A Review of Natural Language Processing Techniques for Opinion Mining Systems

Shiliang Sun*, Chen Luo, Junyu Chen

Department of Computer Science and Technology, East China Normal University, 500 Dongchuan Road, Shanghai 200241, P. R. China

---

Abstract

As the prevalence of social media on the Internet, opinion mining has become an essential approach to analyzing so many data. Various applications appear in a wide range of industrial domains. Meanwhile, opinions have diverse expressions which bring along research challenges. Both of the practical demands and research challenges make opinion mining an active research area in recent years. In this paper, we present a review of Natural Language Processing (NLP) techniques for opinion mining. First, we introduce general NLP techniques which are required for text preprocessing. Second, we investigate the approaches of opinion mining for different levels and situations. Then we introduce comparative opinion mining and deep learning approaches for opinion mining. Opinion summarization and advanced topics are introduced later. Finally, we discuss some challenges and open problems related to opinion mining.

Key words:
Opinion mining, sentiment analysis, natural language processing, deep learning, machine learning

---

1. Introduction

With the explosive growth of user-generated texts on the Internet, extraction of useful information automatically from abundant documents receives

*Corresponding author.Tel.: +86-21-54345183; fax: +86-21-54345119.
Email address: shiliangsun@gmail.com, slsun@cs.ecnu.edu.cn (Shiliang Sun)

interests from researchers in many fields, in particular the community of Natural Language Processing (NLP). Opinion mining (also known as sentiment analysis) [1, 2] was firstly proposed in early this century and has become an active research area gradually. Moreover, various practical applications of opinion mining, such as product pricing [3], competitive intelligence [4], market prediction [5, 6], election forecasting [7, 8], nation relationship analysis [9], and risk detection in banking systems [10], draw extensive attentions from industrial communities. On the other hand, the growth of social media, electronic commerce and online review sites, such as *Twitter*, *Amazon*, and *Yelp*, provides a large amount of corpora which are crucial resources for academic research. Interests from both academia and industry promote the development of opinion mining.

We follow the definition of opinion or sentiment from [2] where it is represented as a quintuple

$$(e_i, a_{ij}, s_{ijk\ell}, h_k, t_\ell),$$

in which e_i is the ith entity, a_{ij} is the jth aspect of the ith entity, h_k is the kth opinion holder, t_ℓ is the time when the opinion is expressed, $s_{ijk\ell}$ is the opinion or sentiment towards the jth aspect of the ith entity from opinion holder h_k at time t_ℓ. For example, in the review "*The screen of this mobile phone is good!*", the first three components can be determined: *screen* is an aspect of entity *mobile phone* and a positive sentiment is expressed.

With this definition, the objective of opinion mining is defined as determining the quintuple of given texts. Corresponding sub-tasks are identifying five components, respectively. However, the whole quintuple is not always necessary for different situations. For example, the third component is enough for document level opinion mining while more components are required for the fine-grained level. Based on the quintuple or a part of it, some advanced tasks, such as summarization, can be performed to provide a quick look at some particular opinion targets.

Machine learning approaches play a significant role for opinion mining. Generally, document and sentence level opinion mining can be formulated as classification problems which determine whether a positive or negative sentiment is expressed. Classifiers are trained to determine the polarities of forthcoming texts. Naïve Bayes classifier, maximum entropy classifier and Support Vector Machine (SVM) [11] are the most commonly used models. However, the requirement of annotated corpora is not easily satisfied, especially for cross-domain and cross-lingual situations. In the cross-domain situation, since there exist differences among different domains, the classifier trained from one domain does not always achieve comparable performance in another domain. It is more severe in the cross-lingual situation than the cross-domain situation. In the cross-lingual situation, trained classifiers can not be applied to texts in another language directly because of the language differences. Semi-supervised methods, which train classifiers on both annotated and unannotated corpora in different domains and languages, are developed to deal with the lack of annotated corpora. In fine-grained level opinion mining, more efforts are required for extraction of opinion targets and their relations. Meanwhile, corpora with annotated opinion targets and sentiment polarities are difficult to obtain. Several unsupervised methods based on Latent Dirichlet Allocation (LDA) [12] have been proposed to release the dependence of annotated corpora.

