Chapter 1

An Introduction to SA

Sentiment Analysis (SA) refers to a broad field of study in Natural Language Processing (NLP) and Artificial Intelligence (AI). In general, Sentiment Analysis is the task of identifying the orientation of opinions expressed, towards a specific entity, in the subjective portions of a given piece of text.

People have a tendency to express their opinion on various entities. Sentiment Analysis deals with evaluating whether this expressed opinion about the entity has a positive or a negative orientation. With the increasing amount of similar user-generated data – in the form of reviews, blogs, etc – online, the need for automated tools for such analysis has increased recently.

Most of the research in Sentiment Analysis till now has focused on in-domain supervised approaches, mainly based on bag-of-words or other similar feature models based on the bag-of-words approach. In this work, we try to provide a different perspective. We believe that the syntax and semantics of the language play an equally important role as the words themselves. All our work through this project has a focus on leveraging the syntactical and semantic information available in the text to determine the sentiment.

In this chapter, we introduce you to the general task of Sentiment Analysis. We try discuss some of the basic tasks and challenges in SA. We start with a comparison of Sentiment Analysis to Standard Fact-based Text Categorization (Section 1.1), followed by the Motivations for Sentiment Analysis (Section 1.2). Sections 1.3 and 1.4 delve into some of the Applications and Tasks in Sentiment Analysis while Section 1.5 discusses some of the Challenges in Sentiment Analysis. Section 1.6 describes the work done at IITB on Sentiment Analysis. We conclude with a summary of the chapter (section 1.7).

1.1 Contrast to Fact-based Text Categorization

Standard Text Categorization deals with assigning every document to one of the classes based on the contents of the document. Based on this definition, Sentiment Analysis seems almost like an extension to this task, there have been several approaches for text categorization devised with extremely good results. However, Sentiment Analysis is much more subtle than text categorization. The two fields differ on multiple grounds. We will discuss a few in this section.

- In standard text categorization, the number of classes is usually arbitrary, their definitions mainly depending on the user and application. For a given task, we might be dealing with as few as two classes (binary classification) or as many as thousands of classes (*e.g.*, classifying documents with respect to a complex taxonomy). In contrast, with sentiment classification, we often have relatively few classes (*e.g.*, positive or 3 stars) that generalize across many domains and users.

- While the different classes in topic-based categorization can be completely unrelated, the sentiment labels that are widely considered, typically represent opposing (if the task is binary classification) or ordinal/numerical categories (if classification is according to a multi-point scale).

- While simple bag-of-words approach does perform well in standard textual analysis, [16] have shown that it fails to produce similar results in case of Sentiment Analysis.

- In textual analysis, we can devise a list of keywords for any given topic, and based on the number of occurrences of these keywords, we can classify the document. In sentiment analysis, the very task of coming up with the right set of keywords is non-trivial even for human users [16].

  This can be attributed to the fact that compared to topic, sentiment can often be expressed in a more subtle manner, making it difficult to be associated with any of sentence's or document's terms when considered in isolation.

- In text-based analysis, the overall topic of a document should be what the majority of the document is focusing on, regardless of the order in which the terms are presented in the document. Thus, a term that appears more frequently can be considered more closely related to the topic than a term that appears with a lesser frequency. This does not hold true for sentiment analysis. In sentiment analysis, in certain cases, the presence of a word, outweighs the frequency information, as shown by [16].

- The order in which potentially different opinion bearing sentences occur, might change the orientation of the document in Sentiment Analysis. This does not hold true for text categorization.

1.2 Motivation: Need for Sentiment Analysis

The motivation for Sentiment Analysis has not just been user-need driven. Neither has it been solely governed by the market's needs. This is one field which has found applications in both the user domain, as well as the vendor domain. Below, we will examine how Sentiment Analysis fits into the needs of both the users as well as the manufacturers.

1.2.1 A User's Perspective

Intelligent beings try and learn from others' mistakes. Thus, *what other people think* about a product carries great influence on one's decision about any product. With the humongous amount of information available on the Internet, it becomes impossible for any user to actually go through these in order to understand the general opinion about a product. Even a very simple query fetches millions of results on any search engine. What people need is a tool that could automatically go through all such reviews and come up with a summarized opinion about the entity being evaluated.

1.2.2 The Vendor's Perspective

With the explosion of Web 2.0 platforms such as blogs, discussion forums, etc, consumers have an unprecedented reach and power. Their opinions about any product ultimately govern its success or failure. Because of the enormity of this data, it is not possible to understand such feedback manually. Thus, they too require an automated engine that can crawl through such inputs across the Internet and decide whether the customers have given the product a *thumbs up* or *thumbs down*. Whenever a new product is launched in the market, such Sentiment Analysis tools can aid the company to identify their product's perception among the consumer base.

The feedback need not be restricted to the product as a whole alone. An organization might require a much more directed feedback trying to figure out which *parts/features* of the product are being disliked by its users, so that it can improve on them and provide better solutions to their customers.

Alternatively, these tools can be used to gauge *brand perception* and even for *reputation management*. While brand perception tends to understand how a brand is being perceived by the audience (consumers), reputation management deals with figuring out ways to maintain or improve the overall view about an already existing brand. It involves methods to identify the reasons for both, the positive as well as negative image of the company, so as to be able to work on the negative aspects while improving the positive aspects as well.

1.3 Applications of Sentiment Analysis

Recently, Sentiment Analysis has found applications in multiple domains. These include applications in review-related websites, applications in business intelligence and across many other domains. We will discuss these one-by-one in this section. This is not a comprehensive list of applications of Sentiment Analysis. We try to cover all the major applications of this field though.

