Knowledge and Information Systems (2019) 60:617–663
https://doi.org/10.1007/s10115-018-1236-4

SURVEY PAPER

# A survey of sentiment analysis in social media

Lin Yue[1,2,3,5] · Weitong Chen[3] · Xue Li[3,4] · Wanli Zuo[5] · Minghao Yin[1,2,5]

Received: 17 November 2016 / Revised: 30 March 2018 / Accepted: 27 April 2018 /
Published online: 4 July 2018


**Abstract**
Sentiments or opinions from social media provide the most up-to-date and inclusive information, due to the proliferation of social media and the low barrier for posting the message. Despite the growing importance of sentiment analysis, this area lacks a concise and systematic arrangement of prior efforts. It is essential to: (1) analyze its progress over the years, (2) provide an overview of the main advances achieved so far, and (3) outline remaining limitations. Several essential aspects, therefore, are addressed within the scope of this survey. On the one hand, this paper focuses on presenting typical methods from three different perspectives (task-oriented, granularity-oriented, methodology-oriented) in the area of sentiment analysis. Specifically, a large quantity of techniques and methods are categorized and compared. On the other hand, different types of data and advanced tools for research are introduced, as well as their limitations. On the basis of these materials, the essential prospects lying ahead for sentiment analysis are identified and discussed.

**Keywords** Sentiment analysis · Social media · Data mining · Machine learning · Survey

## 1 Introduction

Sentiment analysis (SA) sometimes alternatively mentioned as opinion mining is a research area which aims to analyze people's sentiments or opinions toward entities such as topics, events, individuals, issues, services, products, organizations, and their attributes [63]. Senti-

✉ Weitong Chen
w.chen9@uq.edu.au

✉ Minghao Yin
ymh@nenu.edu.cn

1 School of Information Science and Technology, Northeast Normal University, Changchun, China

2 Key Laboratory of Applied Statistics of MOE, Northeast Normal University, Changchun, China

3 School of Information Technology and Electrical Engineering, The University of Queensland, Brisbane, Australia

4 College of Computer Science and Technology, Nanjing University of Aeronautics and Astronautics, Nanjing, China

5 Key Laboratory of Symbolic Computation and Knowledge Engineering of Ministry of Education, Jilin University, Changchun, China

ment analysis should be treated as a branch of machine learning, data mining, natural language processing, and computational linguistics, which also borrows elements from sociology and psychology. Although the history of natural language processing (NLP) starts in the 1950s, little attention had been paid to people's opinions and sentiments analysis until 2005s. For the past few years, the prosperity of the social media propels the development of sentiment analysis. In this section, we will mainly answer three main questions: the importance of sentiment analysis, the differences between our survey and former ones, and the contributions of our survey.

### 1.1 Reasons for caring about sentiment analysis in social media

The growth of information available in social media makes sentiment analysis more crucial; related researches have been filtered into three main aspects of the application.

From the commercial perspective, sentiment analysis can provide online advices and recommendations for both of the customers and merchants. On the one hand, user preferences that the data reveals can be used to help e-commerce platforms analyzing their products and services. On the other hand, for the virtual nature of online shopping, it is not easy to understand a commodity comprehensively and objectively, and whether the consumer is willing to learn about other consumers' comments or opinions.

From the political perspective, the massive demand for political information can be regarded as another important factor. Commercial application is not the only motivation behind people seeking or expressing opinions online. For example, in the analysis of the conversation on Twitter before the European Parliament elections, researchers have collected more than 1.2 million tweets in three languages (English, German, and French) during a two-week period (May 1 to May 14, 2014),[1] in which:

- For the Twitter conversation in English, 39% of the assertions on Twitter are negative toward the European Union (EU), compared with 30% neutral and 31% positive. In French, the result broke down the same basic way, about 39% negative, 28% neutral, and 33% positive. And while in German, conversations on Twitter are much more positive (39%) than negative (5%), these attitudes are embedded in sparse conversation that represent a mere fraction of the Twitter activity in English and French (see Fig. 1a).
- There are low-intensity conversations of English language on Twitter about the five candidates competing for the European Commission presidency: Jean-Claude Juncker from Luxembourg, who is a candidate of the center-right European People's Party; Martin Schulz from Germany, who is a candidate of the center-left Party of European Socialists; Guy Verhofstadt from Belgium, representing the Alliance of Liberals and Democrats for Europe; Alexis Tsipras from Greece, who is a candidate of the European Left Party; and Ska Keller from Germany, who is a candidate of the European Green Party (see Fig. 1b).

From the perspective of public security, sociopolitical events such as the Arab Spring and the London Riots vividly demonstrate the importance of sentiment analysis to public security. In both cases, online social medias such as Twitter and Facebook were regarded as significant contributors to the outbreak and proliferation of the events. Sentiment analysis can help authorities discover those types of sensitive information in advance. If so, an action such as shutting down Internet communication channels would prevent supporters of terrorism from having access to such services.

[1] http://www.journalism.org/2014/05/22/the-eu-elections-on-twitter/.

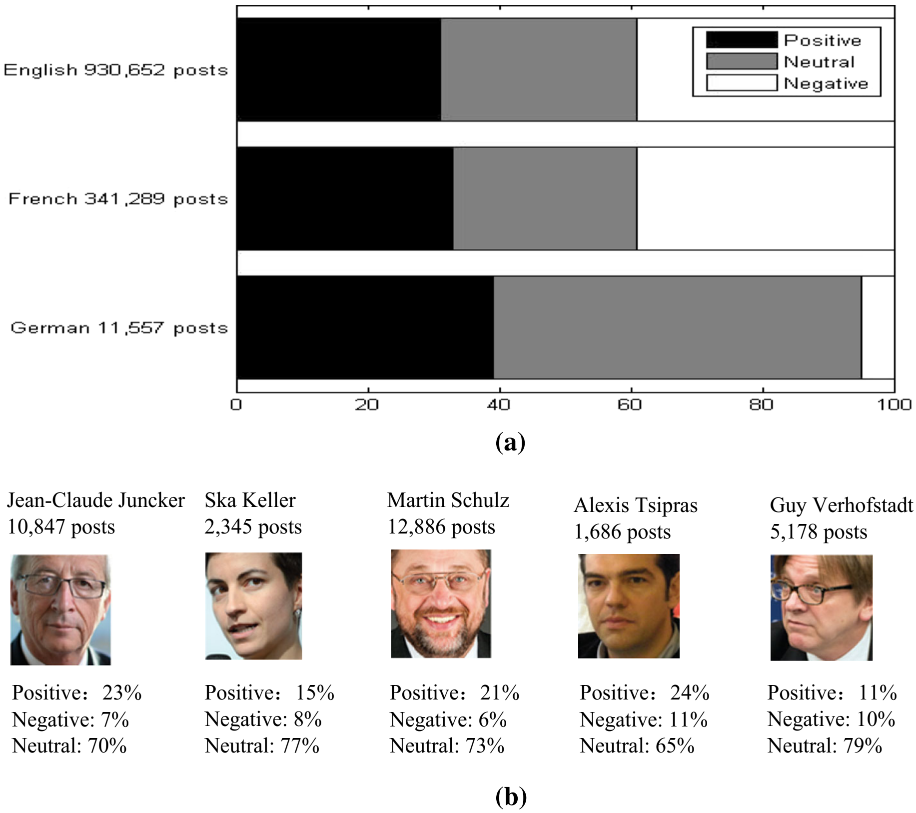



**Fig. 1** The EU elections on Twitter: mixed views about the EU and little passion for the candidates. **a** Mixed Twitter views on the EU in English, French, German. **b** EU commission presidential candidates get very modest attention on Twitter (number of posts in English only)

To sum up, sentiment analysis or opinion mining task is important not only for traditional consumers and companies conducting surveys to gather opinions about corresponding products or services, but also it plays an important role on national security and public opinion analysis.

### 1.2 Differences between this survey and former ones

Plenty of techniques have been developed and conduced to make sentiment analysis viable. However, to the best of our knowledge, sentiment analysis is still in its early stage and suffers changes; few systematic reviews well shape this area and position existing works and current progress. The main barrier is that sentiment analysis is a multi-faceted problem including many subproblems, but not a single task. Although some works have explored sentiment analysis built on multi-perspectives, few have sought to provide an in-depth survey of current researches or detail on some open problems presented in this field. Existing surveys of sentiment analysis studies, either concentrate on enumerating technical details or just focus on a certain aspects of sentiment analysis. Moreover, the rapid development of the field, to some extent, makes those surveys outdated. Besides, there is no survey organized for the beginners getting started quickly with the research progress, the classical algorithm, the recent dynamics, the typical tools, the suitable databases, etc.

Of all those surveys proposed in recent years, Liu [63] summed up all important research topics in the field of sentiment analysis, which is regarded as an encyclopedia on sentiment analysis and opinion mining. As evidence of that, it involves more than 400 bibliographic references from major journals and conferences. Pang et al. [87] covered works that directly enable opinion-oriented information-seeking systems, which sought to address the problems raised by sentiment-aware applications and included material on issues regarding manipulation, privacy, and economic impact that is generated by opinion-oriented information-access services. Medhat et al. [73] provided a categorization of articles according to the used techniques, which can help the researchers who are familiar with certain techniques to use them in the sentiment analysis and choose the appropriate technique for a certain application. Vinodhini et al. [131] presented a short survey that covers the techniques and challenges appeared in the field of sentiment analysis. Wiegand et al. [138] presented a survey on the role of negation in sentiment analysis, in which various computational approaches modeling negation are introduced. In particular, this work focuses on aspects such as negation word detection, scope of negation, and limitation and challenges of negation modeling. Ravi et al. [106] is organized on the basis of sub-tasks to be performed, that is to say, machine learning, natural language processing techniques, and applications of sentiment analysis. Saif et al. [109] presented an overview of eight publicly available manually annotated evaluation datasets for Twitter sentiment analysis and a common limitation of most of these datasets. Tsytsarau et al. [125] reviewed the development of sentiment analysis in recent years and also discussed the gradual progress of a research direction, namely contradiction analysis. Schouten et al. [111] focused on aspect-level sentiment analysis, and the goal of this paper is to find and aggregate sentiment on entities mentioned within documents or aspects of these entities. Tang et al. [121] discussed related issues and main approaches to word sentiment classification, subjectivity classification, opinion extraction, and document sentiment classification. Giachanou et al. [39] investigated and briefly described the algorithms of sentiment analysis in Twitter, in which researchers discussed tasks related to Twitter opinion retrieval, tracking sentiments over time, irony detection, emotion detection, and tweet sentiment quantification.

### 1.3 Contributions of this survey

With the constant advent of novel researches, a new inclusive survey is required for better understanding of current research progress. The goal of this paper is to give an in-depth introduction and present new insights toward this area. Our work is different from the others by proposing novel multiple perspectives, teasing out series of works, organizing tools, and benchmarks datasets used in various work. To our knowledge, this type of study is not available in the literature.

In summary, the contributions of this article are:

1. We reviewed literatures in the area of sentiment analysis from multiple perspectives and provided the strength and weaknesses of these methods and techniques. Various techniques of sentiment analysis are categorized with brief details of the algorithms and their originating references. This work will help beginners aiming at sentiment analysis to have a panoramic view of the entire research field.
2. The available benchmark datasets are discussed and categorized according to the usage in certain applications. We have analyzed types of tools and lexicons that can be used in sentiment analysis research as well as their limitations.

3. On the basis of these materials, the essential prospects lying ahead for sentiment analysis are identified and discussed. Specifically, we have summarized multimodal sentiment analysis (MSA) techniques in sentiment analysis, which we thought will be the main challenge for future development.

This survey is organized as follows: In Sect. 2, we introduce main concepts and definitions used throughout this paper. In Sect. 3, we present researches from new perspectives of task-oriented, granularity-oriented, methodology-oriented sentiment analysis and explore the motivations, benefits, and limitations of these techniques and methods. In Sect. 4, we discuss some of the benchmark datasets and tools that have been used in sentiment analysis research as well as their limitations. Section 5 highlights possible trends and challenges in sentiment analysis. Finally, Sect. 6 concludes this survey with final remarks.

## 2 Terminology and background concepts

Before we dive into the details of this survey, we start with an introduction of the terminologies and background concepts, which concerns of emotion and sentiment in the field of psychology, the definition of sentiment analysis in social media, and different types of opinions.

### 2.1 Emotion and sentiment from psychological perspective

The words emotion and sentiment are usually used interchangeably in daily life. While in the field of psychology, they are treated as two different concepts. Watson and Tellegen's well-known psychological model showed a two-dimensional structure of emotions (see Fig. 2). Emotion is defined as a complex psychological state, which plays a key role of operating motivators. Psychologist Paul Eckman proposed six basic emotions, i.e., happiness, sadness, anger, fear, surprise, and disgust. Later on, other psychologists enriched this list by adding emotions such as pride, excitement, embarrassment, contempt, shame. There are three main components in emotion, which respectively are the subjective experience, the physical response, and the behavioral response. Sentiment, on the other hand, refers to a mental attitude, which is created based on emotion. The definition highlights that sentiment is used to convey the thoughts of the individual deriving from his emotion. Unlike emotion, sentiment goes a step further, which is not confined to the psychological dimensions. Sentiment is usually not a primary emotion but is highly organized. McDougall stated that sentiment usually connects primary emotion with action. To sum up, the differences between emotion and sentiment can be seen from four aspects (see Table 1):

All types of emotion can be treated as multiple factors together for affecting sentiment. Most of the sentiment analysis researches are based on natural language processing so far. These traditional works focus on textual content, while we are increasingly making use of audios, images, and videos to air our opinions on social media platforms. Thus, it is highly crucial to analyze opinions and identify sentiments from multiple modalities, which is called emotion analysis. While the field of multimodal sentiment analysis (MSA) has not received much attention, the scatter of multimodal sentiment expressions requires a higher demand for more analysis to derive useful data. We will present an overview of recent research of different models and individually and jointly find the gaps in terms of tasks, approaches theories, and applications in Sect. 5.

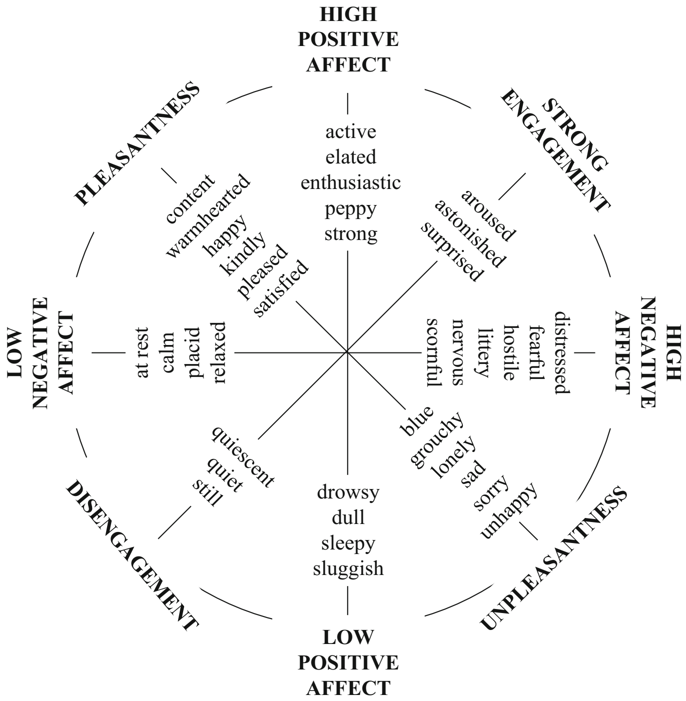



**Fig. 2** The two-dimensional structure of emotions

**Table 1** The differences between emotion and sentiment

| | Definition | Connection | Dimension | Nature |
|---|---|---|---|---|
| Emotion | Complex psychological states | – | Psychological dimension | Raw and natural |
| Sentiment | Mental attitudes or thoughts | Expression of emotions | Social dimension | Highly organized |

## 2.2 Sentiment analysis

The terms of sentiment analysis and opinion mining are usually used interchangeably in research papers of data mining, machine learning, etc., although these concepts are not equivalent. The meaning of sentiment itself is still ambiguous. We will distinguish them when needed. To simplify the presentation, we will use the term of sentiment to denote corresponding opinion, sentiment, appraisal, evaluation, attitude, and emotion. As sentiment analysis was proposed to deal with product or service review, the corresponding definition was based on this kind of reviews as shown below:

| |
|---|
| 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 the observation, an opinion consists of some key components:

1. *Opinion holder* For sentence (1), (2), (3), (4), the opinion holder is John Smith; while for sentence (5), the holder of opinion is the wife of John Smith.
2. *The date of the review* September 10, 2011.
3. *A Aarget* Canon G12 camera.
4. *Aspects or features of the target* The picture quality and the battery life are different aspect of the same target.
5. *Sentiment on the target* This review has a number of both positive and negative opinions about Canon G12 camera. A positive opinion about the Canon camera is expressed by Sentence (2); a positive opinion about its picture equality is expressed by Sentence (3); a positive opinion about its battery life is expressed by Sentence (4); a negative opinion about the weight of the camera is expressed by Sentence (5).

