

Opinion Mining and Sentiment Analysis: A Survey

Mohammad Sadegh
Hajmohammadi
Faculty of Computer Science
and
Information Systems,
Universiti Teknologi Malaysia,
Malaysia

Roliana Ibrahim
Faculty of Computer Science
and
Information Systems,
Universiti Teknologi Malaysia,
Malaysia

Zulaiha Ali Othman
Centre of Artificial Intelligent
Technology (CAIT)
Faculty of Technology and
Information Science
Universiti Kebangsaan Malay-
sia

ABSTRACT

In the past few years, a great attention has been received by web documents as a new source of individual opinions and experience. This situation is producing increasing interest in methods for automatically extracting and analyzing individual opinion from web documents such as customer reviews, weblogs and comments on news. This increase was due to the easy accessibility of documents on the web, as well as the fact that all these were already machine-readable on gaining. At the same time, Machine Learning methods in Natural Language Processing (NLP) and Information Retrieval were considerably increased development of practical methods, making these widely available corpora. Recently, many researchers have focused on this area. They are trying to fetch opinion information and analyze it automatically with computers. This new research domain is usually called Opinion Mining and Sentiment Analysis. . Until now, researchers have developed several techniques to the solution of the problem. This paper try to cover some techniques and approaches that be used in this area.

General Terms

Text mining, Opinion mining, Sentiment analysis.

Keywords

Opinion mining, Sentiment analysis, Sentiment classification

1. INTRODUCTION

Human life consists of emotions and opinions; we cannot imagine the world without them. Emotions and opinions manage how humans communicate with each other and how they motivate their actions. Emotions and opinions play a role in nearly all human actions. Emotions and opinions influence the way humans think, what they do, and how they act.

In the past few years, a great attention has been received by web documents as a new source of individual opinions and experience. This situation is producing increasing interest in methods for automatically extracting and analyzing individual opinion from web documents such as customer reviews, weblogs and comments on news.

This increase was due to the easy accessibility of documents on the web, as well as the fact that all these were already machine-readable on gaining. At the same time, Machine Learning methods in Natural Language Processing (NLP) and Information Retrieval were considerably increased development of practical methods, making these widely available corpora.

Recently, many researchers have focused on this area. They are trying to fetch opinion information and analyze it automatically with computers. As we know, there are large amounts of information created by users on the Internet, including product re-

views, movie reviews, forum entries, blog and so on. How to analyze and summarize the opinions expressed in these documents is a very interesting domain for researchers. This new research domain is usually called Opinion Mining and Sentiment Analysis. Until now, researchers have developed several techniques to the solution of the problem. Current-day opinion mining and sentiment analysis is a field of study at the crossroad of information retrieval and natural language processing and shares some characteristics with other disciplines such as text mining and information extraction.

This paper try to cover some techniques and approaches that be used in this area. At the first, the definition of some terms in this field are introduced and then several problems that related to sentiment analysis and some related works that try to solve these problems are presented.

2. OPINION DEFINITION

When we search a dictionary for opinion, we can find following definition:

1. A view or judgment formed about something, not necessarily based on fact or knowledge.
2. The beliefs or views of a large number or majority of people about a particular thing.

In general, opinion refers to what a person thinks about something. In other words, opinion is a subjective belief, and is the result of emotion or interpretation of facts.

3. OPINION MINING AND SENTIMENT ANALYSIS

Opinion mining and sentiment analysis is a technique to detect and extract subjective information in text documents. In general, sentiment analysis tries to determine the sentiment of a writer about some aspect or the overall contextual polarity of a document. The sentiment may be his or her judgment, mood or evaluation. A key problem in this area is sentiment classification, where a document is labeled as a positive or negative evaluation of a target object (film, book, product, etc.).

4. DOCUMENT-LEVEL SENTIMENT CLASSIFICATION

The binary classification task of labeling a document as expressing either an overall positive or negative opinion is called document-level sentiment classification. Document-level sentiment classification assumes that the opinionated document expresses opinions on a single target and the opinions belong to a single person. It is clear that this assumption is true for customer review of products documents which usually focus on one product and single reviewer writes it. A movie review, restaurant review, or product review consists of a document written by the review-

er, explaining what he/she felt was principally positive or negative about the product. The task of document-level sentiment classification is to predict whether reviewer wrote a positive or negative review, based on an analysis of the text of the review.

Two type of classification techniques have been used in document-level sentiment classification, supervised method and unsupervised method. Based on the pro-posed taxonomy, Table 1 shows selected previous studies dealing with document-level sentiment classification. Some of these related studies were introduced in detail next.

