Chapter 1

A SURVEY OF OPINION MINING AND SENTIMENT ANALYSIS

Bing Liu
University of Illinois at Chicago
Chicago, IL
liub@cs.uic.edu

Lei Zhang
University of Illinois at Chicago
Chicago, IL
lzhang32@gmail.com

Abstract Sentiment analysis or opinion mining is the computational study of people's opinions, appraisals, attitudes, and emotions toward entities, individuals, issues, events, topics and their attributes. The task is technically challenging and practically very useful. For example, businesses always want to find public or consumer opinions about their products and services. Potential customers also want to know the opinions of existing users before they use a service or purchase a product.

With the explosive growth of *social media* (i.e., reviews, forum discussions, blogs and social networks) on the Web, individuals and organizations are increasingly using public opinions in these media for their decision making. However, finding and monitoring opinion sites on the Web and distilling the information contained in them remains a formidable task because of the proliferation of diverse sites. Each site typically contains a huge volume of opinionated text that is not always easily deciphered in long forum postings and blogs. The average human reader will have difficulty identifying relevant sites and accurately summarizing the information and opinions contained in them. Moreover, it is also known that human analysis of text information is subject to considerable biases, e.g., people often pay greater attention to opinions that are consistent with their own preferences. People also have difficulty, owing to their mental and physical limitations, producing consistent

results when the amount of information to be processed is large. Automated opinion mining and summarization systems are thus needed, as subjective biases and mental limitations can be overcome with an objective sentiment analysis system.

In the past decade, a considerable amount of research has been done in academia [58,76]. There are also numerous commercial companies that provide opinion mining services. In this chapter, we first define the opinion mining problem. From the definition, we will see the key technical issues that need to be addressed. We then describe various key mining tasks that have been studied in the research literature and their representative techniques. After that, we discuss the issue of detecting opinion spam or fake reviews. Finally, we also introduce the research topic of assessing the utility or quality of online reviews.

1. The Problem of Opinion Mining

In this first section, we define the opinion mining problem, which enables us to see a structure from the intimidating unstructured text and to provide a unified framework for the current research. The abstraction consists of two parts: opinion definition and opinion summarization [31].

1.1 Opinion Definition

We use the following review segment on iPhone to introduce the problem (an id number is associated with each sentence for easy reference):

"(1) I bought an iPhone a few days ago. (2) It was such a nice phone. (3) The touch screen was really cool. (4) The voice quality was clear too. (5) However, my mother was mad with me as I did not tell her before I bought it. (6) She also thought the phone was too expensive, and wanted me to return it to the shop ..."

The question is: what we want to mine or extract from this review? The first thing that we notice is that there are several opinions in this review. Sentences (2), (3) and (4) express some positive opinions, while sentences (5) and (6) express negative opinions or emotions. Then we also notice that the opinions all have some targets. The target of the opinion in sentence (2) is the iPhone as a whole, and the targets of the opinions in sentences (3) and (4) are "touch screen" and "voice quality" of the iPhone respectively. The target of the opinion in sentence (6) is the price of the iPhone, but the target of the opinion/emotion in sentence (5) is "me", not iPhone. Finally, we may also notice the holders of opinions. The holder of the opinions in sentences (2), (3) and (4) is the author of the review ("I"), but in sentences (5) and (6) it is "my mother".

With this example in mind, we now formally define the opinion mining problem. We start with the *opinion target.*

In general, opinions can be expressed about anything, e.g., a product, a service, an individual, an organization, an event, or a topic, by any person or organization. We use the *entity* to denote the target object that has been evaluated. Formally, we have the following:

Definition 1.1 (Entity) *An* entity e *is a product, service, person, event, organization, or topic. It is associated with a pair,* $e : (T, W)$*, where* T *is a hierarchy of* components *(or parts),* sub-components*, and so on, and* W *is a set of attributes of* e*. Each component or sub-component also has its own set of attributes.*

An example of an entity is as follows:

Example 1.2 *A particular brand of cellular phone is an entity, e.g.,* iPhone*. It has a set of components, e.g., battery and screen, and also a set of attributes, e.g.,* voice quality*,* size*, and* weight*. The battery component also has its own set of attributes, e.g.,* battery life*, and* battery size*.*

Based on this definition, an entity is represented as a tree or hierarchy. The root of the tree is the name of the entity. Each non-root node is a component or sub-component of the entity. Each link is a part-of relation. Each node is associated with a set of attributes. An opinion can be expressed on any node and any attribute of the node.

In practice, it is often useful to simplify this definition due to two reasons: First, natural language processing is a difficult task. To effectively study the text at an arbitrary level of detail as described in the definition is very hard. Second, for an ordinary user, it is too complex to use a hierarchical representation. Thus, we simplify and flatten the tree to two levels and use the term *aspects* to denote both components and attributes. In the simplified tree, the root level node is still the entity itself, while the second level nodes are the different aspects of the entity.

For product reviews and blogs, opinion holders are usually the authors of the postings. Opinion holders are more important in news articles as they often explicitly state the person or organization that holds an opinion [5, 13, 49]. Opinion holders are also called *opinion sources* [107].

There are two main types of opinions: *regular opinions* and *comparative opinions*. Regular opinions are often referred to simply as opinions in the research literature. A comparative opinion expresses a relation of similarities or differences between two or more entities, and/or a preference of the opinion holder based on some of the shared aspects of the entities [36, 37]. A comparative opinion is usually expressed using the *comparative* or *superlative* form of an adjective or adverb, although not always. The discussion below focuses only on regular opinions. Comparative opinions will be discussed in Sect. 6. For simplicity, the terms *regular opinion* and *opinion* are used interchangeably below.

An opinion (or regular opinion) is simply a positive or negative sentiment, attitude, emotion or appraisal about an entity or an aspect of the entity from an opinion holder. Positive, negative and neutral are called *opinion orientations* (also called *sentiment orientations*, *semantic orientations*, or *polarities*). We are now ready to define an opinion [58].

DEFINITION 1.3 (OPINION) *An opinion (or regular opinion) is a quintuple,* $(e_i, a_{ij}, oo_{ijkl}, h_k, t_l)$, *where* e_i *is the name of an entity,* a_{ij} *is an aspect of* e_i, oo_{ijkl} *is the orientation of the opinion about aspect* a_{ij} *of entity* e_i, h_k *is the opinion holder, and* t_l *is the time when the opinion is expressed by* h_k. *The opinion orientation* oo_{ijkl} *can be positive, negative or neutral, or be expressed with different strength/intensity levels. When an opinion is on the entity itself as a whole, we use the special aspect GENERAL to denote it.*

These five components are essential. Without any of them, it can be problematic in general. For example, if one says "*The picture quality is great*", and we do not know whose picture quality, the opinion is of little use. However, we do not mean that every piece of information is needed in every application. For example, knowing each opinion holder is not necessary if we want to summarize opinions from a large number of people. Similarly, we do not claim that nothing else can be added to the quintuple. For example, in some applications the user may want to know the sex and age of each opinion holder.

An important contribution of this definition is that it provides a basis for transforming unstructured text to structured data. The quintuple gives us the essential information for a rich set of qualitative and quantitative analysis of opinions. Specifically, the quintuple can be regarded as a schema for a database table. With a large set of opinion records mined from text, database management systems tools can be applied to slice and dice the opinions for all kinds of analyses.

Objective of opinion mining: Given a collection of opinionated documents D, discover all opinion quintuples $(e_i, a_{ij}, oo_{ijkl}, h_k, t_l)$ in D.

To achieve this objective, one needs to perform the following tasks:

- **Task 1** (entity extraction and grouping): Extract all entity expressions in D, and group synonymous entity expressions into entity clusters. Each entity expression cluster indicates a unique entity e_i.

- **Task 2** (aspect extraction and grouping): Extract all aspect expressions of the entities, and group aspect expressions into clusters. Each aspect expression cluster of entity e_i indicates a unique aspect a_{ij}.
- **Task 3** (opinion holder and time extraction): Extract these pieces of information from the text or unstructured data.
- **Task 4** (aspect sentiment classification): Determine whether each opinion on an aspect is positive, negative or neutral.
- **Task 5** (opinion quintuple generation): Produce all opinion quintuples $(e_i, a_{ij}, oo_{ijkl}, h_k, t_l)$ expressed in D based on the results of the above tasks.

We use an example blog to illustrate these tasks (a sentence id is associated with each sentence):

EXAMPLE 1.4 (BLOG POSTING) **Posted by: bigXyz on Nov-4-2010:** *(1) I bought a Motorola phone and my girlfriend bought a Nokia phone yesterday. (2) We called each other when we got home. (3) The voice of my Moto phone was unclear, but the camera was good. (4) My girlfriend was quite happy with her phone, and its sound quality. (5) I want a phone with good voice quality. (6) So I probably will not keep it.*

Task 1 should extract the entity expressions, "Motorola", "Nokia" and "Moto", and group "Motorola" and "Moto" together as they represent the same entity. Task 2 should extract aspect expressions "camera", "voice", and "sound", and group "voice" and "sound" together as they are synonyms representing the same aspect. Task 3 should find the holder of the opinions in sentence (3) to be bigXyz (the blog author), and the holder of the opinions in sentence (4) to be bigXyz's girlfriend. It should also find the time when the blog was posted, which is Nov-4-2010. Task 4 should find that sentence (3) gives a negative opinion to the voice quality of the Motorola phone, but a positive opinion to its camera. Sentence (4) gives positive opinions to the Nokia phone as a whole and also its sound quality. Sentence (5) seemingly expresses a positive opinion, but it does not. To generate opinion quintuples for sentence (4), we also need to know what "her phone" is and what "its" refers to. All these are challenging problems. Task 5 should finally generate the following four opinion quintuples:

(Motorola, voice_quality, negative, bigXyz, Nov-4-2010)
(Motorola, camera, positive, bigXyz, Nov-4-2010)

(Nokia, GENERAL, positive, bigXyz's_girlfriend, Nov-4-2010)
(Nokia, voice_quality, positive, bigXyz's_girlfriend, Nov-4-2010)

Before going further, let us discuss two other important concepts related to opinion mining and sentiment analysis, i.e., *subjectivity* and *emotion.*

DEFINITION 1.5 (SENTENCE SUBJECTIVITY) *An objective sentence presents some factual information about the world, while a subjective sentence expresses some personal feelings, views or beliefs.*

For instance, in the above example, sentences (1) and (2) are objective sentences, while all other sentences are subjective sentences. Subjective expressions come in many forms, e.g., opinions, allegations, desires, beliefs, suspicions, and speculations [87, 103]. Thus, a subjective sentence may not contain an opinion. For example, sentence (5) in Example 4 is subjective but it does not express a positive or negative opinion about anything. Similarly, not every objective sentence contains no opinion. For example, "*the earphone broke in two days*", is an objective sentence but it implies a negative opinion. There is some confusion among researchers to equate subjectivity with opinion. As we can see, the concepts of subjective sentences and opinion sentences are not the same, although they have a large intersection. The task of determining whether a sentence is subjective or objective is called subjectivity classification [105], which we will discuss in Sect. 3.

DEFINITION 1.6 (EMOTION) *Emotions are our subjective feelings and thoughts.*

According to [80], people have 6 primary emotions, i.e., love, joy, surprise, anger, sadness, and fear, which can be sub-divided into many secondary and tertiary emotions. Each emotion can also have different intensities. The strengths of opinions are related to the intensities of certain emotions, e.g., joy, anger, and fear, as these sentences show: (1) "I am very angry with this shop," (2) "I am so happy with my iPhone," and (3) "with the current economic condition, I fear that I will lose my job." However, the concepts of emotions and opinions are not equivalent. Many opinion sentences express no emotion (e.g., "the voice of this phone is clear"), which are called *rational evaluation sentences*, and many emotion sentences give no opinion, e.g., "I am so surprised to see you."

1.2 Aspect-Based Opinion Summary

Most opinion mining applications need to study opinions from a large number of opinion holders. One opinion from a single person is usually not sufficient for action. This indicates that some form of summary of opinions is desirable. Opinion quintuples defined above provide an excellent source of information for generating both qualitative and quantitative summaries. A common form of summary is based on aspects, and is called *aspect-based opinion summary* (or *feature-based opinion summary*) [31, 60]. Below, we use an example to illustrate this form of summary, which is widely used in industry.