1.3.1 Applications to Review-related Websites

Reviews and feedbacks on almost everything, ranging from product reviews, to feedbacks on political issues are abundantly available over the Internet. Much like a search engine, a *sentiment engine* can be built to utilize this information. It will provide a consolidated feedback or rating for the given topic. Such sites themselves do not contain any opinions, but they crawl the opinionated text from various resources and provide an effective polarity about the searched query.

Another application of Sentiment Analysis is in *automatic summarization of user reviews*. Automatic summarization is essentially the creation of a concise version of text by a computer program. In review domain, it has become exceedingly difficult for any user to go through all the reviews – about a product – available online. What people need is some kind of summarizing application that can proide succinct information about the polarity of the reviews, say *thumbs up* or *thumbs down* for the given topic.

We generally assume that all the user ratings are accurate. However, that is not always the case. There are cases where users accidentally or, at times, intentionally select a low rating when their review indicates a positive evaluation, or vice versa. Moreover, there is some evidence that user ratings can be biased, or otherwise, in need of correction. Automated sentiment classifiers can help to rectify such cases.

1.3.2 Applications as a Sub-component Technology

SA tools can easily be used as an augmentation to many existing technologies and applications. A recommender system, for example, should not suggest any item which has been frequently criticized. This can be taken care of, by augmenting an SA engine to the already existing recommender system.

Identification of "Flames" has been a point of concern for all e-mail service providers. Flames in emails—or other types of communications—are basically hostile and insulting interactions between Internet users. These can generally be characterized by overly-heated or antagonistic language, which should typically be identifiable by a sentiment analyzer.

In online systems displaying ads as sidebars, it is helpful to detect web-pages that contain sensitive content inappropriate for advertisement placement. For more sophisticated systems, it could be useful to bring up product ads when relevant positive sentiments are detected, and perhaps more importantly, nix the ads when relevant negative statements are discovered.

1.3.3 Applications in Business Intelligence

"*For many businesses, online opinion has turned into a kind of virtual currency that can make or break a product in the marketplace*" [38]. This statement highlights the importance of sentiment analysis in businesses.

The most obvious usage of Sentiment Analysis in business intelligence lies in understanding the user reviews to improve their products, and in turn, their reputation.

Another application of Sentiment Analysis in business domain is in creating *Influence Networks* [2], that try to correlate entities with their impact on the product.

Sentiment Analysis can also be used in trend prediction. By tracking public viewpoints, we might be able to predict trends in sales or in some other relevant domains.

1.3.4 Applications across Different Domains

The applications mentioned above are but a small portion of the applications of Sentiment Analysis. The interest in this field has risen spectacularly in the past few years, and as a result, it has applications across multiple domains. Its applications range from political and legal domains to even sociology domain, wherein, various applications based on sentiment analysis have been used to aid understanding. Discussion of such applications is out of the scope of this report.

1.4 Tasks in Sentiment Analysis

The tasks in Sentiment Analysis can be divided based on multiple criteria. One way is to differentiate based on the task at hand: whether it is a classification task or a scoring task, or even different types of classification. Another way to differentiate tasks in Sentiment Analysis is based on the level at which the classification is taking place. Although there might exist other ways to define tasks in SA, we will focus mainly on the above mentioned tasks.

1.4.1 Tasks based on Classification

Determining text SO-polarity: This task deals with identifying whether a given piece of text is Subjective or Objective (SO). In other words, it aims to identify whether the text expressed has a factual nature or does it express an opinion, given its subjective nature.

Determining text PN-polarity: This is a two-class classification problem that deals with deciding whether a given Subjective text expresses a Positive or Negative (PN) opinion on its subject matter.

Determining text PNO-polarity: Quite like the previous task, this tends to classify the given Subjective text into three classes, namely, Positive, Negative, or Objective (PNO).

Determining the strength of text PN-polarity: This deals with generating a regression score to the given piece of text. For example, given a piece of text expressing some Positive opinion, we need to identify whether the sentiment is Weakly Positive, Mildly Positive or Strongly Positive. Alternatively, PN-polarity can be assigned on a 5-star scale.

1.4.2 Tasks based on Levels of Sentiment Analysis

Word level: At the word level, we assign every word its sentiment score, based on its usage. Taking the context of the word into account yields better results.

Over the years, quite a few lexical resources have been developed that map individual words to their respective sentiment. WordNet Affect [29] and SentiWordNet [5] are a couple of such resources.

Phrase level: Phrase-level Sentiment Analysis deals with tagging phrases with their respective sentiment polarities. While we decide onto the polarity of individual words in word-level SA, phrase-level SA aims at finding the polarity of a group of words, that generally carry a single sentiment. The general approach followed is to find the sentiment orientation of individual words in the phrase and combine them to come up with the sentiment of the complete phrase. Other approaches, like considering discourse structure of the text or utilizing contextual dependencies, have also been explored.

Sentence level: Sentence level SA is quite similar to phrase-level SA with only a subtle difference. While phrases tend to carry a single sentiment, a sentence might not have a uniform sentiment. A sentence might not just contain multiple sentiment-bearing words, but even multiple entities.

Document level: As the name suggests, document-level sentiment analysis maps individual documents to their respective sentiment. The general approach is to find the sentiment polarity of individual sentences and combine them together to find the polarity of the document. Different approaches use different weighting schemes to combine the sentence-level polarity scores to come up with a document-level score.

1.4.3 Some other Tasks

Other than the general classification and sentiment-level tasks, there are some more higher level tasks related to Sentiment Analysis and Classification. Based on whether the corpus belongs to a single domain or multiple domains, the SA can be carried out in a Domain-dependent way or at a generic level. Researchers till now have mainly focused on domain-dependent

SA. Generic SA approaches typically are unable to replicate the accuracies produced by domain-dependent methods.

