

International Journal of Computer Science and Mobile Computing

A Monthly Journal of Computer Science and Information Technology

ISSN 2320-088X

IJCSMC, Vol. 4, Issue. 3, March 2015, pg.301 – 307

SURVEY ARTICLE

CHALLENGES AND TECHNIQUES FOR SENTIMENT ANALYSIS: A SURVEY

K.S.Ilakiya¹, Mrs. M.Lovelin Ponn Felciah²

¹Department of Computer Science, Bishop Heber College, Tiruchirapalli, TN, India

²Department of Computer Application, Bishop Heber College, Tiruchirapalli, TN, India

¹ ilakiya02@gmail.com; ² lovelinsathya@gmail.com

Abstract— The current research is play a dominant role in opinion mining is also called as sentiment analysis due to clear volume of opinion rich web resource such as discussion form, review sites, blogs, tweets are available in digital form. Sentiment analysis is the field of study that analyzes customer opinion, feedback, sentiment evaluation, attitudes and emotion from written language. Within fraction second, we classify the text in different manner in different seconds. It is one of the active research areas in Natural Language Processing [NLP]. There are number of techniques we need to classify the Opinion reviews. The main problem in the sentiment analysis is to understand the usage of negation and the classification of positive and negative sentiments recorded by the users in the social group. This paper presents a survey regarding the presently available technique, application and problem that appear in the field of opinion mining.

Keywords— Sentiment analysis, Aspect extraction, Navies bayes, Support vector machine, Positive and negative

I. INTRODUCTION

Sentiment analysis is the type of natural language processing for tracking the mood of the public about a particular product or topic. Sentiment analysis which is also called as opinion mining. It involves in building a system to collect and examine opinions about the product made in blogs post, comments, reviews or tweets in publicly. Sentiment analysis is the subfields of machine learning. In Fig-1 show the work flow diagram of sentiment analysis. Opinions are very important in the current scenario because, lots of user opinionated texts are available in the web resources. There is a hard problem to be solved because natural language is highly unstructured in nature. There are many challenges we have to analysis in Sentiment analysis. The first is an opinion word considered positive in one situation and negative in another. The second challenge is that people express opinions in various ways. Conventional text processing is based on the fact that limited differences can be identified between two text pieces which does not change meaning much. People's opinions and experience are very valuable information in decision making process, but to get benefits from these opinion and experience, the accumulated content should be extracted and analyzed properly. This extracted and analyzed sentiments or

opinion are useful for consumer as well as manufactures as consumers can get a chance to evaluate others opinion and experience related to some product or services before purchasing them. Similarly the manufactures can take these opinions as feedback from the consumers and thereby they can improve and collect their opinions effectively.

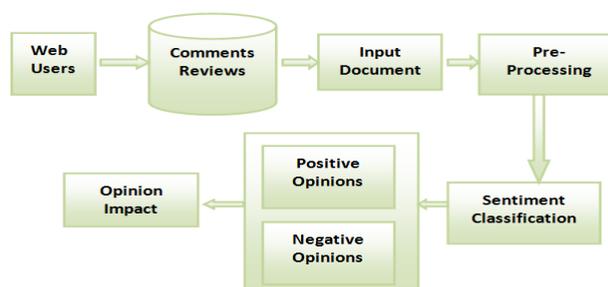


Fig-1 Work flow of sentiment Analysis

II. LITERATURE REVIEW

A. Research Challenges For Sentiment Analysis

Sentiment analysis classifies text as positive, negative or neutral, so it can be called as text classification task such as classic text classification and opinion classification. In text classification it has many classes and many topics but in sentiment analysis it has only three classes. There are many problems that make sentiment analysis difficult when compared to text classification. The main challenges that are faced by sentiment analysis are the following:

1) *Domain dependency*

The main challenge faced by sentiment analysis is the domain dependent nature of sentiment words such as, One features set may give very good performance in one domain, at the same time it perform very poor in some other domain.

2) *Detection of spam and fake reviews:*

The web contains both authentic and spam contents. For effective Sentiment classification, this spam content should be eliminated before processing. This can be done by identifying duplicates, by detecting outliers and by considering reputation of review.

3) *Classification filtering:*

For better sentiment classification result this classification should be reduced. The risk of filter bubble gives inappropriate opinion sets and it results false summarization of sentiment.

