


Sentiment Analysis: An Overview

Yelena Mejova

Related papers

[Download a PDF Pack](#) of the best related papers 



[Hybrid Sentiment Analysis Framework for a Morphologically Rich Language](#)

Jelena Mitrović, Miljana Mladenovic

[Subgroup detection in ideological discussions](#)

Mona Diab

[Sentiment Analysis of Events from Twitter Using Open Source Tool](#)

Vishal Shirsat, Rajkumar S Jagdale

Sentiment Analysis: An Overview

Comprehensive Exam Paper

Yelena Mejova
Computer Science Department,
University of Iowa
yelena-mejova@uiowa.edu

November 16, 2009

Abstract

As a response to the growing availability of informal, opinionated texts like blog posts and product review websites, a field of Sentiment Analysis has sprung up in the past decade to address the question *What do people feel about a certain topic?* Bringing together researchers in computer science, computational linguistics, data mining, psychology, and even sociology, Sentiment Analysis expands the traditional fact-based text analysis to enable opinion-oriented information systems.

This paper is an overview of Sentiment Analysis, its basic tasks and the latest techniques developed to address the challenges of working with emotionally-charged text.

<i>CONTENTS</i>	2
-----------------	---

Contents

1 Introduction	3
2 Definitions	3
2.1 What is Sentiment?	3
2.2 Sentiment Analysis	5
2.3 Examples of Sentiment Research	6
3 Goals of Sentiment Analysis	6
4 Methodologies	8
4.1 Classification	9
4.2 Identifying the semantic orientation of words	10
4.3 Identifying semantic orientation of sentences and phrases	15
4.4 Identifying the semantic orientation of documents	15
4.5 Object feature extraction	17
4.6 Comparative Sentence Identification	18
4.7 Performance Achieved	20
5 General Questions	20
6 Commercial Uses	23
7 Open Research Directions	24
8 Conclusions	26
A A selection of lists of “fundamental” or “basic” emotions	28

1 Introduction

When conducting serious research or making every-day decisions, we often look for other people's opinions. We consult political discussion forums when casting a political vote, read consumer reports when buying appliances, ask friends to recommend a restaurant for the evening. And now Internet has made it possible to find out the opinions of millions of people on everything from latest gadgets to political philosophies. The latest Pew study on Internet and Civic engagement says that "just under one in five internet users (19%) have posted material about political or social issues or a used a social networking site for some form of civic or political engagement"¹. Another study shows that a third (33%) of internet users read blogs, with 11% doing so on daily basis². Internet is increasingly both the forum for discussion and source of information for a growing number of people.

Ready availability of opinionated text has created a new area in text analysis, expanding the subject of study from traditionally fact- and information-centric view of text to enable sentiment-aware applications. In the past decade, extraction of sentiment from text has been getting a lot of attention in both industry and academia. Increasingly businesses realize the importance of Internet users' opinions about their product and services.

This paper is an overview of the area of Sentiment Analysis, which deals with subjective texts. Our first task is to define sentiment and delineate its relation to text.

2 Definitions

2.1 What is Sentiment?

One of the challenges of Sentiment Analysis is defining the objects of the study – opinions and subjectivity. Originally, subjectivity was defined by linguists, most prominently, Randolph Quirk (R. Quirk and Svartvik, 1985). Quirk defines *private state* as something that is not open to objective observation or verification. These private states include emotions, opinions, and speculations, among others. The very definition of a private state foreshadows difficulties in analyzing sentiment. Subjectivity is often implied in conversation, it is highly context-sensitive, and its expression is often peculiar to each person. Note, however, that *subjective* does not imply *not true* (Wiebe et al., 2004). The sentence "Mary loves chocolate"

¹<http://www.pewinternet.org/Reports/2009/15-The-Internet-and-Civic-Engagement.aspx>

²<http://www.pewinternet.org/Commentary/2008/July/New-numbers-for-blogging-and-blog-readership.aspx>

expresses a sentiment of Mary towards chocolate, but it doesn't mean it's not true. Likewise, not all objective sentences are true.

To underline the ambiguity of the concept, Pang and Lee (Pang and Lee, 2008) list the definitions of terms closely linked to the notion of sentiment:

- *Opinion* implies a conclusion thought out yet open to dispute (*"each expert seemed to have a different opinion"*).
- *View* suggests a subjective opinion (*"very assertive in stating his views"*).
- *Belief* implies often deliberate acceptance and intellectual assent (*"a firm belief in her party's platform"*).
- *Conviction* applies to a party's firmly and seriously held belief (*"the conviction that animal life is as sacred as human"*).
- *Persuasion* suggests a belief grounded on assurance (as by evidence) of its truth (*"was of the persuasion that everything changes"*).
- *Sentiment* suggests a settled opinion reflective of one's feelings (*"her feminist sentiments are well-known"*).

Wiebe, a prominent Natural Language Processing (NLP) researcher, used Quirk's definition of the *private state* when tracking point of view in narrative (Wiebe, 1994). She defines private state as a tuple

$$(p, \text{experiencer}, \text{attitude}, \text{object})$$

relating experiencer's state p to his/her attitude possibly toward an object. In practice, a simplified version of this model, where we look only at polarity and the target of the sentiment, is usually used. In fact, many researchers define sentiment loosely, as a negative or positive opinion (Pang and Lee, 2002; Hu and Liu, 2005; Melville et al., 2009).

Sentiment also has several unique properties that set it apart from other qualities that we may want to track in text. Often we want to categorize text by topic, which may involve dealing with whole taxonomies of topics. Sentiment classification, on the other hand, usually deals with two classes (positive vs. negative), a range of polarity (e.g. star ratings for movies), or even a range in strength of opinion (Pang and Lee, 2008). These classes span many topics and users and kinds of documents. Although dealing with only a few classes may seem like an easier task than standard text analysis, it couldn't be further from the truth.

2.2 Sentiment Analysis

As a field of research, it is closely related to (or can be considered a part of) computational linguistics, natural language processing, and text mining. Proceeding from the study of affective state (psychology) and judgment (appraisal theory), this field seeks to answer questions long studied in other areas of discourse using new tools provided by data mining and computational linguistics.

Sentiment Analysis has many names. It's often referred to as *subjectivity analysis*, *opinion mining*, and *appraisal extraction*, with some connections to *affective computing* (computer recognition and expression of emotion) (Pang and Lee, 2008). The field usually studies *subjective elements*, defined by Wiebe et. al. as "linguistic expressions of private states in context" (Wiebe et al., 2004). These are usually single words, phrases, or sentences. Sometimes whole documents are studied as a sentiment unit (Turney and Littman, 2003; Agrawal et al., 2003), but it's generally agreed that sentiment resides in smaller linguistic units (Pang and Lee, 2008). Since *sentiment* and *opinion* often refer to the same idea, this paper will use the terms interchangeably.

Sentiment that appears in text comes in two flavors: *explicit* where the subjective sentence directly expresses an opinion ("It's a beautiful day"), and *implicit* where the text implies an opinion ("The earphone broke in two days") (Liu, 2006). Most of the work done so far focuses on the first kind of sentiment, since it is the easier one to analyze.

Sentiment polarity is a particular feature of text. It is usually dichotomised into two – positive and negative – but polarity can also be thought of as a range. A document containing several opinionated statements would have a mixed polarity overall, which is different from not having a polarity at all (being objective). Furthermore, a distinction must be made between the polarity of sentiment and of its strength. One may feel strongly about a product being *OK*, not particularly good or bad; or weakly about a product being very good (because perhaps one owned it for too short time to form a strong opinion).

Another important part of sentiment is its target - an object, a concept, a person, anything. Most work has been done on product and movie reviews, where it is easy to identify the topic of the text. But it is often useful to pay attention to which feature of this object the writer is talking about: is it the camera display or battery life that troubles consumers the most? Because of ready availability of product review datasets, feature extraction has been closely studied in the past decade (Liu, 2006; Hu and Liu, 2005; Popescu and Etzioni, 2005). The mention of these features in text can also be *explicit* ("Battery life is too short") or *implicit* ("Camera is too large") (Liu, 2006).

Unlike in usual topical analysis, sentiment statement authorship can be integral to the problem. One of the main problems is quotation. It is important to

know that the sentiment expressed in the document is representative of the actual intent of the author. Political commentaries and news are full of quotations and opinion citations, and can have a convoluted structure that is difficult to discern. For example, a news article about a political debate would have a mix of quotations from the debaters, the pundits commenting on the debate, and perhaps even the author's stance on the issues.

