International Journal of Scientific Engineering and Technology (ISSN : 2277-1581)
Volume 2 Issue 3, PP : 164-170 1 April 2013

Analysis of different approaches to Sentence-Level Sentiment Classification

V. S. Jagtap, Karishma Pawar

Department of Computer Engineering, Maharashtra Institute of Technology, Pune, India
vandana.jagtap@mitpune.edu.in, kvppawar@gmail.com

Abstract: Sentiment classification is a way to analyze the subjective information in the text and then mine the opinion. Sentiment analysis is the procedure by which information is extracted from the opinions, appraisals and emotions of people in regards to entities, events and their attributes. In decision making, the opinions of others have a significant effect on customers ease, making choices with regards to online shopping, choosing events, products, entities. The approaches of text sentiment analysis typically work at a particular level like phrase, sentence or document level. This paper aims at analyzing a solution for the sentiment classification at a fine-grained level, namely the sentence level in which polarity of the sentence can be given by three categories as positive, negative and neutral.

Keywords - document level, phrase level, polarity, sentence level, sentiment analysis.

1. Introduction

With the rapid development of the Internet, especially the emergence of Web 2.0 technology, more and more people begin to express their views and perspectives about all kinds of things on the Internet, this is because of its rapid growth due to the increasing phenomenon of social network contacts, online Discussion forums, blogs, movie reviews, digitals libraries and quick streaming news stories. Lots of researches focus on finding how to effectively conduct of subjectivity text. Sentiment classification is a new field of NLP that classifies subjectivity text into positive or negative. Sentiment classification could be done in word/phrase level, sentence level and document level.

There are different approaches to classify the document level sentiment which includes the machine learning methods. These machine learning methods include Naïve Bayes, Maximum Entropy classification, and Support Vector Machines (SVM). Various studies on sentiment classification have been conducted and evolved at different levels word), sentence level and document level [2, 3, 4, 5, 6, 7].

Sentiment analysis has now become the dominant approach used for extracting sentiment and appraisals from online sources. Subjectivity analysis focuses on dividing language units into two categories: objective and subjective, whereas sentiment analysis attempts to divide the language units into three categories; negative, positive and neutral.

With the passage of time and a need for better understanding and extraction, momentum slowly increased towards sentiment classification and semantic orientation. In this paper, various methods proposed for sentence-level sentiment classification have been analyzed.

2. Sentiment Classification

Sentiment classification is a new field of Natural Language Processing that classifies subjectivity text into positive or negative.

A. Sentiment Levels

To underline the ambiguity of the concept, Pang and Lee (Pang and Lee, 2008) list the definitions of terms closely linked to the notion of sentiment. Opinion implies a conclusion thought out yet opens to dispute (“each expert seemed to have a different opinion”). View suggests a subjective opinion (“very assertive in stating his views”). Belief implies often deliberate acceptance and intellectual assent (“a firm belief in her party’s platform”). Conviction applies to a party’s firmly and seriously held belief (“the conviction that animal life is as sacred as human”). Persuasion suggests a belief grounded on assurance (as by evidence) of its truth (“was of the persuasion that everything changes”). Sentiment suggests a settled opinion reflective of one’s feelings (“her feminist sentiments are well-known”). Sentiment classification is performed at different levels.

Classifying a document (e.g., a review, blogs) is based on the overall sentiment expressed by opinion holder. It assumes that each document focuses on a single object and contains opinions from a single opinion holder. The main task in document level sentiment classification is to determine the overall sentiment orientation of the document depends on classes which can be Positive, or negative and neutral. The sentence level classification considers each sentence as a separate unit and assumes that sentence should contain only one opinion. Sentence-level sentiment analysis has two tasks as subjectivity classification and sentiment classification.

The goal of feature level classification is to produce a feature-based opinion summary of multiple reviews. It has mainly three tasks. The first task is to identify and extract object features that have been commented on by an opinion holder (e.g. "picture", "battery life"). The second task is to determine the polarity of opinions on features classes: positive, negative and neutral and third task is related to the group feature synonyms.

B. 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 [3] and negation tagging. After this phase, sentiment classification is performed to get the results.

