

A SURVEY ON SENTIMENT ANALYSIS METHODS AND APPROACH

Ms.A.M.Abirami

Dept. of Information Technology
Thiagarajar College of Engineering
Madurai, Tamilnadu, India

abiramiam@tce.edu

Abstract— 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. Sentiment analysis (or) opinion mining plays a significant role in our daily decision making process. 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 huge impact on the daily life. Sentiment Analysis is dealing with various issues such as Polarity Shift, accuracy related issues, Binary Classification problem and Data sparsity problem. However various methods were introduced for performing sentiment analysis, still that are not efficient in extracting the sentiment features from the given content of text. Naïve Bayes, Support Vector Machine, Maximum Entropy are the machine learning algorithms used for sentiment analysis which has only a limited sentiment classification category ranging between positive and negative. Especially supervised and unsupervised algorithms have only limited accuracy in handling polarity shift and binary classification problem. Even though the advancement in sentiment Analysis technique there are various issues still to be noticed and make the analysis not accurately and efficiently. So this paper presents the survey on various sentiment Analysis methodologies and approaches in detailed. This will be helpful to earn clear knowledge about sentiment analysis methodologies. At last the comparison is made between various paper's approach and issues addressed along with the metrics used.

Keywords—Data Analytics, sentiment Analysis, Decision making.

I. INTRODUCTION

DATA 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 modeling 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. 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. These decisions may

Ms.V.Gayathri

Dept. of Information Technology
Thiagarajar College of Engineering Madurai,
Tamilnadu, India

gayathrivairam2015@gmail.com

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.

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 naïve Bayes, Neural Networks, Support Vector Machine. 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 are 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.

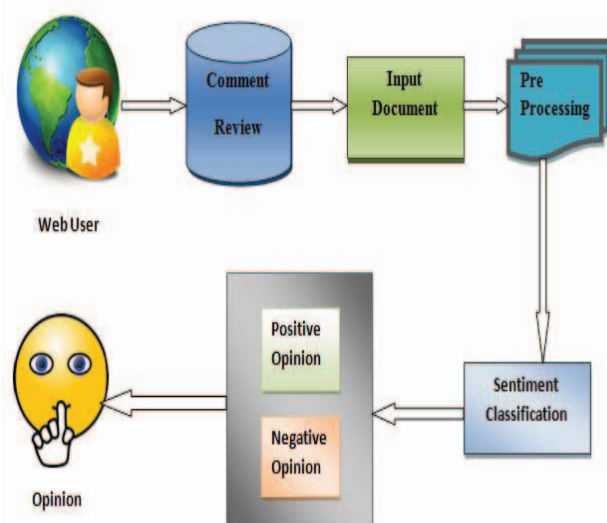


Fig 1.1: Sentiment Analysis concepts

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.

II. SENTIMENT ANALYSIS APPROACHES

Sentiment Analysis systems are generally classified on the basis of following Techniques

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 make 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).

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 is no new 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 means 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. The table.2.1 shows the pros and cons of various sentiment analysis techniques

III. SENTIMENT ANALYSIS CHALLENGES

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

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 this car". Here the polarity shifter is "Don't".

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 as

"Positive", "Strong Positive", "Negative", "Strong Negative" and "Neutral". By increasing the classification category it is possible to improve the accuracy of the opinion mining.

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.

Table.2.1. Comparison of Methods

	Approach	Advantage	Limitation
Machine Learning	Make use of supervised and unsupervised techniques to perform sentiment analysis.	It can be customized to deal with different domains.	Classification is only limited to 2 categories such as "positive", "Negative".
Dictionary Based	It make use of reference dictionary to classify the individual sentence.	Computation overhead is lesser as there is no training of dataset.	It is not suitable for context /domain specific classification.
Ontology Based	Used in feature extraction stage of sentiment analysis.	Mainly take cares of semantic relationship between the features.	Updating ontology is difficult task

