ISSN: 2350-0328

International Journal of Advanced Research in Science, Engineering and Technology

Vol. 2, Issue 10 , October 2015

A Survey: Sentiment Analysis and Sentiment Classification Based On Machine Learning Methods

Senthamarai.R , Mary shamala.L

P.G. Student, Department of Computer Science and Engineering, IFET College of Engineering, Villupuram, Tamilnadu, India

Associate Professor, Department of Information Technology , IFET College of Engineering, Villupuram, Tamilnadu, India

ABSTRACT: Sentiment Analysis is an emerging field for research which deals with information extraction and knowledge discovery from text using Natural Language Processing (NLP) and Data Mining (DM) techniques. When you need to make some decision, you address some opinion from others. Such opinions or Sentiments of each and every individual are useful for people and organizations to make several decisions. Opinion Mining or Sentiment Analysis is an information extraction task that identifies the user's views or opinions explained in the form of positive, negative or neutral comments about the text. Supervised learning and unsupervised learning is the type of machine learning techniques used for sentiment classifications. Some methodologies are used for sentiment classification like Naive Bayes, Maximum Entropy Decision tree learning and support vector machine. This paper is a survey on sentiment analysis, sentiment analysis levels, tools and machine learning methods that is used for mining sentiment analysis.

KEYWORDS: Sentiment analysis, Sentiment classifications, Sentiment analysis types, Machine learning methods.

I. INTRODUCTION

In the last few years the growth and use of internet increases and sharing of user's opinions also increases. social media platforms like Twitter, Facebook, MySpace and other forums provides a platform for people to express their emotions in the digital world, which provide valuable information[1]. By using this information we can make several decisions which are suitable to our work. Opinion can be collected from any individual in the world about anything through review sites, blogs, web forums. Organizations and product owners who makes this feedback of users or customers to improve their products/services [8]. Generally used sources for finding opinion are Blogs, review sites, raw dataset, and Micro-blogging web sites. Online messages that are posted by individual in the internet are mostly informal. Analysis and handling of this kind of text is often more difficult when compared with formal texts. The main difference between formal and informal text is in data preprocessing for formal text require less preprocessing whereas informal text contains emotions, sarcasm, utilization of weak grammar, and non lexicon- standard words. Therefore, extraction of informal content is more troublesome [8].

People frequently ask their friends, relatives, and field specialists for suggestion during the decision-making procedure, and their opinions and perspectives are based on experiences and observations. One's point of view a subject can either be positive or negative, which is known as the polarity detection of the sentiment. During sentiment analysis process, it requires very fast and concise information so any individual can make quick and accurate decisions [8]. Nowadays, large attention has been given to opinion mining because of its wide range of possible applications. This mining helps to collect information about the positive and negative aspects of a particular topic [1]. Some sentiment analysis algorithms are designed and already available. We can make a modification on that algorithms which suits for our needs. Here various machine learning algorithm is used for sentiment classification such as Naïve Byes, Maximum Entropy, Decision tree learning and support vector machine. In this paper, we examine the basic concept of sentiment analysis and their machine learning techniques to the sentiment classifications.

ISSN: 2350-0328
International Journal of Advanced Research in Science, Engineering and Technology
Vol. 2, Issue 10 , October 2015

II.ANALYSIS LEVELS

In sentiment analysis, the information or data collected from the reviews has been investigated mainly at three sentiment analysis level.

A. Document sentiment level
B. Sentence Sentiment level.
C. Entity and Aspect sentiment level

A. Document Sentiment Level

The task at this level is to identify whether an entire sentiment document expresses a positive or negative sentiment.
.

B. Sentence Sentiment Level

The tasks at this level goes to the sentences and find out if each sentence expressed a positive, negative, or neutral sentiment.

C. Entity and Aspect Sentiment Level

Above described both the document sentiment level and the sentence sentiment level do not analyze what exactly people like and did not like. Aspect level helps to derive polarity (positive or negative) and a target of sentiment. A sentiment without its target being recognized is of restricted use. Finding out the target of sentiment helps to understand the sentiment analysis problem better.

