RESEARCH ARTICLE OPEN ACCESS

A Survey on Sentence Level Sentiment Analysis

Vrushali K. Bongirwar [1]
Department of Computer Science and Engineering
Ramdeobaba College of Engineering & Management Nagpur
India

ABSTRACT

Sentiment Analysis, also called Opinion Mining, is one of the most recent research topics within the field of Information Processing. Textual information retrieval techniques are mainly focused on processing, searching or mining factual information. Textual information also have some objective as well as subjective characteristics. These elements are mainly opinions, sentiments, appraisals, attitudes, and emotions, which are the focus of Sentiment Analysis.
Text sentiment analysis typically work at a particular level like phrase, sentence or document level. This paper presents survey of various sentiment analysis methods on different levels. Further it extends the literature on sentence level.
***Keywords:-* Document Level, Phrase Level, Polarity, Sentence Level, Sentiment Analysis.**

I. INTRODUCTION

Sentiment analysis refers to the inference of people's views, positions and attitudes in their written or spoken texts. Before the coining of the term, the field was studied under names such as subjectivity, point of view and opinion mining. Nowadays, the field is rapidly evolving due to the rise of new platforms such as blogs, social media and user-generated reviews. A large body of work exists on the analysis of latent sentiment in social media platforms such as Twitter. The goal of these studies is to extract timely and relevant information as well as to judge widespread opinions and sentiment.

Sentiment Analysis offers many opportunities to develop new applications, especially due to the huge growth of available information in sources such as blogs and social networks. For example, recommendations of items proposed by any recommender system can be computed taking into account aspects such as positive or negative opinions about those items. Review- and opinion-aggregation websites could collect information from different sources in order to summary or compose an opinion about a candidate, product, etc., thus replacing systems which require explicitly opinions or summaries. Therefore one of the most important fields where Sentiment Analysis has a greater impact is in the industrial field. Small and big companies, as well as other organizations such as governments, desire to know what people say about their marques, products or members.

II. LEVELS OF ANALYSIS

Sentiment analysis has been handled as a Natural Language Processing task at many levels of granularity. Depending on whether the target of study is a whole text or document, one or several linked sentences, or one or several entities or aspects of those entities, different NLP and Sentiment Analysis tasks can be performed. Hence, it is necessary to distinguish three levels of analysis that will clearly determine the different tasks of Sentiment Analysis: (i) document level, (ii) sentence level and (iii) entity/aspect level [1].

A. Document Level Analysis: Document level considers that a document is an opinion on an entity or aspect of it. This level is associated with the task called document-level sentiment classification [10], [11],[12]. The task is to classify whether a whole opinion document expresses a positive or negative sentiment For example, given a product review, the system determines whether the review expresses an overall positive or negative opinion about the product. This task is commonly known as document-level sentiment classification. [19][25].

B. Sentence Level Analysis: The task at this level goes to the sentences and determines whether each sentence expressed a positive, negative, or neutral opinion. Neutral usually means no opinion. This level of analysis is closely related to subjectivity classification which distinguishes sentences (called objective sentences) that express factual information from sentences (called subjective sentences) that express subjective views and opinions. However, we should note that subjectivity is not equivalent to sentiment as many objective sentences can imply opinions. [13], [14].

C. Feature Level Analysis: Both the document level and the sentence level analyses do not discover what exactly people liked and did not like. Aspect level performs finer-grained analysis. Aspect level was earlier called feature level (feature-based opinion mining and summarization).Instead of looking at language constructs (documents, paragraphs, sentences, clauses or phrases), aspect level directly looks at the opinion itself. It is based on the idea that an opinion consists of a sentiment (positive or negative) and a target (of opinion). It is closely related to tasks like Feature-based Opinion Mining and Opinion Summarization [15], [16],[17][25].

III.SENTIMENT ANALYSIS MODEL

The typical Sentiment Analysis Model is shown in the figure 1. The data preparation step performs necessary data preprocessing and cleaning on the dataset for the subsequent analysis. Some commonly used preprocessing steps include removing non-textual contents and markup tags (for HTML pages), and removing information about the reviews that are not required for sentiment analysis, such as review dates and reviewers' names. The review analysis step analyzes the linguistic features of reviews so that interesting information, including opinions and/or product features, can be identified. Two commonly adopted tasks for review analysis are POS tagging [1] and negation tagging. After this phase, sentiment classification is performed to get the results.

