

39. Opinion mining and sentiment analysis

Eric Breck and Claire Cardie

Abstract

Opinions are ubiquitous in text, and readers of on-line text — from consumers to sports fans to news addicts to governments — can benefit from automatic methods that synthesise useful opinion-orientated information from the sea of data. In this chapter on **opinion mining** and **sentiment analysis**, we introduce an idealised, end-to-end **opinion analysis system** and describe its components, including constructing opinion lexica, performing sentiment analysis, and producing opinion summaries.

Keywords Facets, opinion analysis, opinion lexicon, opinion mining, opinion summarisation, opinion-orientated question answering, perspective, polarity, private state, semantic orientation, sentiment analysis, sentiment classification, subjectivity analysis.

39.1 Introduction

Human beings love to express their opinions. This car is better than that one. This political candidate is dishonest. That restaurant's food is delectable. And the World Wide Web has provided countless forums for collecting these opinions and presenting them to the world. But they also present a challenge — as with other textual information, there is now so much of it that humans have difficulty navigating the sea of data. To address this issue, the field of opinion mining and sentiment analysis has arisen to provide automatic and semi-automatic methods for taking expressions of opinion in text and providing useful analysis and summarisation for users.

The potential users for an opinion mining or sentiment analysis system are many. Vendors could benefit from rapid feedback from customers on their own and their competitors' products. Consumers could benefit from easy navigation through their peers' evaluation of products. Citizens could benefit from analysis of the opinions expressed by and about politicians, candidates, and their policies. Governments could benefit from the analysis of opinions expressed by hostile entities. In all of these cases, the issue is not simply collecting textual opinions, but analysing and presenting them in useful ways for the needs of the user in question.

In this article, we provide a survey of state-of-the-art methods in opinion mining and sentiment analysis in the context of an idealised end-to-end system that will be described below.

39.1.1 A bit of history

Beginning in the mid-to-late 1990s, work began to emerge in natural language processing that, rather than extracting factual information from text, considered opinionated information instead. Wiebe and Bruce (1995), for example, designed classifiers to track point of view. Hatzivassiloglou and McKeown (1997) developed a machine learning approach to predict the semantic orientation of adjectives (positive or negative) and Argamon et al. (1998), to distinguish among news collections based on style. Subasic and Huettnner (2001) performed an analysis of affect in text. Over the course of the first decade of the 21st century, work in this area defined a number of computational tasks related to the analysis of opinionated text, explored effective approaches to each, and settled on a more consistent terminology.

39.1.2 Terminology

As this field has evolved, the terminology used to describe it has changed. At present, three terms are worth clarifying – subjectivity analysis, sentiment analysis, and opinion mining:

Subjectivity analysis refers to the identification of text that divulges someone’s thoughts, emotions, beliefs and other “private state” information (Quirk et al. 1985) that is not objectively visible. For example, the sentence “international officers believe the EU will prevail” gives the reader insight into the internal mental state of the officers, i.e. one of their beliefs. Subjectivity

analysis can be applied at the sentence level — does the sentence contain subjective content; or at the expression level — which words or phrases express subjective content.

Sentiment analysis is one component of subjectivity analysis. Technically, it refers to the task of identifying the valence — positive or negative — of a snippet of text. The identification can be done at a wide variety of granularities, from a word type — either in or out of context — to a phrase, sentence, paragraph, or entire document. For example, one variation of this task is distinguishing positive words like “hopeful” or “excited” from negative words like “awful” or “insipid.” At the other end of the scale is classifying reviews, e.g. distinguishing a positive movie review from a negative one. At any granularity, the task can be a simple binary one (positive vs. negative) or an ordinal one (e.g. 1, 2, 3, 4, or 5 stars).

Opinion analysis is a term that is most often used as a shorthand for systems that are doing both subjectivity analysis in conjunction with sentiment analysis. For the sentence “international officers believe the EU will prevail,” an opinion analysis system might determine that the sentence is subjective (it divulges a belief) and has positive sentiment (the belief is a positive one with respect to the EU). Occasionally, the term “sentiment analysis” is used as a synonym for “opinion analysis.”

Opinion mining generally refers to the corpus-level task of canvassing all available sources of opinions about a topic of interest to produce a coherent

summary. For example, given all the reviews published about a digital camera, produce a summary for the vendor of customer satisfaction with the camera. Or, given news stories about a political candidate, describe how different constituencies feel about the candidate's views on various topics.

39.1.3 A unified system

State-of-the-art research in opinion mining and sentiment analysis typically targets individual subproblems rather than presenting a comprehensive user solution. To facilitate presentation, however, we discuss here an idealised, unified system, and then investigate how each component has been addressed in the literature.

Before interacting with a user, our unified system (Figure 39.1) begins by collecting a lexicon of words that express positive or negative opinions (labelled CONSTRUCT OPINION LEXICON in Figure 39.1). How this is done will be described in Section 39.2. The system then allows users to specify a general topic of interest, for example a political candidate or a consumer product. The system will then canvass all available resources to compile a set of documents containing opinions expressed about the topic (OPINION-ORIENTATED INFORMATION RETRIEVAL in Figure 39.1, discussed in Section 39.3). Next, the system will discover FACETS of the topic about which parties express separate opinions, such as the *acting* in a movie or the *safety features* of an automobile (Section 39.4). Focusing on the words within each document that indicate opinions on the topic of interest, the system determines the overall degree of positive or negative opinion in the document (DETERMINE SENTIMENT, Section 39.5). The unified system

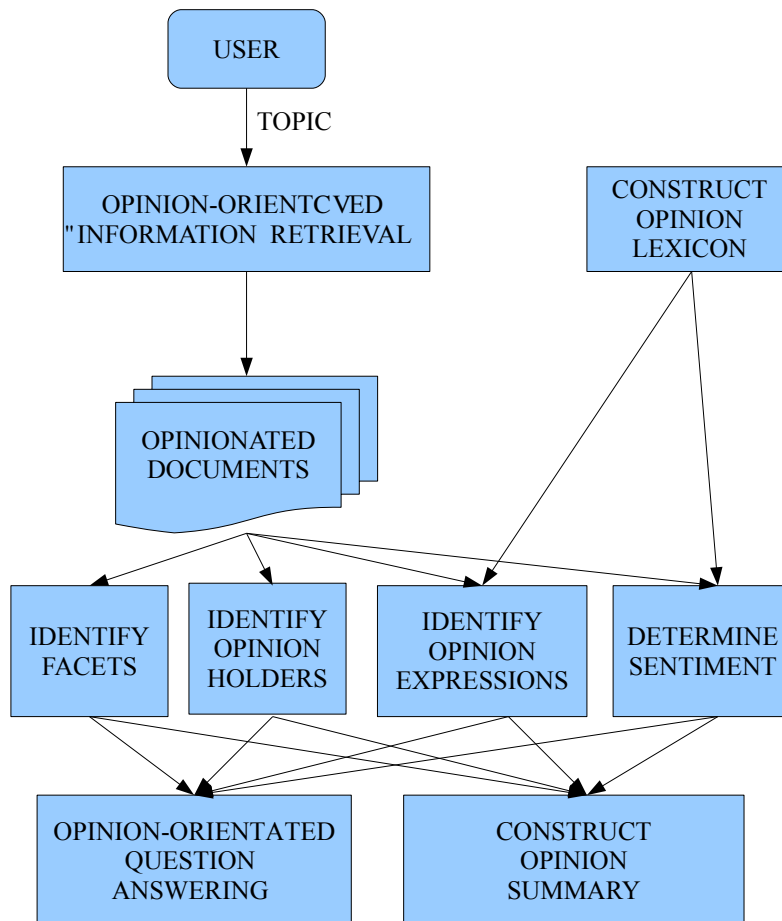


Figure 39.1: Architecture of a unified end-to-end system for opinion and sentiment analysis.

also identifies individual opinion expressions for tracking more fine-grained opinions with respect to the topics of interest (IDENTIFY OPINION EXPRESSIONS, Section 39.6): for each of the opinion expressions, the system attempts to determine the entity — consumers, professional reviewers, government leaders, political pundits, etc. — that is expressing the opinion (IDENTIFY OPINION HOLDERS, Section 39.7) as well as the specific topic of interest that is the target of the opinion. Finally, the system collects all of this information into a single database, and presents the user with an interface for viewing it. One interface might present a summary view, providing an overview of all the opinions about the topic or by particular parties, but also allowing the user to drill down to specific opinions (CONSTRUCT OPINION SUMMARY, Section 39.8). Another interface might allow queries to be executed over the database, to extract specific information about the canvassed opinions. For example, a user might want to know which facet of a hotel was mentioned most in negative reviews, or identify the publications expressing the strongest negative sentiment about a particular political regime (OPINION-ORIENTATED QUESTION ANSWERING, Section 39.9).

