Decision Support Systems-2016 .
- [4] V.K. Singh, R. Piryani, A. Uddin, P. Waila," Sentiment Analysis of Movie Reviews", conference on IEEE-2013
- [5] Isidro Peñalver-Martinez, Francisco Garcia-Sanchez , Rafael Valencia-Garcia," Feature-based opinion mining through ontologies", Expert Systems with Applications-2014.
- [6] Efstratios Kontopoulo, Christos Berberidis, Theologos Dergiades , Nick Bassiliades," Ontology- based sentiment analysis of twitter posts", Expert Systems with Applications 40(2013)4065-4074.
- [7] Ziang Lia, Wei Xu, Likuan Zhang, Raymond Y.K. Lau," An Ontology-based Web Mining Method for nemployment Rate Prediction", Decision Support Systems -2014.
- [8] Jian Ma, Wei Xu, Yong-hong Sun, Efraim Turban, Shouyang Wang, and Ou Liu," An Ontology-Based Text-Mining Method to Cluster Proposals for Research Project Selection", IEEE TRANSACTIONS ON SYSTEMS, MAN, AND CYBERNETICS—PART A: SYSTEMS AND HUMANS, VOL. 42, NO. 3, MAY-2012
- [9] Farman Alia, Kyung-Sup Kwaa,Yong-Gi Kimb," Opinion mining based on fuzzy domain ontology and Support Vector Machine: A proposal to automate online review classification", Applied Soft Computing-2016.
- [10] Doug Lundquist, Kunpeng Zhang, and Aris Ouksel,"Ontology- Driven Cyber-Security Threat Assessment Based on Sentiment Analysis of Network Activity Data", 2014 IEEE International Conference on Cloud and Autonomic Computing.
- [11] Frye, L., Cheng, L., and Kaplan, R.,"A Methodology to Identify Complex Network Attacks," International Conference on Security and Management, Las Vegas, July 2011.