Lexicon approaches determine the sentiment score of text according to sentiment lexicons in an unsupervised manner. A lexicon is a dictionary of sentiment words and phrases with their polarities and strengths. For each document or sentence, the corresponding polarity is determined by a sentiment score which is computed by the occurred words or phrases and their sentiment polarities and strengths. Comparing to machine learning approaches, lexicon approaches require less resources, which makes it suitable for the situation that no annotated corpora are available. In addition, a sentiment lexicon can be adopted in machine learning approaches to construct sentiment related features which are helpful for better performances.

Recently, there are several reviews on NLP [42], opinion mining [2] and information fusion [13]. Khaleghi et al. [13] presented a review on multi-sensor information fusion and discussed soft data fusion, which relates to natural language processing and opinion mining. They also indicated the complexity of soft data fusion. Cambria and White [42] focused on the evolution of NLP, including the historical background of NLP, syntax and semantics technologies and others. Brief contents about opinion mining were mentioned in this review. Liu [2] provided a comprehensive review of NLP and opinion mining on the definition of opinion mining, different levels of opinion mining, summarization, lexicon creation, spam detection, opinion quality measurement and so on. However, the relations between information fusion and opinion mining are not covered. Besides the above reviews on specific areas, Balazs and Velásquez [14] investigated related work on opinion mining as well as information fusion, and proposed a conceptual framework of applying information fusion for opinion mining. However, the contents on opinion mining are not comprehensive.

Different from the above reviews, we not only investigate existing NLP techniques for opinion mining and relations between information fusion and opinion mining, but also discuss open problems and challenges in opinion mining. Techniques for different levels and settings of opinion mining as well as advanced topics are introduced. Some recent work, such as deep learning for opinion mining, is provided as well.

The rest of the paper is organized as follows. We first discuss information fusion and opinion mining in Section 2. Section 3 introduces some NLP techniques for text preprocessing. Section 4 presents existing approaches of opinion mining for different levels and settings. Section 5 presents comparative opinion mining. Then we introduce deep learning related studies in Section 6. Opinion summarization is introduced in Section 7. Advanced topics including opinion spam detection and usefulness measurement are presented in Section 8. Finally, we make conclusion after discussing the challenges and open problems in opinion mining.

2. Information Fusion and Opinion Mining

Information fusion is a kind of technology that integrates information from multiple sources to solve a certain task. It is now widely used in various areas including sensor networks, video and image processing, and intelligent system design [13]. Information fusion also exists extensively in opinion mining. Below we discuss the information fusion in opinion mining from aspects of data sources, sentiment resources, and techniques [14].

The fusion of data sources in opinion mining is integration of raw data from different sources. For product reviews, user and product information are adopted to capture clues about an overall sentiment polarity [15]. In opinion spam detection and usefulness measurement, various related information about reviewers, raters, and products is taken into account together [16, 17, 18, 19, 20, 21, 22].

The fusion of sentiment resources is mainly about the fusion of corpora and lexicons. In the cross-domain situation, for example, annotated and unannotated corpora from different domains are integrated to train a model in a semi-supervised manner, which can deal with the lack of annotated corpora as well as the unrobustness of classification performance between different domains [23, 24, 25]. Similarly, in the cross-lingual situation, bilingual dictionaries, machine translation technology, and parallel corpora make it possible to exploit the abundant English resources [26, 27, 28, 29]. Lexicon resources are valuable for both machine learning and lexicon approaches. Various existing sentiment lexicons are combined to form a rich lexicon in order to obtain robust performances [30]. In the lexicon creation, some external synonym or antonym dictionaries are used, which is also a kind of fusion of resources [31, 32, 33].

The fusion of techniques is the fusion of methods. For example, a well-designed combination of different models could outperform a single model [34]. Information retrieve methods are commonly used to construct feature space for machine learning approaches, such as unigram and bigram with tf-idf weights [35, 36, 37]. The features in machine learning are usually extracted by various methods which can be regarded as fusion of methods [35, 38, 39, 40, 41].