III. SENTIMENT CLASSIFICATIONS

Sentiment Classification consists basically two approaches machine learning and lexicon based approaches.

A. Machine based Learning Approaches [7]

It is all about learning structures and involves text classification techniques. It is a supervised learning algorithm which is capable of acquiring and integrating the knowledge automatically.

Naive Bayes

The basic idea is to estimate the probabilities of categories to a given test document by using the joint probabilities of words and categories. The algorithm is widely used for document classification.

Maximum Entropy

It is a probabilistic classifier which belongs to the class of exponential models. The Maximum Entropy is based on the Principle of Maximum Entropy and from all the models that fit our training data, selects the one which has a largest entropy.

Support vector machine

It is considered to be the best text classification method. The main idea of SVM for sentiment classification is to discover a hyper plane which separates the documents as per the sentiment, and the margin between the classes. Then it also focused around the Structural Risk Minimization principle.

B. Lexicon Based Approaches [7]

Words in a sentence express different opinion.

ISSN: 2350-0328
International Journal of Advanced Research in Science, Engineering and Technology
Vol. 2, Issue 10 , October 2015

The corpus-based techniques

Corpus-driven method is to determine the emotional affinity of words which is meant to learn their probabilistic affective scores from large corpus.

The dictionary-based techniques

It uses lexical resources such as Word Net it automatically retrieve similar words from Word Net. This techniques use "synonyms, antonyms and hierarchies" in Word Net to find word sentiments.

Bootstrapping approach

A high precision classifier is first used to automatically identify some subjective and objective sentences. The learned patterns are then used to extract more subject and objective sentences.

IV.TOOLS USED IN OPINION MINING

The tools which are used to track the opinion from the user generated contents are:

Review Seer tool – This tool is used to automate the work done by aggregation sites. The Naive Bayes classifier approach is used to collect positive and negative opinions for assigning a score to the extracted feature terms. The results are simple opinion sentence.

Web Fountain - It uses the beginning definite Base Noun Phrase (BNP) heuristic approach for extracting the product features. It is possible to develop a simple web interface.

Red Opal –It is a tool that enables the users to determine the opinion orientations of products based on their features. It assigns the scores to each product based on features extracted from the customer reviews. The results be web based interface.

Opinion observer-This is an opinion mining system for analyzing and comparing opinions on the Internet using user generated contents. This results in a graph format showing opinion of the product feature by feature.

NTLK-The natural language toolkit is a tool for text processing, classification, tokenization, stemming, tagging, parsing etc.

Open NLP-perform the most common NLP tasks, such as POS tagging, named entity extraction, chunking and co-reference resolution.

Opinion Finder-It aims to identify subjective sentences and to mark various aspects of subjectivity in these sentences, including the opinion holder of the subjectivity and words that are included in phrases expressing positive or negative sentiments.

V.RELATED WORK

"Thumbs up? Sentiment Classification using Machine Learning Techniques" [2].

This paper deals with the problem of classifying documents not by topic, but by overall sentiments like positive, negative or neutral comments. Here they used three machine learning methods such as Naive Bayes, Support Vector Machine and Maximum Entropy. The results produced via machine learning techniques are quite good in comparison to the human generated baselines. In terms of relative performance, Naive Bayes tends to worst and SVMs tend to the best. On the other hand, we were not able to achieve accuracies on the sentiment classification problem comparable to those reported for standard topic-based categorization. Lastly they conclude by examining factors that make the sentiment classification problem more challenging.

ISSN: 2350-0328

International Journal of Advanced Research in Science, Engineering and Technology

Vol. 2, Issue 10 , October 2015

"Using Appraisal Groups for Sentiment Analysis" [5].

It presents a new method for sentiment classification based on extracting and analyzing appraisal groups such as "very good" or "not terribly funny". An appraisal group is represented as a set of attribute values in several task independent semantic taxonomies, based on Appraisal Theory. The first goal is to extract appraisal groups, from which they derive useful features for machine learning and assign four main types of attributes to appraisal groups: Attitude, Orientation, Graduation, and Polarity. Semi-automated technique to construct a lexicon giving appraisal attributes values for relevant terms. Classifying movie reviews used standard bag-of-words techniques with limited success. In addition of typed features of semantic orientation has been shown to improve results. Semantic orientation has also been useful for classifying more general product reviews be easier to classify than movie reviews.

