Opinion Dependent Aspect Based Sentiment Analysis

Dr Y. Nagesh
Associate Professor
Dept of Information Technology
Mettu University
Mettu,
Ethiopia
nageshyagnam1@gmail.com

Dr. K. Raja Shekar
Assistant Professor
Dept. Of Computer & Science
Mettu University
Mettu,
Ethiopia
rajashekar562@gmail.com

Dr. B. Bhoomeshwar
Associate Professor
Dept. Of Computer & Science
Mettu University
Mettu
Ethiopia
eshwarmsis@gmail.com

Abstract

Sentiment analysis has been emerging research field in Computational Linguistics, Text Analysis and Natural Language Processing (NLP). This is the computational study of people's opinion towards entities and their aspects. Entities refer to individuals, events, topics, products and organizations. Aspects are attributes or components of entities. Presently the social media has become an excellent source to express and share people's opinion on entities and their aspects. In the form of comments, reviews, blogs, tweets, status updates, etc., it is harder for people to analyze all opinions at a time to make good decisions. So, there is a need for effective automated systems to evaluate opinions and generate accurate results.

Keywords:

Sentiment Analysis, Emotion, Analysis, Subjectivity Detection, Polarity detection, Text Analysis

I. Introduction

Textual information in the world can be broadly classified into two types: Facts and Opinions. Facts are objective expressions about entities, events and their properties. Opinions are usually subjective expressions that describe people's sentiments, appraisals or feeling toward entities, events and their properties.

The concept of opinion is very broad; we will focus on opinion expressions that convey people's positive or negative sentiments. Much of the existing research on textual information processing has been focused on mining and retrieval of factual information, e.g., information retrieval, Web search, text classification, text clustering and many other text mining and natural language processing tasks. Little work had been done on the processing of opinions until only recently. Yet, opinions are so important that whenever we need to make a decision, we want to hear others' opinions. This is not only true for individuals but also true for organizations.

One of the main reasons for the lack of study on opinions is the fact that there was little opinionated text available before the World Wide Web. Before the Web, when an individual needed to make a decision, one has to typically ask for opinions from friends and families. When an organization wanted to find the opinions or sentiments of the general public about its products and services, it conducted opinion polls, surveys, and focus groups. However, with the Web, especially with the explosive growth of the user generated content on the Web in the past few years, the world has been transformed.

II. APPROACHES IN ASPECT BASED SENTIMENT ANALYSIS

A. Sentiment and subjectivity classification

Two important sub-topics are useful while analyzing sentiment analysis they are: (1) classifying an opinionated document as expressing a positive or negative opinion, and (2) classifying a sentence or a clause of the sentence as subjective or objective, and for a subjective sentence or clause classifying it as expressing a positive, negative or neutral opinion. The first topic, commonly known as sentiment classification or document-level sentiment classification, aims to find the general sentiment of the author in an opinionated text. For example, given a product review, it determines whether the reviewer is positive or negative about the product. The second topic goes to individual sentences to determine whether a sentence expresses an opinion or not (often called subjectivity classification), and if so, whether the opinion is positive or negative (called sentence-level sentiment classification).

B. Feature-based sentiment analysis

This model first discovers the targets on which opinions have been expressed in a sentence, and then determines whether the opinions are positive, negative or neutral. The targets are objects, and their components, attributes and features. An object can be a product, service, individual, organization, event, topic, etc. For instance, in a product review sentence, it identifies product features that have been commented on by the reviewer and determines whether the comments are positive or negative. For example, in the sentence, "The battery life of this camera is too short," the comment is on "battery life" of the camera object and the opinion is negative. Many real life applications require this level of detailed analysis because in order to make product improvements one needs to know what components and/or features of the product are liked and disliked by consumers. Such information is Not discovered by sentiment and subjectivity classification.

C. Sentiment analysis of comparative sentences

Evaluation of an object can be done in two main ways, direct appraisal and comparison. Direct appraisal, called direct opinion, gives positive or negative opinion about the object without mentioning any other similar objects. Comparison means to compare the object with some other similar objects (e.g., competing products). For example, "The picture quality of this camera is poor" expresses a direct opinion, while "The picture quality of this camera is better than that of Camera-x." expresses a comparison. Clearly, it is useful to identify such sentences, extract comparative opinions expressed in them and determine which objects are preferred by the sentence authors (in the above example, Camera-x is preferred with respect to the picture quality).

