

Literature Survey

Vinita Sharma

123050055

June 29, 2014

Chapter 1

Sentiment Analysis

Sentiment Analysis (SA) is one of the most widely studied applications of Natural Language Processing (NLP) and Machine Learning (ML). This field has grown tremendously with the advent of the Web 2.0. The Internet has provided a platform for people to express their views, emotions and sentiments towards products, people and life in general. Thus, the Internet is now a vast resource of opinion rich textual data. The goal of Sentiment Analysis is to harness this data in order to obtain important information regarding public opinion, that would help make smarter business decisions, political campaigns and better product consumption. Sentiment Analysis focuses on identifying whether a given piece of text is subjective or objective and if it is subjective, then whether it is negative or positive.

Liu (2010) define sentiment or opinion as a quintuple-

“ $\langle o_j, f_{jk}, so_{ijkl}, h_i, t_l \rangle$, where o_j is a target object, f_{jk} is a feature of the object o_j , so_{ijkl} is the sentiment value of the opinion of the opinion holder h_i on feature f_{jk} of object o_j at time t_l , so_{ijkl} is +ve, -ve, or neutral, or a more granular rating h_i is an opinion holder, t_l is the time when the opinion is expressed.”

The recent trends in Sentiment Analysis techniques have moved towards building generative models that can capture complex contextual phenomena. Conversely, due to the unavailability of annotated data, the focus is moving towards unsupervised approaches that use the power of co-occurrence to solve the problem. Since, the web has a huge amount of opinionated data, in the form of blogs, reviews, *etc.*, the unsupervised approaches flourish.

1.1 Motivation

According to Ramteke et al. (2012) motivation for Sentiment Analysis is two-fold. Both consumers and producers highly value “customer’s opinion” about products and services. Thus, Sentiment Analysis has seen a considerable effort from industry as well as academia.

The Consumer's Perspective

While taking a decision it is very important for us to know the opinion of the people around us. Earlier this group used to be small, with a few trusted friends and family members. But, now with the advent of Internet we see people expressing their opinions in blogs and forums. These are now actively read by people who seek an opinion about a particular entity (product, movie etc.). Thus, there is a plethora of opinions available on the Internet.

From a consumers' point of view extracting opinions about a particular entity is very important. Trying to go through such a vast amount of information to understand the general opinion is impossible for users just by the sheer volume of this data. Hence, the need of a system that differentiates between good reviews and bad reviews. Further, labeling these documents with their sentiment would provide a succinct summary to the readers about the general opinion regarding an entity.

The Producer's Perspective

With the explosion of Web 2.0 platforms such as blogs, discussion forums, etc., consumers have at their disposal, a platform to share their brand experiences and opinions, positive or negative regarding any product or service. According to Pang and Lee (2008) these consumer voices can wield enormous influence in shaping the opinions of other consumers and, ultimately, their brand loyalties, their purchase decisions, and their own brand advocacy.

Since the consumers have started using the power of the Internet to expand their horizons, there has been a surge of review sites and blogs, where users can perceive a product's or service's advantages and faults. These opinions thus shape the future of the product or the service. The vendors need a system that can identify trends in customer reviews and use them to improve their product or service and also identify the requirements of the future.

The Societies' Perspective

Recently, certain events, which affected Government, have been triggered using the Internet. The social networks are being used to bring together people so as to organize mass gatherings and oppose oppression. On the darker side, the social networks are being used to insinuate people against an ethnic group or class of people, which has resulted in a serious loss of life. Thus, there is a need for Sentiment Analysis systems that can identify such phenomena and curtail them if needed.

1.2 Applications of Sentiment Analysis

Sentiment Analysis has many applications in various fields. According to Ramteke (2012) the application from a user's standpoint is the application related to review websites. Tools that help summarize the

sentiment regarding a product or service help users in identifying their product of choice. Similarly, vendors build tools that analyze customer feedback which help improve user experience. The future might see applications wherein a system gauges the human emotion through sensory means and then creates an environment that helps improve the human life in general. This section describes a few of these applications that have been built or are possibilities in the near future.

Applications to Review-related Websites

Today Internet has an entire gamut of reviews and feedbacks on almost everything. These include product reviews; feedbacks on political issues etc. Thus there is a need for a sentiment engine that can extract sentiments about a particular entity. It will provide a consolidated feedback or rating for the given topic. Such applications would not themselves contain any opinions, but they would fetch the opinionated text from various resources and provide an effective polarity. This would serve the need of both the users and the vendors.

Another application of Sentiment Analysis is in automatic summarization of user reviews. Automatic summarization is the creation of a summary of the entire review using an automated program. In case of user reviews, it is difficult for a new user to look at all the reviews thoroughly and understand what aspect of the product is not appreciated. Thus, there is a need of a summarizing application that will briefly inform the user about the polarity of the reviews, for example, thumbs up or thumbs down for the topic.

It is assumed that all user ratings are accurate. However, there are cases where users have accidentally selected a low rating when their review indicates a positive evaluation, or vice versa. Moreover, there is some evidence that the user ratings can be biased, based on a previous experience or otherwise in need of correction. Automated sentiment classifiers can help us correct such cases by identifying sentiments corresponding to the relevant features of the product.

Applications as a Sub-component Technology

A sentiment predictor system can be naturally considered to aid a recommender system. The recommender system will not recommend items that receive a lot of negative feedback.

In online communication a hostile and insulting interaction between Internet users is termed as “Flames”. This involves abusive language and other negative elements. These can thus be detected simply by identifying a highly negative sentiment.

While placing advertisements in sidebars it is important to understand the sensitivity of the users. A further improvement would be to detect the sentiment expressed in the page and thus bring up adver-

tisements relevant to the sentiment. For example, on a positive review about a product an advertisement about a related product from the same manufacturer will improve the sales. Conversely, if a negative sentiment is detected then an advertisement from a competitor would be appreciated.

Applications in Business Intelligence

It has been observed that more and more people nowadays tend to look upon reviews of products online before buying them. And for many businesses the online opinion can make or break their product. Thus, Sentiment Analysis finds an important role in businesses. Businesses wish to understand the online reviews in order to improve their products and in turn their reputation.

Sentiment Analysis can also be used in trend prediction. By tracking public opinion, vital data regarding sales trends and customer satisfaction can be extracted.

Applications across different Domains

So far we have mentioned only applications pertaining to a business setting. But, Sentiment Analysis finds various applications in other fields. Studies in sociology and other fields have been aided by Sentiment Analysis systems that show trends in human emotions especially on social networks.