EXAMPLE 1.7 *Assume we summarize all the reviews of a particular cellular phone, cellular phone 1. The summary looks like that in Fig. 1.1, which was proposed in [31] and is called a structured summary. In the figure,* GENERAL *represents the phone itself (the entity). 125 reviews expressed positive opinions about the phone and 7 expressed negative opinions.* Voice quality *and* size *are two product aspects. 120 reviews expressed positive opinions about the voice quality, and only 8 reviews expressed negative opinions. The <individual review sentences> link points to the specific sentences and/or the whole reviews that give the positive or negative opinions. With such a summary, the user can easily see how existing customers feel about the phone. If he/she is interested in a particular aspect, he/she can drill down by following the <individual review sentences> link to see why existing customers like it and/or dislike it.*

Cellular phone 1:
Aspect: GENERAL
Positive: 125 <individual review sentences>
Negative: 7 <individual review sentences>
Aspect: Voice quality
Positive: 120 <individual review sentences>
Negative: 8 <individual review sentences>
...

Figure 1.1. An aspect-based opinion summary

The aspect-based summary in Fig. 1.1 can be visualized using a bar chart and opinions on multiple products can also be compared in a visualization (see [60]).

Researchers have also studied opinion summarization in the tradition fashion, e.g., producing a short *text summary* [4, 11, 51, 89, 91]. Such a summary gives the reader a quick overview of what people think about a product or service. A weakness of such a text-based summary is that it is not quantitative but only qualitative, which is usually not suitable for analytical purposes. For example, a traditional text summary may say "*Most people do not like this product*". However, a quantitative summary may say that 60% of the people do not like this product and 40% of them like it. In most opinion mining applications, the quantitative side is crucial just like in the traditional survey research. In survey research, aspect-based summaries displayed as bar charts or pie charts are commonly used because they give the user a concise, quantitative and visual view. Recently, researchers also tried to produce text summaries similar to that in Fig. 1.1 but in a more readable form [73, 81, 96].

2. Document Sentiment Classification

We are now ready to discuss some main research topics of opinion mining. This section focuses on *sentiment classification*, which has been studied extensively in the literature (see a survey in [76]). It classifies an opinion document (e.g., a product review) as expressing a positive or negative opinion or sentiment. The task is also commonly known as the *document-level sentiment classification* because it considers the whole document as the basic information unit.

Definition 1.8 (Document Level Sentiment) *Given an opinionated document d evaluating an entity e, determine the opinion orientation oo on e, i.e., determine oo on aspect GENERAL in the quintuple $(e, GENERAL, oo, h, t)$. e, h, and t are assumed known or irrelevant.*

An important assumption about sentiment classification is as follows:
Assumption: Sentiment classification assumes that the opinion document d (e.g., a product review) expresses opinions on a single entity e and the opinions are from a single opinion holder h.

This assumption holds for customer reviews of products and services because each such review usually focuses on a single product and is written by a single reviewer. However, it may not hold for a forum and blog posting because in such a posting the author may express opinions on multiple products, and compare them using comparative sentences.

Most existing techniques for document-level sentiment classification are based on supervised learning, although there are also some unsupervised methods. We give an introduction to them below.

2.1 Classification based on Supervised Learning

Sentiment classification obviously can be formulated as a supervised learning problem with three classes, positive, negative and neutral. Training and testing data used in the existing research are mostly product reviews, which is not surprising due to the above assumption. Since each review already has a reviewer-assigned rating (e.g., 1-5 stars), training and testing data are readily available. For example, a review with 4 or 5 stars is considered a positive review, a review with 1 or 2 stars is considered a negative review and a review with 3 stars is considered a neutral review.

Any existing supervised learning methods can be applied to sentiment classification, e.g., naive Bayesian classification, and support vector machines (SVM). Pang et al. [78] took this approach to classify movie reviews into two classes, positive and negative. It was shown that using unigrams (a bag of individual words) as features in classification performed well with either naive Bayesian or SVM.

Subsequent research used many more features and techniques in learning [76]. As most machine learning applications, the main task of sentiment classification is to engineer an effective set of features. Some of the current features are listed below.

- *Terms and their frequency:* These features are individual words or word n-grams and their frequency counts. In some cases, word positions may also be considered. The TF-IDF weighting scheme from information retrieval may be applied too. These features have been shown quite effective in sentiment classification.

- *Part of speech:* It was found in many researches that adjectives are important indicators of opinions. Thus, adjectives have been treated as special features.

- *Opinion words and phrases:* Opinion words are words that are commonly used to express positive or negative sentiments. For example, *beautiful, wonderful, good, and amazing* are positive opinion words, and *bad, poor, and terrible* are negative opinion words. Although many opinion words are adjectives and adverbs, nouns (e.g., *rubbish, junk, and crap*) and verbs (e.g., *hate* and *like*) can also indicate opinions. Apart from individual words, there are also opinion phrases and idioms, e.g., *cost someone an arm and a leg.* Opinion words and phrases are instrumental to sentiment analysis for obvious reasons.

- *Negations:* Clearly negation words are important because their appearances often change the opinion orientation. For example, the sentence "*I don't like this camera*" is negative. However, negation words must be handled with care because not all occurrences of such words mean negation. For example, "*not*" in "*not only but* *also*" does not change the orientation direction (see opinion shifters in Sect. 5.1).

- *Syntactic dependency:* Word dependency based features generated from parsing or dependency trees are also tried by several researchers.

Instead of using a standard machine learning method, researchers have also proposed several custom techniques specifically for sentiment classification, e.g., the score function in [15] based on words in positive and negative reviews. In [74], feature weighting schemes are used to enhance classification accuracy.

Manually labeling training data can be time-consuming and label-intensive. To reduce the labeling effort, *opinion words* can be utilized in the training procedure. In [95], Tan et al. used opinion words to label a portion of informative examples and then learn a new supervised classifier based on labeled ones. A similar approach is also used in [86]. In addition, opinion words can be utilized to increase the sentiment classification accuracy. In [68], Melville et al. proposed a framework to incorporate lexical knowledge in supervised learning to enhance accuracy.

Apart from classification of positive or negative sentiments, research has also been done on predicting the rating scores (e.g., 1-5 stars) of reviews [77]. In this case, the problem is formulated as regression since the rating scores are ordinal. Another interesting research direction is transfer learning or domain adaptation as it has been shown that sentiment classification is highly sensitive to the domain from which the training data is extracted. A classifier trained using opinionated documents from one domain often performs poorly when it is applied or tested on opinionated documents from another domain. The reason is that words and even language constructs used in different domains for expressing opinions can be quite different. To make matters worse, the same word in one domain may mean positive, but in another domain may mean negative. Thus, domain adaptation is needed. Existing research has used labeled data from one domain and unlabeled data from the target domain and general opinion words as features for adaptation [2, 7, 75, 112].

2.2 Classification based on Unsupervised Learning

It is not hard to imagine that opinion words and phrases are the dominating indicators for sentiment classification. Thus, using unsupervised learning based on such words and phrases would be quite natural. For example, the method in [93] uses known opinion words for classification, while [100] defines some phrases which are likely to be opinionated. Below, we give a description of the algorithm in [100], which consists of three steps:

Step 1: It extracts phrases containing adjectives or adverbs as adjectives and adverbs are good indicators of opinions. However, although an isolated adjective may indicate opinion, there may be insufficient context to determine its opinion orientation (called semantic orientation in [100]). For example, the adjective "*unpredictable*" may have a negative orientation in an automotive review, in such a phrase as "*unpredictable steering*", but it could have a positive orientation in a movie review, in a phrase such as "*unpredictable plot*". Therefore, the algorithm extracts two consecutive words, where one member of the pair is an adjective or adverb, and the other is a context word.

Two consecutive words are extracted if their POS tags conform to any of the patterns in Table 1.1. For example, the pattern in line 2 means that two consecutive words are extracted if the first word is an adverb and the second word is an adjective, but the third word cannot be a noun. NNP and NNPS are avoided so that the names of entities in the review cannot influence the classification.

EXAMPLE 1.9 *In the sentence "This camera produces beautiful pictures", "beautiful pictures" will be extracted as it satisfies the first pattern.*

Step 2: It estimates the semantic orientation of the extracted phrases using the pointwise mutual information (PMI) measure given in Equation 1.1:

$$PMI(term_1, term_2) = \log_2 \left(\frac{Pr(term_1 \wedge term_2)}{Pr(term_1) \cdot Pr(term_2)} \right) \tag{1.1}$$

Here, $Pr(term_1 \wedge term_2)$ is the co-occurrence probability of $term_1$ and $term_2$, and $Pr(term_1) \cdot Pr(term_2)$ gives the probability that the two terms co-occur if they are statistically independent. The ratio between $Pr(term_1 \wedge term_2)$ and $Pr(term_1) \cdot Pr(term_2)$ is thus a measure of the degree of statistical dependence between them. The log of this ratio is the amount of information that we acquire about the presence of one of the words when we observe the other. The semantic/opinion orientation (SO) of a phrase is computed based on its association with the positive reference word "*excellent*" and its association with the negative reference word "*poor*":

$$SO(phrase) = PMI(phrase, \text{“}excellent'') - PMI(phrase, \text{“}poor'') \qquad (1.2)$$

	First Word	Second Word	Third Word (Not Extracted)
1	JJ	NN or NNS	anything
2	RB, RBR, or RBS	JJ	not NN nor NNS
3	JJ	JJ	not NN nor NNS
4	NN or NNS	JJ	not NN nor NNS
5	RB, RBR, or RBS	VB, VBD, VBN, or VBG	anything

Table 1.1. Patterns of tags for extracting two-word phrases

The probabilities are calculated by issuing queries to a search engine and collecting the number of hits. For each search query, a search engine usually gives the number of relevant documents to the query, which is the number of hits. Thus, by searching the two terms together and separately, we can estimate the probabilities in Equation 1.1. The author of [100] used the AltaVista search engine because it has a NEAR operator, which constrains the search to documents that contain the words within ten words of one another in either order. Let hits(query) be the number of hits returned. Equation 1.2 can be rewritten as follows:

$$SO(phrase) = \log_2 \left(\frac{hits(\text{phrase NEAR “excellent”})hits(\text{“poor”})}{hits(\text{phrase NEAR “poor”})hits(\text{“}excellent'')} \right) \qquad (1.3)$$

To avoid division by 0, 0.01 is added to the hits.
Step 3: Given a review, the algorithm computes the average SO of all phrases in the review, and classifies the review as recommended if the average SO is positive, not recommended otherwise.

Final classification accuracies on reviews from various domains range from 84% for automobile reviews to 66% for movie reviews.

To summarize, we can see that the main advantage of document level sentiment classification is that it provides a prevailing opinion on an entity, topic or event. The main shortcomings are that it does not give details on what people liked and/or disliked and it is not easily applicable to non-reviews, e.g., forum and blog postings, because many such postings evaluate multiple entities and compare them.

3. Sentence Subjectivity and Sentiment Classification

Naturally the same document-level sentiment classification techniques can also be applied to individual sentences. The task of classifying a sentence as subjective or objective is often called *subjectivity classification* in the existing literature [30, 87, 88, 106, 109, 110, 113]. The resulting subjective sentences are also classified as expressing positive or negative opinions, which is called *sentence-level sentiment classification*.

DEFINITION 1.10 *Given a sentence s, two sub-tasks are performed:*

1 Subjectivity classification: *Determine whether s is a subjective sentence or an objective sentence,*

2 Sentence-level sentiment classification: *If s is subjective, determine whether it expresses a positive, negative or neutral opinion.*

Notice that the quintuple (e, a, oo, h, t) is not used in defining the problem here because sentence-level classification is often an intermediate step. In most applications, one needs to know what entities or aspects of the entities are the targets of opinions. Knowing that some sentences have positive or negative opinions but not about what, is of limited use. However, the two sub-tasks are still useful because (1) it filters out those sentences which contain no opinions, and (2) after we know what entities and aspects of the entities are talked about in a sentence, this step can help us determine whether the opinions about the entities and their aspects are positive or negative.

Most existing researches study both problems, although some of them focus only on one. Both problems are classification problems. Thus, traditional supervised learning methods are again applicable. For example, one of the early works reported in [104] performed subjectivity classification using the naive Bayesian classifier. Subsequent researches also used other learning algorithms.

Much of the research on sentence-level sentiment classification makes the following assumption:

Assumption: The sentence expresses a single opinion from a single opinion holder.

