



Survey of Classification Methods and Summarization of Sentiment Analysis and Opinion Analysis

Shobha.D

Assistant Professor, Department of Studies in Computer Science, Pooja Bhagavat Memorial Mahajana Post Graduate
Centre, K.R.S. Road, Metagalli, Mysuru, Karnataka, India

ABSTRACT: The growth of social website and electronic media contributes vast amount of user generated content such as comments and opinions, customer reviews, Social media gives users a platform to communicate effectively with friends, family, and colleagues, and also gives them a platform to talk about their favorite. The digital ecosystem has itself paved way for use of huge volume of opinionated data recorded. Sentiment analysis is basically concerned with analysis of emotions and opinions from text. We can refer sentiment analysis as opinion mining. Opinion mining/sentiment analysis is a multidisciplinary and multifaceted Artificial intelligence problem. Its aim is to minimize the gap between human and computer.

KEYWORDS: Sentiment Analysis, Opinion, Machine Learning, big data, mining.

I. INTRODUCTION

Sentiment is an attitude, thought, or judgment prompted by feeling. Sentiment analysis [1-8], which is also known as opinion mining, studies people's sentiments towards certain entities. Sentiment Analysis includes branches of Computer Science like Natural Language Processing, Machine Learning, Text Mining and Information Theory and Coding. The analyzing and summarizing of others opinions, expressed in huge opinionated text data is a very interesting new field for researchers. This area of research is called Sentiment Analysis or Opinion Mining of others. By using approaches, methods, techniques and models of defined branches, we can categorize our unstructured data which may be in the form of news articles, blogs, tweets, movie reviews, product reviews etc. into positive, negative or neutral sentiment according to the sentiment expressed in them. Sentiment analysis is the aggregation of public opinion towards useful information, also known as opinion mining.

Sentiment Analysis has lead to development of better products and good business management. This research area has provided more importance to the mass opinion instead of word-of-mouth. Usually these kinds of reviews are written by customers who have used the particular product or service. An individual's interests, opinions and perceptions greatly influence the nature of the review Opinion mining/sentiment analysis is a multidisciplinary and multifaceted Artificial intelligence problem. Its aim is to minimize the gap between human and computer.

II. LITERATURE SURVEY

Machine Learning consists of various different classifiers such as Ensemble classifier, k-means, Artificial Neural Network etc. These are used to classify reviews [8]. Y.Mejova et al [1] in his research work proposed that we can use presence of each character, frequency of occurrences of each character, word which is considered as negation etc. as features for creating feature vector. He also shows that we can effectively use unigram and bigram approaches to make feature vector in Sentiment analysis.

Domingos et al suggested that Naive Bayes works well for dependent features for certain problem. Zhen Niu et al found a new model. This model is based on Bayesian algorithm. In this model, some efficient approaches are used for selecting feature, computation of weight and classification. Barbosa et al designed a 2 step analysis method which is an automatic sentiment analysis for classifying tweets. In the first step, tweets are classified into subjective and objective



International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 4, Issue 5, May 2016

tweets. After that, in a second step, subjective tweets are classified as positive and negative tweets. Celikyilmaz et al developed one method as pronunciation based word clustering. This method normalizes noisy tweets.

There are some words which have the same pronunciation but having different meanings. In this mentioned method, words having same pronunciation are clustered and assigned common tokens. Wu et al in his paper recommended model, namely, the influence probability to analysis the sentiment tweets. In this, if @username is found in the tweet, it takes influencing action and helps to influencing probability. By collecting automatic tweets, Pak et al developed a method for sentiment analysis by creating twitter corpus. He used a Naïve Bayesian classifier to do the sentiment analysis. Some researches are made to identify the public opinion about movies, news etc. from twitter tweets. V.M. Kiran et al had taken the information from other publicly available databases like IMDB and Blippr. There are two techniques widely used to detect the sentiments from text. They are Symbolic techniques and Machine Learning techniques [3].

III. SENTIMENT TECHNIQUES

In general, various symbolic techniques and machine learning techniques are used to analyze the sentiment from the twitter data. So in another way we can say that a sentiment analysis is a system or model that takes the documents that analyzed the input, and generates a detailed document summarizing the opinions of the given input document. In the first step pre-processing is done. In the pre-processing we are removing the stop words, white spaces, repeating words, emoticons and #hash tags.

There are two techniques widely used to detect the sentiments from text. They are

- Symbolic techniques and
- Machine Learning techniques [3].