Applications in smart homes

Smart homes are supposed to be the technology of the future. It is speculated by leading scientists in all fields that eventually the entire homes would be networked and people would be able to control any part of the home using a tablet device. In such homes, Sentiment Analysis would also find its place. Based on the current sentiment or emotion of the user, the home could alter its ambiance to create a soothing and peaceful environment.

1.3 Dimensions of Sentiment Analysis

Figure 1.3.1 shows the various dimensions of Sentiment Analysis. The tasks in Sentiment Analysis may be classified based on the complexity of the problem, the degree of detail required, the approaches used, *etc.* This section describes some of these tasks.

Tasks based on Classification

Identifying Subjectivity:

The basic question asked in Sentiment Analysis is whether a given piece of text contains any subjective content (opinions, emotions, *etc.*) or not. This task aims to tackle this problem of differentiating between

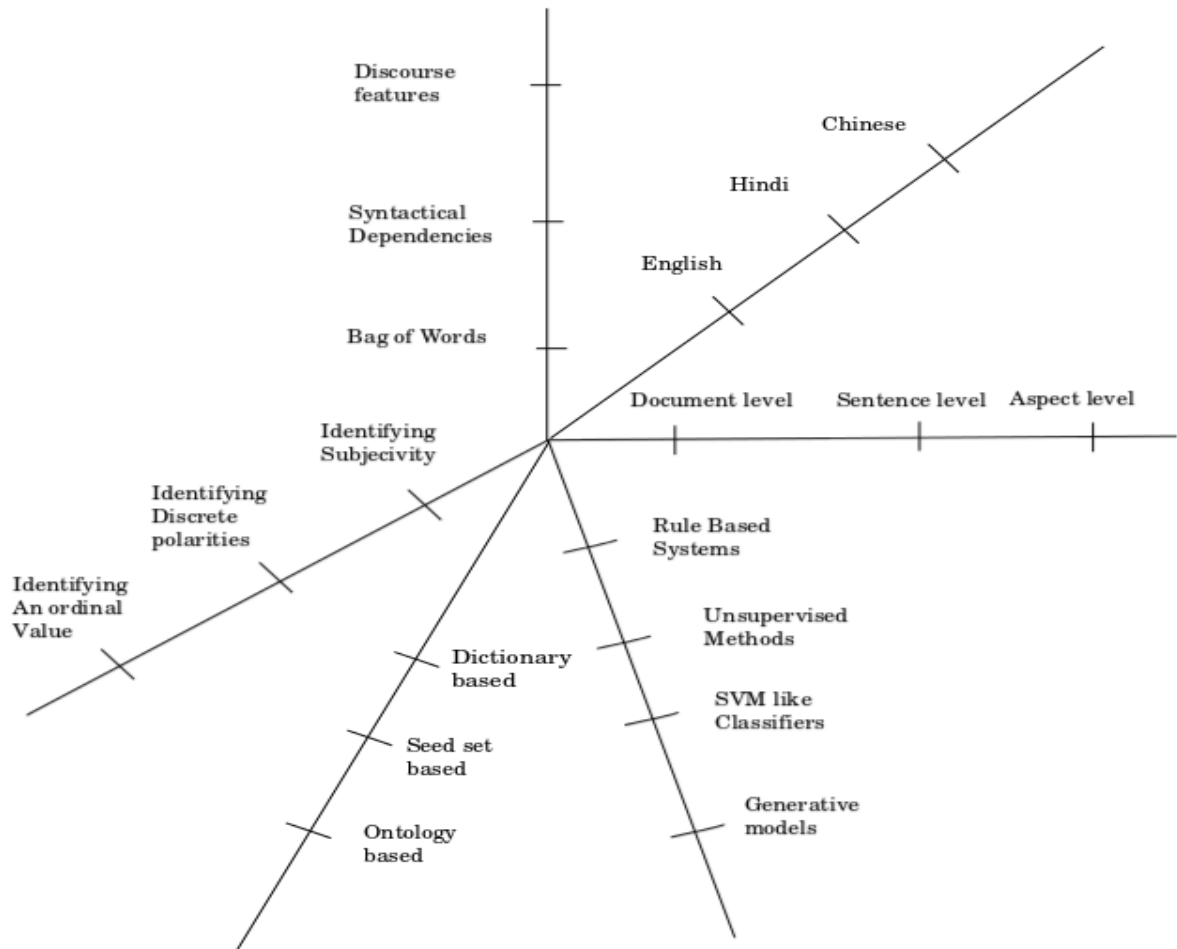


Figure 1.3.1: Dimensions of Sentiment Analysis

subjective and objective content.

Identifying discrete polarities:

Once the subjective part is determined, the next task is to determine if the content is positive or negative. This problem can be looked upon as a classification problem

Identifying an ordinal value:

Some applications require not just the type of polarity but the intensity as well. For example, movies are typically rated on a 5 point scale. Thus, this task aims at identifying such an ordinal value.

Tasks based on Levels of Sentiment Analysis

Document level:

As the name suggests, document-level Sentiment Analysis tags individual documents with their sentiment. The general approach here is to find the sentiment polarities of individual sentences or words and combine

them together to find the polarity of the document. These techniques may involve complex linguistic phenomena like co-reference resolution, pragmatics, *etc.*

Sentence or phrase level:

Sentence-level Sentiment Analysis deals with tagging individual sentences with their respective sentiment polarities. The general approach that is followed is to find the sentiment orientation of individual words in the sentence/phrase and then to combine combine them to determine the sentiment of the whole sentence or phrase. Other approaches like considering discourse structure of the text, have also been considered.

Aspect level:

These methods not only concern themselves with tagging individual words with their sentiment but they also aim at identifying the entity towards which the sentiment is directed. These methods heavily use techniques like dependency parser and discourse structures.

1.4 Challenges

Sentiment Analysis is a very challenging task. It requires deep understanding of the problem. We discuss some of the challenges faced in Sentiment Analysis.

- **Identifying subjective portions of text:** The same word can be treated as subjective in one context, while it might be objective in some other. This makes it difficult to identify the subjective (sentiment-bearing) portions of text. For example:

- *The language of the author was very **crude**.*
- ***Crude** oil is extracted from the sea beds.*

The same word “crude” is used as an opinion in first sentence, while it is completely objective in the second sentence.

- **Associating sentiment with specific keywords:** Many sentences indicate an extremely strong opinion, but it is difficult to pinpoint the source of these sentiments. Hence an association to a keyword or phrase is extremely difficult. For example:

- *Every time I read ‘Pride and Prejudice’ I want to dig her up and beat her over the skull with her own shin-bone.*

In this example, “her” refers to the character in the book “Pride and Prejudice”, which is not explicitly mentioned. In such cases the negative sentiment must be associated with the character in the book.