Table 1. Selected previous studies in document-level sentiment classification

Paper	Technique	Features	Dataset	Type
[1]	SVM, Naive Bayes	Unigram, Bi-grams, Trigrams	Restaurant Review	Supervised
[2]	SVM, Naive Bayes, Maximum Entropy	Unigrams, bi-grams, adjective, position of words	Movie Review	Supervised
[3]	SVM, Naive Bayes, character based N-gram model	Unigram Frequency	reviews to travel destination	Supervised
[4]	SVM, Rule-based Classifier	POS tag, N-grams	movie reviews, product reviews, MySpace comments	Supervised Learning and Rule-based classification
[5]	SVM	Unigram, bigram and extraction pattern feature	Movie Review, MPQA dataset	Supervised
[6]	SVM	Adjective word frequency, percentage of appraisal groups	Movie Review	Supervised
[7]	PMI-IR	adjectives and adverbs	Automobile, bank, movie, travel reviews	Unsupervised
[8]	Association Rule	adjectives and adverbs	Movie Review	Unsupervised
[9]	Dictionary based approach	Adjectives, Nouns, verbs, Adverbs, Intensifier, Negation	Movie Review, Camera Review, Epinions	Unsupervised

4.1 Supervised Methods

Sentiment classification task can be formulated as a supervised learning problem with two classes, positive and negative. Product reviews mostly are used as training and testing data. Any existing supervised learning techniques can be used to sentiment classification, such as naïve Bayes and support vector machines (SVM). In most cases, SVMs have shown marginal improvement over Naïve Bayes classifiers. Cui et al. [10] have argued that discriminative classifiers such as SVMs are more appropriate for sentiment classification than generative models because they can better differentiate mixed sentiments (i.e., both positive and negative words are used in the same review). However, when the set of training data is small, a Naïve Bayes classifier

might be more appropriate since SVMs must be exposed to a large set of data in order to build a high-quality classifier. Several techniques and features are used by researchers in learning process. One of the most important tasks in sentiment classification is selecting an appropriate set of features. The most commonly used features in sentiment classification are introduced below.

1. Terms and their frequency: these features consist of single words or word n-grams and their frequency or presence. These features have been widely and successfully used in sentiment classification and shown quite effective for this task.
2. Part of speech information: POS information is a very important indicator of sentiment expression. For example adjectives carry a great deal of information regarding a document’s sentiment [2].
3. Opinion words: Opinion words (or sentiment words) and phrases are words and phrases that express positive or negative emotions. For instance, good, fantastic, amazing and brilliant are words with positive emotion and bad, boring, slow, worst and poor are words with negative emotion. Though almost opinion words are adjectives and adverb, nouns and verbs can also express an opinion. For example rubbish (noun), hate and like (verb) can indicate opinion in some documents.
4. Negations: Obviously, negation words are very important to evaluate the polarity of a sentence because they can transform the sentiment orientation in a sentence. For instance, the sentence “I don’t like this mobile” has negative orientation.
5. Syntactic dependency: Several research work in this area used word dependency based features generated from dependency tree or parsing.

One of the earliest works which used supervised method to solve sentiment classification problem is [2]. In this paper, authors used three machine learning techniques to classify sentiment of movie review documents. To implement these machine learning techniques on movie review documents, they used the standard bag of features frame work. They test several features to find optimal feature set. Unigrams, bigrams, adjective and position of words were used as features in these techniques. To reduce the number of features, they used only unigrams appearing at least four times in all document corpus and bigrams occurring at least seven times. The results show that the best performance is achieved when the unigrams are used in SVM classifier. As they show in this paper, better performance is reached by using only presence of feature instead of feature frequency.

In [6] authors augmented bag-of-words classification with a technique which performed shallow parsing to find opinion phrases, classified by orientation and by a taxonomy of attitude types from appraisal theory [11], specified by a hand-constructed attitude lexicon. Text classification was performed using a support vector machine, and the feature vector for each corpus included word frequencies (for the bag-of-words), and the percentage of appraisal groups that was classified at each location in the attitude taxonomy, with particular orientations. They achieved 90.2% accuracy classifying the movie reviews in Pang et al.’s [2] corpus.

Ye et al. [3] incorporated sentiment classification techniques into the domain of destination review. They used three supervised learning algorithms of support vector machine, NB and the character based N-gram model to classify destination reviews. The information gain (IG) was used to selecting feature set. They used the frequency of words to represent a document instead of word presence. They found SVM outperforms the other two classifiers with an accuracy peak at about 86% when the training corpora contain 700 reviews.