D. Opinion search and retrieval

Since the general Web search has been so successful in many aspects, it is not hard to imagine that opinion search will be very useful as well. For example, given a keyword query "gay marriage", one wants to find positive and negative opinions on the issue from an opinion search engine. For such a query, two tasks need to be performed: (1) retrieving documents or sentences that are relevant to the query, and (2) identifying and ranking opinionated documents or sentences from these retrieved. Opinion search is thus a combination of information retrieval and sentiment analysis.

E. Opinion spam and utility of opinions

As opinions on the Web are important for many applications, it is no surprise that people have started to game the system. Opinion spam refers to fake or bogus opinions that try to deliberately mislead readers or automated systems by giving undeserving positive opinions to some target objects in order to promote the objects and/or by giving malicious negative opinions to some other objects in order to damage their reputations. Detecting such spam is very important for applications. The utility of opinions refers to the usefulness or quality of opinions. Automatically assigning utility values to opinions is useful as opinions can then be ranked based on their utility values. With the ranking, the reader can focus on those quality opinions. We should note, however, that spam and utility are very different concepts.

F. Opinion Lexicon Generation

In the research literature, opinion words are also known as polar words, opinion-bearing words, and sentiment words. Positive opinion words are used to express desired states while negative opinion words are used to express undesired states. Examples of positive opinion words are: beautiful, wonderful, good, and amazing. Examples of negative opinion words are bad, poor, and terrible. Apart from individual words, there are also opinion phrases and idioms, e.g., cost someone an arm and a leg. Collectively, they are called the opinion lexicon. They are instrumental for sentiment analysis for obvious reasons.

Opinion words can, in fact, be divided into two types, the base type and the comparative type. All the examples above are of the base type. Opinion words of the comparative type are used to express comparative and superlative opinions. Examples of such words are better, worse, best, worst, etc, which are comparative and superlative forms of their base adjectives or adverbs, e.g., good and bad. Unlike opinion words of the base type, the words of the comparative type do not express a direction opinion/sentiment on an object, but a comparative opinion/sentiment on more than one object, e.g., “Car-x is better than Car-y”. This sentence tells something quite interesting. It does not express an opinion that any of the two cars is good or bad. It just says that comparing to Car-y, Car-x is better, and comparing to Car-x, Car-y is worse. Thus, although still it is possible to assign a comparative word as positive or negative based on whether it represents a desirable or undesirable state, it is also not possible to use it in the same way as an opinion word of the base type. Opinion words can, in fact, be divided into two types, the base type and the comparative type. Opinion words of the comparative type are used to express comparative and superlative opinions. Examples of such words are better, worse, best, worst, etc, which are comparative and superlative forms of their base adjectives or adverbs.

G. Feature-Based Sentiment Analysis

Although classifying opinionated texts at the document level or at the sentence level is useful in many cases, they do not provide the necessary detail needed for some other applications. A positive opinionated document on a particular object does not mean that the author has positive opinions on all aspects or features of the object. Likewise, a negative opinionated document does not mean that the author dislikes everything. In a typical opinionated text, the author writes both positive and negative aspects of the object, although the general sentiment on the object may be positive or negative.

Document-level and sentence-level classification does not provide such information. At the feature level, the mining task is to discover every quintuple (oj, fjk, ooijkl, hi, tl) and identify all the synonyms (Wjk) and feature indicators Ijk of feature fjk. It is necessary to focus on two key mining tasks: 1. Identify object features that have been commented on. For instance, in the sentence, “The picture quality of this camera is amazing,” the object feature is “picture quality”. 2. Determine whether the opinions on the features are positive, negative or neutral. In the above sentence, the opinion on the feature "picture quality" is positive.

III. ASPECT BASED SENTIMENT ANALYSIS TASKS

The most important objective of Aspect Based Sentiment Analysis is to identify the aspects of the given target entities and sentiment expressed for each aspect. The objectives of Aspect Based Sentiment Analysis can be done through the following tasks. The first task is the extraction of aspect terms and grouping aspect terms into aspect categories. The second task is about identification of polarity of the aspect terms and polarity of the aspect categories of each sentence. The above tasks are divided into four sub tasks namely: Aspect Term Extraction (ATE), Aspect Term Polarity (ATP), Aspect Category Detection (ACD) and Aspect Category Polarity (ACP).