2.3 Examples of Sentiment Research

As mentioned before, some of the most studied texts in Sentiment Analysis are product and movie reviews (Hu and Liu, 2005; Popescu and Etzioni, 2005). The advantage is that they already have a clearly specified topic, and it is often (reasonably) assumed that the sentiments expressed in the reviews have to do with the topic. Many also have a star rating system, which serves as a quantitative indication of the opinion. Such data is often used as gold standard while evaluating sentiment extraction/identification. A general task aimed at sentiment research would be to find opinions on a given product in any web content. Several companies offer services in brand tracking and market perception use Sentiment Analysis techniques. For example, OpSec Security³ provides “monitoring, measuring, and analyzing consumer feedback” to their customers, helping them understand the market needs, target customer segments, and their position against competitors.

On the other hand, one of the most difficult areas for Sentiment Analysis methods is that of politics. Political discussions are fraught with quotations, sarcasm, and complex references to persons, organizations, and ideas (Gamon et al., 2008). Some work has been done on determining whether a political speech is in support of or in opposition to the issue under debate (Bansal et al., 2008; Thomas and B. Pang, 2006). There is a related work on categorizing election forums into “likely to win” and “unlikely to win” (Kim and Hovy, 2007). This problem of complex discussions will be further addressed in **Open Research Directions** section.

3 Goals of Sentiment Analysis

Because of the complexity of the problem (underlying concepts, expressions in text, etc.), Sentiment Analysis encompasses several separate tasks. These are usually combined to produce some knowledge about the opinions found in text. This section provides an overview of these tasks, and the next will discuss some of the tools that are used for each.

³<http://www.opsecsecurity.com/>

The first task is **sentiment or opinion detection**, which may be viewed as classification of text as objective or subjective. Usually opinion detection is based on the examination of adjectives in sentences. For example, the polarity of “this is a beautiful picture” can be determined easily by looking at the adjective. An early study by Hatzivassiloglou (Hatzivassiloglou and Wiebe, 2000) examines the effects of adjectives in sentence subjectivity. More recent studies (Benamara et al., 2007) have shown that adverbs may be used for similar purpose. A survey of subjectivity recognition techniques can be found in (Wiebe et al., 2004).

The second task is that of **polarity classification**. Given an opinionated piece of text, the goal is to classify the opinion as falling under one of two opposing sentiment polarities, or locate its position on the continuum between these two polarities (Pang and Lee, 2008). When viewed as a binary feature, polarity classification is the binary classification task of labeling an opinionated document as expressing either an overall positive or an overall negative opinion. Most of this research was done on product reviews, where the definitions of “positive” and “negative” are clear. Other tasks, such as classifying news as “good” or “bad” present some difficulty. A news article may contain “bad” news without actually using any subjective terms. Furthermore, these classes usually appear intermixed when a document expresses both positive and negative sentiments. Then the task can be to identify the main sentiment of the document.

To distinguish between different mixtures of the two opposites, polarity classification uses a **multi-point scale** (such as the number of stars for a movie review). This is where the task becomes a multi-class text categorization problem. But unlike the topic-based multi-class classification problems where vocabularies differ for each class (or overlap slightly), the vocabularies for positive, neutral, and negative classes can be very much alike, and differ only in few crucial words. Since many documents have a “mixed” opinion, this class is actually a combination of positive and negative. Negations, which tend to be disregarded in much of text analysis as unimportant, play an important role in sentiment, flipping an originally positive term into negative, and vice versa (see section 4.1.5 for more on negations).

The above two tasks can be done at several **levels**: term, phrase, sentence, or document level. It is common to use the output of one level as the input for the higher layers (Turney and Littman, 2003; Dave et al., 2003; Kanayama et al., 2004). For instance, we may apply sentiment analysis to phrases, and then use this information to evaluate sentences, then paragraphs, etc. Different techniques are suitable for different levels. Techniques using n-gram classifiers or lexicons usually work on term level, whereas Part-Of-Speech tagging is used for phrase and sentence analysis. Heuristics are often used to generalize the sentiment to document level.

A third task that is complementary to sentiment identification is the **discovery**

of the opinion’s target. The difficulty of this task depends largely on the domain of the analysis. As mentioned earlier, it is usually safe to assume that product reviews usually talk about the specified product. On the other hand, general writing such as webpages and blogs don’t always have a pre-defined topic, and often mention many objects. **Another lively area of research is feature extraction, given an object or topic of the text** (Liu et al., 2005; Popescu and Etzioni, 2005; Hu and Liu, 2005). Liu et. al. define features as either *components* or *attributes* of an object (Liu, 2006), which is a definition that is mostly used in practice. An example of features extracted for a scanner, for example, can be found in Table 1 in section 4.5. Breaking down the discussion into features allows for a more precise analysis of the sentiments, and for a more detailed summarization of the results.

Sometimes there is more than one target in a sentiment sentence, which is the case in **comparative sentences**. A subjective comparative sentence orders objects in order of preferences, for example, “this camera is better than my old one”. These sentences can be identified using comparative adjectives and adverbs (*more, less, better, longer*), superlative adjectives (*most, least, best*) and other words such as *same, differ, win, prefer*, etc. (Liu, 2006) Once the sentences have been retrieved, the objects can be put in an order that is most representative of their merits, as described in text.

One of the peculiarities of sentiment is that even though the notion of positive and negative opinion is a general one, the expression of these opinions differs widely across the spectrum of topical domains. Thus, topic-specific and cross-topic sentiment analysis is studied in order to improve performance in a particular domain. Here, combining general knowledge about the expression of sentiment and topic-specific one is an important issue. In cross-topic analysis, the idea is to use the knowledge gathered about one domain in another (perhaps from one with labeled data to one without) (Nigam and Hurst, 2004; Blitzer et al., 2007).

4 Methodologies

A wide range of tools and techniques are used to tackle the goals described above. This section describes some of the most common and interesting ones. First, Machine Learning and Part-Of-Speech tagging will be discussed, since these are very powerful tools that are most often used in Sentiment Analysis. Then specific techniques and approaches for tackling each of the tasks described in the previous section will be addressed.

4.1 Classification

Many of the tasks in Sentiment Analysis can be thought of as classification. Machine Learning offers many algorithms designed to do just that, but this task of classifying text according to its sentiment presents many unique challenges. These can be formulated in one question: “*What kinds of features do we use?*”

4.1.1 Term Presence Vs Frequency

Traditional Information Retrieval systems have long emphasized the importance of term frequency. The famous TF-IDF (Term Frequency - Inverse Document Frequency) measure is well-used in modeling documents (Jones, 1972). The intuition is that terms that often appear in the document but seldom in the whole collection are more informative as to what the document is about as compared to the terms mentioned just once. In the field of Sentiment Analysis we find that instead of paying attention to most frequent terms, it is more beneficial to seek out the most unique ones. Pang et al. (Pang and Lee, 2002) improve the performance of their system using term presence instead of frequency. Document representation emphasizing term presence contain 1 if the term appears in the document at least once, 0 otherwise. Wiebe (Wiebe et al., 2004) writes “apparently people are creative when they are being opinionated”, implying that increased importance of low-frequency terms in opinionated texts.

4.1.2 n-grams

Term positions are also important in document representation for Sentiment Analysis. The position of terms determines, and sometimes reverses, the polarity of the phrase. So, position information is sometimes encoded into the feature vector (Pang and Lee, 2002; Kim and Hovy, 2006). Wiebe (Wiebe et al., 2004) selects n-grams ($n=1,2,3,4$) based on precision calculated using annotated documents. The n-grams are a word-stem, part-of-speech pair. For instance (*in-prep the-det can-noun*) is a 3-gram. The importance of part-of-speech tagging is discussed in the next section.

4.1.3 Part-of-Speech

As mentioned earlier, it has been determined that adjectives are good indicatives of sentiment in text (Hatzivassiloglou and Wiebe, 2000; Benamara et al., 2007), and in the past decade they have been commonly exploited in Sentiment Analysis (Mullen and Collier, 2004; Whitelaw et al., 2005). This is true for other fields in textual analysis, since part-of-speech tags can be considered to be a crude form of word sense disambiguation (Wilks and Stevenson, 1998). For example, Turney

(Turney, 2002) uses part-of-speech patterns, most including an adjective or an adverb, for sentiment detection at the document level.

4.1.4 Syntax

Syntax information has also been used in feature sets, though there is still discussion about the merit of this information in Sentiment classification (Pang and Lee, 2008). This information may include important text features such as negation, intensifiers, and diminishers (Kennedy and Inkpen, 2006). Kudo et. al. (Kudo and Matsumoto, 2004) used subtree-based boosting algorithm with dependency-tree-based features for polarity classification, and show that it outperforms the bag-of-words baseline.