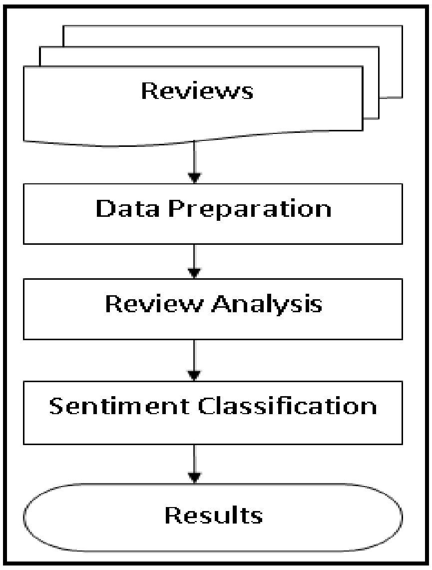

Figure 1 Typical Sentiment Analysis Model

C. General Challenges

- **Contrasts with Standard Fact-Based Textual analysis**

With sentiment classification there are relatively few classes (e.g., "positive" or "3 stars") that generalize across many domains and users. In addition, while the different classes in topic-based categorization can be completely unrelated, the sentiment labels that are widely considered typically represent opposing (if the task is binary classification) or ordinal/numerical categories (if classification is according to a multi-point scale). In fact, the regression-like nature of strength of feeling, degree of positivity, and so on seems rather unique to sentiment categorization (although one could argue that the same phenomenon exists with respect to topic-based relevance).

There are also many characteristics of answers to opinion-oriented questions that differ from those for fact-based questions. As a result, opinion-oriented information extraction, as a way to approach opinion-oriented question answering, naturally differs from traditional information extraction (IE). Interestingly, in a manner that is similar to the situation for the classes in sentiment-based classification, the templates for opinion-oriented IE also often generalize well across different domains, since the interest is in roughly the same set of fields for each opinion expression (e.g., holder, type, strength) regardless of the topic. In contrast, traditional IE templates can differ greatly from one domain to another, the typical template for recording information relevant to a natural disaster is very different from a typical template for storing bibliographic information. These distinctions might make the problems appear deceptively simpler than their counterparts in fact based analysis, but this is far from the truth. Few examples have been sampled to show what makes these problems difficult compared to traditional fact-based text analysis.

The challenge in this topic is to determine whether a document or portion (e.g. paragraph or statement) is subjective. A single keyword can be used to convey three different opinions, positive, neutral and negative respectively. In order to arrive at sensible conclusions, sentiment analysis has to understand context. For example, "fighting" and "disease" is negative in a war context but positive in a medical one.

- **Factors That Make Sentiment Analysis Difficult**

Begin with a sentiment polarity text-classification example. Suppose an opinionated text as either positive or negative, according to the overall sentiment expressed by the author within it has to be classified. Is this a difficult task? To answer this question, first consider the following example, consisting of only one sentence (by Mark Twain):

"Jane Austen's books madden me so that I can't conceal my frenzy from the reader".

Just as the topic of this text segment can be identified by the phrase "Jane Austen", the presence of words like "madden" and "frenzy" suggests negative sentiment. So one might think this is an easy task, and hypothesize that the polarity of opinions can generally be identified by a set of keywords. But, the results of an early study by Pang et al. [1] on movie reviews suggest that coming up with the right set of keywords might be less trivial than one might initially think.

The purpose of Pang et al.'s pilot study was to better understand the difficulty of the document-level sentiment-polarity classification problem. Two human subjects were asked to pick keywords that they would consider to be good indicators of positive and negative sentiment. As shown in table, the use of the subjects' lists of keywords achieves about 60% accuracy when employed within a straightforward classification policy.

In contrast, word lists of the same size but chosen based on examination of the corpus' statistics achieves almost 70% accuracy, even though some of the terms, such as "still", might not look that intuitive at first. The table shows sentiment classification using keyword lists created by human subjects ("Human 1" and "Human 2"), with corresponding results using keywords selected via examination of simple statistics of the test data ("Statistics-based"). The table is adapted from Figures 1 and 2 in Pang et al. [1].

TABLE I. SENTIMENT CLASSIFICATION USING KEYWORD LISTS CREATED BY HUMAN SUBJECTS

	Word lists	Accuracy	Ties
Human 1	positive: dazzling, brilliant, phenomenal, excellent, fantastic negative: suck, terrible, awful, unwatchable, hideous	58%	75%
Human 2	positive: gripping, mesmerizing, riveting, spectacular, cool negative: bad, clichéd, sucks, boring, stupid, slow	64%	39%
Statistics-based	positive: love, wonderful, best, great, superb, still, beautiful Negative: bad, worst, stupid, waste, boring,	69%	16%

D. Applications

- **Applications to Review-related Websites**

Summarizing user reviews is an important problem. In blog analysis, it is used to perform subjectivity and polarity classification on blog posts, discover irregularities in temporal mood patterns (fear, excitement, etc) appearing in a large corpus of blogs, use link polarity information to model trust and influence in the blogosphere, analyze Blog sentiments about movies and correlate it with its sales [8].