In [63], the above five components are considered as essential. Researcher defines opinion as a quintuple $(e_i, a_{ij}, s_{ijkl}, h_k, t_l)$, where $e_i$ is the name of an entity, $a_{ij}$ is an aspect of $e_i$, $s_{ijkl}$ is the sentiment on 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 sentiment $s_{ijkl}$ is positive, negative, or neutral, or expressed with different strength/intensity levels, e.g., 1–5 stars as used by most review online.

Actually, current sentiment analysis collects all relevant data, public thoughts, opinions and feelings, from all sorts of publicly accessible sources. Besides, it automatically predicts outcomes or trends on different aspects of the information collected worldwide in real time. As the aforementioned examples illustrate, coordinating a series of sociopolitical events such as the London riots and Arab Spring, or election prediction, to a large extent, are based on spatial information. We have achieved some statistical results as shown in Table 2; Dataset 1 and Dataset 2 are separately crawled from Twitter posted around Brisbane Australia from March 17 to March 25, 2012 (219,933 messages) and the messages posted around USA from June 21 to June 27, 2012 (196,834 messages), in which spatial information could appear in geo-tag, user profile, IP address, and even tweet content.

**Table 2** Different ways of spatial information appearing in Twitter

| List | Dataset 1 #msg | % | Dataset 2 #msg | % |
|---|---|---|---|---|
| Messages with geo-tagged location | 3121 | 1.42 | 5055 | 2.57 |
| Messages with user profile location | 170,334 | 77.45 | 132,661 | 67.40 |
| Messages with IP address-based location | 46,478 | 21.13 | 59,118 | 30.03 |
| Messages that contain geographical location in contents | 16,875 | 7.67 | 19,529 | 9.92 |
| Messages that contain geographical locations in contents more than 1 location | 744 | 4.41 | 1337 | 6.85 |

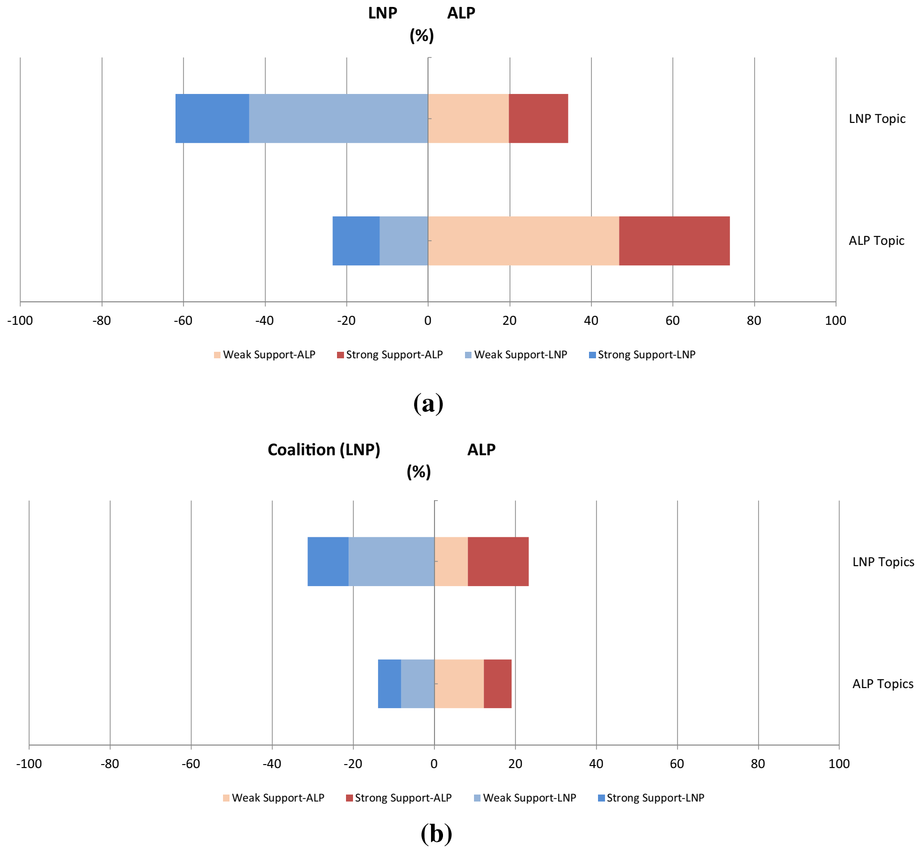



**Fig. 3** Opinion analysis on Twitter for the Australian election (LNP = Liberal National Party; ALP = Australian Labor Party). **a** Queensland election prediction. **b** Australian federal election prediction

As different users from different geo-locations will post different tweets for their local issues, spatial information is a baseline to catch the prominent and statistical features of tweets, such as to detect emerging local events, to guess the location of a user, to reveal behavioral features of local users. Spatial–temporal word spectrum (STWS) model is a linguistic fingerprint of a geo-location on social media, which opens a new way of studying social media. With STWS model, we have also discovered a series of interesting events on Twitter, e.g., people are dissatisfied with Orange polices' green-lighting for drug dealers (June 25, 2015); people's appealing for homosexual marriage legitimation in Australia (June 27, 2015); people from Brisbane tend to have bad mood and talk billingsgate the most on Tuesday (data collected from Twitter on an average of 2 min interval from April 23 17:00 EST 2012- April 26 11:00 EST 2012).

Besides, we noticed a few important points, that is to say, Twitter has been used to study national election results; however, very few people, to our knowledge, have yet conducted an in-depth investigation into how well this platform predicts the outcomes of the election with location-based information. For that reason, we performed opinion analysis on Twitter for the Australian Election Prediction (see Fig. 3):[2]

[2] http://www.queenslandimage.com/.

*Queensland election prediction* We collected a total number of 14,922 tweets (Tweets collected between 00 middle night 01/01/2015 and 12 noon 30/01/2015) for predicting QLD Election Parties results. Based on the data crawled from Twitter, the Labour Party is expected to win slightly. Specifically, the public are positive to Labor which is 12.39% higher than LNP. In addition, others (Greens, etc.) are at 13%.

*Australian federal election prediction* Twitter data are collected between 00:00 12/31/2015 and 03:55 28/06/2016, in a statewide scope of Australia. About 968,966 tweets are related to politics. A proportion of 63.13% tweets (611,691) was found relevant to this event. Among them, a proportion of 34.07% tweets (208,413) were subjective opinions. Based on our analysis on Twitter, there seems a 10% ahead of support to Coalition than ALP in Australian Federal Election 2016.

The problem that we addressed is how to identify emerging events with location sensitivity from a given set of Twitter messages. We have considered a set of messages, in which each message is associated with an event. However, due to the characteristics of Twitter messages, several issues should be considered such as the abbreviation identification, the short tweets text with the semantic loss, the dynamic topic tracking. And, probably most important of all is location-based sentiment analysis. People might share various types of content such as conversation topics, advertisements, events, opinions, and others. Our goal is to detect only emerging hotspot events that are happening in a particular area within a given period of time. From this review, we need to perform location identification based on check-in information, user profile, and geographical name in Twitter textual content before we analyze users' support over a specific political party.

From this review, we add spatial information element and redefine opinion as a six-tuple:

**Definition** (*opinion*) An opinion is a six-tuple,

$$(e_i,\ a_{ij},\ s_{ijklt},\ h_k,\ t_l,\ p_t)$$

where $e_i$ denotes the name of an object, $a_{ij}$ is an aspect of $o_i$, $s_{ijklt}$ is the sentiment on an aspect $a_{ij}$ of object $o_i$, $h_k$ is the sentiment holder, $t_l$ is the time when the sentiment is expressed by $h_k$, and $p_t$ is the place when the sentiment is expressed by $h_k$. Note that objects can be issues, products, organizations, services, events, individuals, topics, etc.

## 2.3 Different types of opinions

In [63], various types of opinions in opinion mining are introduced and compared, including regular and comparative opinions, explicit and implicit opinions, subjectivity and emotion. In this section, we will briefly sketch these basic conceptions combined with instances.

*Regular and comparative opinions* Direct sentiment and indirect sentiment, as two main sub-types, are included in regular sentiment. Direct sentiment expresses opinions on an object or an aspect of the object directly. For the indirect sentiment, opinions on an object or an aspect of an object are expressed indirectly. For example, the sentence "After the election campaign, all candidates felt exhausted" describes an undesirable effect of the election on "all candidates," which indirectly gives a negative opinion or sentiment to the election campaign. In this case, the entity is the election campaign and the aspect is the effect on all candidates. Comparative sentiment expresses relative or similar attribute between two objects. The superlative or comparative form of an adjective or an adverb is usually contained in comparative opinion, e.g., "Donald Trump has higher approval ratings than Hillary Clinton does".

*Explicit and implicit opinions* A subjective statement that gives a comparative or a regular opinion is treated as an explicit opinion. While an objective statement that gives a comparative or regular opinion is treated as an implicit opinion, e.g., "Stop the boat!" is a domain-specific phrase, which implies negative attitude to ALP and positive to LNP in Australian government election in 2013. We can intuitively get the conclusion that explicit opinions are easier to detect and classify than implicit opinions.

*Subjectivity and emotion* A subjective sentence expresses some personal feelings, views, or beliefs. Emotions are subjective feelings and thoughts, which are closely related to sentiments. The strength of a sentiment or an opinion is typically related to the intensity of certain emotions, e.g., "joy and anger."

Other than all types of opinions mentioned above, there are still some challenging examples as shown below. From the observation: The review in sentence (1) has a target restaurant,while food and service are two different aspects of the same target with positive and negative opinions, constructing the corresponding relationship between one aspect of the target and its opinion is one important preparatory step; sentence (2) implies a negative attitude about a certain event or location of current speaker; sentence (3) seems like a negative opinion literally. But under some particular situations, such as between two friends, it could be positive tone as well. All these examples show more complicated proposals should be involved in opinion analysis task, especially in natural language process, text mining, etc.

(1) "The food was great but the service was awful."
–Object with features: [Restaurant, Food, Services]
(2) "I really think I shouldn't be here."
–Negative to the implied event or location of the current speaker.
(3) "You are terrible! : - )"
–Positive to an object: [a friend of speaker]
(4) "Stop the boat!"
–Domain specific: negative to ALP & positive to LNP in Australian government election in 2013.

## 3 Sentiment analysis research

Existing surveys of sentiment analysis studies either concentrate on enumerating technical details in the areas of social media analysis, text mining, natural language processing, and data mining or just focus on a certain aspects of sentiment analysis researches. Moreover, the rapid development of the field, to some extent, makes those surveys outdated. Besides, there is no one survey organized for the beginners getting started quickly with the research progress, the classical algorithm, the recent dynamics. In this survey, we reviewed the literatures in sentiment analysis from multiple perspectives and provided the strength and weaknesses of these methods and techniques (see Fig. 4).

### 3.1 Task-oriented sentiment analysis

In this section, five different aspects in task-oriented sentiment analysis, as depicted in Fig. 4, will be introduced and compared. These five aspects are polarity classification, the level of valence or arousal at specific scale, beyond polarity, subjectivity/objectivity identification, and feature/aspect-based sentiment analysis.

#### 3.1.1 Polarity classification

Basic research of sentiment analysis is polarity classification, which explores whether the expressed opinion in a document or a sentence, on a certain feature or aspect of a target,

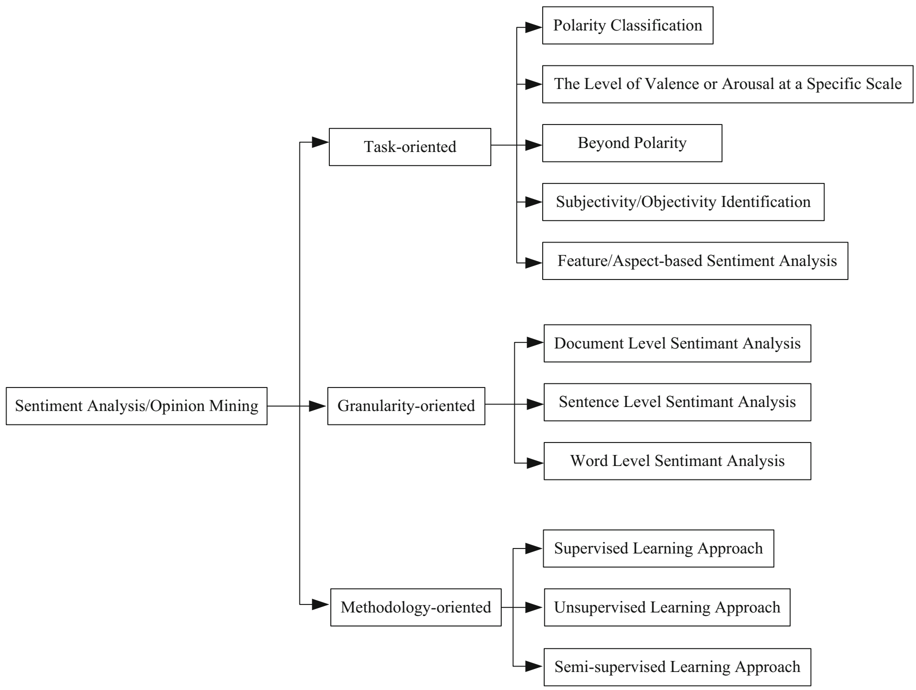



**Fig. 4** Sentiment analysis researches from multiple perspectives

is positive, negative, or neutral. Early work in this area includes [89, 126], who applied different methods for detecting the polarity of product reviews and movie reviews. Pang et al. [89] first explored the effectiveness of applying naïve Bayes and support vector machine to the sentiment classification task on movie reviews. Sentiment classification can usually be seen as a two-class classification problem. In this type of research, sentiment analysis is essentially a text classification problem, in which feature selection has important effects on the performance of classification algorithm. As in machine learning and statistics, feature selection is the process of selecting a subset of relevant features for model construction. In text classification, feature selection techniques are used for simplifying the models to make them easier to be interpreted, reducing training times, and enhancing generalization by overcoming the problem of overfitting. In sentiment classification, feature words should be sentiment words that indicate a certain polarity, which needs to process more complex semantics. Moreover, researchers studied the several types of features and found that using unigrams features gets better performance compared to other different feature selection strategies, although two machine learning methods on sentiment classification task do not perform as well as on traditional text classification task. Like traditional classification task, the key to sentiment classification task is also selecting effective features. These effective features can be seen in Table 3.

In recent years, Khan et al. [56] presented a multilingual framework for polarity classification with two aspects of contributions. On the one hand, this approach can serve as a baseline to compare other classification systems. It considers techniques for text representation such as spelling features, emoticons, word-based n-grams, character-based q-grams, and language dependent features. On the other hand, it is a framework for practitioners looking for a bootstrapping sentiment classifier method to build more elaborated systems. As

**Table 3** The candidate features for sentiment classification

| Feature types | Descriptions | Examples |
|---|---|---|
| Terms and their frequency | Frequency counts of individual words and their n-grams | TF-IDF weighting scheme |
| Part of speech | As important indicators of opinions, adjectives are treated as special features | – |
| Sentiment words and phrases | Words that are used to express negative or positive sentiments | Positive sentiment words such as good, wonderful, and amazing<br>Negative sentiment words such as bad, poor, and terrible<br>Nouns such as rubbish, junk, and crap<br>Verbs such as hate and love |
| Rules and opinions | Individual sentiment words with implied sentiments or compound expressions and their orientations determined by domain knowledge | Compositional semantics |
| Sentiment shifters | Expressions that could change the sentiment orientations | Negation words of the most common type of sentiment shifters such as not, never, none, nobody, nowhere, neither, and cannot<br>Modal auxiliary verbs such as would, should, could, might, must, and ought<br>Words such as fail, omit, neglect, less<br>Sarcasm often changes orientations |
| Syntactic dependency | Features with word dependency generated from parsing or dependency trees | – |

SentiWordNet (SWN) has been extensively used as a lexical resource for opinion mining, Tellez et al. [122] proposed a framework named enhanced sentiment analysis and polarity classification (eSAP), which incorporates SWN as the labeled training corpus, where the sentiment scores are extracted based on the part of speech information. Then, SWN-based vocabulary SWN-V generated from SWN with revised sentiment scores is used for support vector machine model learning and classification process.

