Chapter 2
Many Facets of Sentiment Analysis

Bing Liu

Abstract Sentiment analysis or opinion mining is the computational study of people's opinions, sentiments, evaluations, attitudes, moods, and emotions. It is one of the most active research areas in natural language processing, data mining, information retrieval, and Web mining. In recent years, its research and applications have also spread to management sciences and social sciences due to its importance to business and society as a whole. This chapter defines the sentiment analysis problem and its related concepts such as sentiment, opinion, emotion, mood, and affect. The goal is to abstract a structure from the complex unstructured natural language text related to the problem and its pertinent concepts. The definitions not only enable us to see a rich set of inter-related sub-problems, but also a common framework that can unify existing research directions. They also help researchers design more robust solution techniques by exploiting the inter-relationships of the sub-problems.

Keywords Sentiment analysis • Opinion mining • Emotion • Mood • Affect • Subjectivity

Many people thought that sentiment analysis is just the problem of classifying whether a document or a sentence expresses a positive or negative sentiment or opinion. It is in fact a much more complex problem than that. It involves many facets and multiple sub-problems. In this chapter, I define an abstraction of the sentiment analysis problem. The definitions will enable us to see a rich set of inter-related sub-problems. It is often said that if we cannot structure a problem, we probably do not understand the problem. The objective of the definitions is to abstract a structure from the complex unstructured natural language text. The structure serves as a common framework to unify existing research directions and enable researchers to design more robust solution techniques by exploiting the inter-relationships of the sub-problems.

B. Liu (✉)
Department of Computer Science, University of Illinois at Chicago, Chicago, IL, USA
e-mail: liub@cs.uic.edu

E. Cambria et al. (eds.), *A Practical Guide to Sentiment Analysis*,
Socio-Affective Computing 5, DOI 10.1007/978-3-319-55394-8_2

Unlike factual information, sentiment and opinion have an important characteristic, namely, being subjective. The subjectivity comes from many sources. First of all, different people may have different experiences and thus different opinions. Different people may also have different interests and/or different ideologies. Due to such different subjective experiences, views, interests and ideologies, it is important to examine a collection of opinions from many people rather than only one opinion from a single person because such an opinion represents only the subjective view of a single person, which is usually not sufficient for action. With a large number of opinions, some form of summary becomes necessary (Hu and Liu 2004). Thus, the problem definitions should also state what kind of summary may be desired. Along with the problem definitions, the chapter also discusses different types of opinions and the important concepts of affect, emotion and mood.

Throughout this chapter, I mainly use product reviews and sentences from such reviews as examples to introduce the key concepts, but the ideas and the resulting definitions are general and applicable to all forms of formal and informal opinion text such as news articles, tweets (Twitter posts), forum discussions, blogs, and Facebook posts, and all kinds of domains including social and political domains. The content of this chapter is mainly taken from my book "Sentiment Analysis: Mining Opinions, Sentiments, and Emotions" (Liu 2015).

2.1 Definition of Opinion

Sentiment analysis mainly studies opinions that express or imply positive or negative sentiment. We define the problem in this context. We use the term *opinion* as a broad concept that covers sentiment, evaluation, appraisal, or attitude, and its associated information such as opinion target and the person who holds the opinion, and use the term *sentiment* to mean only the underlying positive or negative feeling implied by opinion. Due to the need to analyze a large volume of opinions, in defining opinion we consider two levels of abstraction: *a single opinion* and *a set of opinions*. In this section, we focus on defining a single opinion and describing the tasks involved in extracting an opinion. Section 2.2 focuses on a set of opinions, where we define *opinion summary*.

2.1.1 Opinion Definition

We use the following review (Review A) about a camera to introduce the problem (an id number is associated with each sentence for easy reference):

Review A

Posted by John Smith
Date: September 10, 2011

> (1) *I bought a Canon G12 camera six months ago*. (2) *I simply love it*. (3) *The picture quality is amazing*. (4) *The battery life is also long*. (5) *However, my wife thinks it is too heavy for her*.

From this review, we can make the following important observation:

Opinion, sentiment and target: Review A has several opinions with positive or negative sentiment about the Canon G12 camera. Sentence (2) expresses a positive sentiment about the Canon camera as a whole. Sentence (3) expresses a positive sentiment about its picture equality. Sentence (4) expresses a positive sentiment about its battery life. Sentence (5) expresses a negative sentiment about the camera's weight.

From these opinions, we can make a crucial observation about sentiment analysis. That is, an opinion has two key components: a *target* g and a *sentiment* s on the target, i.e., (g, s), where g can be any entity or aspect of the entity on which an opinion has been expressed, and s can be a positive, negative, or neutral sentiment, or a numeric rating. *Positive*, *negative* and *neutral* are called *sentiment* or *opinion orientations*. For example, the target of the opinion in sentence (2) is the *Canon G12 camera*, the target of the opinion in sentence (3) is the *picture quality of Canon G12*, and the target of sentence (5) is the *weight of Canon G12* (*weight* is indicated by *heavy*). Target is also called *topic* by some researchers.

Opinion holder: Review A contains opinions from two persons, who are called *opinion sources* or *opinion holders* (Kim and Hovy 2004; Wiebe et al. 2005). *The* holder of the opinions in sentences (2), (3), and (4) is the author of the review ("John Smith"), but for sentence (5), it is the wife of the author.

Time of opinion: The date of the review was September 10, 2011. This date is useful because one often wants to know how opinions change over time or the opinion trend.

With this example, we can define opinion as a quadruple.

Definition 1 (Opinion) An *opinion* is a quadruple,

$$(g, s, h, t),$$

where g is the *sentiment target*, s is the *sentiment* of the opinion about the target g, h is the *opinion holder* (the person or organization who holds the opinion), and t is the *time* when the opinion is expressed.

The four components here are essential. It is generally problematic if any of them is missing. For example, the time component is important in practice because

an opinion two years ago is not the same as an opinion today. Not having an opinion holder is also problematic. For example, an opinion from a very important person (e.g., the US President) is probably more important than that from the average Joe on the street.

One thing that we want to stress about the definition is that *opinion has target*. Recognizing this is important for two reasons: First, in a sentence with multiple targets, we need to identify the specific target for each positive or negative sentiment. For example, "*Apple is doing very well in this poor economy*" has a positive sentiment and a negative sentiment. The target for the positive sentiment is *Apple* and the target for the negative sentiment is *economy*. Second, words or phrases such as *good*, *amazing*, *bad* and *poor* that express sentiments (called *sentiment* or *opinion terms*) and opinion targets often have some syntactic relations (Hu and Liu 2004; Qiu et al. 2011; Zhuang et al. 2006), which allow us to design algorithms to extract both sentiment terms and opinion targets, which are two core tasks of sentiment analysis (see Sect. 2.1.6).

The opinion defined here is just one type of opinion, called a *regular opinion* (e.g., "*Coke taste great*"). Another type is *comparative opinion* (e.g., "*Coke tastes better than Pepsi*") which needs a different definition (Jindal and Liu 2006b; Liu 2006). Section 2.1.4 will further discuss different types of opinions. For the rest of this section, we focus on only regular opinions, which, for simplicity, we will just call opinions.

2.1.2 Sentiment Target

Definition 2 (Sentiment Target) The *sentiment target*, also known as the *opinion target*, of an opinion is the entity or a part or attribute of the entity that the sentiment has been expressed upon.

For example, in sentence (3) of Review A, the target is the *picture quality of Canon G12*, although the sentence mentioned only the *picture quality*. The target is not just the *picture quality* because without knowing that the picture quality belongs to the Canon G12 camera, the opinion in the sentence is of little use.

An entity can be decomposed and represented hierarchically (Liu 2006).

Definition 3 (Entity) An *entity e* is a product, service, topic, person, organization, issue or event. It is described with a pair, *e*: (*T*, *W*), where *T* is a hierarchy of *parts*, *sub-parts*, and so on, and *W* is a set of *attributes* of *e*. Each part or sub-part also has its own set of attributes.

For example, a particular camera model is an entity, e.g., Canon G12. It has a set of attributes, e.g., *picture quality*, *size*, and *weight*, and a set of parts, e.g., *lens*, *viewfinder*, and *battery*. *Battery* also has its own set of attributes, e.g., *battery life* and *battery weight*. A topic can be an entity too, e.g., *tax increase*, with its sub-topics or parts '*tax increase for the poor*,' '*tax increase for the middle class*' and '*tax increase for the rich*.'