Prabowo and Thelwall [4] took a combined approach to sentiment analysis with a hybrid classifier, applying different classifiers in series, until acceptable results are obtained. If this cannot be achieved with one classifier, the system passes the task on to the next in line, until no more classifiers exist. For this, they use a combination of rule-based classification, supervised learning, and machine learning. For rule-based classification and supervised learning, authors use three different rules from existing research. Two of the used rule sets were also combined with two preexisting induction algorithms, ID3 and RIPPER, to generate two induced rule sets, which were also tested. For machine learning-based classification, they used a Support Vector Machine using two pre-classified training sets, positive and negative, and have the SVM create a hyper plane to best separate the two planes. The hybrid classifier was tested on a combination of movie reviews, product reviews and MySpace comments, and yielded anywhere from 72.77% F1 score to 90% F1 score, depending on the corpus.

The biggest limitation associated with supervised learning is that it is sensitive to the quantity and quality of the training data and may fail when training data are biased or insufficient. Sentiment classification at the sub-document level raises additional challenges for supervised learning based approaches because there is little information for the classifier.

4.2 Unsupervised Methods

Obviously, sentiment words and phrases are the main indicators of sentiment classification. Therefore several works have been done by using unsupervised learning methods based on such words and phrases.

Turney [7] presented a simple unsupervised learning algorithm for classifying a review as recommended or not recommended. He determined whether words are positive or negative and how strong the evaluation is by computing the words' pointwise mutual information (PMI) for their co-occurrence with a positive seed word ("excellent") and a negative seed word ("poor"). He called this value the word's semantic orientation. This method scanned through a review looking for phrases that match certain part of speech patterns (adjectives and adverbs), computed the semantic orientation of those phrases, and added up the semantic orientation of all of those phrases to compute the orientation of a review. He achieved 74% accuracy classifying a corpus of product reviews.

Harb et al. [8] performed blog classification by starting with the 2 sets of seed words with positive and negative semantic orientations respectively, as in [7] and used Google's search engine to create association rules that find more. They then counted the numbers of positive versus negative adjectives in a document to classify the documents. They achieved 0.717 F1 score identifying positive documents and 0.622 F1 score identifying negative documents. However, these approaches rely on only the labeled

seed words ("excellent", "poor") to construct a domain-oriented sentiment lexicon, and cannot to discover the mutual relationship between the words and the documents.

A lexicon-based method to sentiment classification was presented by Taboada et al. [9]. They used dictionaries of positive or negative polarized words to this classification task. A semantic orientation calculator (SO-CAL) was build based on these dictionaries by incorporating intensifiers and negation words. This lexicon-based approach has been shown to have 59.6% to 76.4% accuracy on 1900 documents of movie review dataset.

In brief, the main advantage of sentiment classification in document-level is that it provides predominant opinion on a topic, entity or event. The main weakness is that it does not provide details about people's interests and of course it is not easily applicable to non-reviews, such as blog and forum postings, because these posts evaluate and compare multiple entities.

5. SENTENCE-LEVEL SENTIMENT CLASSIFICATION

The first step to sentiment classification in sentence level is classifying a sentence as objective or subjective. This task is called subjectivity classification in literature. After this step, subjective sentences are classified as positive or negative orientation. This classification is called sentence-level sentiment classification. One of the most important issues that must be considered in this classification is that which target or aspect has been mentioned in the sentence. Actually, without knowing target of a sentence, the polarity detected for the sentence cannot be useful.

McDonald et al. [12] developed a model for sentiment analysis at different levels of granularity simultaneously. They use graphical models in which a document level sentiment is linked to several paragraph level sentiments, and each paragraph level sentiment is linked to several sentence level sentiments (in addition to being linked sequentially). They apply the Viterbi algorithm to infer the sentiment of each text unit, constrained to ensure that the paragraph and document parts of the labels are always the same where they represent the same paragraph/document. They report 62.6% accuracy at classifying sentences when the orientation of the document is not given, and 82.8% accuracy at categorizing documents. When the orientation of the document is given, they report 70.2% accuracy at categorizing the sentences.

In [13], authors developed a conditional random field model structured like the dependency parse tree of the sentence they are classifying to determine the polarity of sentences, taking into account opinionated words and polarity shifters in the sentence. They report 77% to 86% accuracy at categorizing sentences, depending on which corpus they tested against.

Neviarouskaya et al. [14] developed a system for computing the sentiment of a sentence based on the words in the sentence, using Martin and White's [11] appraisal theory and Izard's [15] affect categories. They used a complicated set of rules for composing attitudes found in different places in a sentence to come up with an overall label for the sentence. They achieved 62.1% accuracy at determining the fine-grained attitude types of each sentence in their corpus, and 87.9% accuracy at categorizing sentences as positive, negative, or neutral.