### 3.1.2 The level of valence or arousal at specific scale

People also classify document polarity on multiple scales; Pang et al. [86] and Snyder et al. [113] expanded the basic task of two-class classification as either positive or negative for predicting star ratings on either a 3 or a 4 star scale, while Snyder et al. [113] performed an in-depth review analysis by predicting ratings for various aspects of the given target. Although in most statistical classification algorithms, the neutral category is ignored under the assumption that neutral objects lie near the boundary of the binary classifier, a few researchers suggest that, as in every polarity classification task, three categories must be identified. Moreover, it can be proven that specific classifiers like max entropy and SVMs will benefit from adding neutral class and improve the effectiveness of classification.

In opinion mining, unigram and n-gram representations of text are common ways. However, important expressions like "could have been better" that are essential for prediction models of ratings cannot be captured by unigrams. N-grams representations, on the other hand, are capable of capturing such phrases, but typically occur too sparsely in the training data and thus fail to yield robust performance. Qu et al. [105] addressed the problem of numeric rating prediction from another perspective. They overcome the limitations of unigrams and n-grams representations, through introducing a novel bag-of-opinions representation method, where the opinion in a review is constituted with three components: a root word, a modifier word set from the same sentence, and one or more negative words. The prior polarity of the opinion is determined by opinion root. Then the strength of the prior polarity will be intensified or weakened by modifiers. Negation words reverse or strongly reduce the prior polarity. For each opinion, the set of negation words consists of at most one negation valence shifter like "not" and its intensifiers like capitalization of the valence shifter. Each opinion component is associated with a score. By using a score function, the scores of opinion elements are assembled into an opinion-score.

This type of sentiment analysis can also well characterize user preferences, which can be helpful for improving the recommendation performance. Traditional recommender systems (RS) should consider factors of user's purchase records, product category, and geographical location. In [59], researchers proposed a sentiment-based rating prediction method (RPS) to improve prediction accuracy in recommender systems, including a social user sentimental measurement approach on calculating each user's sentiment on items or products, an important mechanism of considering a user's own sentimental attributes and interpersonal sentimental influences. Recommender system has made an accurate rating prediction with considering these factors, which is also a novel perspective of applying sentiment analysis techniques. Karyotis et al. [54] presented a novel emotion modeling methodology for incorporating human emotion into intelligent computer systems, which includes eliciting emotion information from users, a new representation of emotion (AV-AT model) with a genetically optimized adaptive fuzzy logic technique, and a framework for predicting and tracking user's affective trajectory over time. Compared to other existing machine learning approaches, fuzzy technique is evaluated in terms of its ability to model affective states.

### 3.1.3 Beyond polarity (entity, opinion holder, spatial information, and temporal information analysis)

The goal of researches in the area of fine-grained opinion analysis is to identify subjective expressions in text, along with their associated sources and targets. More specifically, fine-grained opinion analysis aims to identify types of opinion entities such as opinion holders [55], opinion expressions [61], opinion targets [139], aspects of a target [98], opinion sources [31]. In addition, we have redefined opinion as a six-tuple, which brings us news sub-task of sentiment analysis such as temporal and spatial information pattern analysis [101].

Opinion holder extraction is a typical task in this section. Several researches have investigated on opinion holder extraction based on conditional random fields (CRF) [26], maximum entropy (ME) model [57], convolution kernels [140], and dependency parser method [68], or automatic semantic role labeling (ASRL) [108]. Recently, Katiyar et al. [55] investigated the use of deep bidirectional LSTMs for joint extraction of opinion entities and the IS-FROM and IS-ABOUT relations that connect them, which is the first attempt at using a deep learning approach and explores LSTM-based models for the joint extraction of opinion entities and relations such as opinion expression, opinion target, and opinion holder.

Liao et al. [61] embedded both semantic structures and language expression features in the representation of opinion object and its corresponding attribute and proposed the FREERL framework. This framework is domain-independent and no manually crafted linguistic rules are needed. The framework can freely fuse any type of language expression features like statistical co-occurrence or dependency syntax into the structure-based embedding of each entity.

Wiegand et al. [139] presented a method for the new task of opinion holder and target extraction on opinion compounds. Opinion compounds are noun compounds whose head is an opinion noun. Researchers did not only examine features known to be effective for noun compound analysis, such as paraphrases and semantic classes of heads and modifiers, but also proposed novel features tailored, in which the paraphrases of holders and targets are jointly considered, noun heads are replaced by related verbs, inferencing between different compounds is allowed by a global head constraint, etc.

Poria et al. [98] presented the first deep learning approach for aspect extraction, which identifies opinion targets and detects the specific aspects of a service or a product that the opinion holder is either praising or complaining about. In their model, a 7-layer deep convolutional neural network is used to tag each word in opinionated sentences as either aspect or non-aspect word. For the same purpose, a set of linguistic patterns and the neural network combined with them are developed. The ensemble classifier coupled with a word-embedding model allows this approach to obtain significantly better accuracy than the state-of-the-art methods.

Deng et al. [31] improved recognizing sources of opinions based on a new categorization of opinions, i.e., non-participant opinion or participant opinion, in which transductive SVM is built to classify an opinion utilizing existing limited resources. The categorization information is then utilized by a probabilistic soft logic model to jointly recognize sources of the two types of opinions in a single model.

As people might create or share their content and interact with other users with many location-based services, usually in a real-time manner, there are some special factors needed to be considered such as "spatial factor," "temporal information." Zhao et al. [152] studied two types of user preferences, i.e., topical-region preference and category-aware topical-aspect preference. With a large amount of geo-tagged review data, researchers performed deep analysis of user preferences by proposing a unified probabilistic model to capture these two preferences simultaneously. This model is capable of capturing the interaction of different factors, including topical aspect, sentiment, and spatial information. Moreover, it can investigate whether people like an aspect of an object or whether people like a topical aspect of some objects in a region, which offer explanation for recommendations. Location-based sentiment analysis extracts, identifies, or characterizes the sentiment content of a "text unit," according to the location of origin of the text unit. In many application scenarios, it is often interesting, and in some cases critical, to discover patterns and trends based on geographical and/or temporal partitions, and keep track of how they will change overtime. Almatrafi et al. [4] studied the application of location-based sentiment analysis using Twitter for identifying trends and patterns toward the Indian general elections 2014. In this work, researchers illustrated how sentiments, both positive and negative, change from one location to the other. Paul et al. [90] investigated trends based on geographical and/or temporal partitions through US Election 2016. This work aims to discover sentiment on Twitter toward either the Democratic or the Republican Party at US county and state levels over any arbitrary temporal intervals,

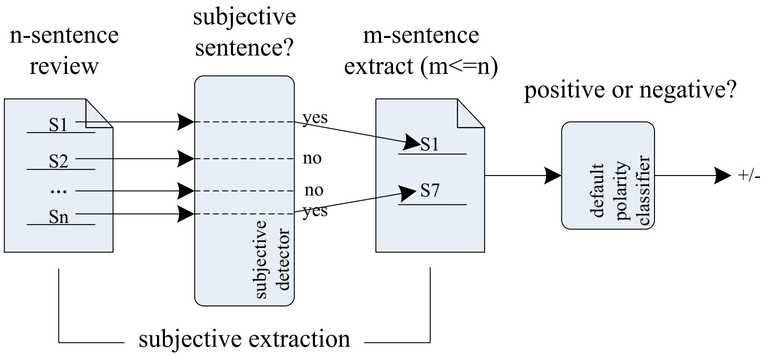



**Fig. 5** Subjective extraction for improving polarity classification

using a large collection of geo-tagged tweets from a period of 6 months leading up to the US Presidential Election in 2016.

### 3.1.4 Subjectivity or objectivity identification

Another research direction is subjectivity or objectivity identification. Although it seems to be an identification problem, in fact, is often solved as a classification problem. Pang et al. [88] defined this task as allocating a given text (usually a sentence) into one of two categories: subjective or objective. Sometimes, it can be more difficult than polarity classification problem. For example, the subjectivity of words or phrases usually depends on the context, but an objective chapter such as a news article may quote people's opinions which are subjective sentences. Moreover, as mentioned in [115], analysis results largely depend on the definition of subjectivity annotated texts. However, Pang et al. [85] showed that it could be helpful to remove objective sentences from a document before classifying its sentiment polarity (as shown in Fig. 5).

In recent years, Wiebe et al. [136] proposed a rule-based classifier, which classifies a sentence with strong subjective clues into a subjective category. This system is capable of learning new patterns of objective sentences using the information extraction system called AutoSlog-TS, which identifies patterns in terms of some syntactic templates. Barbosa et al. [12] proposed an approach automatically detecting sentiments on Tweets, which explores some traits of how tweets are written, and meta-information of the words composing these messages. Moreover, they utilized the sources of noisy labels as training data. The labels are provided by a few sentiment detection Web sites over Twitter data. Some Twitter-specific clues such as hashtags, retweets, links, emoticons, uppercase words, exclamation, and question marks are used for subjectivity classification. Benamara et al. [13] leveraged a subjectivity classification method at the segment level, which is more appropriate for discourse-based sentiment analysis. This approach automatically uses local and global context features to distinguish between implicit and explicit opinions, and between subjective non-evaluative and objective segments. Karimi et al. [53] proposed a language model-based structure that can facilitate dealing with subjectivity detection by introducing a difference based scoring formula, which leads to minimizing the effect of common topic relevant words in the process of distinguishing subjective documents from objective ones.

Banea et al. [11] conducted a case study seeking to assess subjectivity transfer across languages following sense aligned resources. In this model, the framework is able to predict subjectivity labeling for unseen senses by using either cross-lingual or multilingual training enhanced with bootstrapping. Fu et al. [36] proposed an adaptive semi-supervised extreme learning machine (ASELM) to solve the data imbalance between subjective and objective questions, which employs the different impacts on identification performance caused by the imbalanced data. This method introduces the unlabeled data and builds a model about the ratio between the number of labeled and unlabeled data based on Gaussian model, which is applied to automatically generate the constraint on the unlabeled data.

### 3.1.5 Feature or aspect-based sentiment analysis

A more fine-grained analysis task is known as the feature or aspect-based sentiment analysis [46], which refers to determining the opinions expressed on different aspects or features of an entity. In the context of sentiment analysis, aspect extraction is different from entity extraction. A feature or an aspect is an attribute or a component of an entity. This problem involves subproblems such as identifying relevant entities, extracting their features or aspects, and determining whether an opinion expressed on each aspect or feature is positive, negative, or neutral [64]. More detailed discussions about this problem also can be found in "*Sentiment Analysis and Subjectivity*" chapter of Handbook of Natural Language Processing [62]. Basically, aspect-based analysis needs to find explicit aspect expressions, which usually are nouns and noun phrases from the text of given domain. These nouns and phrases can be identified by a part-of-speech tagger (POS).

All of these types of methods are based on an assumption: When people comment on different aspects of an entity, the words that they use usually converge together. Researchers will be more interested in frequently used nouns and consider them as significant. This method is quite simple, but actually effective. On this basis, many researchers tried to improve the precision of this algorithm. Among all these work, a classic algorithm has been proposed in [96], which tried to remove redundancy noun phrases by calculating a pointwise mutual information score (PMI) between each discovered noun phrase and meronymy discriminator related to the entity class. Given a candidate aspect that is identified by using the frequency approach and a discriminator, the computed PMI score is denoted as:

$$\mathrm{PMI}(a, d) = \frac{\mathrm{hits}(a \wedge d)}{\mathrm{hits}(a)\mathrm{hits}(d)}$$

The algorithm needs to find the number of hits of individual terms and their co-occurrences. Once the value of PMI is too low, the candidate aspect may not be a component of the entity.

Zhu et al. [153] proposed a Cvalue measure-based method to extract multi-word aspects, which is also a frequency-based method by considering the frequency of multi-word term $t$, the length of term $t$, and other terms that contain term $t$. The Cvalue score of a multi-word term $t$ can be calculated:

$$C\ \mathrm{value}(t) = \begin{cases} \log(|t|) \times \mathrm{frequency}(t), & \text{if } t \text{ is not contained by any other terms} \\ \log(|t|)(\mathrm{frequency}(t) - \frac{1}{n(L)} \sum_{l \in L} \mathrm{frequency}(l)), & \text{otherwise} \end{cases}$$

where $|t|$ represents the number of words contained in $t$, frequency($t$) represents the frequency of occurrence of $t$ in the corpus, $L$ is the set of multi-word terms containing $t$, and n($L$) represents the number of terms in $S$. However, Cvalue is only capable of finding a set of candidates. It still needs to be refined using bootstrapping technique with a set of seed aspects. The co-occurrence of each candidate with the seeds is further considered in a refinement process.

Since opinion always has a target, their relationships can be exploited to extract aspect. This method is based on the assumption: Of those sentences with some sentiment words and no frequent aspect, the nearest noun or noun phrase to the sentiment word is extracted. Such dependency relations were further generalized into double propagation for extracting sentiment words or aspects, at the same time, exploiting certain syntactic relations between sentiments and targets, and a small set of seed sentiment words [104]. This method is also based on bootstrapping technique.

Besides, two kinds of aspects that have been introduced in Sect. 2 were identified: explicit and implicit aspects. Limited research has been done on mapping implicit aspects to explicit aspects, although explicit aspect extraction has been studied extensively. Hai et al. [43] utilized two-phase co-occurrence association rule mining to identify implicit features. Firstly, in rule generation phase, for the occurrence of each opinion word in an explicit sentence in the corpus, they mined a significant set of association rules of the form [opinion-word, explicit-feature] from a co-occurrence matrix. Secondly, in rule application phase, they clustered rule consequents (explicit features) to generate more robust rules for each opinion word mentioned above. Given a new opinion word with no explicit feature, they then searched a matched list of robust rules, among which the rule containing the feature cluster with the highest frequency weight is fired, and accordingly, they assigned the representative word of the cluster as the final identified implicit feature.

Recently, Penalver-Martinez et al. [91] proposed an innovative method that takes advantage of new Semantic Web-guided solutions to enhance the results obtained with traditional sentiment analysis processes and natural language processing techniques. This method improves feature-based opinion mining by using ontologies at the feature selection stage, providing a new vector analysis-based method for sentiment analysis. Yang et al. [149] proposed a combined approach, which integrates local context information (LCI) and global context information (GCI) to extract and rank features based on feature score and frequency, in which LCI uses the HITS algorithm with considering direct links between opinion words and nouns in a sentence, GCI uses SimRank with considering the LCI (direct) links and indirect links in the corpus. Manek et al. [70] proposed a statistical method using weight by Gini index method for feature selection in sentiment analysis with SVM classifier.

### 3.2 Granularity-oriented sentiment analysis

The rapid growth in the scale of existing applications or a surge in demand for particular services may result in additional needs for sentiment analysis at different granularity. Current researches broadly handle this issue by classifying the task of sentiment analysis into three levels: document level, sentence level, and word level. That is, models built for multi-granularity sentiment analysis assume fully labeled corpus at the fine-grained level or coarse-grained level or both. A huge amount of online reviews is not fully labeled at any particular level; instead, they are partially labeled at both levels. Actual usage patterns of many real-world application granularities vary with strategy, and most of the time, in unpredictable ways. Therefore, as we stated earlier, unexpected granularity can potentially overburden a single task and lead to unreliable and interrupted services.

#### 3.2.1 Document-level sentiment analysis

The major challenges on document-level sentiment analysis are cross-domain sentiment analysis and cross-language sentiment analysis. It has been shown that specific domainoriented sentiment analysis has achieved remarkable accuracy, which is highly sensitive to the domain. The feature vector used in these tasks contains a bag of words, which should be specific to a particular domain and are limited. Sentiment classifier is applied as it is costly to annotate data for each new domain. Spectral feature alignment, structural correspondence learning, and sentiment-sensitive thesaurus are three classical techniques. They are different in terms of feature vector expansion, words relatedness measurement, and finally classifier used for classification.

Many methods used for cross-domain classification usually utilize labeled or unlabeled data or both of them. Hence, the techniques give different results for different domains as well as for different purposes. Bollegala et al. [15] developed a technique which uses sentiment-sensitive thesaurus (SST) for performing cross-domain sentiment analysis. They proposed a cross-domain sentiment classifier using an automatically extracted sentiment-sensitive thesaurus. To handle the mismatch between features in cross-domain sentiment classification, they utilized labeled data from multiple source domains and unlabeled data from target domains to compute the relatedness of features and construct a sentiment-sensitive thesaurus. Then the created thesaurus is used to expand feature vectors during training and testing process for a binary classifier. A relevant subset of the features is selected using L1 regularization. Pan et al. [84] first proposed spectral feature alignment (SFA). In SFA model, the mutual information between a feature and a domain label is used to classify features into the domain-specific or the domain-independent. SFA represents a review by considering both unigrams and bigrams as features. Next, a bipartite graph between domain-independent and domain-specific features is constructed. In this graph, an edge is formed between a domain-independent and a domain-specific feature, when those two features co-occur in any feature vector. After that, feature clusters are identified by conducting spectral clustering, and a binary classifier based on these feature clusters is trained for positive and negative sentiment classification. The structural correspondence learning (SCL) is proposed in [14]. It selects pivots using the mutual information between a feature (unigrams or bigrams) and the domain label. Next, linear classifiers are learned for predicting the existence of those pivots. The learned weight vectors are arranged in a matrix, and singular value decomposition (SVD) is performed to reduce the number of dimensions. Finally, this lower-dimensional matrix is used as project features for training a binary sentiment classifier.