As indicated above, we will describe state-of-the-art approaches to tackling each of the parts of our unified system in the sections that follow. Before concluding, we also briefly mention a few aspects of research in opinion mining and sentiment analysis that our idealised system ignores — work on multilingual sentiment analysis (Section 39.10) and recent trends in explicitly compositional accounts of sentiment analysis (Section 39.11).

39.2 Building an opinion lexicon

Reading through text, some words quickly signal an author’s opinion, even without knowing yet exactly what the opinion is about. Describing something as “excellent” or “outstanding” is clearly positive, while “atrocious” and “horrific” are clearly negative. Researchers in this field have found that possessing an extensive **opinion lexicon** of such terms is invaluable in building automatic opinion mining and sentiment analysis systems. In this section we discuss what such lexicons look like, and how they can be acquired.

Even before knowing what sort of opinion a word denotes, it is useful to know that it suggests an opinion is being expressed at all. So, one type of useful lexicon is simply a list of those words that indicate *subjectivity*, i.e. that divulge someone’s thoughts, emotions, opinions. Many subjective words can further be categorised by their typical *sentiment orientation*, either positive or negative. Less commonly, other features of such words may also be listed in a lexicon, such as their intensity (weak, medium, strong).

A straightforward way to collect an opinion lexicon is to build it by hand, asking human annotators to list relevant words, and to mark them with the desired features (usually subjectivity or sentiment orientation). This has been done many times, both for general language (e.g. the General Inquirer lexicon (Stone 1968)) and for specific domains (e.g. Kanayama et al. (2004)). Such lists are highly valuable and generally high-precision with respect to subjectivity in the sense that when the lexicon indicates that a word is subjective, it is correct. There are

inevitably problems caused by context, however: the subjectivity and polarity of words can vary from their a priori meaning depending on the context in which they are used. In addition, subjectivity and sentiment lexicons typically exhibit lower coverage than lists produced by automatic methods.

Many automatic methods for compiling opinion lexicons have been proposed (e.g. Popescu and Etzioni (2005), Esuli and Sebastiani (2006), Kanayama and Nasukawa (2006), Mohammad et al. (2009), Feng et al. (2011), Feng et al. (2013)): some aim to expand an existing opinion lexicon and others aim to acquire the lexicon largely from scratch. We will focus here on the latter approaches, which generally begin with an initial set of “seed” words, chosen by hand to be canonical representatives of the desired categories. The methods then take large sets of unlabelled text, and essentially group together words based on some measure of contextual similarity to the seeds. The method we present here is based on Turney and Littman (2002).

We start with just two seed words, one for positive sentiment orientation (“excellent”) and one for negative (“poor”). The goal is then to find words that are in some sense similar to the seeds. Turney and Littman measure similarity via *Pointwise Mutual Information (PMI)* (Church and Hanks 1990).

$$PMI(word_1, word_2) = \log \frac{p(word_1 \& word_2)}{p(word_1)p(word_2)}$$

This statistic measures what we learn about one word when we see the other

nearby. A word's sentiment orientation will be scored as

$$PMI(word, \text{"excellent"}) - PMI(word, \text{"poor"})$$

— roughly, how much more the word is like “excellent” than it is like “poor.” Given a word, we can estimate its probability of occurrence or cooccurrence using queries to a web search engine. This gives us a way of identifying highly positive or highly negative words that does not depend at all on human labelling.

Evaluation. Evaluating manually and automatically constructed opinion lexicons is difficult as there is no gold standard to compare to. As a result, lexicons are typically evaluated in the context of a larger opinion-oriented task (e.g. sentiment categorisation of reviews) that employs the lexicon. For example, one might have a corpus in which sentences are annotated as subjective vs. objective. Then a subjectivity or sentiment lexicon could be used in a rule-based fashion to predict sentence-level subjectivity: if the sentence contains one or more words that are subjective/polar, based on the lexicon, then the sentence is deemed subjective; otherwise, it is labelled as objective. Performance of the lexicons is judged with respect to the resulting accuracy of the rule-based classifier on the gold standard sentence labels.

This type of evaluation is referred to as an “extrinsic evaluation.” The hope is that a system that employs the lexicon performs better on the task than a system that does not employ the lexicon as well as better than the same system that uses a different opinion lexicon.

39.3 Opinion-orientated information retrieval

For many opinion-mining and sentiment analysis tasks, we have a specific topic in mind at the start. e.g. we might be interested in what people are thinking with respect to a particular movie, sports figure, current event, or political issue. Unless we're lucky enough to be handed a set of documents on the topic, our unified opinion analysis system will need to start with a standard *information retrieval* step (for more on information retrieval, see Chapter 34 of this volume): given a natural language query (that describes the user's topic or domains of interest) and a document collection (possibly the Web), the system must return to the user a (usually ranked) set of those documents that are relevant to the query (i.e. on-topic).

And although there has been extensive research since the 1960's to develop effective information retrieval techniques (e.g. see the yearly SIGIR proceedings of Bruza et al. (2014) and Kelly et al. (2013)), topic-based *opinion* retrieval systems (Macdonald et al. 2008; Ounis et al. 2009) require more — they aim to locate documents that express an opinion or sentiment on a topic of interest, even if the overall focus of the document is not the target topic. In cases where the documents are likely to discuss multiple topics, this topic-only retrieval step should ultimately identify only those snippets or portions of the document that are on-topic. For this, standard *passage retrieval* algorithms can be employed (e.g. Salton et al. (1993), Kaszkiel and Zobel (1997)).

Thus, after an initial topic-only document or passage retrieval step, opinion

retrieval systems employ a second, re-ranking or filtering stage to locate the actual opinions. We discuss two common approaches next.

39.3.1 Dictionary-based approaches

An opinion dictionary, or lexicon, of the sort described in Section 39.2 is used to rank documents and passages based on their relative frequency of opinion lexicon terms and the distance of those terms to occurrences of topic-related words (e.g. Zhou et al. (2008)).

If training data for the opinion retrieval task is available, a different dictionary-based approach can be employed. Using the training data, first induce an opinion lexicon with terms weighted according to their ability to discriminate opinionated vs. non-opinionated documents. Once acquired, such a lexicon can then be used as a separate retrieval query (i.e. the query simply contains all of the opinion terms) to assign an opinion score to each document or passage (e.g. Hannah et al. (2008)).

39.3.2 Text classification approaches

In these approaches, training data consisting of subjective content (e.g. reviews) vs. factual content (e.g. encyclopaedias) is used to train classifiers that can estimate the degree of opinionated content in retrieved documents (e.g. Jia et al. (2009)). The original set of topic-based documents or passages is then re-ranked according to their subjective/objective classification scores — those scoring high-

est with respect to subjectivity at the top and those scoring highest with respect to objectivity at the bottom.

Finally, many opinion retrieval systems also determine the sentiment, i.e. polarity, of the identified opinion passages as one of positive, negative, or mixed (Macdonald et al. 2008; Ounis et al. 2009). Happily, the same dictionary- and classification-based techniques described above can be modified to determine the sentiment of arbitrary text snippets. Details on sentiment classification methods can be found in Section 39.5.

