Opinion mining: Reviewed from word to document level

Malik Muhammad Saad Missen, Mohand Boughanem, Guillaume Cabanac

To link to this article : DOI :10.1007/s13278-012-0057-9
URL : http://dx.doi.org/10.1007/s13278-012-0057-9

To cite this version : Missen, Malik Muhammad Saad and Boughanem, Mohand and Cabanac, Guillaume *Opinion mining: Reviewed from word to document level*. (2013) Social Network Analysis and Mining, vol. 3 (n° 1). pp. 107-125. ISSN 1869-5450

Opinion mining: reviewed from word to document level

Malik Muhammad Saad Missen · Mohand Boughanem · Guillaume Cabanac

Abstract Opinion mining is one of the most challenging tasks of the field of information retrieval. Research community has been publishing a number of articles on this topic but a significant increase in interest has been observed during the past decade especially after the launch of several online social networks. In this paper, we provide a very detailed overview of the related work of opinion mining. Following features of our review make it stand unique among the works of similar kind: (1) it presents a very different perspective of the opinion mining field by discussing the work on different granularity levels (like word, sentences, and document levels) which is very unique and much required, (2) discussion of the related work in terms of challenges of the field of opinion mining, (3) document level discussion of the related work gives an overview of opinion mining task in blogosphere, one of most popular online social network, and (4) highlights the importance of online social networks for opinion mining task and other related sub-tasks.

1 Introduction

Opinion mining is the process of extracting opinions from text documents (Liu 2007). In the literature, this process is also known by expressions like "sentiment analysis", and/or "subjectivity analysis' ' (Pang and Lee 2008). If we look at definitions of opinion mining, sentiment analysis, and subjectivity analysis, these seem to denote the same field of study. The term *sentiment analysis* made its appearance in articles like (Nasukawa and Yi 2003; Yi et al. 2003) with the task of classifying given text into positive or negative classes. However, nowadays this term is used in a broader sense and is meant for computational treatment of opinion, sentiment, and subjectivity in the text (Pang and Lee 2008). Wiebe (1990) defines subjectivity as a function of private states (i.e., the states that are not open to objective observation or verification). Opinions, evaluations, emotions, and speculations all fall into this category (Pang and Lee 2008). The process of analyzing these opinions and emotions is called *Subjectivity Analysis* whose objective is to recognise the opinion-oriented language to distinguish it from objective language. Other commonly used terms for this process are "opinion detection", "sentiment detection", "polarity detection", "opinion finding", and "polarity retrieval". In addition to this, many other terms have been used for opinion-related work [like "affective computing" (Picard 2002), "review mining" (Zhuang et al. 2006), "appraisal extraction" (Bloom et al. 2007), etc.] but in this paper we will limit ourselves to use of the most common terms mentioned above.

Year 2001 was the beginning of widespread awareness of the research problems related to opinion mining which caused hundreds of papers published on this subject (see Fig. 1).[1] The popularity of machine learning, availability of huge opinion data collections in the form of online social networks (e.g., blogging, tweeting, product review forums

M. M. S. Missen (✉)
Department of Computer Science and Information Technology, The Islamia University of Bahawalpur, Bahawalpur, Pakistan
e-mail: Saad.Malik@iub.edu.pk

M. Boughanem · G. Cabanac
Institut de Recherche en Informatique de Toulouse (IRIT), Université de Toulouse, UMR 5505 CNRS, 31062 Toulouse, France
e-mail: Mohand.Boughanem@irit.fr

G. Cabanac
e-mail: Guillaume.Cabanac@irit.fr

[1] Taken from slides of talk by Andrea Esuli on the topic of opinion mining in Istituto di Scienza e Tecnologie dell'Informazione Consiglio Nazionale delle Ricerche, Pisa, Italy, 14th June 2006.

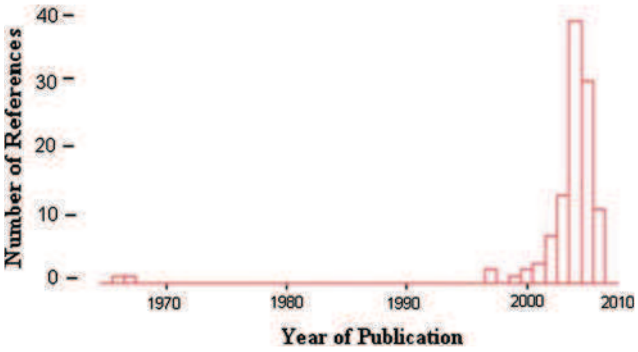

Fig. 1 Emerging trend in number of articles for opinion mining research

etc.) and evolving nature of information needs are described the main factors behind this shift in focus from traditional adhoc information retrieval (IR) to opinion retrieval (Pang and Lee 2008). Adhoc IR focuses on factual information retrieval while purpose of opinion retrieval is to retrieve opinions for a given query. Following subsection briefly highlights major differences between these two tasks.

1.1 Fact-based retrieval versus opinion-based retrieval

To better understand the difference between adhoc IR and opinion retrieval, let us try to understand the term *opinion* itself. Bethard et al. (2004) defines opinion as *A sentence, or a part of a sentence, that would answer the question, "How does X feel about Y?"* This definition suggests that opinions are subjective (i.e., if an individual is asked a question then he/she might give a different answer from another person). For example, many people will agree with the following statement: *This colour is too bright to suit you*, given by a person X for a person Y dressed in black, but others may disagree too because there is no standard defined for *the best colour* for a specific person. It means that *opinions* are different from *facts* because a factual statement (e.g., *July 14 celebrates France National Day*) remains valid for all individuals while opinions might vary from person to person.

The major tools used for searching information on the web are search engines like Google, Yahoo, etc., but they are more focused to retrieve topic-based factual information rather than opinion information (Liu 2008). Pang and Lee (2008) differentiate the treatment of opinionated text from classic topic-based classification. According to them, traditionally text classification seeks to classify documents by topic. While dealing with topics, we can have as few as two classes (like Relevant and Non-Relevant) or as many as thousands of classes (i.e., when classifying w.r.t. a taxonomy) for text classification. But in case of classifying opinions, generally we have few classes (like positive, negative or neutral, etc.). In addition, while dealing with topic-based categorization, different classes can be unrelated to each other but as far as opinion-based categorization is concerned, the classes for categorization are always related somehow (i.e., whether they are opposite or they have some ordinal relation between them). Further, Tang et al. (2009) and Ku and Chen (1838) give similar kind of arguments while differentiating opinion-based information retrieval from classic topic-based factual information retrieval.

In this paper, we review the literature work of opinion mining by summarizing the work in a very unique fashion making this contribution very useful for other researchers. This article also discusses the major challenges of this field and highlights different works that have tried to tackle these challenges.

2 Major opinion mining references

In this section, we list prominent existing works that have summarized the work related to opinion mining with respect to different aspects.

2.1 Work by Tang et al.

Tang et al. (2009) present a detailed survey of work for sentiment detection in product reviews. They identify three kinds of major approaches in the literature for sentiment detection in real-world applications:

- *Machine Learning Approaches* In this type of approaches, generally a machine learning classifier is trained on already annotated data to create a model of the trained data and then this model is used to estimate the classes of documents in the test data.
- *Semantic Analysis Approaches* Lexical resources play a very important role in this type of approaches. Semantic relations of concepts, extracted from some lexical resource, are used to provide some evidences about the subjectivity. Use of synonyms and antonyms has been very common in this regard.
- *Natural Language Processing Approaches* Approaches exploiting the Parts-of-Speech (POS) information, complex syntactical structural information, etc. are part of this type of approaches.

Besides this, Tang et al. also highlight the related work in context of major tasks like "subjectivity classification", "sentiment classification", etc.

2.2 Work by Esuli and Sebastiani

Similarly, Esuli and Sebastiani (2006) have categorized the related work in three classes according to the nature of tasks associated with sentiment detection. These three classes are

- *Determining Text SO-Polarity* The type of approaches belonging to this class focuses on the task of deciding whether a given text is factual or contains opinions on a topic (i.e., a binary text categorization with classes *Subjective and Objective*).
- *Determining Text PN-Polarity* The task this type of approaches focused on is to evaluate the polarity of a subjective text (i.e., whether given subjective text contain positive or negative opinion about the target).
- *Determining the strength of Text PN-Polarity* Once it has been decided whether a given text is positive or negative, then the task of determining the degree of its positivity or negativity becomes active. The approaches in this category of classes calculate this degree of positivity or negativity.

2.3 Work by Pang and Lee

While Esuli and Sebastiani only describe three tasks related to problem of opinion mining, Pang and Lee (2008) identify a set of relatively larger number of opinion-related tasks in the literature. Few major tasks are listed below:

- *Sentiment Polarity Classification* It is a binary classification task in which the polarity of a given opinionated document is estimated to be positive or negative.
- *Likely versus Unlikely* Another related task identified by Pang and Lee (2008) is classifying predictive opinions in election forums into *likely to win* and *unlikely to win* classes.
- *Good versus Bad News* Classifying a news article as a *good news* or *bad news* has also been identified as a sentiment classification task.
- *Reviewer's Evaluation* Another task is to determine reviewer's evaluation with respect to a multi-point scale (e.g., one to five stars for a review). This problem can be seen as a multi-class categorization problem.
- *Agreement Detection* Given a pair of texts, deciding whether they should receive the same or different sentiment-related labels based on the relationship between elements of the pair.
- *Opinion Strength* Another task identified was to determine the clause-level opinion strength (e.g., *How mad are you?*).
- *Viewpoint Classification* Classifying the viewpoints and perspectives into classes like *liberal, conservative, libertarian, etc*. is another task identified.
- *Genre Classification* This task focuses on determining the genre of a given piece of text, i.e. whether the given text is an editorial, advertisement or announcement, etc.
- *Source Classification* Classifying the documents according to their source or source style. Authorship identification is a very good example of such task or similarly classifying the documents according to their publisher (e.g., *The Washington Post or The Daily News*).

3 Granularity-based state-of-the-art

While works by Tang et al. (2009), Esuli and Sebastiani (2006), and Pang and Lee (2008) have organized the related work according to the nature of tasks and type of approaches adopted, we provide a novel granularity-based (word, sentence/passage, and document) classification of the related work for opinion mining. In this section, we describe the opinion mining process in steps and discuss the work related to each step in separate sections. This organization of work is very useful for researchers working on the task of opinion detection at any granularity level.

3.1 Opinion detection process

The process of "opinion detection" can be described in following major steps:

1. retrieve the relevant set of documents for a given topic (Topic Relevance Retrieval) if needed,
2. compute the word-level polarity orientations (determining whether a word is positive or negative) and polarity strengths (determining the strength of the positivity or negativity of a word),
3. combine the word-level subjectivity scores, polarity orientations or strength to calculate the polarity orientations and strengths on sentence-level (or passage-level),
4. combine the sentence-level subjectivity scores, polarity orientations or strengths to compute the polarity orientations and strengths of the given document.
5. Combine the relevance and opinion scores of a document to compute its final score.

Each step of the above process sacks lot of research work. We will discuss the related work in terms of major techniques being used for opinion finding to give an overview of related work from various perspectives.