Another document-level sentiment analysis is cross-language sentiment analysis, which has been studied by several researchers. Most of them focus on sentiment classification at the document level. Wan [132] proposed an approach that first translates Chinese reviews into English reviews by machine translation services and then identifies the sentiment polarity of English reviews by directly leveraging English resources. Furthermore, their approach performs sentiment analysis for both Chinese reviews and English reviews and then uses ensemble methods to combine the individual analysis results. Wan [133] proposed another co-training method which leverages an available English corpus for Chinese sentiment classification by using the English corpus as training data. Besides, limited cross-language sentiment analysis has been done at the aspect level [42]. Xia et al. [146] proposed a three-stage cascade model for the polarity shift problem in the context of document-level sentiment classification, in which each document is split into a set of sub-sentences, and a hybrid model is built up with employing rules and statistical methods to detect explicit and implicit polarity shifts. Then, a polarity shift elimination method is used to remove polarity shift in negations. Finally, different types of polarity shifts are used to train base classifiers. Li et al. [60] proposed a cross-lingual structural correspondence learning SCL based on the distributed representation of words; it can learn meaningful one-to-many mappings for pivot words using large amounts of monolingual data and a small dictionary.

### 3.2.2 Sentence-level sentiment analysis

Sometimes document-level sentiment analysis is too coarse for some special purposes. A lot of early work at sentence-level analysis focuses on identifying subjective sentences. But there will be complex tasks such as dealing with conditional sentences or dealing with sarcastic sentences. In such cases, sentence-level sentiment analysis is desirable.

A conditional sentence consists of the condition clause and the consequence clause, which describes implications or hypothetical situations and their consequences. Sarcasm is a more sophisticated form of speech act, in which the speakers say the opposite of what they really mean. Tsur et al. [124] presented a novel semi-supervised algorithm called SASI for sarcasm identification that recognizes sarcastic sentences in product reviews. SASI has two stages: semi-supervised pattern acquisition and sarcasm classification. For the first stage, they used a small set of labeled sentences as seeds and expanded this seed set through the web search. This enriched training set was then used for learning and classification. González-Ibánez et al. [41] tried to solve the problem of automatically detecting sarcasm in Tweets. They used a corpus annotated by the tweeters themselves as their gold standard and relied on the judgments of tweets because of the relatively poor performance of human coders at this task. They semiautomatically cleaned the corpus to address concerns about corpus noisiness resulted from previous work and explored the contribution of linguistic and pragmatic features of tweets to the automatic separation of sarcastic messages from positive and negative ones. Wu et al. [144] proposed an approach for sentence-level sentiment classification without labeling sentence. It is a unified framework to incorporate two types of weak supervision with document-level and word-level sentiment labels, to learn the sentence-level sentiment classifier. Recently, recursive autoencoder (RAE) methods have been proposed for sentence-level sentiment analysis. Fu et al. [37] proposed a semi-supervised method CHL-PRAE which combines HowNet lexicon to train phrase recursive autoencoders. CHL-PRAE model calculates the sentiment orientation of each node with the HowNet lexicon acting as sentiment and conducts bidirectional training to capture global information. Appel et al. [5] presented using natural language processing (NLP) essential techniques, a sentiment lexicon enhanced with the assistance of SentiWordNet, and fuzzy sets to estimate the semantic orientation polarity and its intensity for sentences. As using deep learning models to solve sentiment analysis of sentences is still a challenging task, Fu et al. [38] proposed a model using rhetorical structure theory (RST) for text parsing. This model builds long short-term memory (LSTM) network on RST parse structure to make full use of LSTM structural characteristics for automatically enhancing the nucleus information and filtering the satellite information. This approach is capable of making the representations concerning the relations between segments of text, which can improve text semantic representations.

### 3.2.3 Word-level sentiment analysis

Document-level analysis focuses on distinguishing the entire document from subjective or objective, and positive or negative, while sentence-level analysis is more effective than document-level analysis, because a document contains both subjective and objective sentences. While word is the basic unit of language, the polarity of a word is closely related to the subjectivity of corresponding sentence or document. There exists a huge possibility that a sentence containing an adjective is a subjective sentence. In addition, choice of word for expression reflects not only the individual's demographic characteristic such as gender, age, but also reflects its motivation, personality, social status, and other psychological or social traits. Therefore, word is the basis of text sentiment analysis. At present, the commonly used methods include: natural language processing technology-based approach and machine learning-based approach. For sentiment analysis of micro-blog text, most researchers suggest that term matching-based technique should be adopted. The emotional term is the link between the emotional orientation of the text and the single word. Each word can be regarded as the collection of certain kinds of viewpoint information, which is a clue to the emotion and subjectivity of the text. There exist several studies on the emotional orientation of word based on using sentiment lexicon. We will discuss typical sentiment lexicons, the bases of word-level sentiment analysis in Sect. 4.2.

### 3.3 Methodology-oriented sentiment analysis

The existing methods can be roughly categorized into three groups: supervised learning, semi-supervised learning, and unsupervised learning. The idea of supervised learning is to study the features from both positive and negative examples. All of the existing approaches consist of a system of reading annotated corpus, memorizing lists of entities, and creating disambiguation rules. The shortcoming of supervised learning is the requirement of a large annotated corpus, which leads to two alternative learning methods: semi-supervised learning and unsupervised learning. The main technique for unsupervised learning "bootstrapping" also involves some degree of supervision such as a set of seeds for starting the learning process. The typical approach in unsupervised learning is clustering such as using a dictionary to compile sentiment words. Basically, these techniques rely on resources, lexical patterns, or statistics calculated on a large unannotated corpus.

#### 3.3.1 Supervised learning approach

In the analysis of emotion orientation, an important task is sentiment classification. The objects for classifying are some subjective factors, which is different from traditional text classification based on themes. Supervised learning considers that the sentiment classification is a standard statistical classification that contains a large number of labeled instances. In the following, several widely used classification algorithms in sentiment analysis are discussed.

*Rule-based algorithm* Rule-based classification can be used to any classification scheme that makes use of IF–THEN rules for class prediction. Asghar et al. [7] analyzed the semantic orientation of words by categorizing them into +ive and -ive classes to identify and classify emoticons, modifiers, general-purpose, and domain-specific words expressed in the public's feedback about the products, which solves the problem of previous studies becoming less efficient due to data sparseness. Lexicon enhancing sentiment analysis-based on rule-based classification scheme, in this work, is an effective alternative approach for improving sentiment classification.

*Decision tree algorithm* Decision tree is a decision support tool that uses a tree-like graph or model of decisions and their possible consequences. Phu et al. [94] proposed a model for the English document-level emotional classification, which presents a new model by using an ID3 algorithm of a decision tree to classify semantics (positive, negative, and neutral) for the English documents. The semantic classification of this model is based on many rules which are generated by applying the ID3 algorithm for dataset training.

*Support vector machine (SVM)* SVM is supervised learning model with associated learning algorithms that analyze data used for classification and regression analysis. Liu et al. [66] presented a multi-class sentiment classification method based on an improved one-vs-one (OVO) strategy and SVM algorithm. For improved OVO strategy, the relative competence weight of each binary classifier is determined according to the $K$ nearest neighbors and the class center of each class in the training sample set concerning the binary classifier. The important features for multi-class sentiment classification are selected using the information gain (IG) algorithm; a binary SVM classifier is then trained on feature vectors training of each pair of sentiment classes.

*Artificial neural networks (ANNs)* ANNs is widely used for classification task, which is inspired by biological neural networks of human's central nervous system. In the area of sentiment analysis, document-level sentiment classification remains a problem of encoding the intrinsic relations between sentences in the semantic meaning of a document. To address this challenge, Tang et al. [118] introduced a neural network model to learn vector-based document representation in a unified, bottom-up fashion, which learns sentence representation with convolutional neural network or long short-term memory, and then encodes the semantics of sentences, and their relations are adaptively in document representation with gated recurrent neural network.

*Deep learning* Deep learning is part of a broader family of machine learning methods based on learning data representations, as opposed to task-specific algorithms. Learning can be supervised, partially supervised, or unsupervised. Moreover, deep learning plays an important role in big data analysis. To date, various types of deep learning models have been applied for sentiment analysis such as deep belief, deep neural networks, deep convolution neural networks. Tang et al. [119] developed a deep learning system called Coooolll for message-level Twitter sentiment classification. Coooolll is built on a supervised learning framework by concatenating the sentiment-specific word embedding (SSWE) features with the state-of-the-art hand-crafted features, in which SSWE is trained from large-scale tweets collected by positive and negative emotions, without any manual annotation. The advantage of this system is that Coooolll can be easily re-implemented with the publicly available sentiment-specific word embedding. As the fact that most existing algorithms for learning continuous word representations typically only model the syntactic context of words but ignore the sentiment of text, Tang et al. [120] presented a method of learning word embedding for Twitter sentiment classification, which addresses this issue by learning sentiment specific word embedding (SSWE), which encodes sentiment information in the continuous representation of words. Moreover, the researchers developed three neural networks to effectively incorporate the supervision from sentiment polarity of text in their loss functions.

*Ensemble* Ensemble methods use multiple learning algorithms to obtain better performance rather than use any of the constituent learning algorithms alone. To improve the performance of deep learning techniques by integrating them with traditional surface approaches based on manually extracted features, Araque et al. [6] developed a deep learning-based sentiment classifier using a word-embedding model and a linear machine learning algorithm, which serves as a baseline to compare to subsequent results. Then, the researchers proposed two ensemble techniques aggregating this baseline classifier with other surface classifiers widely used in sentiment analysis. In this work, combining both surface and deep features to merge information with using different methods is also presented. Akhtar et al. [2] presented a cascaded framework of feature selection and classifier ensemble using particle swarm optimization (PSO) for aspect-based sentiment analysis. Aspect-based sentiment analysis is performed by two steps: aspect term extraction and sentiment classification. This pruned and compact set of features perform better than the baseline model that makes use of the complete set of features for aspect term extraction and sentiment classification.

*Regression* Regression analysis is widely used for prediction and forecasting, and it has substantial overlap with the field of machine learning. Regression analysis is also used to understand, to what extent the independent variables are related to the dependent variables and to explore the forms of these relationships. The text on social media is usually informal and short in length, which may not always express the sentiment clearly. Therefore, multiple raters may assign different sentiments to a tweet. So rather than employing majority voting that ignores the strength of sentiments, the annotation can be enriched with a confidence score assigned for each sentiment. Ertugrul et al. [34] analyzed the effect of using regression on confidence scores in sentiment analysis using Turkish tweets, in which hand-crafted features including lexical features, emoticons and sentiment scores are extracted. Then, the researchers employed word embedding of tweets for regression and classification. The findings in this work reveal that employing regression on confidence scores slightly improves sentiment classification accuracy. Moreover, combining word embedding with hand-crafted features is capable of reducing the feature dimensionality and outperforms alternative feature combinations.

*Statistical model* Above all else, there are several statistical models widely used in sentiment analysis task such as hidden Markov model and Gaussian mixture model, which will be further discussed.

- *Hidden Markov model (HMM)* Kang et al. [52] presented a text-based hidden Markov models (TextHMMs)-based method, for text classification that uses a sequence of words in training texts instead of a predefined sentiment lexicon, which seeks to learn text patterns representing sentiment through ensemble TextHMMs. In this method, the researchers defined hidden variables in TextHMMs by semantic cluster information in consideration of the co-occurrence of words and thus calculated the sentiment orientation of sentences by fitted TextHMMs. An ensemble of TextHMM-based classifiers is applied to reflect diverse patterns.
- *Gaussian mixture model (GMM)* Fattah [35] proposed multiple classifiers for sentiment analysis, in which GMM is used for clustering data by allocating query data points to the multivariate normal components. Assigning data points to clusters is termed as hard clustering. Power of GMM clustering can be noted because it uses soft clustering techniques, assigning score to data point for each cluster.

### 3.3.2 Unsupervised learning approach

Clustering is a process of classifying data into different classes or clusters, so the objects in the same cluster have the extremely high similarity, while the objects of different clusters have very high diversity. From the point of view of machine learning, clusters correspond to hidden patterns, while clustering is the unsupervised learning process of searching clusters. And unlike classification, unsupervised learning is not dependent on a predefined category or training instances with labels. Clustering algorithms are capable of automatically determining latent patterns, which is observing learning, not sample learning. Further information regarding clustering algorithms for sentiment analysis is discussed.

*Partitional algorithm* Partitional algorithm decomposes a data set into a set of disjoint clusters. Given a data set of $N$ points, a partitioning method constructs $K$ ($N \geq K$) partitions of the data, with each partition representing a cluster. Riaz et al. [107] performed sentiment analysis on the customer review *real-world data* at phrase level to find out customer preference by analyzing subjective expressions. The researchers calculated the strength of sentiment words to find out the intensity of each expression and applied k-means clustering for placing the words in various clusters based on their intensity. Suresh [116] presented a fuzzy clustering model to analyze Twitter feeds regarding the sentiments of a particular brand using the real dataset collected over a period of 1 year. Phu et al. [93] proposed a model for big data sentiment classification in the parallel network environment. The proposed model uses the fuzzy c-means (FCM) method for English sentiment classification with Hadoop MAP(M)/REDUCE(R) in Cloudera.

*Hierarchical algorithm* Hierarchical clustering groups data over a variety of scales by creating a cluster tree or *dendrogram*. The tree is not a single set of clusters, but rather a multilevel hierarchy, where clusters at one level are joined as clusters at the next level. This allows people to decide the level or scale of clustering that is most appropriate for particular application. Wu et al. [145] designed a visual analysis system OpinionFlow, which can visually explore and trace opinion diffusion among users on Twitter. Especially, the proposed system uses a hierarchical topic structure built by Bayesian rose tree (BRT), which detects a multibranch hierarchy of topics from a large number of tweets and displays the topic hierarchy by using a stacked topic tree. The tree is linked to opinion flow for facilitating analysis of opinion diffusion on different topics. Chen et al. [23] developed a method called TCOL-Miner, which integrates network structure and leadership quality analysis methods to efficiently find opinion leaders in a large social network. In this work, the two-stage clustering methods are utilized to significantly reduce the impact of overlapping problem in social network. Vaziripour et al. [128] classified the sentiment of individual tweets to determine their changing sentiments over time toward a number of trending political topics. In their work, a hierarchical representation of the vocabulary is created by tracking the merging process of clusters, which is then used as features.

*Density-based algorithm* Given a set of points in some space, a density-based clustering algorithm groups together the points that are closely packed together, marking outliers in low-density regions. Hassan et al. [45] proposed a framework for sentiment analysis by using a combination of density-based clustering and distance based clustering. This proposed framework gives a high level of accuracy and it is useful for performing types of attacks prediction. It gives analyst a feasible way of dealing with terrorism attacks in advance.

### 3.3.3 Semi-supervised learning approach

Semi-supervised learning (SSL) techniques are capable of using unlabeled data in their training processes and improving classification in applications when labeled data are scarce. From this perspective, SSL techniques are of great practical value because a balance between supervised and unsupervised learning is found. Basically, there exist several classic categories of semi-supervised learning approaches: graph-based, wrapper-based, topic-based methods.

*Graph-based method* In the area of graph-based methods, semi-supervised learning can manage either sentence or document. Lu et al. [69] proposed a generic framework for incorporating social context information by adding regularization constraints to the text-based predictor. Tan et al. [117] used social network structures to help sentiment analysis by incorporating link information. These links can correspond to attention, such as when a Twitter user wants to pay attention to another's status updates, or homophily, where people who know each other are connected. Pozzi et al. [103] proposed a framework that estimates user polarities about a given topic by combining post contents and weighted approval relations, which are intended to better represent the contagion on social networks.

*Wrapper-based method* The wrapper-based method labels a certain number of unlabeled instances using the decision function, which is then learned and incorporated into the training data iteratively. The well-known types of wrapper-based method include self-training [67, 151] and co-training [50, 65, 150]. Self-training starts with a supervised learner that is trained on available labeled data and then iterates several times. Instead of using a single supervised learner, co-training uses two learners, which operate on different feature sets from independent views and build more labeled data in turn. Iosifidis et al. [50] presented how to annotate large-scale collections with sentiment labels, which uses self-learning and co-training, batch- and stream-processing. This work solves the problem of lacking large-scale labeled datasets that span large periods of time, which is especially important for stream mining research.