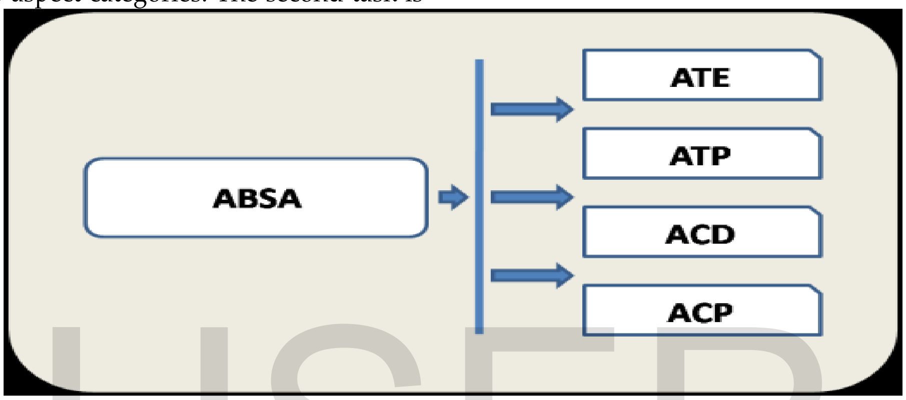

Fig.1. Aspect based sentiment analysis task

A. Aspect Term Extraction

The work of first sub-task Aspect Term Extraction (ATE) is also known as information extraction task is to identify all the aspect terms given in each review sentence. There can be multiple aspects, in a sentence and every aspect need to be extracted. The aspect in the aspect terms of the sentence can be expressed by a noun, verb, adverb and adjective. It is well known that 60% - 70% of aspect terms are explicit nouns. The aspect terms can also consist of multiword entities such as "screen size". These multiword entities and their aspects are considered to be much critical than single word aspects. The researchers have used various processes for extracting aspect terms, like Word N-grams, Bigrams, Word cluster, Casting, POS tagging, Parse dependencies, Relations and Punctuation marks. The various Methods used for extracting aspect terms, such as Conditional Random Fields (CRF), Support Vector Machines (SVM), Random trees and Random Forest.

The method of Hu and Liu (2004) extracts all the different nouns and noun phrases from the reviews of each dataset and consider them as candidate distinct aspect terms. In a co-occurrence based method [12] for category discovery a dictionary based sentiment classification algorithm is used through which aspects can be identified by annotation process. On the other hand, by using the training set to count how frequently each word appears within an aspect, a simple probability should be computed , which specifies the chance that this word is an aspect word or not. The above probability is also used for filtering a set of noun phrases, such that the noun phrases remaining will have at least one word. The aspect probability for the noun phrases holding the value greater than or equal to 0.05 and the noun phrases with the probability below 0.05 are removed. This process will remove some of the determiner words from the initial noun phrase, as those are excluded.

B. Aspect Term Polarity

The second sub-task is aspect term polarity is that, within a sentence for a given set of aspect terms, the task is to determine the polarity of each aspect term: positive, negative, neutral or conflict (i.e., both positive and negative). Here in the identification of Aspect term polarity different features like Word N-grams, Polarity of neighboring adjectives, Neighboring POS tags and Parse dependencies and relations have been widely used by researchers. The sentiment of aspect in [12] is computed by using sentiment value of each n-gram and distance between the n-gram and the aspect. In [13] Aspect lexicon based on additional information such as POS for polarity identification was developed by the author. In [14] a new class called conflict has been introduced along with they developed a method called RFC (Random Forest Classification. They have used many futures in this classification like local context, POS, chunk, prefix and suffix. In [15] aspect term polarity, they have extracted it by using various features like word N-grams, polarity of neighboring adjectives, neighboring POS tags and parse dependencies and relations.In [16] the author reusing the generated Word2Vec model, developed a polarity lexicon for the corresponding domain with the perception that a polarity word in a domain should be more "similar" to a set of "very positive" words than to a set of "very negative" words, and vice versa. This is engaged the in-domain generated Word2Vec models since the polarity of words may differ between domains and wanted to detain the polarity for each particular domain. In [17] the words that affect the sentiment of the aspect term are assumed to be close in most of cases and thus used a context window of 10 words in both directions around the target aspect term.

C. Aspect Category Detection

The third sub-task is Aspect Category Detection, in which the task is to identify the majority of categories that are discussed in each sentence. Aspect categories are usually difficult to find than the aspect terms as defined in Aspect Term Extraction, and at times they do not even occur as terms in the sentence. Aspect category detection [17] is based on a set of binary Maximum Entropy classifiers. The final decision is merely calculated from decisions of various individual classifiers. Aspect category classification [11] is based on a set of available binary classifiers, one classifier for each category found in the training set. To create a training example each sentence in the training set, the extracted features from all words in the sentence is taken. The co-occurrence based algorithm [12] is used for category detection. The algorithm is a co-occurrence matrix that captures the frequency of the co-occurrences between words in the sentence and the annotated aspect category, gives mapping from words to aspect categories. Aspect category detection [15] is considered as multi label classification problem. In a given instance, it should predict all labels that instance fit into. In [18] Aspect category detection the authors have used supervised classification approach and each task is done by identifying every entity E and attribute A pair E,A towards which an opinion is express. calculating the distance between n-gram and the corresponding aspect. The aspect category polarity has been detected using just unigram and bigram features in [15].