A. Sentiment analysis using Symbolic Techniques

A symbolic technique uses the availability of lexical resources. Turney [4] suggested an approach for sentiment analysis called 'bag of words'. In the mentioned approach, individual words are neglected and only collections of words are considered. He gathered word having adjectives or adverb for the polarity of review from a search engine AltaVista. A lexical database called Wordnet [6] was used by Kamps et al [5] which determines an emotional matter in a word. WordNet carries synonyms and distance metric to find the orientation of adjectives.

To overcome obstacles in lexical substitution task, Baroni et al [7] developed a system supported by word space model formalism thereby representing local words. EmotiNet conceptually represented the text that stored the structure of real events in a domain. This was introduced by Balahur et al [8].

B. Sentiment analysis using Machine Learning Techniques

Under this technique, there are two sets, namely a training set and a test set. Generally the dataset which is collected from different sources and whose behavior and output values are known to us falls into the category of training data sets. In contrast with this, the datasets whose values or behavior are unknown to us are called as test data sets. Here different classifiers are trained with training data and then unknown data or we can say a test data is given to this model to get desired results.

IV. SENTIMENT CLASSIFICATION

The Sentiment classification is presented in various formats in different domains. Positive/negative, good/bad, like/dislike, buy/don't buy, recommended/not recommended, excellent/boring (film), support/against [5], optimistic/pessimistic [9], favorable/unfavorable [17]. Sentiment classification may be done at different levels. In Document level – whole document classify either into positive or negative class. Sentence level – classifies sentence into positive, negative or neutral class. Aspect or feature level – identifying & extracting product features from the source data.

A. Sentiment Classification Techniques:

In general, various symbolic techniques and machine learning techniques are used to analyze the sentiment from the twitter data. So in another way we can say that a sentiment analysis is a system or model that takes the documents that analyzed the input, and generates a detailed document summarizing the opinions of the given input document. In the



International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 4, Issue 5, May 2016

first step pre-processing is done. In the pre-processing removing the stop words, white spaces, repeating words, emoticons are removed.

The several methods are used to extract the feature from the source text. Feature extraction is done in two phases: In the first phase extraction of data related to twitter is done i.e. twitters specific data is extracted. Now by doing this, the tweet is transformed into normal text. In the next phase, more features are extracted and added to feature vector. Each tweet in the training data is associated with class label. This training data is passed to different classifiers and classifiers are trained. Then test tweets are given to the model and classification is done with the help of these trained classifiers. So finally we get the tweets which are classified into the positive, negative and neutral.

Sentiment classification or Polarity classification is also done by the binary classification task of labeling an opinionated document as expressing either an overall positive or an overall negative opinion. A technique for analyzing subjective information in a large number of texts, and many studies is sentiment classification. A typical approach for sentiment classification is to use machine learning algorithms.

B. Data Source:

People and companies across disciplines exploit the rich and unique source of data for varied purposes. detecting subjective and objective sentences; classifying sentences as positive, negative, or neutral; detecting the person expressing the sentiment and the target of the sentiment; detecting emotions such as joy, fear, and anger; visualizing sentiment in text; and applying sentiment analysis in health, commerce, and disaster management. Surveys by Pang and Lee (2008) and Liu and Zhang (2012) give a summary of many of these approaches. Hu and Liu (2004) manually labeled about 6,800 words and used them for detecting sentiment of customer reviews. Since the opinions contributed by people and companies are to be evaluated, the data consider from blogs, review sites, web discourse and news articles.

Since manual annotation of data is costly, distant supervision techniques have been actively applied in the domain of short informal texts. User-provided indications of emotional content, such as emoticons, emoji, and hashtags, have been used as noisy sentiment labels.

Blogs With an increasing usage of the internet, blogging and blog pages are growing rapidly. Blog pages have become the most popular means to express one's personal opinions. Bloggers record the daily events in their lives and express their opinions, feelings, and emotions in a blog (Chau & Xu, 2007). Many of these blogs contain reviews on many products, issues, etc. Blogs are used as a source of opinion in many of the studies related to sentiment analysis (Martin, 2005; Murphy, 2006; Tang et al., 2009). Most of the work in the field uses movie reviews data for classification. Movie review datas are available as data set .

V. SENTIMENT ANALYSIS: A GENERAL VIEW

Sentiment analysis was initially formulated as the NLP task of retrieval of sentiments expressed in texts. A simple keyword finding will not be appropriate for mining all kinds of opinions. Hence the need for use of sophisticated opinion extraction methods. Sentiment analysis is a natural language processing technique, helps to find and dig out subjective information in source materials. Sentiment analysis aims to establish the approach of the writer with respect to the entire contextual polarity of a document.