- **Applications as a Sub-Component technology**

Sentiment-analysis and opinion-mining systems also have an important potential role as enabling technologies for other systems. In online systems that display ads as sidebars, it is helpful to detect web pages that contain sensitive content inappropriate for ads placement; for more sophisticated systems, it could be useful to bring up product ads when relevant positive sentiments are detected, and perhaps more importantly, mix the ads when relevant negative statements are discovered [9].

- **Applications in Business and Government Intelligence**

The field of opinion mining and sentiment analysis is well-suited to various types of intelligence applications [10]. Indeed, business intelligence seems to be one of the main factors behind corporate interest in the field. Sentiment classification can be applied for marketing intelligence, product and service benchmarking and improvement, to understand the voice of the customer as expressed in everyday communications. Government intelligence is another application that has been considered. For example, it has been suggested that one could monitor sources for increases in hostile or negative communications [11].

- **Applications across different Domains**

Sentiment analysis has specifically been proposed as a key enabling technology in eRulemaking, allowing the automatic analysis of the opinions that people submit about pending policy or government-regulation proposals [12]. On a related note, there has been investigation into opinion mining in weblogs devoted to legal matters, sometimes known as "blawgs". Interactions with sociology promise to be extremely fruitful [13].

3. Approaches to Sentiment Classification

In this section, approaches for sharing the common theme of mapping a given piece of text, such as a document,

International Journal of Scientific Engineering and Technology (ISSN : 2277-1581)
Volume 2 Issue 3, PP : 164-170 1 April 2013

paragraph, or sentence, to a label drawn from a pre-specified finite set or to a real number has been discussed.

The **machine learning approach** involves text classification techniques. This approach treats the sentiment classification problem as a topic-based text classification problem (Liu, 2007). Any text classification algorithm can be employed, e.g., naïve Bayes, SVM, etc. This approach was put forth by Pang et al. (2002) to classify movie reviews into two classes: positive and negative. The study compared naïve Bayes, Maximum Entropy, and SVM. A test bed of 700 positive reviews and 700 negative reviews was used. The highest classification accuracy (82.9%) was achieved using SVM with 3-fold cross validation.

To implement these machine learning algorithms, the following standard bag-of-features framework was used [1]. Let $\{f_1,\dots\dots,f_m\}$ be a predefined set of m *features* that can appear in a document; examples include the word “still" or the bigram “really stinks". Let $n_i(d)$ be the number of times f_i occurs in document d. Then, each document d is represented by the document vector

$$\overrightarrow{d} = (n_1(d), n_2(d), \dots\dots\dots\dots n_m(d))$$

Naïve Bayes

One approach to text classfication is to assign to a given document d the following class which is given as $c* = \arg\max c\ P(c|d)$

We derive the Naïve Bayes (NB) classifier by first observing that by Bayes' rule,

$$P(c|d) = \frac{P(c)\,P(d \mid c)}{P(d)}$$

where $P(d)$ plays no role in selecting c*. To estimate the term $P(d \mid c)$, Naive Bayes decomposes it by as suming the *fi*'s are conditionally independent given *d*'s class:

$$P_{NB}(c\,|d) := \frac{P(c)\sum_{i=1}^{m} P(f_i \mid c)^{ni(d)})}{P(d)}$$

The training method consists of relative-frequency estimation of $P(c)$ and P(*fi* | *c*), using add-one smoothing. Despite its simplicity and the fact that its conditional independence assumption clearly does not hold in real-world situations, Naive Bayes-based text categorization still tends to perform surprisingly well (Lewis, 1998); indeed, Domingos and Pazzani (1997) show that Naive Bayes is optimal for certain problem classes with highly dependent features. On the other hand, more sophisticated algorithms might yield better results; we examine two such algorithms next.