6. SENTIMENT LEXICON CONSTRUCTION

Sentiment words are used in many sentiment classification tasks. These words are also identified by “opinion words” or “opinion bearing words” in literature. Sentiment words are always divided into two categories according their orientation: positive or negative sentiment words. For instance, “excellent” is a positive sentiment words and “poor” is a negative sentiment word. In addition to the single words, there are several sentiment phrases that can be used in sentiment classification tasks. Sentiment words and sentiment phrases form the sentiment lexicon.

There are three methods to construct a sentiment lexicon: manually construction, corpus-based methods and dictionary-based methods. The manual construction of sentiment lexicon is a very hard and time-consuming task and always cannot be used alone but it can be combined with other methods to improve the accuracy of these methods. Two other methods are discussed in following subsections.

6.1 Corpus-based Methods

These methods always use a seed set of sentiment words with known polarity and exploit syntactic patterns or co-occurrence patterns to identify new sentiment words and their polarity in a large corpus. The work of Hatzivassiloglou and McKeown [16] has been the first to deal with the problem of determining the orientation of words. In this work, authors developed a graph-based technique for learning lexicons by reading a corpus. In their technique, they find pairs of adjectives conjoined by conjunctions (e.g. "simple and well-received" or "fair but brutal"), as well as morphologically related adjectives (e.g. "thoughtful" and "thoughtless"), and create a graph where the vertices represent words, and the edges represent pairs (marked as same-orientation or opposite-orientation links). They apply a graph clustering algorithm to cluster the adjectives found into two clusters of positive and negative words. This technique achieved 82% accuracy at classifying the words found.

Another algorithm for constructing lexicons is that of Turney and Littman [17]. They determine whether words are positive or negative and how strong the evaluation is by computing the words' pointwise mutual information (PMI) for their co-occurrence with small set of positive seed words and a small set of negative seed words. Unlike their earlier work [7], which mentioned in Section 4.2, the seed sets contained seven representative positive and negative words each, instead of just one each. This technique had 78% accuracy classifying words in [16] word list.

Corpus-based methods can produce lists of positive and negative words with relatively high accuracy. Most of these methods need very large labeled training data to achieve their full capabilities. Dictionary-based approaches can overcome some of the limitations of corpus-based approaches by using existing lexicographical resources (such as WordNet) as a main source of semantic information about individual words and senses.

6.2 Dictionary-based methods

Dictionary-based methods to sentiment lexicon construction do not require large corpora or search engines with special capabilities. Instead, they exploit available lexicographical resources like WordNet. Accurate, domain-independent and comprehensive lists of words and their senses can be produced by these

methods. The main strategy in these methods is to collect an initial seed set of sentimental words and their orientation manually, and then searching in a dictionary to find their synonyms and antonyms to expand this set. The new seed set are used iteratively to generate new sentiment words.

Esuli and Sebastiani [18] developed a technique for classifying words as positive or negative, by starting with a seed set of positive and negative words, then running WordNet synset expansion multiple times, and training a classifier on the expanded sets of positive and negative words. They found [19] that different amounts of WordNet expansion, and different learning methods had different properties of precision and recall at identifying opinionated words. Based on this observation, they applied a committee of eight classifiers trained by this method (with different parameters and different machine learning algorithms) to create SentiWordNet which assigns each WordNet synset a score for how positive the synset is, how negative the synset is, and how objective the synset is. The scores are graded in intervals of 1-8, based on the binary results of each classifier, and for a given synset, all three scores sum to 1. This version of SentiWordNet was released as SentiWordNet 1.0. Baccianella, Esuli, and Sebastiani [20] improved upon SentiWordNet 1.0, by updating it to use WordNet 3.0 and the Princeton Annotated Gloss Corpus and by applying a random graph walk procedure so related synsets would have related opinion tags. They released this version of SentiWordNet as SentiWordNet 3.0.

In [21] authors used a system that deals with word orientation detection by assigning a positivity score and a negativity score to each word. interestingly, words may be supposed to have both a positive and a negative association, maybe with different degrees, and some words may be deemed to carry a stronger positive (or negative) orientation than others. Their system starts from a set of positive and negative seed words, and expands the positive (negative) seed set by adding to it the synonyms of positive (negative) seed words and the antonyms of negative (positive) seed words. The system classifies a target word w into either Positive or Negative by means of two alternative learning-free methods based on the probabilities that synonyms of w also appear in the respective expanded seed sets. A problem with this method is that it can classify only words that share some synonyms with the expanded seed sets. A similar approach was used by Hu and Liu [22], who used synonymy relations to extract opinion-words from WordNet.

Then main problem of dictionary-based methods is that this methods unable to find sentiment word with domain specific orientation. A sentiment word maybe expresses positive emotion in one domain and negative emotion in another domain. For example, the word “large” has a positive orientation when it is being used for describing a computer screen and it has a negative orientation if it describes a mobile phone.