3.2 Word level processing

We can identify the following three word-level sentiment analysis tasks in the literature (Esuli and Sebastiani 2006):

- to determine subjectivity of words in a document (i.e., whether the word is subjective or objective)
- to determine orientation or polarity of words (i.e., whether the word is positively subjective or negatively subjective)

– to determine strength of orientation (i.e., how much positive or negative a word is)

Most of the approaches found in literature do not explicitly differentiate between these tasks because all of these tasks are inter-related. For example, an approach whose aim is to determine the polarity strength of a word might start with tasks of determining subjectivity and polarity of words. Similarly, an approach that is meant to determine the sentimental orientation of words might use their polarity scores to decide about their polarity. Therefore, in this section we will discuss the approaches which focus on any of these tasks. Generally two kinds of approaches have been proposed for determining the sentiment orientation of words (Andreevskaia and Bergler 2006b): first, *Corpus-Based approaches* and second, *Dictionary-Based Approaches*.

3.2.1 Corpus-based approaches

Corpus-based approaches generally exploit the inter-word relationships (syntactic or co-occurrence relationships) in large corpora to perform any of the three tasks defined above (Grefenstette et al. 2006; Hatzivassiloglou and McKeown 1997; Kim and Hovy 2004; Turney and Littman 2002; Yu and Hatzivassiloglou 2003). We discuss some major works from the proposed approaches by classifying them according to the nature of evidences used.

Using language constructs This type of opinion mining approaches generally take support of language constructs (conjunctions, prepositions, grammar rules, etc.). For example, Hatzivassiloglou and McKeown (1997) proposed a method for automatically tagging the adjectives with a sentimental tag (positive or negative) with the help of conjunctions (and, or, but, either-or, or neither-nor) joining them. The basic principle behind their approach was that adjectives combined with the conjunction *and* (like *beautiful and calm*) are supposed to have same orientation while those joined by conjunction *but* (like *justified but brutal*) generally differ in their sentimental orientations. A classification precision of over 90% was observed for adjectives that occur with modest number of conjunctions in the corpus. Similarly, Wilson et al. (2005b) use conjunctions (in the same manner as used by Hatzivassiloglou and McKeown 1997), local negations (i.e., presence of a negative word before a polar expression) and dependency tree to disambiguate the contextual polarities (discussed in Sect. 4.5 in detail) of polar expressions which helped to significantly improve the baseline for the phrase-level sentiment classification task. Other studies (Hatzivassiloglou and Wiebe 2000; Wiebe 2000) showed that restricting features, used for classification, to those adjectives that come through as strongly dynamic, gradable, or oriented improved performance in the genre-classification task.

Using co-occurrence evidence In this type of opinion mining approaches, opinion score of a word is computed on behalf of its distance from the already known list of opinionated words. For example, Baroni and Vegnaduzzo (2004) used a seed list of subjective adjectives to rank another list of adjectives that are to be ranked in descending order by their subjectivity. The motivating factor behind this work was the intuition that subjective adjectives are most likely to co-occur with other subjective adjectives. They computed the subjectivity score of target adjectives by computing their mutual information with the adjectives of seed set and *Pointwise Mutual Information* (PMI) (Church and Hanks 1990; Grefenstette et al. 2006; Stone and Hunt 1963) technique was used for this purpose. PMI can be defined as (Church and Hanks 1990):

$$\mathrm{PMI}(t, t_i) = \log_2\left(\frac{p(t\&t_i)}{p(t)\cdot p(t_i)}\right) \tag{1}$$

where $p(t \;\&\; t_i)$ is the probability that terms t and t_i occur together. In other words, above equation represents the measure of the degree of statistical dependence between t and t_i.

A similar kind of approach was proposed by Turney and Littman (2002, 2003) wherein they prepared a list of positive terms (i.e., good, nice, excellent, positive, fortunate, correct, superior) and a list of negative terms (i.e., bad,nasty,poor,negative,unfortunate,wrong,inferior) to be used as seed terms. The semantic orientation of a given term t (i.e., $O(t)$) is computed as

$$O(t) = \sum_{t_i\in S_p}\mathrm{PMI}(t, t_i) - \sum_{t_i\in S_n}\mathrm{PMI}(t, t_i) \tag{2}$$

where $\mathrm{PMI}(t,\ t_i)$ is the Pointwise Mutual Information (Church and Hanks 1990) score for term t with each seed term t_i as a measure of their semantic association.

The results showed that this approach required a large data collection for good performance. Even this is understandable because the reliability of the co-occurrence data increases with the number of documents for which co-occurrence is computed but still it is a limitation of this approach. Another drawback with this approach is that it did not deal with ambiguous terms (having both positive and negative senses at a time like the word *mind, unpredictable, etc*.) because the ambiguous terms were deleted from the set of testing words.

3.2.2 Dictionary-based approaches

The second type of approaches for word-level sentiment analysis benefits from the flexibility provided by various lexicons (Esuli and Sebastiani 2006; Miller 1995; Stone et al. 1966) through their nature, structure, and lexical relations. The definitions like terms' glosses (Esuli and Sebastiani 2005) and semantic relations (like synonyms and antonyms) (Kamps et al. 2004) provide enough level of liberties to the researchers to be exploited for finding semantic orientations of words.

Use of semantic relations Use of semantic relations has always been part of classical IR and it has got equal importance in the field of opinion mining and sentiment analysis. There exist a number of publications exploiting lexical semantic relations between concepts to estimate their subjectivity which eventually assists to estimate the subjectivity of a document. For example, Kamp et al. (2004) developed a distance based WordNet measure which determines the semantic orientations of adjectives based on the distance of a given word from two selected reference words, "good" and "bad". WordNet (Miller 1995) is a large lexical database containing about 150,000 words organized in over 115,000 synset entries for a total of 203,000 word-sense pair (Pasca 2005). The concepts in WordNet are related through various semantic relations. Like Kamp et al., Williams and Anand (2009) use lexical relations of WordNet to assign polarity scores to adjectives. They use a small set of reference positive and negative terms to build an adjective graph, using the lexical relations defined in WordNet. To compute the polarity strength of adjectives, they used various combinations of lexical relations. The best results were achieved when using the lexical relations of related words and similar words in addition to the standard synonym relation commonly used.

Use of gloss definitions Each word in WordNet comes along with a short description for all of its senses which is called its gloss definition. The glosses are usually one or two sentences long. For example, gloss definitions for the word *Car* are

– a motor vehicle with four wheels; usually propelled by an internal combustion engine,
– a wheeled vehicle adapted to the rails of railroad,
– the compartment that is suspended from an airship and that carries personnel and the cargo and the power plan,
– where passengers ride up and down,
– a conveyance for passengers or freight on a cable railway.

There are some approaches (Esuli and Sebastiani 2005, 2006; Sebastiani et al. 2006) that make use of the quantitative analysis of the gloss definitions of terms found in online dictionaries to determine their semantic orientations. The motivation behind the work of Esuli and Sebastiani (2005) is the assumption that if a word is semantically oriented in one direction, then the words in its gloss tends to be oriented in the same direction. For instance, the glosses of terms *good* and *excellent* will both contain appreciative expressions, whereas the glosses of *bad* and *awful* will both contain derogative expressions.

Sebastiani et al. (2006) extend the work presented in Esuli and Sebastiani (2005) by including an additional task of determining term subjectivity. Further extension to these works led to the creation of an automatic subjectivity lexicon SentiWordNet (SWN) (Esuli and Sebastiani 2006). SWN assigns three numerical scores ($Obj(s)$, $Pos(s)$, $Neg(s)$) to each synset of the WordNet describing how objective, positive or negative the terms within a synset are. The range of three scores lies in interval [0, 1] and sum of all the scores equals to 1. This process of assigning scores makes the task of determining semantic orientation and semantic strength more precise than the one in which terms are labeled just with tags subjective or objective(for semantic orientation task) or Strong or Weak (for polarity strength task). All of three scores are obtained by combining the results of eight ternary classifiers, all characterized by similar accuracy levels but different classification behavior. A template of SWN is shown in Fig. 2.

Quantitative analysis of the glosses of the synsets is performed to obtain three scores as mentioned above. The basic intuition behind the creation of SWN was that different senses of a term might have different semantic orientations.

However, there are few other works (Andreevskaia and Bergler 2006a; Kim and Hovy 2004; Subasic and Huettner 2001) too who have treated the task of determining

```
# POS offset      PosScore    NegScore    SynsetTerms
a     1000003     0.0   0.125 form-only#a#1
a     1000159     0.25  0.0   dress#a#1 full-dress#a#1
a     1000307     0.0   0.0   titular#a#5 nominal#a#6
a     1000440     0.0   0.0   prescribed#a#4 positive#a#5
a     1000554     0.0   0.25  perfunctory#a#2 pro_forma#a#1
a     1000681     0.0   0.5   semiformal#a#1 black-tie#a#1 semi-formal#a#1
a     10007 0.0   0.625 abstentious#a#1 abstinent#a#1
a     1000859     0.0   0.0   starchy#a#2 buckram#a#1 stiff#a#4
```

Fig. 2 Template of SentiWordNet with *first column*: Parts of Speech (POS) of the Synset; *2nd column*: Offset of the Synset in WordNet; *3rd Column*: Positive Score of the Synset; *4th Column*: Negative Score of the Synset; *5th Column*: Entries of a Synset semantic orientation same as (Esuli and Sebastiani 2006) (i.e., instead of viewing the properties of positivity and negativity as categories, graded versions of these properties have been proposed.)

Using WordNet affect Valitutti (2004) developed a lexicon called *WordNet Affect* for representation of affective knowledge by selecting and tagging a subset of WordNet synsets with the affective concepts like emotion, trait, and feeling, etc. For building this lexicon, a support was taken from another lexicon *WordNet Domains* (Magnini and Cavagli 2000). WordNet Domains is a multilingual extension of the WordNet and provides at least one domain label (like sports, politics, and medicine, etc.) for each of its synset. It has a hierarchy of almost two hundred domain labels. WordNet-Affect is an additional hierarchy of the affective domain labels, independent from the domain hierarchy, wherewith the synsets that represent affective concepts are annotated. Bobicev et al. (2010) have used WordNet-Affect to develop another multilingual (Russian and Romanian) WordNet-Affect lexical resource.

There are very few works though where both of above approaches (i.e., dictionary-based and corpus-based approaches) were combined to improve the results (like Zhang et al. 2009). Generally, it has been observed that corpus-based approaches for word-level subjectivity classification perform better than dictionary-based approaches. However, the performance of corpus-based approaches is badly affected across different domains. On the other hand, most of the dictionary-based approaches generally take support of domain-independent lexical resources (e.g., SentiWordNet, WordNet) and hence avoid the drawback of corpus-based approaches. However, performance of dictionary-based approaches might vary with the nature and scope of the lexicon being used.

3.3 Sentence level processing

Most of the work related to opinion mining on sentence level focuses on following two tasks:

– to determine whether a sentence is subjective or objective,
– to determine whether a sentence is positive or negative.

In this section, we will discuss few major contributions for both tasks.

3.3.1 Sentence subjectivity identification

In this section, we will discuss approaches that have used different types of evidences to determine whether a given sentence is subjective or objective.