4) *Asymmetry in availability of opinion mining software:*

The opinion mining software is very expensive and currently affordable only to large enterprises and government. It is beyond the common citizen's expectation. This should be available to all small and medium size businesses at affordable price, so that everyone gets benefit from it.

5) *Incorporation of opinion with implicit and behavior data:*

For successful sentiment analysis, the opinion words should concatenate with implicit data. The implicit data look like the actual behavior of sentiment words.

6) *Natural Language Processing Language:*

The natural language overhead like ambiguity, co-reference, Implicitness, inference etc. created hindrance in sentiment analysis too.

B. The Process of Sentiment Analysis

Given a set of opinion document such as, the sentiment analysis consists of following 6 main task.

Task1- Entity extraction and categorization:

Extract all entity expression and we can assume whole entity expression into entity categorize. Each entity expression category indicates a unique entity.

Task2-Aspect extraction and categorization:

Extract all aspect expressions of the entities, and categorize these aspect expressions into clusters. Each aspect expression cluster of entity represents unique aspect.

Task 3 -Opinion holder extraction and categorization:

Extract opinion holders for opinions from text or structured data and categorize them.

Task4-Time extraction and standardization:

Extract the times when opinions are given and standardize different time formats.

Task 5 -Aspect sentiment classification:

Determine whether an opinion on an aspect is positive, negative or neutral, or assign a numeric sentiment rating to the aspect in the range of 1 to 5.

Task 6 - Opinion quintuple generation:

Produce all opinion quintuples (ei, aij, sijkl, hk, tl) expressed in document.

C. Preprocessing

The pre-processing phase has been further divided into number of sub phases such as:

Tokenization:

It is a text document has a collection of sentences which is split up into terms or tokens by removing white spaces, commas and other symbols etc.

Stop word Removal:

It removes articles (like, a, an, the).

Stemming:

The goal of stemming is to diminish the inflectional forms and sometimes derivationally related forms of a word to a common root form.[12]

Case Normalization:

It is a process that has English texts to be published in both higher and lowercase characters and turns the entire document or sentences into lowercase/uppercase.

D .Document Sentiment Classification

This approach we consider the whole document as a single entity and classify the sentiment as positive or negative or neutral. Document-level sentiment classification assumes that the opinionated document expresses opinions on a single target and the opinions belong to a single person or product or service. It is clear that this assumption is true for customer reviews of products documents which usually focus on one product and single reviewer writes it.

There are two type of classification techniques that have been used in document-level sentiment classification such as supervised method and unsupervised method.

1) Supervised Method

Sentiment classification task can be spontaneous as a supervised learning problem with two classes, positive and negative. Product reviews are used as training and testing data. Another method such as online reviews have rating scores assigned by their reviewers, e.g., 1-5 stars, the positive and negative classes are determined using the ratings. A review with 4 or 5 stars is considered a positive review, and a review with 1 to 2 stars is considered a negative review.

A training set is used by an automatic classifier to learn the differentiating characteristics of documents, and a test set is used to check how well the classifier performs. The most commonly used features in sentiment classification are introduced below

1.1. Terms and their frequency:

The features can be considered as individual words termed as unigram and their n-grams with associated frequency counts. It can be calculated as weightage form.

1.2. Part of speech (POS):

POS information is a very important indicator of sentiment expression.

1.3. Sentiment words and phrases:

Sentiment words and phrases that is express in positive and negative sentiments. For example good, fantastic, amazing words are positive sentiment and bad, boring, slow, worst and poor are negative sentiment. For example rubbish (noun), hate and like (verb) can indicate opinion in some documents.

1.4. Negations:

Negation words are very important to evaluate the polarity of a sentence because they can transform the sentiment orientation in a sentence. For instance, the sentence “I don’t like this phone” has negative orientation.

1.5. Syntactic dependency:

It is represented as a parse tree and it contains word dependency based features.

SUPERVISED LEARNING TECHNIQUES:

Supervised learning techniques can be used to sentiment classification, such as

- 1) Naive Bayes Classification (NB)
- 2) Maximum Entropy Rule (ME)
- 3) Support Vector Machine (SVM)

In Fig-2 we can analyses the supervised learning techniques. In research works which is used as supervised method to solve sentiment classification problem. In these papers, we used three machine learning techniques - Naïve Bayes, Maximum Entropy classification (ME), and SVM to classify sentiment of product review documents. They test several features to find Unigrams, Bigrams, Parts of Speech (POS) and position of words were used as features in these techniques. The results show that the best performance is achieved is used in SVM classifier.