7. ASPECT-BASED SENTIMENT ANALYSIS

With the continuously increasing volume of e-commerce transactions, the amount of product information and the number of product reviews are increasing on the Web. Many costumers feel that they can make more efficient decisions based on the experiences of others that expressed in product reviews on the web[23]. Therefore, product reviews are very important resource to decision making for selecting a product by a costumer.

However, because the number of reviews increases day by day in review websites, reading all of the relevant review documents is very difficult for users. In order to solve this problem, several approaches have been proposed by researchers to summarize product evaluations from reviews. The summary information that is shown to the users is very important. For example, a detailed summary that consists of evaluations for all product features, such as size and cost, may be more useful than an overall summary that displays an average score for all of the product features.

To produce this detailed summary of product reviews that called Aspect-based sentiment analysis [24], several tasks need to be performed. Two core tasks are explained in the following sub sections, aspect extraction and aspect sentiment orientation detection.

7.1 Aspect extraction

Aspect extraction is one of the most complex tasks in aspect-based sentiment analysis (also known as topic, feature, or target extraction). It needs the use of Natural language processing techniques in order to automatically extract the aspects (features) in the opinionated documents. Some techniques for aspect extraction have been shown in **Error! Reference source not found.**

Table 2. Aspect extraction techniques

Paper	Technique	Type
[25]	Association Mining and Web PMI	Unsupervised
[22]	Association Mining	Unsupervised
[26]	Likelihood Ratio	Unsupervised
[27]	CRF approach	Supervised
[28]	SVM	Supervised
[29]	HMMs	Supervised
[30]	Averaged perception	Supervised
[31]	Double propagation-syntactic relation	Semi-supervised

One of the earliest work to extraction of aspects in costumer reviews was done by Hu and Liu [22]. They used association rule mining combined with pruning strategies to find the candidates of features which frequently used in product reviews. They assumed that the product features are nouns or noun phrases. They first performed Part-of-Speech (POS) parsing. A tag of POS was then given to each word. A transaction was built from the noun words of each sentence. All transactions were fed into the association rule mining algorithm to find the frequent item sets. In their context, an item set is simply a set of words or phrases that occurs together in some sentences. The returned frequent item sets were used to identify product features. Product features was divided into two groups: frequent and infrequent features, dependent on their Frequency. Infrequent feature words are extracted by extracting known opinion words' adjacent noun phrases.

The precision of above algorithm was improved in[25]. In this paper, the authors try to remove those noun phrases that may not be product features. It evaluates each noun phrase by computing a pointwise mutual information (PMI) score between the phrase and some *meronymy discriminators* associated with the product

class. For example, the meronymy discriminators for the scanner class are, "of scanner", "scanner has", "scanner comes with", etc., which are used to find components or parts of scanners by searching the Web. If the PMI value of a candidate aspect is too low, it may not be a component of the product because candidate aspect and discriminator do not co-occur frequently. Querying the web is a main problem for this method.

Another unsupervised aspect extraction technique was introduced in the work of Yi et al. [26]. They introduced a complete system for opinion extraction and developed and tested two feature term selection algorithms based on a mixture language model and likelihood ratio.

Dave et al. [32] examines classification of product reviews from C|net. The studied corpus consists of 10 randomly selected sets of 56 positive and 56 negative reviews from 4 largest categories of C|net (in total, 448 reviews). A review is annotated as *positive* if it is rated in C|net with three or more stars, and as *negative*, otherwise. Before aspect extraction, reviews' texts are preprocessed as follows:

Unique words are substituted with the string `_unique`, product names are substituted with the string `_productname`, and product specific words are substituted with the string `_producttypeword`;

Ambiguous words are disambiguated using POS tags and substituted with their distinct similarities using WordNet.

Negations are identified by the words `not`, and `never`. Negation phrases are substituted with artificial terms resulting from the combination of the corresponding negation and the following word. For instance, the phrase `not good` becomes the `NOT_good` string.

After making the above changes, the approach extracts N-grams (unigrams, bigrams and trigrams) that are evaluated as frequency vectors. The SVM classifier is used for classification and yields the accuracy value of 85.8% using ten-fold cross-validation without stratification.

Kessler and Nicolov [28] exclusively focus on identifying which opinion expression is related to which aspect in a sentence of a product review. They introduced a dataset of car and camera reviews in which opinion expressions and target aspects are annotated. They train a machine learning classifier (SVM) to finding related opinion expression and target aspect. The objective was to learn a model that ranks aspects occurring in the same sentence as an opinion expression such that the ones which are highest ranked are likely to be targeted. The feature vectors were formed based on the syntactic and semantic relationship between the opinion expression and candidate aspect. This classifier was compared to the algorithm presented by Bloom et al. (2007) and showed better performance based on F-measure.