This assumption is only appropriate for simple sentences with a single opinion, e.g., "*The picture quality of this camera is amazing.*" However, for compound and complex sentences, a single sentence may express more than one opinion. For example, the sentence, "*The picture quality of this camera is amazing and so is the battery life, but the viewfinder is too small for such a great camera*", expresses both positive and negative opinions (it has mixed opinions). For "*picture quality*" and "*battery life*", the sentence is positive, but for "*viewfinder*", it is negative. It is also positive for the camera as a whole (i.e., the GENERAL aspect).

Many papers have been published on subjectivity classification and sentence-level sentiment classification. In [113], for subjectivity classification, it applied supervised learning. For sentiment classification of each subjective sentence, it used a similar method to that in Sect. 2.2 but with many more seed words, and the score function was log-likelihood ratio. The same problem was also studied in [30] considering gradable adjectives, and in [23] using semi-supervised learning. In [48-50], researchers also built models to identify some specific types of opinions.

As we mentioned earlier, sentence-level classification is not suitable for compound and complex sentences. It was pointed out in [109] that not only a single sentence may contain multiple opinions, but also both subjective and factual clauses. It is useful to pinpoint such clauses. It is also important to identify the strength of opinions. A study of automatic sentiment classification was presented to classify clauses of every sentence by the *strength* of the opinions being expressed in individual clauses, down to four levels deep (*neutral*, *low*, *medium*, and *high*). The strength of *neutral* indicates the absence of opinion or subjectivity. Strength classification thus subsumes the task of classifying a sentence as subjective versus objective. In [108], the problem was studied further using supervised learning by considering contextual sentiment influencers such as negation (e.g., *not* and *never*) and contrary (e.g., *but* and *however*). A list of influencers can be found in [82]. However, in many cases, identifying only clauses are insufficient because the opinions can be embedded in phrases, e.g., "*Apple is doing very well in this terrible economy.*" In this sentence, the opinion on "*Apple*" is clearly positive but on "*economy*" it is negative.

Besides analyzing opinion sentences in reviews, research has been done in threaded discussions, which includes forum discussions, emails, and newsgroups. In threaded discussions, people not only express their opinions on a topic but also interact with each other. However, the discussions could be highly emotional and heated with many emotional statements between participants. In [115], Zhai et al. proposed a method to identify those *evaluative sentences* from forum discussions, which only express people's opinions on entities or topics and their different aspects. In [28], Hassan et al. proposed an approach to find sentences with attitudes of participants toward one another. That is, it predicts whether a sentence contains an attitude toward a text recipient or not.

Finally, we should bear in mind that not all subjective sentences have opinions and those that do form only a subset of opinionated sentences.

Many objective sentences can imply opinions too. Thus, to mine opinions from text one needs to mine them from both subjective and objective sentences.

4. Opinion Lexicon Expansion

In the preceding sections, we mentioned that opinion words are employed in many sentiment classification tasks. We now discuss how such words are generated. In the research literature, opinion words are also known as opinion-bearing words or sentiment words. Positive opinion words are used to express some desired states while negative opinion words are used to express some undesired states. Examples of positive opinion words are: *beautiful*, *wonderful*, *good*, and *amazing*. Examples of negative opinion words are *bad*, *poor*, and *terrible*. Apart from individual words, there are also opinion phrases and idioms, e.g., *cost someone an arm and a leg*. Collectively, they are called the *opinion lexicon*. They are instrumental for opinion mining for obvious reasons.

To compile or collect the opinion word list, three main approaches have been investigated: manual approach, dictionary-based approach, and corpus-based approach. The manual approach is very time-consuming and thus it is not usually used alone, but combined with automated approaches as the final check because automated methods make mistakes. Below, we discuss the two automated approaches.

4.1 Dictionary based approach

One of the simple techniques in this approach is based on bootstrapping using a small set of seed opinion words and an online dictionary, e.g., WordNet [69] or thesaurus[71]. The strategy is to first collect a small set of opinion words manually with known orientations, and then to grow this set by searching in the WordNet or thesaurus for their synonyms and antonyms. The newly found words are added to the seed list. The next iteration starts. The iterative process stops when no more new words are found. This approach is used in [31, 49]. After the process completes, manual inspection can be carried out to remove and/or correct errors. Researchers have also used additional information (e.g., glosses) in WordNet and additional techniques (e.g., machine learning) to generate better lists [1, 19, 20, 45]. Several opinion word lists have been produced [17, 21, 31, 90, 104].

The dictionary based approach and the opinion words collected from it have a major shortcoming. The approach is unable to find opinion words with domain and context specific orientations, which is quite common. For example, for a speaker phone, if it is quiet, it is usually negative.

However, for a car, if it is quiet, it is positive. The corpus-based approach can help deal with this problem.

4.2 Corpus-based approach and sentiment consistency

The methods in the corpus-based approach rely on syntactic or co-occurrence patterns and also a seed list of opinion words to find other opinion words in a large corpus. One of the key ideas is the one proposed by Hazivassiloglou and McKeown [29]. The technique starts with a list of seed opinion adjectives, and uses them and a set of linguistic constraints or conventions on connectives to identify additional adjective opinion words and their orientations. One of the constraints is about the conjunction AND, which says that conjoined adjectives usually have the same orientation. For example, in the sentence, "*This car is beautiful and spacious*," if "*beautiful*" is known to be positive, it can be inferred that "*spacious*" is also positive. This is so because people usually express the same opinion on both sides of a conjunction. The following sentence is rather unnatural, "This car is beautiful and difficult to drive". If it is changed to "This car is beautiful but difficult to drive", it becomes acceptable. Rules or constraints are also designed for other connectives, OR, BUT, EITHER-OR, and NEITHER-NOR. This idea is called *sentiment consistency*. Of course, in practice it is not always consistent. Learning is applied to a large corpus to determine if two conjoined adjectives are of the same or different orientations. Same and different-orientation links between adjectives form a graph. Finally, clustering is performed on the graph to produce two sets of words: positive and negative. In [46], Kanayama and Nasukawa expanded this approach by introducing the idea of intra-sentential (within a sentence) and inter-sentential (between neighboring sentences) sentiment consistency (called coherency in [46]). The intra-sentential consistency is similar to that in [29]. Inter-sentential consistency applies the idea to neighboring sentences. That is, the same opinion orientation (positive or negative) is usually expressed in a few consecutive sentences. Opinion changes are indicated by adversative expressions such as *but* and *however*. Some criteria to determine whether to add a word to the positive or negative lexicon are also proposed. This study was based on Japanese text. In Sect. 5.4, a related but also quite different method will be described. Other related work includes [43, 44].

In [17], Ding et al. explored the idea of intra-sentential and inter-sentential sentiment consistency further. Instead of finding domain dependent opinion words, they showed that the same word could indicate different orientations in different contexts even in the same domain. This fact was also clearly depicted by the basic rules of opinions in Sect. 5.2. For example, in the digital camera domain, the word "*long*" expresses opposite opinions in the two sentences: "*The battery life is long*" (positive) and "*The time taken to focus is long*" (negative). Thus, finding domain dependent opinion words is insufficient. They then proposed to consider both possible opinion words and aspects together, and use the pair (aspect, opinion_word) as the opinion context, e.g., the pair ("*battery life*", "*long*"). Their method thus determines opinion words and their orientations together with the aspects that they modify. The above rules about connectives are still applied. The work in [24] adopted the same context definition but used it for analyzing comparative sentences. In [63], Lu et al. proposed an optimization framework to learn aspect-dependent sentiments in opinion context based on integer linear programming [14]. In fact, the method in [94, 100] can also be considered as a method for finding context specific opinions, but it does not use the sentiment consistency idea. Its opinion context is based on syntactic POS patterns rather than aspects and opinion words that modify them. All these context definitions, however, are still insufficient as the basic rules of opinions discussed in Sect. 5.2 show, i.e., many contexts can be more complex, e.g., consuming a large amount of resources. In [9], the problem of extracting opinion expressions with any number of words was studied. The Conditional Random Fields (CRF) method [52] was used as a sequence learning technique for extraction. In [84, 85], a double-propagation method was proposed to extraction both opinion words and aspects together. We describe it in Sect. 5.4.

Using the corpus-based approach alone to identify all opinion words, however, is not as effective as the dictionary-based approach because it is hard to prepare a huge corpus to cover all English words. However, as we mentioned above, this approach has a major advantage that the dictionary-based approach does not have. It can help find domain and context specific opinion words and their orientations using a domain corpus. Finally, we should realize that populating an opinion lexicon (domain dependent or not) is different from determining whether a word or phrase is actually expressing an opinion and what its orientation is in a particular sentence. Just because a word or phrase is listed in an opinion lexicon does not mean that it actually is expressing an opinion in a sentence. For example, in the sentence, "*I am looking for a good health insurance*", "*good*" does not express either a positive or negative opinion on any particular insurance. The same is true for opinion orientation. We should also remember that opinion words and phrases are not the only expressions that bear opinions. There are many others as we will see in Sect. 5.2.

5. Aspect-Based Sentiment Analysis

Although classifying opinionated texts at the document level or at the sentence level is useful in many cases, it does not provide the necessary detail needed for many other applications. A positive opinionated document about a particular entity does not mean that the author has positive opinions on all aspects of the entity. Likewise, a negative opinionated document does not mean that the author dislikes everything. In a typical opinionated document, the author writes both positive and negative aspects of the entity, although the general sentiment on the entity may be positive or negative. Document and sentence sentiment classification does not provide such information. To obtain these details, we need to go to the aspect level. That is, we need the full model of Sect. 1.1, i.e., aspect-based opinion mining. Instead of treating opinion mining simply as a classification of sentiments, aspect-based sentiment analysis introduces a suite of problems which require deeper natural language processing capabilities, and also produce a richer set of results.

Recall that, at the aspect level, the mining objective is to discover every quintuple $(e_i, a_{ij}, oo_{ijkl}, h_k, t_l)$ in a given document d. To achieve the objective, five tasks need to be performed. This section mainly focuses on the following two core tasks and they have also been studied more extensively by researchers (in Sect. 7, we will briefly discuss some other tasks):

1 **Aspect extraction:** Extract aspects that have been evaluated. For example, in the sentence, "*The picture quality of this camera is amazing*," the aspect is "*picture quality*" of the entity represented by "*this camera*". Note that "*this camera*" does not indicate the GENERAL aspect because the evaluation is not about the camera as a whole, but about its picture quality. However, the sentence "*I love this camera*" evaluates the camera as a whole, i.e., the GENERAL aspect of the entity represented by "*this camera*". Bear in mind whenever we talk about an aspect, we must know which entity it belongs to. In our discussion below, we often omit the entity just for simplicity of presentation.

2 **Aspect sentiment classification:** Determine whether the opinions on different aspects are positive, negative or neutral. In the first example above, the opinion on the "*picture quality*" aspect is positive, and in the second example, the opinion on the GENERAL aspect is also positive.

5.1 Aspect Sentiment Classification

We study the second task first, determining the orientation of opinions expressed on each aspect in a sentence. Clearly, the sentence-level and clause-level sentiment classification methods discussed in Sect. 3 are useful here. That is, they can be applied to each sentence or clause which contains some aspects. The aspects in it will take the opinion orientation of the sentence or clause. However, these methods have difficulty dealing with mixed opinions in a sentence and opinions that need phrase level analysis, e.g., "*Apple is doing very well in this terrible economy.*" Clause-level analysis also needs techniques to identify clauses which itself is a challenging task, especially with informal text of blogs and forum discussions, which is full of grammatical errors. Here, we describe a *lexicon-based approach* to solving the problem [17, 31], which tries to avoid these problems and has been shown to perform quite well. The extension of this method to handling comparative sentences is discussed in Sect. 6. In the discussion below, we assume that entities and their aspects are known. Their extraction will be discussed in Sect. 5.3, 5.4, and 7.

The lexicon-based approach basically uses an opinion lexicon, i.e., a list of *opinion words* and *phrases*, and a set of rules to determine the orientations of opinions in a sentence [17, 31]. It also considers opinion shifters and but-clauses. The approach works as follows:

1 **Mark opinion words and phrases:** Given a sentence that contains one or more aspects, this step marks all opinion words and phrases in the sentence. Each positive word is assigned the opinion score of +1, each negative word is assigned the opinion score of -1.