4.1.5 Negations

Negations have been long known to be integral in Sentiment Analysis. The usual bag-of-words representation of text disconnects all of the words, and considers sentences like “I like this book” and “I don’t like this book” very similar, since only one word distinguishes one from the other. But when talking about sentiment, a negation flips the polarity of a whole phrase. Negations are often considered in post-processing of results, while the original representation of text ignores them (Hu and Liu, 2005). Or, as in Das and Chen (Das and Chen, 2001a), one could explicitly include the negation in the document representation by appending them to the terms that are close to negations; for example term “like-NOT” would be extracted from “I don’t like this book” (Pang and Lee, 2008). Though using co-location may be too crude a technique. It would be incorrect to negate the sentiment in a sentence such as “No wonder everyone loves it”. To handle such cases, Na et. al. (Na et al., 2004) use specific part-of-speech tags patterns to identify the negations relevant to the sentiment polarity of a phrase.

4.2 Identifying the semantic orientation of words

One of the most basic tasks in Sentiment Analysis is identifying the semantic orientation (the polarity and objectivity) of a word. A variety of techniques have been used, which can be roughly categorized in the following:

- using a *lexicon*, constructed manually or automatically
- using some *statistical techniques* such as looking at concurrence of a word with a word of a known polarity
- using *training documents*, labeled or unlabeled, as a source of knowledge about the polarity of terms within the collection

Each of these techniques has its advantages and difficulties, which will be discussed in detail in this section.

4.2.1 Lexicons

Extended lexicons are a fundamental part of Sentiment Analysis, but not all of them are alike. The simplest ones are ones with binary classification of words into *positive* vs. *negative* polarities or *objective* vs. *subjective*. A more fine distinction between the classes can be made with *fuzzy* lexicons where each label has a score associated with it, conveying the “strength” of the label. A yet more sophisticated approach is to adopt any of the finer-grained affective classifications developed in areas of psychology such as Plutchik’s emotion model (Prinz, 2004). Ortony, Clore, and Collins’s book *The Cognitive Structure of Emotions* (Ortony et al., 1988) provides an overview of several theories of “fundamental” or “basic” emotions (see Appendix A). Finally, even more sophisticated knowledge can be utilized, like commonsense knowledge bases that have been developed by researchers in Artificial Intelligence. This section describes recent work done on sentiment-annotated lexicons.

A variety of lexicons have been created for the use in Sentiment Analysis, often by extending existing general-purpose lexicons. For example, Subasic and Huettner (Subasic and Huettner, 2001) have manually constructed a lexicon associating words with affect categories, specifying an *intensity* (strength of affect level) and *centrality* (degree of relatedness to the category) (Dave et al., 2003). This lexicon can be called “fuzzy” since it is able to handle ambiguity of a term by assigning it to several semantic categories. The system is then designed to work with these “fuzzy” definitions: “After the affect words in a document are tagged, the fuzzy logic part of the system handles them by using fuzzy combination operators, set extension operators and a fuzzy thesaurus to analyze fuzzy sets representing affects” (Subasic and Huettner, 2001).

Besides manual annotation, other resources can be used to build lexicons. Existing lexicons can be augmented to include sentiment information. Princeton University’s WordNet lexicon has been one of the most popular ones to be used for Sentiment Analysis. As described on <http://wordnet.princeton.edu/>,

WordNet® is a large lexical database of English, developed under the direction of George A. Miller. Nouns, verbs, adjectives and adverbs are grouped into sets of cognitive synonyms (synsets), each expressing a distinct concept. Synsets are interlinked by means of conceptual-semantic and lexical relations.

Esuli and Sebastiani (Esuli and Sebastiani, 2006) expand WordNet by adding polarity (Positive-Negative) and objectivity (Subjective-Objective) labels for each

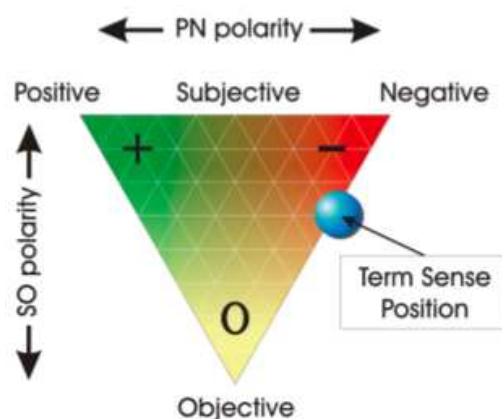


Figure 1: Graphical representation of opinion-related properties of a term

term. To label each term, they classify the synset (a group of synonyms) to which this term belongs using a set of ternary classifiers (a device that attaches to each object exactly one out of three labels), each of them capable of deciding whether a synset is *Positive*, or *Negative*, or *Objective*. The resulting scores range from 0.0 to 1.0, giving a graded evaluation of opinion-related properties of the terms. These can be summed up visually as in Figure A. The edges of the triangle represent one of the three classifications (*positive*, *negative*, and *objective*). A term can be located in this space as a point, representing the extent to which it belongs to each of the classifications.

Another extension to WordNet is WordNet-Affect, developed by Strapparava and Valitutti (Strapparava and Vlitutti, 2004). They label WordNet synsets using affective labels (*a-labels*) representing different affective category like *emotion*, *cognitive state*, *attitude*, *feeling*, etc.

WordNet has been also directly used in Sentiment Analysis. For example, Kim and Hovy (Kim and Hovy, 2004) and Hu and Liu (Hu and Liu, 2005) generate lexicons of positive and negative terms by starting with a small list of “seed” terms of known polarities (e.g. *love*, *like*, *nice*, etc.) and then using the antonymy and synonymy properties of terms to group them into either of the polarity categories.

Other resources have been used to generate lexicons. Extensive work has been done to create common sense knowledge bases in the field of Artificial Intelligence. These are collections of facts and information that an ordinary person is expected to know. Some of the most prominent projects are Cyc⁴, Open Mind Common Sense⁵, and ThoughtTreasure⁶ (Liu et al., 2003). Liu et. al. have used

⁴<http://www.cyc.com/>

⁵<http://openmind.media.mit.edu/>

⁶<http://alumni.media.mit.edu/~mueller/papers/tt.html>

the Open Mind Common Sense knowledge base, containing close to half a million sentences, to create several models mapping different concepts to six “basic” emotions - happiness, sadness, anger, fear, disgust, surprise - based on Ekman’s research on universal facial expressions (Ekman, 1993). Furthermore, Zhou and Chaovalit have developed an ontology-supported polarity mining (OSPM) approach to semantic labeling (Zhou and Chaovalit, 2008). They manually built an ontology for movie reviews and incorporated it into the polarity classification task, significantly improving performance over standard baseline.

There are also ways of determining sentiment orientation of words using statistical analysis of large corpora of text. Turney and Littman (Turney and Littman, 2003), for example, use word co-occurrence to infer semantic orientation of words, for which they explore two methods: Pointwise Mutual Information (PMI) (Church and Hanks, 1989) and Latent Semantic Analysis (LSA) (Landauer and Dumais, 1997). The idea is that “the semantic orientation of a word tends to correspond to the semantic orientation of its neighbors”. The techniques they utilize are actually quite different in nature: PMI calculates word co-occurrence by querying a search engine, and LSA uses a matrix factorization technique Singular Value Decomposition to analyze the statistical relationship between words.

As resources are seldom compared, it is still an open question as to which of these is the most beneficial for building annotated lexicons. The approaches described above vary greatly in the amount of data or human supervision needed. Thus, it is an important task to understand just how much of an improvement in performance we get by using a more sophisticated lexicon instead of a basic one.

4.2.2 Using Training Documents

It is possible to perform sentiment classification using statistical analysis and machine learning tools that take advantage of the vast resources of labeled (manually by annotators or using a star/point system) documents available. Product review websites like C-NET⁷, Ebay⁸, RottenTomatoes⁹ and the Internet Movie Database (IMDB)¹⁰ have all been extensively used as sources of annotated data. The star (or tomato, as it were) system provides an explicit label of the overall polarity of the review, and it is often taken as a gold standard in algorithm evaluation.

A variety of manually labeled data is available through evaluation efforts such as the Text REtrieval Conference (TREC)¹¹, NII Test Collection for IR Systems

⁷<http://www.cnet.com/>

⁸<http://www.ebay.com/>

⁹<http://www.rottentomatoes.com/>

¹⁰<http://www.imdb.com/>

¹¹<http://trec.nist.gov/>

(NTCIR)¹², and Cross Language Evaluation Forum (CLEF)¹³. The datasets these efforts produce often serve as standard in the Information Retrieval community, including for Sentiment Analysis researchers.