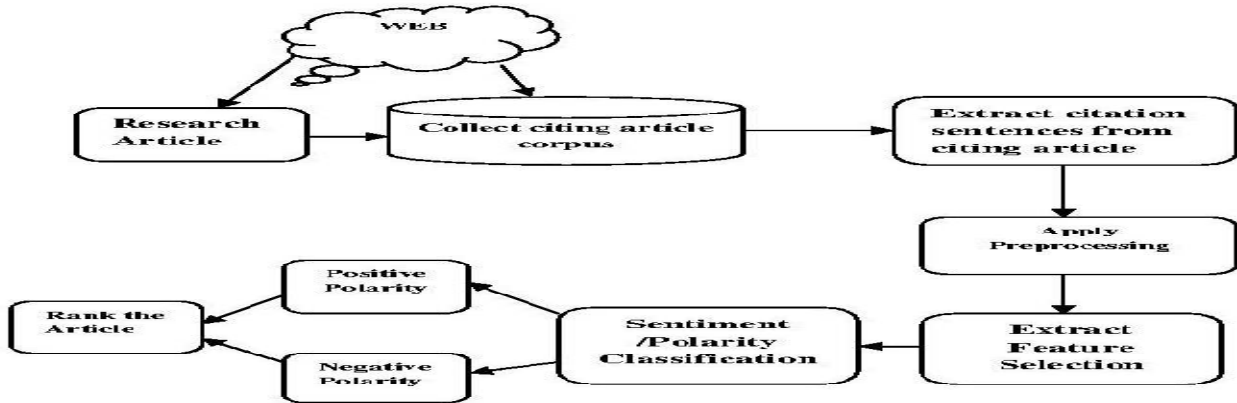
The approach may be his or her decision or valuation, sentimental state or the intended emotional communication. A fundamental process in sentiment analysis is classification of opinion polarity of a specified context at the document; whether the given opinion in an article, a context or an entity feature is positive, negative or neutral. Beyond polarity sentiment classification, the emotional states such as "irritated", "gloomy" and "joyful" are also recognized.

Figure shows the architecture of the framework which extracts the citation from the citing articles. It consists of the following modules: Citation extraction, preprocessing the extracted citations, feature extraction, apply machine learning technique for sentiment classification and rank the result.

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

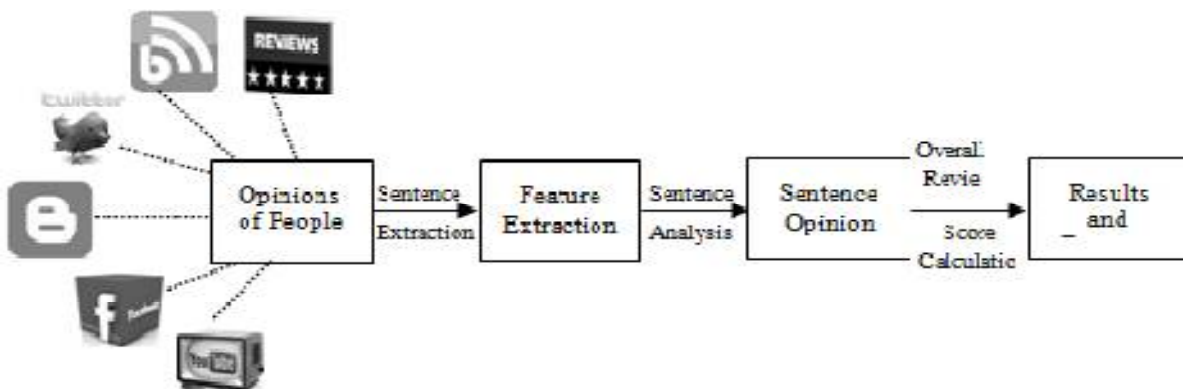
Vol. 4, Issue 5, May 2016



Architecture of sentiment mining for journal citation.

In review there are many user generated reviews available on the internet that aids a customer in buying a product. E-commerce sites such as www.amazon.in, www.flipkart.com and www.reviewcentre.com has millions of customer reviews for products, where as www.rediff.com/movies/reviews, www.indiaglitz.com and www.rottentomatoes.com has reviews for movies and www.yelp.com, www.burrrp.com has restaurant reviews. A blog is a personal website or web page on which an individual records opinions, links to other sites, etc. on a regular basis. They are the fastest growing sections of the emerging communication systems. The simple and no-nonsense style of writing a post and uploading it on the web has made the blogging world an indispensable source of data in the case of sentiment mining [11]. The micro blogging site Twitter is also flooded with opinions that are decisive in determining the election results even [3]. These opinions can also be used for classifying sentiments. People record the daily events in their lives and express their opinions, feelings, and emotions in an on-line journal, or blog or on Twitter [14].