Using presence of subjective words Most of the approaches rely on the evidence of presence of subjective words in a sentence to analyze the subjectivity of that sentence. For example, Hu and Liu (2004) proposed a very simple method of finding the opinionated sentences for summarizing the product reviews in which a sentence is considered as an opinionated sentence if it contains one or more product features and one or more opinion words.

Zhang et al. (2009) found that presence of a single strong opinionated word in a sentence (Model-2) could prove more useful than using total opinion score of all words in that sentence (Model-1) to evidence the subjectivity of that sentence (see Table 1). Hu and Liu (2004) experimented with the same evidence (i.e., if a sentence contains one or more opinion words then the sentence is considered an opinion sentence) which proved to be effective.

However, an interesting relation between presence of adjectives in a sentence and its subjectivity have been explored by many works (Bruce and Wiebe 1999; Hatzivassiloglou and Wiebe 2000; Wiebe 2000; Wiebe et al. 2004). For example, Bruce and Wiebe (1999) proved that adjectives are statistically, significantly, and positively correlated with subjective sentences in the corpus on the basis of the log-likelihood ratio. The probability that a sentence is subjective, simply given that there is at least one adjective in the sentence, is 55.8%, even though there are more objective than subjective sentences in the corpus. Hatzivassiloglou and Wiebe (2000) adapt a very simple method to predict the subjectivity of a sentence. They classified a sentence as subjective if at least one member of a set of adjectives *S* (obtained from previous works like, Bruce and Wiebe 1999; Hatzivassiloglou and McKeown 1997) occurs in the sentence and objective otherwise.

We have seen that most of the earlier work depends on presence of adjectives within a sentence for subjective classification of a sentence, but the work by Riloff et al. (2003) showed the effectiveness of nouns for identification of subjective sentences. With the help of Naive Bayes classifier, they were able to achieve a precision of 81% on subjective classification of sentences.

Use of sentence similarities Similarity approach to classifying sentences as subjective or objective explores the hypothesis that, within a given topic, opinion sentences will be more similar to other opinion sentences than to factual sentences [use of state-of-the-art sentence similarity

Table 1 System performance with different models and cutoff values on TREC 2003 data

Model	System parameter λ	F score
Model-1	0.2	0.398
	0.3	0.425
Model-2	0.2	0.514
	0.3	0.464

algorithm SIMFINDER (Vasileios et al. 2001) by Yu and Hatzivassiloglou (2003)]. The similarity approach generally exploits the evidences like shared words, phrases, and WordNet synsets for measuring similarities (Dagan et al. 1993, 1994; Leacock and Chodorow 1998; Miller and Charles 1991; Resnik 1995; Zhang et al. 2002).

Using presence of subjective words as an evidence for deciding about the subjectivity of a sentence has proved its worth. Even finding similarities between a set of subjective sentences and candidate sentences using some machine learning techniques has shown good results, but performance in this case is prone to drawback of lower performance across different domains.

3.3.2 Sentence polarity tagging

It is to be noted that the performance of an approach developed for predicting the polarity orientation of a sentence is dependent on the performance of the approach proposed to estimate the polarity estimation of words within that sentence. Therefore, it is only an effective combination of techniques on both levels that can eventually give good performance for predicting the sentimental orientation of sentences.

Using number of polar words Hu and Liu (2004) proposed a very simple method for detection of sentence polarity orientation. According to them if a sentence contains more number of positive words than negative words, it is considered as a positive sentence; otherwise negative. In the case where there are equal numbers of positive and negative opinion words in the sentence, they predict the orientation using the average orientation of effective opinions or the orientation of the previous opinion sentence. Their approach performed well by giving an average accuracy of 84% in predicting the sentence sentimental orientation.

Using word-level polarity scores The approach proposed by Yu and Hatzivassiloglou (2003) tags the opinion sentences with polarity tags (i.e., positive or negative). They used a co-occurrence measure including a seed-set of semantically oriented words to estimate the polarity orientations of words in a sentence. This has been discussed in previous section in detail. For evaluation purposes, they aggregated the word-level polarity scores to estimate sentence level polarity orientations with different combinations of parts-of-speeches (i.e., adjectives, adverbs, nouns, verbs). However, maximum accuracy was obtained (90% over baseline of 48%) when they combined word-level evidences for adjectives, adverbs, and verbs.

We have seen that most of the sentence-level work depends on the semantic orientations of the words present in the sentence to compute its semantic orientation. But it should be noted that polarity of a word is likely to change when it is surrounded by other words in a sentence. In other words, polarity of an individual word (*prior polarity*) and polarity of a word in a sentence (*contextual polarity*) are most likely to be different. For example, take the following sentence: *John's house is not beautiful at all*. We know that word *beautiful* has a positive prior polarity but in the above sentence the contextual polarity of the word *beautiful* is negative because of the presence of negation *not* just before the word *beautiful* in the sentence. In rest of the discussion for sentence polarity tagging, we will present some works that have proposed sentence polarity approaches by focusing on the problem of contextual polarity of words.

Using word-level context-aware polarity approaches Contextual polarity of a term is the polarity which is generated after modification of the prior polarity of the term. This modification of the prior polarity occurs because of change in the context. Here we define few major contexts responsible for polarity shift of the terms:

- This type of contextual polarity is defined by the presence of negations (like *not, neither, nor or never*, etc.) in surroundings of a given word. For example, *Good* is a positive word but if preceded by a negation like *not or never*, its contextual polarity is changed from positive to negative.
- The second type of contextual polarities are caused by the senses of a word as found in a everyday dictionary (like WordNet). A word can have many senses. This is called *Polysemy*. For example, *bank* can be used as a *financial institute* or a *river shore*. Similarly, the polarities of words can be different for different senses of a word. For example, while the word *strong* is considered a positive adjective (with positive score of 0.75 and negative score 0.0) when used as sense strong#a#7, it is more likely to highlight its negative aspect (with negative score of 0.5 and positive score of 0.0) when used as sense strong#a#8 in subjective lexicon SWN (Esuli and Sebastiani 2006).
- Third type of contextual polarity is defined by the type of the topic (or query) we are searching for, so we call it *Topic-Dependent contextual polarity*. For example, the word *unpredictable* in an opinion document containing opinion about a film as *unpredictable film plot* will be taken as a positive. On the contrary, if the same word is used in another document containing opinion about a digital camera as *unpredictable functional response* then this time it will be considered as a negative word. Hence, a change in term's semantic orientation is observed with the change in topic-class, i.e. from movie class to product class.

However, there are few works (Grefenstette et al. 2004; Hu and Liu 2004; Kim and Hovy 2004; Ku et al. 2006) that have dealt with the problem of local contextual polarities by focusing on negations like *no, not, never*, etc. However, works like (Nasukawa and Yi 2003; Wilson et al. 2005b; Yi et al. 2003) also focus on other type of contextual polarities. Kim and Hovy (2004) and Ku et al. (2006) use contextual polarities of words to identify the sentence polarity. To identify contextual polarities of words in sentences, they take into account the negations (like *not and never*) and reverse the prior polarity of the words following these negations. Kim and Hovy (2004) further use a window-based approach for sentence polarity detection, whereas Ku et al. (2006) decide the opinionated tendency of a sentence by the function of sentiment words and the opinion holder as follows:

$$S_p = S_{\text{Opinion Holder}} \times \sum_{j=1}^{n} S_{w_j} \tag{3}$$

where S_p, $S_{\text{opinion Holder}}$, and S_{wj} are sentiment score of sentence p, weight of opinion holder, and sentiment score of word w_j, respectively, and n is the total number of sentiment words in p.

Wilson et al. (2005b) propose some features to automatically identify the contextual polarities of sentimental expressions. Further work from Wilson et al. (2003, 2005a) also includes the development of sentence level subjectivity detection tool (i.e., Opinion Finder). Other worth reading orks that focus on identification of contextual polarities include (Ding et al. 2008; Nasukawa and Yi 2003; Yi et al. 2003).

While using presence of polar words in a sentence proved to be useful for computing its subjectivity, using this evidence for computing its polarity could not be as effective because of the contextual polarity problem which directly affects the polarity of the sentence. An effective approach dealing efficiently with contextual polarity problem could perform better as observed in related work.

3.4 Document-level processing

Most of the earlier work on document-level sentiment detection is limited to the use of data collections like news articles and product reviews. However, with the popularity of online social networks, various types of data collections have emerged (like collection of blogs and tweets) that have given boost to the research work in this field. For example, a significant increase in interest for research in opinion mining field has been noticed after start of TREC Blog track in year 2006 (see Fig. 1).

In this section, we will discuss the approaches focusing on identifying opinionated documents and classifying them according to their polarities (i.e., positive, negative or neutral). A two-step approach is generally followed by most of the works for the task of opinion detection with very few exceptions (like Attardi and Simi 2006) that adapt a single-step method. In the first step, called *Topical Relevance Retrieval*, a set of relevant documents is retrieved for a given topic. In the second step, called *Opinion Finding step*, the set of relevant documents retrieved during first step is processed and re-ranked according to their opinionatedness.

It has been seen that many approaches have used various kinds of topic-relevance methods to obtain a set of relevant documents. Knowing that performance of an opinion finding approach depends on the performance of topic-relevance baseline (Soboroff et al. 2007, 2008; Ounis et al. 2006), it becomes meaningless to compare two opinion finding approaches using two different topic-relevance baseline. This is the reason TREC Blog track provided five standard topic relevance baseline runs (chosen from the baselines submitted by participants for topic relevance task) to its participants of TREC 2008 to evaluate the performance of different approaches on common baselines which can give better idea of effectiveness of an approach. The details of these baselines are given below in Table 2.

In this section, we discuss the related work for document-level opinion finding from different perspectives. Globally, we discuss the related work with respect to the lexicon-based and machine learning-based techniques used. However, we also discuss the major data collections used for opinion finding task and the role relevance feedback has performed for this task. In addition to this, we acknowledge the importance of TREC Blog's opinion finding task by summarizing its key findings over years.

3.4.1 Using corpus-based dictionaries

In this section, we discuss the approaches that use an opinion lexicon for identifying opinionated documents. These lexicons may be explicitly prepared using the given test corpus (or some external corpus) or one can use ready-made lexicons (Esuli and Sebastiani 2006; Miller 1995) especially available for such kind of tasks.

Using internal corpus-based dictionaries Gerani et al. (2009) chose not to rely on external lexicons of opinionated terms, but investigate to what extent the list of opinionated terms can be mined from the same corpus of relevance/opinion assessments that are used to train the retrieval system. They calculate the opinion score of a term t by taking ratio (Weighted Log-Likelihood Ratio, Ng et al. 2006; Nigam et al. 2000) of relative frequency of the term t in set of opinionated documents (set O) to the set of relevant documents (set R and $O \subset R$). In order to calculate an opinion score for an entire document, average opinion score over all the words in the document is calculated as

Table 2 TREC provided baselines' relevance and opinion MAP (over all 150 topics from year 2006 to 2008) (Soboroff et al. 2008)

Baseline	Run type	Topics	Relevance		Opinion finding	
			MAP	P10	MAP	P10
Baseline1	Automatic	Title only	0.3701	0.7307	0.2639	0.4753
Baseline2	Automatic	Title-desc	0.3382	0.7000	0.2657	0.5287
Baseline3	Automatic	Title-desc	0.4244	0.7220	0.3201	0.5387
Baseline4	Automatic	Title-desc	0.4776	0.7867	0.3543	0.558
Baseline5	Manual	Title only	0.4424	0.7793	0.3147	0.5307

An *Automatic Run* involves no human intervention at any stage while in a Manual Run, queries could be extended or modified manually. A *Title Only Run* is a run in which only title of the topic is used while in a *Title-desc Run*, information from two parts of the topic i.e., title and description is used to generate the run

$$\text{Opinion}_{\text{avg}}(d) = \sum_{t \in d} \text{Opinion}(t) \cdot p(t|d) \tag{4}$$

where $p(t|d) = c(t,d)/|d|$ is the relative frequency of term t in document d.