- **Domain dependence:** The same sentence or phrase can have different meanings in different domains. The word *unpredictable* is positive in the domain of movies, but if the same word is used in the context of a vehicle's steering, then it has a negative connotation.

- **Sarcasm Detection:** Sarcastic sentences express negative opinion about a target using positive words. For example:

– *Nice perfume. You must marinate in it.*

The sentence contains only positive words but still the sentence expresses a negative sentiment.

- **Thwarted expressions:** There are some sentences wherein a minority of the text determines the overall polarity of the document. Consider the following example:

– *This film should be brilliant. It sounds like a great plot, the actors are first grade, and the supporting cast is good as well, and Stallone is attempting to deliver a good performance. However, it can't hold up.*

Simple bag-of-words approaches will fail drastically in such cases, as most of the words used in here are positive, but the ultimate sentiment is negative.

- **Indirect negation of sentiment:** Sentiment can be negated in subtle ways as opposed to a simple no, not, *etc.* It is non-trivial to identify such negations easily. Consider the following example:

– *It **avoids** all cliches and predictability found in Hollywood movies.*

While the words cliche and predictable bear a negative sentiment, the usage of “avoids” negates their respective sentiments.

- **Order dependence:** While in traditional text classification, the discourse structure does not play any role in classification, since the words are considered independent of each other, discourse analysis is essential for Sentiment Analysis/Opinion Mining. For example:

– *A is better than B, conveys the exact opposite opinion from, B is better than A.*

- **Entity Recognition:** Not everything in a text talks about the same entity. We need to separate out the text about a particular entity and then analyze its sentiment. Consider the following:

– *I hate Nokia, but I like Samsung.*

A simple bag-of-words approach will mark this as neutral, however, it carries a specific sentiment for both the entities present in the statement.

- **Identifying opinion holders:** It is non-trivial to identify the opinion holders in any given piece of text. All that is written in a piece of text is not always the opinion of the author. For example, when the author quotes someone else, it becomes difficult to identify the source of that particular opinion. Such sentences are usually observed in news articles. Consider the following example:

- *Romney accused his rival of overseeing a stagnant economy. “The middle class has been crushed over the last four years and jobs have been too scarce,” the former Massachusetts governor said.*

Even though the comment by Romney is negative, this news item provides an objective opinion.

Chapter 2

Sarcasm

Sarcasm is a form of speech act in which the speakers convey their message in an implicit way. The inherently ambiguous nature of sarcasm sometimes makes it hard even for humans to decide whether an utterance is sarcastic or not. In this chapter, we discuss sarcasm in detail, what are the types of sarcasm and the challenges faced in detection of sarcasm.

2.1 Definition

Sarcasm is the use of words that mean the opposite of what the speaker wants to say with the “hidden” or rather apparent intention of insulting someone, showing irritation or to be funny. Recognition of sarcasm can benefit many Sentiment Analysis NLP applications, such as review summarization, dialogue systems, review ranking systems, *etc.*

Consider the following sentences:

- *Wow GPRS data speeds are blazing fast.*
- *Nice perfume. You must marinate in it.*
- *I like you. You remind me of myself when I was young and stupid.*
- *If I throw a stick, will you leave?*

The sentences listed above have no negative word in them, yet they are all sarcastic. If a bag of words technique is used for Sentiment Analysis on these type of sentences, it would give positive or neutral, but they are actually negative sentences. Unlike a simple negation, a sarcastic sentence conveys a negative opinion using only positive words or intensified positive words. The detection of sarcasm is therefore important, for the development and refinement of Sentiment Analysis.

Sarcasm is “a form of ironic speech commonly used to convey implicit criticism with a particular victim as its target” (McDonald, 1999, 486-87). “Irony” and “sarcasm” are both ways of saying one thing and meaning another but they go about it in different ways. Irony is a rhetorical device, literary technique, or an event characterized by an incongruity, or contrast, between what the expectations of a situation are and what is really the case. Sarcasm is really the use of irony with the added intention to mock, ridicule or express contempt. Sarcasm is broader and more deliberate in its reversal of meanings. For example: A statement like *Great, someone stained my new dress.* is ironic, while *You call this a work of art?* is sarcastic.

Sarcasm is meant to mock people but not in all cases. Banter is an example of positive sarcasm, also known as teasing, or mocking someone gently. In this case, the use of a negatively worded utterance conveys praise (McDonald, 1999, 487). For example, *Jeff is the most selfish individual in the world; you can find him serving at soup kitchens on Saturday nights while his buddies are off dancing in nightclubs.* In this example, positive sentiment about Jeff is conveyed using negative words. One may assume that “Jeff is selfish” means negative about Jeff, but in this case the word “selfish” is used to draw attention to the fact that Jeff is the antithesis of a selfish person.

In this report, we do not consider banter. We only look at sarcasm wherein the speaker uses positive words to convey a negative opinion about a target. We discuss detection of sarcasm in short and noisy text, specifically on Twitter messages called as tweets.

2.2 Types of Sarcasm

There are seven different types of sarcasm as defined by Writers’ Cafe Lamb (2006). We shall look at them in brief.

- Self-deprecating sarcasm: This type of sarcasm shows an exaggerated sense of worthlessness and inferiority. For example: *“Hey Bob, I’m going to need you to work overtime this weekend.” “Yeah, that’s fine. I mean, I was going to get married this weekend but, you know, it’s not a big deal, I will just skip it. She would have left me anyway.”*
- Brooding sarcasm: In brooding sarcasm, the speaker says something polite or subservient in a bitter or irritated tone. For example: *“Hey Bob, I’m going to need you to work overtime this weekend.” “Looking forward to it. I live to serve.”*
- Deadpan sarcasm: Deadpan sarcasm is said without laughter or emotion, so that it’s hard to tell whether or not the speaker is mocking the other person. For example: *“Hey Bob, going to need you to work overtime this weekend.” “Can’t make it. Got a cult meeting. It’s my turn to kill the goat.”*
- Polite sarcasm: Polite sarcasm is a subtle sarcasm, but sounds very polite. This is a kind of sarcasm

that sounds genuine at first, but then it slowly becomes clear. For example: *“Hey Bob, I’m going to need you to work overtime this weekend.” “Ooh, fun! I’ll bring the ice cream!”*