3. NLP Techniques for Text Preprocessing

Opinion mining requires several preprocessing steps for structuring the text and extracting features, including tokenization, word segmentation, Part of Speech (POS) tagging, parsing. Now we give a brief overview of these techniques.

Tokenization is a fundamental technique for most NLP tasks. It splits a sentence or document into tokens which are words or phrases. For English, it is trivial to split words by the spaces, but some additional knowledge should be taken into consideration, such as opinion phrases, named entities. In tokenization, some stop words, such as "*the*", "*a*", will be removed as these words provide little useful information. As a fundamental technique, many tokenization tools are available, such as Stanford Tokenizer[1], OpenNLP Tokenizer[2].

For Chinese, Japanese or other languages which do not have explicit word boundary markers, tokenization is not as trivial as English and word segmentation is required. The word segmentation is a sequential labeling problem. Conditional Random Fields (CRFs) [43] have been applied to this problem and outperformed hidden Markov models and maximum-entropy Markov models [44, 45, 46]. Recently, word embedding and deep learning based approaches have been applied to Chinese word segmentation [47, 48]. Several

[1] `http://nlp.stanford.edu/software/tokenizer.shtml`

[2] `https://opennlp.apache.org/documentation/manual/opennlp.html#tools.tokenizer`

tools are available, such as ICTCLAS[3], THULAC[4], and Stanford Segmenter[5]. For more about CRF, Sutton and McCallum [49] presented a comprehensive introduction.

POS tagging and parsing are techniques that analyze the lexical and syntactic information. POS tagging is used to determine the corresponding POS tag for each word. Similar to word segmentation, it is also a sequential labeling problem. The POS tags, such as *adjective*, *noun*, are quite helpful because opinion words are usually adjectives and opinion targets (i.e., entities and aspects) are nouns or combination of nouns. While POS tagging provides lexical information, parsing obtains syntactic information. Parsing produces a tree which represents the grammatical structure of a given sentence with the corresponding relationship of different constituents. Comparing to POS tagging, parsing provides richer structure information. As the similarity and relevance among word segmentation, POS tagging, and parsing, some approaches are proposed to deal with these tasks simultaneously [50, 51, 52].

3.1. Available Toolkits for NLP

We have mentioned several tools that deal with tokenization and Chinese word segmentation, some of which are integrated in powerful toolkits. Now we present an overview of commonly used toolkits in Table 1.

Table 1: Available Toolkits for NLP

Toolkit	Language	Description
NLTK [53]	Python	Natural Language Toolkit (NLTK) is an open source platform for performing NLP tasks including tokenization, stemming, POS tagging, parsing, and semantic reasoning. It provides interfaces for many corpora and lexicons which are useful for opinion mining and sentiment analysis. `http://www.nltk.org/`

[3]`http://ictclas.nlpir.org`

[4]`http://thulac.thunlp.org`

[5]`http://nlp.stanford.edu/software/segmenter.shtml`

OpenNLP	JAVA	The Apache OpenNLP is a JAVA library for the processing of natural language texts, which supports common tasks including tokenization, sentence segmentation, POS tagging, named entity recognition, parsing, and coreference resolution. `https://opennlp.apache.org`
CoreNLP [54]	JAVA	Stanford CoreNLP is a framework which supports not only basic NLP task, such as POS tagging, named entity recognization, parsing, coreference resolution, but also advanced sentiment analysis [55]. `http://stanfordnlp.github.io/CoreNLP/`
Gensim [56]	Python	Gensim is an open source library for topic modeling which includes online Latent Semantic Analysis (LSA), Latent Dirichlet Allocation (LDA), Random Projection, Hierarchical Dirichlet Process and word2vec. All implemented algorithms support large scale corpora. LSA and LDA have distributed parallel versions. `http://radimrehurek.com/gensim/`
FudanNLP [57]	JAVA	FudanNLP is an open source toolkit for Chinese NLP, which supports word segmentation, POS tagging, named entity recognition, dependency parsing, coreference resolution and so on. `https://code.google.com/archive/p/fudannlp/`
LTP [58]	C++ /Python	The Language Technology Platform (LTP) is an open source NLP system for Chinese, including lexical analysis (word segmentation, POS tagging, named entity recognition), syntactic parsing and semantic parsing (word sense disambiguation, semantic role labeling) modules. `http://www.ltp-cloud.com/intro/en/`
NiuParser [59]	C++	NiuParser is a Chinese Syntactic and Semantic Analysis Toolkit, which supports word segmentation, POS tagging, named entity recognition, constituent parsing, dependency parsing and semantic role labeling. `http://www.niuparser.com/index.en.html`

Besides the above brief information, we also investigate the implementation details of the toolkits.