Evaluation. As in a number of information retrieval scenarios, the quality of opinion retrieval systems is typically judged according to two primary evaluation measures: precision@10 and mean average precision. *Precision@10 (P@10)* is the percentage of correctly identified passages with respect to the 10 top-ranked passages retrieved. The *mean average precision (MAP)* measure is somewhat more complicated. The *average precision* for an individual query is first calculated as the average of the precisions computed at the point of each relevant document in the ranked list of retrieved documents. The mean average precision for a set of queries is then just the mean of the average precision score across all queries. Additional information on information retrieval evaluation metrics can be found in Manning et al. (2008).

39.4 Facets

In commenting on a restaurant, movie, or digital camera, a useful review includes more than just a blanket thumbs-up/thumbs-down recommendation. The reader of a review wants to know about the food quality as well as the price of a restaurant, about the usability as well as image quality of a camera. Therefore, reviewers typically include individual opinions about these “facets” or “aspects” of a topic. Opinion analysis with respect to facets, also called *aspect-based opinion analysis*, is usually restricted to the context of reviews; computational techniques are therefore developed with this genre of text in mind. In this section, we will discuss facets, and describe how to determine an appropriate set of facets for a given topic.

Facets come in two general categories. First, there are physical parts or components of an object, about which a reviewer might comment separately. For example, one might find a car’s seats comfortable, but its steering wheel poorly placed. Second, there are attributes or features of the object and its parts. A chair might be highly comfortable but also very expensive. Here, we will consider these two kinds of facets together.

One way of identifying the appropriate set of facets for a given topic is to simply pre-specify them by hand. For hotels, one might decide, as Hotels.com does, that the relevant facets are service, condition, comfort, and cleanliness. This is feasible for tasks where one type of opinion is to be studied exhaustively. For a general system, though, we need a way of learning the appropriate set of facets

automatically. This problem has been well studied in the context of product and movie reviews (e.g. Hu and Liu (2004), Popescu and Etzioni (2005), Gamon et al. (2005), Carenini et al. (2006), Zhuang et al. (2006), Snyder and Barzilay (2007), Titov and McDonald (2008)); we sketch an early approach here (Hu and Liu 2004).

Facets are generally expressed via noun phrases — “The camera has *a powerful lens*, but produced *fuzzy landscape pictures*” — so we begin by applying a part-of-speech tagger and a noun phrase chunker to a large corpus of reviews of the desired type. We then extract all noun phrases that occur above a particular frequency (say 1% of all reviews).

This set then needs to be pruned to increase the precision of the result. There are a variety of methods that can be used. If a set of words typically used to expressed opinions is known, we can remove noun phrases not modified by one of these opinion expressions. We can also use external resources such as WordNet (Fellbaum 1998) or web statistics to determine whether the extracted set of noun phrases is actually associated with the topic.

For a survey of methods for opinion mining from product reviews, including facet identification, see Liu (2012). Once the facets are assembled, our system should determine the author’s opinion relative to each one. This can be done via a variety of methods, presented below in Section 39.5.

39.5 Determining the sentiment of a passage

The next step for our opinion analysis system is to determine the sentiment of the opinion passage under consideration. A natural entity to consider here is a review. Professional reviewers write reviews of everything from experiences (like concerts or movies) to products (like cars or stereos). Increasingly, consumers are writing reviews too, giving an explosion of textual data. Sometimes, these reviews come with a “star rating” or thumbs-up/thumbs-down flag, indicating the general opinion of the entire passage. But these are not always provided, and so in this section we look at means for automatically classifying a passage of text, like a review, as to whether it is generally positive or negative.

One approach is to exploit an existing opinion lexicon, as described in Section 39.3. Taking a passage, we can compute a summary statistic of the sentiment categories of all the words in the passage. For example, we could count positively orientated words and negatively orientated words and determine which occur more frequently. We could also compute the average sentiment of all words in the passage. This approach provides a natural extension of sentiment classification from the word level to the passage level.

Another approach is to adopt a supervised learning method (see Chapter 13 on Machine Learning). Since many reviews are labelled by their authors with a category (e.g. thumbs-up or thumbs-down), we have a natural source of training data for a machine learning algorithm. Many such algorithms have been proposed, and here we present an approach based on Pang et al. (2002).

When using a machine learning algorithm, a first step is choosing what features will be used to represent the instance (here, a passage or document) to the learning algorithm. Most successful approaches begin with a simple binary bag-of-words feature set — that is, a passage is represented by a vector of features f_i , where each f_i is 1 if the i th word in the vocabulary is present in the passage, and 0 otherwise. Many other more complex feature representations are possible (e.g. bigrams, parts-of-speech, frequency-based feature values), but their utility in this task is questionable (Pang et al. 2002). The next step is to choose a learning algorithm; and many standard algorithms are available in off-the-shelf packages. Commonly adopted algorithms include *support vector machines* (Joachims 2002), *naive Bayes* (Mitchell 1997), and *maximum entropy-based classification* (Ratnaparkhi 1996).

Predicting a star rating for a passage — e.g. 1, 2, 3, or 4 stars — requires substituting the classification-based learning algorithm with one that can predict numeric values (e.g. support vector regression (Zhu et al. 2009)) or ordinal values (e.g. ordered logistic regression). This allows us to produce a sentiment classification system by training on a large corpus of reviews with ratings provided by the author.

Evaluation. Sentiment categorisation systems are evaluated using the same measures as standard text categorisation algorithms — via *accuracy* and category-specific *precision* and *recall*.

39.6 Identifying opinion expressions

We want our opinion analysis system to go deeper than just classifying passages as to their sentiment orientation. We want to be able to extract information about individual opinions. The first step towards doing this is to identify the words and phrases that indicate that an opinion is being expressed.

One approach is simply to once again take an opinion lexicon and simply predict that if, say, the word “awesome” appears in the lexicon, then any appearance of the word “awesome” in a passage of text indicates that an opinion is being expressed there. This method has the advantage of simplicity, but it suffers from a number of drawbacks. First, many potentially opinionated words are ambiguous — a *small* hotel room is bad, a *small* carbon footprint is good — and we need context to determine whether or not the words actually express an opinion in a particular instance. Second, humans are endlessly creative in their expressions of opinion, and a fixed list can never hope to capture all the potential phrases used to express opinions. It should not be surprising then, that state-of-the-art systems again adopt supervised learning approaches to recognise expressions of opinion. The method we present here is based on Breck et al. (2007) but is typical of most opinion extraction systems.

One issue for supervised approaches to opinion expression identification is that they require training data; and unfortunately, such data is not as easy to come by for this task as it is for, say, sentiment categorisation of reviews. Fortunately, some data does exist in which individual expressions of opinion have been anno-

tated (e.g. Wiebe et al. (2005)), allowing a learning approach to proceed.

The choice of learning model is also more complex than in the sentiment categorisation task, as we want to take into account the fact that expressions of opinion often consist of multiple words. Our unified system therefore might use conditional random fields (CRFs) (Lafferty et al. 2001), a standard sequence tagging model (see Chapter 12) employed successfully for identifying part-of-speech, named entities, and other sequential categories. This method requires that a set of features be defined around individual words as well as for cues that link the predicted categories with adjacent words.

Breck et al. (2007) adopts a representation with standard features for context (a window of words around the target word), and syntactic structure (part-of-speech and the previous and subsequent syntactic constituent). To help generalise from the expressions encountered in the training data, the approach also includes features based on the hypernyms of the target word as identified via WordNet (Fellbaum 1998). The resulting system is able to identify the words and phrases expressing opinions in text.

Perhaps surprisingly, better performance can generally be obtained by employing learning methods that aim to *jointly* identify other attributes of the opinion — the opinion holder, the polarity, the target — at the same time as identifying the opinion expression itself. For examples, see Choi et al. (2006), Choi and Cardie (2010), Johansson and Moschitti (2011, 2013), and Yang and Cardie (2013, 2014).

