



Review Paper on Sentiment Analysis Using Web 2.0 by Classification Method

Miss. Siddhi Patni*, Prof. Avinash Wadhe

M.E (CSE) 2nd Semester*, M-Tech (CSE)

GHRCE&M, Amravati

patni.siddhi@gmail.com*, aviwadhe@gmail.com

Abstract: Our day-to-day life has always been influenced by what people think. Ideas and opinions of others have always affected our own opinions. A vital part of the information era has been to find out the opinions of other people. In the pre-web era, it was customary for an individual to ask his or her friends and relatives for opinions before making a decision. Organizations conducted opinion polls, surveys to understand the sentiment and opinion of the general public towards its products or services. In the past few years, web documents are receiving great attention as a new medium that describes individual experiences and opinions. The explosion of Web 2.0 has led to increased activity in Podcasting, Blogging, and Tagging, Contributing to RSS, Social Bookmarking, and Social Networking. As a result there has been an eruption of interest in people to mine these vast resources of data for opinions. The enormous amount of heterogeneous data that is generated by the users of these communities, offers an unprecedented opportunity to create and employ theories & technologies that search and retrieve relevant data from the huge quantity of information available and mine for opinions thereafter. Sentiment Analysis or Opinion Mining is the computational treatment of opinions, sentiments and subjectivity of text.

Keywords- Sentiment Analysis, Tasks, and Levels, Web2.0.

I. INTRODUCTION

Sentiment Analysis is a Natural Language Processing and Information Extraction task that aims to obtain writer's feelings expressed in positive or negative comments, questions and requests, by analyzing a large numbers of documents. Generally speaking, sentiment analysis aims to determine the attitude of a speaker or a writer with respect to some topic or the overall tonality of a document. In recent years, the exponential increase in the Internet usage and exchange of public opinion is the driving force behind Sentiment Analysis today ([1]). The Web is a huge repository of structured and unstructured data. The analysis of this data to extract latent public opinion and sentiment is a challenging task. The goal of sentiment analysis is to detect subjective information contained in various sources and determine the mind-set of an author towards an issue or the overall disposition of a document. Sentiment analysis is done on user generated content on the Web which contains opinions, sentiments or views. An opinionated document can be a product review, a forum post, a blog or a tweet, that evaluates an object. The opinions indicated can be about anything or anybody, for e.g. products, issues, people, organizations or a service.

II. LITERATURE SURVEY

Formally stating Sentiment Analysis is the computational study of opinions, sentiments and emotions expressed in text ([1]). The goal of sentiment analysis is to detect subjective information contained in various sources and determine the mind-set of an author towards an issue or the overall disposition of a document.

Wiebe et al. ([2]) described subjectivity as the linguistic expression of somebody's opinions, sentiments, emotions, evaluations, beliefs and speculations. The words opinion, sentiment, view and belief are used interchangeably but there are subtle differences between them [3].

Opinion: A conclusion thought out yet opens to dispute ("each expert seemed to have a different opinion").

- View: subjective opinion ("very assertive in stating his views").
- Belief: deliberate acceptance and intellectual assent ("a firm belief in her party's platform").
- Sentiment: a settled opinion reflective of one's feelings ("her feminist sentiments are well-known").

Web 2.0 is an evolution from passive viewing of information to interactive creation of user generated data by the collaboration of users on the Web. Every facet of Web 2.0 is driven by contribution and collaboration. The evolution of Web from Web 1.0 to Web 2.0 was enabled by the rise of read/write platforms such as blogging, social networks, and free image and video sharing sites. These platforms have jointly allowed exceptionally effortless content creation and sharing by anyone. The research field of sentiment analysis has been rapidly progressing because of the rich and diverse data provided by Web 2.0 applications. Blogs, review sites, forums, micro blogging sites, wikis and social networks have all provided different dimensions to the data used for sentiment analysis.

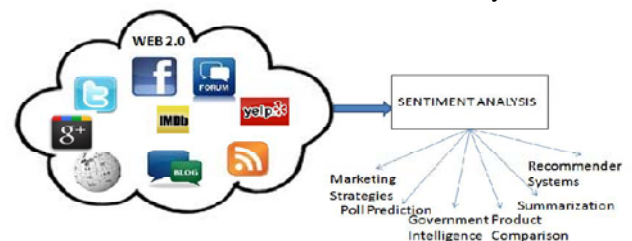


Figure 1: Conceptual model of Sentiment Analysis [1].