Another approach to SA focuses on calculating the sentiment for a single entity. Unlike typical document-level SA where in we assign a sentiment to a document, Entity-level SA aims at associating a sentiment to a specific entity based on the complete corpus. The typical output is a percentage score for positive and negative opinion about the entity.

1.5 Challenges in Sentiment Analysis

All the above discussed tasks and approaches to SA seem pretty much straightforward and intuitive, considering we sub-consciously do all that on a regular basis. However, an in-depth analysis surfaces the problems encountered while trying to implement such concepts on a machine. In this section, we will discuss a few such challenges, not all of which have been fully adressed yet.

- **Identifying subjective portions of text**. The same word can be treated as subjective in one context and objective in some other. This makes it difficult to identify the subjective (sentiment-bearing) portions of text. Consider the following statements:

  - *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 one. Assigning a polarity to a word, thus, does not work until we take the context into account. There have been some works that discuss such approaches.

- **Subtlety of sentiment expression**. This is mainly in cases of *ironic text* or in cases where sentiment is expressed using *neutral words*. Consider the following example:

  - *If you are reading this because it is your darling fragrance, please wear it at home exclusively, and tape the windows shut.*

  While there is no ostensibly negative word being used here, the expression is surely negative. Such subtle expressions are extremely hard to identify for a machine. Not much work has been done to correctly identify the opinion expressed in such sentences.

- **Associating sentiment with specific keywords**. Although many sentences indicate an extremely strong opinion, it is difficult to associate the presence of this strong opinion with specific keywords or phrases in the sentence. 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.*

  While words like *dig, beat, skull* are mildly negative, the opinion expressed here is extremely negative, clearly not being captured by the words. To the best of our knowledge, there has not been any work to identify such phenomena.

- **Domain dependence**. The same sentence or phrase can have different meanings in different domains.

  The word *unpredictable* is positive while talking about some movie-plot, but if the same word is used in the context of a vehicle's steering, then it has a negative connotation.

  There have been some approaches that try to incorporate the contextual meaning of the word in order to capture the correct sentiment.

- **Thwarted expectations**. In certain cases, while the majority of the text has some sentiment orientation, the overall sentiment of the text is quite opposite based on the minor portion of text, which happens to be *more relevant* than the majority. 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 fail drastically in such cases as most of the words used are positive and yet the ultimate sentiment is negative. To the best of our knowledge, there has not been any work on dealing with thwarted expectations.

- **Indirect negation of sentiment**. Typically, we tend to associate the negation of sentiment to words like no, not, never, etc. However, that is not always the case. Certain other words tend to reverse the sentiment polarity implicitly. It is non-trivial to identify such negations in a straight-forward way. Consider the following example:

- *[it]* ***avoids*** *all cliches and predictability found in Hollywood movies.*

  While the words *cliche* and *predictable* bear a negative sentiment in this context, the usage of *avoids* negates their respective sentiments.

- **Order dependence**. Traditional text classification primarily classifies based on the frequencies of the words. The discourse structure does not play a role in classification. However, identifying sentiment while ignoring the syntactical and semantic structure of the sentence does not provide good results. For example:

  - *"A is better than B"* conveys the exact opposite opinion from *"B is better than A"*, even though the words used are exactly the same.

- **Entity Recognition**. Not everything in a text talks about the same entity. When multiple entities are being talked about in a single document, the overall document polarity does not make much sense. 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 statement as neutral, however, it carries a specific sentiment for both the entities being talked about 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. When the author quotes someone else, it becomes difficult to identify whose exactly is that particular opinion.

1.6 Sentiment Analysis at IITB

IIT Bombay has been active in Sentiment Analysis for around half a decade now. Our efforts here have mainly been focussed on generalizing various aspects of SA – to enhance its scope. [1] discussed an approach where they incorporated linguistic knowledge in Sentiment Analysis using WordNet Synonymy Graph. [9], [3], [10] took Sentiment Analysis to a new language, to a new feature space and to a new text form.

Our work on Harnessing Discourse Features takes a leaf from [1]. We try to incorporate more linguistic knowledge in Sentiment Analysis. We also follow the idea developed in [3] by using WordNet synsets instead of words in Sentiment Classification. Our other work on Sentiment Similarity Metric is an extension to the work previously done by [3]. The concept of using a similarity metric for cross-domain sentiment analysis was discussed by [3]. We have extended it by showing that sentiment too is important in measuring such a similarity and should be incorporated, along with semantics to calculate the similarity between word pairs.

We also enhanced the C-Feel-It system discussed in [10] by incorporating various modules like Spam Filter, pragmatics-detector and entity-specific sentiment analyzer.
Most works in SA around the world till now have identified Thwarted Expectations as one of the major unresolved problems. We have worked on various approaches, both statistical as well as rule-based, where in, we try to identify documents with this specific phenomenon. Our work on Review Trees also aims at resolving this hurdle.

1.7 Summary

In this chapter, we understood what basic Sentiment Analysis is. We looked at some typical tasks in Sentiment Analysis and we discussed some of the challenges faced in this field. We also introduced the work being done on Sentiment Analysis here in IIT Bombay.

In the next section, we discuss some of the literature in this field, mainly pertaining to the works that we have done.

Chapter 2

Literature Survey

For the past decade or so, there has been a lot of research that has taken place in Sentiment Analysis. As discussed in chapter 1, most of this research has focussed on developing more and more sophisticated supervised approaches for Sentiment Analysis. We will discuss the works done in various sub-fields of Sentiment Analysis in this chapter.

We start with a discussion of various lexical resources that have been built to aid sentiment prediction in Section 2.1. In section 2.2, we discuss some of the basic supervised and unsupervised approaches that have been implemented in Sentiment Analysis. Section 2.3 delves into work done on Ontology-based Sentiment Analysis. We follow it up with a description of work related to language features and discourse features in Sentiment Analysis in Section 2.4. Finally, section 2.5 describes research on similarity metrics.

