

Study of Different Levels for Sentiment Analysis

Seema Kolkur[†], Gayatri Dantal[‡] and Reena Mahe^{‡*}

[†]Computer Engineering Department, Thadomal Shahani College of Engineering, Maharashtra, India

[‡]Computer Engineering Department, Sh L R Tiwari College of Engineering, Maharashtra, India

Accepted 20 March 2015, Available online 25 March 2015, Vol.5, No.2 (April 2015)

Abstract

With the evolution of web technology, online shopping has become very popular as it is convenient and time saving. It also provides platform to share experiences and provide feedback. These opinions reflect whether the product is good or bad and helps to make decisions by potential customers. Sentiment analysis is the procedure by which information can be extracted from these reviews, analyzed and categorized as positive, negative or neutral. It is the combination of natural language processing and information extraction system. Main goal of Sentiment analysis is to train computer to be able to understand, recognize and generate emotions. This paper gives an overview of different levels on which sentiment analysis can be performed.

Keywords: Reviews, Polarity, Sentiment Analysis, Sentence level, Document level, Feature level etc.

1. Introduction

Human life is filled with emotions and opinions. They play a critical role in how individuals behave, think and act. With the exponential increase in the internet usage, people prefer to express and share information on different topics ranging from hotels and products to health and different services. Due to ever increasing existence of these emotions, opinions, views, feedbacks and suggestions on the web, it becomes necessary to explore, analyze and organize this information for better decision making by subsequent users. This new research domain is called sentiment analysis or opinion mining. It is one of the hottest research areas in computer science nowadays. Many researchers are focusing on this field and several techniques and tools have been evolved over the years. Sentiment classification, subjectivity classification, aspect based sentiment analysis and cross domain sentiment analysis are some of the research directions.

2. Major Challenges in Sentiment Analysis

Language is the most wonderful, dynamic and mysterious phenomenon in the universe (Jaganadh G, 2012). Language and its structure is the primary challenge. There are several open ended issues and challenges in the field of sentiment analysis. Few of them are:

2.1 Word Sense Disambiguation

*Corresponding author: Reena Mahe; Seema Kolkur and Gayatri Dantal are working as Assistant Professors

Often encountered problem is Word Sense Disambiguation WSD (S.ChandraKala *et al*, 2012; Aurangzeb Khan *et al*, 2011). Correct meaning of word based on the context needs to be extracted as word can have different meanings for different domains. For example *small size* can be positive opinion for mobile phones but negative for hotels.

2.2 Comparisons

To determine the polarity for comparative sentences can be a challenge. For example *Battery life of phone X is better than phone Y*. This review has positive word 'better' but the author's preferred object is not easy to determine which is the key piece of information in a comparative reviews.

2.3 Negations

Negations if not handled properly can give completely wrong results. For example *There is a good chance that this phone will not break easily*. This review shows positive polarity but presence of negation changes the effect completely.

2.4 Intensity

Depending upon the intensity of opinion (mild or strong), to obtain result as highly positive or highly negative can also be challenging. It is called as degree of polarity.

2.5 Sarcasm

Another interesting challenge can be of identification of sarcasm and to analyze emotions expressed in text at a more fine-grained level.

3. Different Levels of Sentiment Analysis

Three different levels on which sentiment analysis can be performed depending upon the granularities required are:

3.1 Document level Sentiment Analysis

This is the simplest form of classification. The whole document of opinionated text is considered as basic unit of information. It is assumed that document is having opinion about single object only (film, book or hotel). This approach is not suitable if document contains opinions about different objects as in forums and blogs. Classification for full document is done as positive or negative. Irrelevant sentences need to be eliminated before processing.

A lot of work (Peter D. Turney, 2002; Yan Zhao *et al*, 2014; Richa Sharma *et al*, 2014) has been done on document based sentiment analysis. There are two approaches to do classification.

1. Supervised machine learning approach
2. Unsupervised machine learning approach

In supervised machine learning approach there is finite set of classes for classification. Training dataset is also available. Given the training data, the system classifies the document by using one of the common classification algorithms such as Support Vector Machine, Naïve Bayes, K Nearest Neighbours, and Maximum Entropy etc. Document based classification has been done by (Yan Zhao *et al*, 2014) for news comments using different supervised machine learning approaches. Different approaches can be combined also for effective results. The Naive Bayes and Neural Network classifier are combined by (Lina L. Dhande and Dr. Prof. Girish K. Patnaik, 2014) for classifying movie reviews. They proved that accuracy of sentiment analysis is increased up to 80.65% by combining these two methods.