Evaluation. The extent of an opinion expression is often ambiguous. In the sentence “I pretty much enjoyed the whole movie,” should the system identify “en-

joyed” or “pretty much enjoyed” as denoting the opinion? For problems where the exact span of text to be included in the gold standard annotations will likely vary from one human annotator to the next, systems tend to be evaluated with respect to how well their predictions overlap those in the gold standard, using both a strict (i.e. exact) and a lenient (i.e. partial or head word) matching scheme.

39.7 Identifying the opinion holder

For some opinion analysis tasks, the identity of the person or entity expressing the opinion is not so important. This is the case for most product reviews — we are interested in the sentiment of the review, regardless of the reviewer. Other times, knowing the person or organisation or report that has offered the opinion is critical — we would likely have more trust in an opinion about U.S. Secretary of State Hillary Clinton if it came from U.S. president, Barack Obama, than if it emanated from Hollywood bad boy, Charlie Sheen. This section describes methods for the automatic identification of the *opinion holder*, the entity that expresses the opinion. We prefer the term “opinion holder” to “opinion source” because “source” is also used to refer to the news source in which an opinion appears.

Consider the following sentences:

S1: Taiwan-born voters criticised China’s trade policy.

S2: International officers believe that the EU will prevail.

S3: International officers said US officials want the EU to prevail.

In S1, the phrase “Taiwan-born voters” describes the direct (i.e. first-hand) opinion holder of the critical sentiment. Similarly, in S2, we recognise the “international officers” as the group that has directly expressed an opinion regarding the EU. The same phrase in S3, however, denotes an indirect (i.e. second-hand, third-hand, etc.) opinion holder; the first-hand source is “US officials”. Most research in opinion analysis focuses on first-hand opinion holders (e.g. Bethard et al. (2004), Choi et al. (2005), Kim and Hovy (2006), Johansson and Moschitti (2010), Wiegand and Klakow (2010)) largely ignoring cases where opinions are expressed second- or third-hand (Breck and Cardie 2004, Wiebe et al. 2005).

State-of-the-art methods for identifying opinion holders mirror those for identifying opinion expressions: supervised learning methods are used to train classifiers or sequence taggers (see Chapter 13) for the task using a training corpus that is annotated for the task. (See Section 39.6 for details.) Our unified system, for example, might employ a sequence tagging algorithm to identify opinion holder spans. The feature set employed could be largely the same as well, but focus on representing cues associated with opinion holder entities — noun phrases located in the vicinity of an opinion expression that are of a semantic class that can bear sentiment (e.g. a person or an organisation). Wiegand and Klakow (2010) describe features commonly employed for opinion holder identification — at the word level, semantic class level, constituent level, grammatical relation level, and predicate argument level — and also discuss a method for generating them automatically.

Evaluation. The evaluation measures employed are the same as those for opinion

expressions (see Section 39.6).

39.8 Presenting a summary opinion

As discussed in the sections above, research in NLP has addressed issues in the identification and characterisation of opinions and sentiment in text — at the document, passage, sentence, and phrase levels. This section discusses the task of presenting the extracted opinion information to the end user.

For document- and passage-level sentiment analysis, it is generally enough to present to the user the thumbs-up/thumbs-down (positive/negative) classification or star rating predicted for the text. Sometimes, however, users want an explanation for the sentiment decision. This can be something as simple as showing the most important features from the machine learning system's point of view, or highlighting the opinion lexicon words in the text. Some document- and passage-level sentiment classification systems, however, generate useful explanatory material as a side-effect of the learning process. Pang and Lee (2004), for example, present a document-level sentiment analysis approach that identifies the key sentences that support the system's positive or negative prediction. These subjective sentences might also be returned to the user as an opinion-orientated summary of the document.

For fine-grained opinion analysis systems, the situation is somewhat different. Within any single opinionated text snippet, these systems are likely to identify a multitude of opinion expressions. Although this collection of opinions is useful

“ [Topic Delaying of Bulgaria’s accession to the EU] would be a *serious mistake*” [OH Bulgarian Prime Minister Sergey Stanishev] said in an interview for the German daily *Suddeutsche Zeitung*. “[Topic Our country] *serves as a model and encourages* countries from the region to follow despite the difficulties”, [OH he] added.

[Topic Bulgaria] is *criticized* by [OH the EU] because of slow reforms in the judiciary branch, the newspaper notes.

Stanishev was elected prime minister in 2005. Since then, [OH he] has been a *prominent supporter* of [Topic his country’s accession to the EU].

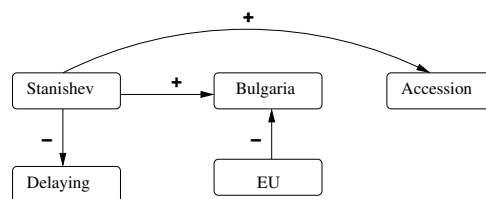


Figure 39.2: Example of text containing fine-grained opinions (above) and a summary of the opinions (below). In the text, opinion holders (OH) and topics (TOPIC) of opinions are marked and opinion expressions are shown in italics. In the summary graph, + stands for an overall positive opinion, and - for negative.

for a number of purposes (see Section 39.9), many users might prefer an overview of the opinion content in the paragraph or document. For these users, our unified system could create a summary of all of the opinions in a paragraph or document by grouping together all opinions from the same opinion holder and/or on the same topic and aggregating their polarities and intensities (Cardie et al. 2004). See, for example, Figure 39.2, which shows one possible graph-based summary of the opinions in the paragraph above it.

Generating this type of summary requires the ability to identify references to each opinion holder and each topic even though they are mentioned using different words. In Figure 39.2, for example, the phrases “Prime Minister Sergey Stanishev”, “he”, “Stanishev”, and “a prominent supporter” all refer to opinion holder Sergey Stanishev. For a survey of state-of-the-art methods for this task of *noun phrase coreference resolution* (see Chapters 6 and 27), see Ng (2010). For methods specifically designed for detecting expressions denoting the same opinion holder, see Stoyanov and Cardie (2006).

For the review genre, *multi-aspect sentiment summarisation* techniques are a focus of much current research (e.g. Zhuang et al. (2006), Blair-Goldensohn et al. (2008), Lerman et al. (2009)).

39.9 Opinion-orientated question answering

Given the opinions extracted using the techniques outlined in the sections above, one option is to summarise them (Section 39.8); another is to access the opin-

ions in direct response to a user's questions. Opinion-orientated questions appear to be harder than fact-based questions to answer. Their answers are often much longer, require combining partial answers from one or more documents, and benefit from finer-grained semantic distinctions among opinion types (Stoyanov and Cardie (2008); Somasundaran et al. (2007)). But research has addressed opinion-orientated question answering. The TAC QA track, for example, is a performance evaluation that focuses on finding answers to opinion questions (e.g. Dang (2008)). And our unified system might employ the methods from these evaluations to provide a question-answering interface for users: first, use the opinion questions to retrieve passages or sentences that are both topic-relevant and contain subjective material; then choose the answer candidate with the highest topic+opinion score (see Section 39.3). More recent approaches begin to consider the relationships between different answer candidates, incorporating opinion and sentiment information into PageRank- and HITS-style graph models (e.g. Li et al. (2009)). And Wang et al. (2014) explicitly treat opinion-orientated question answering as a summarisation task, proposing a submodular function-based framework to ensure topic coverage and diverse viewpoints in the system-generated answer.

Alternatively, when fine-grained opinions are identified, the unified system might store them in a database as 5-tuples (opinion expression, opinion holder, topic, polarity, intensity). End users could then access the extracted opinion content via simple database queries.

The next two sections cover two important and emerging areas of research in sentiment analysis and opinion mining: systems for languages other than English

and systems that treat sentiment analysis explicitly as a task in compositional semantics.

39.10 Multilingual Sentiment Analysis

We have focused, thus far, entirely on research in sentiment analysis and opinion mining involving English text. However, there is a growing body of work on multilingual sentiment analysis.