Jin et al.[29] introduced a machine learning approach build under the framework of lexicalized Hidden Markov Model (L-HMMs). This approach which they called it "OpinionMiner", combine multiple significant linguistic features (e.g. part of speech, phrases internal information patterns, surrounding contextual clues) into an automatic learning process. To labeling training data, authors designed a bootstrapping approach which could extract high confidence labeled data through self-learning.

In [27], authors modeled the problem of aspect extraction as an information extraction task and used a Conditional Random

Fields (CRF)-based approach for opinion target extraction. They used several features as input for their CRF-based approach such as POS tags, short dependency path, word distance and opinion sentence. They employ datasets from three different sources to evaluate this method and show that CRF-based method for opinion target extraction can be used effectively in single-domain and cross-domain setting.

Stoyanov and Cardie [30] treated aspect extraction task as a topic co reference resolution problem. Their method tried to cluster opinions sharing the same target together. They proposed to train a classifier to judge if two opinions are on the same target, which indicates that their approach is supervised.

Yi et al.[26] introduced an approach to opinion mining that classifies subjective phrases as positive/negative in topic or non-topic documents. The positive/negative orientation of documents was evaluated similarly to the naïve algorithm. To assess the sentiment orientation of text pieces, the approach detected phrases that match the pattern <predicate>-<sentiment_category>-<target> where <predicate> is a verb, <sentiment_category> is a relation between the source and the target of the emotional phrase (either positive, or negative, or opposite), <target> refers to the target of an emotional phrase. 120 patterns of the proposed form were collected automatically and adjusted manually. For instance, using the proposed pattern the approach extracts the phrase <impress>-<positive>-<by; with object> that occurs in the sentence I'm impressed by the picture quality describing a camera. Emotion words necessary for opinion mining, e.g. impress, are extracted from GI, DAL, or WordNet. The approach by Yi and colleagues was evaluated in experiments on reviews from the digital camera domain, using 485 manually annotated as topic documents and 1,838 annotated as non-topic documents. The documents are collected on the Internet. A review is considered as recommended if the number of positive patterns in review's text pieces exceeds the number of negative patterns and as not recommended, otherwise. This classification algorithm yields the recall value of 56% and the precision value of 87%.

7.2 Aspect sentiment orientation detection

Determining the sentiment orientation expressed on each aspect in a sentence is the second task in aspect based sentiment analysis. It must determine whether the sentiment orientation on each aspect is positive, negative or neutral. This task can be divided into the following sub tasks:

1. Extracting opinion words or phrases.
2. Identifying the polarity of each opinion words or phrases.
3. Handling opinion shifters (such as no, not, don't) and opinion intensifiers (such as very, extremely)
4. Handling but clauses.
5. Aggregating opinions (if there is more than one opinion word or phrase in a sentence).

In [22] a distance based approach was used to extract opinion words and phrases after extracting aspects. In this paper, adjacent adjective words (e.g. within the 3-words distance to the aspect) were considered as opinion words. Authors used a WordNet lexicon to calculate the polarity of each extracted opinion word. The negation words were considered in this paper

but intensifiers were not extracted. For a sentence that contains a *but* clause which implies sentimental change for aspects in the clause, they used the effective opinion in the clause to select the orientation of the features. The opposite orientation of the sentence was used when no opinion appeared in the clause.

Popescu and Etzioni [25] used also extracted aspects to identify opinion words. Their idea was similar to that of [22] but instead of using a distance based approach, they used 10 “syntactic dependency rule templates” over a dependency tree to relate identified product features to potential opinion words. Then, the potential opinion phrases were examined in order to find the actual opinion phrases. Every phrase whose words had a positive or negative sentiment orientation was considered as an opinion phrase. A novel relaxation- labeling technique was used to determine the semantic orientation of potential opinion words in the context of the extracted product features and specific review sentences. The presence of negation modifiers was taken into consideration in this work but intensifiers were ignored.

In a similar, but less sophisticated technique, Godbole et al. [33] construct a sentiment lexicon by using a WordNet based technique, and associate sentiments with entities by assuming that a sentiment word found in the same sentence as an entity is describing that entity.

In [31] , Authors proposed a propagation based method to extract opinion and aspect simultaneously. This method is based on the fact that there are natural relation between opinion words and aspects because opinion words are used to describe aspects. They used a bootstrapping approach. Their approach start with an initial opinion word seeds and by using several syntactic relation that linked opinion words and aspects, it try to find new aspects. Then these new aspects and available opinion words are used to identify another aspects and opinion words. The process terminates until no more new aspects or opinion words can be identified.