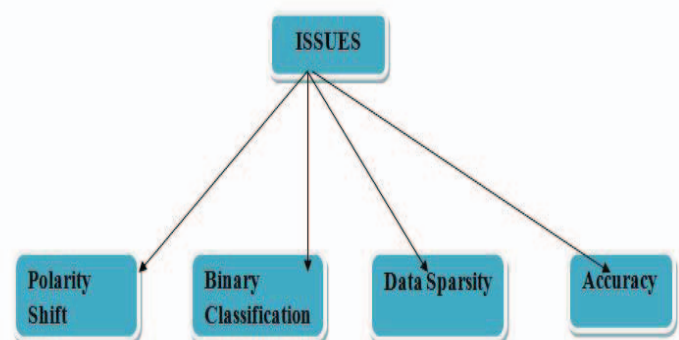


Fig 3.1:Issues in Sentiment Analysis

IV. RELATED WORK

This paper discusses about different facets and aspects of methodologies and approaches used in Sentiment Analysis.

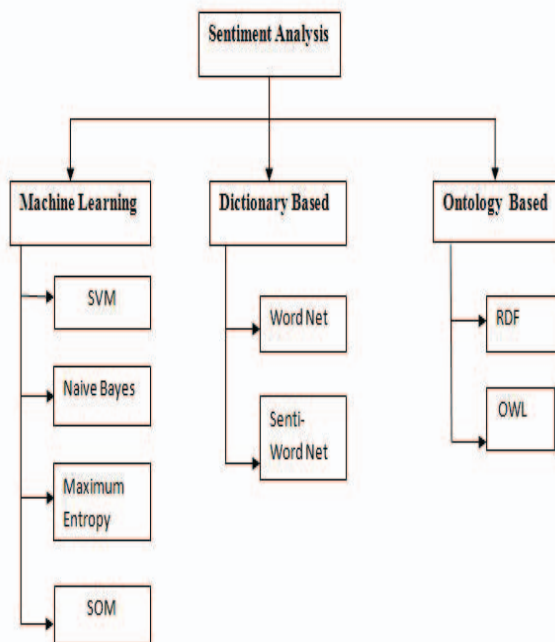


Fig.4.1. Sentiment Analysis methodology taxonomy

A. Machine Learning

[1] In this paper the proposed framework provide a clear understanding about the polarity shift problem. Sentiment Analysis is affected by many factors. Among that polarity shift problem is considered to be very dangerous factor that destroys the complete classification performance of traditional machine learning based sentiment classification. Usually the review data is represented in the form of Bag Of Words(BOW) that entirely terminates the semantic correlation between the text. The existing system make use of term counting method addressing the polarity shift problem. The proposed polarity shift Detection, Elimination and Ensemble(PSDEE) performs detection of hybrid polarity shifts. To perform hybrid polarity shift detection it make use of 3 levels of cascading model. Polarity shift problem arises if there is a polarity shifters or valance shifters such as negation (implicit or explicit),contrast, sentiment inconsistency in the text review. Proposed methodology make use of Rule based Method is used for detecting negations and contrast polarity shift and statistical methods are used for detecting implicit inconsistency. The proposed PSDEE was examined in four domains which are extracted from the Amazon website.

[2] This paper proposes a framework for aspect/feature based sentiment analysis along with the sentence compression technique. Aspect based sentiment analysis is performed based on syntactic features which poses a chance for over natural problem. This type of issue makes the sentiment analysis too difficult to handle the syntactic parsers used in the opinion mining technique. The proposed framework develops a innovated sentence compression technique before the sentiment analysis. For compressing a text for sentiment analysis 2 schemes are used. That is syntactic compression

and extractive compression technique. Compared to extractive compression technique syntactic is considered to be more efficient because it compress the text by removing the unimportant words. The proposed technique makes use of A-P (Aspect-Polarity) collection based sentiment analysis. Most of the aspect based sentiment analysis focus on the relationship between the aspects and the polarity words which extremely affects the efficiency. To solve this problem the proposed framework makes use of syntactic patterns.