Most approaches focus on methods to adapt sentiment resources (e.g. lexicons) from resource-rich languages (typically English) to other languages with few sentiment resources. Mihalcea et al. (2007), for example, produced a subjectivity lexicon for Romanian by translating an existing English subjectivity lexicon. They then used the lexicon to build a rule-based sentence-level subjectivity classifier (as in Riloff and Wiebe (2003)) that can determine if a sentence in Romanian is subjective or objective.

The bulk of research for multilingual sentiment and subjectivity analysis, however, has focused on building resources that support supervised learning techniques in the desired target language — techniques that require training data annotated with the appropriate sentiment labels (e.g. document-level or sentence-level positive vs. negative polarity). This data is difficult and costly to obtain, and must be acquired separately for each language under consideration. Mihalcea et al. (2007), for example, also investigated the creation of a (sentence-level) subjectivity-annotated Romanian corpus by manually translating one from En-

glish and (automatically) projecting the subjectivity class labels for each English sentence to its Romanian counterpart. With this corpus in hand, they then used a standard supervised learning approach (as in Section 39.5) to obtain a classifier directly from the Romanian text. Their experiments found the parallel-corpus approach to work better than their lexicon translation method described above.

In earlier work, Kim and Hovy (2006) performed similar studies for German and English: they manually translated the target corpus (German or English) into the second language (English or German, respectively), and used an existing sentiment lexicon in the source language to determine sentiment polarity for the target corpus.

More recently, others have employed automatic machine translation engines to obtain the necessary subjectivity- or sentiment-labelled corpus. (For more on Machine Translation, see Chapter 36 of this volume.) Banea et al. (2008, 2010) did so for the task of sentence-level subjectivity classification. The Banea et al. (2010) study, for example, translated an English corpus into five different languages, mapping the sentence-level labels to the translated text. They found that the approach works consistently well regardless of the target language.

Approaches that do not explicitly involve resource adaptation include Wan (2009), which uses a weakly supervised learning technique called *co-training* (Blum and Mitchell 1998). Their co-training approach employs unlabelled Chinese data and a labelled English corpus, and independent “views” comprised of English vs. Chinese features to improve Chinese sentiment classification. Another notable approach is the work of Boyd-Graber and Resnik (2010), which presents

a generative model — supervised multilingual latent Dirichlet allocation — that jointly models topics that are consistent across languages, and employs them to better predict sentiment ratings.

In recent years, however, sentiment-labelled data is gradually becoming available for languages other than English. And there is still much room for improvement in existing monolingual (including English) sentiment classifiers, especially at the sentence level (Pang and Lee 2008). With this in mind, Lu et al. (2011) tackled the task of bilingual sentiment analysis: they assumed that some amount of sentiment-labelled data is available for each language in the pair under study, and aimed to simultaneously improve sentiment classification for both languages. Given the labelled data in each language, they developed an approach that exploits an unlabelled parallel corpus and the intuition that two sentences or documents that are parallel (i.e. translations of one another) should exhibit the same sentiment — their sentiment labels (e.g. polarity, subjectivity) should be similar. Their solution is a maximum entropy-based EM approach (see Chapter 12) that jointly learns two monolingual sentiment classifiers by treating the sentiment labels in the unlabelled parallel text as unobserved latent variables and maximising the regularised joint likelihood of the language-specific labelled data together with the inferred sentiment labels of the parallel text.

39.11 Compositional Approaches to Phrase-Level Sentiment Analysis

A key component of systems that perform fine-grained sentiment (see Section 39.6), is the ability to identify subjective expressions. To date, this task has for the most part been accomplished by sequence-tagging approaches that rely on sentiment lexicons as well as a number of syntactic and semantic features of the sentence. A recent trend in sentiment analysis harkens back to early work in computational linguistics on computational semantics (Montague 1974).

The semantic compositionality principle (see Chapter 5) states that the meaning of a phrase is composed from the meaning of its words and the rules that combine them. In the context of phrase-level sentiment analysis, a key effect is a change in polarity (e.g. flip, increase, decrease) when combining one word with other words in the phrase. Consider the following examples:

- prevent war
- limiting freedom
- absolutely delicious

In all of these phrases we observe changes in sentiment with respect to the underlined word when the preceding word is considered. In the first example, “war” has a negative sentiment; however, the word “prevent” essentially flips the polarity of the phrase to positive (i.e. preventing war is good). In the second, “freedom” has positive sentiment; however, “limiting freedom” makes the resulting sentiment of

the phrase negative. And in the final third example, the presence of the adverb “absolutely” strengthens the already positive sentiment of “delicious.” Clearly, the computation of phrase-level sentiment follows compositional rules of some sort.

According to the semantic compositionality principle in the context of sentiment analysis, the sentiment of a phrase depends on the sentiment of the words used in the phrase and the rules to combine them. The sentiment of individual words might be determined by a sentiment lexicon of the type discussed in Section 39.2. But what are these compositional rules? One might look at a number of sentiment-bearing phrases and provide a set of hand-written compositional rules for a sentiment analysis system (e.g. Moilanen and Pulman (2007), Choi and Cardie (2008)). Such rules are typically based on the output of a parser: the sentiment of a phrase or a sentence is computed from a parse tree in a bottom-to-top manner by starting from the sentiments of the individual lexical items and computing sentiment values in the intermediate nodes of the parse tree and, finally, at the root, according to hand-written compositional rules.

However, writing the rules by hand is tedious. For example, to obtain a set of rules such as “IF the syntactic pattern is VB NP and the verb is prevent and the noun phrase has a negative sentiment, THEN the resulting sentiment of a phrase is positive”, one has to consider various syntactic patterns and observe how the resulting sentiment changes when composed with specific lexical items.

While some learning-based methods based on compositional semantics have been proposed (e.g. Choi and Cardie (2008), Nakagawa et al. (2010)), recent

years have seen the emergence of distributional methods for phrase-level sentiment analysis. One option, for example, is to represent the meaning of each word as a matrix and then use general-purpose matrix multiplication or addition in lieu of composition rules (e.g. Baroni and Zamparelli (2010), Rudolph and Giesbrecht (2010), Yessenalina and Cardie (2011)). These models additionally allow the sentiment value for a phrase to be an ordinal rather than a binary value. The basic idea (from Yessenalina and Cardie (2011)) is as follows.

Consider combining an adverb like “very” with a polar adjective like “good”. “Good” has an a priori positive sentiment, so “very good” should be considered more positive even though “very”, on its own, does not bear sentiment. Combining “very” with a negative adjective, like “bad”, results in a phrase (“very bad”) that should be characterised as more negative than the original adjective. Thus, it is convenient to think of the effect of combining an intensifying adverb with a polar adjective as being multiplicative in nature, if we assume the adjectives (“good” and “bad”) to have positive and a negative sentiment scores, respectively.

We can also consider adverbial negators, e.g. “not”, combined with polar adjectives. When modeling only binary (positive and negative) labels for sentiment, negators are generally treated as flipping the polarity of the adjective it modifies. However, distributional approaches using an ordinal sentiment scale model negators as dampening the adjectives polarity rather than flipping it. For example, if “perfect” has a strong positive sentiment, then the phrase “not perfect” is still positive, though to a lesser degree. And while “not terrible” is still negative, it is less negative than “terrible”. For these cases, it is convenient to view “not” as shifting

polarity to the opposite side of polarity scale by some value, which is essentially an additive effect.

In addition to the above methods, an alternative framework for representing and applying compositionality has emerged in recent years in the form of new connectionist architectures (Bengio 2009) employed in conjunction with learned *word embeddings* that represent a single word as a dense, low-dimensional vector in a (distributed) meaning space (Collobert and Weston 2008; Mnih and Hinton 2007; Mikolov et al. 2013; Turian et al. 2010). Recursive neural networks, for example, operate on structured inputs and have been very successfully applied to the task of phrase- and sentence-level sentiment analysis (Socher et al. 2011; Socher et al. 2013). Given the structural representation of a sentence, e.g. a parse tree, they recursively generate parent representations in a bottom-up fashion, by combining tokens to produce representations for phrases, eventually producing the whole sentence. The sentence-level representation (or, alternatively, its phrases) can then be used to make a final classification for a given input sentence — e.g. whether it conveys a positive or a negative sentiment. Recently, “deep” (Bengio 2009; Hermans and Schrauwen 2013) versions of bidirectional recurrent nets (Schuster and Paliwal 1997) have been proposed for the same task and shown to outperform recursive nets while requiring no parse tree representation of the input sentence (Irsoy and Cardie 2014).