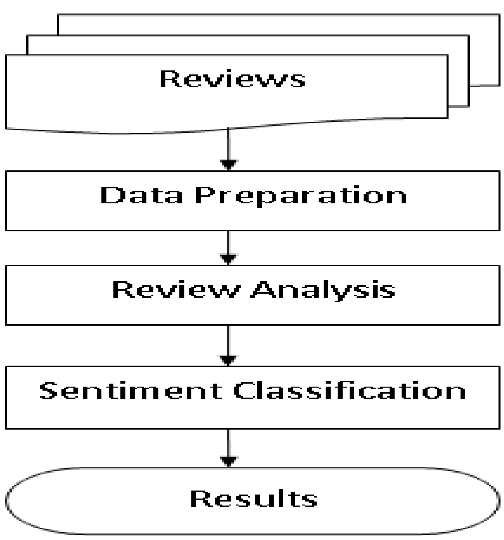

Figure 1 Typical Sentiment Analysis Model

IV. APPROACHES TO SENTIMENT CLASSIFICATION

In this section, approaches for sharing the common theme of mapping a given piece of text, such as a document, paragraph, or sentence, to a label drawn from a pre-specified finite set or to a real number has been discussed.

a) Machine learning approaches:
They can be grouped in two main categories: supervised and unsupervised techniques. The success of both is mainly based on the selection and extraction of the appropriate set of features used to detect sentiments. In this task Natural Language Processing techniques play a very important role. With respect to supervised techniques, support vector machines (SVM), Naive Bayes, Maximum Entropy are some of the most common techniques used. [2],[3],[5]

b) Semantic Orientation Approach
The semantic orientation approach performs classification based on positive and negative sentiment words and phrases contained in each evaluation text . It does not require prior training in order to mine the data. Two types of techniques have been used in previous sentiment classification research using the semantic orientation approaches.[25]

The corpus-based techniques: Corpus-based techniques try to find co-occurrence patterns of words to determine their sentiments. Turney calculated a phrase's semantic orientation to be the mutual information between the phrase and the word "excellent" (as positive polarity) minus the mutual information between the phrase and the word "poor" (as negative polarity). The overall polarity of an entire text was predicted as the average semantic orientation of all the phrases that contained adjectives or adverbs. Riloff and Wiebe used a bootstrapping process to learn linguistically rich patterns of subjective expressions in order to classify subjective expressions from objective expressions.[25]

The dictionary-based techniques : Dictionary-based techniques use synonyms, antonyms and hierarchies in WordNet (or other lexicons with sentiment information) to determine word sentiments.[20]

c) Lexicon-based approaches:
Lexicon-based approaches mainly rely on a sentiment lexicon, i.e., a collection of known and precompiled sentiment terms, phrases and even idioms, developed for traditional genres of communication, such as the Opinion Finder lexicon; but, even more complex structures like ontologies, or dictionaries measuring the semantic orientation of words or phrases can be used for this purpose [7],[8],[9].

d) Other Unsupervised Approaches
Bootstrapping is another approach. The idea is to use the output of an available initial classifier to create labeled data, to which a supervised learning algorithm may be applied. This method was used in conjunction

with an initial high-precision classifier to learn extraction patterns for subjective expressions. Pang and Lee [19] experiment with a different type of unsupervised approach. The problem they consider is to rank search results for review-seeking queries so that documents that contain evaluative text are placed ahead of those that do not. They propose a simple "blank slate" method based on the rarity of words within the search results that are retrieved (as opposed to within a training corpus). The intuition is that words that appear frequently within the set of documents returned for a narrow topic (the search set) are more likely to describe objective information, since objective information should tend to be repeated within the search set; in contrast, it would seem that people's opinions and how they express them may differ. Counter intuitively, though, Pang and Lee find that when the vocabulary to be considered is restricted to the most frequent words in the search set (as a noise-reduction measure), the subjective documents tend to be those that contain a higher percentage of words that are less rare, perhaps due to the fact that most reviews cover the main features or aspects of the object being reviewed. (This echoes our previous observation that understanding the objective information in a document can be critical for understanding the opinions and sentiment it expresses.) The performance of this simple method is on par with that of a method based on a state-of-the-art subjectivity detection system, Opinion Finder [21],[22][25]

V. ANALYSIS AND CONCLUSION

Most existing techniques for document level sentiment classification are based on supervised learning, for Example, n-gram features, Naïve Bayes, Maximum Entropy classification, and SVM can be applied to achieve the high performance [19]. The results produced via machine learning techniques are quite better in comparison to the baselines Naïve Bayes tends to do the worst and SVMs tend to do the best, although the differences are not very large.

It is clear that although we may be able to build comprehensive lexicons of sentiment-annotated words, it is still a challenge to accurately locate it in text. Few studies have been done outside the realm of short documents like product reviews, and especially in difficult domains like political commentaries. This is true partially because there is little annotated data available for realms outside reviews. The Method called syntax tree pruning and tree kernel based approach to sentiment classification to sentence-level sentiment classification can be applied [23].

If the opinion words are included in dictionary then it must contain all words. It is important for lexicon based approach. Because it will reduce the performance if there are fewer words present in dictionary. Another significant challenge to this approach is that the polarity of many words is domain and context dependent. For example, 'funny movie' is positive in movie domain and 'funny taste' is negative in food domain. Such words are associated with sentiment in a particular domain.

REFERENCES

[1] Apoorv Agarwal, Fadi Biadsy, and Kathleen Mckeown. 2009. Contextual phrase-level polarity analysis using lexical affect scoring and syntactic n-grams. Proceedings of the 12th Conference of the European Chapter of the ACL (EACL 2009), pages 24–32, March.

[2] Y. He, D. Zhou, Self-training from labeled features for sentiment analysis, Inform. Process. Manage. 47 (4) (2011) 606–616.