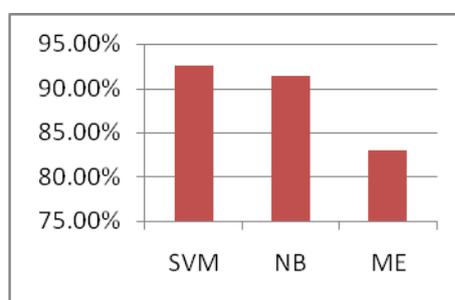


FIG -2 SUMMARY OF OPINION MINING TECHNIQUES

2) Unsupervised Method:

The Unsupervised learning method has a set of inputs, like clustering, labels are not known during training. Classification is performed using some fixed viti patterns which are used to express opinions. The part-of-speech (POS) tags are used to compose viti patterns. The unsupervised learning can be done by extracting the opinion words inside a document .The point-wise mutual information can be made use of to find the semantics of the extracted words

E. Sentence-Level Sentiment Classification:

In the sentence level sentiment analysis, the polarity of each sentence is calculated. The same document level classification methods can be applied to the sentence level classification problem. Objective and subjective sentences must be found out. The subjective sentences contain opinion words which help in determining the sentiment about the entity. After which the polarity classification is done into positive and negative classes.

Knowing that a sentence is positive or negative is of lesser use than knowing the polarity of a particular feature of a product. The advantage of sentence level analysis lies in the subjectivity/ objectivity classification.

F. Sentiment Lexicon Construction:

Sentiment lexicon words are identified by opinion words are also known as opinion-bearing words. Sentiment words are always divided into two categories such as: positive or negative. Positive opinion words are used to express some desired states while negative opinion words are used to express some undesired states.

There are three methods to construct a sentiment lexicon: manually construction, corpus-based methods and dictionary-based methods.

1) Manually Construction:

The manual construction of sentiment lexicon is a very hard and time-consuming task and always cannot be used alone but it can be combined with other methods to improve the accuracy of these methods.

Table -3 Summary of Algorithm

ALGORITHM	BENEFITS	LIMITATION
Navies Bayes Algorithm	It is very easy to construct, not needing any complicated iterative parameter estimation schemes. It may be readily applied to huge data sets. It Handles real and discrete data and streaming data.	Assume independence of features
Support Vector Machine	They are robust in high dimensional spaces and any feature is relevant. It has a sound theoretical foundation, requires only a dozen examples for training, and is insensitive to the number of dimensions.	Its high algorithmic complexity and extensive memory requirements. It limitation is speed and size both training and testing the optimal design for multiclass SVM classifier in further area for research.
Maximum Entropy	It has proven effective in a number of natural language processing applications. It also classifies text. Conditional random fields is that training can be considerably more efficient.	It has no assumption about the relationship between features and it performs better when conditional independence assumption are not.

2) *Corpus –Based Method:*

Corpus-based methods can produce opinion words with relatively high accuracy. Most of these corpus based methods need very large labeled training data. This approach has a major advantage that the dictionary-based approach does not have. It can help find domain specific opinion words and their orientations.

3) *Dictionary -Based Method:*

In this methods is to collect an initial seed set of sentimental words and their orientation manually, and then searching in a dictionary to find their synonyms and antonyms to expand this set. The new seed set are used iteratively to generate new sentiment words.

G. *Aspect- Based Sentiment Analysis:*

The major steps involved to produce this detailed summery of product reviews are

- (1) Aspect extraction
- (2) Aspect sentiment orientation detection

1) *Aspect extraction:*

Aspect extraction indicate as feature extraction is one of the key tasks in aspect-based sentiment analysis.

This method proposed is based on information extraction approach that identifies frequently occurring noun phrases. This approach is generally useful in finding aspects which are strongly associated with a single noun. The disadvantage of this approach is that it cannot detect the aspect terms which are of low frequency and noun phrases.