III. WEB 2.0 AND SENTIMENT ANALYSIS

As per literature survey infer about sentiment analysis require information gathering from various data source. In this paper various approaches use for data collection some of them are:

A. Review Sites:

A review site is a website which allows users to post reviews which give a critical opinion about people, businesses, products, or services. Most sentiment analysis work has been done on movie and product review sites ([4, 5]). The purpose of a review is to appraise a specific object, thus it is a single domain problem.

B. Blogs:

The term web-log or blog refers to a simple webpage consisting of brief paragraphs of opinion, information, personal diary entries, or links, called posts, arranged chronologically with the most recent first, in the style of an online journal ([6]). Sentiment analysis on blogs ([7]) has been used to predict movie sales, political mood and sales analysis.

C. Forums:

Forums or message boards allow its members to hold conversations by posting on the site. Forums are generally dedicated to a topic and thus using forums as a database allows us to do sentiment analysis in a single domain.

D. Social Networks:

Social networking is online services or sites which try to emulate social relationships amongst people who know each other or share a common interest. Social networking sites allow users to share ideas, activities, events, and interests within their individual networks.

IV. TASKS OF SENTIMENT ANALYSIS

Sentiment analysis is a challenging interdisciplinary task which includes natural language processing, web mining and machine learning. It is a complex task and encompasses several separate tasks.

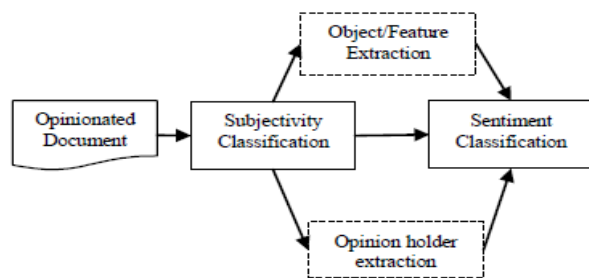


Figure: 2 Tasks of Sentiment Analysis [8].

A. Subjectivity Classification :

Typically, any given document will contain sentences that express opinion and some that do not. That is, a document is a collection of objective sentences, sentences that state a fact, and subjective sentences, sentences that represents the author's opinion, point of view or emotion. Subjectivity classification is the task of classifying sentences as opinionated or not opinionated ([1]).

B. Sentiment Classification

Once the task of finding whether a piece of text is opinionated is over we have to find the polarity of the text i.e., whether it expresses a positive or negative opinion. Sentiment classification can be a binary classification (positive or negative), multi-class classification (extremely

negative, negative, neutral, positive or extremely positive), regression or ranking.

C. Opinion Holder Extraction:

Sentiment Analysis also involves elective tasks like opinion holder extraction, i.e. the discovery of opinion holders or sources. Detection of opinion holder is to recognize direct or indirect sources of opinion. They are vital in news articles and other formal documents because multiple opinions can be expressed in the same article corresponding to different opinion holders. In documents like these, the multiple opinion holders may explicitly be mentioned by name. In social networks, review sites and blogs the opinion holder is usually the author who may be identified by the login credentials.

D. Object/Feature Extraction:

An additional task is the discovery of the target entity. In contrast with review sites, blogs and social media sites tend not have a set intention or predefined topic and are thus, inclined to discuss assorted topics. In such platforms it becomes necessary to know the target entity. Also as mentioned before target entities can have features or components that are being reviewed. A reviewer can have differing opinions about the different features or components of the target entity. As a result, feature based sentiment analysis, i.e. extraction of object feature and the related opinion, is an optional task of sentiment analysis ([8]).

V. LEVELS OF SENTIMENT ANALYSIS

The tasks described in the previous section can be done at several levels of granularity, namely, word level, phrase or sentence level, document level and feature level. The following Fig. 3 depicts the levels of granularity of sentiment analysis. The sentiment analysis tasks can be accomplished at the following levels of granularity.