A very basic step of opinion mining and sentiment analysis is feature extraction is as shown in Figure .



process of opinion minning and sentiment analysis

Sentiment Analysis is done on three levels [1].

1. Document level
2. Sentence level
3. Aspect or Entity level

Document Level Sentiment Analysis is performed for the whole document and then decide whether the document express positive or negative sentiment [1].

Entity or Aspect Level Sentiment Analysis performs fine-grained analysis. The goal of entity or aspect level Sentiment Analysis is to find sentiment on entities and/or aspect of those entities. For example consider a statement “My HTC



International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 4, Issue 5, May 2016

Wildfire S phone has good picture quality but it has low phone memory storage.” so sentiment on HTC's camera and display quality is positive but the sentiment on its phone memory storage is negative.

Sentence level Sentiment Analysis is related to find sentiment from sentences whether each sentence expressed a positive, negative or neutral sentiment. Sentence level Sentiment Analysis is closely related to subjectivity classification. Many of the statements about entities are factual in nature and yet they still carry sentiment. Current Sentiment Analysis approaches express the sentiment of subjective statements and neglect such objective statements that carry sentiment [1].

In some sentiment analysis system, they utilize three freely available, manually created, general-purpose sentiment lexicons. Experiments are carried out to assess both, the performance of the overall sentiment analysis system as well as the quality and value of the automatically created tweet-specific lexicons.

Till today different sentiment analysis techniques are being implemented with different aspects of evaluation for big data opinion analysis.

Sr. No.	Sentiment Analysis Technique	Takes into account
1.	Document level sentiment analysis	Classifying the whole document as positive or negative
2	Supervised learning techniques Unsupervised learning techniques.	‘terms and their frequency’, ‘parts of speech’, ‘sentiment words and phrases’, ‘sentiment shifters’ Use of fixed syntactic patterns that occur in an opinion.
3.	Sentence level Sentiment Analysis	Associated with a phrase or sentence
	Aspect Based Sentiment Analysis	sentiment on entities and/or aspect of those entities
	Oracle Advanced Analytics	database into a comprehensive advanced analytics

The above table no. 1 shows comparative analysis of different sentiment analysis techniques and what it takes into account for implementation base. Polarity, subjective detection and opinion identification all are very important things in this kind of sentiment analysis. Document level sentiment analysis classifies the whole document as positive and negative statement documents. Supervised learning verifies terms and frequency, ‘parts of speech’, ‘sentiment words and phrases’, ‘sentiment shifters’. Unsupervised learning technique uses fixed syntactic patterns.

Sentiment Analysis Tasks Most of the researchers focus on specific tasks: sentiment analysis of words, subjective expressions, subjective sentences and topic. These approaches find sentiment at opinion/facts level and used to enhance the effectiveness of a sentiment classification Pang & Lee (2004). Choi et al. (2005) focus on finding the sources of opinions (e.g. Finding who play a crucial role the person or organizations in influencing other individuals’ opinion) instead of carrying out a sentiment classification. There was various types of Data sources where used that include The Multi-Perspective Question Answering (MPQA) corpus, user feedback, The Wall Street Journal (WSJ) corpus and the Document Understanding Conference (DUC) corpus.

VI. SEMANTIC ORIENTATION

The Semantic orientation approach to Sentiment analysis is unsupervised learning because it does not require prior training in order to mine the data. Instead, it measures how far a word is inclined towards positive and negative. Much of the research in unsupervised sentiment classification makes use of lexical resources available.

A few recent studies in this field explained the use of neural networks in sentiment classification. Zhu Jian (2010) proposed an individual model based on Artificial neural networks to divide the movie review corpus into positive, negative and fuzzy tone which is based on the advanced recursive least squares back propagation training algorithm.



International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 4, Issue 5, May 2016

Long-Sheng Chen (2011) proposed a neural network based approach, which combines the advantages of the machine learning techniques and the information retrieval techniques.

VII. APPLICATIONS AND TOOLS

Some of the applications of sentiment analysis includes online advertising, hotspot detection in forums etc. Online advertising has become one of the major revenue sources of today. Internet ecosystem. Sentiment analysis find its recent application in Dissatisfaction oriented online advertising Guang Qiu(2010) and Blogger-Centric Contextual Advertising (Teng-Kai Fan, Chia-Hui Chang ,2011), which refers to the assignment of personal ads to any blog page. This has motivated the research on identification of online forum hotspots, where useful information is quickly exposed to those seekers. Nan Li (2010) used sentiment analysis approach to provide a comprehensive and timely description of the interacting structural natural groupings of various forums, which will dynamically enable efficient detection of hotspot forums.