There are few works (Gerani et al. 2009; He et al. 2008a, b) that have used the target collection itself to build the opinion lexicons which were to be used for opinion finding task. For example, He et al. (2008a) automatically created a lexicon of opinionated words with the help of Skewed Query Model (Cacheda et al. 2005) from the document collection (TREC Blog 2006 collection) they used for experimentation. Skewed Query Model was used to filter out too frequent or too rare terms in the collection. The terms are ranked in descending order by their collection frequencies using the skewed model. The final ranking of the documents is done with combination of opinion and relevance score (obtained with original unexpanded query) of the documents. This approach managed to improve the TREC strongest baseline of that time (Ernsting et al. 2007) and further all improvements were statistically significant according to the Wilcoxon test at 0.01 level.

Using external corpus-based dictionaries There are many who took the support of external opinionated data collections for building their own lexicons. There is always a trade-off between domain independency and performance in building a lexicon from external data collections (i.e., a lexicon built using external data collection tend to be more generalized but a bit poor in performance relative to a lexicon built from the given test data collection). Hui Yang and Si (2006) created the simplest form of dictionary created through web. This dictionary was composed of positive and negative verbs and adjectives were downloaded from the web. Finally, manual selection was used to shorten the list so that it is short enough to not to lengthen the retrieval time too much.

Similar to Hui Yang and Si (2006), Seki et al. (2007) adopt a very simple approach to build a lexicon of opinion terms from reviews of http://www.amazon.com. They explored to use 27,544 positive/negative customer reviews harvested from http://www.amazon.com to find good sentiment terms as features. Another work that makes use of external sources for building an opinion lexicon is He et al. (2007). They prepared a lexicon of 12,000 English words derived from various linguistic sources which gave an improvement of 15.8% over its baseline.

3.4.2 *Using ready-made dictionaries*

Use of domain-independent ready-made dictionaries is very common in the field of opinion mining. Dictionaries like General Inquirer, SentiWordNet, and WordNet Affect, etc., are available to researchers for this task. Many (Kennedy and Inkpen 2006; Mishne 2006; Wang et al. 2007) have used the lexicon General Inquirer (GI) for their work related to opinion finding. General Inquirer is a large-scale, manually-constructed lexicon. It assigns a wide range of categories[2] to more than 10,000 English words. The categories assigned are Osgood's semantic dimensions and emotional categories.

Mishni (2006) selects opinion sentences from a subset of topical sentences using words from GI categories. Opinion score of a document is computed on behalf of these opinion sentences selected and on several other modules. Wang et al. (2007) and Kennedy and Inkpen (2006) use GI categories for document-level opinion finding task.

The use of sentiment lexicons is very helpful for the tasks related to opinion detection but there is a need for more sophisticated lexicons and techniques that can get benefit of the information these lexicons are providing. Simply counting the occurrences of the opinion words in a document to calculate the document's subjectivity is not an optimal solution and is subject to many drawbacks. Given two subjective words, one might be stronger in its subjectivity than the other one. Intuitively, a document

[2] A complete list of the General Inquirer categories is given at http://www.wjh.harvard.edu/inquirer/homecat.htm.

containing stronger subjective words should be ranked higher than a document with equal number of subjective words but with lesser subjectivity. Therefore, such a lexicon is needed that not only categorizes the words as positive or negative but also assigns subjectivity scores to the words to indicate their polarity strength and avoid the problem mentioned above.

SentiWordNet (Esuli and Sebastiani 2006) solves the problem mentioned above by providing objective and subjective (i.e., positive and negative subjectivity scores) scores for each synset of the WordNet. Some works (Bermingham et al. 2008, Missen and Boughanem 2009; Zhang and Zhang 2006; Zhao et al. 2007) showed their interest in using SWN as a lexical resource. All of these approaches sum the opinion scores of the words in a document to compute the opinion score for that document. Zhang and Zhang (2006) fixed a threshold value of 0.5 for an adjective to be considered as subjective. Zhao et al. (2007) follow a similar approach but on document level (i.e., the document is positive if its subjectivity score $P(d) \geq 0.4$, negative if $P(d) \leq 0.2$, neutral if $0.2 < P(d) < 0.4$).

3.4.3 Text classification approaches

Text classification approaches (Aue and Gamon 2005; Beineke et al. 2004; Boiy and Moens 2009; Claire Fautsch 2008; He et al. 2008c; Java et al. 2006; Pang et al. 2002; Salvetti et al. 2004; Seki et al. 2007; Zhang and Zhang 2006) generally make use of some machine learning classifier trained on already annotated opinionated data and then are tested on test data. Most of the commonly used classifiers for opinion detection in blogs are Support Vector Machines (SVM) (Aue and Gamon 2005; Gerani et al. 2008; GuangXu et al. 2007; Hemant Joshi and Xu 2006; Java et al. 2006; Jia et al. 2008; Kennedy and Inkpen 2006; Mukras et al. 2007; Rui et al. 2007; Seki et al. 2007; Zhang et al. 2006; Zhang and Yu 2007), Logistic Regression Classifier (Claire Fautsch 2008; Zhang and Zhang 2006) and Maximum Entropy classifier (He et al. 2008c). SVM has been the most preferred machine learning classifier because SVMs are reported to perform better as compared with other machine learning algorithms. Most of the approaches have proposed very simple features for the opinion related tasks. The major ones used are

- the number of subjective words in a document d,
- the number of positive and negative words in a document d,
- the number of subjective sentences in a document d,
- the number of positive and negative sentences in a document d,
- the proximity approach (i.e., a fixed number of sentimental words around the topic words in a document or the fixed number of words around adjectives, verbs or adverbs),
- the use of punctuations like smiley faces : or sad faces 9, etc.,
- the sum of the classification scores of the sentences in d that are classified to be positive relevant,
- the sum of the classification scores of the sentences in a document d that are classified to be negative relevant,
- average score of classified positive relevant sentences in d,
- average score of classified negative relevant sentences in d,
- the ratio of the number of the classified positive relevant sentences in d, to the number of the classified negative relevant sentences in d,
- the ratio of the sum of the scores of the classified positive relevant sentences in d to the sum of the scores of the classified negative relevant sentences in d,

Hui Yang and Si (2006) used a passage-based retrieval approach and retrieved 1,000 passages for each query. Logistic regression was used to predict the subjectivity of each sentence in a passage. Logistic regression model was trained using the Pang and Lee movie review data (Pang and Lee 2005; Pang et al. 2002) and Hu and Liu (2004) customer review data. Similarly Zhang et al. (2007) calculate the subjectivity score of each sentence using a CME classifier trained on movie review data (Pang and Lee 2005; Pang et al. 2002) using unigram, bigram features of a sentence. SVM classifier then predicts the opinion score of each blogpost on behalf of the subjective sentences contained in a blogpost. Almost similar kind of approach was used by Robin Anil (2008) for opinion finding task using movie review data of Pang et al. with other data sources with Nave Bayes Classifier. Table 3 provides an overview of different machine learning classifiers used in several opinion finding approaches.

3.4.4 Role of external data collections as a tool for query expansion and training data

Many opinion finding approaches seek help of some external data collection whether for query expansion or for training the classifier for opinion detection task. An external data collection means the data collection other than the one used for evaluation of an approach. The most common and popular data collections used for training the machine learning classifiers are movie review data provided by Pang and Lee (2005) and Pang et al. (2002) and customer Review Data provided by Hu and Liu (2004). The movie review data includes 5,000 subjective sentences and 5,000 objective

Table 3 Document-level summarization of related work in context of external data collections and ML-classifiers used

Title of the paper	ML-classifier	Data collection
Customizing sentiment classifiers to new domains: a case study (Aue and Gamon 2005)	Naive-Bayes and Support Vector Machines	Pang and Lee (2004) movie review data set (2,000 reviews), book review data (100 positive, 100 negative reviews), Product Support Services web survey data (2,564 examples of positive and 2,371 examples of negative feedback), Knowledge Base web survey data (6,035 examples of ad feedback and 6,285 examples of good feedback)
The sentimental factor: improving review classification via human-provided information (Beineke et al. 2004)	Naive-Bayes	Pang et al. (2002) movie review data set (1,400 reviews)
Sentiment classification of movie reviews using contextual valence shifters (Alistair 2006)	Support Vector Machines	Pang and Lee (2004) movie review data set (2,000 reviews)
Which side are you on? identifying perspectives at the document and sentence levels (Lin et al. 2006)	Naive-Bayes	http://www.bitterlemons.org 591 articles
Sentiment classification using word sub-sequences and dependency sub-trees (Matsumoto et al. 2005)	Support Vector Machines	Pang et al. (2002) movie review data set (1,400 reviews) and Pang and Lee (2004) movie review data set (2,000 reviews)
Sentiment analysis using support vector machines with diverse information sources (Mullen and Collier 2004)	Support Vector Machines	Pang et al. (2002) movie review data set (1,400 reviews)
A sentimental education: sentiment analysis using subjectivity summarization based on minimum cuts (Pang and Lee 2004)	Naive-Bayes, SVM	5,000 movie reviews nippets (e.g., bold, imaginative, and impossible to resist?) from http://www.rottentomatoes.com 5,000 sentences from plot summaries available from the Internet Movie Database (http://www.imdb.com)
Thumbs up? Sentiment classificationusing machine learning techniques (Pang et al. 2002)	(Naive Bayes, maximum entropyclassification, and support vector machines	Pang et al. (2002) movie review data set (1,400 reviews)
Using emotions to reduce dependency in machine learning techniques for sentiment classification (Read 2005)	Naive-Bayes, SVM	Pang et al. (2002) movie review data set (1,400 reviews), Internet Movie Review Database archive of movie reviews, Emoticon corpus
Automatic opinion polarity classification of movie reviews (Salvetti et al. 2004)	Naive Bayes and Markov Model	Pang et al. (2002) movie review data set (1,400 reviews)
Using appraisal taxonomies for sentiment analysis (Whitelaw et al. 2005)	SVM	Pang and Lee (2004) movie review data set (2,000 reviews)
Sentiment extraction from unstructured textusing tabu search-enhanced Markov blanket (Bai et al. 2004)	Markov Blanket Classifier, SVM, Naive-Bayes, Max. Entropy, voted Perceptron	Pang et al. (2002) movie review data set (1,400 reviews)
Automatic extraction of opinion propositions and their holders (Bethard et al. 2004)	Naive-Bayes	FrameNet: a corpus of over 100,000 sentences, PropBank: a million word corpus
Mining the peanut gallery: opinion extraction and semantic classification of product reviews (Dave et al. 2003)	Naive-Bayes	C-Net and Amazon customer reviews

sentences. The subjective sentences are sentences expressing opinions about a movie. The objective sentences are descriptions or the storytelling of a movie. The customer review data contains 4,258 sentences in total with 2,041 positive examples and 2,217 negative examples. The customer reviews are from http://www.amazon.com about 5 electronic products including digital cameras, DVD players and jukeboxes. Table 3 gives an overview of role of different data collections in different opinion finding approaches.