"Recognizing Contextual Polarity in Phrase-Level Sentiment Analysis" [8].

It presented a new approach to phrase level sentiment analysis that first determines whether an expression is neutral or polar expressions. With this approach, the system is able to automatically identify the contextual polarity for a large subset of sentiment expressions, achieving results that are significantly better than baseline. Some research classifies the sentiments of sentences. All begin by first creating prior-polarity lexicons. Then assign a sentiment to a sentence by averaging the prior semantic orientations of instances of lexicon words in the sentence. But they do not identify the contextual polarity of individual phrases containing clues, as we do in this paper.

In addition, their systems assign one sentiment per sentence; our system assigns contextual polarity to individual expressions. As seen above, sentences often contain more than one sentiment expression. They classify expressions that are about specific items, and use manually developed patterns to classify polarity. These patterns are high-quality, yielding quite high precision.

"Support Vector Machine for Text Categorization" [7].

This paper introduces support vector machines for text categorization. It provides both theoretical and empirical evidence that SVMs are very well suited for text categorization. The theoretical analysis tells that SVMs acknowledge the particular properties of text: (a) high dimensional feature spaces, (b) few irrelevant features (dense concept vector), and (c) sparse instance vectors. Then the experimental results show that SVMs consistently achieve good performance on text categorization tasks.

With their ability to generalize well in high dimensional feature spaces, SVMs eliminate the need for feature selection and make the application of text categorization be easier. Advantage of SVMs over the conventional methods is their robustness and also avoids CA strophic failure. SVMs do not require any parameter tuning, since they can find good parameter settings automatically. All this makes SVMs a very promising and easy-to-use method for learning text classifiers.

VI.CONCLUSION

Sentiment analysis is vast research are with several challenges. It has a wide variety of applications in e-commerce. It helps in classifying, summarizing reviews and in other real time applications. This paper focuses on sentiment classification, classification techniques and what tools are available for sentiment analysis. Then described about Machine learning approaches works well for classifying sentiment analysis. Even though these approaches faces some challenges.

Our idea of future wok focusses on using a unsupervised machine learning methods for sentiment classification and it will provide better results and performance.

REFERENCES

ISSN: 2350-0328

International Journal of Advanced Research in Science, Engineering and Technology

Vol. 2, Issue 10 , October 2015

[1] Gaurav S. Chavan, Sagar Manjare, "A survey of various Machine Learning Techniques for Text Classification", International Journal of Engineering Trends and Technology, Volume 15, 2014.

[2] Bo Pang, Lillian Lee, "Thumbs up Sentiment Classification Using Machine Learning Technique", Proceedings of the ACL-02 conference on Empirical methods in natural language processing, Volume 10, 2002.

[3] Bing Liu, "Sentiment Analysis and Opinion Mining", Morgan & Claypool Publishers, May 2012.

[4] C. Whitelaw, N. Garg, and S. Argamon, "Using Appraisal Groups for Sentiment Analysis", Presented at the Proceedings of the 14th ACM International Conference on Information and Knowledge Management, Bremen, Germany, 2005.

[5] A. Basu, C. Watters, and M. Shepherd. "Support Vector Machines for Text Categorization," Proceedings of the 36th Hawaii International Conference on System Sciences (HICSS'03).

[6] Theresa Wilson, Janyce Wiebe, Paul Hoffmann, "Recognizing Contextual Polarity in Phrase-Level Sentiment Analysis", An International Journal, July 2011.

[7] Angu Lakshmi, Dr. R. Manickachezian, "An Analysis on Opinion Mining Techniques and Tools", International Journal of Advanced Research in Computer and Communication Engineering.

[8] Dharmesh Ramani, Hazari Prasun, "A survey: Sentiment Analysis of Online Review", November 2013