39.12 Conclusion

In this chapter, we have presented a unified model of research in opinion mining and sentiment analysis. We believe this captures the central ideas in the field, although it necessarily leaves some research out. We have assumed that the topics of opinions are provided by a user, but we could instead identify them automatically (e.g. Yi et al. (2003); Bethard et al. (2004); Kim and Hovy (2006); Stoyanov and Cardie (2008); Somasundaran and Wiebe (2009)). The distinction between positive and negative sentiment is usually clear, but determining neutral sentiment is difficult, and under-explored (Koppel and Schler 2006). And we have sometimes assumed that words have a fixed polarity, but of course many words require context to disambiguate their polarity (Wilson et al. 2005).

To recap, our model begins with the creation of an opinion lexicon. Next, the user identifies a set of documents containing opinions on a topic of interest. Opinions are then extracted from these documents, as we consider the overall sentiment of the document as well as the opinion holders and topics of each opinion expression. Finally, the resulting collection of opinions is presented to the user both as a queryable database and as a holistic summary.

Opinion mining and sentiment analysis is a relatively new area of natural language processing, but it is growing quickly. With applications to real-world business problems and fascinating research questions to explore, we expect it will continue to yield insights in the years to come.

Further reading and relevant resources

This chapter is necessarily brief; for a thorough survey of the field, see Pang and Lee (2008) or Liu (2012). There are frequent conferences and workshops in opinion mining and sentiment analysis or that often include research in this area. Some examples are the Text Analysis Conference (TAC) held by NIST, the International AAAI Conference on Weblogs and Social Media (ICWSM), and the Workshop on Computational Approaches to Subjectivity and Sentiment Analysis (WASSA). The associated conference proceedings are generally available on-line.

Finally, although sentiment analysis and opinion mining are among the most active research areas in natural language processing today, they are now also widely studied in other subareas of computer science — e.g., in data mining (see the proceedings of the ICDM and KDD conferences), Web science (see the proceedings of WWW and WSDM), and human-computer interaction (see the proceedings of CHI).

References

- Argamon, Shlomo, Moshe Koppel, and Galit Avneri (1998). ‘Routing documents according to style’. In *In First International Workshop on Innovative Information Systems*.
- Banea, Carmen, Rada Mihalcea, and Janyce Wiebe (2010). ‘Multilingual subjectivity: Are more languages better?’ In *Proceedings of the 23rd Interna-*

- tional Conference on Computational Linguistics (Coling 2010)*, 28–36. Beijing, China.
- Banea, Carmen, Rada Mihalcea, Janyce Wiebe, and Samer Hassan (2008). ‘Multilingual subjectivity analysis using machine translation’. In *Proceedings of the 2008 Conference on Empirical Methods in Natural Language Processing*, 127–135.
- Baroni, Marco and Roberto Zamparelli (2010). ‘Nouns are vectors, adjectives are matrices: Representing adjective-noun constructions in semantic space’. In *Proceedings of the 2010 Conference on Empirical Methods in Natural Language Processing*, 1183–1193.
- Bengio, Yoshua (2009). ‘Learning deep architectures for ai’. *Foundations and trends® in Machine Learning*, 2(1), 1–127.
- Bethard, Steven, Hong Yu, Ashley Thornton, Vasileios Hativassiloglou, and Dan Jurafsky (2004). ‘Automatic extraction of opinion propositions and their holders’. In *Proceedings of AAAI Spring Symposium on Exploring Attitude and Affect in Text*.
- Blair-Goldensohn, Sasha, Kerry Hannan, Ryan McDonald, Tyler Neylon, George A. Reis, and Jeff Reynar (2008). ‘Building a sentiment summarizer for local service reviews’. In *WWW Workshop on NLP in the Information Explosion Era (NLPIX08)*.

- Blum, Avrim and Tom Mitchell (1998). ‘Combining Labeled and Unlabeled Data with Co-Training’. In *Proceedings of the 11th Annual Conference on Computational Learning Theory (COLT-98)*.
- Boyd-Graber, Jordan and Philip Resnik (2010). ‘Holistic sentiment analysis across languages: Multilingual supervised latent dirichlet allocation’. In *Proceedings of the 2010 Conference on Empirical Methods in Natural Language Processing*, 45–55. Association for Computational Linguistics.
- Breck, Eric and Claire Cardie (2004). ‘Playing the telephone game: Determining the hierarchical structure of perspective and speech expressions’. In *The 20th International Conference on Computational Linguistics (COLING 2004)*, 120–126.
- Breck, Eric, Yejin Choi, and Claire Cardie (2007). ‘Identifying expressions of opinion in context’. In *Proceedings of the Twentieth International Joint Conference on Artificial Intelligence (IJCAI-2007)*. Hyderabad, India.
- Bruza, Peter, Charles L.A. Clarke, and Kal Jrvlin (2014). ‘Sigir ’14: Proceedings of the 37th international acm sigir conference on research & development in information retrieval’.
- Cardie, Claire, Janyce Wiebe, Theresa Wilson, and Diane Litman (2004). ‘Low-level annotations and summary representations of opinions for multi-perspective question answering’. In Maybury, Mark, editor, *New Directions in Question Answering*.

- Carenini, Giuseppe, Raymond Ng, and Adam Pauls (2006). ‘Multi-document summarization of evaluative text’. In *Proceedings of the European Chapter of the Association for Computational Linguistics (EACL)*, 305–312.
- Choi, Yejin, Eric Breck, and Claire Cardie (2006). ‘Joint extraction of entities and relations for opinion recognition’. In *Proceedings of the 2006 Conference on Empirical Methods in Natural Language Processing*, 431–439. Association for Computational Linguistics.
- Choi, Yejin and Claire Cardie (2008). ‘Learning with compositional semantics as structural inference for subsentential sentiment analysis’. In *Proceedings of the 2008 Conference on Empirical Methods in Natural Language Processing*, 793–801.
- Choi, Yejin and Claire Cardie (2010). ‘Hierarchical sequential learning for extracting opinions and their attributes’. In *Proceedings of the ACL 2010 Conference Short Papers*, 269–274. Association for Computational Linguistics.
- Choi, Yejin, Claire Cardie, Ellen Riloff, and Siddarth Patwardhan (2005). ‘Identifying sources of opinions with conditional random fields and extraction patterns’. In *Proceedings of the conference on Human Language Technology and Empirical Methods in Natural Language Processing*, 355–362. Association for Computational Linguistics.
- Church, Kenneth W. and Patrick Hanks (1990). ‘Word association norms, mutual

- information, and lexicography’. *Computational linguistics*, 16(1), 22–29. ISSN 0891-2017.
- Collobert, Ronan and Jason Weston (2008). ‘A unified architecture for natural language processing: Deep neural networks with multitask learning’. In *Proceedings of the 25th international conference on Machine learning*, 160–167. ACM.
- Dang, Hoa Trang (2008). ‘Overview of the tac 2008 opinion question answering and summarization tasks’.
- Esuli, Andrea and Fabrizio Sebastiani (2006). ‘Sentiwordnet: A publicly available lexical resource for opinion mining’. In *Proceedings of Language Resources and Evaluation (LREC)*.
- Fellbaum, Christiane (1998). *WordNet: An electronic lexical database*. The MIT press. ISBN 026206197X.
- Feng, Song, Ritwik Bose, and Yejin Choi (2011). ‘Learning general connotation of words using graph-based algorithms’. In *Proceedings of the 2011 Conference on Empirical Methods in Natural Language Processing*, 1092–1103. Association for Computational Linguistics, Edinburgh, Scotland, UK.
- Feng, Song, Jun Seok Kang, Polina Kuznetsova, and Yejin Choi (2013). ‘Connotation lexicon: A dash of sentiment beneath the surface meaning’. In *Proceedings of the 51st Annual Meeting of the Association for Computational Lin-*