- Obnoxious sarcasm: Obnoxious sarcasm is usually spoken in a whiny tone of voice. For example: *“Hey Bob, I’m going to need you to work overtime.” “Oh, well that’s just great. Just what I wanted to do this weekend. Awesome.”*
- Manic sarcasm: In manic sarcasm, the speaker expresses unnatural extreme happiness. For example: *“Hey Bob, I’m going to need you to work overtime.” “God, you are the best boss EVER! Have I ever told you how much I love this job? I wish I could live here! Somebody get me a tent, I never wanna leave!”*
- Raging sarcasm: Raging sarcasm relies heavily on hyperbole and has threats of violence. For example: *“Hey Bob, I’m going to need you to work overtime.” “Oh, don’t worry! I’ll be there! Want me to shine your shoes while I’m at it?! Hell, I’ll come to your house tonight and wash your goddamn Ferrari! Actually, you know what? Forget it. I’m just gonna go home and blow my brains out.”*

2.3 Challenges in Sarcasm Detection

Sarcasm detection is a very challenging task. Following are some of the challenges faced in sarcasm detection.

- In spoken interaction, sarcasm is often marked with a special intonation or an incongruent facial expression. In written communication, authors do not have clues like “a special intonation” or “an incongruent facial expression” at their disposal. Therefore detection of sarcasm from text requires much deeper insight.
- Sarcastic sentences convey a negative opinion using only positive words or intensified positive words. So it is not possible to use a simple bag-of-words approach for Sentiment Analysis on such sentences.
- We discuss sarcasm detection in short and noisy text (tweets). The tweets are short and constrained to a length of 140 characters. The detection of sarcasm in such contextless tweets becomes very challenging.
- In some cases of sarcasm, incorporation of world knowledge is required. For example, *Thank you Janet Jackson for yet another year of Super Bowl classic rock!*. The given example is sarcastic because of the fact that Janet Jackson gave a bad performance in the year 2010 and then another scandalous performance in the next year. Incorporation of universal knowledge is itself a big task.
- Research has shown that sarcasm is often signaled by hyperbole. Hyperbole is the use of exaggeration. For example, *Wow GPRS data speeds are blazing fast*. In this sentence, “blazing” is the

hyperbole. Hyperbole detection would help in sarcasm detection, but this itself is an NLP problem that requires much more research.

- Not much research has been done in the field of sarcasm detection. Various new features need to be explored. Sarcasm detection may involve going deeper into semantics.

2.4 Hyperbole Detection

In this section we discuss hyperbole detection which is one of the challenges faced in sarcasm detection. Hyperbole is the use of exaggeration as a rhetorical device or figure of speech. Research has shown that sarcasm is often signaled by hyperbole, using intensifiers and exclamations. For example: *Your dad is the smartest guy in the world.* Such sentences make use of hyperbole to be sarcastic. So the aim is to detect hyperbole as a subtask which will aid in sarcasm detection.

Hyperbole and irony both can be used to express surprise but they do so differently. Hyperbole is understood because it inflates the discrepancy between the expected and ensued situation. When a speaker's expectations about some event are not known explicitly and then a negative event ensues, then using hyperbole to describe that situation expresses more surprise than using irony. For example, *Kerri broke the strings of her guitar right before she was to perform.* Using hyperbole to describe this situation, *This is the worst situation that anyone could ever be in* and using irony, *This is a great situation.* However, if the speaker's expectations are explicitly stated prior to the event (for example, *Kerri expected her performance to go off without a hitch*), irony expresses more surprise than hyperbole.

We tried to handle hyperbole by creating a list of hyperbolic words like “blazing”, “fantastic”, “astounding”, etc. This approach failed to capture situations like the following:

- *My mom is going to kill me for breaking the vase.*
- *She can have any boy that she wants.*
- *I can smell pizza from a mile away, etc.*

So we need to explore new approaches in order to handle hyperbole.

Another proposed approach to handle hyperbole is as follows:

- The first step is to analyze the list of adjectives manually.
- Then run a concordancer to get the combinations of noun and adjective. A concordancer gives a list of several words, phrases, or distributed structures along with immediate contexts, from a corpus or other collection of texts assembled for language study.

- Search for the obtained adjective-noun pairs in lexical resources like conceptnet, HowNet, WordNet, verbocean, framenet. If a new combination of adjective-noun is found, it is most likely to be sarcastic.

This may handle some types of hyperboles but not all. Hyperbole detection is itself a very challenging task which needs to be explored at greater depths.

2.5 Incorporation of World Knowledge

Certain sarcastic sentences require the incorporation of universal knowledge in order to classify it as being sarcastic. Consider the following sentences:

- *Thank you Janet Jackson for yet another year of Super Bowl classic rock!*
- *I Love The Cover* (book, amazon)
- *Defective only by design* (music player, amazon)

The first example is sarcastic because of the fact that Janet Jackson gave a bad performance in the year 2010 and then another scandalous performance in the next year. The second sentence is a review about a book given by some user in Amazon. If we consider the expression “do not judge a book by its cover” and apply this expression in our example, we realize that it is actually a sarcastic sentence. The third sentence may appear to be a positive sentence but it is actually sarcastic because of the fact that “design” is the most celebrated feature of Apple’s products, and if design itself is defective then the product is not liked by the consumer. Such sentences cannot be detected as being sarcastic unless world knowledge is incorporated in the system.

One proposed approach to incorporate world knowledge into the system is as follows:

- For a particular input sentence, find the entities in the sentence.
- Crawl the web for that entity and collect the facts. The best place to search for facts would be Wikipedia.
- Compare the facts with the input sentence. If they contradict, then most likely the sentence under consideration is a sarcastic sentence.

One more situation that needs to be handled is when, tweets on a particular topic are majority of the times negative, and a new tweet appears with inflated positive words, then it is most likely to be sarcastic. This will require fetching tweets related to the entity identified in the sentence under consideration. Then we need to compare if all the fetched tweets or a majority of them, have a negative polarity and the sentence under consideration uses highly positive words, then the sentence is most probably sarcastic.

Chapter 3

Related Work

For the past decade or so, there has been a lot of research that has taken place in Sentiment Analysis. We will discuss some of those works in this chapter.

3.1 Lexical Resources

There are various lexical resources in use, for Sentiment Analysis. We discuss dictionary and SentiWordNet in this section.

Dictionary

All sentiment analysis tools require a list of words or phrases with positive and negative connotation, and such a list of words is referred to as a dictionary ¹. Dictionary is an important lexical resource for Sentiment Analysis.