2.1 Lexical Resources in Sentiment Analysis

Even though Sentiment Analysis is a relatively new field, the first lexical resources date as early as 1966 [26]. Over the time, many resources have been developed that can provide a polarity value for words. Lexical resources can be developed either *manually* or *automatically*. While manually built lexical resources tend to be more accurate, automatically built resources can attain much higher coverage than their manual counterparts.

However, both types of lexical resources face some similar constraints. As discussed in 1.5, even the same word can carry different sentiment in different contexts. This has been one of the major hurdles faced in creating any lexical resource. Some resources, thus, try to leverage the already existing meaning based word ontologies like WordNet [13]. WordNet is a lexical hierarchical database with nodes represented by word meaning instead of word itself, and relationships between different synonym sets (*synsets*) representing the edges between the nodes. A single node might contain multiple words and a word might be present in more than one nodes. In this section, we will discuss some of the existing sentiment lexicons. We start with a discussion of manually built sentiment lexicons, followed by a discussion on Taboada lexicon, and finally concluding with discussing a couple of automatically constructed lexical resources: SentiWordNet and WordNet Affect. Both these resources are based on WordNet. While there is not much to discuss about the manually built resources, we will delve into the details of the construction of the automatically constructed resources.

2.1.1 Manually Constructed Lexical Resources

One of the first manually created lexical resources is the **General Inquirer** [26]. It consists of 11000+ words compiled from two different sources with roughly 2000 positive and 2000 negative words. Each word is hand tagged as either positive, negative or none. Even though the precision with which Inquirer tags a word is very high, the recall rate is extremely low because of its low coverage.

Another manually annotated and much recent sentiment lexicon is the Bing-Liu Lexicon [7]. It consists of around 2000 positive and 4000 negative words that are used frequently in social media. Each word is manually tagged as either positive or negative. Bing-Liu lexicon scores over General Inquirer in terms of coverage as it contains more sentiment-bearing words and the words in this lexicon are the most frequently appearing sentiment-bearing words on the Internet. However, 6000 is still a very small number of words and hence, has a low precision.

Subjectivity Lexicon [36] is another such resource developed from multiple sources and sentiment tagged manually. It contains 8000+ sentiment words, each tagged as either positive or negative.

Apart from the coverage issue, another problem with manually-tagged sentiment lexicons is that they have binary/tertiary rating. There is no concept of something like *mildly positive* or *strongly negative*, or similar expressions. This becomes possible in case of automatically constructed resources.

2.1.2 Taboada

Taboada [27] is created out of a huge corpus measuring the frequencies of a list of words in sentiment context. It contains roughly 1700 words with three scores, namely, the occurences of that word in positive, negative and objective context. The scores can be normalized to get scores similar to SentiWordNet, all ranging from 0.0 to 1.0, and adding up to one.
The unique feature of Taboada is that. while the words appearing in it are manually selected, the scores are assigned automatically using corpus frequency.

2.1.3 SentiWordNet

SentiWordNet[5] is a lexical resource that is basically an extension to WordNet. It appends sentiment information to every synset in WordNet. It tags each synset with three scores: a Positive, a Negative and an Objective score. The three scores sum up to 1.0. The basic intuition is that the same word can have different meanings, and correspondingly, different sentiment orientations, but the words in the same synset should reflect a specific sentiment, as they define a particular meaning.

Each of the three scores ranges from 0.0 to 1.0, and sum up to 1.0 for each synset. The advantage of this approach is that it does not fix a synsets sentiment to a single category. This aids the fuzzy concept that the same synset can be positive to some extent, but can also have a bit of negative flavour at the same time. It can also be interpreted as saying that the same synset is positive in certain context, and negative in some other context, and thus, has a non-zero score for both the classes.

Building SentiWordNet

A pool of ternary classifiers (total eight in number) classify each synset into one of the three categories of Positive, Negative and Objective. The classification is based on a quantitative analysis of the glosses associated to synsets, using the resulting vectorial term representations for semi-supervised synset classification.

The three scores are derived by combining the results produced by combining the results produced by the committee of these **eight** ternary classifiers.

Each ternary classifier differs from the others in the training set used to train it and in the learning device used to train it, thus producing different classification results of the WordNet synsets.

If all the ternary classifiers agree in assigning the same label to a synset, that label will have the maximum score for that synset, otherwise each label will have a score proportional to the number of classifiers that have assigned it.

Training a Classifier

The classifiers are trained using semi-supervised approach. A semi-supervised approach is a learning process whereby only a small subset $L \subset Tr$ of the training data Tr have been manually labelled. The remaining $U = Tr - L$ instances in the training data are instead unlabelled. These instances are then automatically labelled by the algorithm, based on the labels of the labelled data L.

The set L itself is a union of three seed sets L_p, L_n, and L_o of known Positive, Negative and Objective synsets, respectively. L_p and L_n are two small sets, which are defined by manually selecting the intended synsets of 144 paradigmatic Positive and Negative terms.

Seed-set Expansion

L_p and L_n are iteratively expanded in K iterations, into the final training sets Tr_K^p and Tr_K^n. At each iteration step k, two sets Tr_k^p and Tr_k^n are generated, where $Tr_k^p \supset Tr_{k-1}^n \supset \ldots \supset Tr_1^p = L_p$ and $Tr_k^n \supset Tr_{k-1}^n \supset \ldots \supset Tr_1^n = L_n$. The expansion at iteration step k consists of adding:

- to Tr_k^p (resp. Tr_k^n) all the synsets that are connected to synsets in Tr_{k-1}^p (resp. Tr_{k-1}^n) by WordNet lexical relations (e.g. also-see) such that the two related synsets can be taken to have the same PN-polarity. E.g., if the word goodness is in the set Tr_{k-1}^p then words like good (similarity relationship) and morality (also-see relationship) will be added to the set Tr_k^p.