Individual researchers and research groups have also produced many interesting data sets. Here are some of these:

- *Congressional floor-debate transcripts*¹⁴ - published by Thomas et al. (Thomas and B. Pang, 2006) contains political speeches that are labeled to indicate whether the speaker supported or opposed the legislation discussed.
- *Economining*¹⁵ - published by Stern School at New York University, consisting of feedback postings for merchants at Amazon.com.
- *Cornell movie-review datasets*¹⁶ - introduced by Pang and Lee (Pang and Lee, 2008), containing 1000 positive and 1000 negative automatically derived document-level labels, and 5331 positive and 5331 negative sentences /snippets
- *MPQA Corpus*¹⁷ - Corpus and Opinion Recognition System corpus contains 535 manually annotated news articles from a variety of news sources containing labels for opinions and private states (beliefs, emotions, speculations, etc.)
- *Multiple-aspect restaurant reviews*¹⁸ - introduced by Snyder and Barzilay (Snyder and Barzilay, 2007), contains 4,488 reviews with an explicit 1-to-5 rating for five different aspects - food, ambiance, service, value, and overall experience

Once a desirable data set has been obtained, a variety of machine learning algorithms can be used to train sentiment classifiers. Some of the most popular algorithms are Support Vector Machines (Pang and Lee, 2002; Dave et al., 2003; Gamon, 2004; Matsumoto et al., 2005; Airoidi et al., 2006), Naive Bayes (Wiebe et al., 1999; Yu and Hatzivassiloglou, 2003; Melville et al., 2009), and maximum-entropy-based classifiers (Nigam et al., 1999; Pang and Lee, 2002). A comparison between these can be found in (Pang and Lee, 2002).

¹²<http://research.nii.ac.jp/ntcir/>

¹³<http://www.clef-campaign.org/>

¹⁴<http://www.cs.cornell.edu/home/llee/data/convote.html>

¹⁵<http://economining.stern.nyu.edu/datasets.html>

¹⁶<http://www.cs.cornell.edu/people/pabo/movie-review-data/>

¹⁷<http://www.cs.pitt.edu/mpqa/databaserelease/>

¹⁸<http://people.csail.mit.edu/bsnyder/naacl07>

4.3 Identifying semantic orientation of sentences and phrases

Once the semantic orientation of individual words has been determined, it is often desirable to extend this to the phrase or sentence the word appears in.

One of the most straightforward ways to accomplish this is to take an average of the polarities of words in the sentence. Hu and Liu (Hu and Liu, 2005) write: “if positive/negative opinion prevails, the opinion sentence is regarded as a positive/negative one”. In the case that the number of positive and negative opinion words is the same, they take the orientation of the closest opinion sentence.

Yu and Hatzivassiloglou (Yu and Hatzivassiloglou, 2003) train a Naive Bayes classifier using sentences and documents labeled as opinionated or factual as examples of the two categories. The features include words, bigrams, and trigrams, as well as the parts of speech in each sentence. They also use the presence of words with known polarities in a sentence as an indication that the sentence is subjective. And they take into consideration the effect of negation words such as “no”, “not”, and “yet” appearing in the window of 5 words around the word in question. Although simplistic, this heuristic has been shown to work for most of the cases.

An even more sophisticated combination of sentiment labels is possible by taking advantage of syntactic relationships between words. For example, Popescu and Etzioni (Popescu and Etzioni, 2005) use an unsupervised classification technique Relaxation Labeling that extends the label attributed to the word to the sentence it appears in. This approach takes, among other things, the *negation modifiers*, the significance of which is discussed in Section 4.1.5.

A novel application of machine translation techniques is described in (Kanayama et al., 2004). Here, using a Japanese to English translation engine, the researchers were able to build sentence trees and then apply pattern matching to discover sentiment orientation of sentences. Because of the sophistication of the method, they were able to incorporate many linguistic cues into the process, including negations, the use of which has been discussed earlier.

4.4 Identifying the semantic orientation of documents

Although most of the work is done in determining the semantic orientation of words and phrases, some tasks like summarization and text retrieval may require semantic labeling of the whole document. It may not make much sense to do this for long documents such as articles or books, which have been a key form in traditional Information Retrieval. But in the age of social networking and internet commerce, we see a vastly increasing number and variety of short documents, often containing only a few sentences. These may be product reviews, emails, blog posts, etc. Much like approaches for identifying semantic orientation of

words, those for documents also range from simple statistical ones to ones using elaborate knowledge structures to guide the process.

One of the most popular, and simple, methods is a linear combination of all polarities. For example, Dave et al. (Dave et al., 2003) and Turney et al. (Turney and Littman, 2003) use averaging to determine the polarity of documents. The polarity of the document can be expressed as

$$class(d_i) = \begin{cases} C, & eval(d_i) > 0 \\ C', & eval(d_i) < 0 \end{cases} \quad (1)$$

$$eval(d_i) = \sum_j score(t_j) \quad (2)$$

where class C is determined by the sum of scores of all terms: if the sum adds to a positive number, the document gets a positive label, otherwise a negative one. Notice that this approach is strictly binary - it does not take into consideration the fact that if a document may have strong opinions both ways, in all it should be considered to have a *mixed* opinion label, instead of relying on minute differences in measurement to assign either of the polarities.

As a part of the Text REtrieval Conference (TREC) task for Blog track in 2007, the teams had to classify documents as having either negative, positive, or mixed opinion (Macdonald et al., 2007). The team at University of Illinois at Chicago (Voorhees and Buckland, 2007) used a set of rules with thresholds to label the documents:

- Firstly, if both positive and negative opinions are strong in the document, the document should be mixed.
- Otherwise, if one type of the opinions is strong, the document is labeled to that type.
- Finally, if there are no strong opinions either way, the document is labeled as mixed.

Document labeling can incorporate information other than the semantic orientation of its constituent parts. Agrawal et al. (Agrawal et al., 2003) uses citations in newsgroup postings to divide the group into subgroups that are for or against an issue. An explicit assumption is made (and tested) that citations represent antagonistic standpoints, i.e. it is more likely that a reply would be in disagreement with the previous post than otherwise. Although probably an oversimplification of the actual workings of newsgroups, these are first steps in evaluating documents in their larger contexts.

4.5 Object feature extraction

Now we move on to another important part of sentiment - its target. In shorter, more focused documents it is often safe to assume that the author is only talking about the topic of the document. Product reviews, for example, usually contain opinions about that product, and movie reviews talk about the movies in question. Yet it is often not enough to know the general topic of the writing. A company making a product would certainly want to know not only what people think about this product in general, but which features they like/dislike in particular. Thus, the task of feature extraction (where feature can be any target of an opinionated statement) has been gaining popularity in the field of Sentiment Analysis.

A common approach is to use the part-of-speech (POS) tags to construct templates of how sentiment is applied to objects. For example, Bing Liu et al. (Liu et al., 2005) use this process for a phrase “included memory is stingy”:

1. Perform part-of-speech (POS) tagging and remove digits:

“<V>included <N>memory <V>is <Adj>stingy”

2. Replace the actual feature words in a sentence with [feature]:

“<V>included <N>[feature] <V>is <Adj>stingy”

3. Use n-gram to produce shorter segments from long ones:

“<Adj>included <N>[feature] <V>is”

“<N>[feature] <V>is <Adj>stingy”

4. Distinguish duplicate tags by giving them numbers:

“<Adj1>included <N1>[feature] <V1>is”

5. Perform word stemming

They then use association mining system CBA (Liu et al., 1998) to extract the rest of the features. Once the features are found, they are grouped using aforementioned WordNet synsets. For example, words “photo”, “picture”, and “image” all refer to the same feature in the digital camera. So, if they’re found to be synonymous, they become known synonyms of the same feature.

Popescu and Etzioni (Popescu and Etzioni, 2005) improve on Hu and Liu’s algorithm by removing those noun phrases that may not be product features. Their algorithm evaluates each noun phrase by computing a Pointwise Mutual Information (PMI) score between the noun phrase and meronymy (property of being a

part of something) discriminators associated with the product class. Here, PMI is defined as

$$MPI(f, d) = \frac{hits(f \wedge d)}{hits(f)hits(d)} \quad (3)$$

Given a set of relations of interest, their system calculates PMI between each feature and automatically generated discriminator phrases. For example, “scanner” class would be compared with phrases like “of scanner”, “scanner has”, “scanner comes with”, etc. which are used to find components or parts of scanners by searching the Web. The PMI scores are then converted to binary features for a Naive Bayes Classifier, which outputs a probability associated with each feature (Etzioni et al., 2005). In the end, a rich system of features is developed, a part of which is shown in Figure 1.

Table 1: Feature Information

Explicit Features	Examples
Properties	Scanner Size
Parts	Scanner Cover
Features of Parts	Battery Life
Related Concepts	Scanner Image
Related Concept’s Features	Scanner Image Size

But the approaches above discover only *explicit* features, ones that are mentioned in text. There are many implicit features in sentences like “this camera is too large”, which refers to the camera’s *size*. These can be extracted using the context of already known features. The rule mining technique described in (Liu et al., 2005) can be extended to implicit features by tagging each feature-specific template with its respective feature.