2) *Aspect sentiment orientation detection:*

It resolves the sentiment orientation expressed on each aspect in a sentence is the second task in aspect based sentiment analysis. It must determine whether the sentiment orientation on each aspect is positive, negative or neutral. This task can be divided into the following sub tasks:

1. Extracting opinion words or phrases.
2. Identifying the polarity of each opinion words or phrases
3. Handling opinion shifters (such as no, not, don't) and opinion intensifiers (such as very, extremely)
4. Handling but clauses.
5. Aggregating opinions (if there is more than one opinion word or phrase in a sentence).

III. CONCLUSION

Opinion is very important for everybody and who is going to take decision for product review. In this paper we have introduced sentiment classification problem in different level i.e. document-level, sentence-level and aspect-level. Also, some techniques that have been used to solve these problems have been introduced. We have discussed about proposed method, Data set, Accuracy and Approaches used in recently developed stemming algorithms for Indian and Non-Indian languages. These techniques are not fully produce 100% output, but are good enough to use for classification method..In future, more research is needed to improve the performance metrics of methods and techniques introduced in this paper.

ACKNOWLEDGEMENT

I would like to express my special thanks to God and my guide **Mrs.M.Lovelin Ponn Felciah** she give full support & encourage for doing my research work successfully and my dear friends who are all supporting with us for doing my project.

REFERENCES

- [1] Richa Sharma¹, Shweta Nigam² and RekhaJain,”Supervised Opinion Mining Techniques: A Survey”. International Journal of Information and Computation Technology. ISSN 0974-2239 Volume 3, Number 8 (2013), pp. 737-742.
- [2] Sindhu C1, Dr. S. ChandraKala, “A Survey on opinion mining and sentiment polarity classification”. International Journal of Emerging Technology and Advanced Engineering. , Volume 3, Special Issue 1, January 2013).
- [3] Mahmoud Othman¹, Hesham Hassan², Ramadan Moawad¹ and Abeer El-Korany²,”Opinion Mining and Sentimental Analysis Approaches: A Survey”.
- [4] BlessySelvam¹ ,S.Abirami,”A Survey On Opinion Mining framework”. International Journal of Advanced Research in Computer and Communication Engineering Vol. 2, Issue 9, September 2013.
- [5] MR.S.M.Vohra,² PROF.J.B.Teraiya,”A Comparative Study OF Sentiment Analysis Techniques”. Journal Of Information , Knowledge And Research In Computer Engineering .
- [6] S. Veeramani¹, S. Karuppusamy,”A Survey on Sentiment Analysis Technique in Web Opinion Mining”. International Journal of Science and Research (IJSR).
- [7] HaseenaRahmath P, “Opinion Mining and Sentiment Analysis - Challenges and Applications“.International Journal of Application or Innovation in Engineering & Management (IJAIEM) Volume 3, Issue 5, May 2014.
- [8] RaisaVarghese¹, Jayasree M², “A Survey on opinion mining and sentiment analysis”. International Journal of Research in Engineering and Technology. eISSN: 2319-1163 | pISSN: 2321-7308
- [9] Bo Pang¹ and Lillian Lee², “Opinion mining and sentiment analysis”.
- [10] G.Vinodhini RM.Chandrasekaran”Sentiment Analysis and Opinion Mining: Survey”. International Journal of Advanced Research in Computer Science and Software Engineering Volume 2, Issue 6, June 2012.
- [11] Bing Liu, Lei Zhang, “A Survey On Opinion Mining And Sentiment Analysis”.
- [12] Kasthuri,” Comprehensive Analyze of Stemming Algorithms for Indian and Non-Indian Languages’, International Journal of Computer Engineering Applications (IJCEA), Volume 7, Issue 3 September 2014.
- [13] Mahmoud Othman¹, Hesham Hassan², Ramadan Moawad¹ and Abeer El-Korany²,” Opinion Mining and Sentimental Analysis Approaches: A Survey”, Life Science Journal 2014;11(4).<http://www.lifesciencesite.com>
- [14] T. Ramani¹, M. Ramzan Begam²,” Survey: A Techniques implemented on Opinion Mining”, International Journal of Computer Science & Engineering Technology (IJCSSET).
- [15] Ms. Vaishali Mehta Prof. Ritesh K Shah,” Approaches of Opinion Mining and Performance Analysis: A Survey”, International Journal of Advanced Research in Computer Science and Software Engineering. Volume 4, Issue 10, October 2014.