NLTK, OpenNLP and Stanford CoreNLP are widely used general NLP toolkits which support most of basic NLP tasks, such as POS tagging, named entity recognition, parsing.

NLTK provides industrial-strength NLP libraries. For each basic NLP task, it provides implementations of various techniques. In POS tagging module, regular expression-based tagger, language model-based tagger, HMM tagger, CRF tagger as well as rule-based tagger are provided. It also provides APIs for Stanford POS tagger and Senna tagger. In named entity recognition module, maximum entropy classifier is adopted. The pre-trained model contains multiple labels, including PERSON, ORGANIZATION, and GPE (e.g., geo-political entities such as city, state/province, and country). In the parsing module, various parsers, including chart parser, shift-reduce parser and transition parser, are provided. NLTK also supports semantic reasoning which is mainly about natural language understanding and provides tools for analyzing expressions of the first order logic and evaluating such expressions.

OpenNLP is an open source framework which supports basic NLP tasks. Maximum entropy classifiers are adopted for named entity recognition, POS tagging, chunking and coreference resolution. Chunking parser is implemented for the parsing module. The most of pre-trained models in NLTK and openNLP are limited on English. However, training APIs are provided which makes it capable for training new models on additional corpora in English and other languages.

Stanford CoreNLP is a collection of NLP tools based on their own publications. The implementation details and supported languages are described in [54]. More information is provided on the website. In this toolkit, CRFs, maximum entropy models and deep learning are adopted in different modules. Compared with NLTK and OpenNLP, Stanford CoreNLP supports more languages, including Arabic, Chinese, French, Spanish and German. Moreover, CoreNLP provides deep learning-based sentiment analysis tools which support English.

FudanNLP, LTP and NiuParser are developed for Chinese specifically. CRFs are commonly used models in these toolkits, especially for NiuParser.

In NiuParser, CRFs are adopted for Chinese word segmentation, POS tagging, named entity recognition and chunking. Variants of shift-reduce parser are adopted in the parsing module. A two stage classification approach with maximum entropy classifier is provided for semantic role labelling. More details can be found in [59].

In FudanNLP, CRFs are adopted for Chinese word segmentation and

POS tagging. Named entity recognition module is based on POS tagging, language model and rule-based strategies. Shift-reduce method is provided in the parsing module.

LTP provides SVM and maximum entropy model based approaches for POS tagging and named entity recognition, which is different from FudanNLP and NiuParser. In word sense disambiguation and semantic role labeling modules, SVM and maximum entropy model are adopted, respectively. A high order graph-based method is provided for dependency parsing. More details can be found in [58].

4. Opinion Mining and Sentiment Analysis: Methods and Resources

Opinion mining and sentiment analysis aim to extract the sentiment orientation of given texts. In general, opinion mining can be divided into three levels: document level, sentence level, and fine-grained level. The task of the document level is determining the overall polarity of a document which includes multiple sentences. This task makes the assumption that only a single opinion target is discussed in one document. Similar to document level opinion mining, the task of the sentence level is also a classification problem but focuses on each sentence in documents. It determines whether a sentence expresses positive, negative or neutral orientation. Subjectivity classification is another task at the sentence level, which extracts subjective and objective parts of documents. The above tasks pay no attention to extracting the detail information of opinions, such as opinion target and opinion holder. For example, "*The screen of this mobile phone is good.*" expresses a positive polarity to aspect "*screen*" of entity "*mobile phone*". Fine-grained opinion mining deals with this problem beyond classification techniques. From the document level to the fine-grained level, the complexities of problems are increasing. The lack of annotated corpora at the finer level makes it worse. Usually, supervised methods outperform unsupervised methods. However, requirements of annotated corpora are not always satisfied, especially for fine-grained level opinion mining, which propels researchers to develop semi-supervised or unsupervised methods. Moreover, the lack of annotated corpora is common in the cross-domain and cross-lingual situations, which we will discuss in Section 4.4 and Section 4.5, respectively.