In the end, features can be used to effectively summarize sentiment found in text. In Figure 2, for example, the various features of two cameras are compared (Liu et al., 2005). Here, each bar indicates the range of opinions (ranging from negative to positive) on a camera’s feature. At a glance we can tell that there are more positive opinions on Camera 1’s picture quality than Camera 2’s.

4.6 Comparative Sentence Identification

One last major research area in Sentiment Analysis is the study of comparative sentences. In (Liu, 2006) Liu defines comparative sentence as “a sentence that

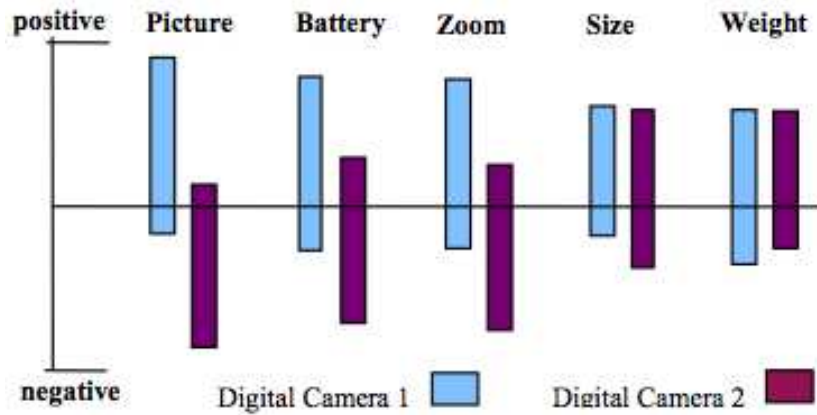


Figure 2: Visual comparison of consumer opinions on two products

expresses a relation based on similarities or differences of more than one object”. These can be classified into types, such as *gradable* and *non-gradable* comparisons. A gradable comparison is based on the relationship of greater, equal to, or less than. For example, “Intel chip is faster than the AMD one” ranks objects in quality. A non-gradable comparison the features are compared, but not ranked in the order of preference: “Coke tastes differently from Pepsi”. Both types of sentences tell us something about the relationships between different objects. Thus, one of the outputs of a comparative sentence analysis system could be a rank of products, as determined by the opinion holders. So far though, identification of comparative sentences has been the primary focus of the computational linguistics community.

Jindal and Liu (Jindal and Liu, 2006) take a data mining approach to solve this problem. They use class sequential rule (CSR) mining to identify comparative sentences in customer reviews, forum discussions, and news articles. They use a relatively small list of words (using WordNet¹⁹) as a first step in identifying the sentences, since it successfully identifies almost all of the comparative sentences (high recall: 94%), though also getting lots of false positives (low precision: 32%).

Hou and Li (Hou and Li, 2008) apply another data mining technique, Conditional Random Fields (CRF) to a manually annotated corpus of Chinese comparative sentences. They identify six semantic parts of comparative opinion: Holder, Entity 1, Comparative predicates, Entity 2, Attributes, and Sentiment, and extract them using Semantic Role Labeling, a statistical machine learning technique (Gildea and Jurafsky, 2002).

¹⁹<http://wordnet.princeton.edu/>

4.7 Performance Achieved

An overview of the work done in the most popular task of Sentiment Analysis, polarity classification, is shown in Table 4.7, which is extended from (Zhou and Chaovalit, 2008). This table is meant to offer a sample of work done and is not a comprehensive overview of the works published on the topic of Sentiment Analysis.

The work in this area started around 2000 and is still strong today. As mentioned earlier, a lot of work has been done on movie and product reviews, especially popular are the Internet Movie Database (IMDb) and product reviews downloaded from Amazon. The performance achieved by these method is difficult to judge, since each method uses a variety of resources for training and different collections of documents for testing. Many studies, such as Blitzer et al. (2007), deal with several domains, some more “challenging” for their algorithms than others. Notice especially how much the results vary across domains: in a recent study by Melville et al. (2009) the performance of their system on blogs and on political commentary differs by nearly 30%. Some studies, such as Godbole et al. (2007) work on the level of words, sometimes achieving accuracy of over 90%. Others, working on longer documents, such as blog posts and full web pages, have in general performance of around 65–85%.

It is clear that although we may be able to build comprehensive lexicons of sentiment-annotated words, it is still a challenge to accurately locate it in text. Few studies have been done outside the realm of short documents like product reviews, and especially in difficult domains like political commentaries. This is true partially because there is little annotated data available for realms outside reviews. Finally, although relatively high accuracy in document polarity labeling has been achieved, it is still a challenge to extract the full *private state*, complete with the emotion’s intensity, its holder, and its target.

5 General Questions

After exploring in detail the tasks of Sentiment Analysis, it is worth to first talk briefly about the general questions Sentiment Analysis brings up.

- First and foremost, the precise definition of *sentiment* still an open question. Can sentiment be ascribed to a word, a phrase, or can it be extended to the whole document? Must it have a target, and how granular is this target? By choosing a representation of sentiment, each researcher implicitly defines the scope and nature of the particular flavor of *sentiment* they are working with.

Studies	Polarity mining techniques used	Text granularity	Features	Data sources/Domains	Performance (accuracy)
Hatzivassiloglou and McKeown (1997)	log-linear regression model	document	conjunctions, part-of-speech	Wall Street Journal corpus	adjectives: precision: >90%
Das and Chen (2001b)	lexicons and grammar rules	document	words	financial news	62%
Pang and Lee (2002)	Nave Bayes, maximum entropy classification support vector machines	document	unigram, bi-gram, contextual effect of negation, feature presence or frequency, position	IMDb (Movie review)	82.9%
Turney (2002)	pointwise mutual information	document	bi-grams	Known positive terms such as excellent and negative terms such as poor movies, cars, banks	66–84%
Morinaga et al. (2002)	decision tree induction	document	characteristic words, co-occurrence words, and phrases	cellular phones, PDAs and Internet service providers	N/A
Dave et al. (2003)	support vector machines	document	semantic features based on substitutions and proximity	Amazon Cnn.Net	88.9%
Yi et al. (2003)	sentiment lexicon and semantic pattern	subject-spot terms	feature lexical semantics	digital cameras, music albums	85.6%
Turney and Littman (2003)	SO-LSA(Latent Semantic Analysis), SO-PMI (Pointwise Mutual Information) General Inquirer	document	words and phrases	TASA-ALLcorpus (from sources such as novels and newspaper articles)	65.27% (SO-LSA) 61.26% (SO-PMI)
Pang and Lee (2004)	Nave Bayes support vector machines	document	sentence-level subjectivity summarization based on minimum cuts	IMDb	86.4%
Hu and Liu (2005)	Opinion word extraction and aggregation enhanced with WordNet	product features	opinion words opinion sentences	Amazon Cnn.Net	cameras: 93.6% DVD plyr: 73% MP3 plyr: 84.2% cellphone: 76.4%
Nigam and Hurst (2004)	syntactic rules based chunking	sentence	a lexicon of polar phrases and their parts-of-speech, syntactic patterns	online resources (e.g., Usenet, online message boards) in a particular domain	general polarity analysis: precision: 77% (positive), 84% (negative); recall: 43% (positive), 16% (negative)
Hiroshi et al. (2004)	transfer-based machine translation, principal patterns auxiliary/nominal, patterns polarity lexicon	sentiment unit	full parsing semantic analysis	bulletin boards forums on digital cameras	precision: 89% recall 43%
Bai et al. (2005)	two-stage Markov Blanket Classifier	document	dependence among words, minimal vocabulary	IMDb, Infonic	movie: 87.5%, news: 89-96%
Gamon et al. (2005)	Nave Bayes classifier	sentence	stemmed terms, their frequency and weights, go list (salient words in a domain)	car reviews	recall: 96% (positive) 5- 24% (negative and other)
Popescu and Etzioni (2005)	relaxation labeling clustering	phrase	Syntactic dependency templates, conjunctions and disjunctions, WordNet	Amazon Cnn.Net	Opinion phrase polarity: precision: 86% recall: 97% relationships