A single dictionary for all the domains, is difficult to generate. This happens because of the domain specificity of words. Certain words convey different sentiments in different domains. For example:

- Word like “fingerprints” conveys a major breakthrough in a criminal investigation whereas it will be negative for smartphone manufacturers.
- “Freezing” is good for a refrigerator but pretty bad for software applications.
- We want the movie to be “unpredictable” but not our cell phones.

A few popular dictionaries are discussed in the following sections.

- The Loughran and McDonald Financial Sentiment Dictionary:
Loughran and McDonald (2011) show how, applying a general sentiment word list to accounting and

¹<http://provalisresearch.com/products/content-analysis-software/wordstat-dictionary/sentiment-dictionaries>

financial topics, will lead to a high rate of misclassification. They found that around three-fourths of the negative word in a general sentiment dictionary were not negative in the financial domain. So they created the dictionary, “The Loughran and McDonald Financial Sentiment Dictionary”. It is a publicly available domain-specific dictionary and it contains custom lists of positive and negative words specific to the accounting and financial domain.

- Lexicoder Sentiment Dictionary (LSD):

Lexicoder Sentiment Dictionary (LSD) is also a domain-specific dictionary. It expands the score of coverage of existing sentiment dictionaries, by removing neutral and ambiguous words and then extracting the most frequent ones. Some important features of this dictionary are the implementation of basic word sense disambiguation with the use of phrases, truncation and preprocessing, as well as the effort to deal with negations.

- WordStat Sentiment Dictionary:

The WordStat Sentiment Dictionary was formed by combining words from the Harvard IV dictionary, the Regressive Imagery dictionary (Martindale, 2003) and the Linguistic and Word Count dictionary (Pennebaker, 2007). It contains a list of more than 4733 negative and 2428 positive word patterns. Sentiment is not predicted by these word patterns but by a set of rules that take into account negations.

Sentiwordnet

SentiwordNet is a lexical resource in which each wordnet synset ‘s’ is associated to three numerical scores Obj(s), Pos(s) and Neg(s), which describe how objective, positive and negative the terms contained in the synset are. Each of the three scores range from 0.0 to 1.0, and their sum is 1.0 for each synset. A graded evaluation of opinion, as opposed to hard evaluation, proves to be helpful in the development of opinion mining applications.

For example: The synset “estimable”, corresponding to the sense “may be computed or estimated”, has an Obj score of 1.0 and Pos and Neg scores 0.0. Whereas, the synset “estimable”, corresponding to the sense, “deserving of high respect or high regard”, has a Pos score of 0.75, a Neg score of 0.0, and an Obj score of 0.25. For such cases hard evaluation will not be efficient. Through SentiWordNet we can efficiently represent the scores of each synset in the wordnet.

3.2 Feature Engineering

The efficiency of classifiers depends upon the selection of features. Under feature engineering, we discuss feature selection in Sentiment Analysis.

- Sense based feature: The traditional approaches to sentiment analysis have been using lexeme and

syntax based features. Balamurali et al. (2011) focus on a new approach to sentiment analysis by using “word senses” as “semantic features” for sentiment classification. In his paper, he used WordNet 2.1 (Fellbaum, 1998) as the sense repository. Each word is mapped to a synset based on its sense.

The motivation behind this is that a word can have multiple senses. It may carry some polarity in one sentence at the same time, it may have some other polarity in another sentence. For example:

- *Her face fell when she heard the bad news.*
- *The apple fell off the tree.*

In the two sentences, the same word “fell” is used but with different senses. The first has a negative sentiment whereas the second sentence is objective and carries no sentiment. Hence, incorporating word senses is the need of the hour.

- **Term Presence vs. Term Frequency:** Traditionally term frequency was used as a feature in all the sentiment classification tasks. Later, Pang et al. (2002) showed that term presence is more important than term frequency. Term presence is a binary-valued feature which shows whether a term is present or not, unlike the term frequency feature which kept a count of the terms. This has been proved experimentally that term presence gives better results than term frequency.
- **Term Position as feature:** An observation that has been made is that term position plays an important role in determining the sentiment. For example, in movie reviews, the review might begin with some sentiment, discuss about the movie and at the end summarize it with the author’s view. We see that the sentiment is mainly present in the concluding sentences. Hence, we conclude that term position is very important. The sentiment of the initial sentences might not be the sentiment of the whole review. So term position is also included as a feature.
- **Part-Of-Speech features:** Part-Of-Speech plays a very vital role in all Natural Language Processing tasks. We describe some of the POS features:
 - **Adjectives only:**

Adjectives are considered to be the sentiment bearing words in any sentence. There is a strong correlation between adjectives and subjectivity. People use adjectives to reveal their sentiment. For example:

 - * *The movie was awesome*
 - * *I had a terrible day*

In the above two sentences, “awesome” and “terrible” are the adjectives and they are the ones deciding the sentiment of the whole sentence.

– Adjective-Adverb Combination:

Adverbs alone may not have any sentiment bearing property. But when used along with adjectives, they play an important role in sentiment analysis. Adverbs of degree, on the basis of the extent to which they modify sentiment, are classified as:

- * Adverbs of affirmation: certainly, totally
- * Adverbs of doubt: maybe, probably
- * Strongly intensifying adverbs: exceedingly, immensely
- * Weakly intensifying adverbs: barely, slightly
- * Negation and minimizers: never

For example: *I will never watch that awful movie*

The word “never” which is an adverb shows that the sentence is a strong negative sentence.

So we conclude that POS proves to be very efficient features.

- Unigram features: The unigrams, *i.e.*, the individual words, can be included as a feature. Pang et al. (2002) in their paper analysed the performance of unigram as features. The results showed that unigram presence taken as feature turns out to be the most efficient. We can also have bigram features such as “awesome plot”, “phenomenol acting” *etc.* In general, we can have n-gram as features in order to capture the context. But the paper’s experimental results showed that bigram as feature did not improve the performance of the sentiment classifier any further. So, unigram features are preferred over n-gram features.

3.3 Machine Learning techniques

Using movie reviews as data, Pang et al. (2002) show that standard Machine Learning techniques outperform human-produced baselines. The movie-review domain has been chosen because there are large on-line collections of such reviews, and reviewers often summarize their overall sentiment using stars ratings *etc.*, which are easily extractable by machine. We discuss a few of the Machine Learning Approaches.

- Naïve Bayes: A Naïve Bayes classifier is a simple probabilistic classifier based on applying Bayes’ theorem with strong (naïve) independence assumptions.¹ The Bayes’ rule is given below:

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

Naïve Bayes holds a strong independence assumption which states that features are independent of each other. By applying conditional independence assumption of features we get:

$$P_{NB}(c|d) = \frac{P(c) \prod_{i=1}^N P(f_i|c)^{n_i(d)}}{P(d)}$$

Where,

$P(c|d)$: probability of data d belonging to class c.