8. EVALUATION OF SENTIMENT CLASSIFICATION

Generally, the performance of sentiment classification is evaluated by using four indexes: Accuracy, Precision, Recall and F1-score. The common way for computing these indexes is based on the confusion matrix shown in Table 3.

Table 3. Confusion Matrix

	Predicted positives	Predicted negatives
Actual positive instances	# of True Positive instances (<i>TP</i>)	# of False Negative instances (<i>FN</i>)
Actual negative instances	# of False Positive instances (<i>FP</i>)	# of True Negative instances (<i>TN</i>)

These indexes can be defined by following equations:

$$Accuracy = \frac{TP+TN}{TP+TN+FP+FN} \quad (1)$$

$$Precision = \frac{TP}{TP+FP} \quad (2)$$

$$Recall = \frac{TP}{TP+FN} \quad (3)$$

$$F1 = \frac{2 \times Precision \times Recall}{Precision + Recall} \quad (4)$$

Accuracy is the portion of all true predicted instances against all predicted instances. An accuracy of 100% means that the predicted instances are exactly the same as the actual instances. *Precision* is the portion of true positive predicted instances against all positive predicted instances. *Recall* is the portion of true positive predicted instances against all actual positive instances. *F1* is a harmonic average of precision and recall.

9. CONCLUSION

Sentiment analysis has many applications in information systems, including review classification, review summarization, synonyms and antonyms extraction, opinions tracking in online discussions and etc. this paper try to introduce sentiment classification problem in different level i.e. document-level, sentence-level, word-level and aspect-level. Also, some techniques that have been used to solve these problems have been introduced.

In future, more research is needed to improve methods and techniques introduced in this paper.

10. REFERENCES

- [1] Z. Zhang, Q. Ye, Z. Zhang, and Y. Li, "Sentiment classification of Internet restaurant reviews written in Cantonese," *Expert Syst. Appl.*, vol. 38, pp. 7674-7682, 2011.
- [2] B. Pang, L. Lee, and S. Vaithyanathan, "Thumbs up?: sentiment classification using machine learning techniques," presented at the Proceedings of the ACL-02 conference on Empirical methods in natural language processing - Volume 10, 2002.
- [3] Q. Ye, Z. Zhang, and R. Law, "Sentiment classification of online reviews to travel destinations by supervised machine learning approaches," *Expert Systems with Applications*, vol. 36, pp. 6527-6535, 2009.
- [4] R. Prabowo and M. Thelwall, "Sentiment analysis: A combined approach," *Journal of Informetrics*, vol. 3, pp. 143-157, 2009.
- [5] E. Riloff, S. Patwardhan, and J. Wiebe, "Feature subsumption for opinion analysis," presented at the Proceedings of the 2006 Conference on Empirical Methods in Natural Language Processing, Sydney, Australia, 2006.
- [6] C. Whitelaw, N. Garg, and S. Argamon, "Using appraisal groups for sentiment analysis," presented at the Proceedings of the 14th ACM international conference on Information and knowledge management, Bremen, Germany, 2005.
- [7] P. D. Turney, "Thumbs up or thumbs down?: semantic orientation applied to unsupervised classification of reviews," presented at the Proceedings of the 40th Annual Meeting on Association for Computational Linguistics, Philadelphia, Pennsylvania, 2002.
- [8] A. Harb, M. Planti, G. Dray, M. Roche, Fran, o. Troussset, and P. Poncellet, "Web opinion mining: how to extract opinions from blogs?," presented at the Proceedings of the 5th international conference on Soft computing as trans-disciplinary science and technology, Cergy-Pontoise, France, 2008.
- [9] M. Taboada, J. Brooke, M. Tofiloski, K. Voll, and M. Stede, "Lexicon-based methods for sentiment analysis," *Comput. Linguist.*, vol. 37, pp. 267-307, 2011.
- [10] H. Cui, V. Mittal, and M. Datar, "Comparative experiments on sentiment classification for online product reviews," presented at the proceedings of the 21st national conference on Artificial intelligence - Volume 2, Boston, Massachusetts, 2006.
- [11] J. R. Martin and P. R. R. White, *The Language Of Evaluation: Appraisal In English*. London: Palgrave, 2005.
- [12] R. McDonald, K. Hannan, T. Neylon, M. Wells, and J. Reynar, "Structured Models for Fine-to-Coarse Sentiment Analysis," in Proceedings of the 45th Annual Meeting of the Association of Computational Linguistics, 2007.
- [13] T. Nakagawa, K. Inui, and S. Kurohashi, "Dependency tree-based sentiment classification using CRFs with hidden variables," presented at the Human Language Technologies: The 2010 Annual Conference of the North American Chapter of the Association for Computational Linguistics, Los Angeles, California, 2010.
- [14] A. Neviarouskaya, H. Prendinger, and M. Ishizuka, "Recognition of affect, judgment, and appreciation in text," presented at the Proceedings of the 23rd International Conference on Computational Linguistics, Beijing, China, 2010.
- [15] C. E. Izard, *The Face of Emotion: Appleton Century Crofts*, 1971.
- [16] V. Hatzivassiloglou and K. R. McKeown, "Predicting the semantic orientation of adjectives," presented at the Proceedings of the eighth conference on European chapter of the Association for Computational Linguistics, Madrid, Spain, 1997.
- [17] P. D. Turney and M. L. Littman, "Measuring praise and criticism: Inference of semantic orientation from association," *ACM Trans. Inf. Syst.*, vol. 21, pp. 315-346, 2003.
- [18] A. Esuli and F. Sebastiani, "Determining the semantic orientation of terms through gloss classification," presented at the Proceedings of the 14th ACM international conference on Information and knowledge management, Bremen, Germany, 2005.
- [19] A. Esuli and F. Sebastiani, "Determining Term Subjectivity and Term Orientation for Opinion Mining Andrea Esuli," In Proceedings of the 11th Conference of the European Chapter of the Association for Computational Linguistics (EACL'06), 2006.
- [20] S. Baccianella, A. Esuli, and F. Sebastiani, "SentiWordNet 3.0: An Enhanced Lexical Resource for Sentiment Analysis and Opinion Mining," in Proceedings of the Seventh conference on International Language Resources and Evaluation (LREC'10), 2010, pp. 2200-2204.
- [21] S.-M. Kim and E. Hovy, "Determining the sentiment of opinions," presented at the Proceedings of the 20th international conference on Computational Linguistics, Geneva, Switzerland, 2004.