This definition describes an entity hierarchy based on the *part-of* relation. The root node is the name of the entity, e.g., Canon G12 Review A. All the other nodes are parts and sub-parts, etc. An opinion can be expressed on any node and any attribute of the node. For instance, in Review A, sentence (2) expresses a positive opinion about the entity Canon G12 as a whole, and sentence (3) expresses a positive opinion about the picture quality attribute of the camera. Clearly, we can also express opinions about any part or component of the camera.

In the research literature, entities are also called *objects*, and attributes are also called *features* (as in product features) (Hu and Liu 2004; Liu 2010). The terms *object* and *feature* are not used in this Chapter because object can be confused with the term *object* used in grammar and feature can be confused with *feature* used in machine learning as data attribute. In recent years, the term *aspect* has become popular, which covers both *part* and *attribute* (see Sect. 2.1.4).

Entities may be called other names in specific application domains. For example, in politics, entities are usually *political candidates*, *issues*, and *events*. There is no term that is perfect for all application domains. The term *entity* is chosen because most current applications of sentiment analysis study opinions about various forms of named entities, e.g., products, services, brands, organizations, events, and people.

2.1.3 Sentiment and Its Intensity

Definition 4 (Sentiment) *Sentiment* is the underlying feeling, attitude, evaluation, or emotion associated with an opinion. It is represented as a triple,

$$(y, o, i),$$

where y is the *type* of the sentiment, o is the *orientation* of the sentiment, and i is the *intensity* of the sentiment.

Sentiment type: Sentiment can be classified into several types. There are linguistic-based, psychology-based, and consumer research-based classifications. Here I choose to use a consumer research-based classification as it is simple and easy to use in practice. Consumer research classifies sentiment broadly into two categories: *rational sentiment* and *emotional sentiment* (Chaudhuri 2006).

Definition 5 (Rational Sentiment) *Rational sentiments* are from rational reasoning, tangible beliefs, and utilitarian attitudes. They express no emotions.

We also call opinions expressing rational sentiment the *rational opinions*. The opinions in the following sentences imply rational sentiment: "*The voice of this phone is clear*," and "*This car is worth the price.*"

Definition 6 (Emotional Sentiment) *Emotional sentiments* are from non-tangible and emotional responses to entities which go deep into people's psychological state of mind.

We also call opinions expressing emotional sentiment the *emotional opinions*. The opinions in the following sentences imply emotional sentiment: "*I love iPhone*," "*I am so angry with their service people*," "*This is the best car ever*" and "*After our team won, I cried*."

Emotional sentiment is stronger than rational sentiment, and is usually more important in practice. For example, in marketing, to guarantee the success of a new product in the market, the positive sentiment from a large population of consumers has to reach the emotional level. Rational positive may not be sufficient.

Each of these broad categories can be further divided into smaller categories. For example, there are many types of emotions, e.g., *anger*, *joy, fear*, and *sadness.* We will discuss some possible sub-divisions of rational sentiment in Sect. 2.4.2 and different emotions in Sect. 2.3. In applications, the user is also free to design their own sub-categories.

Sentiment orientation: It can be *positive*, *negative*, or *neutral*. Neutral usually means the absence of sentiment or no sentiment or opinion. Sentiment orientation is also called *polarity*, *semantic orientation*, or *valence* in the research literature.

Sentiment intensity: Sentiment can have different levels of strength or intensity. People often use two ways to express intensity of their feelings in text. The first is to choose sentiment terms (words or phrases) with suitable strengths. For example, *good* is weaker than *excellent*, and *dislike* is weaker than *detest*. *Sentiment words* are words in a language that are often used to express positive or negative sentiments. For example, *good*, *wonderful*, and *amazing* are positive sentiment words, and *bad*, *poor*, and *terrible* are negative sentiment words. The second is to use *intensifiers* and *diminishers*, which are terms that change the degree of the expressed sentiment. An intensifier increases the intensity of a positive/negative term, while a diminisher decreases the intensity of that term. Common English intensifiers include *very*, *so*, *extremely*, *dreadfully*, *really*, *awfully*, *terribly*, etc., and common English diminishers include *slightly*, *pretty*, *a little bit*, *a bit*, *somewhat*, *barely*, etc.

Sentiment rating: In applications, we commonly use some discrete ratings to express sentiment intensity. Five levels (e.g., 1–5 stars) are commonly employed, which can be interpreted as follows based on the two types of sentiment in Definitions 5 and 6:

- *emotional positive* (+2 or 5 stars)
- *rational positive* (+1 or 4 stars)
- *neutral* (0 or 3 stars)
- *rational negative* (−1 or 2 stars)
- *emotional negative* (−2 or 1 star)

Clearly, it is possible to have more rating levels, which, however, become difficult to differentiate based on the natural language text alone due to the highly subjective nature and the fact that people's spoken or written expressions may not fully match with their psychological states of mind. For example, the sentence "*This is an excellent phone*" expresses a rational evaluation of the phone, while "*I love this*

phone" expresses an emotional evaluation about the phone. However, whether they represent completely different psychology states of mind of the authors is hard to say. In practice, the above five levels are sufficient for most applications. If these five levels are not enough in some applications, I suggest dividing *emotional positive* (and, respectively, *emotional negative*) into two levels. Such applications are likely to involve sentiment about personal, social or political events or issues, for which people can be highly emotional.

2.1.4 Opinion Definition Simplified

Opinion as defined in Definition 1, although concise, may not be easy to use in practice especially in the domain of online reviews of products, services, and brands. Let us first look at the sentiment (or opinion) target. The central concept here is *entity*, which is represented as a hierarchy with an arbitrary number of levels. This can be too complex for practical applications because NLP is a very difficult task. Recognizing parts and attributes of an entity at different levels of details is extremely hard. Most applications also do not need such a complex analysis. Thus, we simplify the hierarchy to two levels and use the term *aspect* to denote both *part* and *attribute*. In the simplified tree, the root node is still the entity itself and the second level (also the leaf level) nodes are different aspects of the entity.

The definition of sentiment in Definition 4 can be simplified too. In many applications, positive (denoted by $+1$), negative (denoted by -1) and neutral (denoted by 0) orientations alone are already enough. In almost all applications, 5 levels of ratings are sufficient, e.g., 1–5 stars. In both cases, sentiment can be represented with a single value. The other two components in the triple can be folded into this value.

This simplified framework is what is typically used in practical sentiment analysis systems. We now redefine the concept of opinion (Hu and Liu 2004; Liu 2010).

Definition 7 (Opinion) An *opinion* is a quintuple,

$$(e, a, s, h, t),$$

where e is the target entity, a is the target aspect of entity e on which the opinion has been expressed, s is the sentiment of the opinion on aspect a of entity e, h is the opinion holder, and t is the opinion posting time. s can be *positive*, *negative*, or *neutral*, or a *rating* (e.g., 1–5 stars). When an opinion is only on the entity as a whole, the special aspect GENERAL is used to denote it. Here, e and a together represent the opinion target.

Sentiment analysis (or opinion mining) based on this definition is often called *aspect-based sentiment analysis,* or *feature-based sentiment analysis* as it was called earlier in (Hu and Liu 2004; Liu 2010).

We should note that due to the simplification, the quintuple representation of opinion may result in information loss. For example, *ink* is a part of *printer*. A printer review might say "*The ink of this printer is expensive*." This sentence does not say that the printer is expensive (*expensive* here indicates the aspect *price*). If one does not care about any attribute of the ink, this sentence just gives a negative opinion about the ink (which is an aspect of the printer entity). This results in information loss. However, if one also wants to study opinions about different aspects of the ink, then the ink needs to be treated as a separate entity. The quintuple representation still applies, but an extra mechanism will be required to record the part-of relationship between ink and printer. Of course, conceptually we can also extend the flat quintuple relation to a *nested relation* to make it more expressive. However, as we explained above, too complex a definition can make the problem extremely difficult to solve in practice. Despite this limitation, Definition 4 does cover the essential information of an opinion sufficiently for most applications.

In some applications, it may not be easy to distinguish entity and aspect or there is no need to distinguish them. Such cases often occur when people discuss political or social issues, e.g., "*I hate property tax increases*." We may deal with them in two ways. First, since the author regards '*property tax increase*' as a general issue and it thus does not belong to any specific entity. We can treat it as an entity with the aspect GENERAL. Second, we can regard '*property tax*' as an entity and '*property tax increases*' as one of its aspects to form a hierarchical relationship. Whether treating an issue/topic as an aspect or an entity can also depend on the specific context.

For example, in commenting about a local government, one says "*I hate the proposed property tax increase*." Since it is the local government that imposes and levies property taxes, the specific local government may be regarded as an entity and '*the proposed property tax increase*' as one of its aspects.