¹https://en.wikipedia.org/wiki/Naïve_Bayes_classifier

$P(d)$: probability of data.

$P(f_i|c)$: probability of the i th feature belonging to class c .

So, we see that the probability is expressed in terms of product of features. The conditional independence assumption may or may not hold in every situation. Yet, Naïve Bayes has proven to give good results.

- **Maximum Entropy:** Maximum entropy classifiers are commonly used as alternatives to Naïve Bayes classifiers because they do not assume statistical independence of the independent variables (commonly known as features).² (Nigam, 1999) shows that it sometimes, but not always, outperforms Naïve Bayes classifier. Here is the exponential form taken by Maximum Entropy classifier:

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

Where,

$Z(d)$ is a normalization function.

$F_{i,c}$ is a feature /class function for f_i and class c , defined as follows:

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

The $\lambda_{i,c}$ s are feature-weight parameters. Its large value shows that f_i is a strong indicator of the class c . Maximum Entropy classifier focuses on choosing parameters which maximize the the distribution. It does not make any assumptions about the relationships between features and so performs better than Naïve Bayes when conditional independence assumptions are not met.

- **Support Vector Machines:** Support Vector Machines(SVM), also called Large Margin classifiers are non-probabilistic classifier. It constructs a hyperplane or a set of hyperplanes in a high-dimensional space, which can be used for classification, regression, *etc.* A good separation is one in which the hyperplane, represented by \vec{w} , has the largest distance to the nearest training data point of any class.

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

Where,

α_j is a parameter

\vec{d}_j for which $\alpha_j > 0$ are called support vectors.

3.4 Sarcasm

Sarcasm is the use of positive words to express a negative opinion about some target. In this section, we look at the different features that can be used in sarcasm detection.

²https://en.wikipedia.org/wiki/Maximum_entropy

- Intensifiers as features:

Liebrecht et al. (2013) introduce a sarcasm detection system for tweets, messages on the micro-blogging service offered by Twitter. In micro-blogging sites such as Twitter, tweets are often explicitly marked with the #sarcasm hashtag to indicate that it is sarcastic. Research has shown that sarcasm is often signaled by hyperbole, using intensifiers and exclamations. In contrast to this, non-hyperbolic sarcastic messages often have an explicit marker. Unlike a simple negation, a sarcastic message conveys a negative opinion using only positive words or intensified positive words. According to Gibbs and Izett (2005), sarcasm divides its addressees into two groups; a group of people who understand sarcasm (the so-called group of wolves) and a group of people who do not understand sarcasm (the so-called group of sheep). On Twitter, the senders use the hashtag in order to ensure that the addressees detect the sarcasm in their utterance.

This paper focuses on the use of intensifiers as features. Hyperbolic words which strengthen the evaluative utterance are called intensifiers. For example: (when it rains)

- *The weather is good.*
- *The weather is fantastic.*

Both the sentences convey a literally positive attitude towards the weather, however, the utterance with the hyperbolic “fantastic” may be easier to interpret as sarcastic than the utterance with the non-hyperbolic “good”. Senders use such intensifiers in their tweets to make the utterance hyperbolic and thereby sarcastic. An experiment was performed using uni-, bi- and trigrams were used as features. Balanced Winnow was used as the classification algorithm. A set of 77,948 tweets was collected for training the classifier. The results show that some intensifiers are strong predictors of sarcasm, such as “awesome”, “lovely”, “wonderful”, “of course”, “fortunately”, “soooo”, “most fun”, “fantastic”, and “veeery”.

- Lexical and pragmatic features:

Gonzalez-Ibanez et al. (2011) throw light upon the impact of lexical and pragmatic factors for effectively identifying sarcastic utterances in Twitter. He also compares the performance of machine learning techniques and human judges on this task. Sarcastic tweets were collected using the #sarcasm hashtag and automatic filtering was done to remove retweets, duplicates, quotes, spam, tweets written in languages other than English, and tweets with URLs. Since hashtagged tweets are noisy, all the tweets were filtered where the hashtags of interest were not located at the very end of the message.

Two kinds of lexical features have been used, unigram and dictionary based. The dictionary-based features were derived from Pennebaker et al.’s LIWC (2007) dictionary, WordNet Affect (WNA) (Strapparava and Valitutti, 2004) and list of interjections and punctuation. Three pragmatic

features have been used, positive emoticons, negative emoticons and ToUser which is used as a reply to some other user. Two classifiers were used for the classification task : support vector machine with sequential minimal optimization (SMO) and logistic regression (LogR). The following features were used: unigrams, presence of dictionary-based lexical and pragmatic factors and frequency of dictionary-based lexical and pragmatic factors. Bigrams and trigrams were also used as features but the results were not very good. The results show that lexical and pragmatic features do not provide sufficient information to efficiently distinguish sarcastic from positive and negative sentences. The results obtained were compared with the performance of humans by allowing humans to classify 10% of the test dataset and it was observed that humans do not perform significantly better than the machine.

- Pattern-based features:

Davidov et al. (2010) make use of pattern-based features. Words are divided into two categories, high frequency words(HFW) and content words(CW). A word whose corpus frequency is higher than FH is HFW and a word whose corpus frequency is lower than FC is a content word. A pattern is an ordered sequence of high frequency words and slots for content words. After extracting patterns, hundreds of patterns are obtained out of which some may be very specific and some may be very general. In order to reduce the feature space, pattern selection is done. Then for each pattern, feature value is calculated.

Along with pattern based features following features were also used:

- Sentence length in words
- Number of “!” characters in the sentence
- Number of “?” characters in the sentence
- Number of quotes in the sentence
- Number of capitalized/all capitals words in the sentence

These features were all normalized. K-nearest neighbor strategy is used to assign label to an instance in the test set. The model was trained using the tweets collected from #sarcasm hashtag but the results were not promising. In order to solve this issue, cross domain corpus was built using positive reviews from the Amazon dataset and negative tweets from Twitter. An accuracy of 90.2% is achieved on the Twitter dataset with F-Score of 0.505.

3.5 Hyperbole

Hyperbole is the use of exaggeration as a rhetorical device or figure of speech. Colston and Keller (1998) compare irony with hyperbole and the extent to which they express surprise. The authors test the inflation hypothesis which states that hyperbole is understood because it inflates the discrepancy

between the expected and ensued situation. If hyperbole is understood because of this inflation then it should not matter that a speaker's surprise is obvious when explicitly stated expectations are violated. Inflating that discrepancy should still express surprise.