Studies	Polarity mining techniques used	Text granularity	Features	Data sources/Domains	Performance (accuracy)
Wilson et al. (2005)	AdaBoost	phrase	subjectivity lexicon	multiperspective Question, Answering Opinion Corpus	contextual polarity: 65.7%
Kennedy and Inkpen (2006)	support vector machines, term-counting method, a combination of the two	document	term frequencies	General Inquirer dictionary, CTRWdictionary & Adj, IMDb (Movie review)	enhanced combined method: 86.2%
Chesley et al. (2006)	Support Vector Machines Wiktionary	document	textual features (e.g., exclamation points and question marks) and lexical semantics	Web sites of CNN, NPR, Atlanta Journal and Constitution, newspaper columns, reviews, political blogs, etc.	positive: 84.2% negative: 80.3% objective: 72.4%
Thomas and B. Pang (2006)	support vector machines	speech segment	reference classification	2005 U.S. floor debate in the House of Representatives	with same-speaker links and agreement links: 71.16%
Kaji and Kitsuregawa (2007)	phrase trees and word co-occurrence, Pointwise Mutual Information	phrase	lexical relationships, word co-occurrence	HTML documents	62.7–92.9%
Blitzer et al. (2007)	Structural Correspondence Learning	document	word frequencies and co-occurrences, part-of-speech	book, DVD, electronics and kitchen appliance product reviews	66.1–86.6%
Godbole et al. (2007)	lexical (WordNet)	word	graph distance measurements between words based on relationships of synonymity and anonymity, commonality of a words	newspapers, blog posts	82.7–95.7%
Annett and Kondrak (2008)	lexical (WordNet) & Support Vector Machines	document	number of positive/negative adjectives/adverbs, presence, absence or frequency of words, minimum distance from pivot words in WordNet	movie reviews, blog posts	65.4–77.5%
Zhou and Chaovalit (2008)	ontology-supported polarity mining	document	n-grams, words, word senses	movie reviews	72.2%
Hou and Li (2008)	Conditional Random Fields	sentence	POS tags, comparative sentence elements	product reviews, forum discussions; labeled manually and automatically	precision: man.: 89% aut.: 75% recall: man.: 81%, aut.: 71%
Ferguson et al. (2009)	Multinomial Naive Bayes (MNB)	phrase	binary word feature vectors	financial blog articles	75.25%
Tan et al. (2009)	Naive Bayes Classifier with feature adaptation using Frequently Co-occurring Entropy	document	words	Education reviews, stock reviews, and computer reviews	F1 score: 69–91%
Wilson et al. (2009)	boosting, memory-based learning, rule learning, and support vector learning	phrase	words, negation, polarity modification features	MPQA Corpus	83.6%
Melville et al. (2009)	Bayesian classification with lexicons and training documents	document	words	Blog posts reviewing software, political blogs, movie reviews	Blogs: 91.21% Political: 63.61% movies: 81.42%

- Once the notion of *sentiment* is settled, we need to find out how it is expressed in text. Is it just in the emotionally-charged words, or also in sentence structure? Can misspelling or punctuation tell us something about the sentimental nature of the passage? Does the document's sentiment spread to other related documents, say by links or co-authorship?
- Finally, a variety of cross-domain considerations need to be examined. What is the difference between expressing an opinion on zoom of a camera or an ambiance of a restaurant or fairness of a law? Are there cultural differences in the ways people express their opinions? Can people be grouped by the way they express their opinions on a subject? Is it possible to determine emotion in real time? It appears that each research project reviewed here presents its own flavor of emotional discourse.

6 Commercial Uses

Although the field of Sentiment Analysis is relatively young, there are already numerous businesses that use the techniques developed in this field to customers interested in brand tracking and market perception. For instance, as a part of its anti-counterfeiting and online brand abuse services, OpSec Security²⁰ provides sentiment analysis services such as “monitoring, measuring, and analyzing consumer feedback” so that their customers are “better informed to understand market needs, target customer segments, and position against competitors”. Specifically, these are the types of activities that may be involved:

- Tracking collective user opinions and ratings of products and services
- Analyzing consumer trends, competitors, and market buzz
- Measuring response to company-related events and incidents
- Monitoring critical issues to prevent negative viral effects
- Evaluating feedback in multiple languages

As a source of opinionated discourse, these companies look at

- Online communities
- Discussion boards

²⁰<http://www.opsecsecurity.com/>

- Weblogs
- Product rating sites
- Chatrooms
- Price comparison portals
- Newsgroups

By aggregating, evaluating, and interpreting the data found on these web sites, OpSec promises to “provide insights and recommendations” and “forecast product and brand trends”. Text analysis vendor Lexalytics²¹, on the other hand, worked with Cisco, where they “used a sentiment engine to determine which executives have the highest correlation to positively moving the stock price” (Grimes, 2008). The discovery of the “opinion leaders”, they claim, helps companies discover their strengths.

These services may also be helpful to a government intelligence agency. Monitoring communications for spikes in negative sentiment may be of use to agencies like Homeland Security. But besides companies and government agencies, general web users can benefit from sentiment-aware tools. There are several opinion-oriented search engines available online such as Opinmind²². By pre-labeling web pages and blogs, these services provide clustered view of the results, which enhances the users understanding of the results. The topics need not be restricted to product reviews - these can be political issues or opinions about candidates running for office (Pang and Lee, 2008).

Finally, opinion discovery can be a useful subcomponent of another service. Recommender systems can greatly benefit from extracting user ratings from text. Information retrieval systems can also use subjectivity measures when dealing with a certain type of information, such as when objectivity is desired in a scientific literature search.

7 Open Research Directions

A relatively new field, in its brief history Sentiment Analysis has used natural language processing, data mining, and text retrieval tools to tackle the problem of extracting opinions from text. The initial attempts at solving this problem borrowed techniques from the related areas of research: statistical methods have been used

²¹<http://www.lexalytics.com>

²²<http://www.opinmind.com>

to track opinionated words, machine learning algorithms have been applied to labeled text to produce polarity classifiers, etc. But the complex nature of the task requires even more sophisticated approaches (perhaps a combination of known ones). Two major problems remain:

- A lot of studies have been done on controlled collections of text like movie or product reviews, but algorithms that work for these collections fail miserably in a more **complex** setting. Extracting sentiment from more discursive texts such as political commentary (Bansal et al., 2008)(Thomas and B. Pang, 2006) or news articles (Koppel and Shtrimberg, 2004), where general topic is already known in advance, is still difficult.
- **Sentiment is topic-specific. The meaning of words changes, and they sometimes become reversed with context differences.** A phrase “go read the book” would be a positive statement in a book review, but if said in a review about a movie it may suggest that the book is better than the movie, and thus have an opposite effect (Pang and Lee, 2008). **General lexicons and algorithms must be adjusted and extended to accommodate each topic and its peculiarities.**

As mentioned earlier, many Sentiment Analysis approaches are lexicon-based, and many use the well-known “bag of words” text representation that disregards lexical relationships between words. Given the complexities of human language, these techniques can take us only so far. Gamon et al. (2008), for example, show several examples of political discussions involving news article links:

- *If you liked last term’s Supreme Court, you’re going to love the sequel -* Negative sentiment towards a state of affair expressed in ironic disguise
- *Leftard policy at its finest. \$100,000 a year and they’re in public housing? [news link] I am shocked the Washington Post would even report it. -* Negative sentiment expressed towards a state of affairs as reported in the news link, negative sentiment in ironic disguise towards the news provider
- *Taking a break from not getting anywhere on Iraq, Congress looks to not get anywhere on domestic issues for a little while. [news link] -* Negative sentiment towards a state of affair, news article cited in support

Because of the complex nature of sentiments, more sophisticated tools are needed to fully take advantage of semantic information in text. Some work has been done in adopting tools developed for other tasks for Sentiment Analysis.

A step above pure statistical analysis is lexical analysis, which includes part-of-speech tagging and phrase-structure trees. The structure of the text is often represented in the form of POS rules (Popescu and Etzioni, 2005) and tree templates (Kanayama et al., 2004). These are created manually, and may be supplemented by bootstrapping from text (Liu et al., 2005). Yet another step above deep lexical analysis, we can use a knowledge base to get closer to the meaning of the text. Zhou and Chaovalit (2008) and Liu et al. (2003) have used knowledge bases in constructing and supplementing lexicons in the task of sentiment polarity classification.

As it is time to turn to more sophisticated tools, it is also the time to turn to more interesting texts. Among untapped emotionally-charged discussions are political commentaries, inter-personal communication, and a wide variety of online discussion forums. Sentiment Analysis can be used to address larger questions on topics like

- *Hate Speech*. How are different kinds of hate speech expressed? How does the “hate” lexicon gets developed? Which documents use this kind of language?
- *Online Bullying*. When does correspondence becomes personal? Which blog posts, emails, or articles use combative or threatening language?
- *Opinion Tracking*. How are opinions spread in discussions? How is a sentiment adopted and changed from author to author? Is there an “opinion drift” (like a “topic drift”)?