Not all applications need all five components of an opinion. In some applications, the user may not need the aspect information. For example, in brand management, the user typically is interested in only opinions about product brands (entities). This is sometimes called *entity-based sentiment analysis*. In some other applications, the user may not need to know the opinion holder or time of opinion. Then these components can be ignored.

2.1.5 Reason and Qualifier for Opinion

We can in fact perform an even finer-grained analysis of opinions. Let us use the sentence "*This car is too small for a tall person*" to explain. It expresses a negative sentiment about the *size* aspect of the car. However, only reporting the negative sentiment for size does not tell the whole story because it can mean *too small* or *too big*. In the above sentence, we call "*too small*" the *reason* for the negative sentiment about size. Furthermore, the sentence does not say that the car is too small for everyone, but only *for a tall person*. We call "*for a tall person*" the *qualifier* of the opinion. We now define these concepts.

Definition 8 (Reason for Opinion) A reason for an opinion is the cause of the opinion.

In practical applications, discovering the reasons for each positive or negative opinion can be very important because it may be these reasons that enable one to perform actions to remedy the situation. For example, the sentence "*I do not like the picture quality of this camera*" is not as useful as "*I do not like the picture quality of this camera because the pictures are quite dark*." The first sentence does not give the reason for the negative sentiment about the picture quality and it is thus difficult to know what to do to improve the picture quality. The second sentence is more informative because it gives the reason or cause for the negative sentiment. The camera manufacturer can make use of this piece of information to improve the picture quality of the camera. In most industrial applications, such reasons are called *problems* or *issues*. Knowing the issues allows businesses to find ways to address them.

Definition 9 (Qualifier of Opinion) A qualifier of an opinion limits or modifies the meaning of the opinion.

Knowing the qualifier is also important in practice because it tells what the opinion is good for. For example, "*This car is too small for a tall person*" does not say that the car is too small for everyone, but just for tall people. For a person who is not tall, this opinion does not apply.

However, as we have seen, not every opinion comes with an explicit reason and/or an explicit qualifier. "*The picture quality of this camera is not great*" does not have a reason or a qualifier. "*The picture quality of this camera is not good for night shots*" has a qualifier "*for night shots*," but does not give a specific reason for the negative sentiment. "*The picture quality of this camera is not good for night shots as the pictures are quite dark*" has a reason for the negative sentiment ('*the pictures are quite dark*') and also a qualifier ('*for night shots*'). Sometimes, the qualifier and the reason may not be in the same sentence and/or may be quite implicit, e.g., "*The picture quality of this camera is not great. Pictures of night shots are very dark*" and "*I am 6 feet 5 inches tall. This car is too small for me*." An expression can also serve multiple purposes. For example, '*too small*' in the above sentence indicates the *size* aspect of the car, a *negative sentiment* about the size, and also the *reason* for the negative sentiment/opinion.

2.1.6 Objective and Tasks of Sentiment Analysis

With the definitions in Sects. 2.1.1, 2.1.2, 2.1.3 and 2.1.4, we can now present the core objective and the key tasks of (aspect-based) sentiment analysis.

Objective of Sentiment Analysis Given an opinion document d, discover all opinion quintuples (e, a, s, h, t) in d. For more advanced analysis, discover the reason and qualifier for the sentiment in each opinion quintuple.

Key Tasks of Sentiment Analysis The key tasks of sentiment analysis can be derived from the five components of the quintuple (Definition 7). The first component is the entity and the first task is to extract entities. The task is similar to named entity recognition (NER) in information extraction (Hobbs and Riloff 2010; Sarawagi 2008). However, as defined in Definition 3, an entity can also be an event, issue, or topic, which is usually not a named entity. For example, in "*I hate tax increase*," the entity is '*tax increase*,' which is an issue or topic. In such cases, entity extraction is basically the same as aspect extraction and the difference between entity and aspect becomes blurry. In some applications, there may not be a need to distinguish them.

After extraction, we need to categorize the extracted entities as people often write the same entity in different ways. For example, Motorola may be written as Mot, Moto, and Motorola. We need to recognize that they all refer to the same entity (see (Liu 2015) for details).

Definition 10 (Entity Category and Entity Expression) An *entity category* represents a unique entity, while an *entity expression* or *mention* is an actual word or phrase that indicates an entity category in the text.

Each entity or entity category should have a unique name in a particular application. The process of grouping or clustering entity expressions into entity categories is called *entity resolution* or *grouping*.

For aspects of entities, the problem is basically the same as for entities. For example, *picture*, *image*, and *photo* refer to the same aspect for cameras. We thus need to extract aspect expressions and resolve them.

Definition 11 (Aspect Category and Aspect Expression) An *aspect category* of an entity represents a unique aspect of the entity, while an *aspect expression* or *mention* is an actual word or phrase that indicates an aspect category in the text.

Each aspect or aspect category should also have a unique name in a particular application. The process of grouping aspect expressions into aspect categories (aspects) is called *aspect resolution* or *grouping*.

Aspect expressions are usually nouns and noun phrases but can also be verbs, verb phrases, adjectives, and adverbs. They can also be explicit or implicit (Hu and Liu 2004).

Definition 12 (Explicit Aspect Expression) Aspect expressions that appear in an opinion text as nouns and noun phrases are called *explicit aspect expressions*.

For example, '*picture quality*' in "*The picture quality of this camera is great*" is an explicit aspect expression.

Definition 13 (Implicit Aspect Expression) Aspect expressions that are not nouns or noun phrases but indicate some aspects are called *implicit aspect expressions*.

For example, *expensive* is an implicit aspect expression in "*This camera is expensive*." It implies the aspect *price*. Many implicit aspect expressions are adjectives and adverbs used to describe or qualify some specific aspects, e.g., *expensive* (price), and *reliably* (reliability). They can also be verb and verb phrases, e.g., "*I can install the software easily*." *Install* indicates the aspect *installation*.

Implicit aspect expressions are not just individual adjectives, adverbs, verbs and verb phrases; they can be very complex. For example, in "*This camera will not easily fit in my pocket*," '*fit in my pocket*' indicates the aspect *size* (and/or *shape*). In the sentence "*This restaurant closes too early*," '*closes too early*' indicates the aspect of *closing time* of the restaurant. In both cases, some commonsense knowledge may be needed to recognize them.

Aspect extraction is a very challenging problem, especially when it involves verbs and verb phrases. In some cases, it is even very hard for human beings to recognize and to annotate. For example, in a vacuum cleaner review, one wrote "*The vacuum cleaner does not get the crumbs out of thick carpets*," which seems to describe only one very *specific* aspect, '*get the crumbs out of thick carpets*.' But in practice, it may be more useful to decompose it into three different aspects indicated by (1) '*get something out of*,' (2) *crumbs*, and (3) '*thick carpets*.' (1) represents the suction power of the vacuum cleaner in general, (2) represents suction related to *crumbs*, and (3) represents *suction* related to '*thick carpets*.' All three are important and useful because the user may be interested in knowing whether the vacuum can suck crumbs, and whether it works well with thick carpets.

The third component in the opinion definition is the sentiment. For this, we need to perform sentiment classification or regression to determine the sentiment orientation or score on the involved aspect and/or entity. The fourth component and fifth components are opinion holder and opinion posting time respectively. They also have expressions and categories as entities and aspects. I will not repeat their definitions. Note that opinion holders (Bethard et al. 2004; Choi et al. 2005; Kim and Hovy 2004) are also called *opinion sources* in (Wiebe et al. 2005).

Based on the above discussions, we can now define a model of entity and a model of opinion document (Liu 2006) and summarize the main sentiment analysis tasks.

Model of Entity An entity e is represented by itself as a whole and a finite set of its aspects $A = \{a1, a2, \ldots, an\}$. e can be expressed in text with any one of a finite set of its entity expressions $\{ee1, ee2, \ldots, ees\}$. Each aspect $a \in A$ of entity e can be expressed with any one of its finite set of aspect expressions $\{ae1, ae2, \ldots, aem\}$.

Model of Opinion Document An opinion document d contains opinions about a set of entities $\{e1, e2, \ldots, er\}$ and a subset of aspects of each entity. The opinions are from a set of opinion holders $\{h1, h2, \ldots, hp\}$ and are given at a particular time point t.

Given a set of opinion documents D, sentiment analysis performs the following eight (8) main tasks:

Task 1 (entity extraction and resolution): Extract all entity expressions in D, and group synonymous entity expressions into entity clusters (or categories). Each entity expression cluster refers to a unique entity e.