2 **Handle opinion shifters:** Opinion shifters (also called valence shifters [82]) are words and phrases that can shift or change opinion orientations. Negation words like not, never, none, nobody, nowhere, neither and cannot are the most common type. Additionally, sarcasm changes orientation too, e.g., "*What a great car, it failed to start the first day.*" Although it is easy to recognize such shifters manually, spotting them and handling them correctly in actual sentences by an automated system is by no means easy. Furthermore, not every appearance of an opinion shifter changes the opinion orientation, e.g., "*not only . . . but also*". Such cases need to be dealt with carefully.

3 **Handle but-clauses:** In English, *but* means contrary. A sentence containing *but* is handled by applying the following rule: the opinion orientation before *but* and after *but* are opposite to each

other if the opinion on one side cannot be determined. As in the case of negation, not every *but* means contrary, e.g., "*not only but also*". Such non-but phrases containing "*but*" also need to be considered separately. Finally, we should note that contrary words and phrases do not always indicate an opinion change, e.g., "*Car-x is great, but Car-y is better*". Such cases need to be identified and dealt with separately.

4 **Aggregating opinions:** This step applies an opinion aggregation function to the resulting opinion scores to determine the final orientation of the opinion on each aspect in the sentence. Let the sentence be s, which contains a set of aspects $\{a_1 \dots a_m\}$ and a set of opinion words or phrases $\{ow_1 \dots ow_n\}$ with their opinion scores obtained from steps 1, 2 and 3. The opinion orientation for each aspect a_i in s is determined by the following opinion aggregation function:

$$score(a_i, s) = \sum_{ow_j \in s} \frac{ow_j \cdot oo}{dist(ow_j, a_i)} \tag{1.4}$$

where ow_j is an opinion word/phrase in s, $dist(ow_j, a_i)$ is the distance between aspect a_i and opinion word ow_j in s. $ow_j.oo$ is the opinion score of ow_i. The multiplicative inverse is used to give lower weights to opinion words that are far away from aspect a_i. If the final score is positive, then the opinion on aspect a_i in s is positive. If the final score is negative, then the opinion on the aspect is negative. It is neutral otherwise.

This simple algorithm can perform quite well in many cases, but it is not sufficient in others. One main shortcoming is that opinion words and phrases do not cover all types of expressions that convey or imply opinions. There are in fact many other possible opinion bearing expressions. Most of them are also harder to deal with. Below, we list some of them, which we call the basic rules of opinions [58, 59].

5.2 Basic Rules of Opinions

An opinion rule expresses a concept that implies a positive or negative opinion [58, 59]. In actual sentences, the concept can be expressed in many different ways in natural language. We present these rules using a formalism that is similar to the BNF form. The top level rules are as follows:

1. POSITIVE ::= P

2.			— PO
3.			— orientation shifter N
4.			— orientation shifter NE
5.	NEGATIVE	::=	N
6.			— NE
7.			— orientation shifter P
8.			— orientation shifter PO

The non-terminals P and PO represent two types of *positive opinion expressions*. The non-terminal N and NE represent two types of *negative opinion expressions*. 'opinion shifter N' and 'opinion shifter NE' represent the negation of N and NE respectively, and 'opinion shifter P' and 'opinion shifter PO' represent the negation of P and PO respectively. We can see that these are not expressed in the actual BNF form but a pseudo-language stating some concepts. The reason is that we are unable to specify them precisely because for example, in an actual sentence, the opinion shifter may be in any form and can appear before or after N, NE, P, or PO. POSITIVE and NEGATIVE are the final orientations used to determine the opinions on the aspects in a sentence.

We now define N, NE, P and PO, which contain no opinion shifters. These opinion expressions are grouped into 6 conceptual categories based on their characteristics:

1 *Opinion word or phrase:* This is the most commonly used category, in which opinion words or phrases alone can imply positive or negative opinions on aspects, e.g., "*great*" in "*The picture quality is great*". These words or phrases are reduced to P and N.

9.	P	::= a positive opinion word or phrase
10.	N	::= an negative opinion word or phrase

Again, the details of the right-hand-sides are not specified (which also apply to all the subsequent rules). It is assumed that a set of positive and negative opinion words/phrases exist for an application.

2 *Desirable or undesirable fact:* In this case, it is a factual statement, and the description uses no opinion words, but in the context of the entity, the description implies a positive or negative opinion. For example, the sentence "*After my wife and I slept on it for two*

weeks, I noticed a mountain in the middle of the mattress" indicates a negative opinion about the mattress. However, the word "mountain" itself does not carry any opinion. Thus, we have the following two rules:

11.	P	::= desirable fact
12.	N	::= undesirable fact

3 *High, low, increased and decreased quantity of a positive or negative potential item:* For some aspects, a small value/quantity of them is negative, and a large value/quantity of them is positive, e.g., "*The battery life is short*" and "*The battery life is long.*" We call such aspects *positive potential items (PPI)*. Here "*battery life*" is a positive potential item. For some other aspects, a small value/quantity of them is positive, and a large value/quantity of them is negative, e.g., "*This phone costs a lot*" and "*Sony reduced the price of the camera.*" We call such aspects *negative potential items (NPI)*. "*cost*" and "*price*" are negative potential items. Both positive and negative potential items themselves express no opinions, i.e., "*battery life*" and "*cost*", but when they are modified by quantity adjectives or quantity change words or phrases, positive or negative opinions are implied.

13.	PO	::= no, low, less or decreased quantity of NPI
14.		— large, larger, or increased quantity of PPI
15.	NE	::= no, low, less, or decreased quantity of PPI
16.		— large, larger, or increased quantity of NPI
17.	NPI	::= a negative potential item
18.	PPI	::= a positive potential item

4 *Decreased and increased quantity of an opinionated item (N and P):* This set of rules is similar to the negation rules 3, 4, 7, and 8 above. Decreasing or increasing the quantity associated with an opinionated item (often nouns and noun phrases) can change the orientation of the opinion. For example, in the sentence "*This drug reduced my pain significantly*", "*pain*" is a negative opinion word, and the reduction of "*pain*" indicates a desirable effect of the drug. Hence, decreased pain implies a positive opinion on the drug. The concept of decreasing also extends to removal and disappearance, e.g., "*My pain has disappeared after taking the drug.*"

19.	PO	::= less or decreased N
20.		— more or increased P
21.	NE	::= less or decreased P
22.		— more or increased N

Rules 20 and 22 may not be needed as there is no change of opinion orientation, but they can change the opinion intensity. The key difference between this set of rules and the rules in the previous category is that no opinion words or phrases are involved in the previous category.

5 *Deviation from the norm or some desired value range:* In some application domains, the value of an aspect may have a desired range or norm. If it is above or below the normal range, it is negative, e.g., "*This drug causes low (or high) blood pressure*" and "*This drug causes my blood pressure to reach 200.*" Notice that no opinion word appeared in these sentences.

23.	PO	::= within the desired value range
24.	NE	::= above or below the desired value range

6 *Producing and consuming resources and wastes:* If an entity produces a lot of resources, it is positive. If it consumes a lot of resources, it is negative. For example, water is a resource. The sentence, "*This washer uses a lot of water*" gives a negative opinion about the washer. Likewise, if an entity produces a lot of wastes, it is negative. If it consumes a lot of wastes, it is positive.

25.	PO	::= produce a large quantity of or more resource
26.		— produce no, little or less waste
27.		— consume no, little or less resource
28.		— consume a large quantity of or more waste
29.	NE	::= produce no, little or less resource
30.		— produce some or more waste
31.		— consume a large quantity of or more resource
32.		— consume no, little or less waste

We should note that these rules are not the only rules that govern expressions of positive and negative opinions. With further research, additional new rules may be discovered.

5.3 Aspect Extraction

Existing research on *aspect extraction* (more precisely, *aspect expression extraction*) is mainly carried out in online reviews. We thus focus on reviews here. We describe some unsupervised methods for finding aspect expressions that are nouns and noun phrases. The first method is due to [31]. The method consists of two steps:

1 Find frequent nouns and noun phrases. Nouns and noun phrases (or groups) are identified by a POS tagger. Their occurrence frequencies are counted, and only the frequent ones are kept. A frequency threshold can be decided experimentally. The reason for using this approach is that when people comment on different aspects of a product, the vocabulary that they use usually converges. Thus, those nouns that are frequently talked about are usually genuine and important aspects. Irrelevant contents in reviews are often diverse, i.e., they are quite different in different reviews. Hence, those infrequent nouns are likely to be non-aspects or less important aspects.

2 Find infrequent aspects by exploiting the relationships between aspects and opinion words. The above step can miss many genuine aspect expressions which are infrequent. This step tries to find some of them. The idea is as follows: The same opinion word can be used to describe or modify different aspects. Opinion words that modify frequent aspects can also modify infrequent aspects, and thus can be used to extract infrequent aspects. For example, "*picture*" has been found to be a frequent aspect, and we have the sentence,
"*The pictures are absolutely amazing.*"
If we know that "*amazing*" is an opinion word, then "*software*" can also be extracted as an aspect from the following sentence,
"*The software is amazing.*"
because the two sentences follow the same dependency pattern and "*software*" in the sentence is also a noun. This idea of using the modifying relationship of opinion words and aspects to extract aspects was later generalized to using dependency relations [120], which was further developed into the double-propagation method for simultaneously extracting both opinion words and aspects [85]. The double-propagation method will be described in Sect. 5.4.

The precision of step 1 of the above algorithm was improved in [83]. Their algorithm tries to remove those noun phrases that may not be product aspects/features. It evaluates each noun phrase by computing a pointwise mutual information (PMI) score between the phrase and some *meronymy discriminators* associated with the product class, e.g., a scanner class. The meronymy discriminators for the scanner class are, "of scanner", "scanner has", "scanner comes with", etc., which are used to find components or parts of scanners by searching the Web.

$$PMI(a, d) = \frac{hits(a \wedge d)}{hits(a) \cdot hits(d)} \tag{1.5}$$

where a is a candidate aspect identified in step 1 and d is a discriminator. Web search is used to find the number of hits of individual terms and also their co-occurrences. The idea of this approach is clear. If the PMI value of a candidate aspect is too low, it may not be a component of the product because a and d do not co-occur frequently.

Other related works on aspect extraction use existing knowledge, supervised learning, semi-supervised learning, topic modeling and clustering. For example, many information extraction techniques can also be applied, e.g., Conditional Random Fields (CRF) [33, 52], and Hidden Markov Models (HMM) [22, 34, 35], and sequential rule mining [60]. Wu et al. [111] used dependency tree kernels. Su et al. [92] proposed a clustering method with mutual reinforcement to identify implicit aspects.

Topic modeling methods have also been attempted as an unsupervised and knowledge-lean approach. Titov and McDonald [99] showed that global topic models such as LDA (Latent Dirichlet allocation [6]) might not be suitable for detecting rateable aspects. They proposed multi-grain topic models to discover local rateable aspects. Here each discovered aspect is a unigram language model, i.e., a multinomial distribution over words. Such a representation is thus not as easy to interpret as aspects extracted by previous methods, but its advantage is that different words expressing the same or related aspects (more precisely aspect expressions) can usually be automatically grouped together under the same aspect. However, Titov and McDonald [99] did not separate aspects and opinion words in the discovery. Lin and He [57] proposed a joint topic-sentiment model also by extending LDA, where aspect words and opinion words were still not explicitly separated. To separate aspects and opinion words using topic models, Mei et al. [67] proposed to use a positive sentiment model and a negative sentiment model in additional to aspect models. Brody and Elhadad [10] proposed to first identify aspects using topic models and then identify aspect-specific opinion words by considering adjectives only. Zhao et al. [119] proposed a MaxEnt-LDA hybrid model to jointly discover both aspect words and aspect-specific opinion words, which can leverage syntactic features to help separate aspects and opinion words. Topic modeling based approaches were also used by Liu et al. [62] and Lu et al. [65].

Another line of work is to associate aspects with opinion/sentiment ratings. It aims to predict ratings based on learned aspects or jointly model aspects and ratings. Titov and McDonald [98] proposed a statistical model that is able to discover aspects from text and to extract textual evidence from reviews supporting each aspect rating. Lu et al. [66] defined a problem of rated aspect summarization. They proposed to use the structured probabilistic latent semantic analysis method to learn aspects from a bag of phrases, and a local/global method to predict aspect ratings. Wang et al. [102] proposed to infer both aspect ratings and aspect weights at the level of individual reviews based on learned latent aspects. Jo and Oh [41] proposed an *Aspect* and *Sentiment Unification Model* (ASUM) to model sentiments toward different aspects.