Maximum Entropy

Maximum entropy classification (MaxEnt, or ME, for short) is an alternative technique which has proven effective in a number of natural language processing applications (Berger et al., 1996). Nigam et al. (1999) show that it sometimes, but not always, outperforms Naive Bayes at standard text classification. Its estimate of *P*(*c* | *d*) takes the following exponential form:

$$P_{ME}(c\,|d) := \frac{1}{Z(d)} \exp\left(\sum_i \lambda_{i,c} F_{i,c}(d,c)\right)$$

where $Z(d)$ is a normalization function. $F_{i,c}$ is a feature/class function for feature *fi* and class c, defined as follows:

$$F_{i,c}(d,c') = \begin{cases} 1, & n_i(d) > 0 \text{ and } c' = c \\ 0, & \text{otherwise} \end{cases}$$

For instance, a particular feature/class function might fire if and only if the bigram “still hate" appears and the document's sentiment is hypothesized to be negative. Importantly, unlike Naive Bayes, MaxEnt makes no assumptions about the relationships between features, and so might potentially perform better when conditional independence assumptions are not met.

The $\lambda_{i,c}$'s are feature-weight parameters; inspection of the definition of PME shows that a large $\lambda_{i,c}$ means that *fi* is considered a strong indicator for class c. The parameter values are set so as to maximize the entropy of the induced distribution (hence the classifier's name) subject to the constraint that the expected values of the feature/class functions with respect to the model are equal to their expected values with respect to the training data: the underlying philosophy is that we should choose the model making the fewest assumptions about the data while still remaining consistent with it, which makes intuitive sense. We use ten iterations of the improved iterative scaling algorithm (Della Pietra et al., 1997) for parameter training (this was a sufficient number of iterations for convergence of training-data accuracy), together with a Gaussian prior to prevent overfitting (Chen and Rosenfeld, 2000).

Support Vector Machines

Support vector machines (SVMs) have been shown to be highly effective at traditional text categorization, generally outperforming Naive Bayes (Joachims, 1998). They are large-margin, rather than probabilistic, classifiers, in contrast to Naive Bayes and MaxEnt. In the two-category case, the basic idea behind the training procedure is to find a hyperplane, represented by vector $\vec{w}$, that not only separates the document vectors in one class from those in the other, but for which the separation, or margin, is as large as possible. This search corresponds to a constrained optimization problem; letting $c_j \in \{1, -1\}$(corresponding to positive and negative) be the correct class of document dj , the solution can be written as

$$\vec{w} := \sum_j \alpha_j c_j \vec{d}_j, \qquad \alpha_j \geq 0$$

Where, the α_j 's are obtained by solving a dual optimization problem. Those $\vec{d}_j$ such that α_j is greater than zero are called support vectors, since they are the only document vectors contributing to $\vec{w}$. Classification of test instances consists simply of determining which side of $\vec{w}$'s hyperplane they fall on. Joachim's (1999) SVMlight package8 for training and testing, with all parameters set to their default values has been used.

Semantic Orientation Approach

The semantic orientation approach performs classification based on positive and negative sentiment words and phrases contained in each evaluation text (Liu, 2007). 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.

- The corpus-based techniques:

Corpus-based techniques try to find co-occurrence patterns of words to determine their sentiments. Turney (2002) 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 (2003) used a bootstrapping process to learn linguistically rich patterns of subjective expressions in order to classify subjective expressions from objective expressions.

- The dictionary-based techniques :

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

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 [1] 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 [13, 14].

4. Analysis of Various Methods

The work in this area started around 2000 and is still strong today. As mentioned earlier, a lot of work has been done on movie and product reviews, especially popular are the Internet Movie Database (IMDb) and product reviews downloaded from Amazon. The performance achieved by various methods is difficult to judge, since each method uses a variety of resources for training and different collections of documents for testing.

Most existing techniques for document level sentiment classification are based on supervised learning, for example, n-gram features and three machine learning methods (Naïve Bayes, Maximum Entropy classification, and SVM) can be applied to achieve the high performance [1]. The results produced via machine learning techniques are quite good in comparison to the human generated baselines as discussed in section 2. Naïve Bayes tends to do the worst and SVMs tend to do the best, although the differences are not very large.

A kernel-based approach for classifying sentiment for a sentence incorporates multiple features from lexical and syntactic levels [15]. It outperforms the n-gram based method.

The Method called syntax tree pruning and tree kernel based approach to sentiment classification to sentence-level sentiment classification can be applied [6]. Tree kernel-based method is not as good as basic kernel-based method since the tree kernel can't capture some information that is important for the sentiment classification.