Task 2 (aspect extraction and resolution): Extract all aspect expressions of the entities, and group these aspect expressions into clusters. Each aspect expression cluster of entity e represents a unique aspect a.

Task 3 (opinion holder extraction and resolution): Extract the holder expression of each opinion from the text or structured data and group them. The task is analogous to tasks 1 and 2.

Task 4 (time extraction and standardization): Extract the posting time of each opinion and standardize different time formats.

Task 5 (aspect sentiment classification or regression): Determine whether an opinion about an aspect a (or entity e) is positive, negative or neutral (classification), or assign a numeric sentiment rating score to the aspect (or entity) (regression).

Task 6 (opinion quintuple generation): Produce all opinion quintuples (e, a, s, h, t) expressed in D based on the results from tasks 1–5. This task is seemingly very simple but it is in fact quite difficult in many cases as Review B below shows.

For more advanced analysis, we also need to perform the following two additional tasks, which are analogous to task 2:

Task 7 (opinion reason extraction and resolution): Extract reason expressions for each opinion, and group all reason expressions for each aspect or entity and each sentiment orientation into clusters. Each cluster for an aspect (or entity) and a sentiment orientation represents a unique reason for the aspect (or entity) and the orientation.

Task 8 (opinion qualifier extraction and resolution): Extract qualifier expressions for each opinion, and group all qualifier expressions for each aspect (or entity) and each sentiment orientation into clusters. Each cluster for an aspect (or entity) and a sentiment orientation represents a unique qualifier for the aspect (or entity) and the orientation.

Although reasons for and qualifiers of opinions are useful, their extraction and categories are very challenging. Little research has been done about them so far.

We use an example review to illustrate the tasks (a sentence id is again associated with each sentence) and the mining results.

Review B

Posted by: bigJohn
Date: Sept. 15, 2011

> (1) *I bought a Samsung camera and my friend brought a Canon camera yesterday*. (2) *In the past week, we both used the cameras a lot*. (3) *The photos from my Samy are not clear for night shots, and the battery life is short too*. (4) *My friend was very happy with his camera and loves its picture quality*. (5) *I want a camera that can take good photos*. (6) *I am going to return it tomorrow*.

Task 1 should extract the entity expressions, *Samsung*, *Samy*, and *Canon*, and group *Samsung* and *Samy* together because they represent the same entity. Task 2 should extract aspect expressions *picture*, *photo*, and *battery life*, and group *picture* and *photo* together as they are synonyms for cameras. Task 3 should find that the holder of the opinions in sentence (3) is bigJohn (the blog author) and that the holder of the opinions in sentence (4) is bigJohn's friend. Task 4 should find that the time when the blog was posted is Sept-15-2011. Task 5 should find that sentence (3) gives a negative opinion to the *picture quality* of the Samsung camera and a negative

opinion also to its *battery life*. Sentence (4) gives a positive opinion to the *Canon camera* as a whole and also to its *picture quality*. Sentence (5) seemingly expresses a positive opinion, but it does not. To generate opinion quintuples for sentence (4) we need to know what '*his camera*' and *its* refer to. Task 6 should finally generate the following opinion quintuples:

1. (Samsung, picture_quality, negative, bigJohn, Sept-15-2011)
2. (Samsung, battery_life, negative, bigJohn, Sept-15-2011)
3. (Canon, GENERAL, positive, bigJohn's_friend, Sept-15-2011)
4. (Canon, picture_quality, positive, bigJohn's_friend, Sept-15-2011)

With more advanced mining and analysis, we also find the reasons and qualifiers of opinions. *None* below means unspecified.

1. (Samsung, picture_quality, negative, bigJohn, Sept-15-2011)
   Reason for opinion: picture not clear
   Qualifier of opinion: night shots
2. (Samsung, battery_life, negative, bigJohn, Sept-15-2011)
   Reason for opinion: short battery life
   Qualifier of opinion: none
3. (Canon, GENERAL, positive, bigJohn's_friend, Sept-15-2011)
   Reason for opinion: none
   Qualifier of opinion: none
4. (Canon, picture_quality, positive, bigJohn's_friend, Sept-15-2011)
   Reason for opinion: none
   Qualifier of opinion: none

2.2 Definition of Opinion Summary

Unlike facts, opinions are subjective (although they may not be all expressed in subjective sentences). An opinion from a single opinion holder is usually not sufficient for action. In almost all applications, the user needs to analyze opinions from a large number of opinion holders. This tells us that some form of summary of opinions is necessary. The question is what an opinion summary should be. On the surface, an opinion summary is just like a multi-document summary because we need to summarize multiple opinion documents, e.g., reviews. It is, however, very different from traditional multi-document summary. Although there are informal descriptions about what a traditional multi-document summary should be, it is never formally defined. A traditional multi-document summary is often just "defined" operationally based on each specific algorithm that produces the summary. Thus different algorithms produce different kinds of summaries. The resulting summaries are also hard to evaluate. An opinion summary in its core form, on the other hand, can be defined precisely based on the quintuple definition of opinion and easily evaluated. That is, all opinion summarization algorithms should aim to produce the

same summary. Although they may still produce different final summaries, that is due to their different accuracies. This core form of opinion summary is called the *aspect-based opinion summary* (or *feature-based opinion summary*) (Hu and Liu 2004; Liu et al. 2005)

Definition 11 (Aspect-Based Opinion Summary) The *aspect-based opinion summary* about an entity e is of the following form:

GENERAL: number of opinion holders who are positive about entity e number of opinion holders who are negative about entity e
Aspect 1: number of opinion holders who are positive about aspect 1 of entity e number of opinion holders who are negative about aspect 1 of entity e
...
Aspect n: number of opinion holders who are positive about aspect n of entity e number of opinion holders who are negative about aspect n of entity e
where GENERAL represents the entity e itself and n is the total number of aspects of e.

The key features of this opinion summary definition are that it is based on positive and negative opinions about each entity and its aspects and that it is quantitative. The quantitative perspective is reflected by the numbers of positive and negative opinions. In an application, the number counts can also be replaced by percentages. The quantitative perspective is especially important in practice. For example, 20% of the people positive about a product is very different from 80% of the people positive about the product.

To illustrate this form of summary, we summarize a set of reviews of a digital camera, called *digital camera* 1, in Figure 2.1. This is called a *structured summary* in contrast to a traditional text summary of a short document generated from one or multiple long documents. In the figure, 105 reviews expressed positive opinions about the camera itself denoted by GENERAL and 12 expressed negative opinions. *Picture quality* and *battery life* are two camera aspects. 75 reviews expressed positive opinions about the picture quality, and 42 expressed negative opinions.

Digital Camera 1:
Aspect: **GENERAL**
Positive: 105 <Individual review sentences>
Negative: 12 <Individual review sentences>
Aspect: **Picture quality**
Positive: 75 <Individual review sentences>
Negative: 42 <Individual review sentences>
Aspect: **Battery life**
Positive: 50 <Individual review sentences>
Negative: 9 <Individual review sentences>
...

Fig. 2.1 An aspect-based opinion summary

We also added <*Individual review sentences*>, which can be a link pointing to the sentences and/or the whole reviews that contain the opinions (Hu and Liu 2004; Liu et al. 2005). With this summary, one can easily see how existing customers feel about the camera. If one is interested in a particular aspect and additional details, one can drill down by following the <*Individual review sentences*> link to see the actual opinion sentences or reviews.

In a more advanced analysis, we can also summarize opinion reasons and qualifiers in a similar way. Based on my experience, qualifiers for opinion statements are rare, but reasons for opinions are quite common. To perform the task, we need another level of summary. For example, in the example of Figure 2.1, we may want to summarize the reasons for the poor picture quality based on the sentences in <*Individual review sentences*>. We may find that 35 people say the pictures are not bright enough and 7 people say that the pictures are blurry.

Based on the idea of aspect-based summary, researchers have proposed many opinion summarization algorithms, and also extended this form of summary to some other more specialized forms (Liu 2015).

2.3 Affect, Emotion, and Mood

A*ffect*, *emotion*, and *mood* have been studied extensively in several fields, e.g., psychology, philosophy, and sociology. However, investigations in these fields are seldom concerned with the language expressions used to express such feelings. Their main concerns are people's psychological states of mind, theorizing what affect, emotion and mood are, what constitute basic emotions, what physiological reactions happen (e.g., heart rate changes, blood pressure, sweating and so on), what facial expressions, gestures and postures are, and measuring and investigating the impact of such mental states. These mental states have also been exploited extensively in application areas such as marketing, economics, and education.