5.4 Simultaneous Opinion Lexicon Expansion and Aspect Extraction

In [84, 85], a method was proposed to extract both opinion words and aspects simultaneously by exploiting some syntactic relations of opinion words and aspects. The method needs only an initial set of opinion word seeds as the input and no seed aspects are required. It is based on the observation that opinions almost always have targets. Hence there are natural relations connecting opinion words and targets in a sentence due to the fact that opinion words are used to modify targets. Furthermore, it was found that opinion words have relations among themselves and so do targets among themselves too. The opinion targets are usually aspects. Thus, opinion words can be recognized by identified aspects, and aspects can be identified by known opinion words. The extracted opinion words and aspects are utilized to identify new opinion words and new aspects, which are used again to extract more opinion words and aspects. This propagation or bootstrapping process ends when no more opinion words or aspects can be found. As the process involves propagation through both opinion words and aspects, the method is called double propagation. Extraction rules are designed based on different relations between opinion words and aspects, and also opinion words and aspects themselves. Specifically, four subtasks are performed:

1 extracting aspects using opinion words;

2 extracting aspects using the extracted aspects;

3 extracting opinion words using the extracted aspects;

4 extracting opinion words using both the given and the extracted opinion words.

Dependency grammar [97] was adopted to describe the relations. The algorithm uses only a simple type of dependencies called *direct dependencies* to model the relations. A direct dependency indicates that one word depends on the other word without any additional words in their dependency path or they both depend on a third word directly. Some constraints are also imposed. Opinion words are considered to be adjectives and aspects nouns or noun phrases.

For example, in an opinion sentence "*Canon G3 produces great pictures*", the adjective "*great*" is parsed as directly depending on the noun "*pictures*" through relation *mod*. If we know "*great*" is an opinion word and are given the rule 'a noun on which an opinion word directly depends through *mod* is taken as an aspect', we can extract "*pictures*" as an aspect. Similarly, if we know "*pictures*" is an aspect, we can extract "*great*" as an opinion word using a similar rule.

6. Mining Comparative Opinions

Directly or indirectly expressing positive or negative opinions about an entity and its aspects is only one form of evaluation. Comparing the entity with some other similar entities is another. Comparisons are related to but are also quite different from regular opinions. They not only have different semantic meanings, but also different syntactic forms. For example, a typical regular opinion sentence is "*The picture quality of this camera is great*", and a typical comparative sentence is "*The picture quality of Camera-x is better than that of Camera-y.*" This section first defines the problem, and then presents some existing methods to solve it [15, 18, 24, 37].

In general, a comparative sentence expresses a relation based on similarities or differences of more than one entity. The comparison is usually conveyed using the comparative or superlative form of an adjective or adverb. A comparative sentence typically states that one entity has more or less of a certain attribute than another entity. A superlative sentence typically states that one entity has the most or least of a certain attribute among a set of similar entities. In general, a comparison can be between two or more entities, groups of entities, and one entity and the rest of the entities. It can also be between an entity and its previous versions.

Two types of comparatives: In English, comparatives are usually formed by adding the suffix *-er* and superlatives are formed by adding the suffix *-est* to their base *adjectives* and *adverbs*. For example, in "*The battery life of Camera-x is longer than that of Camera-y*", "*longer*" is the comparative form of the adjective "*long*". In "*The battery life of this camera is the longest*", "*longest*" is the superlative form of the adjective "*long*". We call this type of comparatives and superlatives as *Type 1* comparatives and superlatives. Note that for simplicity, we often use comparative to mean both comparative and superlative if superlative is not explicitly stated.

Adjectives and adverbs with two syllables or more and not ending in *y* do not form comparatives or superlatives by adding *-er* or *-est*. Instead, *more*, *most*, *less* and *least* are used before such words, e.g., more beautiful. We call this type of comparatives and superlatives as *Type 2* comparatives and superlatives. Both Type 1 and Type 2 are called regular comparatives and superlatives. In English, there are also irregular comparatives and superlatives, i.e., *more, most, less, least, better, best, worse, worst, further/farther* and *furthest/farthest*, which do not follow the above rules. However, they behave similarly to Type 1 comparatives and are thus grouped under Type 1.

Apart from these standard comparatives and superlatives, many other words or phrases can also be used to express comparisons, e.g., *prefer* and *superior*. For example, the sentence, "*Camera-x's quality is superior to Camera-y*", says that "*Camera-x is better or preferred.*" In [36], Jindal and Liu identified a list of such words. Since these words behave similarly to Type 1 comparatives, they are also grouped under Type 1.

Types of comparative relations: Comparative relations or comparisons can be grouped into four main types. The first three types are called the *gradable comparisons* and the last one the *non-gradable comparisons.*

1 *Non-equal gradable comparisons:* Relations of the type *greater or less than* that express an ordering of some entities with regard to some of their shared aspects, e.g., "*The Intel chip is faster than that of AMD*". This type also includes user preferences, e.g., "*I prefer Intel to AMD*".

2 *Equative comparisons:* Relations of the type *equal to* that state two or more entities are equal with regard to some of their shared aspects, e.g., "*The performance of Car-x is about the same as that of Car-y.*"

3 *Superlative comparisons:* Relations of the type *greater or less than all others* that rank one entity over all others, e.g., "*The Intel chip is the fastest*".

4 *Non-gradable comparisons:* Relations that compare aspects of two or more entities, but do not grade them. There are three main sub-types:

- Entity A is similar to or different from entity B with regard to some of their shared aspects, e.g., "*Coke tastes differently from Pepsi.*"
- Entity A has aspect a_1, and entity B has aspect a_2 (a_1 and a_2 are usually substitutable), e.g., "*Desktop PCs use external speakers but laptops use internal speakers.*"
- Entity A has aspect a, but entity B does not have, e.g., "*Phone-x has an earphone, but Phone-y does not have.*"

Comparative words used in non-equal gradable comparisons can be further categorized into two groups according to whether they express increased or decreased quantities, which are useful in opinion analysis.
• *Increasing comparatives:* Such a comparative expresses an increased quantity, e.g., *more* and *longer.*
• *Decreasing comparatives:* Such a comparative expresses a decreased quantity, e.g., *less* and *fewer.*

Objective of mining comparative opinions: Given a collection of opinionated documents D, discover in D all comparative opinion sextuples of the form (E_1, E_2, A, PE, h, t), where E_1 and E_2 are the entity sets being compared based on their shared aspects A (entities in E_1 appear before entities in E_2 in the sentence), $PE(\in \{E1, E2\})$ is the preferred entity set of the opinion holder h, and t is the time when the comparative opinion is expressed.

EXAMPLE 1.11 *Consider the comparative sentence* "Canon's optics is better than those of Sony and Nikon." *written by John in 2010. The extracted comparative opinion is:*
({Canon}, {Sony, Nikon}, {optics}, preferred: {Canon}, John, 2010)
The entity set E_1 is {Canon}, the entity set E_2 is {Sony, Nikon}, their shared aspect set A being compared is {optics}, the preferred entity set is {Canon}, the opinion holder h is John and the time t when this comparative opinion was written is 2010.

To mine comparative opinions, the tasks of extracting entities, aspects, opinion holders and times are the same as those for mining regular opinions. In [37], a method based on label sequential rules (LSR) is proposed to extract entities and aspects that are compared. A similar approach is described in [54] for extracting the compared entities. Clearly, the approaches discussed in previous sections are applicable as well, and so are many other information extraction methods. See [37, 24, 18] for some existing methods for performing sentiment analysis of comparative sentences, i.e., identifying comparative sentences and identifying the preferred entity set.

7. Some Other Problems

Besides the problems discussed in previous sections, there are many other challenges in opinion mining. This section gives an introduction to some of them. As we will see, most of these problems are related to their general problems that have been studied before but the opinion text provides more clues for their solutions and also has additional requirements.

Entity, opinion holder, and time extraction: In some applications, it is useful to identify and extract entities, opinion holders, and the times when opinions are given. These extraction tasks are collectively called Named Entity Recognition (NER). They have been studied extensively in the literature.

In the case of social media on the Web, the opinion holders are often the authors of the discussion postings, bloggers, or reviewers, whose login ids are known although their true identities in the real world may be unknown. The date and time when an opinion is submitted are also known and displayed on the page, so their extraction is easy [59].

For entity name extraction, there is a difference from NER. In a typical opinion mining application, the user wants to find opinions on some competing entities, e.g., competing products or brands. However, he/she often can only provide a few names because there are so many different brands and models. Furthermore, Web users also write names of the same product brands in many ways. For example, "*Motorola*" may be written as "*Moto*" or "*Mot*", and "*Samsung*" may be written as "*Sammy*". Product model names have even more variations. It is thus important for a system to automatically discover them from a relevant corpus. The key requirement is that the discovered entities must be of the same type as entities provided by the user (e.g., phone brands and models). In [55], this problem was modeled as a *set expansion problem* [25, 79], which expands a set of given seed entities (e.g., product names). Formally, the problem is stated as follows: Given a set Q of seed entities of a particular class C, and a set D of candidate entities, we wish to determine which of the entities in D belong to C. That is, we "grow" the class C based on the set of seed examples Q. Although this is a classification problem, in practice, the problem is often solved as a ranking problem, i.e., to rank the entities in D based on their likelihoods of belonging to C. It was shown that learning from positive and unlabeled examples provides a more effective method than the traditional distributional similarity methods [53, 79] and the machine learning technique *Bayesian Sets* [25] which was designed specifically for set expansion.

Objective expressions implying sentiments: Much of the research on sentiment analysis focuses on subjective sentences, which are regarded as opinion bearing. However, many objective sentences can bear opinions as well. For example, in a mattress review, the sentence "*Within a month, a valley formed in the middle of the mattress*" is not a subjective sentence, but an objective sentence. However, it implies a negative opinion about the mattress. Specifically, "*valley*" in this context indicates the quality of the mattress (a product aspect) and implies a negative opinion. Objective words (or sentences) that imply opinions are very difficult to recognize because their recognition typically requires the commonsense or world knowledge of the application domain. In [116], a method was proposed to deal with the problem of product aspects which are nouns and imply opinions using a large corpus. Our experimental results show some promising results. However, the accuracy is still low, and much further research is still needed.

Grouping aspect expressions indicating the same aspects: It is common that people use different words or phrases (which are called aspect expressions in Sect. 1) to describe the same aspect. For example, *photo* and *picture* refer to the same aspect in digital camera reviews. Identifying and grouping aspect expressions indicating the same aspect are essential for applications. Although WordNet [69] and other thesaurus dictionaries help to some extent, they are far from sufficient due to the fact that many synonyms are domain dependent. For example, *picture* and *movie* are synonyms in movie reviews, but they are not synonyms in digital camera reviews as *picture* is more related to *photo* while *movie* refers to *video*. It is also important to note that although most aspect expressions of an aspect are domain synonyms, they are not always synonyms. For example, "*expensive*" and "*cheap*" can both indicate the aspect price but they are not synonyms of price.

Carenini et al [12] proposed the first method to solve this problem in the context of opinion mining. Their method is based on several similarity metrics defined using string similarity, synonyms and distances measured using WordNet. It requires a taxonomy of aspects to be given for a particular domain. The algorithm merges each discovered aspect expression to an aspect node in the taxonomy. Experiments based on digital camera and DVD reviews showed promising results.

In [114], Zhai et al. proposed a semi-supervised learning method to group aspect expressions into the user specified aspect groups. Each group represents a specific aspect. To reflect the user needs, he/she first manually labels a small number of seeds for each group. The system then assigns the rest of the discovered aspect expressions to suitable groups using semi-supervised learning based on labeled seeds and unlabeled examples. The method used the Expectation-Maximization (EM) algorithm. Two pieces of prior knowledge were used to provide a better initialization for EM, i.e., (1) aspect expressions sharing some common words are likely to belong to the same group, and (2) aspect expressions that are synonyms in a dictionary are likely to belong to the same group.