- guistics (Volume 1: Long Papers)*, 1774–1784. Association for Computational Linguistics, Sofia, Bulgaria.
- Gamon, Michael, Anthony Aue, Simon Corston-Oliver, and Eric Ringger (2005). ‘Pulse: Mining customer opinions from free text’. *Advances in Intelligent Data Analysis VI*, 121–132.
- Hannah, David, Craig Macdonald, Jie Peng, Ben He, and Iadh Ounis (2008). ‘University of glasgow at trec 2007: Experiments in blog and enterprise tracks with terrier’. In *The Sixteenth Text REtrieval Conference (TREC 2007) Proceedings*. NIST.
- Hatzivassiloglou, Vasileios and Kathleen R. McKeown (1997). ‘Predicting the semantic orientation of adjectives’. In *ACL97*.
- Hermans, Michiel and Benjamin Schrauwen (2013). ‘Training and analysing deep recurrent neural networks’. In *Advances in Neural Information Processing Systems*, 190–198.
- Hu, Mingqing and Bing Liu (2004). ‘Mining and summarizing customer reviews’. In *Proceedings of the tenth ACM SIGKDD international conference on Knowledge discovery and data mining*, 168–177. ACM. ISBN 1581138881.
- Irsoy, Ozan and Claire Cardie (2014). ‘Opinion mining with deep recurrent neural networks’. In *Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, 720–728. Association for Computational Linguistics, Doha, Qatar.

- Jia, Lifeng, Clement Yu, and Wei Zhang (2009). ‘Uic at trec 2008 blog track’. In *The Seventeenth Text REtrieval Conference (TREC 2008) Proceedings*. NIST.
- Joachims, T. (2002). *Learning to Classify Text using Support Vector Machines*. Kluwer.
- Johansson, Richard and Alessandro Moschitti (2010). ‘Reranking models in fine-grained opinion analysis’. In *Proceedings of the 23rd International Conference on Computational Linguistics (Coling 2010)*, 519–527.
- Johansson, Richard and Alessandro Moschitti (2011). ‘Extracting opinion expressions and their polarities: exploration of pipelines and joint models’. In *Proceedings of the 49th Annual Meeting of the Association for Computational Linguistics: Human Language Technologies: short papers-Volume 2*, 101–106. Association for Computational Linguistics.
- Johansson, Richard and Alessandro Moschitti (2013). ‘Relational features in fine-grained opinion analysis’. *Computational Linguistics*, 39(3), 473–509.
- Kanayama, Hiroshi and Tetsuya Nasukawa (2006). ‘Fully automatic lexicon expansion for domain-oriented sentiment analysis’. In *Proceedings of the 2006 Conference on Empirical Methods in Natural Language Processing*, 355–363. Association for Computational Linguistics.
- Kanayama, Hiroshi, Tetsuya Nasukawa, and Hideo Watanabe (2004). ‘Deeper sentiment analysis using machine translation technology’. In *Proceedings of Coling 2004*, 494–500. COLING.

- Kaszkiel, Marcin and Justin Zobel (1997). ‘Passage retrieval revisited’. In *Proceedings of the 20th annual international ACM SIGIR conference on Research and development in information retrieval, SIGIR ’97*, 178–185. ACM, New York, NY, USA.
- Kelly, Diane, Maarten de Rijke, and Tetsuya Sakai (2013). ‘Sigir ’13: Proceedings of the 36th international acm sigir conference on research and development in information retrieval’. ACM, New York, NY, USA. ISBN 978-1-4503-2034-4.
- Kim, Soo-Min and Eduard Hovy (2006). ‘Extracting opinions, opinion holders, and topics expressed in online news media text.’ In *Proceedings of ACL/COLING Workshop on Sentiment and Subjectivity in Text*.
- Koppel, Moshe and Jonathan Schler (2006). ‘The importance of neutral examples in learning sentiment’. *Computational Intelligence*, 22(2).
- Lafferty, John, Andrew McCallum, and Fernando Pereira (2001). ‘Conditional random fields: Probabilistic models for segmenting and labeling sequence data’. In *Proceedings of the 18th International Conference on Machine Learning*. Morgan Kaufmann, San Francisco, CA.
- Lerman, Kevin, Sasha Blair-Goldensohn, and Ryan McDonald (2009). ‘Sentiment summarization: Evaluating and learning user preferences’. In *Proceedings of EAACL09*.
- Li, Fangtao, Yang Tang, Minlie Huang, and Xiaoyan Zhu (2009). ‘Answering opinion questions with random walks on graphs’. In *Proceedings of the Joint*

Conference of the 47th Annual Meeting of the ACL and the 4th International Joint Conference on Natural Language Processing of the AFNLP, 737–745. Association for Computational Linguistics.

Liu, Bing (2012). *Sentiment Analysis and Opinion Mining*. Morgan & Claypool Publishers.

Lu, Bin, Chenhao Tan, Claire Cardie, and Benjamin K. Tsou (2011). ‘Joint bilingual sentiment classification with unlabeled parallel corpora’. In *Proceedings of the 49th Annual Meeting of the Association for Computational Linguistics: Human Language Technologies*, 320–330.

Macdonald, Craig, Iadh Ounis, and Ian Soboroff (2008). ‘Overview of trec-2007 blog track’. In *Proceedings of TREC-2007*. Gaithersburg, USA.

Manning, Christopher D., Prabhakar Raghavan, and Hinrich Schtze (2008). *Introduction to Information Retrieval*. Cambridge University Press.

Mihalcea, Rada, Carmen Banea, and Janyce Wiebe (2007). ‘Learning multilingual subjective language via cross-lingual projections’. In *Proceedings of the 45th Annual Meeting of the Association of Computational Linguistics*, 976–983. Association for Computational Linguistics, Prague, Czech Republic.

Mikolov, Tomas, Ilya Sutskever, Kai Chen, Greg S Corrado, and Jeff Dean (2013). ‘Distributed representations of words and phrases and their compositionality’. In *Advances in Neural Information Processing Systems*, 3111–3119.

- Mitchell, T. (1997). *Machine Learning*. McGraw-Hill, New York.
- Mnih, Andriy and Geoffrey Hinton (2007). ‘Three new graphical models for statistical language modelling’. In *Proceedings of the 24th international conference on Machine learning*, 641–648. ACM.
- Mohammad, Saif, Cody Dunne, and Bonnie Dorr (2009). ‘Generating high-coverage semantic orientation lexicons from overtly marked words and a thesaurus’. In *Proceedings of the 2009 Conference on Empirical Methods in Natural Language Processing*, 599–608. Association for Computational Linguistics.
- Moilanen, Karo and Stephen Pulman (2007). ‘Sentiment composition’. In *Proceedings of RANLP 2007*. Borovets, Bulgaria.
- Montague, Richard (1974). *Formal Philosophy: Selected Papers of Richard Montague*, ed. Richmond Thomason. Yale University Press, New Haven.
- Nakagawa, Tetsuji, Kentaro Inui, and Sadao Kurohashi (2010). ‘Dependency tree-based sentiment classification using crfs with hidden variables’. In *Human Language Technologies: The 2010 Annual Conference of the North American Chapter of the Association for Computational Linguistics*, 786–794.
- Ng, Vincent (2010). ‘Supervised noun phrase coreference research: The first fifteen years’. In *Proceedings of the 48th Annual Meeting of the Association for Computational Linguistics*, 1396–1411.
- Ounis, Iadh, Craig Macdonald, and Ian Soboroff (2009). ‘Overview of trec-2008 blog track’. In *Proceedings of TREC-2008*. Gaithersburg, USA.