*Topic-based method* Topic information has been applied in different domains of sentiment analysis. Si et al. [112] utilized a continuous Dirichlet process mixture model to learn the daily topic set from Twitter to help predicting the stock market. Xiang et al. [147] presented a topic-based sentiment mixture model with topic-specific data to improve sentiment analysis on Twitter data, which generates topic information through topic modeling based on the implementation of latent Dirichlet allocation (LDA).

### 3.4 Summary of different sentiment analysis studies

Analyzing the sentiments contained in the subjective statement is one of the most active research domains. This research focuses on the opinions, emotions, and subjectivity of the information being processed. In all kinds of emotional analysis, polarity classification is the main task of emotional analysis, which is to analyze the positive and negative feelings of emotion, namely to analyze and judge the positive and negative meanings. Polarity classification has great commercial value and public service value, which will be very time-consuming, high cost, if such amount of research is performed artificially. Sentiment analysis, especially the polarity classification research, perfectly meets the potential huge demand for the customer. However, simple dichotomy point of sentiment analysis ignores the richness and diversity of human emotions, and more emotional types, dimensions, or granularities. Multiple-level sentiment analysis has been developed to address this problem to a certain extent. Moreover, scale system calculates the positive and negative emotional strength scores of the text, which puts sentiment into an interval. This approach has promoted the understanding of textual emotions to a more intelligent level. Subjectivity or objectivity identification is sometimes more difficult than the problem of polarity classification. Subjective words and phrases may be based on the context, while an objective document may contain subjective statements. Feature or aspect-based sentiment analysis determines the opinions or sentiments expressed on different features or aspects of entities, which involves the extraction of entity, opinion holder, etc. Basically, almost all task orientation sentiment analysis studies have the same limitation, which ignores the richness and diversity of human emotions, and more emotional types, dimensions or granularities, mainly based on the positive and negative direction of sentiment that should be further explored with considering more sentiment factors.

The rapid growth in the scale of existing applications or a surge in demand for particular services may result in additional needs for sentiment analysis at different granularity. Current researches broadly handle this issue by classifying the task of sentiment analysis into three levels: document level, sentence level, and word level. That is, models built for multi-granularity sentiment analysis assume fully labeled corpus at fine-grained level or coarse-grained level or both. We will discuss both the advantages and disadvantages from multi-granularity in this section. Specifically, we show the advantages and disadvantages of cross-domain sentiment analysis and cross-language sentiment analysis on document level, and analysis on sentence level and word level.

In recent years, a large amount of studies has been conducted on data mining and machine learning that bring up both advantages and disadvantages for sentiment analysis research. Machine learning is capable of extracting models and patterns in complicated data sets, which includes supervised learning, unsupervised learning, and semi-supervised learning methods. Supervised learning algorithms can produce sentiment classification by training model with labeled data, while unsupervised learning algorithms perform this procedure with unlabeled data and get hidden structure from unlabeled data. Semi-supervised learning takes advantage of training model with using unlabeled data and can improve classification when labeled data are scarce. All the algorithms in three types of machine learning techniques are analyzed separately in Table 4, including both pros and cons.

## 4 Available datasets and tools

In this section, we introduce benchmark datasets, tools and lexicons for sentiment analysis. We aim to identify the most notable and promising ones rather than offer an exhaustive list.

### 4.1 Benchmark datasets

*Facebook and Twitter* As two main online social media, they were launched in 2004 and 2006, respectively. Since then, Twitter service rapidly gained worldwide popularity with more than 500 million users, out of which more than 284 million are active users as of September 2014. Facebook had over 1.3 billion active users as of June 2014. A large volume of data that users submit to the service should be the best data resource. Ideally, researchers also like to conduct research with these direct data available on social network media. For the reasons of user privacy and corporate policies, Facebook's Terms of Service prohibit the distribution of any content to third parties; therefore, any collected data is usually prevented from being given to the researchers. Nonetheless, there are studies providing significant insights because of the massive scale of the analysis that is typically conducted by employees of such services [58, 129]. Recently, Ortigosa et al. [80] presented a new method for sentiment analysis in Facebook. They have asked for the user's permission by following a strict policy-oriented toward protecting the user's privacy. Therefore, the data accessed by their application depends on the level of user authorization. On the other hand, the majority of the contents in Twitter can be accessed since most of these contents are public. The contents from Twitter have been used by researchers [16, 17, 81, 82]. The Twitter API provides easy access to the site's content through programming. Therefore, such data can be collected and analyzed by individual researcher.

*Digg* Paltoglou et al. [83] presented another two publicly available datasets extracted from BBC discussion forums and the Digg. The former is a complete crawl of a subsection of the BBC message boards spanning 4 years, from June 2005, when the forum went online, until June 2009, when part of it was shut down. It provides 97,946 separate discussion threads and 2592,745 distinct comments about ethical, religious, and news-related issues. The second is a complete crawl of the Digg Web site spanning the months of February, March, and April in 2009, which provides 1195,808 submitted stories, 1646,153 individual comments, and 14,376,742 approvals (users are allowed to publicly approve or disapprove submitted stories and comments, processes, respectively, called "digging" and "burying"). Both datasets are available from the official CyberEmotion project web site. They have been extensively used in researches [24, 25, 74]. Corresponding details about how to obtain these datasets can be found in [83].

*Cornell movie review* As research in sentiment analysis has mainly focused on datasets extracted from review sites, many prominent datasets such as movie review and general

**Table 4** Summary of different sentiment analysis studies

| Taxonomy | Categories | Pros | Cons |
|---|---|---|---|
| Task-oriented sentiment analysis | Polarity classification | Provide a two-class classification with positive and negative opinions | Ignore the richness and diversity of human emotions, and more emotional types, dimensions or granularities |
| | The level of valence or arousal at specific scale | Provide opinion classification at specific scale | Ignore the richness and diversity of human emotions, and more emotional types, dimensions or granularities |
| | Beyond polarity | Provide specific details of opinion | Depend on methods |
| | Subjectivity or objectivity identification | Removing objective sentences from a document is helpful for improving the performance of sentiment analysis | Depend on methods |
| | Feature or aspect-based sentiment analysis | Provide opinions on different aspects of what object has been evaluated | Depend on methods |
| Granularity-oriented sentiment analysis | Document-level sentiment analysis | Provide an overall opinion | Ignore details of specific aspect; not easily applicable to non-reviews |
| | Sentence level sentiment analysis | Provide an intermediate opinion | Ignore details of specific aspect<br>Cannot deal with opinions in comparative sentences<br>Complex sentence has different sentiments on different targets |
| | Word-level sentiment analysis | Provide a local opinion | Cannot detect linguistic nuances such as irony or sarcasm<br>Fail to capture the context of language |

**Table 4** continued

| Taxonomy | Categories | Pros | Cons |
|---|---|---|---|
| Methodology-oriented sentiment analysis | Supervised learning approach | Rule based: Generate descriptive models that are easier to interpret<br>Decision tree: easy to implement<br>SVM: high accuracy<br>Neural network: handle noisy data<br>Deep learning: process large dataset; deal with deep architecture; support multi-task learning<br>Ensemble: overcome overfitting<br>Generalization; predictive; high performance<br>Regression: computational cost is not high; easy to understand and implement<br>Statistical model: depend on methods | Rule based: Suitable for handling imbalanced data<br>Decision Tree: Space limitation, overfitting<br>SVM: Slow training; computationally expensive<br>Neural network: slow; computationally<br>expensive; black-box models; low accuracy<br>Deep learning: Difficult to interpret; computationally expensive<br>Ensemble: hard to analyze; computationally expensive<br>Regression: easy to produce underfitting<br>Statistical model: depend on methods |
| | Semi-supervised learning approach | Capable of using unlabeled data in their training processes and improving classification in applications when labeled data are scarce | Graph-based methods: Propagate labels to unlabeled data, while the effectiveness of label propagation process depends on the computation of similarities among the data instances<br>Wrapper-based methods: The method only works if the highest confidence predictions are effectively correct<br>Topic-based method: Depend on the choice of the confidence threshold and the inference of the number of topics (clusters) |
| | Unsupervised learning approach | Partitioning: Handle large datasets, fast, simple<br>Hierarchical: visualization<br>Capability<br>Density-based: detect outliers and arbitrary shapes; handle nonstatic and complex data | Partitioning: High sensitivity to initialization, noise and outliers<br>Hierarchical: Poor visualization for large data; slow; use huge amount of memory; low accuracy<br>Density-based: not well for large datasets; slow; tricky parameter selection |

product reviews were conducted. Pang et al. [89] selected movie reviews where the author rating was expressed either with stars or some numerical value. Moreover, they limited fewer than 20 reviews per author per sentiment category, yielding a corpus of 752 negative and 1301 positive reviews, with a total of 144 reviewers represented.

*Amazon product reviews* Blitzer et al. [14] conducted a dataset by selecting the Amazon product reviews for books, DVDs, electronics and kitchen applications, where 1000 positive and 1000 negative examples are involved for each domain. Hu et al. [47] conducted experiments on the customer reviews of five electronics products: 2 digital cameras, 1 DVD player, 1 mp3 player, and 1 cellular phone. The two Web sites that they collected the reviews are www.amazon.com and www.CNETdownload.com. Products in these sites have a large number of reviews. Each of the reviews includes a text review and a title. Additional information includes date, time, author name, and location (for Amazon reviews), and ratings are also available. Besides, almost every e-commerce Web site has reviews for their product. Some of the famous review sites have made their data available for research purpose.

*MPQA* Multi-perspective question answering (MPQA) is a dataset containing news articles and other text documents manually annotated for opinions and other private states (e.g., opinions, emotions, sentiments, speculations, evaluations, and other private states), which consists of 692 documents (15,802 sentences). In [137, 141], more details about producing MPQA are described.

*ICWSM Spinn3r datasets* There are two versions of ICWSM Spinn3r Datasets [18], i.e., version 2009 and a more recent version 2011. Both of the two versions include several million blog posts crawled by Spinn3r. The former dataset comprises of 44 million blog posts published between 08/2008 and 10/2008. The time span includes a number of significant worldwide news events, such as the U.S. presidential nominating conventions, the Olympic Games, the beginning of the credit crisis. The version 2011 [19] comprises of 386 million blog posts, new articles, forum posts, and social media communication published between 13/01/2011 and 14/02/2011. In this span of time, there exist significant events related to the Arab Spring, including the Tunisian revolution, the Egyptians protests, etc. They have been used in sentiment analysis studies and other corresponding research in [22].

*TSentiment15* It is a collection of 228 million tweets without retweets and 275 million tweets with retweets, collected from Twitter using its public streaming API which spans the whole year of 2015.

In Table 5, we summarize the research instances based on the aforementioned datasets. We list at least one paper for each dataset, the research content of the corresponding paper, and the type of these work. For researchers who try to find a proper dataset, they can take these instances as references. Details of these approaches have already been given throughout the text and will not be repeated again.

### 4.2 Tools and lexicons

Document-level analysis focuses on distinguishing the entire document from subjective to objective, or from positive to negative. Sentence-level analysis is more effective than documentation, because a document contains both subjective and objective statements. And word is the basic unit of language; the polarity of the word is closely related to the subjectivity of a sentence or document. Word-based matching techniques judge emotional tendency by matching the words contained in the text and emotional words in the emotional lexicon. Emotional lexicon is undoubtedly the foundation of this approach.

**Table 5** Research instances based on different datasets

| Datasets | References | Applications | Availability |
|---|---|---|---|
| Facebook | [54, 80] | Get information about the users' sentiment polarity (positive, neutral, or negative) according to the messages they write | Facebook's sentiment analysis API<br>The data accessed by their application depends on the level of user authorization |
| Twitter | [16, 17] | Utilize Twitter mood to predict the stock market | The Twitter API (https://dev.twitter.com/) |
| | [81] | Build a sentiment classifier to determine positive, negative and neutral sentiments | |
| | [82] | The actual levels of rainfall in a given location and time are inferred from the content of tweets. Infer regional influenza-like illness rates in the effort of detecting timely an emerging epidemic disease | |
| Digg | [24] | The emotions of a community member may influence the emotions of others | http://www.digg.com/ |
| BBC Discussion Forums | [25] | A majority of posts contain negative sentiments that sustain online discussions | http://www.bbc.co.uk/messageboards/ |
| Cornell Movie Reviews | [89] | Sentiment classification | Available online (http://cs.cornell.edu/people/pabo/movie-review-data/) |
| Amazon Product Reviews | [14] | Sentiment classification | Available online (http://www.cs.jhu.edu/mdredze/datasets/sentiment) |

**Table 5** continued

| Datasets | References | Applications | Availability |
|---|---|---|---|
| CNET Product Reviews | [47] | Mine the specific features of the product that customers have opinions on and analyze the opinions are positive or negative | www.CNETdownload.com |
| Restaurant Reviews from Yelp | [51, 59] | Aspect discovery; senti-aspect discovery; sentiment classification | www.yelp.com |
| Product Reviews | [95] | Sentiment classification | www.reviewcentre.com |
| MPQA | [49] | Subjectivity classification | Available online http://mpqa.cs.pitt.edu/ |
| ICWSM Spinn3r Datasets | [22] | The spread of media content through blogs | version 2009 (http://www.icwsm.org/2009/data/index.s html) and a more recent version 2011(http://icwsm.or g/data/index.php) |
| TSentiment15 | [50] | How to annotate large-scale collections with sentiment labels | Available online (https://l3 s.de/ %7Eiosifidis/TSentiment15/) |

There are lexicons and tools that are freely available for identifying subjective content or named-entity, performing various levels of affective analysis, and generating multi-dimensional mood, etc. Sentiment lexicon is an important connection between textual content and individual word, in which each word can be treated as a collection of opinion information. Some sentiment lexicons can be used to classify emotional orientation based on the corpus, where semantic collocations such as coordinating relation, progressive relation, and transitive relation divide adjectives into positive and negative. The shortcoming of this method is that algorithm rule is limited to adjectives and adverbs. On the other hand, the polarity of a new word only can be judged by utilizing conjunctions. Other methods utilize the synonyms or antonyms of seed words, the definition, and annotations of words to find out related information.

Lexicon-based approaches can be applied without any training or human labor. There are two common ways to construct sentiment lexicon: existing lexicon-based construction method that artificially creates lexicon, and automatical or semiautomatical construction method. *General Inquirer (GI)* is regarded as the earliest one emotional thesaurus and affective analysis program, in which emotional words derives from Harvard IV-4 Dictionary and Lasswell's Dictionary. GI labels each word with polarity, strength, parts of speech, etc, and can be used in sentiment analysis task flexibly. General Inquirer takes a pre-set list of positive and negative words and analyzes the polarity of documents based on the prevalence of words from each category. The existing lexicon-based method is time-consuming and only includes fewer vocabularies and certain deviations. With the development of techniques, automatical or semiautomatical construction of an emotion lexicon becomes the dominant way. *SentiWordNet* is a lexical resource for opinion mining, which assigns each synset of WordNet three sentiment scores: positivity, negativity, objectivity. The current official version of SentiWordNet is 3.0, which is based on WordNet 3.0. *OpinionFinder* automatically identifies subjective sentences as well as various aspects of subjectivity within sentences, including sources of opinion, direct subjective expressions and speech events, and sentiment expressions. Two versions of OpinionFinder can be found online. The OpinionFinder system seeks to address deviation problem by analyzing polarity in the context, rather than simply using single-word triggers. *National Taiwan University Sentiment Dictionary (NTUSD)* is constructed based on the Chinese translation of GI and Chinese network sentiment dictionary (CNSD), including 2812 positive words and 8276 negative words. *Bing Liu's Opinion Lexicon* includes positive words and negative 4783 words.