Besides being used as training data for classifiers, these external sources have been used for expanding the queries or for generating a list of opinionated words (individual terms or phrases) (Kiduk Yang and Yu 2007; Li et al. 2008). In addition to these data collection, there are others too which have been providing support for several approaches for opinion detection. Few of them are listed below:

– Yahoo Movie Review Data (used in Zhang and Zhang 2006)
– Epinion Digital Camera Review data (used in Zhang and Zhang 2006)
– Reuters Newswire Data (used in Zhang and Zhang 2006])
– Reviews from http://www.Rateitall.com (used in Zhang and Yu 2006, 2007)
– Reviews from http://www.amazon.com (used in Seki et al. 2007)
– AQUAINT-2 news corpus (used in Ernsting et al. 2007; He et al. 2007)
– Internet Movie Database plot summaries (used in Robin Anil 2008; Yang 2008)
– Reviews from Rotten Tomatoes (used in Robin Anil 2008)

It is hard to conclude that which external data source has performed well because no data collection has as such given distinctive results consistently. Therefore, we believe that it is not the type of data collection which improves the system's performance but more the way that data collection is being used. After an analysis of the top performing opinion finding approaches, it can be concluded that systems using data collections as a way to expanding the given query or creating an opinion lexicons have performed well.

3.4.5 Role of relevance feedback

A general overview of opinion finding approaches reveals an interesting observation about the use of relevance feedback. If we look at the topmost effective opinion finding approaches then it can be noted that most of the top performing approaches (Ernsting et al. 2007; Lee et al. 2008; Mishne 2006; Weerkamp and de Rijke 2008; Zhang and Yu 2006, 2007) have benefited from the use of Pseudo Relevance Feedback on topical retrieval step to have improved topic relevance MAP. Knowing already that the performance of the opinion finding task is dominated by the performance of topic relevance task, it can be suggested that use of Pseudo Relevance Feedback at retrieval step can influence the performance of opinion finding phase.

Lexicon-based approaches and machine learning approaches are the two most effective type of approaches for document level opinion finding task. While subjectivity lexicons could yield into some effective domain-independent approaches by providing a set of opinionated words, machine learning-based approaches cannot provide this level of domain independency. A machine learning classifier when trained on a data of one domain is most likely not to perform well when applied across domain data set. Machine learning approaches, however, are well tuned to give better performance under given circumstances but these approaches require large amount of data for training their classification models.

4 Challenges for opinion mining

Most of the opinion detection approaches model the presence of subjective words in a given document. They use several methods to identify subjective words and process this information to identify and retrieve opinionated sentences or documents (as discussed above). However, proposing approaches that can process subjective information effectively requires overcoming a number of challenges. In this section, we discuss the major problems that researchers working in this domain are facing.

4.1 Identifying comparative sentences

Use of comparative sentences is very common while expressing one's opinion especially when writing product reviews. For example, the comparative sentences, *Mobile phone A is better than B* and *Mobile phone B is better than A*, convey total opposite opinions. To understand well the meanings of these comparative phrases, an effective modeling of sequential information and discourse structure is required. Product reviews are generally subjective but, on the other hand, comparisons can be subjective or objective. Jindal and Liu (2006) explain this by giving the following examples of an opinion sentence, a subjective comparison sentence and an objective comparison sentence as shown in Table 4.

Identification of comparison sentences is challenging because although there are few indicators that can help to identify such sentences (i.e. comparative adverbs and comparative adjectives like *better, longer,more*, etc.) such indicators are also present in sentences that are not comparative, e.g., *I do not love you any more*. Similarly, many

Table 4 A comparison of opinion, subjective comparative and objective comparative sentences

Car X is very ugly	Opinion sentence
Car X is much better than Car Y	Subjective comparison
Car X is 2 feet longer than Car Y	Objective comparison

sentences that do not contain such indicators are comparative sentences, e.g., *Cellphone X has Bluetooth, but cellphone Y does not have* (Jindal and Liu 2006). Jindal and Liu (2006) take a data mining approach to identify the comparison sentences. They use class sequential rule (CSR) mining with supervised learning approach to identify comparative sentences in customer reviews, forum discussions, and news articles. Their approach successfully identifies almost all of the comparative sentences with precision of 79% and recall of 81%. Hou and Li (2008) applied another data mining technique, *Conditional Random Fields* (CRF) with Semantic Role Labeling (SRL), a statistical machine learning technique (Gildea and Jurafsky 2002) and achieved maximum precision of 93% for recognizing and labeling of comparative predicates.

4.2 Leveraging domain-dependency

The performance of effective opinion mining approaches (Blitzer et al. 2007; Engstrm 2004; Hui Yang and Si 2006; Owsley et al. 2006; Read 2005) differ from domain to domain (Liu 2007). For example, the opinion finding approach of Seki and Uehara (2009) performed exceptionally well for "products" related topics while it fails to give good results for topics of type "politics" and "organization". The one major and obvious reason is the difference in vocabularies across different domains. Developing domain-based approaches (or topic-based approaches) might give an edge as far as their performance is concerned but this performance is achieved on cost of its generalization. On the other hand, a domain-independent approach (or topic-independent approaches) is more generalized but might suffer from low performance. Therefore, developing an opinion finding approach that maintains its generalization and gives better performance is a big challenge for researchers working in this domain.

There exist a lot of work in the literature for both kind of approaches. Owsley et al. (2006) show the importance of building a domain-specific classifier. Read (2005) reports that standard machine learning techniques for opinion analysis are domain-dependent (with domains ranging from movie reviews to newswire articles). Na et al. (Na et al. 2009) proved that building a query-specific subjectivity lexicons helps improving the results for opinion finding task.

Similarly, there exist some approaches that exploit domain-independent features for the task of opinion mining. Hui Yang and Si (2006) propose a set of domain-independent features and performed evaluations on movie and product domains. Blitzer et al. (2007) explicitly address the domain transfer problem for sentiment classification by achieving an average of 46% improvement over a supervised baseline of product reviews. Domain-independent approach of Zhang and Zhang (2006), Wang et al. (2008), Liao et al. (2006), Mishne (2006) and Seki et al. (2007) perform well with the help of different machine learning techniques.

4.3 Opinion–topic association

A document can contain information about many topics and might have opinions on many of them too. In this situation, determining the opinion on a given topic requires a very effective approach which should not only separate opinionated information from factual information but also look for opinion–topic associations in the documents. Processing the documents on sentence and passage level might be a good idea to help solve this problem of finding opinion–topic associations. Various techniques have been used in the past to find this association between the given topic and the corresponding opinion; here we will discuss some prominent work done in this regard.

Natural Language Processing (NLP) techniques (like *POS Tagging* and *Syntactic Parsing*) have been used to identify opinion expressions and analyze their semantic relationships with the topic (Nasukawa and Yi 2003; Yi et al. 2003). POS tagging can be helpful to disambiguate polysemous expressions (such as the word *like*) which assists in identifying the correct sense of an ambiguous word to relate opinion expression with the topical terms. Similarly, syntactic parsing is used to identify relationships between sentiment expressions and the subject term. Besides NLP techniques, there exist approaches (like Attardi and Simi 2006; Java et al. 2006; Santos et al. 2009) that have been using proximity-based techniques for finding the opinion–topic associations in textual documents. For example, Santos et al. (2009) hypothesized that the proximity of the query terms to the subjective sentences in the document helps to find that level of opinion–topic association necessary for opinion finding task.

Relative to approach proposed by Java et al. (2006), simpler proximity approaches were adopted by Java et al. (2006) and Attardi and Simi (2006) where they just check for occurrences of opinionated terms around the query terms. However, comparison between all these approaches is not possible because all of these approaches use different data collections and baselines. Similarly, a comparison of results for approaches of Java et al. (2006) and Attardi and

Simi (2006) cannot be justified because both approaches use different topic-relevance baselines. Another very effective technique used for finding opinion–topic associations is to process only selective segments of a document (i.e., sentences or passages) instead of the whole document. The approaches adopted by Hui Yang and Si (2006), Missen et al. (2010) and Lee et al. (2008) use passages for this purpose.

4.4 Feature-based opinion mining

A document might be overall positive about a certain topic while it may also contain some negative opinions about few aspects of the topic. For example in review of a digital camera, a reviewer might be overall satisfied with the camera, but it is possible that he is not happy with one or two of its features (e.g., size of the screen or optical zoom). Feature-based opinion mining is considered a big challenge for opinion mining and it involves two tasks, *Feature Extraction* and *Feature-Sentiment Association*. To explain these tasks, let us take the example of the following sentence: *I love picture quality of this camera*. In this sentence, *picture quality* is a product feature and *love* is the sentiment associated with it. If a feature appears in subjective text, it is called *explicit feature*. If a feature appears in text other than subjective and is implied then it is called *implicit feature*. For instance, in sentence given above, the feature *picture quality* is an explicit feature while *size* is an implicit feature in the sentence given below as it does not appear in the sentence, but is implied (Liu 2007): *This camera is too large*. Mining implicit feature is harder than mining explicit feature because the feature word is not explicitly mentioned in the text. Zhuang et al. (2006) use a dependency grammar graph to extract explicit features and they defined classes of movie domain-related features with a set of opinion words allocated to each class for extraction of implicit features. Therefore, when such opinion word is found in a sentence, corresponding feature class can be decided even when a feature word is not mentioned in the sentence.

Yi et al. (2003) proposed two feature term selection algorithms based on a mixture language model and likelihood ratio. Likelihood Test method performed better than language model in their experimentation. Liu et al. (2005) proposed a method to extract product features from product reviews (pros and cons) based on association rules. They use association mining system *CBA (Classification based on Association)* (Hu et al. 1999) to perform this task. However, association rule mining is not suitable for this task because association rule mining is unable to consider the sequence of words, which is very important in natural language texts. Thus, many complex ad hoc post-processing methods are used in order to find patterns to extract features. Hu and Liu (2006) propose a more principled mining method based on sequential pattern mining. In particular, they mine a special kind of sequential patterns called *Class Sequential Rules* (CSR). As its name suggests, the sequence of words is considered automatically in the mining process. Unlike standard sequential pattern mining, which is unsupervised, they mine sequential rules with some fixed targets or classes. Thus, the new method is supervised. If we compare the results of work by Hu and Liu (2006) with work of Liu et al. (2005), we observe that the technique proposed by Hu and Liu (2006) generates comparable results as the association rules of Liu et al. (2005). The work of Liu et al. (2005) was further improved by Popescu and Etzioni (2005) by proposing an algorithm based on PMI method (Eq. 1).

4.5 Contextual polarity of words

An accurate identification of polarity of words requires a deep analysis of their contexts. The prior polarity of a word is always subject to changes under the context defined by its surrounding words. The new polarity of the word defined by its context is called its *contextual polarity*. Let us take an example to understand contextual polarity:

Information Secretary of National Environment **Trust**, *Robin Hood, said that Ricky is not a* **good** *guy*.