8 Conclusions

This paper describes the field of Sentiment Analysis and its latest developments. Bringing together researchers from computer science, data mining, text retrieval, and computational linguistics, this field provides ample opportunities for both quantitative and qualitative work. Tackling the blurry definition of sentiment and the complexity of its manifestation in text, it opens doors for novel uses of techniques already developed for data mining and text analysis and brings up new questions, prompting development of yet better tools.

Internet provides us with an unlimited source of the most diverse and opinionated text, and as of yet only a small part of the existing domains have been explored. Much work has been done on product reviews – short documents that have a well-defined topic. More general writing, such as blog posts and web pages, have recently been receiving more attention. Still, the field is struggling

with more complex texts like sophisticated political discussions and formal writings. Future work in expanding existing techniques to handle more linguistic and semantic patterns will surely be an attractive opportunity for researchers and business people alike.

A A selection of lists of “fundamental” or “basic” emotions

Theorist	Fundamental emotions	Basis for selection	Reference
Arnold, M. B.	anger aversion courage dejection desire despair fear hate hope love sadness	relation to action tendencies	Arnold (196)
Ekman, P.	anger disgust fear joy sadness surprise	universal facial expressions	Ekman, Friesen & Ellsworth (1982)
Frijda, N.	desire joy pride surprise distress anger aversion contempt fear shame	forms of action readiness	Frijda (1987, and personal communication)
Gray, J.	rage/terror anxiety joy	hardwired	Gray (1982)
Izard, C. E.	anger contempt disgust distress fear guilt interest joy shame surprise	hardwired	Izard (1972)
James, W. McDougall, W.	fear grief love rage anger disgust elation fear subjection tender-emotion wonder	bodily involvement relation to instincts	James (1884) McDougall (1926)
Mowrer, O. H.	pain pleasure	unlearned emotional states	Mowrer (1960)
Oatley, K., and Johnson-Laird, P. N.	anger disgust fear happiness sadness	do not require propositional content	Oatley & Johnson-Laird (1987)
Panksepp, J.	expectancy fear rage panic	hardwired	Panksepp (1982)
Plutchik, R.	acceptance anger anticipation disgust joy fear sadness surprise	relation to adaptive biological processes	Plutchik (1980)
Tomkins, S. S.	anger interest contempt disgust distress fear joy shame surprise	density of neural firing	Tomkins (1984)
Watson, J. B. Weiner, B.	fear love rage happiness sadness	hardwired attribution-independent	Watson (1930) Weiner & Graham (1984)

Note. Not all the theorists represented in this table are equally strong advocates of the idea of basic emotions. For some it is a crucial notion (e.g., Izard, 1977; Panksepp, 1982; Plutchik, 1980; Tomkins, 1984), while for others it is of peripheral interest only and their discussions of basic emotions are hedged (e.g., Mowrer, 1960; Weiner & Graham, 1984).

Source: *The Cognitive Structure of Emotions* by Ortony, Clore, and Collins. 1988.

References

- Agrawal, Rajagopalan, Srikant, and Xu (2003). Mining newsgroups using network arising from social behavior. *Twelfth international World Wide Web Conference*.
- Airoidi, E. M., Bai, X., and Padman, R. (2006). Markov blankets and meta-heuristic search: Sentiment extraction from unstructured text. *Lecture Notes in Computer Science*, 3932:167–187.
- Annett, M. and Kondrak, G. (2008). A comparison of sentiment analysis techniques: Polarizing movie blogs. *Advances in Artificial Intelligence*, 5032:25–35.
- Bai, X., Padman, R., and Airoidi, E. (2005). On learning parsimonious models for extracting consumer opinions. *Proceedings of the Hawaii International Conference on System Sciences*.
- Bansal, M., Cardi, C., and Lee, L. (2008). The power of negative thinking: Exploring label disagreement in the min-cut classification framework. *Proceedings of the International Conference in Computational Linguistics (COLING)*.
- Benamara, F., Cesarano, C., Picariello, A., Reforgiato, D., and Subrahmanian, V. (2007). Sentiment analysis: Adjectives and adverbs are better than adjectives alone. *In Proceedings of the International Conference in Weblogs and Social Media (ICWSM)*.
- Blitzer, J., Dredze, M., and Pereira, F. (2007). Biographies, bollywood, boom-boxes and blenders: Domain adaptation for sentiment classification. *Proceedings of the 45th Annual Meeting of the Association of Computational Linguistics*, pages 440–447.
- Chesley, P., Vincent, B., Xu, L., and Srihari, R. K. (2006). Using verbs and adjectives to automatically classify blog sentiment. *Proceedings of the AAAI Spring Symposium on Computational Approaches to Analyzing Weblogs*.
- Church, K. W. and Hanks, P. (1989). Word association norms, mutual information and lexicography. *Proceedings of the 27th Annual Conference of the Association of Computational Linguists*, pages 76–83.
- Das, S. and Chen, M. (2001a). Yahoo! for amazon: Extracting market sentiment from stock message boards. *Proceedings of the Asia Pacific Finance Association Annual Conference (APFA)*.
- Das, S. and Chen, M. (2001b). Yahoo! for amazon: Sentiment parsing from small talk on the web. *Proceedings of the 8th annual Conference of the Asia Pacific Finance Association (APFA)*.
- Dave, K., Lawrence, S., and Pennock, D. M. (2003). Mining the peanut gallery: Opinion extraction and semantic classification of product reviews. *Proceedings of the World Wide Web Conference*.

- Ekman, P. (1993). Facial expression of emotion. *American Psychologist*, (48):384–392.
- Esuli, A. and Sebastiani, F. (2006). Sentiwordnet: A publicly available lexical resource for opinion mining. *Proceedings of the 5th Conference on Language Resources and Evaluation (LREC)*.
- Etzioni, O., Cafarella, M., Downey, D., Kok, S., Popescu, A., Shaked, T., Soderland, S., Weld, D., and Yates, A. (2005). Unsupervised named-entity extraction from the web: An experimental study. *Artificial Intelligence*, 165(1):91–134.
- Ferguson, P., O’Hare, N., Davy, M., Bermingham, A., Tattersall, S., Sheridan, P., Gurrin, C., and Smeaton, A. F. (2009). Exploring the use of paragraph-level annotations for sentiment analysis in financial blogs. *1st Workshop on Opinion Mining and Sentiment Analysis (WOMSA)*.
- Gamon, M. (2004). Sentiment classification on customer feedback data: Noisy data, large feature vectors, and the role of linguistic analysis. *Proceedings of the International Conference on Computational Linguistics (COLING)*.
- Gamon, M., Aue, A., Corston-Oliver, S., and Ringger, E. (2005). Pulse: Mining customer opinions from free text. *Proceedings of the 6th International Symposium on Intelligent Data Analysis*.
- Gamon, M., Basu, S., Belenko, D., Fisher, D., Hurst, M., and Konig, A. C. (2008). Blews: Using blogs to provide context for news articles. *Proceedings of the International Conference in Weblogs and Social Media*.
- Gildea, D. and Jurafsky, D. (2002). Automatic labeling of semantic roles. *Computational Linguist*, 28(3):245–288.
- Godbole, N., Srinivasaiah, M., and Skiena, S. (2007). Large-scale sentiment analysis for news and blogs. *Proceedings of the International Conference in Weblogs and Social Media*.
- Grimes, S. (2008). Sentiment analysis: A focus on applications.
- Hatzivassiloglou, V. and McKeown, K. R. (1997). Predicting semantic orientation of adjectives. *Proceedings of the 8th Conference of the European Chapter of the Association for Computational Linguistics*.
- Hatzivassiloglou, V. and Wiebe, J. (2000). Effects of adjective orientation and gradability on sentence subjectivity. In *Proceedings of the International Conference on Computational Linguistics (COLING)*.
- Hiroshi, K., Tetsuya, N., and Hideo, W. (2004). Deeper sentiment analysis using machine translation technology. *Proceedings of the International Conference on Computational Linguistics (COLING)*.