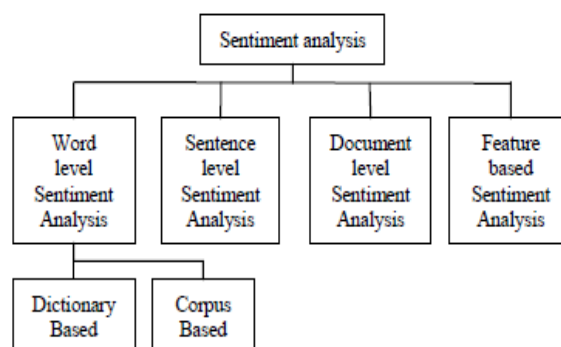


Figure: 3 Levels of Sentiment Analysis [8].

A. Document Level Sentiment Analysis:

Document-level sentiment analysis considers the whole document as the basic unit whose sentiment orientation is to be determined. To simplify the task, it is presumed that each text's overall opinion is completely held by a single opinion holder and is about a single object. The difficulty lies in the fact that there could be mixed opinions in a document, and with the creative nature of natural language, people may express the same opinion in vast ways, sometimes without using any opinion words. Also as stated earlier, a text is equally likely to contain objective sentences along with

subjective sentences. Therefore, tools are required to extract useful information from subjective sentences instead of objective ones. This leads to sentence level sentiment analysis.

B. Sentence Level Sentiment Analysis:

At sentence level, research has been done on detection of subjective sentences in a document from a mixture of objective and subjective sentences and then, the sentiment orientation of these subjective sentences is determined.

C. Word Level Sentiment Analysis:

The work to find semantic orientation at phrase level is an important task of sentiment analysis. Most works use the prior polarity ([9]) of words and phrases for sentiment classification at sentence and document levels. The two methods of automatically annotating sentiment at the word level are: Dictionary-based approaches and Corpus-based approaches.

a. Dictionary based Methods:

In this method, a small seed list of words with known prior polarity is created. This seed list is then extended by extracting synonyms or antonyms iteratively from online dictionary sources like WordNet1.

b. Corpus based Methods:

Corpus based methods rely on syntactic or statistical techniques like co-occurrence of word with another word whose polarity is known.

D. Feature Based Sentiment Analysis:

In a review, its author talks about the positives and negatives of a product. The reviewer may like some features and dislike some, even though the general opinion of the product may be positive or negative. This kind of information is not provided by document level or sentence level sentiment classification. Thus, feature based opinion sentiment analysis ([8, 10]) is required. This involves extracting product feature and the corresponding opinion about it.

VI. APPLICATION

Pang-Lee et al. (2002) broadly classifies the applications into the following categories.

- Applications to Review-Related Websites Movie Reviews, Product Reviews etc.
- Applications as a Sub-Component Technology Detecting antagonistic, heated language in mails, spam detection, context sensitive information detection etc.
- Applications in Business and Government Intelligence Knowing Consumer attitudes and trends
- Applications across Different Domains Knowing public opinions for political leaders or their notions about rules and regulations in place etc.

VII. CHALLENGES FOR SENTIMENT ANALYSIS

Sentiment Analysis approaches aim to extract positive and negative sentiment bearing words from a text and classify the text as positive, negative or else objective if it cannot find any sentiment bearing words. The general challenges can be summarized as:

A. Implicit Sentiment and Sarcasm:

A sentence may have an implicit sentiment even without the presence of any sentiment bearing words. Consider the following examples.

How can anyone sit through this movie?

One should question the stability of mind of the writer who wrote this book.

Both the above sentences do not explicitly carry any negative sentiment bearing words although both are negative sentences.

B. Domain Dependency:

There are many words whose polarity changes from domain to domain. Consider the following examples.

The story was unpredictable.

The steering of the car is unpredictable.

Go read the book.

In the first example, the sentiment conveyed is positive whereas the sentiment conveyed in the second is negative. The third example has a positive sentiment in the book domain but a negative sentiment in the movie domain.

C. Word Knowledge:

Often world knowledge needs to be incorporated in the system for detecting sentiments. Consider the following examples:

He is a Frankenstein.

Just finished Doctor Zhivago for the first time and all I can say is Russia sucks.

The first sentence depicts a negative sentiment whereas the second one depicts a positive sentiment. But one has to know about Frankenstein and Doctor Zhivago to find out the sentiment.

D. Subjectivity Detection:

This is to differentiate between opinionated and non-opinionated text. This is used to enhance the performance of the system by including a subjectivity detection module to filter out objective facts. But this is often difficult to do. Consider the following examples:

I hate love stories.