Mapping implicit aspect expressions to aspects: There are many types of implicit aspect expressions. Adjectives are perhaps the most common type. Many adjectives modify or describe some specific attributes or properties of entities. For example, the adjective "heavy" usually describes the aspect weight of an entity. "Beautiful" is normally used to describe (positively) the aspect look or appearance of an entity. By no means, however, does this say that these adjectives only describe such aspects. Their exact meanings can be domain dependent. For example, "*heavy*" in the sentence "*the traffic is heavy*" does not describe the weight of the traffic. One way to map implicit aspect expressions to aspects is to manually compile a list of such mappings during training data annotation, which can then be used in the same domain in the future. However, we should note that some implicit aspect expressions are very difficult to extract and to map, e.g., "*fit in pockets*" in the sentence "*This phone will not easily fit in pockets*".

Coreference resolution: This problem has been extensively studied in the NLP community. However, the sentiment analysis context has additional needs. In [16], the problem of entity and aspect coreference resolution was proposed. It determines which mentions of entities and/or aspects refer to the same entities. The key interesting points were the design and testing of two opinion-related features for machine learning. The first feature is based on opinion analysis of regular sentences and comparative sentences, and the idea of sentiment consistency. For example, we have the sentences, "*The Sony camera is better than the Canon camera. It is cheap too.*" It is clear that "*It*" means "*Sony*" because in the first sentence, the opinion about "*Sony*" is positive (comparative positive), but it is negative (comparative negative) about "*Canon*", and the second sentence is positive. Thus, we can conclude that "*It*" refers to "*Sony*" because people usually express sentiments in a consistent way. It is unlikely that "*It*" refers to "*Canon*". As we can see, to obtain this feature, the system needs to have the ability to determine positive and negative opinions expressed in regular and comparative sentences.

The second feature considers what entities and aspects are modified by what opinion words. Consider these sentences, "*The picture quality of the Canon camera is very good. It is not expensive either.*" The question is what "*It*" refers to, "*Canon camera*" or "*picture quality*". Clearly, we know that "*It*" refers to "*Canon camera*" because "*picture quality*" cannot be expensive. To obtain this feature, the system needs to identify what opinion words are usually associated with what entities or aspects, which means that the system needs to discover such relationships from the corpus. These two features can boost the coreference resolution accuracy.

Cross lingual opinion mining: This research involves opinion mining for a language corpus based on the corpora from other languages. It is needed in following scenarios. Firstly, there are many English sentiment corpora on the Web nowadays, but for other languages (e.g. Chinese), the annotated sentiment corpora are limited [101]. And it is not a trivial task to label them manually. Utilizing English corpora for opinion mining in Chinese can relieve the labeling burden. Secondly, there are many situations where opinion mining results need to be multilanguage-comparable. For example, global companies need to analyze customer feedback for their products and services from many countries in different languages [47]. Thus, cross-lingual opinion mining is necessary. The basic idea of the current research is to utilize available language corpora to train sentiment classifiers for the target language data. Machine translation is typically used [3, 8, 27, 47, 101].

8. Opinion Spam Detection

It has become a common practice for people to find and to read opinions on the Web for many purposes. For example, if one wants to buy a product, one typically goes to a merchant or review site (e.g., amazon.com) to read some reviews of existing users of the product. If one sees many positive reviews of the product, one is very likely to buy the product. However, if one sees many negative reviews, he/she will most likely choose another product. Positive opinions can result in significant financial gains and/or fames for organizations and individuals. This, unfortunately, gives good incentives for opinion spam, which refers to human activities (e.g., write spam reviews) that try to deliberately mislead readers or automated opinion mining systems by giving undeserving positive opinions to some target entities in order to promote the entities and/or by giving unjust or false negative opinions to some other entities in order to damage their reputation. Such opinions are also called fake opinions, bogus opinions, or fake reviews. The problem of detecting fake or spam opinions was introduced by Jindal and Liu in [38, 39].

Individual Spammers and Group Spammers: A spammer may act individually (e.g., the author of a book) or as a member of a group (e.g., a group of employees of a company).

Individual spammers: In this case, a spammer, who does not work with anyone else, writes spam reviews. The spammer may register at a review site as a single user, or as many fake users using different user-ids. He/she can also register at multiple review sites and write spam reviews.
Group spammers: A group of spammers works together to promote a target entity and/or to damage the reputation of another. They may also register at multiple sites and spam on these sites. Group spam can be very damaging because they may take control of the sentiment on a product and completely mislead potential customers.

8.1 Spam Detection Based on Supervised Learning

In general, spam detection can be formulated as a classification problem with two classes, spam and non-spam. However, manually labeling the training data for learning is very hard, if not impossible. The problem is that identifying spam reviews by simply reading the reviews is extremely difficult because a spammer can carefully craft a spam review that is just like any innocent review.

Since manually labeling training data is hard, other ways have to be explored in order to find training examples for detecting possible fake reviews. In [38], it exploited duplicate reviews. In their study of 5.8 million reviews, 2.14 million reviewers and 6.7 million products from amazon.com, a large number of duplicate and near-duplicate reviews were found. Certain types of duplicate and near-duplicate reviews were regarded as spam reviews, and the rest of the reviews as non-spam reviews.

In [38, 39], three sets of features were identified for learning:

1 *Review centric features:* These are features about the content of reviews. Example features include actual words in a review, the number of times that brand names are mentioned, the percentage of opinion words, the review length, and the number of helpful feedbacks.

2 *Reviewer centric features:* These are features about each reviewer. Example features include the average rating given by the reviewer, the standard deviation in rating, the ratio of the number of reviews that the reviewer wrote which were the first reviews of the products to the total number of reviews that he/she wrote, and the ratio of the number of cases in which he/she was the only reviewer.

3 *Product centric features:* These are features about each product. Example features include the price of the product, the sales rank of the product (amazon.com assigns sales rank to 'now selling products' according to their sales volumes), the average review rating of the product, and the standard deviation in ratings of the reviews for the product.

Logistic regression was used for model building. Experimental results showed some interesting results.

8.2 Spam Detection Based on Abnormal Behaviors

Due to the difficulty of manually labeling training data, treating opinion spam detection as a supervised learning problem is problematic because many non-duplicated reviews can be spam too. Here, we describe two techniques that try to identify atypical behaviors of reviewers for detecting spammers. For example, if a reviewer wrote all negative reviews for a brand but other reviewers were all positive about the brand, then this reviewer is naturally a spam suspect.

The first technique [56] identifies several unusual reviewer behavior models based on different review patterns that suggest spamming. Each model assigns a numeric spamming behavior score to a reviewer by measuring the extent to which the reviewer practices spamming behavior of the type. All the scores are then combined to produce a final spam score for each reviewer.

The second technique [40] identifies unusual reviewer behavior patterns via unexpected rule discovery. This approach formulates the problem as finding unexpected class association rules [59] from data. Four types of unexpected rules are found based on four unexpectedness definitions. Below, an example behavior is given for each type of unexpectedness definition. Their detailed definitions for these types of unexpectedness are involved [40]. Below, we briefly introduce them by giving an example behavior for each unexpectedness.

- **Confidence Unexpectedness:** Using this measure, we can find reviewers who give all high ratings to products of a brand, but most other reviewers are generally negative about the brand.
- **Support Unexpectedness:** Using this measure, we can find reviewers who write multiple reviews for a single product, while other reviewers only write one review.
- **Attribute Distribution Unexpectedness:** Using this measure, we can find that most positive reviews for a brand of products are from only one reviewer although there are a large number of reviewers who have reviewed the products of the brand.
- **Attribute Unexpectedness:** Using this measure, we can find reviewers who write only positive reviews to one brand, and only negative reviews to another brand.

Experimental results of both papers [40, 56] using amazon.com reviews showed that many spammers can be detected based on their behaviors.

8.3 Group Spam Detection

A group spam detection algorithm was reported in [72]. It finds groups of spammers who work together to promote or demote some products. The method works in two steps:

1. **Frequent pattern mining:** First, it extracts the review data to produce a set of transactions. Each transaction represents a unique product and consists of all the reviewers (their ids) who have reviewed that product. Using all the transactions, it performs frequent pattern mining. The patterns thus give us a set of candidate groups who might have spammed together. The reason for using frequent pattern mining is as follows: If a group of reviewers who only worked together once to promote or to demote a single product, it can be hard to detect based on their collective or group behavior. However, these fake reviewers (especially those who get paid to write) cannot be just writing one review for a single product because they would not make enough money that way. Instead, they work on many products, i.e., write many reviews about many products, which unfortunately also give them away. Frequent pattern mining can be used to find them working together on multiple products.

2 **Rank groups based on a set of group spam indicators:** The groups discovered in step 1 may not all be spammer groups. Many of the reviewers are grouped together in pattern mining simply due to chance. Then, this step first uses a set of indicators to catch different types of unusual group behaviors. These indicators including writing reviews together in a short time window, writing reviews right after the product launch, group content similarity, group rating deviation, etc (see [72] for details). It then ranks the discovered groups from step 1 based on their indicator values using SVM rank (also called Ranking SVM) [42].

9. Utility of Reviews

A related problem that has also been studied in the past few years is the determination of the usefulness, helpfulness or utility of each review [26, 50, 61, 118, 64, 70, 117]. This is a meaningful task as it is desirable to rank reviews based on utilities or qualities when showing reviews to the user, with the most useful reviews first. In fact, many review aggregation sites have been practicing this for years. They obtain the helpfulness or utility score of each review by asking readers to provide helpfulness feedbacks to each review. For example, in amazon.com, the reader can indicate whether he/she finds a review helpful by responding to the question "*Was the review helpful to you?*" just below each review. The feedback results from all those responded are then aggregated and displayed right before each review, e.g., "*15 of 16 people found the following review helpful*". Although most review sites already provide the service, automatically determining the quality of a review is still useful because many reviews have few or no feedbacks. This is especially true for new reviews.

Determining the utility of reviews is usually formulated as a regression problem. The learned model assigns a utility value to each review, which can be used in review ranking. In this area of research, the ground truth data used for both training and testing are usually the user-helpfulness feedback given to each review, which as we discussed above is provided for each review at many review sites. So unlike fake review detection, the training and testing data here is not an issue.

Researchers have used many types of features for model building. Example features include review length, review rating (the number of stars), counts of some specific POS tags, opinion words, tf-idf weighting scores, wh-words, product attribute mentions, comparison with product specifications, comparison with editorial reviews, and many more. Subjectivity classification was also applied in [26]. In [61], Liu et al.

formulated the problem slightly differently. They made it a binary classification problem. Instead of using the original helpfulness feedback as the target or dependent variable, they performed manual annotation based on whether the review evaluates many product aspects or not.

Finally, we should note that review utility regression/classification and review spam detections are different concepts. Not-helpful or low quality reviews are not necessarily fake reviews or spam, and helpful reviews may not be non-spam. A user often determines whether a review is helpful or not based on whether the review expresses opinions on many aspects of the product. A spammer can satisfy this requirement by carefully crafting a review that is just like a normal helpful review. Using the number of helpful feedbacks to define review quality is also problematic because user feedbacks can be spammed too. Feedback spam is a sub-problem of click fraud in search advertising, where a person or robot clicks on some online advertisements to give the impression of real customer clicks. Here, a robot or a human spammer can also click on helpfulness feedback button to increase the helpfulness of a review. Another important point is that a low quality review is still a valid review and should not be discarded, but a spam review is untruthful and/or malicious and should be removed once detected.

10. Conclusions

This chapter introduced and surveyed the field of sentiment analysis and opinion mining. Due to many challenging research problems and a wide variety of practical applications, it has been a very active research area in recent years. In fact, it has spread from computer science to management science. This chapter first presented an abstract model of sentiment analysis, which formulated the problem and provided a common framework to unify different research directions. It then discussed the most widely studied topic of sentiment and subjectivity classification, which determines whether a document or sentence is opinionated, and if so whether it carries a positive or negative opinion. We then described aspect-based sentiment analysis which exploits the full power of the abstract model. After that we briefly introduced the problem of analyzing comparative sentences. Last but not least, we discussed opinion spam, which is increasingly becoming an important issue as more and more people are relying on opinions on the Web for decision making. Several initial algorithms were described. Finally, we conclude the chapter by saying that all the sentiment analysis tasks are very challenging. Our understanding and knowledge of the problem and its solution are still limited. The main reason is that it is a natural language processing task, and natural language processing has no easy problems. However, many significant progresses have been made. This is evident from the large number of start-up companies that offer sentiment analysis or opinion mining services. There is a real and huge need in the industry for such services because every company wants to know how consumers perceive their products and services and those of their competitors. These practical needs and the technical challenges will keep the field vibrant and lively for years to come.