- to Tr_k^p (resp. Tr_k^n) all the synsets that are connected to synsets in Tr_{k-1}^n (resp. Tr_{k-1}^p) by WordNet lexical relations (e.g. direct antonymy) such that the two related synsets can be taken to have opposite PN-polarity. E.g., if the word goodness is in the set Tr_{k-1}^p then words like evil (antonymy relationship) will be added to the set Tr_k^n.

The WordNet relations that have been considered in this iterative algorithm are *direct antonymy, similarity, derived from, pertains-to, attribute, and also-see.*

Forming the Objective Set L_o

The objective set L_o has to be handled differently. The objective synsets are complementary in nature. Any synset which is neither positive, nor negative is considered objective. A synset is considered Objective if it satisfies the following conditions.

- It neither belongs to the set Tr_k^p, nor to the set Tr_k^n
- None of the words it contains are tagged Positive or Negative in the General Inquirer [26]

The set Tr_k^o is then said to coincide with the set L_o.

Defining the Committee of Classifiers

[5] have used a total of eight ternary classifiers differing in the training sets and the learning algorithm used. Four different training sets were used, by choosing four different values of K (0, 2, 4, 6), and with each training set, two learners (Rocchio and SVM) were used to yield the eight ternary classifiers.

2.1.4 WordNet Affect

WordNet Affect by [29] is a lexical resource that represents the affective content of synsets by dividing them into affective categories. Thus, it gives more affective information as compared to SentiWordNet and is used when analysis to be done is with respect to emotions like anger, joy, *etc.*

WordNet Affect is constructed using the following steps:

- Select the core synsets that represent the affective content
- Expand these synsets by attaching to them, synsets related to them in the WordNet. The relations considered could be antonymy, derived-from, *etc*
- Label the clusters of synsets by affective domain labels (a-labels) that represent the affect that they belong to

2.2 Basic Approaches to Sentiment Classification

Sentiment Classification is the task of classifying the given documents into various classes. Any general Sentiment Classification task can be formulated as follows:

"Given an opinionated piece of text, wherein it is assumed that the overall opinion in it is about one single issue or item, we need to classify the opinion as falling under one of two opposing sentiment polarities, or locate its position on the continuum between these two polarities."

Some of the earliest works in Sentiment Classification were done by [28] and [16]. While [28] came up with an unsupervised approach defining the *semantic orientation* of words to aid sentiment-based classification of reviews, [16] implemented a bunch of supervised approaches for sentiment classification in order to maximize the accuracies obtained. In this section, we will discuss these approaches and a few more similar approaches to Sentiment Classification.

2.2.1 An Unsupervised Approach to Sentiment Classification

[28] deivsed an unsupervised learning algorithm that can classify any given review as either *thumbs up* or *thumbs down*. The algorithm can be divided into three steps:

- Identify phrases in the input text that contain adjectives or adverbs. The assumption here is that the modifier words like adjectives and adverbs are the most important sentiment-bearing words.
- The next step is to estimate the *Semantic Orientation* (SO) of each extracted phrase. Semantic Orientation is calculated using the PMI-IR algorithm. Both Semantic Orientation and PMI-IR algorithm are described later.
- The final step is to assign the review to one of the classes, *recommended* or *not recommended*, based on the average semantic orientation of the phrases extracted from the review.

Semantic Orientation

The semantic orientation of any phrase is said to be positive, if it has good association (e.g. romantic ambience) and it is said to be negative, if the associations are bad (e.g. horrific events).

Domain	Entity	Accuracy
Automobiles	Honda Accord	83.78%
	Volkswagen Jetta	84.21%
Banks	Bank of America	78.33%
	Washington Mutual	81.67%
Movies	The Matrix	66.67%
	Pearl Harbour	65.00%
Travel Destinations	Cancun	64.41%
	Puerto Vellarta	80.56%
All		74.39%

Table 2.1: Results for Semantic Orientation based Sentiment Classifcation

Semantic Orientation is calculated using the Point-wise Mutual Information (PMI) and Information Retrieval (IR), which give a similarity of pairs of words or phrases.

$$SO(phrase) = PMI(phrase, excellent) - PMI(phrase, poor)$$

[28] experimented on reviews on eight different topics from four varied domains. The accuracies ranged from 65.83% on reviews in movie domain to 84.00% on reviews in automobile domain. The complete set of results can be found in table 2.1[1].

2.2.2 Supervised Approaches to Sentiment Classification

[16] carried out one of the first set of experiments in Sentiment Analysis using basic supervised approaches, quite similar to techniques for standard text-based categorization. They implemented Naive Bayes', MaxEnt as well as an SVM-based classifier on a simple bag-of-words feature set for sentiment classification. They tested these systems on a Movie-review dataset generated from IMDB. With these experiments, they received a best-case accuracy of **82.9%** in case of unigram features with an SVM classifier. During this work, they also tried out bigram features and POS-specific features, however, SVMs with unigrams provided the best results. This work has served as the

[1]The table has been adopted from [28]

baseline for majority of the future works in supervised sentiment classification.

[15] came up with a more sophisticated approach that was based on the intuition that only the subjective portions of the text actually affect the polarity of the document. Hence, they implemented a min-cut based subjectivity classifier before actual sentiment analysis could take place. The subjectivity classifier produces a min-cut based classification using a Naive Bayesian classifier to separate the subjective portions from the objective portions of the text. The subjectivity extract (all subjective sentences from the subjectivity classifier) are then passed on to a general polarity classifier (typically SVM) to decide the overall polarity of the text. The basic flow of the algorithm is described in figure 2.1.