Different researches done by (Peter D. Turney, 2002; Richa Sharma *et al*, 2014; Gautam Kumar *et al*, 2012) analysed data using unsupervised machine learning approach. In unsupervised approach, Sentiment Orientation (SO) of opinion words in document is determined. If the SO of these words is positive then the document is classified as positive otherwise negative. In the most prominent work done by (Turney, 2002), two words Poor and Excellent have been used. The semantic orientation determines whether meaning of opinion is closer to positive word Excellent or negative word Poor. Point Wise Mutual information method is used to calculate the semantic orientation. Lexicon based method has been used to perform sentiment classification by (M. Taboada *et al*, 2011). The unsupervised dictionary based technique (WordNet) is used by (Richa Sharma *et al*, 2014) to determine the polarity of the movie reviews at

document level. In this paper Seed list contains opinion words along with their polarity. The overview of this system is:

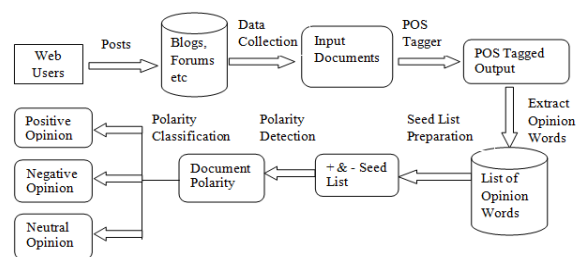


Fig. 1 Document based Sentiment Orientation System

3.2 Sentence Level Sentiment Analysis

Sentence level sentiment analysis is the most fine-grained analysis of the document. In this, polarity is calculated for each sentence as each sentence is considered as separate unit and each sentence can have different opinion. Sentence level sentiment analysis has two tasks:

3.2.1 Subjectivity Classification

A sentence can be either subjective sentence or objective sentence. Objective sentence contains the facts. It has no judgement or opinion about the object or entity while subjective sentence has opinions (e.g.), *India's economy is heavily dependent on tourism and IT industry. It is an excellent place to live in.* The first sentence is a factual one and does not convey any sentiment towards India. Hence this should not play any role in deciding on the polarity of the review and should be filtered out (S. ChandraKala and C. Sindhu, 2012). The advantage of sentence level analysis lies in the subjectivity/ objectivity classification.

3.2.2 Sentiment Classification

Sentence can be classified as positive, negative or neutral depending upon the opinion words present in it.

A number of researches (V.S. Jagtap *et al*, 2013; Raisa V *et al*, 2013; Gizem Gezici *et al*, 2012) focus on finding how to classify the text effectively. The same document level classification methods can be applied to the sentence level classification problem. A number of different methods are discussed and compared under supervised machine learning approach by (S Padmaja and Prof. S Sameen Fatima, 2013). Research done by (V. S. Jagtap and Karishma Pawar, 2013) defines pure, short and no unrealistic sentences in addition to the subjective sentences. They observe that first and last lines of a review are often indicative of review polarity. They performed an in depth analysis on different sentence types. A rule based domain independent sentiment analysis and lexical approach for classification of objective and subjective sentences

has been proposed by (Aurangzeb Khan and Baharum Baharudin, 2011). The semantic score of subjective sentence is extracted from SentiwordNet lexical resource.

3.3 Feature Level Sentiment Analysis

Feature engineering is an extremely basic and essential task for Sentiment Analysis (S. ChandraKala and C. Sindhu, 2012). The basic step in Feature Level sentiment analysis is to identify the piece of text as a feature of some product. For example Battery life is very long lasting. In this review Battery is product feature (noun) and 'very long lasting' is opinion word (adjective). The basic steps for feature based sentiment analysis are (Padmapani P. Tribhuvan et al, 2014):

1. Preparing Review Database
2. POS Tagging
3. Feature Extraction
4. Opinion Word Extraction
5. Opinion Word Polarity Identification
6. Opinion Sentence Polarity Identification
7. Summary Generation

Minqing Hu and Bing Liu's work is the most pioneering in this field (Minqing Hu and Bing Liu, 2004). They presented different techniques based on data mining and NLP methods like frequent features, compactness pruning, P-support pruning and infrequent feature identification. Experiments have been conducted using these techniques for various electronic products. These techniques further have been outperformed by frequent pattern mining algorithm called H-mine proposed by (Seyed Hamid Ghorashi et al, 2012). They state that using Apriori algorithm leads to increase the execution time while dealing with large databases. Various techniques and tools have been proposed by different researchers to enhance the performance and precision of the system.