[3] A.C. Konig, E. Brill, Reducing the human overhead in text categorization, in: Proceedings of the 12th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining – KDD '06, ACM Press, 2006, pp. 598–603.

[4] A. Montejo-Raez, E. Martinez-Camara, M.T. Martin-Valdivia, L.A. Urena Lopez, Ranked WordNet graph for sentiment polarity classification in twitter, Comput. Speech Lang. 28 (1) (2014) 93–107.

[5] M. Rushdi-Saleh, M. Martin-Valdivia, A. Montejo-Raez, L. Urena Lopez, Experiments with SVM to classify opinions in different domains, Expert Syst.Appl. 38 (12) (2011) 14799–14804.

[6] Q. Ye, Z. Zhang, R. Law, Sentiment classification of online reviews to travel destinations by supervised machine learning approaches, Expert Syst. Appl. 36 (3) (2009) 6527–6535.

[7] E. Kontopoulos, C. Berberidis, T. Dergiades, N. Bassiliades, Ontology-based sentiment analysis of twitter posts, Expert Syst. Appl. 40 (10) (2013) 4065–4074.

[8] T. Wilson, P. Hoffmann, S. Somasundaran, J. Kessler, J. Wiebe, Y. Choi, C. Cardie, E. Riloff, S. Patwardhan, OpinionFinder, in: Proceedings of HLT/EMNLP on Interactive Demonstrations, Association for Computational Linguistics, 2005, pp. 34–35.

[9] D. Wang, S. Zhu, T. Li, SumView: a Web-based engine for summarizing product reviews and customer opinions, Expert Syst. Appl. 40 (1) (2013) 27–33.

[10] C. Zhang, D. Zeng, J. Li, F.-Y. Wang, W. Zuo, Sentiment analysis of Chinese documents: from sentence to document level, J. Am. Soc. Inform. Sci. Technol. 60 (12) (2009) 2474–2487.

[11] R. Moraes, J.F. Valiati, W.P. Gaviao Neto, Document-level sentiment classification: an empirical comparison between SVM and ANN, Expert Syst. Appl. 40 (2) (2013) 621–633.

[12] G. Li, F. Liu, Application of a clustering method on sentiment analysis, J. Inform. Sci. 38 (2) (2012) 127–139.

[13] T. Wilson, J. Wiebe, P. Hoffmann, Recognizing contextual polarity in phrase-level sentiment analysis, in: Proceedings of the Conference on Human Language Technology and Empirical Methods in Natural Language Processing (HLT '05), Association for Computational Linguistics, Morristown, NJ, USA, 2005, pp. 347–354.

[14] T. Wilson, J. Wiebe, P. Hoffmann, Recognizing contextual polarity: an exploration of features for phrase-level sentiment analysis, Comput. Linguist. 35 (3) (2009) 399–433.

[15] B.A. Ojokoh, O. Kayode, A feature-opinion extraction approach to opinion mining, J. Web Eng. 11 (1) (2012) 51–63

[16] T.T. Thet, J.-C. Na, C.S.G. Khoo, Aspect-based sentiment analysis of movie reviews on discussion boards, J. Inform. Sci. 36 (6) (2010) 823–848.

[17] O. Vechtomova, Facet-based opinion retrieval from blogs, Inform. Process. Manage. 46 (1) (2010) 71–88.

[18] Q. Li, T. Wang, P. Li, L. Liu, Q. Gong, Y. Chen, The effect of news and public mood on stock movements, Inform. Sci. 278 (2014) 826–840.

[19] Pang B., Lee, L., And Vaithyanathain. S. 2002.Thumbs up? Sentiment classification using machine learning techniques, In Proceedings of the Conference on Empirical Methods in Natural Language Processing , 79-86.

[20] V. S. Jagtap, Karishma Pawar "Analysis of different approaches to Sentence-Level Sentiment Classification" International Journal of Scientific Engineering and Technology Volume 2 Issue 3, April 2013, PP : 164-170

[21] Jack G. Conrad and Frank Schilder, "Opinion mining in legal blogs", In Proceedings of the International Conference on Artificial Intelligence and Law (ICAIL), New York, NY, USA, 2007, pages 231–236.

[22] Janyce M. Wiebe and Ellen Riloff, "Creating subjective and objective sentence classifiers from unannotated texts", In Proceedings of the Conference on Computational Linguistics and Intelligent Text Processing (CICLing), number 3406 in Lecture Notes in Computer Science, 2005 PP 486–497.

[23] Z. Wei, L. Perifeng, Z. Qiaoming, "Sentiment Classification Based on Syntax Tree Pruning and Tree Kernel", In Proceedings of the Conference on Web Information Systems and Applications, China, 2010.

[24] V. S. Jagtap, Karishma Pawar "Analysis of different approaches to Sentence-Level Sentiment Classification" International Journal of Scientific Engineering and Technology, (2013), PP : 164-170

[25] Neha S. Joshi, Suhasini A. Itkat "A Survey on Feature Level Sentiment Analysis", (IJCSIT) International Journal of Computer Science and Information Technologies, Vol. 5 (4) , (2014),PP: 5422-5425