The author tests whether or not hyperbole expresses surprise when a speaker's expectations are explicitly known. This is done by combining irony and hyperbole to see how they together express surprise. For example:

- Hyperbole: *I see we got 10 feet of snow last night*
- Irony: *I see we got some slight flurries last night*
- Combination of irony and hyperbole: *I see we didn't get any snow at all last night*

The latter utterance is both ironic and hyperbolic because it goes back to what was expected (a slight amount of snow) and because it inflates the discrepancy between what was expected and what ensued.

The author conducts experiments to see whether hyperbole is sensitive to how events can turn out unexpected. This is done by investigating the degree of surprise expressed, when expectations concerning quantities of substances are violated. Lesser than expected is constrained because it cannot be less than zero whereas more than expected can go upto infinity. The experiments performed involved fifty-six undergraduates from University of California. They were asked to provide a rating to the set of sentences presented to them. A seven point rating scale is used which ranges from, not at all expected to completely expected. The test cases involved two quantity types: less than expected and more than expected. Three experiments were conducted:

- First experiment: In the first experiment, four sets of scenarios are considered: only hyperbole, only irony, hyperbole and irony and literal comment on an unexpected situation. The results show that for both cases, hyperbole and irony together have a low rating (more surprised) as compared to irony alone and hyperbole. For the test case of "less than expected", hyperbole and literal comment have the same score because hyperbole gets contained. For the case of "more than expected", hyperbole has a lower score.
- Second experiment: Second experiment was performed to observe the degree of surprise when different levels of hyperbole were used. For this, three sets were created.
 - Set 1: realistic version of hyperbole.
 - Set 2: possible but improbable version.
 - Set 3: impossible version of hyperbole.
 The results prove that any level of hyperbole express the same degree of surprise.
- Third experiment: The range in degree of inflation available to hyperbole exists and is used by speakers. One would suspect that it would serve some pragmatic function. Another possibility is

that the range of inflation is used to make a speaker's expression of surprise easier to understand. So the third experiment was conducted. Even when a speaker's expectations are explicitly stated, the range of inflation available to hyperbole serves as a pragmatic function for interpreting the hyperbole.

The conclusion is that hyperbole is comprehended because of the inflation even when the speaker's surprise is obvious.

3.6 Thwarting

Thwarting is the phenomenon in which a minority of sentences decide the polarity of the whole piece of text or document. There are two approaches to detect whether a document is thwarted or not. One is rule-based approach and the other is statistical approach. Ramteke et al. (2011) describe both the approaches. The author has made use of domain ontology to handle thwarting. Domain ontology comprises of features and entities from the domain and the relationships between them depicted by a hierarchy. For building the ontology the features and entities need to be identified and then linked in the form of an hierarchy. The author has built the domain ontology manually.

The rule-based approach makes use of a domain ontology . An ontology gives weightings to entities related to a domain. Ramteke et al.(2012) built up a system using the domain ontology for camera review domain. The word polarities were found using four different lexicons namely SentiWordNet, Inquirer, BL Lexicon and Taboada. The entity specific polarities were found by considering the dependencies found using Stanford Dependency parser. The weighing scheme gave a weight of 1 to the leaf nodes and then subsequently increased the weights by 1 for the higher levels. A review is said to be thwarted, if the root node has a different polarity from its leaf nodes.

The drawback of rule-based approach is that it gives equal weightage to all the features in the domain ontology. This drawback is overcome by assigning weights to features. This approach also makes use of the domain ontology . It aims at finding the features and their weights that are used for training the classifier, from the domain ontology. The review is represented as a sequence of weighted polarities. The review is linearly scanned and if a word is encountered belonging to the ontology, its polarity and weight are extracted using the corresponding node in the ontology. The sequence of occurrence of words is maintained since position is vital to determining thwarting. The features are extracted from the sequence and fed to the classifier, which classifies the review to be thwarted or not.

Chapter 4

Sentiment Analysis efforts at IIT Bombay

There has been a lot of effort put into the field of Sentiment Analysis in the recent years. IIT Bombay has been making efforts in this field since around half a decade.

4.1 Sense-id based Sentiment Analysis

Balamurali et al. (2011) focus on a new approach to Sentiment Analysis by using wordnet senses as semantic features for sentiment classification, instead of the traditional lexeme based features. The motivation behind this work can be understood from the following examples:

- *Her face **lit** up.*
- *The fire was **lit**.*

In this example, we see that the word “lit” has two different senses. In the first sentence it conveys a positive sense whereas in the second sentence it conveys a neutral sense.

- *The tornado **destroyed** the city.*
- *Sachin Tendulkar **destroyed** the opposition with his amazing skills.*

Here we see that the word “destroyed” has two totally opposite polarities in its two sentences. The sense of destroyed in the first sentence makes it a negative sentence. Whereas the second sentence is a positive sentence.

Thus, it is quite apparent that using senses of words instead of the words themselves is important in Sentiment Analysis. The authors used a state of the art iterative Word sense disambiguation system to identify the senses of the words and then performed sentiment classification. They observed that the

accuracies improved when the senses were used as opposed to simple word features. They also proposed a method to handle unknown words using word senses. Thus, while in general these words would be missed, using the WordNet senses such words are also captured thus improving the overall accuracy. The accuracies noted were as high as 85%.

The following feature representations were used by the author and their performance were compared to that of lexeme based features:

1. Word senses that have been manually annotated (M)
2. Word senses that have been annotated by an automatic WSD (I)
3. A group of manually annotated word senses and words (both separately as features)
(Sense + Words (M))
4. A group of automatically annotated word senses and words (both separately as features)
(Sense + Words (I))

If a synset encountered in a test document is not found in the training corpus, it is replaced by one of the synsets present in the training corpus. This is termed as the synset-replacement strategy. The substituent synset is determined on the basis of similarity with the unknown synset, calculated using similarity metrics. The metrics used by the author are LIN, LESK, LCH. The dataset used by the author is the dataset by (Ye et al., 2009). The experiments were performed using C-SVM on the different feature representations.

Following are the observations made from the experiment:

- The combined model of words and manually annotated senses (Words + Senses(M)) gives the best performance with an accuracy of 90.2percentage.
- Negative class detection is more difficult compared to positive class detection. It has been shown that adverb and verb synsets play an important role in negative class detection. Therefore, these synsets must be taken care of.
- The lexeme space requires a larger number of training samples in order to achieve the accuracy which synset space can achieve in lesser number of training samples.
- Partial disambiguation performs better than no disambiguation.
- Lesk gives the best classification accuracy compared to the other two metrics.