A new approach is proposed by (D D Chaudhari et al, 2013) which uses feature oriented appraisal words lexicon. It is fine grained approach in which review categorization is based on attitude and polarity of the adjectival words for the frequent features of the product.

Conclusions

Sentiment Analysis is an emerging field of data mining. It is becoming more important as more people like to buy online and give feedback, reviews and comments on the products. This paper discusses about an overview of Sentiment Analysis and its various level on which this analysis can be conducted in form of Sentence, Document and Feature level sentiment analysis. Various challenges are also discussed that make sentiment analysis a difficult task.

References

- Jaganadh G Opinion, (2012), Mining and Sentiment Analysis, *CSI Communications*
- S. ChandraKala and C. Sindhu, (2012), Opinion Mining And Sentiment Classification: A Survey, *ICTACT Journal on Soft Computing*, Vol- 03, ISSUE: 01, ISSN: 2229-6956
- Aurangzeb Khan, Baharum Baharudin, (2011), Sentiment Classification by Sentence Level Semantic Orientation using SentiWordNet from Online Reviews and Blogs, *Int. J Comp Sci. Emerging Tech*, Vol-2 No 4
- Richa Sharma, Shweta Nigam and Rekha Jain, (2014), Opinion Mining Of Movie Reviews At Document Level, *International Journal on Information Theory (IJIT)*, Vol.3, No.3
- Peter D. Turney, (2002), Thumbs Up or Thumbs Down? Semantic Orientation Applied to Unsupervised Classification of Reviews, *Proceedings of the 40th Annual Meeting of the Association for Computational Linguistics (ACL)*, Philadelphia, pp. 417-424
- M. Taboada, J. Brooke, M. Tofiloski, K. Voll, and M. Stede, (2011), Lexicon-based methods for sentiment analysis, *Association for Computational Linguistics*, vol. 37
- Gautam Kumar, Pawan kumar Goel, Sanjeev kumar Chauhan, Anand kumar Pandey, (2012), Opinion mining and summarization for customer reviews, *International Journal of Engineering Science and Technology (IJEST)*, Vol. 4 No.08, ISSN: 0975-5462
- Yan Zhao, Suyu Dong and Leixiao Li, (2014), Sentiment Analysis on News Comments Based on Supervised Learning Method, *International Journal of Multimedia and Ubiquitous Engineering*, Vol.9, No.7 pp.333-346
- Lina L. Dhande and Dr. Prof. Girish K. Patnaik, (2014), Analyzing Sentiment of Movie Review Data using Naive Bayes Neural Classifier, *International Journal of Emerging Trends & Technology in Computer Science (IJETTC)*, Volume 3, Issue 4, ISSN 2278-6856
- S Padmaja and Prof. S Sameen Fatima, (2013), Opinion Mining and Sentiment Analysis –An Assessment of Peoples' Belief: A Survey, *International Journal of Ad hoc, Sensor & Ubiquitous Computing (IJASUC)*, Vol.4, No.1
- V. S. Jagtap, Karishma Pawar, (2013), Analysis of different approaches to Sentence-Level Sentiment Classification, *International Journal of Scientific Engineering and Technology*, PP : 164-170
- Raisa Varghese, Jayasree M, (2013) A Survey on Sentiment Analysis and Opinion Mining, *International Journal of Research in Engineering and Technology (IJRET)*, eISSN: 2319-1163 pISSN: 2321-7308
- Gizem Gezici, Berrin Yanikoglu, Dilek Tapucu, and Yucel Saygin, (2012), New Features for Sentiment Analysis: Do Sentences Matter? *First International Workshop on Sentiment Discovery from Affective Data (SDAD)*
- Padmapani P. Tribhuvan, S.G. Bhirud, Amrapali P. Tribhuvan, (2014), A Peer Review of Feature Based Opinion Mining and Summarization, *International Journal of Computer Science and Information Technologies*, Vol. 5 (1), ISSN: 0975-9646
- Minqing Hu and Bing Liu, (2004), Mining and Summarizing Customer Review, *ACM*
- Seyed Hamid Ghorashi, Roliana Ibrahim, Shirin Noekhah and Niloufar Salehi Dastjerdi, (2012), A Frequent Pattern Mining Algorithm for Feature Extraction of Customer Reviews *International Journal of Computer Science Issues*, Vol. 9, Issue 4, No 1, ISSN: 1694-0814
- D D Chaudhari, R A Deshmukh, A B Bagwan, P K Deshmukh, (2013), Feature based approach for Review Mining Using Appraisal Words, *IEEE*, ISBN 978-1-4799-1082-3