However, even with such extensive research, understanding these concepts is still slippery and confusing because different theorists often have somewhat different definitions for them and even do not completely agree with each other about what emotion, mood, and affect are. For example, about emotion, diverse theorists have proposed that there are from two to twenty basic human emotions and some even do not believe there is such a thing called basic emotions (Ortony and Turner 1990). In most cases, emotion and affect are regarded as synonymous, and indeed, all three terms are sometimes used interchangeably. Affect is also used as an encompassing term covering all topics related to emotion, feeling, and mood. To make matters worse, in applications, researchers and practitioners use these concepts loosely in whatever way they feel like to without following any established definitions. Thus one is often left puzzled by just what an author means when the word emotion, mood, or affect is used. In most cases, the definition of each term also uses one or more of the other terms resulting in circular definitions, which causes further confusion. The good news for natural language processing researchers and practitioners

is that in practical applications of sentiment analysis, we needn't be too concerned with such an unsettled state of affair because in practice we can pick up and use whatever emotion or mood states that are suitable for the applications at hand.

This section first tries to create a reasonable understanding of these concepts and their relationships for our tasks of natural language processing in general and sentiment analysis in particular. It then puts these three concepts in the context of sentiment analysis and discusses how they can be handled in sentiment analysis.

2.3.1 Affect, Emotion, and Mood in Psychology

We start the discussion with the dictionary definitions of affect, emotion, and mood[1]. The concept of *feeling* is also included as all three concepts are about human feelings. From the definitions, we can see how difficult it is to explain or to articulate these concepts:

- **Affect**: Feeling or emotion, especially as manifested by facial expression or body language.
- **Emotion**: A mental state that arises spontaneously rather than through conscious effort and is often accompanied by physiological changes.
- **Mood**: A state of mind or emotion.
- **Feeling**: An affective state of consciousness, such as that resulting from emotions, sentiments, or desires.

These definitions are confusing from a scientific point of view because we do not see a clear demarcation for each concept. We turn to the field of psychology to look for a better definition for each of them. The convergence of views and ideas among theorists in the past twenty years gives us a workable classification scheme.

An *affect* is commonly defined as an neurophysiological state consciously accessible as the simplest raw (nonreflective) feeling evident in moods and emotions (Russell 2003). The key point here is that such a feeling is primitive and not directed at an object. For example, you are watching a scary movie. If you are affected, it moves you and you experience a feeling of being scared. Your mind further processes this feeling and expresses it to yourself and the world around you. The feeling is then displayed as an *emotion*, such as crying, shock, and scream.

Emotion is thus the indicator of affect. Due to cognitive processing, emotion is a compound (rather than primitive) feeling concerned with a specific object, such as a person, an event, a thing, or a topic. It tends to be intense and focused and lasts a short period of time. *Mood*, like emotion, is a feeling or affective state but it typically lasts longer than emotion and tends to be more unfocused and diffused. Mood is also less intense than emotion. For example, you may wake up feeling happy and stay that way for most of the day.

[1] http://www.thefreedictionary.com/subjective

In short, emotions are quick and tense, while moods are more diffused and prolonged feelings. For example, we can get very angry very quickly, but it is difficult to stay very angry for a long time. The anger emotion may subside into an irritable mood that can last quite a long time. An emotion is usually very specific, triggered by noticeable events, which means that an emotion has a specific target. In this sense, emotion is like a rational opinion. On the other hand, a mood can be caused by multiple events, and sometimes it may not have any specific targets or causes. Mood typically also has a dimension of future expectation. It can involve a structured set of beliefs about general expectations of a future experience of pleasure or pain, or of positive or negative affect in the future (Batson et al. 1992).

Since sentiment analysis is not so much concerned with affect as defined above, below we focus only on *emotion* and *mood* in the psychological context. Let us start with emotion. Emotion has been frequently mentioned in sentiment analysis. Since it has a target or an involved entity, it fits the sentiment analysis context naturally. Almost all applications are interested in opinions and emotions about some target entities.

Theorists in psychology have grouped emotions into categories. However, as we mentioned earlier, there is still not a set of agreed basic (or primary) emotions among theorists. In (Ortony and Turner 1990), the basic emotions proposed by several theorists were compiled to show there is a great deal of disagreement. We reproduce them in Table 2.1.

In (Parrott 2001), apart from the basic emotions, secondary and tertiary emotions were also proposed (see Table 2.2). These secondary and tertiary are useful in some

Table 2.1 Basic emotions from different theorists

Source	Basic emotions
Arnold (1960)	Anger, aversion, courage, dejection, desire, despair, fear, hate, hope, love, sadness
Ekman et al. (1982)	Anger, disgust, fear, joy, sadness, surprise
Gray (1982)	Anxiety, joy, rage, terror
Izard (1971)	Anger, contempt, disgust, distress, fear guilt, interest, joy, shame, surprise
James (1884)	Fear, grief, love, rage
McDougall (1926)	Anger, disgust, elation, fear, subjection, tender-emotion, wonder
Mowrer (1960)	Pain, pleasure
Oatley and Jobnson-Laird (1987)	Anger, disgust, anxiety, happiness, sadness
Panksepp (1982)	Expectancy, fear, rage, panic
Plutchik (1980)	Acceptance, anger, anticipation, disgust, joy, fear, sadness, surprise
Tomkins (1984)	Anger, interest, contempt, disgust, distress, fear, joy, shame, surprise
Watson (1930)	Fear, love, rage
Weiner and Graham (1984)	Happiness, sadness
Parrott (2001)	Anger, fear, joy, love, sadness, surprise

Table 2.2 Primary, Secondary and Tertiary emotions from Parrott (2001)

Primary emotion	Secondary emotion	Tertiary emotion
Anger	Disgust	Contempt, loathing, revulsion
	Envy	Jealousy
	Exasperation	Frustration
	Irritability	Aggravation, agitation, annoyance, crosspatch, grouchy, grumpy
	Rage	Anger, bitter, dislike, ferocity, fury, hatred, hostility, outrage, resentment, scorn, spite, vengefulness, wrath
	Torment	Torment
Fear	Horror	Alarm, fear, fright, horror, hysteria, mortification, panic, shock, terror
	Nervousness	Anxiety, apprehension (fear), distress, dread, suspense, uneasiness, worry
	Cheerfulness	Amusement, bliss, gaiety, glee, jolliness, joviality, joy, delight, enjoyment, gladness, happiness, jubilation, elation, satisfaction, ecstasy, euphoria
Joy	Contentment	Pleasure
	Enthrallment	Enthrallment, rapture
	Optimism	Eagerness, hope
	Pride	Triumph
	Relief	Relief
	Zest	Enthusiasm, excitement, exhilaration, thrill, zeal
Love	Affection	Adoration, attractiveness, caring, compassion, fondness, liking, sentimentality, tenderness
	Longing	Longing
	Lust/sexual desire	Desire, infatuation, passion
	Disappointment	Dismay, displeasure
	Neglect	Alienation, defeatism, dejection, embarrassment, homesickness, humiliation, insecurity, insult, isolation, loneliness, rejection
Sadness	Sadness	Depression, despair, gloom, glumness, grief, melancholy, misery, sorrow, unhappy, woe
	Shame	Guilt, regret, remorse
	Suffering	Agony, anguish, hurt
	Sympathy	Pity, sympathy
Surprise	Surprise	Amazement, astonishment

sentiment analysis applications because the set of basic emotions may not be fine-grained enough. For example, in one of the applications that I worked on, the client was interested in detecting *optimism* in the financial market. Optimism is not a basic emotion in the list of any theorist, but it is a secondary emotion for *joy* in Table 2.2. Note that although the words in Table 2.2 describe different emotions or states of mind, they can also be used as part of an emotion lexicon in sentiment analysis to

spot different kinds of emotions. Of course, they need to be significantly expanded to include those synonymous words and phrases to form a reasonably complete emotion lexicon. In fact, there are some emotion lexicons that have been compiled by researchers, see (Liu 2015). Note also that for sentiment analysis, we do not need to be concerned with the disagreement of theorists. For a particular application, we can choose the types of emotion that are useful to the application. We also do not need to worry about whether they are primary, second or tertiary.

The *emotion annotation and representation language* (EARL) proposed by the Human-Machine Interaction Network on Emotion (HUMAINE) (HUMAINE 2006) has classified 48 emotions into different kinds of positive and negative orientations or valences (Table 2.3). This is useful to us because sentiment analysis is mainly interested in expressions with positive or negative orientations or polarities (also called *valences*). However, we should take note that some emotions do not have positive or negative orientations, e.g., *surprise* and *interest*. Some psychologists felt that these should not be regarded as emotions (Ortony and Turner 1990) simply because they do not have positive or negative orientations or valences. For the same reason, they are not commonly used in sentiment analysis.

