

Sentiment Analysis Methods and Approach: Survey

Saurabh Dorle¹, Dr. Nitin N. Pise²

¹ME-II year, Department of Computer Engineering,

Maharashtra Institute of Technology, Pune, Maharashtra, India

sdd.saurabhd@gmail.com

²Department of Computer Engineering, Maharashtra institute of Technology, Pune, Maharashtra, India

nitin.pise@mitune.edu.in

ABSTRACT

Nowadays, social media present a valuable source for business decision support and Data Analytics is widely used in many industries and organization to make a better Business decision. By applying analytics to the 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 or Opinion analysis is performed to identify the opinion of peoples. It can be performed using Lexicon Based approach or Machine Learning based approach. Some methods are still not efficient in extracting the sentiment features from the given content of text. Naive Bayes, Support Vector Machine are the machine learning algorithms used for sentiment analysis which has only a limited sentiment classification category ranging between positive and negative. 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. This will be helpful to earn clear knowledge about sentiment analysis methodologies.

Keywords: *Data Analytics, sentiment/opinion 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 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-Commerce sites and social media to get a feedback about the specific product or service in market.

Sentiment Analysis (SA) or Opinion Mining (OM) is the computational study of people's opinions, attitudes and emotions toward an entity. The entity

can represent individuals, events or topics. 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, and 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 these types 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.

II. SENTIMENT ANALYSIS APPROACHES

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 (Self-Organizing Map i.e. 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. Lexicon dictionary can be created by adding a set of sentiment words. Then retrieved text is compared with the words in the lexicon dictionary and the score is assigned to the statement. For positive statement (1), for negative statement (-1) and for neutral (0) are assigned.

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 Ontology Web Language (OWL) and Resource Description Framework (RDF). Most commonly ontology was developed by using the tool protégé. Ontology includes several components such as Individuals, classes and properties.

III. CHALLENGES IN SENTIMENT ANALYSIS:

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 bike") and contrast (e.g. "good, but it's not my style"). In the above mentioned example the sentence "I don't like this bike" is very similar to "I like t his bike". 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 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.

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.

IV. RELATED WORK:

Different methodologies and approaches used in Sentiment Analysis.

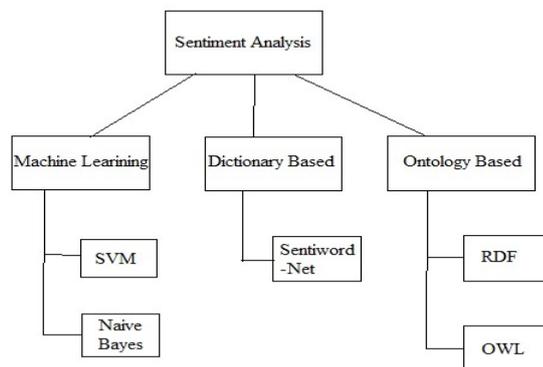


Fig.1 Sentiment Analysis Methods

[5] 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 texts. The existing system makes 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 makes use of 3 levels of cascading model. Polarity shift problem arises if there is a polarity shifters or valance shifters such as negation, 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.

[6] 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 an 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 Aspect-Polarity (A-P) 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.

A. Dictionary Based:

[1] The proposed system develops an 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 lexicons are not performing well for micro blogs because all the reviews in the blogs contains a user specific words such as "Thnx", "gud". These types 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.

2] 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 SentiWordNet (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 Sentiment Analysis 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 are compared with Alchemy API and the SentiWordNet provides accurate result than the Alchemy API.

B. Ontology Based:

[4] 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 makes use of domain ontology at feature extraction stage. This enhancement makes 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. Dataset of Movie Reviews is used to check the performance of proposed model.

[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 finds 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 table IV.I 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 analysis.

Table 1:

No.	Title	Issues	Proposed Technique	Dataset	Limitation
1	Fangzhao Wu, Yongfeng Huang, Yangqiu Song, Shixia Liu," Towards building a high quality micro blog-specific Chinese sentiment lexicon", Decision Support Systems-2016	Not Accurate	Micro blog specific sentiment lexicon	Dataset from Tecent Webio with 2013 labeled messages on 20 topics	Only for Chinese blog
2	V.K. Singh, R. Piryani, A. Uddin, P. Waila," Sentiment Analysis of Movie Reviews", conference on IEEE-2013	Polarity shift and Accuracy	Lexicon based Approach	Movie Review Dataset	Difficult to update the dictionary
3	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	Binary Classification Problem and Accuray	Fuzzy Ontology with Machine learning Technique	Hotel Review	Increased Complexity
4	Isidro Peñalver-Martinez, Francisco Garcia-Sanchez , Rafael Valencia-Garcia," Feature-based opinion mining through ontologies", Expert Systems with Applications-2014.	Binary Classification problem,	Ontology at feature extraction stage	Twitter Dataset	Not suitable for all Domains
5	RuiXia,FengXu,JianfeiYu," Polarity	Polarity Shift	PSDEE Approach	Movie Review	Limited in

	shift detection, elimination and ensemble: A three stage model for document-level sentiment analysis" Information Processing and Management 52 (2016) 36–45.	Problem	based on Rule based and statistical method		Accuracy
6.	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	Challenge to syntactic parsers	Sentiment compression technique before the aspect based sentiment analysis	Chinese Blog Review Dataset	Extractive compression technique fails to achieve accuracy
7.	Afef Walha, Faiza Ghazzi and Faïez Gargouri "A Lexicon Approach to Multidimensional Analysis of Tweets Opinion"	Opinion extraction from informal text	Opinion lexicon building method, POLSentiment algorithm, based on lexical resources	SentiwordNet, Sanders twitter corpus.	Increased Complexity

V.CONCLUSION

Sentiment analysis or opinion analysis 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 survey paper presented an overview on the sentiment analysis methods such as Machine learning, Lexicon/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. Fangzhao Wu, Yongfeng Huang, Yangqiu Song, Shixia Liu," Towards building a high quality micro blog-specific Chinese sentiment lexicon", Decision Support Systems-2016.
2. V.K. Singh, R. Piryani, A. Uddin, P. Waila," Sentiment Analysis of Movie Reviews", conference on IEEE-2013
3. 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.
4. Isidro Peñalver-Martinez, Francisco Garcia-Sanchez, Rafael Valencia-Garcia," Feature-based opinion mining through ontologies", Expert Systems with Applications-2014.
5. 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.
6. 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
7. Afef Walha, Faiza Ghazzi and Faïez Gargouri "A Lexicon Approach to Multidimensional Analysis of Tweets Opinion" (2016)