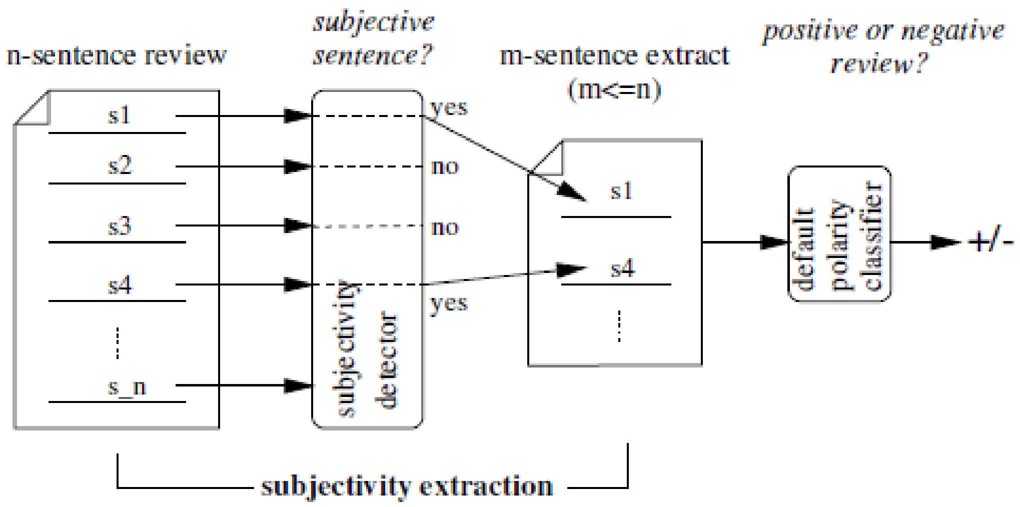

Figure 2.1: Flow chart for min-cut based subjectivity detection

Using this approach, [15] showed that taking the subjectivity extracts out and applying sentiment analysis to it alone gives better accuracies than the baseline, *i.e.*, applying sentiment analysis to the complete text. They achieved an accuracy of around **86%**, which is an improvement of 4% over the baseline. In their experiments, they also showed that a subjectivity extract that produces this accuracy contains only 60% of the original text size. Thus, this approach benefits on two fronts, namely,

- improving the accuaracy over the baseline

- summarizing the sentiment content of the review

Both these approaches tend to be naive, though, as they focus on using simple bag-of-words based features for sentiment classification. Instead, [12] discuss the usage of syntactic relations between words in the same sentence for sentiment classification. The underlying philosophy is that other than n-grams (a general bag-of-words), word order and syntactic relations between words, too, are intuitively important in sentiment classification. They use word sub-sequences and dependency sub-trees to represent word order and syntactic relationships respectively. The results show an increase of around 5% in the accuracy over the baseline (bag-of-words).

The approaches mentioned above give a flavor of some of the common approaches that have been explored in sentiment analysis. A lot more modifications and innovations have happened since, however, a discussion on those is out of scope of this report. In the following sections, we will discuss the literature specifically relevant to our work during this project.

2.3 Ontology based Sentiment Analysis

There has not been a lot of work done in the field of sentiment analysis pertaining to the use of ontology trees. There have been some works that involve both sentiment analysis, as well as, ontology-based learning, however, they do not exactly utilize the ontology structure for sentiment analysis.

[19] present a method of ontology-based sentiment classification to classify and analyse online product reviews. They implement and experiment with Support Vector Machine based on the lexical variation ontology. They implement the Mixed Min and Max Model (MMM Model) to extract the *important* features out of all the bag-of-words features and apply SVM for classification. They utilize the lexical variation ontology to reduce all words refering to the same term to a single entity.

[31] propose a novel HL-SOT approach to labeling a products attributes and their associated sentiments in product reviews by a Hierarchical Learning (HL) process with a defined Sentiment Ontology Tree (SOT). The tree is constructed manually with 35 feature terms pertaining to cameras. They define an extra couple of leaf nodes for every feature node corresponding to the node's positive and negative sentiment. They assign the whole review to a single leaf node depending on if the review talks about all the nodes

between that particular node and the root.

While the HL-SOT approach is similar to our review-tree based approach as far as the construction of the ontology tree is concerned, the approach differs in classifying reviews to a parameter of the product. We use the ontology tree for the general task of sentiment analysis and for identifying thwarted documents. While their approach tends to classify a review as positive or negative pertaining to a single feature, we believe that a sentiment analysis approach should provide an overall score for the review, along with the sentiment scores for *each* feature present in the review.

2.4 Discourse Features for Sentiment Analysis

In Rhetorical Structure Theory [25] developed probabilistic models for identifying elementary discourse units at clausal level and generating trees at the sentence level using lexical and syntactic information from discourse-annotated corpus of RST.

[32] considered the problem of automatically identifying arguments of discourse connectives in the PDTB. They modeled the problem as a predicate-argument identification where the predicates were discourse connectives and arguments served as anchors for discourse segments.

[21] report the development of a Conditional Random Field based model for automatic extraction of discourse connectives in the biomedical domain. They explored two collections of annotated data (Penn Discourse Tree Bank (PDTB) and The Biomedical Discourse Relation Bank (BioDRB)) for training models to identify discourse connectives. They used standard bag-of-words features, morphological and n-gram features for the task.

[14] performed an extensive study of the effect of conditional sentences in Sentiment Analysis. They found that around 8% sentences in a typical document consist of conditional sentences which are difficult to analyze from the point of view of SA due to the conditional clauses. The authors built a supervised learning model to determine if the sentiment expressed by the conditional statements is positive, negative or objective. They used words as features with their POS tags, position in the head or body of the clause, tense pattern, length, negation words and some other syntactic features.