Table 2.3 HUMAINE polarity annotations of emotions

Negative and forceful	**Negative and passive**	**Quiet positive**
Anger	Boredom	Calm
Annoyance	Despair	Content
Contempt	Disappointment	Relaxed
Disgust	Hurt	Relieved
Irritation	Sadness	Serene
Negative and not in control	**Positive and lively**	**Caring**
Anxiety	Amusement	Affection
Embarrassment	Delight	Empathy
Fear	Elation	Friendliness
Helplessness	Excitement	Love
Powerlessness	Happiness	
Worry	Joy	
	Pleasure	
Negative thoughts	**Positive thoughts**	**Reactive**
Doubt	Courage	Interest
Envy	Hope	Politeness
Frustration	Pride	Surprised
Guilt	Satisfaction	
Shame	Trust	
Agitation		
Stress		
Shock		
Tension		

We now turn to mood. The types of mood are similar to those of emotion except that the types of emotion that last only momentarily will not usually be moods, e.g., *surprise* and *shock*. Thus, the words or phrases used to express moods are similar to those for emotions too. However, since mood is a feeling that lasts a relatively long time, is diffused, and may not have a clear cause or target object, it is hard to recognize unless a person explicitly says it, e.g., *I feel sad today*. We can also monitor one's writings over a period of time to assess his/her prevailing mood in the period, which can help discover people with prolonged mental or other medical conditions (e.g., chronicle depression) and even the tendency to commit suicides or crimes.

It is also interesting to discover the mood of the general population, e.g., public mood, and the general atmosphere between organizations or countries, e.g., the mood of US and Russian relations, by monitoring the traditional news media and/or social media over a period of time.

2.3.2 Affect, Emotion, and Mood in Sentiment Analysis

The above discussions are only about people's states of mind, which are the subjects of study of psychologists. However, for sentiment analysis, we need to know how such feelings are expressed in natural language and how they can be recognized. This leads us to the linguistics of affect, emotion and mood. Affect as defined by psychologists as a primitive response or feeling with no target is not much of interest to us as almost everything written in text or displayed in the form of facial expressions and other visible signs have already gone through some cognitive processing to become emotion or mood. However, we note that the term affect is still commonly used in linguistics and many other fields to mean emotion and mood.

Wikipedia has a good page describing the linguistic aspect of emotion and mood. There are two main ways that human beings express themselves, speech and writing. In addition to choices of grammatical and lexical expressions, which are common to both speech and writing (see below), speaker emotion can also be conveyed through paralinguistic mechanisms such as intonations, facial expressions, body movements, biophysical signals or changes, gestures, and postures. In writing, special punctuations (e.g., repeated exclamation marks, !!!!), capitalization of all letters of a word, emoticons, and lengthening of words (e.g., *slooooooow*) are frequently used, especially in social media.

Regarding choices of grammatical and lexical expressions, there are several common ways that people often employ to express emotions or moods:

1. use emotion or mood words or phrases such as love, disgusting, angry, and upset.
2. describe emotion-related behaviors, e.g., "He cried after he saw his mother" and "After received the news, he jumped up and down for a few minutes like a small boy."
3. use intensifiers. As we discussed in Sect. 2.1.3, common English intensifiers include very, so, extremely, dreadfully, really, awfully (e.g., *awfully bad*), *terribly*

(e.g., *terribly good*), *never* (e.g., "I will never buy any product from them again"), *the sheer number of, on earth* (e.g., "What on earth do you think you are doing?"), *the hell* (e.g., "What the hell are you doing?"), *a hell of a*, etc. To emphasize further, intensifiers may be repeated, e.g., "This car is very very good."
4. use superlatives. Arguably, many superlative expressions also express emotions, e.g., "This car is simply the best."
5. use pejorative (e.g., "*He is a fascist.*"), laudatory (e.g., "He is a saint."), and sarcastic expressions (e.g., "What a great car, it broke the second day").
6. use swearing, cursing, insulting, blaming, accusing, and threatening expressions.

My experience is that using these clues is sufficient to recognize emotion and mood in text, although in linguistics, adversative forms, honorific and deferential language, interrogatives, tag questions, and the like may also be employed to express emotional feelings, but their uses are rare and are also hard to recognize computationally.

To design emotion detection algorithms, in addition to considering the above clues, we should be aware that there is a cognitive gap between people's true psychological states of mind and the language that they use to express such states. There are many reasons (e.g., being polite, and do not want people to know one's true feeling) that they may not fully match. Thus, language does not always represent psychological reality. For example, when one says "*I am happy with this car*," one may not have any emotional reaction towards the car although the emotion word *happy* is used. Furthermore, emotion and mood are very difficult to distinguish in written text (Alm 2008). We normally do not distinguish them. When we say emotion, we mean emotion or mood.

Since emotions have targets and most of them also imply positive or negative sentiment, they can be represented and handled in very much the same way as rational opinions. Although a rational opinion emphasizes a person's evaluation about an entity and an emotion emphasizes a person's feeling caused by an entity, emotion can essentially be regarded as sentiment with a stronger intensity (see Sect. 2.1.3). It is often the case that when the sentiment of a person becomes so strong, he/she becomes emotional. For example, "*The hotel manager is not professional*" expresses a rational opinion, while "*I almost cried when the hotel manager talked to me in a hostile manner*" indicates that the author's sentiment reached the emotional level of *sadness* and/or *anger*. The sentiment orientation of an emotion naturally inherits the polarity of the emotion, e.g., *sad*, *anger*, *disgust*, and *fear* are negative, and *love* and *joy* are positive. At the emotional level, sentiment becomes more fine-grained. Additional mechanisms are needed to recognize different types of emotions in writing.

Due to the similarity of emotion and rational opinion, we can still use the quadruple or quintuple representation of opinion (Definitions 1 and 7) to represent emotion. However, if we want to be more precise, we can give it a separate definition based on the quadruple (Definition 1) or quintuple (Definition 7) definitions as the meanings of some components in the tuple are not the exactly same as they were in the opinion definition because emotions focus on personal feelings, while rational opinions focus on evaluations.

Definition 14 (Emotion) An *emotion* is a quintuple,

$$(e, a, m, f, t),$$

where e is the target entity, a is the target aspect of e that is responsible for the emotion, m is the emotion type or a pair representing an emotion type and an intensity level, f is the feeler of the emotion, and t is the time when the emotion is expressed.

For example, for the emotion expressed in the sentence "*I am so upset with the manager of the hotel*," the entity is '*the hotel*,' the aspect is '*the manager*' of the hotel, the emotion type is *anger*, and the emotion feeler is *I* (the author). If we know the time when the emotion was expressed we can add it to the quintuple representation. As another example, in "*After hearing his brother's death, he burst into tears.*" the target entity is '*his brother's death*,' which is an event, and there is no aspect. The emotion type is *sadness* and the emotion feeler is *he*.

In practical applications, we should integrate the analysis of rational opinions and emotions, we may also want to add the sentiment orientation or polarity of an emotion, i.e., whether it is positive (desirable) or negative (undesirable) for the feeler. If that is required, a sentiment component can be included to Definition 14 to make it a sextuple.

Cause of Emotion In Sect. 2.1.5, we discussed the reasons for opinions. In a similar way, emotions have causes as emotions are usually caused by some internal or external events. Here we use the word *cause* instead of *reason* because an emotion is an effect produced by a cause (usually an event), rather than a justification or explanation in support of an opinion. In the above sentence, '*his brother's death*' is the cause for his *sadness* emotion. Actually, '*his brother's death*' is both the target entity and the cause. In many cases, the target and the cause of an emotion are different. For example, in "*I am so mad with the hotel manager because he refused to refund my booking fee,*" the target entity is the *hotel*, the target aspect is the *manager* of the hotel, and the cause of the *anger* emotion is '*he refused to refund my booking fee.*' There is a subtle difference between '*his brother's death*' and '*he refused to refund my booking fee.*' The latter states an action performed by *he* (the hotel manager) that causes the sadness emotion (negative). *He* is the agent of the undesirable action. The sentiment on the hotel manager is negative. The sentence also explicitly stated the *anger* is toward the hotel manager, In the case of '*his brother's death,*' '*his brother*' or *death* alone is not the target of the emotion. It is the whole event that is the target and the cause of the sadness emotion.