In above-mentioned lexicons, emotion words are simply classified into positive and negative based on dichotomy theory, which ignores the richness and diversity of human emotion. While constructing a lexicon based on the theory of emotional structure, subdivide human emotion into more dimensions. *SentiStrength* can estimate the strength of positive and negative sentiment in short texts, even for informal language. It has human-level accuracy for short social web texts in English except for political texts. SentiStrength reports two sentiment strengths (− 1 (not negative) to − 5 (extremely negative): 1 (not positive) to 5 (extremely positive)). It can also report binary (positive/negative), trinary (positive/negative/neutral), and single scale (−4 to +4) results. Moreover, SentiStrength can be configured for other languages and contexts by changing its input files. *WordNet-Affect* expands key words based on WordNet, achieving sentiment lexicon of 4787 words, in which there are 4 types of basic emotion with happy, sad, anger, and fear. *Affective Norms for English Words (ANEW)* grades emotional words with nine levels based on PAD model, rather than a simple two-level partition with positive or negative polarity. *Google-Profile of Mood States (GPOMS)* can generate public mood with six dimensions (calm, alert, sure, vital, kind, and happy). GPOMS's mood dimensions and lexicon are derived from the Profile of Mood States (POMS-bi). Profile of

Mood States (POMS) identifies and assesses transient, fluctuating affective mood states. The POMS measures six identifiable mood or affective states, namely tension-anxiety, vigor-activity, depression-dejection, fatigue-inertia, anger-hostility, and confusion-bewilderment. *Linguistic Inquiry and Word Count (LIWC)* is a text analysis software program. LIWC calculates the degree to which people use different categories of words across a wide array of texts, including emails, speeches, poems, or transcribed daily speech. LIWC2007 were developed to assess emotional, cognitive, and structural components of text samples using a psychometrically validated internal dictionary. Two versions can be found online.

The sentiment lexicons mentioned above are domain-general, which is insensitive to some special domain issues, e.g., positive words in one area may be negative in others. Therefore, there should be emotional lexicon which is sensitive in some special fields. For example, *Financial Sentiment Dictionary (FSD)* deriving from Harvard IV-4 Dictionary is used in the accounting or finance fields. *Lexicoder Sentiment Dictionary (LSD)* deriving from Regressive Imagery Dictionary (RID) is used in the political field. *DICTION* is a computer program used for analyzing text on 5 aspects of the feature such as vigor, optimism, certainty, reality, citizenship. *TAS/C* is a computer program used for psychological treatment, including more than 2000 emotional words divided into three dimensions of pleasure, approval, and attachment.

Besides, *LingPipe* is a tool kit for processing text using computational linguistics, which can be used to find the names of people, organizations, or locations in the news, to classify Twitter search results into categories, or to suggest correct spellings of queries. *Mallet and Apache OpenNLP* are another two fledged toolkits for processing text using natural language processing and machine learning techniques. They both provide a wealth of tools for part-of-speech tagging, sentence segmentation, etc. Although the method of analyzing lexical polarity based on sentiment lexicon is simple and easy to implement without training data, there are still some limitations about this type of approach (see Table 6).

## 5 Perspectives about the past and future

Research of sentiment analysis has studied almost all main aspects of the problems. The most well-studied subproblem is opinion classification on different granularity. But in other ways, existing solutions are still far from being perfect. Not much work has been done to address the problem on finer details. The area of sentiment analysis gives us many clues of the main tasks and challenges. Based on current developments, it is believed that it needs to conduct more in-depth and refined investigations aiming at multimodal sentiment analysis (MSA).

Multimodal content, as the medium of user expressions on the web, has evolved from text content to multimedia data such as videos from YouTube, Vimeo, VideoLectures, images from Flickr, Picasa, Facebook, and audios from podcasts. Multimodal form of expression has become the mainstream information resource for different institutions in strategy-setting and decision-making. But the scatter of multimodal sentiment expressions requires more analysis to derive useful data. We present an overview on recent researches of different modes individually and jointly to find the gaps in terms of approaches, theories, tasks, and applications.

So far, most of the sentiment analysis researches are based on natural language processing and computational linguistics. These traditional works focus on textual content, while people increasingly take advantage of videos, images, and audios to air their opinions on social media platforms. Thus, it is highly crucial to mine opinions and identify sentiments from the

**Table 6** Summary of tools and dictionaries for opinion mining

| Tools and Lexicons | Brief Descriptions | Limitations | Availability |
|---|---|---|---|
| General Inquirer (GI) | Take a pre-set list of positive and negative words and analyze the polarity of documents based on the prevalence of words from each category | A + B+C | http://www.wjh.harvard.edu/~inquirer/ |
| SentiWordNet | SentiWordNet assigns positive or negative numerical sentiment values to WordNet synsets | A + B+C | http://sentiwordnet.isti.cnr.it/ http://sentiwordnet.isti.cnr.it/download.php |
| OpinionFinder | OpinionFinder automatically identifies subjective sentences and marks various aspects of the subjectivity in these sentences, including the source (holder) of the subjectivity and words that are included in phrases expressing positive or negative sentiments. | A + B | http://mpqa.cs.pitt.edu/opinionfinder/ |
| National Taiwan University Sentiment Dictionary (NTUSD) | Include 2812 positive words and 8276 negative words. | A + B+C | http://academiasinicanlplab.github.io/ |
| Bing Liu's Opinion Lexicon | A lexicon includes positive words and negative 4783 words with useful properties. | A + B+C | https://www.cs.uic.edu/~liub/FBS/sentiment-analysis.html |
| SentiStrength | Texts are classified with the largest positive or negative scores of any constituent word unless these are modified by any of the additional classification rules. | A + B+C | http://sentistrength.wlv.ac.uk/ |
| WordNet-Affect | WordNet-Affect is an extension of WordNet Domains, including a subset of synsets suitable to represent affective concepts correlated with affective words. | A + B+C | http://wndomains.fbk.eu/wnaffect.html |

**Table 6** continued

| Tools and Lexicons | Brief Descriptions | Limitations | Availability |
|---|---|---|---|
| Affective Norms for English Words (ANEW) | The ANEW provides a set of normative emotional ratings for a large number of words in the English language. This set of verbal materials have been rated in terms of pleasure, arousal, and dominance in order to create a standard for use in studies of emotion and attention. | A + B+C | http://csea.phhp.ufl.edu/media/anewmessage.html |
| GPOMS | GPOMS can measure human mood states in terms of 6 different mood dimensions, namely calm, alert, sure, vital, kind and happy. | A + B+C | / |
| LIWC | A proprietary database of classified common terms. | A + B+C | http://www.liwc.net/ |
| Financial Sentiment Dictionary (FSD) | A lexicon used in the accounting or finance fields. | A + B+C | http://www.tandfonline.com/doi/full/10.1080/10584609.2012.671234 |
| Lexicoder Sentiment Dictionary (LSD) | A lexicon used in the political field. | A + B+C | http://www.lexicoder.com/ |
| DICTION | A computer programs used for analyzing text on 5 aspects of feature. | A + B+C | http://www.dictionsoftware.com/ |
| TAS/C | A computer programs used for psychological treatment. | A + B+C | / |

**Table 6** continued

| Tools and Lexicons | Brief Descriptions | Limitations | Availability |
|---|---|---|---|
| LingPipe | LingPipe can perform topic classification, named entity recognition, part-of speech tagging, sentence detection, query spell checking, interseting phrase detection, clustering, character language modeling, MEDLINE downloading, parsing and indexing, database text mining, Chinese word segmentation, sentiment analysis, language identification, etc. | Fail to get better result when training corpus and test corpus are from different data resource; need labeled entities and type information in training corpus, when performing named entity recognition. | http://alias-i.com/lingpipe/ |
| Apache OpenNLP | OpenNLP supports the most common NLP tasks, such as tokenization, sentence segmentation, part-of-speech tagging, named entity extraction, chunking, parsing, language detection and coreference resolution. | Sentence detector cannot detect sentence boundary based on content. | http://incubator.apache.org/opennlp/ |

*A* cannot detect linguistic nuances such as irony or sarcasm, *B* fail to capture the context of language, *C* fail to understand a negative polarity to the phrase like “not good” diverse modalities. However, the field of multimodal sentiment analysis has not received much attention, there are only two well-known state-of-the-art methods in multimodal sentiment analysis [21, 75], and there seems to be no single approach, theory, and tool for MSA task. Actually, the majority of such state-of-the-art frameworks rely on processing single modality. These studies showed that each mode presents different problematic issues that have not been fully solved yet. The ensemble application of feature extraction from different types of data and modalities enhances the performance of multimodal sentiment system [99]. Besides, feature fusion is important for the development of a multimodal sentiment analysis system. In general, existing research on MSA can be categorized into two broad categories, namely feature extraction from each individual modality and feature fusion coming from different modalities.

### 5.1 Video modality: emotion analysis based on facial expressions

The earliest work used anger, sadness, surprise, fear, disgust, and joy as six basic emotion categories to describe most emotions from facial expressions. However, later on, researchers added other emotions into this list such as pride, excitement, embarrassment, contempt, shame. Related systems were developed such as facial action coding system (FACS) based on the reconstruction of facial expressions in terms of action units (AU) and emotional facial action coding system (EFACS). The active appearance model (AAM) [30] and optical flow, probabilistic principal component analysis (PCA), independent component analysis (ICA), Gabor wavelet are common methods using FACS to understand expressed facial expressions. A comparison of these techniques found that Gabor wavelet and ICA performed better on most datasets. Moreover, Bayesian networks, hidden Markov models (HMM), nearest neighbor, support vector machine (SVM), AdaBoost classifiers, and artificial neural networks (ANN) have helped many researchers to infer emotions from facial expressions. It is worth mentioning that 3D morphable models (3DMM), muscle-based models, 3D wireframe models [27], elastic net model, geometry-based shape models [130], 3D constrained local model (CLM-Z) [10], generalized adaptive view-based appearance model (GAVAM) [76] are important facial expression recognition techniques.

With the advent of deep learning, we can extract features without prior intervention now. Soleymani et al. [114] presented an approach on detecting the emotions of video viewers' emotions from electroencephalogram (EEG) signals and facial expressions, in which long short-term memory recurrent neural networks (LSTM-RNN) and continuous conditional random fields (CCRF) were utilized on detecting emotions automatically and continuously. On this basis, it will be beneficial to record multimodal data with the sole aim of understanding the correlation and causation relationships between facial expressions and EEG signals. Research on facial expression analysis has been developed from acted facial expressions to spontaneous facial expressions. Huang et al. [48] proposed an approach for multimodal video-induced emotion recognition based on facial expression and electroencephalogram (EEG) technologies. In this paper, spontaneous facial expression is utilized as an external channel. A new feature, formed by percentage of nine facial expressions, is proposed for analyzing the valence and arousal classes. Furthermore, EEG is used as internal channel supplementing facial expressions for more reliable emotion recognition. Finally, these two channels are fused on feature level and decision-level for multimodal emotion recognition. Valstar et al. [127] pointed out that it is hard to tell how existing expression recognition approaches perform under conditions, where faces appear in a wide range of poses (or camera views). This is also another challenge on automatic recognition of facial expressions, which dedicates to FACS action unit detection and intensity estimation on the highly challenging data.

### 5.2 Audio modality: emotion recognition based on speech

Basic researches on emotion recognition from speech aimed at identifying several acoustic features such as frequency, intensity of utterance, bandwidth. Further studies are based on types of features including formants, Mel frequency cepstral coefficients (MFCC), pause, teager energy operated based features, log frequency power coefficients (LFPC), linear prediction cepstral coefficients (LPCC), etc. Existing methods can be categorized into two categories: speaker-dependent approach and speaker-independent approach. In general, the speaker-dependent approach gave much better results than the speaker-independent approach as shown in [77]; higher accuracy was achieved by using the Gaussian mixture model (GMM) as a classifier and Mel frequency cepstral coefficients (MFCC) as speech features. However, the speaker-dependent approach is not feasible in many applications. For the speaker-independent application, the best result achieved so far is only 81% [8], in which features are selected by the sequential floating forward selection algorithm (SFFS).

Recently, Deep learning approaches are gaining increasing attention in audio research. Dahake et al. [29] extracted the features with hybrid of pitch, formants, zero crossing, MFCC and its statistical parameters. The pitch detection is done by cepstral algorithm after comparing it with autocorrelation and AMDF. The training and testing part of the SVM classifier is compared with different kernel function like linear, polynomial, quadratic, and RBF. Trigeorgis et al. [123] proposed a strategy for "context-aware" emotional relevant feature extraction, which combines convolutional neural networks (CNNs) with LSTM networks. In this novel work, the best representation of the speech signal directly from the raw time representation is automatically learned. Usually, deep neural networks trained on simple acoustic features achieve good performance on this type of task, while continuous dimensional emotion recognition from audio has not been fully taken into account. It is treated as a sequential regression problem, in which the goal is to maximize the correlation between sequences of regression outputs and continuous-valued emotion contours, while minimizing the average deviation [135]. Han et al. [44] proposed a prediction-based learning framework for a continuous prediction task, where the main goal is to utmostly exploit the individual advantage of different regression models cooperatively. Researchers concatenate the two models in a tandem structure by different ways, forming a united cascaded framework. The outputs predicted by the former model are combined together with the original features as the input of the latter model for final predictions.

### 5.3 Text modality: emotion recognition based on textual data

This type of emotion recognition is a rapidly developing area of nature language processing and computational linguistics, which is emphatically introduced in former chapters and sections. Basically, researches of sentiment analysis revolve mainly around this type of emotion recognition, and will no longer be covered again here.

**Table 7** Summary of multimodal sentiment analysis (A = Audio; V = Video; T = Text)

| Datasets | References | Modality | Features | Sentiments | Availability |
|---|---|---|---|---|---|
| HUMAINE | [33] | A + V | Emotion words, authenticity, core affect dimensions, context labels | NA | http://emotion-research.net/download/pilot-db |
| Belfast database | [32] | A + V | Wide range of emotions | NA | http://www.belfast-naturalistic-db.sspnet.eu |
| SEMAINE | [72] | A + V | Angry, happy, fear, disgust, disgust, sadness, contempt, and amusement | NA | http://semaine-db.eu |
| IEMOCAP | [20] | A + V | Happiness, anger, sadness, frustration, and neutral state | NA | http://sail.usc.edu/iemocap |
| eNTERFACE | [71] | A + V | Happiness, anger, sadness, surprise, disgust, and fear | NA | http://enterface.net |
| ICT-MMMO | [143] | A + T+V | 1000 linguistic +1941 acoustic +20 visual | Strongly negative, weakly negative, neutral, Weakly positive, strongly positive | By sending mail to Giota Stratou (stratou@ict.usc.edu) |
| MOUD | [92] | A + T+V | 28 acoustic +40 visual | Positive, negative, neutral | http://web.eecs.umich.edu/mihalcea/downloads.html |
| YouTube dataset | [75] | A + T+V | 1000 linguistic +1941 acoustic +20 visual | Polarized words, smile, look away, Pauses and Pitch | By sending mail to Giota Stratou (stratou@ict.usc.edu) |

### 5.4 Multimodal fusion

The ability to perform multimodal fusion is the most critical step for emotion recognition. To date, there are two main levels of fusion studied: feature-level fusion or called early fusion [100, 110, 134] and decision-level fusion or called late fusion [3, 40, 148]. Feature-level fusion fuses features from various modalities before any operations are performed; decision-level fusion examines and classifies the features of each modality independently, then fuses them as a decision vector to achieve the final decision.

Besides, there are novel fusion types such as hybrid multimodal fusion combining both of feature-level fusion and decision-level fusion methods [97, 99], model-level fusion using the correlation between data under different modalities [9, 142], rule-based fusion [28], classification-based fusion using a range of classification algorithms to classify the multi-modal information [1, 78], estimation-based fusion including Kalman filter, extended Kalman filter and particle filter [79, 102]. Mainstream multimodal emotion recognition datasets and related works are summarized in Table 7.

## 6 Conclusion

Sentiment analysis became a very popular research domain and a lot of excellent researches have been accomplished toward to this area. In this survey, a series of the state-of-the-art literatures have been reviewed. In particular, this survey categorized and classified sentiment analysis researches from multiple perspectives, i.e., task-oriented, granularity-oriented, and methodology-oriented. In addition, we explored different types of data and tools that can be used in sentiment analysis research and suggested their strength and limitations. Finally, we emphasized the prospects for future development, suggestions for possible extensions, and specially presented an overview of multimodal sentiment analysis (MSA). However, since MSA methods are, in general, not being used widely in sentiment analysis and related NLP research area, there are significant and timely opportunities for future research in the multi-disciplinary field of multimodal fusion.

This survey established a common terminology across various researches, enabling people from different background knowledge to easily understand, and laid a foundation for advanced research in sentiment analysis. Current studies are intended to pave the way for further researches and development activities by identifying weaknesses and deriving guidelines toward a holistic approach.

**Acknowledgments** This research has been supported by Australian Research Council Discovery Project (Grant NO. DP160104075), the Fundamental Research Funds for the Central Universities (Grant NO. 2412017QD028), China Postdoctoral Science Foundation (Grant No. 2017M621192), the Scientific and Technological Development Program of Jilin Province (Grant No. 20180520022JH). Besides, Dr. Lin YUE has been awarded a scholarship under the State Scholarship Fund to finish this research at the University of Queensland; this work also has been awarded by China Scholarship Council (CSC). We feel grateful to Prof. Xiaofang Zhou at the University of Queensland and Prof. Ivor Tsang at University of Technology Sydney, who once offered us valuable suggestions during the study period. Our sincere thanks are also given to the anonymous reviewers, from whose comments we have benefited greatly.