References

[1] Andreevskaia, A. and S. Bergler. Mining WordNet for fuzzy sentiment: Sentiment tag extraction from WordNet glosses. In *Proceedings of Conference of the European Chapter of the Association for Computational Linguistics (EACL-06)*, 2006.

[2] Aue, A. and M. Gamon. Customizing sentiment classifiers to new domains: a case study. In *Proceedings of Recent Advances in Natural Language Processing (RANLP-2005)*, 2005.

[3] Banea, C., R. Mihalcea and J. Wiebe. Multilingual subjectivity: are more languages better?. In *Proceedings of International Conference on Computational Linguistics (COLING-2010)*, 2010.

[4] Beineke, P., T. Hastie, C. Manning, and S. Vaithyanathan. An exploration of sentiment summarization. In *Proceedings of AAAI Spring Symposium on Exploring Attitude and Affect in Text: Theories and Applications,* 2003.

[5] Bethard, S., H. Yu, A. Thornton, V. Hatzivassiloglou, and D. Jurafsky. Automatic extraction of opinion propositions and their holders. In *Proceedings of the AAAI Spring Symposium on Exploring Attitude and Affect in Text*, 2004.

[6] Blei, D., A. Ng, and M. Jordan. Latent dirichlet allocation. The *Journal of Machine Learning Research*, 2003, 3: p. 993-1022.

[7] Blitzer, J., M. Dredze, and F. Pereira. Biographies, bollywood, boomboxes and blenders: Domain adaptation for sentiment classification. In *Proceedings of Annual Meeting of the Association for Computational Linguistics (ACL-2007)*, 2007.

[8] Boyd-Graber, J. and P. Resnik. Holistic sentiment analysis across languages: multilingual supervised latent dirichlet allocation. In *Proceedings of the Human Language Technology Conference and the Conference on Empirical Methods in Natural Language Processing (HLT/EMNLP-2010)*, 2010.

[9] Breck, E., Y. Choi, and C. Cardie. Identifying expressions of opinion in context. In *Proceedings of the International Joint Conference on Artificial Intelligence (IJCAI-2007)*, 2007.

[10] Brody, S. and S. Elhadad. An unsupervised aspect-sentiment model for online reviews. In *Proceedings of The 2010 Annual Conference of the North American Chapter of the ACL*, 2010.

[11] Carenini, G., R. Ng, and A. Pauls. Multi-document summarization of evaluative text. In *Proceedings of the European Chapter of the Association for Computational Linguistics (EACL-2006)*, 2006.

[12] Carenini, G., R. Ng, and E. Zwart. Extracting knowledge from evaluative text. In *Proceedings of Third Intl. Conf. on Knowledge Capture (K-CAP-05)*, 2005.

[13] Choi, Y., C. Cardie, E. Riloff, and S. Patwardhan. Identifying sources of opinions with conditional random fields and extraction patterns. In *Proceedings of the Human Language Technology Conference and the Conference on Empirical Methods in Natural Language Processing (HLT/EMNLP-2005)*, 2005.

[14] Choi, Y. and C. Claire. Adapting a polarity lexicon using integer linear programming for domain-specific sentiment classification. In *Proceedings of the Conference on Empirical Methods in Natural Language Processing (EMNLP-2009)*, 2009.

[15] Dave, K., S. Lawrence, and D. Pennock. Mining the peanut gallery: Opinion extraction and semantic classification of product reviews. In *Proceedings of International Conference on World Wide Web (WWW-2003)*, 2003.

[16] Ding, X. and B. Liu. Resolving object and attribute coreference in opinion mining. In *Proceedings of International Conference on Computational Linguistics (COLING-2010)*, 2010.

[17] Ding, X., B. Liu, and P. Yu. A holistic lexicon-based approach to opinion mining. In *Proceedings of the Conference on Web Search and Web Data Mining (WSDM-2008)*, 2008.

[18] Ding, X., B. Liu, and L. Zhang. Entity discovery and assignment for opinion mining applications. In *Proceedings of ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD-2009)*, 2009.

[19] Esuli, A. and F. Sebastiani. Determining term subjectivity and term orientation for opinion mining. In *Proceedings of Conf. of the European Chapter of the Association for Computational Linguistics (EACL-2006)*, 2006.

[20] Esuli, A. and F. Sebastiani. Determining the semantic orientation of terms through gloss classification. In *Proceedings of ACM International Conference on Information and Knowledge Management (CIKM-2005)*, 2005.

[21] Esuli, A. and F. Sebastiani. SentiWordNet: a publicly available lexical resource for opinion mining. In *Proceedings of Language Resources and Evaluation (LREC-2006)*, 2006.

[22] Freitag, D. and A. McCallum. Information extraction with HMM structures learned by stochastic optimization. In *Proceedings of National Conf. on Artificial Intelligence (AAAI-2000)*, 2000.

[23] Gamon, M., A. Aue, S. Corston-Oliver, and E. Ringger. Pulse: Mining customer opinions from free text. *Advances in Intelligent Data Analysis VI*, 2005: p. 121-132.

[24] Ganapathibhotla, M. and B. Liu. Mining opinions in comparative sentences. In *Proceedings of International Conference on Computational Linguistics (COLING-2008)*, 2008.

[25] Ghahramani, Z. and K. Heller. Bayesian sets. *Advances in Neural Information Processing Systems (NIPS-2005)*, 2005.

[26] Ghose, A. and P. Ipeirotis. Designing novel review ranking systems: predicting the usefulness and impact of reviews. In *Proceedings of the International Conference on Electronic Commerce*, 2007.

[27] Guo, H., H. Zhu., X. Zhang., and Z. Su. OpinionIt: a text mining system for cross-lingual opinion analysis. In *Proceedings of ACM International Conference on Information and knowledge management (CIKM-2010)*, 2010.

[28] Hassan, A., V. Qazvinian., D. Radev. What's with the attitude? identifying sentences with attitude in online discussion. In *Proceedings of the Conference on Empirical Methods in Natural Language Processing (EMNLP-2010)*, 2010.

[29] Hatzivassiloglou, V. and K. McKeown. Predicting the semantic orientation of adjectives. In *Proceedings of Annual Meeting of the Association for Computational Linguistics (ACL-1997)*, 1997.

[30] Hatzivassiloglou, V. and J. Wiebe. Effects of adjective orientation and gradability on sentence subjectivity. In *Proceedings of International Conference on Computational Linguistics (COLING-2000)*, 2000.

[31] Hu, M. and B. Liu. Mining and summarizing customer reviews. In *Proceedings of ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD-2004)*, 2004.

[32] Huang, X. and W. B. Croft. A unified relevance model for opinion retrieval. In *Proceedings of ACM International Conference on Information and knowledge management (CIKM-2009)*, 2009.

[33] Jakob, N. and I. Gurevych. Extracting opinion targets in a single- and cross-domain setting with conditional random fields. In *Proceedings of Conference on Empirical Methods in Natural Language Processing (EMNLP-2010)*, 2010.

[34] Jin, W. and H. Ho. A novel lexicalized HMM-based learning framework for web opinion mining. In *Proceedings of International Conference on Machine Learning (ICML-2009)*, 2009.

[35] Jin, W. and H. Ho. OpinionMiner: a novel machine learning system for web opinion mining and extraction. In *Proceedings of ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD-2009)*, 2009.

[36] Jindal, N. and B. Liu. Identifying comparative sentences in text documents. In *Proceedings of ACM SIGIR Conf. on Research and Development in Information Retrieval (SIGIR-2006)*, 2006.

[37] Jindal, N. and B. Liu. Mining comparative sentences and relations. In *Proceedings of National Conf. on Artificial Intelligence (AAAI-2006)*, 2006.

[38] Jindal, N. and B. Liu. Opinion spam and analysis. In *Proceedings of the Conference on Web Search and Web Data Mining (WSDM-2008)*, 2008.

[39] Jindal, N. and B. Liu. Review spam detection. In *Proceedings of International Conference on World Wide Web (WWW-2007)*, 2007.

[40] Jindal, N., B. Liu, and E. Lim. Finding unusual review patterns using unexpected rules. In *Proceedings of ACM International Conference on Information and Knowledge Management (CIKM-2010)*, 2010.

[41] Jo, Y. and A. Oh. Aspect and sentiment unification model for online review analysis. In *Proceedings of the Conference on Web Search and Web Data Mining (WSDM-2011)*, 2011.

[42] Joachims, T. Optimizing search engines using clickthrough data. In *Proceedings of the ACM Conference on Knowledge Discovery and Data Mining (KDD-2002)*, 2002.

[43] Kaji, N. and M. Kitsuregawa. Automatic construction of polarity-tagged corpus from HTML documents. In *Proceedings of COLING/ACL 2006 Main Conference Poster Sessions (COLING-ACL-2006)*, 2006.

[44] Kaji, N. and M. Kitsuregawa. Building lexicon for sentiment analysis from massive collection of HTML documents. In *Proceedings of the Joint Conference on Empirical Methods in Natural Language Processing and Computational Natural Language Learning (EMNLP-2007)*, 2007.

[45] Kamps, J., M. Marx, R. Mokken, and M. De Rijke. Using WordNet to measure semantic orientation of adjectives. In *Proc. of LREC-2004*, 2004.

[46] Kanayama, H. and T. Nasukawa. Fully automatic lexicon expansion for domain-oriented sentiment analysis. In *Proceedings of Conference on Empirical Methods in Natural Language Processing (EMNLP-2006)*, 2006.

[47] Kim, J., J. Li., and J. Lee. Evaluating multilanguage-comparability of subjective analysis system. In P*roceedings of Annual Meeting of the Association for Computational Linguistics (ACL-2010)*, 2010.

[48] Kim, S. and E. Hovy. Crystal: analyzing predictive opinions on the web. In Proceedings of the Joint Conference on Empirical Methods in Natural Language Processing and Computational Natural Language Learning (EMNLP/CoNLL-2007), 2007.

[49] Kim, S. and E. Hovy. Determining the sentiment of opinions. In *Proceedings of Interntional Conference on Computational Linguistics (COLING-2004)*, 2004.

[50] Kim, S., P. Pantel, T. Chklovski, and M. Pennacchiotti. Automatically assessing review helpfulness. In *Proceedings of the Conference on Empirical Methods in Natural Language Processing (EMNLP-2006)*, 2006.

[51] Ku, L., Y. Liang, and H. Chen. Opinion extraction, summarization and tracking in news and blog corpora. In *Proceedings of AAAI-CAAW-2006*, 2006.

[52] Lafferty, J., A. McCallum, and F. Pereira. Conditional random fields: probabilistic models for segmenting and labeling sequence data. In *Proceedings of International Conference on Machine Learning (ICML-2001)*, 2001.

[53] Lee, L. Measures of distributional similarity. In *Proceedings of Annual Meeting of the Association for Computational Linguistics (ACL-1999)*, 1999.

[54] Li, S., C. Lin, Y. Song, and Z. Li. Comparable entity mining from comparative questions. In *Proceedings of Annual Meeting of the Association for Computational Linguistics (ACL-2010)*, 2010.

[55] Li, X., L. Zhang, B. Liu, and S. Ng. Distributional similarity vs. PU learning for entity set expansion. In *Proceedings of Annual Meeting of the Association for Computational Linguistics (ACL-2010)*, 2010.

[56] Lim, E., V. Nguyen, N. Jindal, B. Liu, and H. Lauw. Detecting product review spammers using rating behaviors. In *Proceedings of ACM International Conference on Information and Knowledge Management (CIKM-2010)*, 2010.

[57] Lin, C. and Y. He. Joint sentiment/topic model for sentiment analysis. In *Proceedings of ACM International Conference on Information and Knowledge Management (CIKM-2009)*, 2009.

[58] Liu, B. Sentiment analysis and subjectivity. In Handbook of Natural Language Processing, Second Edition, N. Indurkhya and F.J. Damerau, Editors. 2010.

[59] Liu, B. Web Data Mining: Exploring Hyperlinks, Contents, and Usage Data. Second Edition, Springer, 2011.

[60] Liu, B., M. Hu, and J. Cheng. Opinion observer: analyzing and comparing opinions on the web. In *Proceedings of International Conference on World Wide Web (WWW-2005)*, 2005.