- Hou, F. and Li, G.-H. (2008). Mining chinese comparative sentences by semantic role labeling. *Proceedings of the Seventh International Conference on Machine Learning and Cybernetics*.
- Hu, M. and Liu, B. (2005). Mining and summarizing customer reviews. *Proceedings of the conference on Human Language Technology and Empirical Methods in Natural Language Processing*.
- Jindal, L. and Liu, B. (2006). Identifying comparative sentences in text documents. *Proceedings of the 29th annual international ACM SIGIR conference on Research and Development in Information Retrieval*.
- Jones, K. S. (1972). A statistical interpretation of term specificity and its application in retrieval. *Journal of Documentation*, 28:11–21.
- Kaji, N. and Kitsuregawa, M. (2007). Building lexicon for sentiment analysis from massive collection of html documents building lexicon for sentiment analysis from massive collection of html documents. *Proceedings of the Conference on Empirical Methods in Natural Language Processing*.
- Kanayama, H., Nisukawa, T., and Watanabe, H. (2004). Deeper sentiment analysis using machine translation technology. *Proceedings of the International Conference on Computational Linguistics*.
- Kennedy, A. and Inkpen, D. (2006). Sentiment classification of movie reviews using contextual valence shifters. *Computational Intelligence*, 22:110–125.
- Kim, S.-M. and Hovy, E. (2004). Determining the sentiment of opinions. *Proceedings of the 20th International Conference on Computational Linguistics*.
- Kim, S.-M. and Hovy, E. (2006). Automatic identification of pro and con reasons in online reviews. In *Proceedings of the COLING/ACL 2006 Main Conference Poster Sessions*, pages 483–490. Association for Computational Linguistics.
- Kim, S.-M. and Hovy, E. (2007). Crystal: Analyzing prediction opinions on the web. *Proceedings of the Joint Conference on Empirical Methods in Natural Language Processing and Computational Natural Language Learning (EMNLP-CoNLL)*.
- Koppel, M. and Shtrimberg, I. (2004). Good news or bad news? let the market decide. *Proceedings of the AAAI Spring Symposium on Exploring Attitude and Affect in Text: Theories and Applications*, pages 86–88.
- Kudo, T. and Matsumoto, Y. (2004). A boosting algorithm for classification of semi-structured text. In *Proceedings of the Conference on Empirical Methods in Natural Language Processing (EMNLP)*.

- Landauer, T. K. and Dumais, S. T. (1997). A solution to plato's problem: The latent semantic analysis theory of the acquisition, induction, and representation of knowledge. *Psychology Review*.
- Liu, B. (2006). *Web Data Mining*, chapter Opinion Mining. Springer.
- Liu, B., Hsu, W., and Ma, Y. (1998). Integrating classification and association rule mining. *Proceedings of the Conference on Knowledge Discovery and Data Mining*.
- Liu, B., Hu, M., and Cheng, J. (2005). Opinion observer: analyzing and comparing opinions on the web. *Proceedings of the International Conference on World Wide Web*.
- Liu, H., Lieberman, H., and Selker, T. (2003). A model of textual affect sensing using real-world knowledge. *Proceedings of the Seventh International Conference on Intelligent User Interfaces*, pages 125–132.
- Macdonald, C., Ounis, I., and Soboroff, I. (2007). Overview of the trec-2007 blog track. *Proceedings of the Sixteenth Text REtrieval Conference (TREC 2007)*.
- Matsumoto, S., Takamura, H., and Okumara, M. (2005). Sentiment classification using word sub-sequences and dependency sub-trees. *Proceedings of PAKDD'05, the 9th Pacific-Asia Conference on Advances in Knowledge Discovery and Data Mining*.
- Melville, P., Gryc, W., and Lawrence, R. D. (2009). Sentiment analysis of blogs by combining lexical knowledge with text classification. *Proceedings of the Conference on Knowledge Discovery and Data Mining 2009*.
- Morinaga, S., Yamanishi, K., Tateishi, K., and Fukushima, T. (2002). Mining product reputations on the web. *Proceedings of the 8th ACM SIGKDD international Conference on Knowledge Discovery and Data Mining*.
- Mullen, T. and Collier, N. (2004). Sentiment analysis using support vector machines with diverse information sources. *Proceedings of the Conference on Empirical Methods in Natural Language Processing (EMNLP)*, pages 412–418.
- Na, J.-C., Sui, H., Khoo, C., Chan, S., and Zhou, Y. (2004). Effectiveness of simple linguistic processing in automatic sentiment classification of product reviews. *Conference of the International Society of Knowledge Organization (ISKO)*, pages 49–54.
- Nigam, K. and Hurst, M. (2004). Towards a robust metric of opinion. *The AAAI Spring Symposium on Exploring Attitude and Affect in Text*.
- Nigam, K., Lafferty, J., and McCallum, A. (1999). Using maximum entropy for text classification. *Proceedings of the IJCAI-99 Workshop on Machine Learning for Information Filtering*, pages 61–67.

- Ortony, A., Clore, G., and Collins, A. (1988). *The Cognitive Structure of Emotions*. Cambridge University Press.
- Pang, B. and Lee, L. (2002). Thumbs up?: sentiment classification using machine learning techniques. *Proceedings of the ACL-02 Conference on Empirical Methods in Natural Language Processing*, 10:79–86.
- Pang, B. and Lee, L. (2004). A sentimental education: Sentiment analysis using subjectivity summarization based on minimum cuts. *Proceedings of the annual meeting for the Association of Computational Linguists*.
- Pang, B. and Lee, L. (2008). Opinion mining and sentiment analysis. *Foundation and Trends in Information Retrieval*, 2(1-2):1–135.
- Popescu, A.-M. and Etzioni, O. (2005). Extracting product features and opinions from reviews. *Proceedings of the conference on Human Language Technology and Empirical Methods in Natural Language Processing*.
- Prinz, J. (2004). *Gut Reactions: A Perceptual Theory of Emotion*. Oxford University Press.
- R. Quirk, S. Greenbaum, G. L. and Svartvik, J. (1985). *A comprehensive grammar of the English language*. Longman.
- Snyder, B. and Barzilay, R. (2007). Multiple aspect ranking using the good grief algorithm. *Proceedings of the Joint Human Language Technology/North American Chapter of the ACL Conference (HLT-NAACL)*, pages 300–307.
- Strapparava, C. and Vlitutti, A. (2004). Wordnet-affect: and affective extension of wordnet. *In Proceedings of the 4th International Conference on Language Resources and Evaluation*.
- Subasic, P. and Huettner, A. (2001). Affect analysis of text using fuzzy semantic typing. *IEEE-FS*, (9):483–496.
- Tan, S., Cheng, Z., Wang, Y., and Xu, H. (2009). Adapting naive bayes to domain adaptation for sentiment analysis. *Advances in Information Retrieval*, 5478:337–349.
- Thomas, M. and Pang, L. L. (2006). Get out the vote: Determining support or opposition from congressional floor-debate transcripts. *Proceedings of the Conference on Empirical Methods in Natural Language Processing (EMNLP)*, pages 327–335.
- Turney, P. (2002). Thumbs up or thumbs down? semantic orientation applied to unsupervised classification of reviews. *Proceedings of the Association for Computational Linguistics (ACL)*, pages 417–424.

- Turney, P. D. and Littman, M. L. (2003). Measuring praise and criticism: Inference of semantic orientation from association. *ACM Transactions on Information Systems (TOIS)*, 21(4):315–346.
- Voorhees, E. M. and Buckland, L. P. (2007). Uic at trec 2007 blog track. *Proceedings of the Sixteenth Text REtrieval Conference (TREC 2007)*.
- Whitelaw, C., Garg, N., and Argamon, S. (2005). Using appraisal groups for sentiment analysis. *Proceedings of the ACM SIGIR Conference on Information and Knowledge Management (CIKM)*, pages 625–631.
- Wiebe, J., Bruce, R., and O’Hara, T. (1999). Development and use of a gold standard data set for subjectivity classifications. *Proceedings of the 37th Annual Meeting of the Association for Computational Linguistics (ACL-99)*, pages 246–253.
- Wiebe, J. M. (1994). Tracking point of view in narrative. *Computational Linguistics*, 20:233–287.
- Wiebe, J. M., Wilson, T., Bruce, R., Bell, M., and Martin, M. (2004). Learning subjective language. *Computational Linguistics*, 30:277–308.
- Wilks, Y. and Stevenson, M. (1998). The grammar of sense: Using part-of-speech tags as a first step in semantic disambiguation. *Journal of Natural Language Engineering*, 4:135–144.
- Wilson, T., Wiebe, J., and Hoffmann, P. (2005). Recognizing contextual polarity in phrase-level sentiment analysis. *Proceedings of HLT-EMNLP*.
- Wilson, T., Wiebe, J., and Hoffmann, P. (2009). Recognizing contextual polarity: an exploration of features for phrase-level sentiment analysis. *Computational Linguistics*, 35(5):399–433.
- Yi, J., Nasukawa, T., Bunescu, R., and Niblack, W. (2003). Sentiment analyzer: Extracting sentiments about a given topic using natural language processing techniques. *Proceedings of the 3rd IEEE International Conference on Data Mining*.
- Yu, H. and Hatzivassiloglou, V. (2003). Towards answering opinion questions: Separating facts from opinions and identifying the polarity of opinion sentences. *Proceedings of the Conference on Empirical Methods in Natural Language Processing (EMNLP)*.
- Zhou, L. and Chaovalit, P. (2008). Ontology-supported polarity mining. *Journal of the American Society for Information Science and Technology*, 69:98–110.