Unlike rational opinions, in many emotion and mood sentences, the authors may not explicitly state the entities (e.g., named entities, topics, issues, actions and events) that are responsible for the emotions or moods, e.g., "*I felt a bit sad this morning*" and "*There is sadness in her eyes*." The reason is that a rational opinion sentence focuses on both the opinion target and the sentiment on the target but the opinion holder is often omitted (e.g., "*The pictures from this camera are great*") while an emotion sentence focuses on the feeling of the feeler (e.g., "*There*

is sadness in her eyes." This means that a rational opinion sentence contains both sentiments and their targets explicitly, but may or may not give the opinion holder. An emotion sentence always has feelers and emotion expressions, but may or may not state the emotion target or the cause (e.g., "*I love this car*" and "*I felt sad this morning*"). This does not mean that some emotions do not have targets or causes. They do, but the targets or the causes may be expressed in previous sentences or implied by the context, which makes extracting targets or causes very difficult. In the case of mood, the causes may be implicit or even unknown and are thus not stated in the text.

2.4 Different Types of Opinions

Opinions can actually be classified along many dimensions. We discuss some main classifications in this section.

2.4.1 Regular and Comparative Opinions

The type of opinion that we have defined is called the *regular opinion* (Liu 2006). Another type is *comparative opinion* (Jindal and Liu 2006b).

Regular Opinion A *regular opinion* is often referred to simply as an *opinion* in the literature. It has two main sub-types (Liu 2006):

Direct opinion: A *direct opinion* is an opinion that is expressed directly on an entity or an entity aspect, e.g., "*The picture quality is great*."

Indirect opinion: An *indirect opinion* is an opinion that is expressed indirectly on an entity or aspect of an entity based on some positive or negative effects on some other entities. This sub-type often occurs in the medical domain. For example, the sentence "*After injection of the drug, my joints felt worse*" describes an undesirable effect of the drug on '*my joints*,' which indirectly gives a negative opinion or sentiment to the drug. In this case, the entity is *the drug* and the aspect is the *effect on joints*. Indirect opinions also occur in other domains, although less frequently. In these cases, they are typically expressed *benefits* (positive) or *issues* (negative) of entities, e.g., "*With this machine, I can finish my work in one hour*, *which used to take me 5 hours*" and "*After switching to this laptop, my eyes felt much better*." In marketing, benefits of a product or service are regarded as the major selling points. Thus, extracting such benefits is of practical interest.

Comparative Opinion 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 shared aspects of the entities (Jindal and Liu 2006a, b). For example, the sentences "*Coke tastes better than Pepsi*" and "*Coke tastes the best*" express

two comparative opinions. A comparative opinion is usually expressed using the *comparative* or *superlative* form of an adjective or adverb, although not always (e.g., *prefer*). The definitions in Sects. 2.1 and 2.2 do not cover comparative opinion. Comparative opinions have many types. See (Liu 2015) for their definitions.

2.4.2 Subjective and Fact-Implied Opinions

Opinions and sentiments are by nature subjective because they are about people's subjective views, appraisals, evaluations, and feelings. But when they are expressed in actual text, they do not have to appear as subjective sentences. People can use objective or factual sentences to express their happiness and displeasure because facts can be desirable or undesirable. Conversely, not all subjective sentences express positive or negative sentiments, e.g., "*I think he went home*," which is a belief and has no positive or negative orientation. Based on subjectivity, we can classify opinions into two types, *subjective opinions* and *fact-implied opinions*. We define them below.

Subjective Opinion An *subjective opinion* is a regular or comparative opinion given in a subjective statement, e.g.,

"*Coke tastes great*."
"*I think Google's profit will go up next month*."
"*This camera is a masterpiece*."
"*We are seriously concerned about this new policy*."
"*Coke tastes better than Pepsi*."

We can broadly classified subjective opinions into two categories: *rational opinions* and *emotional opinions* (Sect. 2.1.3).

Fact-Implied Opinion A *fact implied opinion* is a regular or comparative opinion implied in an objective or factual statement. Such an objective statement expresses a desirable or undesirable fact or action. This type of opinion can be further divided into two subtypes:

1. **Personal fact-implied opinion**: Such an opinion is implied by a factual statement about someone's personal experience, e.g.,

"*I bought the mattress a week ago, and a valley has formed in the middle*."
"*I bought the toy yesterday and I have already thrown it into the trash can*."
"*My dad bought the car yesterday and it broke today*."
"*The battery of this phone lasts longer than that of my previous Samsung phone.*"

Although factual, these sentences tell us whether the opinion holder is positive or negative about the product or his preference among different products. Thus, the opinions implied by these factual sentences are no different from subjective opinions.

2. **Non-personal fact-implied opinion**: This type is entirely different as it does not imply any personal opinion. It often comes from fact reporting and the reported fact does not give any opinion from anyone, e.g.,

   "*Google's revenue went up by 30%*."
   "*The unemployment rate came down last week.*"
   "*Google made more money than Yahoo last month*."

Unlike personal facts, these sentences do not express any experience or evaluation from any person. For instance, the first sentence above does not have the same meaning as a sentiment resulted from a person who has used a Google product and expresses a desirable or undesirable fact about the Google product. Since these sentences do not give any personal opinion, they do not have opinion holders although they do have the sources of information. For example, the source of the information in the first sentence above is likely to be Google itself, but it is a fact, not a Google's subjective opinion.

However, we can still treat them as a type of opinion sentences due to the following two reasons:

1. Each of the sentences above does indicate a desirable and/or undesirable state for the involved entities or topics (i.e., *Google*, *Yahoo* and *unemployment rate*) based on our commonsense knowledge.
2. The persons who post the above sentences might be expressing positive or negative opinions implicitly about the involved entities. For example, the person who posted the first sentence on Twitter is likely to have a positive sentiment about Google; otherwise, he/she would probably not post the fact. This kind of posts occur very frequently on Twitter, where Twitter users pick up some news headlines from the traditional media and post them on Twitter. Many people may also re-tweet them.

As we can see, it is important to distinguish personal facts and non-personal facts as opinions induced from non-personal facts represent a very different type of opinions and need a special treatment. How to deal with such facts depends on applications. My recommendation is to assign it the positive or negative orientation based on our commonsense knowledge whether the sentence is about a fact desirable or undesirable to the involved entity, e.g., Google. Users of the sentiment analysis system should be made aware of the convention so that they can make use the opinion appropriately based on their applications.

Sometimes the author who posts such a fact may also give an explicit opinion, e.g.,

"*I am so upset that Google's share price went up today*."

The clause '*Google's share price went up today*' in the example gives a non-personal fact-implied positive opinion about Google, but the author is negative about it. This is called a *meta-opinion*, an opinion about an opinion.

Subjective opinions are usually easier to deal with because the number of words and phrases that can be used to explicitly express subjective feelings is limited, but this is not the case for fact-implied opinions. There seem to be an infinite number of desirable and undesirable facts and every domain is different. Much of the existing research has focused on subjective opinions. Limited work has been done about fact-implied opinions (Zhang and Liu 2011).

2.4.3 First-Person and Non-First-Person Opinions

In some applications, it is important to distinguish those statements expressing one's own opinions from those statements expressing beliefs about someone else's opinions. For example, in a political election, one votes based on one's belief of each candidate's stances on issues, rather than the true stances of the candidate, which may or may not be the same.

1. **First-person opinion:** Such an opinion states one's own attitude towards an entity. It can be from a person, a representative of a group, or an organization. Here are some example sentences expressing first-person opinions.

   "*Tax increase is bad for the economy*."
   "*I think Google's profit will go up next month*."
   "*We are seriously concerned about this new policy*."
   "*Coke tastes better than Pepsi*."

   Notice that not every sentence needs to explicitly use the first person pronoun "*I*" or "*we,*" or to mention an organization name.

2. **Non-first-person opinion:** Such an opinion is expressed by a person stating someone else's opinion. That is, it is a belief of someone else's opinion about some entities or topics, e.g.,

   "*I think John likes Lenovo PCs*."
   "*Jim loves his iPhone*."
   "*President Obama supports tax increase*."
   "*I believe Obama does not like wars*."

2.4.4 Meta-opinions

Meta-opinions are opinions about opinions. That is, a meta-opinion's target is also an opinion which is usually contained in a subordinate clause. The opinion in the subordinate clause can express either a fact with an implied opinion or a subjective opinion. Let us see some examples:

"*I am so upset that Google's profit went up*"
"*I am very happy that my daughter loves her new Ford car*"
"*I am so sad that Germany lost the game*."