[61] Liu, J., Y. Cao, C. Lin, Y. Huang, and M. Zhou. Low-quality product review detection in opinion summarization. In *Proceedings of the Joint Conference on Empirical Methods in Natural Language Processing and Computational Natural Language Learning (EMNLP-CoNLL-2007)*, 2007.

[62] Liu, Y., X. Huang, A. An, and X. Yu. ARSA: a sentiment-aware model for predicting sales performance using blogs. In *Proceedings of ACM SIGIR Conf. on Research and Development in Information Retrieval (SIGIR-2007)*, 2007.

[63] Lu, Y., M. Castellanos, U. Dayal, and C. Zhai. Automatic construction of a context-aware sentiment lexicon: an optimization approach. In *Proceedings of International Conference on World Wide Web (WWW-2011)*, 2011.

[64] Lu, Y., P. Tsaparas., A. Ntoulas., and L. Polanyi. Exploiting social context for review quality prediction. In *Proceedings of International Conference on World Wide Web (WWW-2010)*, 2010.

[65] Lu, Y. and C. Zhai. Opinion integration through semi-supervised topic modeling. In *Proceedings of International Conference on World Wide Web (WWW-2008)*, 2008.

[66] Lu, Y., C. Zhai, and N. Sundaresan. Rated aspect summarization of short comments. In *Proceedings of International Conference on World Wide Web (WWW-2009)*, 2009.

[67] Mei, Q., X. Ling, M. Wondra, H. Su, and C. Zhai. Topic sentiment mixture: modeling facets and opinions in weblogs. In *Proceedings of International Conference on World Wide Web (WWW-2007)*, 2007.

[68] Melville, P., W. Gryc., R. D. Lawrence. Sentiment analysis of blogs by combining lexical knowledge with text classification. In *Proceedings of ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD-2009)*, 2009.

[69] Miller, G., R. Beckwith, C. Fellbaum, D. Gross, and K. Miller. WordNet: an on-line lexical database. *Oxford Univ. Press.*, 1990.

[70] Mishne, G. and N. Glance. Predicting movie sales from blogger sentiment. In *Proceedings of AAAI Symposium on Computational Approaches to Analysing Weblogs (AAAI-CAAW-2006)*, 2006.

[71] Mohammad, S., C. Dunne., and B. Dorr. Generating high-coverage semantic orientation lexicons from overly marked words and a thesaurus. In *Proceedings of the Conference on Empirical Methods in Natural Language Processing (EMNLP-2009)*, 2009.

[72] Mukherjee, A., B. Liu, J. Wang, N. Glance, and N. Jindal. Detecting group review spam. In *Proceedings of International Conference on World Wide Web (WWW-2011)*, 2011.

[73] Nishikawa, H., T. Hasegawa, Y. Matsuo, and G. Kikui. Optimizing informativeness and readability for sentiment summarization. In *Proceedings of Annual Meeting of the Association for Computational Linguistics (ACL-2010)*, 2010.

[74] Paltoglou, G. and M. Thelwall. A study of information retrieval weighting schemes for sentiment analysis. In *Proceedings of Annual Meeting of the Association for Computational Linguistics (ACL-2010)*, 2010.

[75] Pan, S., X. Ni, J. Sun, Q. Yang, and Z. Chen. Cross-domain sentiment classification via spectral feature alignment. In *Proceedings of International Conference on World Wide Web (WWW-2010)*, 2010.

[76] Pang, B. and L. Lee. Opinion mining and sentiment analysis. *Foundations and Trends in Information Retrieval*, 2(1-2): p. 1-135, 2008.

[77] Pang, B. and L. Lee. Seeing stars: Exploiting class relationships for sentiment categorization with respect to rating scales. In *Proceedings of Meeting of the Association for Computational Linguistics (ACL-2005)*, 2005.

[78] Pang, B., L. Lee, and S. Vaithyanathan. Thumbs up?: sentiment classification using machine learning techniques. In *Proceedings of Conference on Empirical Methods in Natural Language Processing (EMNLP-2002)*, 2002.

[79] Pantel, P., E. Crestan, A. Borkovsky, A. Popescu, and V. Vyas. Web-scale distributional similarity and entity set expansion. In *Proceedings of Conference on Empirical Methods in Natural Language Processing (EMNLP-2009)*, 2009.

[80] Parrott, W. Emotions in social psychology: Essential readings. *Psychology Press*, 2001.

[81] Paul, M., C. Zhai, and R. Girju. Summarizing contrastive viewpoints in opinionated text. In *Proceedings of Conference on Empirical Methods in Natural Language Processing (EMNLP-2010)*, 2010.

[82] Polanyi, L. and A. Zaenen. Contextual valence shifters. In *Proceedings of the AAAI Spring Symposium on Exploring Attitude and Affect in Text*, 2004.

[83] Popescu, A. and O. Etzioni. Extracting product features and opinions from reviews. In *Proceedings of Conference on Empirical Methods in Natural Language Processing (EMNLP-2005)*, 2005.

[84] Qiu, G., B. Liu, J. Bu, and C. Chen. Expanding domain sentiment lexicon through double propagation. In *Proceedings of International Joint Conference on Artificial Intelligence (IJCAI-2009)*, 2009.

[85] Qiu, G., B. Liu, J. Bu, and C. Chen. Opinion word expansion and target extraction through double propagation. *Computational Linguistics*, 2011.

[86] Qiu, L., W. Zhang., C. Hu., and K. Zhao. SELC: A self-supervised model for sentiment classification. In *Proceedings of ACM International Conference on Information and knowledge management (CIKM-2009)*, 2009.

[87] Riloff, E., S. Patwardhan, and J. Wiebe. Feature subsumption for opinion analysis. In *Proceedings of the Conference on Empirical Methods in Natural Language Processing (EMNLP-2006)*, 2006.

[88] Riloff, E. and J. Wiebe. Learning extraction patterns for subjective expressions. In *Proceedings of Conference on Empirical Methods in Natural Language Processing (EMNLP-2003)*, 2003.

[89] Seki, Y., K. Eguchi, N. Kando, and M. Aono. Opinion-focused summarization and its analysis at DUC 2006. In *Proceedings of the Document Understanding Conference (DUC)*, 2006.

[90] Stone, P. The general inquirer: a computer approach to content analysis. *Journal of Regional Science*, 8(1), 1968.

[91] Stoyanov, V. and C. Cardie. Partially supervised coreference resolution for opinion summarization through structured rule learning. In *Proceedings of Conference on Empirical Methods in Natural Language Processing (EMNLP-2006)*, 2006.

[92] Su, Q., X. Xu, H. Guo, Z. Guo, X. Wu, X. Zhang, B. Swen, and Z. Su. Hidden sentiment association in chinese web opinion mining. In *Proceedings of International Conference on World Wide Web (WWW-2008)*, 2008.

[93] Taboada, M., J, Brooke, M. Tofiloski, K. Voll, and M. Stede, Lexicon-based methods for sentiment analysis. *Computational Intelligence*, 2010.

[94] Takamura, H., T. Inui, and M. Okumura. Extracting semantic orientations of phrases from dictionary. In *Proceedings of the Joint Human Language Technology/North American Chapter of the ACL Conference (HLT-NAACL-2007)*, 2007.

[95] Tan, S., Y. Wang., and X. Cheng. Combining learn-based and lexicon-based techniques for sentiment detection without using labeled examples. In *Proceedings of ACM SIGIR Conference on Research and Development in Information Retrieval (SIGIR-2008)*, 2008.

[96] Tata, S. and B. Di Eugenio. Generating fine-grained reviews of songs from album reviews. In *Proceedings of Annual Meeting of the Association for Computational Linguistics (ACL-2010)*, 2010.

[97] Tesniere, L. Elements de syntaxe structurale: Prf. de Jean Fourquet. 1959: C. Klincksieck.

[98] Titov, I. and R. McDonald. A joint model of text and aspect ratings for sentiment summarization. In *Proceedings of Annual Meeting of the Association for Computational Linguistics (ACL-2008)*, 2008.

[99] Titov, I. and R. McDonald. Modeling online reviews with multi-grain topic models. In *Proceedings of International Conference on World Wide Web (WWW-2008)*, 2008.

[100] Turney, P. Thumbs up or thumbs down?: semantic orientation applied to unsupervised classification of reviews. In *Proceedings of Annual Meeting of the Association for Computational Linguistics (ACL-2002)*, 2002.

[101] Wan, X. Co-training for cross-lingual sentiment classification. In *Proceedings of Annual Meeting of the Association for Computational Linguistics (ACL-2009)*, 2009.

[102] Wang, H., Y. Lu, and C. Zhai. Latent aspect rating analysis on review text data: a rating regression approach. In *Proceedings of ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD-2010)*, 2010.

[103] Wiebe, J. Learning subjective adjectives from corpora. In *Proceedings of National Conf. on Artificial Intelligence (AAAI-2000)*, 2000.

[104] Wiebe, J., R. Bruce, and T. O'Hara. Development and use of a gold-standard data set for subjectivity classifications. In *Proceedings of the Association for Computational Linguistics (ACL-1999)*, 1999.

[105] Wiebe, J. and E. Riloff. Creating subjective and objective sentence classifiers from unannotated texts. *Computational Linguistics and Intelligent Text Processing*, p. 486-497, 2005.

[106] Wiebe, J., T. Wilson, R. Bruce, M. Bell, and M. Martin. Learning subjective language. *Computational Linguistics*, 30(3): p. 277-308, 2004.

[107] Wiebe, J., T. Wilson, and C. Cardie. Annotating expressions of opinions and emotions in language. *Language Resources and Evaluation*, 39(2): p. 165-210, 2005.

[108] Wilson, T., J. Wiebe, and P. Hoffmann. Recognizing contextual polarity in phrase-level sentiment analysis. In *Proceedings of the Human Language Technology Conference and the Conference on Empirical Methods in Natural Language Processing (HLT/EMNLP-2005)*, 2005.

[109] Wilson, T., J. Wiebe, and R. Hwa. Just how mad are you? finding strong and weak opinion clauses. In *Proceedings of National Conference on Artificial Intelligence (AAAI-2004)*, 2004.

[110] Wilson, T., J. Wiebe, and R. Hwa. Recognizing strong and weak opinion clauses. *Computational Intelligence*, 22(2): p. 73-99, 2006.

[111] Wu, Y., Q. Zhang, X. Huang, and L. Wu. Phrase dependency parsing for opinion mining. In *Proceedings of Conference on Empirical Methods in Natural Language Processing (EMNLP-2009)*, 2009.

[112] Yang, H., L. Si, and J. Callan. Knowledge transfer and opinion detection in the TREC2006 blog track. In *Proceedings of TREC*, 2006.

[113] Yu, H. and V. Hatzivassiloglou. Towards answering opinion questions: separating facts from opinions and identifying the polarity of opinion sentences. In *Proceedings of Conference on Empirical Methods in Natural Language Processing (EMNLP-2003)*, 2003.

[114] Zhai, Z., B. Liu, H. Xu, and P. Jia. Grouping product features using semi-supervised learning with soft-constraints. In *Proceedings of International Conference on Computational Linguistics (COLING-2010)*, 2010.

[115] Zhai, Z., B. Liu., L. Zhang., H. Xu., and P. Jia. Identifying evaluative sentences in online discussions. In *Proceedings of National Conf. on Artificial Intelligence (AAAI-2011)*, 2011.

[116] Zhang, L. and B. Liu. Identifying noun product features that imply opinions. In *Proceedings of Annual Meeting of the Association for Computational Linguistics (ACL-2011)*, 2011.

[117] Zhang, M. and X. Ye. A generation model to unify topic relevance and lexicon-based sentiment for opinion retrieval. In *Proceedings of ACM SIGIR Conference on Research and Development in Information Retrieval (SIGIR-2008)*, 2008.

[118] Zhang, Z. and B. Varadarajan. Utility scoring of product reviews. In *Proceedings of ACM International Conference on Information and Knowledge Management (CIKM-2006),* 2006.

[119] Zhao, W., J. Jiang, H. Yan, and X. Li. Jointly modeling aspects and opinions with a MaxEnt-LDA hybrid In *Proceedings of Conference on Empirical Methods in Natural Language Processing (EMNLP-2010) ,* 2010.

[120] Zhuang, L., F. Jing, and X. Zhu. Movie review mining and summarization. In *Proceedings of ACM International Conference on Information and Knowledge Management (CIKM-2006),* 2006.