Although the word *trust* has many senses that express a positive sentiment, in above sentence, the word *trust* is not being used to express a sentiment at all and is part of the name of the organization *National Environment Trust*. In other words, the contextual polarity of the word *trust* is neutral in this case relative to its prior polarity which is generally positive. Similarly because of the presence of negation word *not* just before the word *good* which is positive in its prior polarity, the contextual polarity of word *good* is negative.

The context can be defined by negations (like *not good, never right*, etc), by word senses (like the word *plant* can be used as *nuclear plant* or *biological plant*), by the syntactic role of words around the given word (like *killers* versus *they are killers*), by intensifiers (like *very beautiful*), by diminishers (like *little problem*), or even by the domain/topic (like *unpredictable movie plot* is positive while *unpredictable camera functions* is negative) (Wilson 2008). Polanyi and Zaenen (2006) give a detailed discussion of many of the above types of polarity influencers.

There exist few works that have proposed approaches to identify the contextual polarities in opinion expressions (Popescu and Etzioni 2005; Suzuki et al. 2006; Yi et al. 2003). Yi et al. (2003) used a lexicon and manually developed high-quality patterns to classify contextual polarity. Their approach shows good results with high precision over the set of expressions that they evaluated.

Popescu and Etzioni (2005) use an unsupervised classification technique called *relaxation labelling* (Hummel 1987) to recognize the contextual polarity of words. They adopt a three-stage iterative approach to assign final polarities to words. They use features that represent conjunctions and dependency relations between polarity words. Suzuki et al. (2006) use a bootstrapping approach to classify the polarity of tuples of adjectives and their target nouns in Japanese blogs. Negations (only *not*) were taken into account when identifying contextual polarities.

The problem with above approaches is their limitation to specific items of interest, such as products and product features, or to tuples of adjectives and nouns. In contrast, the approach proposed by Wilson et al. (2005b) seeks to classify the contextual polarity of all instances of the words in a large lexicon of subjectivity clues that appear in the corpus. Included in the lexicon are not only adjectives, but nouns, verbs, adverbs, and even models. They dealt with negations on both local and long-distance levels. Besides this they also include clues from surrounding sentences. It was the first work to evaluate the effects of neutral instances on the performance of features for discriminating between positive and negative contextual polarities.

4.6 Use of social features for opinion detection

With the spread of opinionated content in online social networks, the latter have become an important source of opinions. It does not only provide researchers with an opportunity to have a huge amount of real-world opinion data but also a chance to exploit the social and networked structure of these networks for the task of opinion detection. However, identifying potential social evidences in online social networks (like blogosphere) and implementing them for the task of opinion detection remain a big challenge for researchers working in this domain.

Most of the related work for opinion mining in blogs have been using content-based evidences (Soboroff et al. 2007, 2008; Ounis et al. 2006). However, there exist some works who have exploited the network structure of blogosphere (Hui and Gregory 2010; Kale et al. 2007; Song et al. 2007). Song et al. (2007) propose an algorithm *InfluenceRank* to identify the most influential leaders within blogosphere. This algorithm ranks blogs according to their importance in the network and how novel is the information they provide. The top blogs ranked by *InfluenceRank* tend to be more influential and informative in the network and thus are more likely to be opinion leaders. Another work regarding quantification of trust and influence in the blogosphere includes the contribution by Kale et al. (2007). Their approaches use the link structure of a blog graph to associate sentiments with the links connecting two blogs. Sentiments associated with the links are named as *link polarities*. Similarly, Hui and Gregory (2010) also propose an approach for quantifying sentiment and influence in blogosphere. According to them, an influential blog

- has a non-trivial number of followers,
- generates a non-trivial amount of user feedback, in the form of comments on posts, and
- has a large proportion of posts on the topic being analyzed.

Most of the approaches that we have discussed so far regarding the use of social network features are limited to the tasks of quantification of trust and influence in blogosphere. There is a need of such approach that could exploit characteristics of blogosphere to perform opinion related tasks. In next section, we give brief overview of our work that aims at development of such framework.

5 Online social network features for opinion mining

Our work (reference masked because of blind review) is one of the most significant contributions regarding the task of opinion detection using social network evidences of the blogosphere. We propose a framework which focuses on the use of network of bloggers, blogpost content, blogger's profiles, and links between different blogposts for the tasks of opinion detection, opinion prediction, and multidimensional ranking. Even this contribution lacks experimental work but provides enough list of evidences for further work and improvements. We exploit four kinds of evidences for this work that includes trust, polarity, quality, and opinion score. For each type of variable, we provide a list of content-based and social network (i.e. blogosphere) based evidences.

6 Conclusion

In this paper, we discussed the related work for opinion mining in detail. We classified the work on word, sentence and document level in accordance with the opinion mining process. We also highlighted the TREC Blog track, its tasks and topics.

We have seen that most of the literature for opinion detection is prevailed by the lexicon-based approaches. These subjectivity lexicons whether are already available to the researchers (e.g., General Inquirer, SentiWordNet) or they are readily prepared from the target data collection or some external data collection are used for this task. Lexicon based on external data collections have played a very important role to improve the performance of opinion finding approaches, but this advantage is traded with loss of generalization of the approach because most of the external data collection used in several approaches are domain-specific. However, a very good choice is available now in the form of TREC Blog data collections that contain data from various domains ranging from sports to politics. Researchers are taking full advantage of these data collections by focusing on different challenges of the field of opinion mining. At the end of the paper, we discussed the related work in context of these major challenges. In this section, we discussed how different approaches have tried to tackle challenges of this field. This section concludes on need of exploiting features of online social networks for opinion mining. Last section describes the details of such framework which utilises various evidences of the blogosphere to perform different tasks.

References

Andreevskaia A, Bergler S (2006a) Mining wordnet for fuzzy sentiment: sentiment tag extraction from wordnet glosses. In: Proceedings of the 11th conference of the European Chapter of the Association for Computational Linguistics (EACL-06), pp 209–216

Andreevskaia A, Bergler S (2006b) Semantic tag extraction from wordnet glosses. In: Proceedings of the 5th conference on language resources and evaluation (LREC-2006)

Attardi G, Simi M (2006) Blog mining through opinionated words. In: Proceedings of TREC 2006, the fifteenth text retrieval conference, Gaithersburg, US. National Institute of Standards and Technlogy (NIST)

Aue A, Gamon M (2005) Customizing sentiment classifiers to new domains: a case study. In: RANLP-05: Proceedings of international conference on recent advances in natural language processing

Bai X, Padman R, Airoldi E (2007) Sentiment extraction from unstructured text using tabu search-enhanced Markov Blanket. In: Workshop on mining the semantic web 10th ACM SIGKDD conference (2004), pp 24–35

Baroni M, Vegnaduzzo S (2004) Identifying subjective adjectives through web-based mutual information. In: Proceedings of the 7th Konferenz zur Verarbeitung Natrlicher Sprache (German Conference on Natural Language Processing) KONVENS04, pp 613–619

Beineke P, Hastie T, Vaithyanathan S (2004) The sentimental factor: improving review classification via human-provided information. In: Proceedings of the 42nd annual meeting on association for computational linguistics, ACL '04, Morristown, NJ, USA. Association for Computational Linguistics

Bermingham A, Smeaton A, Foster J, Hogan D (2008) DCU at the TREC 2008 blog track. In: Proceedings of the Text REtrieval Conference (TREC-2008), USA. National Institute of Standards and Technology (NIST)

Bethard S, Yu H, Thornton A, Hatzivassiloglou V, Jurafsky D (2004) Automatic extraction of opinion propositions and their holders. In: Shanahan JG, Wiebe J, Qu Y (eds) Proceedings of the AAAI Spring symposium on exploring attitude and affect in text: theories and applications, pp 22–24, Stanford, USA

Blitzer J, Dredze M, Pereira F (2007) Biographies, bollywood, boomboxes and blenders: domain adaptation for sentiment classification. In: Proceedings of the 45th annual meeting of the association of computational linguistics, Prague, Czech Republic. Association for Computational Linguistics, pp 440–447

Bloom K, Stein S, Argamon S (2007) Appraisal extraction for news opinion analysis at NTCIR-6

Bobicev V, Maxim V, Prodan T, Burciu N, Anghelus V (2010) Emotions in words: developing a multilingual wordnet-affect. In: Gelbukh A (ed) Computational linguistics and intelligent text processing. Lecture notes in computer science, vol 6008. Springer, Berlin, pp 375–384

Boiy E, Moens M-F (2009) A machine learning approach to sentiment analysis in multilingual web texts. Inf Retrieval 12:526–558. doi: 10.1007/s10791-008-9070-z

Bruce RF, Wiebe JM (1999) Recognizing subjectivity: a case study in manual tagging. J Nat Lang Eng 5(2):187–205

Cacheda F, Plachouras V, Ounis I (2005) A case study of distributed information retrieval architectures to index one terabyte of text. J Inf Process Manag 41(5):1141–1161

Church KW, Hanks P (1990) Word association norms, mutual information, and lexicography. Comput Linguist 16(1):22–29

Claire Fautsch JS (2008) UniNE at TREC-2008: fact and opinion retrieval in the blogsphere. In: Proceedings of the text retrieval conference (TREC-2008), USA. National Institute of Standards and Technology (NIST)

Dagan I, Marcus S, Markovitch S (1993) Contextual word similarity and estimation from sparse data. In: Proceedings of ACL-93, pp 164–171

Dagan I, Pereira F, Lee L (1994) Similarity-based estimation of word cooccurrence probabilities. In: Proceedings of the 32nd annual meeting on Association for Computational Linguistics, ACL '94, Stroudsburg, PA, USA. Association for Computational Linguistics, pp 272–278

Dave K, Lawrence S, Pennock DM (2003) Mining the peanut gallery: opinion extraction and semantic classification of product reviews. In: Proceedings of WWW-03, 12th international conference on the World Wide Web, Budapest, HU. ACM Press, pp 519–528

Ding X, Liu B, Yu Philip S (2008) A holistic lexicon-based approach to opinion mining. In: WSDM '08: Proceedings of the international conference on Web search and web data mining, New York, NY, USA. ACM, pp 231–240

Engstrm C (2004) Topic dependence in sentiment classification. Master's thesis, University of Cambridge

Ernsting B, Weerkamp W, de Rijke M (2007) Language modeling approaches to blogpost and feed finding. In: Proceedings of the text retrieval conference (TREC-2007), USA. National Institute of Standards and Technology (NIST)

Esuli A, Sebastiani F (2005) Determining the semantic orientation of terms through gloss classification. In: Proceedings of fourteenth conference on information and knowledge management (CIKM-05), pp 617–624

Esuli A, Sebastiani F (2006) SentiWordNet: a publicly available lexical resource for opinion mining. In: Proceedings of the 5th conference on language resources and evaluation (LREC-06), pp 417–422

Gerani S, Keikha M, Carman M, Gwadera R, Taibi D, Crestani F (2008) University of lugano at TREC-2008 blog track. In: Proceedings of the 17th text retrieval conference (TREC-2008), USA, National Institute of Standards and Technology (NIST)