These sentences look quite different from opinion sentences before. But they still follow the same opinion definition in Definition 7. It is just that the target of the meta-opinion in the main clause is now an opinion itself in the subordinate clause. For example, in the first sentence, the author is negative about '*Google's profit went up*,' which is the target of the meta-opinion in the main clause. So the meta-opinion is negative, but its target is a regular positive opinion about '*Google's profit*.' In practice, these two types of opinions should be treated differently. Since meta-opinions are rare, there is little research or practical work about them.

2.5 Author and Reader Standpoint

We can look at an opinion from two perspectives, that of the author (opinion holder) who posts the opinion, and that of the reader who reads the opinion. Since opinions are subjective, naturally the author and the reader may not see the same thing in the same way. Let us use the following two example sentences to illustrate the point:

"*This car is too small for me*."
"*Google's profits went up by 30%*."

Since the author or the opinion holder of the first sentence felt the car is too small, a sentiment analysis system should output a negative opinion about the size of the car. However, this does not mean that the car is too small for everyone. A reader may actually like the small size, and feel positive about it. This causes a problem because if the system outputs only a negative opinion about size, the reader will not know whether it is too small or too large and then he/she would not see this positive aspect for him/her. Fortunately, this problem can be dealt with by mining and summarizing opinion reasons (see Sect. 2.1.2). Here '*too small*' not only indicates a negative opinion about the size but also the reason for the negative opinion. With the reason, the reader can see a more complete picture of the opinion.

The second sentence represents a non-personal fact-implied opinion. As discussed in Sect. 2.4.1, the person who posts the fact is likely to be positive about Google. However, the readers may have different feelings. Those who have financial interests in Google should feel happy, but Google's competitors will not be thrilled. In Sect. 2.4.2, we choose to assign positive sentiment to the opinion because our commonsense knowledge says that the fact is desirable for Google. Users can decide how to use the opinion based on their application needs.

2.6 Summary

This chapter described many facets of sentiment analysis. It started with the definitions of the concepts of opinion, sentiment, and opinion summary. The definitions abstracted a structure from the unstructured natural language text, and also showed that sentiment analysis is a multi-faceted problem with many interrelated sub-

problems. Researchers can exploit the inter-relationships to design more robust and accurate solution techniques. This chapter also classified and discussed different types of opinions. Along with these definitions and discussions, the important concepts of affect, emotion and mood were introduced and defined too. They are closely related to, but are also different from conventional rational opinions. Opinions emphasize evaluation or appraisal of some target objects, events or topics (which are collectively called entities in this chapter), while emotions emphasize people's feelings caused by such entities.

After reading this chapter, I am sure that you would agree with me that on the one hand, sentiment analysis is a challenging area of research involving many different tasks and perspectives, and on the other, it is also highly subjective in nature. Thus, I do not expect that you completely agree with me on everything in the chapter. I also do not claim that this chapter covered all important aspects of sentiment and opinion. My goal is to present a reasonably precise definition of sentiment analysis (or opinion mining) and its related concepts, issues, and tasks. I hope I have succeeded to some extent.

References

Alm, Ebba Cecilia Ovesdotter. 2008. *Affect in text and speech*: ProQuest.

Arnold, Magda B. 1960. *Emotion and personality*. New York: Columbia University Press.

Batson, C. Daniel, Laura L. Shaw, and Kathryn C. Oleson. 1992. Differentiating affect, mood, and emotion: Toward functionally based conceptual distinctions. *Emotion Review of Personality and Social Psychology* 13: 294–326.

Bethard, Steven, Hong Yu, Ashley Thornton, Vasileios Hatzivassiloglou, and Dan Jurafsky. 2004. Automatic extraction of opinion propositions and their holders. In *Proceedings of the AAAI spring symposium on exploring attitude and affect in text*.

Chaudhuri, Arjun. 2006. *Emotion and reason in consumer behavior*. Oxford: Elsevier Butterworth-Heinemann.

Choi, Yejin, Claire Cardie, Ellen Riloff, and Siddharth Patwardhan. 2005. 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)*.

Ekman, P., W.V. Friesen, and P. Ellsworth. 1982. What emotion categories or dimensions can observers judge from facial behavior? In *Emotion in the human face*, ed. P. Ekman, 98–110. Cambridge: Cambridge University Press.

Gray, Jeffrey A. 1982. *The neuropsychology of anxiety*. Oxford: Oxford University Press.

Hobbs, Jerry R., and Ellen Riloff. 2010. Information extraction. In *In handbook of natural language processing*, ed. N. Indurkhya and F.J. Damerau, 2nd ed. London: Chapman & Hall/CRC Press.

Hu, Minqing and Bing Liu. 2004. Mining and summarizing customer reviews. In *Proceedings of ACM SIGKDD international conference on Knowledge Discovery and Data Mining (KDD-2004)*.

HUMAINE. 2006. *Emotion annotation and representation language*. Available from: http://emotion-research.net/projects/humaine/earl

Izard, Carroll Ellis. 1971. *The face of emotion*. New York: Appleton-Century-Crofts.

James, William. 1884. What is an emotion? *Mind* 9: 188–205.

Jindal, Nitin and Bing Liu. 2006a. Identifying comparative sentences in text documents. In *Proceedings of ACM SIGIR conference on research and development in information retrieval (SIGIR-2006)*.

———. 2006b. Mining comparative sentences and relations. In *Proceedings of national conference on artificial intelligence (AAAI-2006)*.

Kim, Soo-Min and Eduard Hovy. 2004. Determining the sentiment of opinions. In *Proceedings of interntional conference on computational linguistics (COLING-2004)*.

Liu, Bing. 2006. *Web data mining: Exploring hyperlinks, contents, and usage data*. Berlin: Springer.

———. 2010. Sentiment analysis and subjectivity, in *Handbook of natural language processing, Second Edition*, N. Indurkhya and F.J. Damerau, Editors.

———. 2015. *Sentiment analysis: Mining opinions, sentiments, and emotions*. Cambridge: Cambridge University Press.

Liu, Bing, Minqing Hu, and Junsheng Cheng. 2005. Opinion observer: Analyzing and comparing opinions on the web. *Proceedings of international conference on world wide web (WWW-2005)*.

McDougall, William. 1926. *An introduction to social psychology*. Boston: Luce.

Mowrer, Orval Hobart. 1960. *Learning theory and behavior*. New York: Wiley.

Oatley, K., and P.N. Jobnson-Laird. 1987. Towards a cognitive theory of emotions. *Cognition and Emotion* 1: 29–50.

Ortony, Andrew, and Terence J. Turner. 1990. What's basic about basic emotions? *Psychological Review* 97 (3): 315–331.

Panksepp, Jaak. 1982. Toward a general psychobiological theory of emotions. *Behavioral and Brain Sciences* 5 (3): 407–422.

Parrott, W. Gerrod. 2001. *Emotions in social psychology: Essential readings*. Philadelphia: Psychology Press.

Plutchik, Robert. 1980. A general psychoevolutionary theory of emotion. In *Emotion: Theory, research, and experience: Vol. 1. Theories of emotion*, ed. R. Plutchik and H. Kellerman, 3–33. New York: Academic Press.

Qiu, Guang, Bing Liu, Bu Jiajun, and Chun Chen. 2011. Opinion word expansion and target extraction through double propagation. *Computational Linguistics* 37 (1): 9–27.

Russell, James A. 2003. Core affect and the psychological construction of emotion. *Psychological Review* 10 (1): 145–172.

Sarawagi, Sunita. 2008. Information extraction. *Foundations and Trends in Databases* 1 (3): 261–377.

Tomkins, Silvan. 1984. Affect theory. In *Approaches to emotion*, ed. K.R. Scherer and P. Ekman, 163–195. Hillsdale: Eribaum.

Watson, John B. 1930. *Behaviorism*. Chicago: Chicago University Press.

Weiner, B., and S. Graham. 1984. An attributional approach to emotional development. In *Emotion, cognition and behavior*, ed. C.E. Izard, J. Kagan, and R.B. Zajonc, 167–191. New York: Cambridge University Press.

Wiebe, Janyce, Theresa Wilson, and Claire Cardie. 2005. Annotating expressions of opinions and emotions in language. *Language Resources and Evaluation* 39 (2): 165–210.

Zhang, Lei and Bing Liu. 2011. Identifying noun product features that imply opinions. In *Proceedings of the annual meeting of the Association for Computational Linguistics (short paper) (ACL-2011)*.

Zhuang, Li, Feng Jing, and Xiaoyan Zhu. 2006. Movie review mining and summarization. In *Proceedings of ACM international conference on information and knowledge management (CIKM-2006)*.