B. Dictionary Based

[3] The proposed system develops a innovated micro blog specific sentiment lexicon which is based on data driven approach. Sentiment lexicon is considered to be one of the most important components of sentiment analysis. Existing sentiment lexicon are not performing well for micro blogs because all the reviews in the blogs contains a user specific words such as “cool”, “tnx”. These type of words can’t be correctly recognized by the existing framework. The proposed framework for handling micro blog based sentiment lexicon is constructed by integrating 3 types of sentiment knowledge such as word opinion knowledge for sentiment score, opinion similarity knowledge for expressing sentiment similarity and primary knowledge which is extracted from the traditional lexicons. The proposed framework also develops a new word detection method by using a proposed new word detection algorithm and that new word will be added to the dictionary. The proposed framework was validated using a Chinese micro blog of 17.2 million messages and the results were compared with the existing sentiment lexicons in terms of subjectivity identification and opinion polarity classification in both sentence and document level opinion mining.

[4] The Author Proposes a dictionary based technique for domain specific sentiment analysis on the movie review dataset. The author make use of lexicon known as sentiword net(SWN-publically available dictionary) including adjectives, adverbs, and verbs. Document level analysis involves by using linguistic features ranging from adverb+adjective to adverb+adjective+verb combination. For Aspect based SA it make use of 3 steps such as 1.)Identification of Aspect from the review 2.)locating the aspect 3.)Detecting the sentiment polarity of review. The results of proposed lexicon based approach is compared with Alchemy API and the sentiwordnet provides accurate result than the Alchemy API.

C. Ontology Based

[5] This paper proposes a advanced framework for opinion mining that correlates all the merits of semantic web guided solutions to tremendously improve the overall results of traditional NLP(Natural Language Processing). The proposed framework make use of domain ontology at feature extraction stage. This enhancement makes a huge changes in the feature based sentiment classification. Existing machine learning techniques classify the words into limited category such as positive/negative. Existing system also performs sentiment classification at the document level (i.e.) If the document includes huge no of positive than negative terms, then it will be considered to be a positive document otherwise negative document. They make of Movie Review as a data set to check the performance of proposed model.

Table4.2 Comparison of methodology and approach

S.no	Title	Issues Addressed	Proposed Technique	Dataset used	Accuracy	Limitation
1	RuiXia, FengXu, JianfeiYu(2016)	Polarity Shift Problem	PSDEE Approach based on Rule based and statistical method	Movie Review	87.1%	Limited in Accuracy
2	Yanyan Zhao, Honglei Guo,Zhong su and Ting Liu (2015)	Challenge to syntactic parsers	Sentiment-compression technique before the aspect based sentiment analysis	Chinese Blog Review Dataset	No_comp_ssc =88.78% Manual_comp_ssc = 88.95 Auto_comp_scc=87.95	Extractive compression technique fails to achieve accuracy
3	Fangzhao Wu, Yongfeng Huang,Yangqiu Song,Shixia Liu(2016)	To improve the accuracy	Micro blog specific sentiment lexicon	Dataset from Tecent Webio with 2013 labeled messages on 20 topics	84.3%	Only for Chinese blog
4	V.K. Singh, R. Piryani, A. Uddin, P. Waila	Polarity shift and Accuracy	Lexicon based Approach	Movie Review Dataset	SWN(ACC) = 77.6% SWN(AAAVC) = 78.7% Alchemy API = 77.4%	Difficult to update the dictionary
5	Isidro Peñalver-Martinez,Francisco Garcia-Sanchez Rafael Valencia Garcia(2014)	Binary Classification problem,	Ontology at feature extraction stage	Twitter Dataset	82.9%	Not suitable for all Domains
6	Efstratios Kontopoulos,Christos Berberidis , Theologos Dergiades Nick,Bassiliades(2015)	Data Sparsity Problem	Domain Ontology	Twitter Dataset	82.4%	Limited Accuracy when no of tweets gets increased
7	Jian Ma,Wei Xu,Yong-hong sun,Efraim Turban,shouyang wang and ou Liu(2012)	To improve the efficiency of research project selection	Ontology Based Text Mining Method(OTMM) and SOM	110000 proposals from NSFC(National Natural Science Foundation of China)	91.2%	Time Consuming Technique
8	Ziang Lia, Wei Xu,Likuran Zhang(2014)	To improve the accuracy of unemployment rate prediction	Domain Ontology	UIC(Unemployment Initial Claims) values between jan.2004 and Mar.2012 from the official website of the US Dept of Labor	81.7%	Rate Prediction is Not Accurate
9	Farman Ali, Kyung-SupKwak, Yong-Gi Kim(2016)	Binary Classification Problem and Accuracy	Fuzzy Ontology with Machine learning Technique	Hotel Review	82.70%	Increased Complexity

Proposed model provides 89.6% accuracy as compared to other methods. In future they will extend their work to make use of some other domains such as product reviews and tourism.