aspect category discussed in review sentence. The sentiment of aspect category [12] is computed by from both train and test data and implemented them on variety of classifiers (like Stochastic Gradient Descent, SVM, Adaboost) multiple times and stored the confidence scores obtained from decision functions of each of these classifiers.

services. There is a real and huge need in the industry for such services because every company wants to know how consumers perceive their products and services and those of their competitors. It clearly shows Sentiment analysis is the key to future events and predicting future outcomes of public opinions

[18] for sentiment polarity classification, authors have extracted Bag of Words and Word net Synset features

D. Aspect Category Polarity

The final sub-task is Aspect Category Polarity is which it takes the information from the previous task (Aspect Category Detection) to determine the polarity of each

REFERENCES

1. Handbook of Natural Language Processing, Second Edition, (editors: N. Indurkhya and F. J. Damerau), 2010
2. International Journal of Computer Applications (0975 – 8887) Volume 106 – No.3, November 2014, Aspect-based Opinion Mining: A Survey
3. International Journal of Innovative Research in Computer and Communication Engineering Volume 5,Issue 4,April 2017, A Survey on Opinion based Mining
4. International Advanced Research in Computer Science and Software Engineering Volume 6,Issue 4,April 2016, A Comprehensive Survey on Aspect based Sentiment Analysis
5. International Journal of Current Engineering and Technology Volume 5,Issue 6,Dec 2015,Aspect based Opinion Mining and Ranking: Survey
6. Bing Liu(2012).”Sentiment Analysis and Opinion Mining”, Synthesis Lectures on Human Language Technologies, Morgan and Clay pool Publishers
7. M. Hu and B. Liu, “Mining Opinion Features in Customer Reviews,” in Proceedings of the 19th National Conference on Artificial Intelligence (AAAI 2004). AAAI,
8. 2004, pp. 755–760. “Mining and Summarizing Customer Reviews,” in Proceedings of 10th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD 2004). ACM, 2004, pp. 168–177

IV. Conclusion

All the sentiment analysis tasks are very challenging. Understanding and knowledge of the problem and its

evident from the large number of start-up companies that offer sentiment analysis or opinion mining

9. C. Scaffidi, K. Bierhoff, E. Chang, M. Felker, H. Ng, and C. Jin, “Red Opal: Product-Feature Scoring from Reviews,” in Proceedings of the 8th ACM Conference on Electronic Commerce (EC 2007). ACM, 2007, pp. 182–191
10. Z. Li, M. Zhang, S. Ma, B. Zhou, and Y. Sun, “Automatic Extraction for Product Feature Words from Comments on the Web,” in Proceedings of the 5th Asia Information Retrieval Symposium on Information Retrieval Technology (AIRS 2009). Springer, 2009, pp. 112–123
11. P.T Ngoc and M.Yoo. “The lexicon-based sentiment analysis for fan page ranking in face book”, In Information Networking(ICOIN),2014,International Conference on IEEE,2014,pp. 444448
12. Schouten, Kim, Flavius Frasincar, and Franciska de Jong. "Commit-p1wp3: A co-occurrence based approach to aspect-level sentiment analysis." Proceedings of the 8th International Workshop on Semantic Evaluation (SemEval 2014). 2014.
13. Guha, Satarupa, Aditya Joshi, and Vasudeva Varma. "SIEL: Aspect Based Sentiment Analysis in Reviews." SemEval-2015 (2015): 759.
14. Gupta, De1epak Kumar, and Asif Ekbal. "IITP: Supervised Machine Learning for Aspect based Sentiment Analysis." SemEval 2014 (2014): 319.

15. Malhotra, Nishtha, et al. "SAP-RI: A Constrained and Supervised Approach for Aspect-Based Sentiment Analysis." SemEval 2014 (2014): 517.

16. Garcıa-Pablos, Aitor, MontseCuadros, and German Rigau. "V3: Unsupervised Aspect Based Sentiment Analysis for SemEval-2015 Task 12."

17. Brychcın, Tomas, Michal Konkol, and Josef Steinberger. "UWB: Machine Learning Approach to Aspect-Based Sentiment Analysis." SemEval 2014 (2014): 817.

18. Guha, Satarupa, Aditya Joshi, and Vasudeva Varma. "SIEL: Aspect Based Sentiment Analysis in Reviews." SemEval-2015 (2015): 759.