All the previous works, except for [14], focus on extracting discourse segments or connectives from the text and identifying their span. We view the problem from the angle of the effect these discourse elements have in analyzing the sentiment in the document. All discourse elements are not essential for SA. Some of them enhance the sentiment, some express hypothetical situations, some reverse the sentiment polarity and some of them express cause-effect relations. We give more importance to those linguistic constructs that either reinforce sentiment or reverse them while ignoring those that express irrealis events.

[37] present a set of discourse structure relations and way to code or represent them. They report a method for annotating discourse coherent structures and found different kinds of crossed dependencies.

In the work, Contextual Valence Shifters [18], the authors investigate the effect of intensifiers, negators, and modals and connectors that change the prior polarity or valence of the words and brings out a new meaning or perspective. They also talk about pre-suppositional items, like irony and present a simple weighting scheme to deal with them. The authors look at the effect of multi-entity evaluation and the genre or attitude assessment of the speaker. They also present their analysis on reported speech and sub-topics.

[33], [34], [35] presented an extensive analysis of tracking discourse in narrative. But many of those were quite difficult and complicated to express computationally.

Our work on discourse coherent features builds on the work of [18] and [37] and carries the idea further.

2.5 Similarity Metrics using Sentiment Analysis

Various approaches for evaluating the similarity between two words can be broadly classified into two categories: edge-based methods and information-content-based methods. One of the earliest works in edge-based calculation of similarity is by [20], where in, they propose a metric "Distance" over a semantic net of hierarchical relations as the shortest path length between the two nodes. This has been the basis for all the metrics involving simple edge-counting to calculate the distance between two nodes. However, the simple edge-counting fails to consider the variable density of nodes across the taxonomy. It also fails to include relationships other than the is-a relationship, thus, missing out on important information in a generic semantic ontology, like WordNet.

In contrast to edge-based methods, [24] and [22] propose a node-based approach to find the semantic similarity. They approximate conceptual similarity between two WordNet concepts as the maximum information content among classes that subsume both the concepts. [23] advanced this idea by defining the information content of a concept based on the probability of encountering an instance of that concept. Alternatively, [39] compare two concepts based on the length of the path between the root of the hierarchy and the least common subsumer of the concepts.

[8] and [11] combine the above two approaches by using the information content as weights for the edges between concepts. They further reinforce the definition of information content of a concept by adding corpus statistical information.

Instead of measuring the similarity of concepts, some other approaches measure their relatedness. [6] introduce an additional notion of direction along with the length of paths for measuring the relatedness of two concepts. [4] and [17] leverage the gloss information present in WordNet in order to calculate the relatedness of two concepts. [4] assigns relatedness scores based on the overlap between the gloss of the two concepts. [17] use a vector representation of the gloss, based on the context vector of the terms in the gloss. The relatedness is then the cosine between the gloss vectors of the two concepts.

Our work is most related to the work of [30] which improves on [4] and [17] by including relations other than the is-a relationship too. They use an extended gloss definition for a concept which is defined as the original gloss appended by the gloss of all the concepts related to the given concept. They create concept vectors for each sense, based on which, they create context vectors which are an order higher to the concept vectors. Finally, they use cosine of the angle between the vectors of the different concepts to find their relatedness. This approach is better than other approaches in that it captures the context of the concepts to a much larger extent. However, all these methods lack on a common ground. They fail to incorporate sentiment information in calculating the similarity/relatedness of two concepts. We postulate that sentiment information is crucial in finding the similarity between two concepts.

2.6 Summary

We started this chapter with a discussion on the lexical resources and some of the works done in general Sentiment Analysis. We moved on to discussing literature pertaining to our specific works. With the next chapter, we start discussing the work that we did over the course of this project. We start with a discussion on *Thwarting* and some of our approaches to tackle this phenomenon in the next chapter.

References

[1] A. Agarwal and P. Bhattacharyya. Sentiment Analysis: A New Approach for Effective Use of Linguistic Knowledge and Exploiting Similarities in a Set of Documents to be Classified. In *International Conference on Natural Language Processing (ICON*, IIT Kanpur, India, December 2005.

[2] A. Anangnostopoulos, R. Kumar, and M. Mohammad. Influence and Correlation in Social Networks. In *KDD 2008*, Las Vegas, Nevada, USA, 2008.

[3] A. R. Balamurali, A. Joshi, and P. Bhattacharyya. Harnessing WordNet Senses for Supervised Sentiment Classification. In *EMNLP*, Edinburgh, UK, 2011.

[4] S. Banerjee and T. Pedersen. Extended gloss overlaps as a measure of semantic relatedness. In *Proceedings of the Eighteenth International Joint Conference on Artificial Intelligence*, pages 805–810, 2003.

[5] A. Esuli and F. Sebastiani. SentiWordNet: A Publicly Available Lexical Resource for Opinion Mining. In *International Conference on Language Resources and Evaluation (LREC)*, Genoa, 2006.

[6] G. Hirst and D. St-Onge. Lexical chains as representation of context for the detection and correction malapropisms. *Fellbaum C.*, 1997.

[7] M. Hu and B. Liu. Mining and summarizing customer reviews. Seattle, Washington, USA, August 2004.

[8] J. J. Jiang and D. W. Conrath. Semantic Similarity Based on Corpus Statistics and Lexical Taxonomy. In *International Conference Research on Computational Linguistics (ROCLING X)*, page 9008, 1997.

[9] A. Joshi, A. R. Balamurali, and P. Bhattacharyya. A Fall-back Strategy for sentiment analysis in Hindi: a case study. In *International*

Conference on Natural Language Processing (ICON), Kharagpur, India, December 2010.