- Pang, Bo and Lillian Lee (2004). ‘A sentimental education: Sentiment analysis using subjectivity summarization based on minimum cuts’. In *Proceedings of the 42nd Annual Meeting of the Association for Computational Linguistics (ACL)*, 271–278.
- Pang, Bo and Lillian Lee (2008). ‘Opinion mining and sentiment analysis’. *Foundations and Trends in Information Retrieval*, 2, 1–135.
- Pang, Bo, Lillian Lee, and Shivakumar Vaithyanathan (2002). ‘Thumbs up? Sentiment classification using machine learning techniques’. In *Proceedings of the 2002 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, 79–86.
- Popescu, Ana-Maria and Oren Etzioni (2005). ‘Extracting product features and opinions from reviews’. In *Proceedings of Human Language Technology Conference and Conference on Empirical Methods in Natural Language Processing*, 339–346. Association for Computational Linguistics.
- Quirk, Randolph, Sidney Greenbaum, Geoffrey Leech, and Jan Svartvik (1985). *A Comprehensive Grammar of the English Language*. Longman, New York.
- Ratnaparkhi, Adwait (1996). ‘A maximum entropy model for part-of-speech tagging’. In *Proceedings of the Conference on Empirical Methods in Natural Language Processing*, 133–142.
- Riloff, Ellen and Janyce Wiebe (2003). ‘Learning extraction patterns for subjective expressions’. In *Proceedings of the 2003 conference on Empirical methods*

in natural language processing, 105–112. Association for Computational Linguistics, Morristown, NJ, USA.

Rudolph, Sebastian and Eugenie Giesbrecht (2010). ‘Compositional matrix-space models of language’. In *Proceedings of the 48th Annual Meeting of the Association for Computational Linguistics*, 907–916.

Salton, Gerard, J. Allan, and Chris Buckley (1993). ‘Approaches to passage retrieval in full text information systems’. In *Proceedings of the 16th annual international ACM SIGIR conference on Research and development in information retrieval*, SIGIR ’93, 49–58. ACM, New York, NY, USA.

Schuster, Mike and Kuldeep K Paliwal (1997). ‘Bidirectional recurrent neural networks’. *Signal Processing, IEEE Transactions on*, 45(11), 2673–2681.

Snyder, Benjamin and Regina Barzilay (2007). ‘Multiple aspect ranking using the good grief algorithm’. In *Proceedings of NAACL HLT*, 300–307.

Socher, Richard, Jeffrey Pennington, Eric H Huang, Andrew Y Ng, and Christopher D Manning (2011). ‘Semi-supervised recursive autoencoders for predicting sentiment distributions’. In *Proceedings of the Conference on Empirical Methods in Natural Language Processing*, 151–161. Association for Computational Linguistics.

Socher, Richard, Alex Perelygin, Jean Y Wu, Jason Chuang, Christopher D Manning, Andrew Y Ng, and Christopher Potts (2013). ‘Recursive deep models for semantic compositionality over a sentiment treebank’. In *Proceedings of the*

Conference on Empirical Methods in Natural Language Processing, EMNLP
'13.

Somasundaran, Swapna and Janyce Wiebe (2009). 'Recognizing stances in online debates'. In *Proceedings of the Joint Conference of the 47th Annual Meeting of the ACL and the 4th International Joint Conference on Natural Language Processing of the AFNLP*, 226–234. Association for Computational Linguistics.

Somasundaran, Swapna, Theresa Wilson, Janyce Wiebe, and Veselin Stoyanov (2007). 'Qa with attitude: Exploiting opinion type analysis for improving question answering in online discussions and the news'. In *Proceedings of the International Conference on Weblogs and Social Media*.

Stone, P.J. (1968). 'The general inquirer: A computer approach to content analysis'. *JOURNAL OF REGIONAL SCIENCE*, 8(1). ISSN 1467-9787.

Stoyanov, Veselin and Claire Cardie (2006). 'Partially supervised coreference resolution for opinion summarization through structured rule learning'. In *Proceedings of EMNLP*.

Stoyanov, Veselin and Claire Cardie (2008). 'Topic identification for fine-grained opinion analysis'. In *Proceedings of the 22nd International Conference on Computational Linguistics (Coling 2008)*, 817–824.

Subasic, Pero and Alison Huettner (2001). 'Affect analysis of text using fuzzy semantic typing'. In *IEEE-FS*, volume 9, 483–496.

- Titov, Ivan and Ryan McDonald (2008). ‘Modeling online reviews with multi-grain topic models’. In *Proceeding of the 17th international conference on World Wide Web*, 111–120. ACM.
- Turian, Joseph, Lev Ratinov, and Yoshua Bengio (2010). ‘Word representations: a simple and general method for semi-supervised learning’. In *Proceedings of the 48th Annual Meeting of the Association for Computational Linguistics*, 384–394. Association for Computational Linguistics.
- Turney, Peter D. and Michael L. Littman (2002). ‘Unsupervised learning of semantic orientation from a hundred-billion-word corpus’. Technical report, National Research Council Canada. NRC Technical Report ERB-1094.
- Wang, Lu, Hema Raghavan, Claire Cardie, and Vittorio Castelli (2014). ‘Query-focused opinion summarization for user-generated content’. In *Proceedings of COLING 2014, the 25th International Conference on Computational Linguistics: Technical Papers*, 1660–1669. Dublin City University and Association for Computational Linguistics, Dublin, Ireland.
- Wiebe, Janyce and Rebecca Bruce (1995). ‘Probabilistic classifiers for tracking point of view’. In *Proceedings of the AAAI Spring Symposium on Empirical Methods in Discourse Interpretation and Generation*, 181–187.
- Wiebe, Janyce, Theresa Wilson, and Claire Cardie (2005). ‘Annotating expressions of opinions and emotions in language’. *Language Resources and Evaluation*, 39(2–3), 165–210.

- Wiegand, Michael and Dietrich Klakow (2010). ‘Convolution kernels for opinion holder extraction’. In *Human Language Technologies: The 2010 Annual Conference of the North American Chapter of the Association for Computational Linguistics*, 795–803. Association for Computational Linguistics.
- Wilson, Theresa, Janyce Wiebe, and Paul Hoffmann (2005). ‘Recognizing contextual polarity in phrase-level sentiment analysis’. In *Proceedings of Human Language Technologies Conference/Conference on Empirical Methods in Natural Language Processing (HLT/EMNLP 2005)*. Vancouver, Canada.
- Yang, Bishan and Claire Cardie (2013). ‘Joint inference for fine-grained opinion extraction’. In *Proceedings of ACL*.
- Yang, Bishan and Claire Cardie (2014). ‘Joint modeling of opinion expression extraction and attribute classification’. *Transactions of the Association for Computational Linguistics*, 2, 505–516.
- Yessenalina, Ainur and Claire Cardie (2011). ‘Compositional matrix-space models for sentiment analysis’. In *Proceedings of the 2011 Conference on Empirical Methods in Natural Language Processing*, 172–182.
- Yi, J., T. Nasukawa, R. Bunescu, and W. Niblack (2003). ‘Sentiment analyzer: Extracting sentiments about a given topic using natural language processing techniques’. In *Proceedings of the 3rd IEEE International Conference on Data Mining*.

Zhou, GuangXu, Hemant Joshi, and Coskun Bayrak (2008). ‘Topic categorization for relevancy and opinion detection’. In *The Sixteenth Text REtrieval Conference (TREC 2007) Proceedings*. NIST.

Zhu, Jun, Amr Ahmed, and Eric P. Xing. (2009). ‘Medlda: Maximum margin supervised topic models for regression and classification’. In *Proceedings of the International Conference on Machine Learning*.

Zhuang, Li, Feng Jing, and Xiao-Yan Zhu (2006). ‘Movie review mining and summarization’. In *Proceedings of the 15th ACM international conference on Information and knowledge management (CIKM)*, 43–50. ACM.