## References

1. Adams W, Iyengar G, Lin CY, Naphade MR, Neti C, Nock HJ, Smith JR (2003) Semantic indexing of multimedia content using visual, audio, and text cues. EURASIP J Adv Signal Process 2:987184

2. Akhtar MS, Gupta D, Ekbal A, Bhattacharyya P (2017) Feature selection and ensemble construction: a two-step method for aspect based sentiment analysis. Knowl Based Syst 125:116–135
3. Alam F, Riccardi G (2014) Predicting personality traits using multimodal information. In: Proceedings of the 2014 ACM multi media on workshop on computational personality recognition, 2014. ACM, pp 15–18
4. Almatrafi O, Parack S, Chavan B (2015) Application of location-based sentiment analysis using Twitter for identifying trends towards Indian general elections 2014. In: Proceedings of the 9th international conference on ubiquitous information management and communication, 2015. ACM, p 41
5. Appel O, Chiclana F, Carter J, Fujita H (2016) A hybrid approach to the sentiment analysis problem at the sentence level. Knowl Based Syst 108:110–124
6. Araque O, Corcuera-Platas I, Sánchez-Rada JF, Iglesias CA (2017) Enhancing deep learning sentiment analysis with ensemble techniques in social applications. Expert Syst Appl 77:236–246
7. Asghar MZ, Khan A, Ahmad S, Qasim M, Khan IA (2017) Lexicon-enhanced sentiment analysis framework using rule-based classification scheme. PLoS ONE 12(2):e0171649
8. Atassi H, Esposito A (2008) A speaker independent approach to the classification of emotional vocal expressions. In: 20th IEEE international conference on tools with artificial intelligence, 2008. ICTAI'08, 2008. IEEE, pp 147–152
9. Baltrušaitis T, Banda N, Robinson P (2013) Dimensional affect recognition using continuous conditional random fields. In: 10th IEEE international conference and workshops on automatic face and gesture recognition (FG), 2013. IEEE, pp 1–8
10. Baltrušaitis T, Robinson P, Morency LP (2012) 3D constrained local model for rigid and non-rigid facial tracking. In: 2012 IEEE conference on computer vision and pattern recognition (CVPR), 2012. IEEE, pp 2610–2617
11. Banea C, Mihalcea R, Wiebe J (2014) Sense-level subjectivity in a multilingual setting. Comput Speech Lang 28(1):7–19
12. Barbosa L, Feng J (2010) Robust sentiment detection on twitter from biased and noisy data. In: Proceedings of the 23rd international conference on computational linguistics: posters, 2010. Association for Computational Linguistics, pp 36–44
13. Benamara F, Chardon B, Mathieu YY, Popescu V (2011) Towards context-based subjectivity analysis. In: IJCNLP, 2011. pp 1180–1188
14. Blitzer J, Dredze M, Pereira F (2007) Biographies, bollywood, boom-boxes and blenders: domain adaptation for sentiment classification. ACL 2007:440–447
15. Bollegala D, Weir D, Carroll J (2013) Cross-domain sentiment classification using a sentiment sensitive thesaurus. IEEE Trans Knowl Data Eng 25(8):1719–1731
16. Bollen J, Mao H, Zeng X (2011) Twitter mood predicts the stock market. J Comput Sci 2(1):1–8
17. Bollen J, Pepe A, Mao H (2009) Modeling public mood and emotion: Twitter sentiment and socio-economic phenomena. arXiv preprint arXiv:09111583
18. Burton K, Java A, Soboroff I (2009) The ICWSM 2009 spinn3r dataset. In: Proceedings of the 3rd annual conference on weblogs and social media (ICWSM 2009), San Jose, CA, 2009
19. Burton K, Kasch N, Soboroff I (2011) The ICWSM 2011 spinn3r dataset. In: Proceedings of the 5th international conference on weblogs and social media (ICWSM 2011), Barcelona, Spain, 2011
20. Busso C, Bulut M, Lee CC, Kazemzadeh A, Mower E, Kim S, Chang JN, Lee S, Narayanan SS (2008) IEMOCAP: interactive emotional dyadic motion capture database. Lang Resour Eval 42(4):335
21. Cambria E, Howard N, Hsu J, Hussain A (2013) Sentic blending: Scalable multimodal fusion for the continuous interpretation of semantics and sentics. In: 2013 IEEE symposium on computational intelligence for human-like intelligence (CIHLI), 2013. IEEE, pp 108–117
22. Cha M, Pérez J, Haddadi H (2009) Flash floods and ripples: The spread of media content through the blogosphere. In: ICWSM 2009: proceedings of the 3rd AAAI international conference on weblogs and social media, 2009
23. Chen YC, Cheng JY, Hsu HH (2016) A cluster-based opinion leader discovery in social network. In: 2016 conference on technologies and applications of artificial intelligence (TAAI), 2016. IEEE, pp 78–83
24. Chmiel A, Sienkiewicz J, Thelwall M, Paltoglou G, Buckley K, Kappas A, Hołyst JA (2011) Collective emotions online and their influence on community life. PLoS ONE 6(7):e22207
25. Chmiel A, Sobkowicz P, Sienkiewicz J, Paltoglou G, Buckley K, Thelwall M, Hołyst JA (2011) Negative emotions boost user activity at BBC forum. Physica A 390(16):2936–2944
26. Choi Y, Breck E, Cardie C (2006) Joint extraction of entities and relations for opinion recognition. In: Proceedings of the 2006 conference on empirical methods in natural language processing, 2006. Association for Computational Linguistics, pp 431–439

27. Cohen I, Sebe N, Gozman F, Cirelo MC, Huang TS (2003) Learning Bayesian network classifiers for facial expression recognition both labeled and unlabeled data. In: 2003. Proceedings. 2003 IEEE computer society conference on computer vision and pattern recognition, 2003. IEEE, pp I–I
28. Corradini A, Mehta M, Bernsen NO, Martin J, Abrilian S (2005) Multimodal input fusion in human-computer interaction. NATO Sci Ser Sub Ser III Comput Syst Sci 198:223
29. Dahake PP, Shaw K, Malathi P (2016) Speaker dependent speech emotion recognition using MFCC and support vector machine. In: International conference on automatic control and dynamic optimization techniques (ICACDOT), 2016. IEEE, pp 1080–1084
30. Datcu D, Rothkrantz LJ (2014) Semantic audio-visual data fusion for automatic emotion recognition. Emotion Recogn Pattern Anal Approach 411–435
31. Deng L, Wiebe J (2016) Recognizing opinion sources based on a new categorization of opinion types. IJCAI 2016:2775–2781
32. Douglas-Cowie E, Cowie R, Schröder M (2000) A new emotion database: considerations, sources and scope. In: ISCA tutorial and research workshop (ITRW) on speech and emotion, 2000
33. Douglas-Cowie E, Cowie R, Sneddon I, Cox C, Lowry O, Mcrorie M, Martin JC, Devillers L, Abrilian S, Batliner A (2007) The HUMAINE database: addressing the collection and annotation of naturalistic and induced emotional data. Affect Comput Intell Interact 488–500
34. Ertugrul AM, Onal I, Acarturk C (2017) Does the strength of sentiment matter? A regression based approach on turkish social media. In: International conference on applications of natural language to information systems, 2017. Springer, pp 149–155
35. Fattah MA (2015) New term weighting schemes with combination of multiple classifiers for sentiment analysis. Neurocomputing 167:434–442
36. Fu H, Niu Z, Zhang C, Yu H, Ma J, Chen J, Chen Y, Liu J (2016) ASELM: adaptive semi-supervised ELM with application in question subjectivity identification. Neurocomputing 207:599–609
37. Fu X, Liu W, Xu Y, Cui L (2017) Combine HowNet lexicon to train phrase recursive autoencoder for sentence-level sentiment analysis. Neurocomputing 241:18–27
38. Fu X, Liu W, Xu Y, Yu C, Wang T (2016) Long short-term memory network over rhetorical structure theory for sentence-level sentiment analysis. Asian Conf Mach Learn 2016:17–32
39. Giachanou A, Crestani F (2016) Like it or not: a survey of twitter sentiment analysis methods. ACM Comput Surv (CSUR) 49(2):28
40. Glodek M, Reuter S, Schels M, Dietmayer K, Schwenker F (2013) Kalman filter based classifier fusion for affective state recognition. In: International workshop on multiple classifier systems, 2013. Springer, pp 85–94
41. González-Ibánez R, Muresan S, Wacholder N (2011) Identifying sarcasm in Twitter: a closer look. In: Proceedings of the 49th annual meeting of the Association for Computational Linguistics: human language technologies: short papers, vol 2, 2011. Association for Computational Linguistics, pp 581–586
42. Guo H, Zhu H, Guo Z, Zhang X, Su Z (2010) OpinionIt: a text mining system for cross-lingual opinion analysis. In: Proceedings of the 19th ACM international conference on Information and knowledge management, 2010. ACM, pp 1199–1208
43. Hai Z, Chang K, Kim JJ (2011) Implicit feature identification via co-occurrence association rule mining. In: Computational linguistics and intelligent text processing. Springer, pp 393–404
44. Han J, Zhang Z, Ringeval F, Schuller B (2017) Prediction-based learning for continuous emotion recognition in speech. In: 42nd IEEE international conference on acoustics, speech, and signal processing, ICASSP 2017, 2017
45. Hassan T, Bajwa IS, Hassan S (2016) Prediction of terrorist activities by using unsupervised learning techniques. J Appl Emerg Sci 6(2):56–60
46. Hu M, Liu B (2004) Mining and summarizing customer reviews. In: Proceedings of the tenth ACM SIGKDD international conference on knowledge discovery and data mining, 2004. ACM, pp 168–177
47. Hu M, Liu B (2004) Mining opinion features in customer reviews. In: AAAI, 2004, vol 4. pp 755–760
48. Huang X, Kortelainen J, Zhao G, Li X, Moilanen A, Seppänen T, Pietikäinen M (2016) Multi-modal emotion analysis from facial expressions and electroencephalogram. Comput Vis Image Underst 147:114–124
49. Huynh T, He Y, Rüger S (2015) Learning higher-level features with convolutional restricted Boltzmann machines for sentiment analysis. In: Advances in information retrieval. Springer, pp 447–452
50. Iosifidis V, Ntoutsi E (2017) Large scale sentiment learning with limited labels. In: Proceedings of the 23rd ACM SIGKDD international conference on knowledge discovery and data mining, 2017. ACM, pp 1823–1832
51. Jo Y, Oh AH (2011) Aspect and sentiment unification model for online review analysis. In: Proceedings of the 4th ACM international conference on web search and data mining, 2011. ACM, pp 815–824

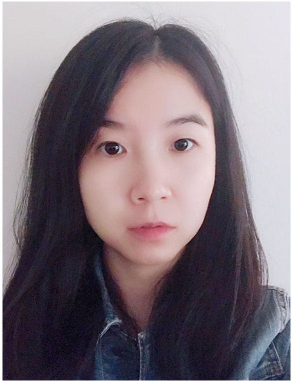
52. Kang M, Ahn J, Lee K (2018) Opinion mining using ensemble text hidden Markov models for text classification. Expert Syst Appl 94:218–227
53. Karimi S, Shakery A (2017) A language-model-based approach for subjectivity detection. J Inf Sci 43(3):356–377
54. Karyotis C, Doctor F, Iqbal R, James A, Chang V (2017) A fuzzy computational model of emotion for cloud based sentiment analysis. Inf Sci 433–434:448–463
55. Katiyar A, Cardie C (2016) Investigating LSTMs for joint extraction of opinion entities and relations. In: ACL (1), 2016
56. Khan FH, Qamar U, Bashir S (2016) eSAP: a decision support framework for enhanced sentiment analysis and polarity classification. Inf Sci 367:862–873
57. Kim SM, Hovy E (2006) Extracting opinions, opinion holders, and topics expressed in online news media text. In: Proceedings of the workshop on sentiment and subjectivity in text, 2006. Association for Computational Linguistics, pp 1–8
58. Kramer AD (2010) An unobtrusive behavioral model of gross national happiness. In: Proceedings of the SIGCHI conference on human factors in computing systems, 2010. ACM, pp 287–290
59. Lei X, Qian X, Zhao G (2016) Rating prediction based on social sentiment from textual reviews. IEEE Trans Multimed 18(9):1910–1921
60. Li N, Zhai S, Zhang Z, Liu B (2017) Structural correspondence learning for cross-lingual sentiment classification with one-to-many mappings. AAAI 2017:3490–3496
61. Liao J, Wang S, Li D, Li X (2017) FREERL: fusion relation embedded representation learning framework for aspect extraction. Knowl Based Syst 135:9–17
62. Liu B (2010) Sentiment analysis and subjectivity. Handb Nat Lang Process 2:627–666
63. Liu B (2012) Sentiment analysis and opinion mining. Synth Lect Hum Lang Technol 5(1):1–167
64. Liu B, Hu M, Cheng J (2005) Opinion observer: analyzing and comparing opinions on the web. In: Proceedings of the 14th international conference on World Wide Web, 2005. ACM, pp 342–351
65. Liu S, Zhu W, Xu N, Li F, Cheng XQ, Liu Y, Wang Y (2013) Co-training and visualizing sentiment evolvement for tweet events. In: Proceedings of the 22nd international conference on World Wide Web, 2013. ACM, pp 105–106
66. Liu Y, Bi JW, Fan ZP (2017) A method for multi-class sentiment classification based on an improved one-vs-one (OVO) strategy and the support vector machine (SVM) algorithm. Inf Sci 394:38–52
67. Liu Z, Dong X, Guan Y, Yang J (2013) Reserved self-training: a semi-supervised sentiment classification method for chinese microblogs. In: Proceedings of the 6th international joint conference on natural language processing 2013. pp 455–462
68. Lu B (2010) Identifying opinion holders and targets with dependency parser in Chinese news texts. In: Proceedings of the NAACL HLT 2010 student research workshop, 2010. Association for Computational Linguistics, pp 46–51
69. Lu Y, Tsaparas P, Ntoulas A, Polanyi L (2010) Exploiting social context for review quality prediction. In: Proceedings of the 19th international conference on World wide web, 2010. ACM, pp 691–700
70. Manek AS, Shenoy PD, Mohan MC, Venugopal K (2017) Aspect term extraction for sentiment analysis in large movie reviews using Gini Index feature selection method and SVM classifier. World wide web 20(2):135–154
71. Martin O, Kotsia I, Macq B, Pitas I (2006) The enterface'05 audio-visual emotion database. In: Proceedings. 22nd international conference on data engineering workshops, 2006, 2006. IEEE, pp 8–8
72. McKeown G, Valstar M, Cowie R, Pantic M, Schroder M (2012) The semaine database: annotated multimodal records of emotionally colored conversations between a person and a limited agent. IEEE Trans Affect Comput 3(1):5–17
73. Medhat W, Hassan A, Korashy H (2014) Sentiment analysis algorithms and applications: a survey. Ain Shams Eng J 5(4):1093–1113
74. Mitrović M, Paltoglou G, Tadić B (2011) Quantitative analysis of bloggers' collective behavior powered by emotions. J Stat Mech Theory Exp 02:P02005
75. Morency LP, Mihalcea R, Doshi P (2011) Towards multimodal sentiment analysis: Harvesting opinions from the web. In: Proceedings of the 13th international conference on multimodal interfaces, 2011. ACM, pp 169–176
76. Morency LP, Whitehill J, Movellan J (2008) Generalized adaptive view-based appearance model: integrated framework for monocular head pose estimation. In: FG'08. 8th IEEE international conference on automatic face & gesture recognition, 2008, 2008. IEEE, pp 1–8
77. Navas E, Hernaez I, Luengo I (2006) An objective and subjective study of the role of semantics and prosodic features in building corpora for emotional TTS. IEEE Trans Audio Speech Lang Process 14(4):1117–1127