[10] A. Joshi, A. R. Balamurali, P. Bhattacharyya, and R. Mohanty. C-Feel-It: A Sentiment Analyzer for Micro-blogs. In *Annual Meeting of the Association of Computational Linguistics (ACL 2011)*, Oregon, USA, June 2011.

[11] C. Leacock, G. A. Miller, and M. Chodorow. Using corpus statistics and wordnet relations for sense identification. *Computational Linguistics*, 24:147–165, 1998.

[12] S. Matsumoto, H. Takamura, and M. Okumura. Sentiment classification using word sub-sequences and dependency sub-trees. In *Proceedings of PAKDD' 05*, 2005.

[13] G. A. Miller, R. Beckwith, C. Fellbaum, D. Gross, and K. Miller. WordNet: An on-line Lexical Database. *International Journal of Lexicography*, 3:235–244, 1990.

[14] R. Narayanan, B. Liu, and A. Choudhary. Sentiment Analysis of Conditional Sentences. In *Proceedings of Conference on Empirical Methods in Natural Language Processing (EMNLP-09)*, 2009.

[15] B. Pang and L. Lee. A sentimental education: Sentiment analysis using subjectivity summarization based on minimum cuts. In *Proceedings of ACL-04, 42nd Meeting of the Association for Computational Linguistics*, pages 271–278, Barcelona, ES, 2004.

[16] B. Pang, L. Lee, and S. Vaithyanathan. Thumbs up? Sentiment Classification using Machine Learning Techniques. In *Proceedings of EMNLP 2002*, pages 79–86, 2002.

[17] S. Patwardhan. Incorporating dictionary and corpus information into a context vector measure of semantic relatedness. *Master's Thesis, University of Minnesota, Duluth*, 2003.

[18] L. Polanyi and A. Zaenen. Contextual Valence Shifters. *James G. Shanahan, Yan Qu, Janyce Wiebe (eds.), Computing Attitude and Affect in Text: Theory and Applications*, pages 1–10, 2004.

[19] J. Polpinij and A. K. Ghose. An Ontology-Based Sentiment Classification Methodology for Online Consumer Reviews. pages 518–524, 2008.

[20] R. Rada, H. Mili, E. Bicknell, and M. Blettner. Development and application of a metric on semantic nets. *IEEE Transactions on Systems Management and Cybernetics*, 19:17–30, 1989.

[21] B. P. Ramesh, R. Prasad, and H. Yu. Identifying explicit discourse connective in biomedical text. *Annual Symposium proceedings, AMIA Symposium*, 2010:657–661, 2010.

[22] P. Resnik. Disambiguating noun groupings with respect to Wordnet senses. *D. Yarovsky and K. Church, eds, Proceedings of the Third Workshop on Very Large Corpora, Association for Computational Linguistics*, pages 54–68, 1995a.

[23] P. Resnik. Using information content to evaluate semantic similarity in a taxonomy. In *Proceedings of the 14th International Joint Conference on Artificial Intelligence*, pages 448–453, 1995b.

[24] R. Richardson, A. F. Smeaton, and J. Murphy. Using wordnet as a knowledge base for measuring semantic similarity between words. In *Technical report, In Proceedings of AICS Conference*, 1994.

[25] R. Soricut and D. Marcu. Sentence level discourse parsing using syntactic and lexical information. In *Proceedings of HLT-NAACL 2003*, 2003.

[26] P. J. Stone, D. C. Dunphy, M. S. Smith, D. M. Ogilvie, and Associates. The General Inquirer: A Computer Approach to Content Analysis. *The MIT Press*, 1966.

[27] M. Taboada and J. Grieve. Analyzing Appraisal Automatically. pages 158–161, Stanford, US, 2004.

[28] P. Turney. Thumbs up or Thumbs down: Semantic orientation applied to unsupervised classification of reviews. In *ACL 2002*, pages 417–424. Association of Computational Linguistics, 2002.

[29] R. Valitutti. Wordnet affect: An affective extension of wordnet. In *4th International Conference on Language Resources and Evaluation*, 2004.

[30] S. Wan and R. A. Angryk. Measuring Semantic Similarity Using WordNet-based Context Vectors. *SMC '07*, pages 908–913, 2007.

[31] W. Wei and J. A. Gulla. Sentiment learning on product reviews via sentiment ontology tree. In *Proceedings of the 48th Annual Meeting of*

the Association for Computational Linguistics, ACL '10, pages 404–413, Stroudsburg, PA, USA, 2010. Association for Computational Linguistics.

[32] B. Wellner, J. Pustejovski, A. Havasi, A. Rumshisky, and R. Suair. Classification of discourse coherence relations: An exploratory study using multiple knowledge sources. In *Proceedings of SIGDIAL2006*, 2006.

[33] J. Wiebe. Identifying Subjective Characters in Text. In *Proc. 13th International Conference on Computational Linguistics (COLING-90)*, pages 410–418, 1990.

[34] J. Wiebe. References in Narrative Text. *Nous (Special issue on Cognitive Science and AI)*, 25:457–486, 1991.

[35] J. Wiebe. Tracking Point of View in Narrative. *Computational Linguistics*, 20:233–287, 1994.

[36] T. Wilson, J. Wiebe, and P. Hoffmann. Recognizing Contextual Polarity in Phrase-Level Sentiment Analysis. 2005.

[37] F. Wolf and E. Gibson. Representing Discourse Coherence: A Corpus-based Study. *Computational Linguistics*, 31:249–287, 2005.

[38] A. Wright. Mining the web for feelings, not facts. *NY Times, http://www.nytimes.com/2009/08/24/technology/internet/24emotion.html?_r=1.*, August 2009.

[39] Z. Wu and M. Palmer. Verb Semantics and Lexical Selection. In *32nd. Annual Meeting of the Association for Computational Linguistics*, pages 133–138, New Mexico State University, Las Cruces, New Mexico, 1994.