In order to identify potential risks, it is important for companies to collect and analyze information about their competitors' products and plans. Sentiment analysis find a major role in competitive intelligence to extract and visualize comparative relations between products from customer reviews, with the interdependencies among relations taken into consideration, to help enterprises discover potential risks and further design new products and marketing strategies.

VIII. OPINION SUMMARIZATION

Summarization of opinion is a main part in opinion mining process. Summary of reviews provided should be established on lineaments or subtopics that are mentioned in reviews. opinions of articles by telling sentiment polarities, degree and the correlated events. With opinion summarization, a customer can easily see how the existing customers feel about a product, and the product manufacturer can get the reason why different stands people like it or what they complain about. Ku, Liang, and Chen (2006) investigated both news and web blog articles. Many works have been executed on condensation of product reviews. The opinion condensation process mainly includes the following two methods. Feature based condensation a type condensation involves returns of frequent terms (features) that are appearing in many reviews. Features present in review text can be identified using Latent Semantic Analysis (LSA) method. Opinion Mining which says how the input is being classified on various steps to summarize the reviews. The issue of relevant sentence selection is discussed, and then topical and opinionated information are summarized. Opinion summarizations are visualized by representative sentences.

IX. CONCLUSION

Knowledge based models are recently introduced to the NLP tasks and is a hot research area. New sub-domains of sentiment analysis and opinion mining are identified that are of interest to the research community including bias analysis, threat /. Sentiment detection has a wide variety of applications in information systems, including classifying reviews, summarizing review and other real time applications. It is found that sentiment classifiers are severely dependent on domains or topics. We can successfully analyze the different schemes for feature selection and their effect on sentiment analysis. The challenge in sentiment analysis of product reviews is to produce a summary of opinions based on product features/attributes.

REFERENCES

- [1] Ion SMEUREANU, Cristian BUCUR, Applying Supervised Opinion Mining Techniques on Online User Reviews, Informatica Economică vol. 16, no. 2/2012.
- [2] R. Prabowo and M. Thelwall, "Sentiment Analysis: A Combind Approach", Journal of Informetric - Elsevier,
- [3] E. Boiy, P. Hens, K. Deschacht and M. Moens, 'Automatic sentiment analysis in on-line text', 11th International Conference on Electronic Publishing, vol. 349360, 2007.
- [4] P. Turney, 'Thumbs Up or Thumbs Down? Semantic orientation applied to unsupervised classification of reviews', 40th annual meeting on association for computational linguistics, vol. 417424, 2002.
- [5] J. Kamps, M. Marx, R. Mokken and M. De Rijke, 'Using wordnet to measure semantic orientations of adjectives', 2004.
- [6] C. Fellbaum, 'Wordnet: An electronic lexical database (language, speech, and communication)', 1998.



ISSN(Online): 2320-9801
ISSN (Print) : 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 4, Issue 5, May 2016

- [7] D. Pucci, M. Baroni, F. Cutugno and A. Lenci, 'Unsupervised lexical substitution with a word space model', Proceedings of EVALITA workshop, 11th Congress of Italian Association for Artificial Intelligence, Citeseer, 2009.
- [8] A. Balahur, J. Hermida and A. Montoyo, 'Building and Exploiting Emotinet, a knowledge base for emotion detection based on the appraisal theory model', Affective Computing, IEEE Transactions, vol. 3, 188101, 2012.
- [9] Kim S-M, Hovy E (2004) Determining the sentiment of opinions. In: Proceedings of the 20th international conference on Computational Linguistics, page 1367. Association for Computational Linguistics, Stroudsburg, PA, USA.
- [10] Bing Liu, Sentiment Analysis and Opinion Mining, Morgan and Claypool Publishers, May 2012, p.18-19,27-28,44-45,47,90-101.
- [11] Liu, H., Lieberman, H., & Selker, T. (2003). A model of textual affect sensing using realworld knowledge. In Proceedings of the 8th International Conference on Intelligent User Interfaces, IUI '03, pp. 125{132, New York, NY. ACM.
- [12] Pang, B., Lee, L., & Vaithyanathan, S. (2002). Thumbs up?: Sentiment classification using machine learning techniques. In Proceedings of the Conference on Empirical Methods in Natural Language Processing, EMNLP '02, pp. 79{86, Philadelphia, PA.
- [13] Ziqiong Zhang, Qiang Ye, Zili Zhang, Yijun Li, "Sentiment classification of Internet restaurant reviews written in Cantonese", Expert Systems with Applications xxx (2011)