[6] This paper proposes a enhanced ontology based technique to improve the accuracy of sentiment analysis in the twitter post. Proposed technique provides solution to deal with imposed character limit problem. Most of the social media such as twitter are limited to certain character limit which leads to inconsistency in the word of sentence. The proposed

framework make use of domain based ontology technique by using OWL. Existing machine learning technique and lexicon based techniques performs sentiment analysis on the tweet by treating each of the tweet as a uniform statement and assigns sentiment score for entire post as whole. Proposed framework splits the entire tweet based on the subject's aspect and assigns sentiment score to each aspect. Proposed model finds some difficulties when the no of tweets get increases. In future they will expand the proposed system to deal with other domains.

[7] The proposed model provides a new enhancement for the project selection process by improving the overall effectiveness and efficiency of various Text mining methods. Existing text mining methods(TMM) for proposal selections are based on keyword matching technique which concentrates only on English proposals and it is very time consuming. Proposed framework suits for handling both the English and Chinese proposals. The framework make use of ontology based text mining method(OTMM) to cluster the research proposals. It make use of Self Organizing Map(SOM) for clustering. The performance of proposed technique was validated based on the proposal selection process in National Natural Science Foundation of China. The proposed model was compared with the traditional text mining in terms of performance. The future work of the proposed framework will be extended to cluster the external reviewers based on their interested area.

[8] In this paper the Author proposed a ontology based technique to improve the unemployment rate prediction problem. This proposed framework was constructed using domain ontology that relates the concepts which are more specific to unemployment with their semantic relationships. Existing system make use of Univariate Time Series Model (ARFIMA- Autoregressive Fractionally Integrated Moving Average) to predict the unemployment rate. But this model is not efficient in rate prediction. This proposed experiment was compared with traditional time series based forecasting models such as ARIMA. In future they will extend this proposed framework to support decision making in other fields.

[9] The proposed framework was developed to increase the accuracy of opinion mining by combing traditional machine learning algorithm(i.e. SVM-Support Vector Machine) along with the Fuzzy domain ontology. Traditional machine learning algorithm founds limitation in classification of sentiment. That is review features will be classified only to limited category such as Positive/Negative. Since the proposed FDO framework make use of fuzzy based ontology the accuracy of feature extraction increase to tremendous level. The proposed system brings a clear understanding that the fuzzy based ontology is more domain specific as compared with traditional crisp ontology. The FDO was constructed by using the protégé OWL-2(Ontology Web Language) and JAVA. The proposed framework also overcomes the binary classification problem by increasing the category to 5 such as positive, strong positive, negative, strong negative. The proposed framework was evaluated in the movie review domain and it is compared with the traditional SVM algorithm. The accuracy of opinion mining increases from 71.87 to 8.70. In future the proposed system will be expanded to make use of ype-2 fuzzy ontology and SVM-based opinion mining for recommendation system.

The table 4.2 gives the clear summary of various approaches used in sentiment Analysis. Among this ontology along with machine learning technique will be very suitable for opinion mining with improved accuracy.

V.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 Kontopoulos, 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 Oukssel,"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.
- [12] J. Ma, W. Xu, E. Turban, S. Wang, O. Liu,"An ontology based text mining method to cluster proposals for research project selection"IEEE Transactions on Systems, Man, and Cybernetics-2012.
- [13] S.Feng, L. Wang, W. Xu, D. Wang, G. Yu,"Unsupervised learning Chinese sentiment lexicon from massive microblog data",Advanced Data Mining and Applications, Springer-Verlag Berlin Heidelberg 2012 .
- [14] F. Bobillo, U. Straccia,"Fuzzy ontology representation using OWL 2",Approx. Reason., 52 - 2011.
- [15] H. Kang, S. Yoo, D. Han," Senti-lexicon and improved Naïve Bayes algorithms for sentiment analysis of restaurant reviews", Expert Systems with Applications, 39 -2012