Gerani S, Carman MJ, Crestani F (2009) Investigating learning approaches for blog post opinion retrieval. In: ECIR '09: Proceedings of the 31th European conference on IR research on advances in information retrieval. Springer, Berlin, pp 313–324

Gildea D, Jurafsky D (2002) Automatic labeling of semantic roles. Comput Linguist 28(3):245–288

Grefenstette G, Qu Y, Shanahan JG, Evans DA (2004) Coupling niche browsers and affect analysis for an opinion mining. In: Proceedings of RIAO, pp 186–194

Grefenstette G, Qu Y, Evans D, Shanahan J (2006) Validating the coverage of lexical resources for affect analysis and automatically classifying new words along semantic axes. In: Shanahan J, Qu Y, Wiebe J (eds) Computing attitude and affect in text: theory and applications. Information retrieval series, vol 20. Springer, Dordrecht, pp 93–107

GuangXu Z, Joshi H, Bayrak C (2007) Topic categorization for relevancy and opinion detection. In: Proceedings of the text retrieval conference (TREC-2007), USA. National Institute of Standards and Technology (NIST)

Hatzivassiloglou V, McKeown KR (1997) Predicting the semantic orientation of adjectives. In: Proceedings of the eighth conference on European chapter of the Association for Computational Linguistics, Morristown, NJ, USA. Association for Computational Linguistics, pp 174–181

Hatzivassiloglou V, Wiebe JM (2000) Effects of adjective orientation and gradability on sentence subjectivity. In: Proceedings of the 18th conference on computational linguistics, Morristown, NJ, USA. Association for Computational Linguistics, pp 299–305

Hatzivassiloglou V, Klavans JL, Holcombe ML, Barzilay R, Kan M-y, McKeown KR (2001) Simfinder: a flexible clustering tool for summarization. In: Proceedings of the NAACL workshop on automatic summarization, pp 41–49

He B, Macdonald C, Ounis I, Peng J, Santos RLT (2007) University of glasgow at TREC-2007: experiments in blog and enterprise tracks with terrier. In: Proceedings of the 16th text retrieval conference (TREC-2007), USA. National Institute of Standards and Technology (NIST)

He B, Macdonald C, He J, Ounis I (2008a) An effective statistical approach to blog post opinion retrieval. In: CIKM '08: Proceeding of the 17th ACM conference on Information and knowledge management, New York, NY, USA. ACM, pp 1063–1072

He B, Macdonald C, Ounis I, Peng J, Santos RLT (2008b) University of glasgow at TREC-2008: experiments in blog and enterprise tracks with terrier. In: Proceedings of the text retrieval conference (TREC-2008), USA. National Institute of Standards and Technology (NIST)

He H, Chen B, Du L, Li S, Gao H, Xu W, Guo J (2008c) PRIS in TREC 2008 Blog Track. In: Proceedings of the text retrieval conference (TREC-2008), USA. National Institute of Standards and Technology (NIST)

Hemant Joshi CB, Xu X (2006) UALR at TREC blog track. In: Proceedings of the text retrieval conference (TREC-2006), USA. National Institute of Standards and Technology (NIST)

Hou F, Li G-H (2008) Mining chinese comparative sentences by semantic role labeling. In: Proceedings of international conference on machine learning and cybernetics, 2008, pp 2563–2568

Hu M, Liu B (2004) Mining and summarizing customer reviews. In: Proceedings of KDD '04, the ACM SIGKDD international conference on knowledge discovery and data mining, Seattle, US. ACM, pp 168–177

Hu M, Liu B (2006) Opinion feature extraction using class sequential rules. In: Proceedings of the Spring symposia on computational approaches to analyzing weblogs

Hu K, Lu Y, Zhou L, Shi C (1999) Integrating classification and association rule mining: a concept lattice framework. In: Proceedings of the 7th international workshop on new directions in rough sets, data mining, and granular-soft computing, RSFDGrC '99, London, UK. Springer, pp 443–447

Hui P, Gregory M (2010) Quantifying sentiment and influence in blogspaces. In: 1st Workshop on social media analytics, SOMA-10, Washington, DC, USA. ACM

Hui Yang JC, Si L (2006) Knowledge transfer and opinion detection in the trec2006 blog track. In: Proceedings of the text retrieval conference (TREC-2006), USA. National Institute of Standards and Technology (NIST)

Hummel RA, Zucker SW (1987) On the foundations of relaxation labeling processes. Morgan Kaufmann, San Francisco, pp 585–605

Java A, Kolari P, Finin T, Joshi A, Martineau J, Mayfield J (2006) The blogvox opinion retrieval system. In: Proceedings of the text retrieval conference (TREC-2006), USA. National Institute of Standards and Technology (NIST)

Jia L, Yu C, Zhang W (2008) UIC at TREC 2008 blog track. In: Proceedings of the 17th text retrieval conference (TREC-2008), USA. National Institute of Standards and Technology (NIST)

Jindal N, Liu B (2006) Identifying comparative sentences in text documents. In: SIGIR '06: Proceedings of the 29th annual international ACM SIGIR conference on Research and development in information retrieval, New York, NY, USA. ACM, pp 244–251

Kale A, Kar A, Kolari P, Java A, Finin T, Joshi A (2007) Modeling trust and influence in the blogosphere using link polarity. In: Proceedings of the international conference on weblogs and social media (ICWSM)

Kamps J, Marx M, Mokken R, de Rijke M (2004) Using wordnet to measure semantic orientation of adjectives. In: Proceedings of the 4th international conference on language resources and evaluation (LREC'04), pp 1115–1118

Kennedy A, Inkpen D (2006) Sentiment classification of movie reviews using contextual valence shifters. Comput Intell 22(2): 110–125

Kiduk Yang HZ, Yu N (2007) WIDIT in TREC-2007 blog track: combining lexicon-based methods to detect opinionated blogs. In: Proceedings of the text retrieval conference (TREC-2007), USA. National Institute of Standards and Technology (NIST)

Kim S-M, Hovy E (2004) Determining the sentiment of opinions. In: COLING '04: Proceedings of the 20th international conference on computational linguistics, Morristown, NJ, USA. Association for Computational Linguistics, pp 1367–1373

Ku L-W, Chen H-H (2007) Mining opinions from the web: beyond relevance retrieval. J Am Soc Inf Sci Technol 58(12):1838–1850

Ku LW, Liang YT, Chen HH (2006) Opinion extraction, summarization and tracking in news and blog corpora. In: Proceedings of AAAI-2006 Spring symposium on computational approaches to analyzing weblogs

Leacock C, Chodorow M (1998) Combining local context and WordNet similarity for word sense identification. In: Fellbaum C (ed) MIT Press, pp 305–332

Lee Y, Na S-H, Kim J, Nam S-H, Jung H-y, Lee J-H (2008) Kle at trec 2008 blog track: Blog post and feed retrieval. In: Proceedings of the text retrieval conference (TREC-2008), USA. National Institute of Standards and Technology (NIST)

Li B, Liu F, Liu Y (2008) UTDallas at TREC 2008 blog track. In: Proceedings of the text retrieval conference (TREC-2008), USA. National Institute of Standards and Technology (NIST)

Liao X, Cao D, Tan S, Liu Y, Ding G, Cheng X (2006) Combining language model with sentiment analysis for opinion retrieval of blog-post. In: Proceedings of the text retrieval conference (TREC-2006), USA. National Institute of Standards and Technology (NIST)

Lin W-H, Wilson T, Wiebe J, Hauptmann A (2006) Which side are you on?: identifying perspectives at the document and sentence levels. In: Proceedings of the tenth conference on computational natural language learning, CoNLL-X '06, Morristown, NJ, USA. Association for Computational Linguistics, pp 109–116

Liu B (2007) Web data mining: exploring hyperlinks, contents, and usage data (Data-Centric Systems and Applications), 1st edn. corr. 2nd printing edn. Springer, Berlin

Liu B (2008) Opinion mining and summarization-sentiment analysis. In: Tutorial in the proceedings of WWW'08. ACM, New York

Liu B, Hu M, Cheng J (2005) Opinion observer: analyzing and comparing opinions on the web. In: Proceedings of WWW '05,