- [22] M. Hu and B. Liu, "Mining and summarizing customer reviews," presented at the Proceedings of the tenth ACM SIGKDD international conference on Knowledge discovery and data mining, Seattle, WA, USA, 2004.
- [23] J.-Y. Yang, H.-J. Kim, and S.-G. Lee, "Feature-based product review summarization utilizing user score," *Journal of information science and engineering*, vol. 26, pp. 1973-1990, 2010.
- [24] B. Liu and L. Zhang, "A Survey of Opinion Mining and Sentiment Analysis," C. C. Aggarwal and C. Zhai, Eds., ed: Springer US, 2012, pp. 415-463.
- [25] A.-M. Popescu and O. Etzioni, "Extracting product features and opinions from reviews," presented at the Proceedings of the conference on Human Language Technology and Empirical Methods in Natural Language Processing, Vancouver, British Columbia, Canada, 2005.
- [26] J. Yi, T. Nasukawa, R. Bunescu, and W. Niblack, "Sentiment Analyzer: Extracting Sentiments about a Given Topic using Natural Language Processing Techniques," presented at the Proceedings of the Third IEEE International Conference on Data Mining, 2003.
- [27] N. Jakob and I. Gurevych, "Extracting opinion targets in a single- and cross-domain setting with conditional random fields," presented at the Proceedings of the 2010 Conference on Empirical Methods in Natural Language Processing, Cambridge, Massachusetts, 2010.
- [28] J. S. Kessler and N. Nicolov, "Targeting sentiment expressions through supervised ranking of linguistic configurations," in *Proceedings of the Third International AAAI Conference on Weblogs and Social Media*, San Jose, California, USA, 2009, pp. 90-97.
- [29] W. Jin, H. H. Ho, and R. K. Srihari, "OpinionMiner: a novel machine learning system for web opinion mining and extraction," presented at the Proceedings of the 15th ACM SIGKDD international conference on Knowledge discovery and data mining, Paris, France, 2009.
- [30] V. Stoyanov and C. Cardie, "Topic identification for fine-grained opinion analysis," presented at the Proceedings of the 22nd International Conference on Computational Linguistics - Volume 1, Manchester, United Kingdom, 2008.
- [31] G. Qiu, B. Liu, J. Bu, and C. Chen, "Opinion word expansion and target extraction through double propagation," *Comput. Linguist.*, vol. 37, pp. 9-27, 2011.
- [32] K. Dave, S. Lawrence, and D. M. Pennock, "Mining the peanut gallery: opinion extraction and semantic classification of product reviews," presented at the Proceedings of the 12th international conference on World Wide Web, Budapest, Hungary, 2003.
- [33] N. Godbole, M. Srinivasaiah, and S. Skiena, "Large-scale sentiment analysis for news and blogs," in *International Conference on Weblogs and Social Media (ICWSM)*, 2007, pp. 219-222.