The Method for domain adaptation for sentiment classifiers (ML based) to sentiment classification outperforms the supervised baseline for sentiment classification [16].

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. Finally, although relatively high accuracy in document polarity labeling has been achieved, it is still a challenge to extract the full private state, complete with the emotion's intensity, its holder, and its target.

5. Conclusion

Sentiment classification is the foundation of sentiment analysis. Sentiment analysis is the process of extracting knowledge from the peoples' opinions, appraisals and emotions toward entities, events and their attributes. In this paper, various methods for sentiment classification have been discussed and their analysis is done. Internet provides us with an unlimited source of the most diverse and opinionated text, and as of yet only a small part of the existing domains have been explored. Much work has been done on product reviews – short documents that have a well-defined topic. More general writing, such as blog posts and web pages, have recently been receiving more attention. Still, the field is struggling with more complex texts like sophisticated political discussions and formal writings.

6. Future work

An important next step is the identification of features indicating whether sentences are on-topic (which is a kind of co-reference problem); so this can be the challenge for future work which can be done by integrating various techniques from machine learning. Data classification methods and machine learning methods can be combined to improve the accuracy. Future work in expanding existing techniques to handle more linguistic and semantic patterns will surely be an attractive opportunity for researchers and business people alike.

REFERENCES

I. B. Pang, L. Lee, and S. Vaithyanathan, "Thumbs up? Sentiment classification using machine learning techniques", In Proceedings of the Conference on Empirical Methods in Natural Language Processing, Philadelphia, Pennsylvania, USA, 2002, pp. 79-86.

II. T. Wilson, J. Wiebe, and P. Hoffmann, "Recognizing Contextual Polarity: An Exploration of Features for Phrase-Level Sentiment Analysis," Computational Linguistics, vol. 35, no. 3, pp. 399-433, 2009.

III. A. Agarwal, F. Biadsy and K. Mckeown, "Contextual Phrase-Level Polarity Analysis using Lexical Affect Scoring and Syntactic Ngrams,", In Proceedings of ECACL 2009, pp. 24–32, 2009.

IV. S. M. Kim and E. Hovy, "Determining the Sentiment of Opinions", In Proceedings of the 20th International Conference on Computational Linguistics (COLING 2004), Geneva, Switzerland, 2004, pp. 1367-1373.

V. B. Li, L. Zhou, S. Feng and K. Wong, "A Unified Graph Model for Sentence-based Opinion Retrieval," In Proceedings of ACL 2010, pp. 1367–1375, 2010.

VI. 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.

VII. P. Turney, "Thumbs Up or Thumbs Down? Semantic Orientation Applied to Unsupervised Classification of Reviews," In Proceeding of ACL 2002, pp. 417-424, 2002.

VIII. Lu´ıs Cabral and Ali Hortac¸su, "The dynamics of seller reputation: Theory and evidence from eBay" Working paper, downloaded version revised in March, 2006. URL http://pages.stern.nyu.edu

IX. Xin Jin, Ying Li, Teresa Mah and Jie Tong, "Sensitive webpage classification for content advertising", In Proceedings of the International Workshop on Data Mining and Audience Intelligence for Advertising, 2007.

X. Lillian Lee. ”I’m sorry Dave, I’m afraid I can’t do that”: Linguistics, statistics, and natural language processing circa 2001. In Committee on the Fundamentals of Computer Science: Challenges, Computer Science Opportunities, and National Research Council Telecommunications Board, editors, Computer Science: Reflections on the Field, Reflections from the Field, pages 111–118. The National Academies Press, 2004

XI. Ahmed Abbasi. “Affect intensity analysis of dark web forums”, In Proceedings of Intelligence and Security Informatics (ISI), pages 282–288, 2007.

XII. Claire Cardie, Cynthia Farina, Thomas Bruce and Erica Wagner, “Using natural language processing to improve eRulemaking”, In Proceedings of Digital Government Research, 2006.

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

XIV. 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, pages 486–497, 2005.

XV. Tianfang Yao and Linlin Li, “Kernel-based Sentiment Classification Approach for Chinese Sentences” In proceedings of World Congress on Computer Science and Information Engineering, 2009

XVI. J. Blitzer, M. Dredze, and F. Pereira. Biographies, “Bollywood, Boom-boxes and Blenders: Domain Adaptation for Sentiment Classification” In ACL 2007, 2007