78. Nefian AV, Liang L, Pi X, Liu X, Murphy K (2002) Dynamic Bayesian networks for audio-visual speech recognition. EURASIP J Adv Signal Process 11:783042
79. Nickel K, Gehrig T, Stiefelhagen R, McDonough J (2005) A joint particle filter for audio-visual speaker tracking. In: Proceedings of the 7th international conference on multimodal interfaces, 2005. ACM, pp 61–68
80. Ortigosa A, Martín JM, Carro RM (2014) Sentiment analysis in Facebook and its application to e-learning. Comput Hum Behav 31:527–541
81. Pak A, Paroubek P (2010) Twitter as a corpus for sentiment analysis and opinion mining. LREC 2010:1320–1326
82. Paltoglou G, Thelwall M (2012) Twitter, MySpace, Digg: unsupervised sentiment analysis in social media. ACM Trans Intell Syst Technol (TIST) 3(4):66
83. Paltoglou G, Thelwall M, Buckely K (2010) Online textual communications annotated with grades of emotion strength. In: Proceedings of the 3rd international workshop of emotion: corpora for research on emotion and affect, 2010. pp 25–31
84. Pan SJ, Ni X, Sun JT, Yang Q, Chen Z (2010) Cross-domain sentiment classification via spectral feature alignment. In: Proceedings of the 19th international conference on World wide web, 2010. ACM, pp 751–760
85. Pang B, Lee L (2004) A sentimental education: Sentiment analysis using subjectivity summarization based on minimum cuts. In: Proceedings of the 42nd annual meeting on Association for Computational Linguistics, 2004. Association for Computational Linguistics, p 271
86. Pang B, Lee L (2005) Seeing stars: Exploiting class relationships for sentiment categorization with respect to rating scales. In: Proceedings of the 43rd annual meeting on Association for Computational Linguistics, 2005. Association for Computational Linguistics, pp 115–124
87. Pang B, Lee L (2008) Opinion mining and sentiment analysis. Found Trends Inf Retr 2(1–2):1–135
88. Pang B, Lee L (2008) Opinion mining and sentiment analysis. Found Trends Inf Retr 2(1–2):1–135
89. Pang B, Lee L, Vaithyanathan S (2002) Thumbs up?: sentiment classification using machine learning techniques. In: Proceedings of the ACL-02 conference on empirical methods in natural language processing, vol 10, 2002. Association for Computational Linguistics, pp 79-86
90. Paul D, Li F, Teja MK, Yu X, Frost R (2017) Compass: Spatio Temporal Sentiment Analysis of US Election What Twitter Says! In: Proceedings of the 23rd ACM SIGKDD international conference on knowledge discovery and data mining, 2017. ACM, pp 1585–1594
91. Penalver-Martinez I, Garcia-Sanchez F, Valencia-Garcia R, Rodriguez-Garcia MA, Moreno V, Fraga A, Sanchez-Cervantes JL (2014) Feature-based opinion mining through ontologies. Expert Syst Appl 41(13):5995–6008
92. Pérez-Rosas V, Mihalcea R, Morency LP (2013) Utterance-level multimodal sentiment analysis. ACL 1(2013):973–982
93. Phu VN, Dat ND, Tran VTN, Chau VTN, Nguyen TA (2017) Fuzzy C-means for english sentiment classification in a distributed system. Appl Intell 46(3):717–738
94. Phu VN, Tran VTN, Chau VTN, Dat ND, Duy KLD (2017) A decision tree using ID3 algorithm for English semantic analysis. Int J Speech Technol 1-21
95. Polpinij J, Ghose AK (2008) An ontology-based sentiment classification methodology for online consumer reviews. In: Proceedings of the 2008 IEEE/WIC/ACM international conference on web intelligence and intelligent agent technology, vol 01, 2008. IEEE Computer Society, pp 518–524
96. Popescu AM, Etzioni O (2007) Extracting product features and opinions from reviews. In: Natural language processing and text mining. Springer, pp 9–28
97. Poria S, Cambria E, Gelbukh A (2015) Deep convolutional neural network textual features and multiple kernel learning for utterance-level multimodal sentiment analysis. In: Proceedings of the 2015 conference on empirical methods in natural language processing, 2015. pp 2539–2544
98. Poria S, Cambria E, Gelbukh A (2016) Aspect extraction for opinion mining with a deep convolutional neural network. Knowl Based Syst 108:42–49
99. Poria S, Cambria E, Howard N, Huang GB, Hussain A (2016) Fusing audio, visual and textual clues for sentiment analysis from multimodal content. Neurocomputing 174:50–59
100. Poria S, Cambria E, Hussain A, Huang GB (2015) Towards an intelligent framework for multimodal affective data analysis. Neural Netw 63:104–116
101. Poria S, Chaturvedi I, Cambria E, Hussain A (2016) Convolutional MKL based multimodal emotion recognition and sentiment analysis. In: 2016 IEEE 16th international conference on data mining (ICDM), 2016. IEEE, pp 439–448
102. Potamitis I, Chen H, Tremoulis G (2004) Tracking of multiple moving speakers with multiple microphone arrays. IEEE Trans Speech Audio Process 12(5):520–529

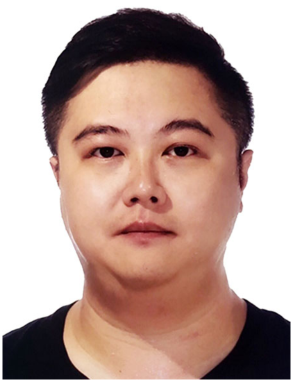
103. Pozzi FA, Maccagnola D, Fersini E, Messina E (2013) Enhance user-level sentiment analysis on microblogs with approval relations. In: Congress of the Italian Association for Artificial Intelligence, 2013. Springer, pp 133–144
104. Qiu G, Liu B, Bu J, Chen C (2011) Opinion word expansion and target extraction through double propagation. Comput Linguist 37(1):9–27
105. Qu L, Ifrim G, Weikum G (2010) The bag-of-opinions method for review rating prediction from sparse text patterns. In: Proceedings of the 23rd international conference on computational linguistics, 2010. Association for Computational Linguistics, pp 913–921
106. Ravi K, Ravi V (2015) A survey on opinion mining and sentiment analysis: tasks, approaches and applications. Knowl Based Syst 89:14–46
107. Riaz S, Fatima M, Kamran M, Nisar MW (2017) Opinion mining on large scale data using sentiment analysis and k-means clustering. Cluster Computing:1-16
108. Ruppenhofer J, Somasundaran S, Wiebe J (2008) Finding the sources and targets of subjective expressions. In: LREC, 2008
109. Saif H, Fernandez M, He Y, Alani H (2013) Evaluation datasets for Twitter sentiment analysis: a survey and a new dataset, the STS-Gold
110. Sarkar C, Bhatia S, Agarwal A, Li J (2014) Feature analysis for computational personality recognition using youtube personality data set. In: Proceedings of the 2014 ACM multi media on workshop on computational personality recognition, 2014. ACM, pp 11–14
111. Schouten K, Frasincar F (2016) Survey on aspect-level sentiment analysis. IEEE Trans Knowl Data Eng 28(3):813–830
112. Si J, Mukherjee A, Liu B, Li Q, Li H, Deng X (2013) Exploiting topic based twitter sentiment for stock prediction. ACL 2(2013):24–29
113. Snyder B, Barzilay R (2007) Multiple aspect ranking using the good grief algorithm. HLT-NAACL 2007:300–307
114. Soleymani M, Asghari-Esfeden S, Fu Y, Pantic M (2016) Analysis of EEG signals and facial expressions for continuous emotion detection. IEEE Trans Affect Comput 7(1):17–28
115. Su F, Markert K (2008) From words to senses: a case study of subjectivity recognition. In: Proceedings of the 22nd international conference on computational linguistics, vol 1, 2008. Association for Computational Linguistics, pp 825–832
116. Suresh H (2016) An unsupervised fuzzy clustering method for twitter sentiment analysis. In: International conference on computation system and information technology for sustainable solutions (CSITSS), 2016. IEEE, pp 80–85
117. Tan C, Lee L, Tang J, Jiang L, Zhou M, Li P (2011) User-level sentiment analysis incorporating social networks. In: Proceedings of the 17th ACM SIGKDD international conference on knowledge discovery and data mining, 2011. ACM, pp 1397–1405
118. Tang D, Qin B, Liu T (2015) Document modeling with gated recurrent neural network for sentiment classification. EMNLP 2015:1422–1432
119. Tang D, Wei F, Qin B, Liu T, Zhou M (2014) Coooolll: a deep learning system for twitter sentiment classification. In: SemEval@ COLING, 2014. pp 208–212
120. Tang D, Wei F, Yang N, Zhou M, Liu T, Qin B (2014) Learning sentiment-specific word embedding for Twitter sentiment classification. ACL 1(2014):1555–1565
121. Tang H, Tan S, Cheng X (2009) A survey on sentiment detection of reviews. Expert Syst Appl 36(7):10760–10773
122. Tellez ES, Miranda-Jiménez S, Graff M, Moctezuma D, Suárez RR, Siordia OS (2017) A simple approach to multilingual polarity classification in twitter. Pattern Recogn Lett
123. Trigeorgis G, Ringeval F, Brueckner R, Marchi E, Nicolaou MA, Schuller B, Zafeiriou S (2016) Adieu features? End-to-end speech emotion recognition using a deep convolutional recurrent network. In: Acoustics, Speech and Signal Processing (ICASSP), 2016 IEEE International Conference on, 2016. IEEE, pp 5200-5204
124. Tsur O, Davidov D, Rappoport A (2010) ICWSM-a great catchy name: semi-supervised recognition of sarcastic sentences in online product reviews. In: ICWSM, 2010
125. Tsytsarau M, Palpanas T (2012) Survey on mining subjective data on the web. Data Min Knowl Disc 24(3):478–514
126. Turney PD (2002) Thumbs up or thumbs down?: semantic orientation applied to unsupervised classification of reviews. In: Proceedings of the 40th annual meeting on association for computational linguistics, 2002. Association for Computational Linguistics, pp 417–424
127. Valstar MF, Sánchez-Lozano E, Cohn JF, Jeni LA, Girard JM, Zhang Z, Yin L, Pantic M (2017) FERA 2017-addressing head pose in the third facial expression recognition and analysis challenge. arXiv preprint arXiv:170204174

128. Vaziripour E, Giraud-Carrier CG, Zappala D (2016) Analyzing the political sentiment of tweets in Farsi. ICWSM 2016:699–702
129. Velikovich L, Blair-Goldensohn S, Hannan K, McDonald R (2010) The viability of web-derived polarity lexicons. In: Human language technologies: the 2010 annual conference of the North American chapter of the Association for Computational Linguistics, 2010. Association for Computational Linguistics, pp 777–785
130. Verma R, Davatzikos C, Loughead J, Indersmitten T, Hu R, Kohler C, Gur RE, Gur RC (2005) Quantification of facial expressions using high-dimensional shape transformations. J Neurosci Methods 141(1):61–73
131. Vinodhini G, Chandrasekaran R (2012) Sentiment analysis and opinion mining: a survey. Int J 2(6):282–292
132. Wan X (2008) Using bilingual knowledge and ensemble techniques for unsupervised Chinese sentiment analysis. In: Proceedings of the conference on empirical methods in natural language processing, 2008. Association for Computational Linguistics, pp 553–561
133. Wan X (2009) Co-training for cross-lingual sentiment classification. In: Proceedings of the joint conference of the 47th annual meeting of the ACL and the 4th international joint conference on natural language processing of the AFNLP: vol 1–vol 1, 2009. Association for Computational Linguistics, pp 235–243
134. Wang S, Zhu Y, Wu G, Ji Q (2014) Hybrid video emotional tagging using users' EEG and video content. Multimed Tools Appl 72(2):1257–1283
135. Weninger F, Ringeval F, Marchi E, Schuller BW (2016) Discriminatively trained recurrent neural networks for continuous dimensional emotion recognition from audio. IJCAI 2016:2196–2202
136. Wiebe J, Riloff E (2005) Creating subjective and objective sentence classifiers from unannotated texts. In: Computational linguistics and intelligent text processing. Springer, pp 486–497
137. Wiebe J, Wilson T, Cardie C (2005) Annotating expressions of opinions and emotions in language. Lang Resour Eval 39(2–3):165–210
138. Wiegand M, Balahur A, Roth B, Klakow D, Montoyo A (2010)A survey on the role of negation in sentiment analysis. In: Proceedings of the workshop on negation and speculation in natural language processing, 2010. Association for Computational Linguistics, pp 60–68
139. Wiegand M, Bocionek C, Ruppenhofer J (2016) Opinion holder and target extraction on opinion compounds—a linguistic approach. HLT-NAACL 2016:800–810
140. Wiegand M, Klakow D (2010) Convolution kernels for opinion holder extraction. In: Human language technologies: the 2010 annual conference of the North American chapter of the Association for Computational Linguistics, 2010. Association for Computational Linguistics, pp 795–803
141. Wilson TA (2008) Fine-grained subjectivity and sentiment analysis: recognizing the intensity, polarity, and attitudes of private states. ProQuest
142. Wöllmer M, Kaiser M, Eyben F, Schuller B, Rigoll G (2013) LSTM-Modeling of continuous emotions in an audiovisual affect recognition framework. Image Vis Comput 31(2):153–163
143. Wöllmer M, Weninger F, Knaup T, Schuller B, Sun C, Sagae K, Morency LP (2013) Youtube movie reviews: sentiment analysis in an audio-visual context. IEEE Intell Syst 28(3):46–53
144. Wu F, Zhang J, Yuan Z, Wu S, Huang Y, Yan J (2017) Sentence-level sentiment classification with weak supervision. In: Proceedings of the 40th international ACM SIGIR conference on research and development in information retrieval, 2017. ACM, pp 973–976
145. Wu Y, Liu S, Yan K, Liu M, Wu F (2014) Opinionflow: visual analysis of opinion diffusion on social media. IEEE Trans Vis Comput Graph 20(12):1763–1772
146. Xia R, Xu F, Yu J, Qi Y, Cambria E (2016) Polarity shift detection, elimination and ensemble: a three-stage model for document-level sentiment analysis. Inf Process Manag 52(1):36–45
147. Xiang B, Zhou L (2014) Improving twitter sentiment analysis with topic-based mixture modeling and semi-supervised training. In: Proceedings of the 52nd annual meeting of the Association for Computational Linguistics, vol 2, short papers, 2014. pp 434–439
148. Yamasaki T, Fukushima Y, Furuta R, Sun L, Aizawa K, Bollegala D (2015) Prediction of user ratings of oral presentations using label relations. In: Proceedings of the 1st international workshop on affect and sentiment in multimedia, 2015. ACM, pp 33–38
149. Yang L, Liu B, Lin H, Lin Y (2016) Combining local and global information for product feature extraction in opinion documents. Inf Process Lett 116(10):623–627
150. Yu N (2014) Exploring Co-training strategies for opinion detection. J Assoc Inf Sci Technol 65(10):2098–2110
151. Zhao J, Lan M, Zhu T (2014) ECNU: expression-and message-level sentiment orientation classification in twitter using multiple effective features. In: SemEval@ COLING, 2014. pp 259–264

152. Zhao K, Cong G, Yuan Q, Zhu KQ (2015) SAR: a sentiment-aspect-region model for user preference analysis in geo-tagged reviews. In: 2015 IEEE 31st international conference on data engineering (ICDE), 2015. IEEE, pp 675–686
153. Zhu J, Wang H, Tsou BK, Zhu M (2009) Multi-aspect opinion polling from textual reviews. In: Proceedings of the 18th ACM conference on information and knowledge management, 2009. ACM, pp 1799–1802



**Lin Yue** is a Postdoctoral Research Fellow in Northeast Normal University. She received the BS and MS degrees in computer science from Northeast Normal University, and Ph.D. degree in computer science from Key Laboratory of Symbolic Computation and Knowledge Engineering of Ministry of Education, Jilin University, China. She was a joint Ph.D. student in The University of Queensland, Australia. Her current main research interests include data mining, natural language processing and machine learning.



**Weitong Chen** is a Ph.D. candidate in the School of Information Technology and Electrical Engineering, The University of Queensland (UQ) in Brisbane, Queensland, Australia. He received the BS and MS degrees from the Griffith University and The University of Queensland, Australia, in 2011 and 2013, respectively. His current main research interests include medical data analytic, deep learning, data mining, pattern recognition, and social computing.

**Xue Li** is a Professor in the School of Information Technology and Electrical Engineering, The University of Queensland (UQ) in Brisbane, Queensland, Australia. He received the MS and Ph.D. degrees from The University of Queensland and Queensland University of Technology, Australia, in 1990 and 1997, respectively. He is honored as one of the most powerful people in Australia on Big Data by the Financial Review-the Power Issue 2015. His major areas of research interests and expertise include data mining, social computing, database systems and intelligent web information systems.

**Wanli Zuo** is a Professor in Jilin University. He received the BS, MS and Ph.D. degrees in computer science from Jilin University, China, in 1982, 1984 and 2005, respectively. His current main research interests include data mining, information retrieval, natural language processing and machine learning.

**Minghao Yin** is a Professor in Northeast Normal University. He received the BS and MS degrees in computer science from Northeast Normal University, China, in 2001 and 2004, respectively, and Ph.D. degree in computer science from Jilin University, China, in 2008. His current main research interests include swarm intelligence, automated reasoning, automated planning and algorithms.