I do not like the movie "I hate stories". The first example presents an objective fact whereas the second example depicts the opinion about a particular movie.

E. Entity Identification:

A text or sentence may have multiple entities. It is extremely important to find out the entity towards which the opinion is directed. Consider the following examples.

Samsung is better than Nokia

Ram defeated Hari in football.

The examples are positive for Samsung and Ram respectively but negative for Nokia and Hari.

F. Negation:

Handling negation is a challenging task in SA. Negation can be expressed in subtle ways even without the explicit use of any negative word. A method often followed in handling negation explicitly in sentences like "I do not like the movie", is to reverse the polarity of all the words appearing after the negation operator (like not). But this does not work for "I do not like the

Acting but I like the direction". So we need to consider the scope of negation as well, which extends only till but

here. So the thing that can be done is to change polarity of all words appearing after a negation word till another negation word appears. But still there can be problems. For example, in the sentence “Not only did I like the acting, but also the direction”,

The polarity is not reversed after “not” due to the presence of “only”. So this type of combinations of “not” with other words like “only” has to be kept in mind while designing the algorithm.

VIII. CONCLUSION

This paper illustrates the research area of Sentiment Analysis and its latest advances. It affirms the major tasks, various challenges, and applications of sentiment analysis. Most work has been done on product reviews – documents that have a definite topic. More general writing with varied domains, such as blog posts, tweets, posts and web pages, have recently been creating & receiving attention. Future work in expanding existing techniques to handle more general writings and crossing domains is an exciting opportunity for both academia and businesses.

IX. REFERENCES

- [1]. Liu B. Sentiment Analysis and Subjectivity. Handbook of Natural Language Processing, Second edition, 2010.
- [2]. Wiebe, J., Wilson, T., Bruce, R., Bell, M., and Martin, M. Learning subjective language. Computational Linguistics, 2004, 30(3):277–308.
- [3]. Pang, B and Lee L. Opinion mining and sentiment analysis. Foundations and Trends in Information Retrieval 2008, (1-2), 1–135.
- [4]. Dave K., Lawrence S, and Pennock D.M. Mining the peanut gallery: Opinion extraction and semantic classification of product reviews. In Proceedings of the 12th international conference on World Wide Web (WWW), 2003, pp.:519–528.
- [5]. Theresa Wilson, Janyce Wiebe, and Rebecca Hwa. Just how mad are you? Finding strong and weak opinion clauses. In Proceedings of AAAI, 2004, pages 761–769.
- [6]. Anderson, P. What is Web 2.0? Ideas, technologies and implications for education. Technical report, JISC, 2007.
- [7]. Mishne G. and Glance N. Predicting movie sales from blogger sentiment. In AAAI Symposium on Computational Approaches to Analyzing Weblogs (AAAI-CAAW), 2006: 155–158.
- [8]. Theresa Wilson, Janyce Wiebe, and Rebecca Hwa. Just how mad are you? Finding strong and weak opinion clauses. In Proceedings of AAAI, 2004, pages 761–769.
- [9]. Hu, M. and Liu, B. Mining opinion features in customer reviews. In Proceedings of AAAI, 2004: 755–760 .
- [10]. Pang B. and Lee L. Seeing stars: Exploiting class relationships for sentiment categorization with respect to rating scales, Proceedings of the Association for Computational Linguistics (ACL), 2005:115–124.
- [11]. Yi, J., Nasukawa, T., Niblack, W., & Bunescu, R., Sentiment analyzer: extracting sentiments about a given topic using natural language processing techniques. Proceedings of the 3rd IEEE international conference on data mining (ICDM 2003):427–434.

Short Bio Data for the Author's



Miss. Siddhi S. Patni is doing M.E (CSE) from G.H Rasoni College of Engineering and Management, Amravati and has done B.E in Information Technology from SGBAU, Amravati.



Prof. Avinash P. Wadhe: Received the B.E and from SGBAU Amravati university and M-Tech (CSE) From G.H Rasoni College of Engineering, Nagpur (an Autonomous Institute). He is currently an Assistant Professor with the G.H Rasoni College of Engineering and Management, Amravati SGBAU Amravati University. His research interest include Network Security, Data mining and Fuzzy system .He has contributed to more than 20 research paper. He had awarded with young investigator award in international conference.