the 14th international conference on World Wide Web, Chiba. ACM Press, pp 342–351
Magnini B, Cavagli G (2000) Integrating subject field codes into wordnet. In: Proceedings of international conference on language resources & evaluation, pp 1413–1418
Matsumoto S, Takamura H, Okumura M (2005) Sentiment classification using word sub-sequences and dependency sub-trees. In: Ho TB, Cheung D, Liu H (eds) Advances in knowledge discovery and data mining. Lecture notes in computer science, vol 3518. Springer, Berlin, pp 301–311
Miller GA (1995) Wordnet: a lexical database for english. Commun ACM 38(11):39–41
Miller GA, Charles WG (1991) Contextual correlates of semantic similarity. Lang Cognit Process 6(1):1–28
Mishne G (2006) Multiple ranking strategies for opinion retrieval in blogs. In: Proceedings of the text retrieval conference (TREC-2006), USA. National Institute of Standards and Technology (NIST)
Missen MMS, Boughanem M (2009) Sentence-level opinion-topic association for opinion detection in blogs. In: IEEE International symposium on mining and web, Bradford, UK. IEEE Computer Society, pp 733–737
Missen MMS, Boughanem M, Cabanac G (2010) Using passage-based language model for opinion detection in blogs. In: Annual ACM symposium on applied computing (SAC), Sierre, Switzerland. ACM, pp 1821–1822
Mukras R, Wiratunga N, Lothian R (2007) The Robert Gordon University at the opinion retrieval task of the 2007 trec blog track. In: Proceedings of the 16th text retrieval conference (TREC-2007), USA. National Institute of Standards and Technology (NIST)
Mullen T, Collier N (2004) Sentiment analysis using support vector machines with diverse information sources. In: Lin D, Wu D (eds) Proceedings of EMNLP 2004, Barcelona, Spain, July 2004. Association for Computational Linguistics, pp 412–418
Na S-H, Lee Y, Nam S-H, Lee J-H (2009) Improving opinion retrieval based on query-specific sentiment lexicon. In: ECIR '09: Proceedings of the 31th European conference on IR research on advances in information retrieval. Springer, Berlin, Heidelberg, pp 734–738
Nasukawa T, Yi J (2003) Sentiment analysis: capturing favorability using natural language processing. In: K-CAP '03: Proceedings of the 2nd international conference on knowledge capture. ACM, New York, pp 70–77
Nasukawa T, Yi J (2003) Sentiment analysis: capturing favorability using natural language processing. In: K-CAP '03: Proceedings of the 2nd international conference on knowledge capture. ACM, New York, pp 70–77
Ng V, Dasgupta S, Niaz Arifin SM (2006) Examining the role of linguistic knowledge sources in the automatic identification and classification of reviews. In: Proceedings of COLING/ACL 2006 main conference poster sessions, pp 611–618
Nigam K, McCallum AK, Thrun S, Mitchell T (2000) Text classification from labeled and unlabeled documents using EM, vol 39, Hingham, MA, USA, May 2000. Kluwer, pp 103–134
Ounis I, de Rijke M, Macdonald C, Mishne G, Soboroff I (2006) Overview of the TREC blog track 2006. In: Proceedings of the 15th Text REtrieval Conference (TREC 2006)
Owsley S, Sood S, Hammond KJ (2006) Domain specific affective classification of documents. In: Proceedings of the AAAI-CAAW06, the spring symposia on computational approaches to analyzing weblogs, pp 181–183
Pang B, Lee L (2004) A sentimental education: sentiment analysis using subjectivity summarization based on minimum cuts. In: Proceedings of the 42nd annual meeting on Association for Computational Linguistics, ACL '04, Morristown, NJ, USA. Association for Computational Linguistics
Pang B, Lee L (2005) Seeing stars: exploiting class relationships for sentiment categorization with respect to rating scales. In: ACL '05: Proceedings of the 43rd annual meeting on Association for Computational Linguistics, Morristown, NJ, USA. Association for Computational Linguistics, pp 115–124
Pang B, Lee L (2008) Opinion mining and sentiment analysis. J Found Trends Inf Retrieval 2(1-2):1–135
Pang B, Lee L, Vaithyanathan S (2002) Thumbs up? Sentiment classification using machine learning techniques. In: EMNLP '02: Proceedings of the ACL-02 conference on empirical methods in natural language processing, Morristown, NJ, USA. Association for Computational Linguistics, pp 79–86
Pasca M (2005) Finding instance names and alternative glosses on the web: Wordnet reloaded. In: Gelbukh A (ed) Computational linguistics and intelligent text processing. Lecture notes in computer science, vol 3406. Springer, Berlin, pp 280–292
Picard RW (2002) Affective computing. User Model User-Adapted Interact 12:85–89
Polanyi L, Zaenen A (2006) Contextual valence shifters. Computing attitude and affect in text: theory and applications, vol 20, pp 1–10
Popescu A-M, Etzioni O (2005) Extracting product features and opinions from reviews. In: HLT '05: Proceedings of the conference on human language technology and empirical methods in natural language processing, Morristown, NJ, USA. Association for Computational Linguistics, pp 339–346
Read J (2005) Using emotions to reduce dependency in machine learning techniques for sentiment classification. In: ACLstudent '05: Proceedings of the ACL student research workshop, Morristown, NJ, USA. Association for Computational Linguistics, pp 43–48
Resnik P (1995) Using information content to evaluate semantic similarity in a taxonomy. In: Proceedings of the 14th international joint conference on artificial intelligence, vol 1, San Francisco, CA, USA. Morgan Kaufmann, pp 448–453
Riloff E, Wiebe J, Wilson T (2003) Learning subjective nouns using extraction pattern bootstrapping. In: Proceedings of the seventh conference on natural language learning at HLT-NAACL 2003, Morristown, NJ, USA. Association for Computational Linguistics, pp 25–32
Robin Anil SS (2008) IIT Kharagpur at TREC-2008 blog track. In: Proceedings of the text retrieval conference (TREC-2008), USA. National Institute of Standards and Technology (NIST)
Rui S, Qin T, Shi D, Lin H, Yang Z (2007) DUTIR at TREC-2007 blog track. In: Proceedings of the 16th text retrieval conference (TREC-2007), USA. National Institute of Standards and Technology (NIST)
Salvetti F, Lewis S, Reichenbach C (2004) Automatic opinion polarity classification of movie reviews. Colorado Res Ling 17(1)
Santos RLT, He B, Macdonald C, Ounis I (2009) Integrating proximity to subjective sentences for blog opinion retrieval. In: ECIR '09: Proceedings of the 31th European conference on IR research on advances in information retrieval. Springer, Berlin, pp 325–336
Sebastiani F, Esuli A, Sebastiani F (2006) 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 (EACL06)
Seki K, Uehara K (2009) Adaptive subjective triggers for opinionated document retrieval. In: WSDM '09: Proceedings of the second ACM international conference on web search and data mining, New York, NY, USA. ACM, pp 25–33
Seki K, Kino Y, Sato S, Uehara K (2007) Trec 2007 blog track experiments at kobe university. In: Proceedings of the text

retrieval conference (TREC-2007), USA. National Institute of Standards and Technology (NIST)
Soboroff I, Macdonald C, Ounis I (2007) Overview of the TREC blog track 2007. In: Proceedings of the 16th Text REtrieval Conference (TREC 2007)
Soboroff I, Macdonald C, Ounis I (2008) Overview of the TREC blog track 2008. In: Proceedings of the 17th Text REtrieval Conference (TREC 2008)
Song X, Chi Y, Hino K, Tseng B (2007) Identifying opinion leaders in the blogosphere. In: Proceedings of the sixteenth ACM conference on information and knowledge management, CIKM-07, pp 971–974
Stone PJ, Hunt EB (1963) A computer approach to content analysis: studies using the general inquirer system. In: AFIPS '63 (Spring): Proceedings of the May 21–23, 1963, spring joint computer conference, New York, NY, USA. ACM, pp 241–256
Stone PJ, Dunphy DC, Smith MS, Ogilvie DM (1966) The general inquirer: a computer approach to content analysis. MIT Press
Subasic P, Huettner A (2001) Affect analysis of text using fuzzy semantic typing. IEEE Trans Fuzzy Syst 9(4):483–496
Suzuki Y, Takamura H, Okumura M (2006) Application of semi-supervised learning to evaluative expression classification. In: Proceedings of CICLing-06, the 7th international conference on computational linguistics and intelligent text processing, Mexico City, pp 502–513
Tang H, Tan S, Cheng X (2009) A survey on sentiment detection of reviews. Expert Syst Appl 36(7):10760–10773
Turney P, Littman M (2002) Unsupervised learning of semantic orientation from a hundred-billion-word corpus. Technical report, Institute for Information Technology, Council Canada, Technical Report ERB-1094
Turney PD, Littman ML (2003) Measuring praise and criticism: Inference of semantic orientation from association. ACM Trans Inf Syst 21(4):315–346
Valitutti R (2004) Wordnet-affect: an affective extension of wordnet. In: Proceedings of the 4th international conference on language resources and evaluation, pp 1083–1086
Wang Y, Liu W, Tan S, Xu H, Liao X, Cao D, Cheng X (2007) Experiments in trec 2007 blog opinion task at CAS-ICT. In: Proceedings of the text retrieval conference (TREC-2007), USA. National Institute of Standards and Technology (NIST)
Wang J, Sun Y, Mukhtar O, Srihari R (2008) TREC 2008 at the university at buffalo: legal and blog track. In: Proceedings of the text retrieval conference (TREC-2008), USA. National Institute of Standards and Technology (NIST)
Weerkamp W, de Rijke M (2008) External query expansion in the blogosphere. In: Proceedings of the text retrieval conference (TREC-2008), USA. National Institute of Standards and Technology (NIST)
Whitelaw C, Garg N, Argamon S (2005) Using appraisal taxonomies for sentiment analysis. In: Proceedings of MCLC-05, the 2nd Midwest computational linguistic colloquium, Columbus, USA
Wiebe JM (1990) Identifying subjective characters in narrative. In: Proceedings of the 13th conference on computational linguistics, Morristown, NJ, USA. Association for Computational Linguistics, pp 401–406
Wiebe J (2000) Learning subjective adjectives from corpora. In: Proceedings of the seventeenth national conference on artificial intelligence and twelfth conference on innovative applications of artificial intelligence. AAAI Press, pp 735–740
Wiebe J, Wilson T, Bruce R, Bell M, Martin M (2004) Learning subjective language. Comput Linguist 30(3):277–308
Williams GK, Anand SS (2009) Predicting the polarity strength of adjectives using wordnet. In: Proceedings of international conference on weblogs and social media (ICWSM-09)
Wilson TA (2008) Fine-grained subjectivity and sentiment analysis: recognizing the intensity, polarity, and attitudes of private states. PhD thesis, Pittsburgh, PA, USA
Wilson T, Pierce DR, Wiebe J (2003) Identifying opinionated sentences. In: NAACL '03: Proceedings of the 2003 conference of the North American Chapter of the Association for Computational Linguistics on Human Language Technology, Morristown, NJ, USA. Association for Computational Linguistics, pp 33–34
Wilson T, Hoffmann P, Somasundaran S, Kessler J, Wiebe J, Choi Y, Cardie C, Riloff E, Patwardhan S (2005a) Opinionfinder: a system for subjectivity analysis. In: Proceedings of HLT/EMNLP on interactive demonstrations, Morristown, NJ, USA. Association for Computational Linguistics, pp 34–35
Wilson T, Wiebe J, Hoffmann P (2005b) Recognizing contextual polarity in phrase-level sentiment analysis. In: HLT '05: Proceedings of the conference on human language technology and empirical methods in natural language processing, Morristown, NJ, USA. Association for Computational Linguistics, pp 347–354
Yang K (2008) WIDIT in TREC-2008 blog track: leveraging multiple sources of opinion evidence. In: Proceedings of the text retrieval conference (TREC-2008), USA. National Institute of Standards and Technology (NIST)
Yi J, Nasukawa T, Bunescu R, Niblack W (2003) Sentiment analyzer: extracting sentiments about a given topic using natural language processing techniques. In: Proceedings of the third IEEE international conference on data mining (ICDM-03), Washington, DC, USA. IEEE Computer Society, pp 427–434
Yu H, Hatzivassiloglou V (2003) Towards answering opinion questions: separating facts from opinions and identifying the polarity of opinion sentences. In: Proceedings of EMNLP-03, pp 129–136
Zhang W, Yu C (2006) UIC at TREC 2006 blog track. In: Proceedings of the text retrieval conference (TREC-2006), USA. National Institute of Standards and Technology (NIST)
Zhang W, Yu C (2007) UIC at TREC 2007 blog track. In: Proceedings of the 16th text retrieval conference (TREC-2007), USA. National Institute of Standards and Technology (NIST)
Zhang E, Zhang Y (2006) UCSC on TREC 2006 blog opinion mining. In: Proceedings of the text retrieval conference (TREC-2006), USA. National Institute of Standards and Technology (NIST)
Zhang Y, Xu W, Callan J (2002) Exact maximum likelihood estimation for word mixtures. In: Proceedings of text learning workshop in international conference on machine learning (ICML2002)
Zhang Q, Wang B, Wu L, Huang X (2007) FDU at TREC-2007: opinion retrieval of blog track. In: Proceedings of the text retrieval conference (TREC-2007), USA. National Institute of Standards and Technology (NIST)
Zhang W, Jia L, Yu C, Meng W (2008) Improve the effectiveness of the opinion retrieval and opinion polarity classification. In: CIKM '08: Proceeding of the 17th ACM conference on information and knowledge management, New York, NY, USA. ACM, pp 1415–1416
Zhang Z, Ye Q, Law R, Li Y (2009) Automatic detection of subjective sentences based on chinese subjective patterns. In: Shi Y, Wang S, Peng Y, Li J, Zeng Y (eds) Cutting-edge research topics on multiple criteria decision making. Communications in computer and information science, vol 35. Springer, Berlin, pp 29–36
Zhao S, Luo Z, Lu W (2007) WHU at TREC blog track 2007. In: Proceedings of the 16th text retrieval conference (TREC-2007), USA. National Institute of Standards and Technology (NIST)
Zhuang L, Jing F, Zhu X-Y (2006) Movie review mining and summarization. In: CIKM '06: Proceedings of the 15th ACM international conference on Information and knowledge management, New York, NY, USA. ACM, pp 43–50