The author concludes that sense-based features prove to be efficient in sentiment classification task. The next important conclusion is that even partial disambiguation performs better than no disambiguation.

4.2 Cross-lingual Sentiment Analysis

Cross-Lingual Sentiment Analysis is the task of predicting the polarity of the opinion expressed in a text in a language L_{text} using a classifier trained on the corpus of another language L_{train} . Popular approaches use Machine Translation (MT) to convert the test document in L_{test} to L_{train} and use the classifier of L_{train} . However, MT systems are resource intensive and do not exist for most pairs of languages and those that exist have a low translation accuracy. Balamurali et al. (2011) present an approach to Cross-Lingual Sentiment Analysis for Indian languages, namely, Hindi and Marathi. The author presents an alternative approach to cross-lingual Sentiment Analysis using WordNet senses as features for supervised sentiment classification. The document to be tested for polarity is preprocessed by replacing the words in it with the corresponding synset identifiers. The document vector created from the sense-based features could belong to any language. The preprocessed document is then given to the classifier coming from L_{train} for polarity detection. Experiments were performed based on a sense marked corpora using an automatic WSD engine. The author suggested that even a low quality word sense disambiguation leads to an improvement in the performance of sentiment classification.

4.3 Discourse based Sentiment Analysis

Mukherjee and Bhattacharyya (2012) propose a lightweight method for using discourse relations for sentiment detection of tweets. The motivation for the work can be seen through the following examples:

- **Violated Expectations:** The direction was (*notthatgreat*)⁻, but still we *loved*⁺ the movie.
Here a simple bag of words approach would tag it as neutral but due to the presence of **but** it eventually turns out to be positive.
- **Violated Expectations:** India managed to *win*⁺ despite the initial *setback*⁻.
Here, the word **despite** works in the opposite fashion as **but** thus the overall polarity is positive as opposed to neutral.
- **Conclusion:** We were (*notmuchsatisfied*)⁻ with the *greatly*⁺*acclaimed*⁺ brand X and subsequently decided to *reject*⁻ it.
Here, the word **subsequently** gives more weight to **reject** thus the overall polarity is negative.
- **Conditional:** If Brand X had *improved*⁺ its battery life, it would have been a *great*⁺ product.
Here, the conditional **if** renders the entire sentence neutral.
- **Modal:** That film might be good. He may be a rising star.
Here, the words **might** and **may** act similar to conditionals and render the sentence neutral.
- I heard the movie is good, so you **must** go to watch that movie.

- You **should** go to watch that awesome movie.

In these examples, we see a difference in degree of certainty. These two examples are more certain than the previous two examples.

- **Negation:** I do (*notlike*)⁻ Nokia but I *like*⁺ Samsung.

This is the conventional negation which is handled in many previous approaches as well. Here, the sentiment towards particular entities are different. The approach used by the authors is to negate all words in a window of size 5 provided the sentence does not end or an instance of violated expectations is encountered (but).

Thus, we see several discourse elements playing a crucial role in determining the word polarities correctly. On incorporating discourse elements in the existing twitter sentiment system, the accuracies were found to be better by 2%. This method is built for the web-based applications that deal with noisy, unstructured text, such as tweets, and cannot use heavy linguistic resources like parsing. This is due to frequent failure of the parsers to handle noisy data. In his work, the author shows how the discourse relations like the connectives and conditionals can be used to incorporate discourse information in any bag-of-words model, in order to improve the sentiment classification accuracy. Also, the author examines the influence of the semantic operators like modal and negations on the discourse relations that affect the sentiment of a sentence. Discourse relations and corresponding rules are identified with minimal processing. A linguistic description of various discourse relations has been given, which leads to conditions in rules and features in SVM. The discourse-based bag-of-words model performs well in a noisy medium such as Twitter. Furthermore, the approach is beneficial to structured reviews. The system is less resource intensive and performs favorably in comparison to the state-of-the-art systems.

4.4 C-Feel-it System

Joshi et al. (2011) developed the C-Feel-It system which is capable of classifying sentiment expressed in tweets. The web-based system developed by the author, categorizes tweets related to the user query, as positive, negative or objective and assigns an aggregate sentiment score to it. C-Feel-It uses a rule-based system to classify the sentiment expressed in tweets using inputs from four sentiment-based knowledge repositories. A weighted majority voting principle is used to predict sentiment of a tweet. An overall sentiment score for the search string is assigned based on the results of predictions for the retrieved tweets. This score is given as a percentage value and it represents the sentiment of users about the topic. Twitter is a very noisy medium where the user posts different forms of slangs, abbreviations, smileys, *etc.* There is also a high occurrence of spams generated by robots. Due to these reasons, the accuracy of the system deteriorated mainly because the words in the post were not present in the lexical resources. The author used a slang and emoticon dictionary for polarity evaluation in the system. Four lexical resources have

been used, namely, Taboada, Inquirer, SentiWordNet and Subjectivity Lexicon. The system categorizes the tweets based on the predictions of these four sentiment based resources.

4.5 Thwarting

Ramteke et al. (2012) used domain ontology to handle thwarting. Domain ontology comprises of features and entities from the domain and the relationships between them depicted by a hierarchy. For building the ontology the features and entities need to be identified and then linked in the form of an hierarchy. The authors have built the domain ontology manually.

The rule-based approach makes use of a domain ontology . An ontology gives weightings to entities related to a domain. Ramteke et al. (2012) built a system using the domain ontology for camera review domain. The word polarities were found using four different lexicons namely SentiWordNet, Inquirer, BL Lexicon and Taboada. The entity specific polarities were found by considering the dependencies found using Stanford Dependency parser. The weighing scheme gave a weight of 1 to the leaf nodes and then subsequently increased the weights by 1 for the higher levels. A review is said to be thwarted, if the root node has a different polarity from its leaf nodes.

The drawback of rule-based approach is that it gives equal weight to all the features in the domain ontology. This drawback is overcome by assigning weights to features. This approach also makes use of the domain ontology . It aims at finding the features and their weights that are used for training the classifier, from the domain ontology. The review is represented as a sequence of weighted polarities. The review is linearly scanned and if a word is encountered belonging to the ontology, its polarity and weight are extracted using the corresponding node in the ontology. The sequence of occurrence of words is maintained since position is vital to determining thwarting. The features are extracted from the sequence and fed to the classifier, which classifies the review to be thwarted or not.