Feature engineering plays an important role for machine learning approaches. We list most commonly used text features below, including n-gram features with Information Retrieve (IR) weighting schemes, syntactic features and semantic features.

N-gram features are the most fundamental text features which are commonly used in NLP. An n-gram is a set of contiguous n items with the corresponding frequency. In opinion mining, binary weights of unigram and bigram are widely adopted. Other IR weighting schemes can take the place of binary weights, such as tf-idf and some other variants [36, 37].

Syntactic features contain POS tags and syntactic information. These features are employed in two ways: constructing a feature space for machine learning approaches [35, 38, 60, 61, 62] and designing rules to extract required information, such as entities and aspects in fine-grained opinion mining [31, 63, 64].

Semantic features are conjunctions which indicate negation, intensification, and diminution. The negation is important for opinion mining as it reverses the sentiment orientation. Intensification and diminution are increasing and decreasing the strength of sentiment, respectively, which are also useful for opinion mining [40, 65, 41, 66].

4.1. Document Level Opinion Mining

Document level opinion mining is a task of extracting the overall sentiment polarities of given documents, such as movie reviews, product reviews, tweets and blogs. According to the definition of opinion mining, the objective of the document level opinion is identifying the third component of the quintuple. From the perspective of text categorization, document level opinion mining can be considered as a special case which takes sentiments into account rather than topics. However, while topic-based classification has achieved a high accuracy, document level opinion mining encounters a more complicated situation [35].

Turney [67] presented an unsupervised method to classify reviews as recommended or not recommended. It determines the polarities by averaging Semantic Orientation (SO) of appeared phrases in reviews. The SO of a pair of phrases is measured by Pointwise Mutual Information and Information Retrieval (PMI-IR) [68], which estimates PMI by the hitting number of a particular query. The SO is estimated as follows

$$\mathrm{SO}(phrase) = \log \frac{\mathrm{hits}(phrase,\ \text{“}excellent\text{”})\mathrm{hits}(\text{“}poor\text{”})}{\mathrm{hits}(phrase,\ \text{“}poor\text{”})\mathrm{hits}(\text{“}excellent\text{”})}. \tag{1}$$

Pang et al. [35] employed machine learning methods to classify reviews as positive or negative. The naïve Bayes classifier, maximum entropy classifier and SVM are trained on unigram, bigram, POS tag and position features. The authors indicated that the occurrences of conflicting opinions make document level opinion mining harder than topic-based text categorization.

Since features greatly affect performances, some work takes more text features into account. Kennedy and Inkpen [40] incorporated three types of valence shifters: negations, intensifiers and diminishers and proposed two methods based on lexicons and SVMs, respectively. Intesifiers and diminishers modify sentiments through addition and subtraction and negations reverse sentiment polarities. They performed experiments on a movie review dataset [39] and the result shows the effectiveness of valence shifters. Taboada et al. [41] presented a lexicon-based method that incorporates intensifications as well as negations and evaluate it on multiple datasets in different domains. The authors constructed a lexicon in which SO of each word ranges from -5 to 5. Intensifications and negations are represented by multiplication operations. Wang and Manning [34] analyzed the influences of features and datasets used in previous work and presented an SVM with naïve Bayes features. The presented model incorporates naïve Bayes log-count ratios as features and combines naïve Bayes and SVMs with naïve Bayes features by an interpolation that trade-offs the confidences of the two models.

Semi-supervised learning approaches have been proposed to reduce the dependence on annotated data. Dasgupta and Ng [69] proposed a semi-supervised method which takes the occurrences of sentimentally ambiguous reviews into consideration. The proposed method firstly determines whether a review is ambiguous or not by the spectral clustering algorithm. Then active learning is employed to label the most 10 uncertain reviews for five iterations. Finally, an ensemble of transductive SVMs is adopted to train a final classifier on labeled reviews. The experiment result shows that manually labeling a small number of ambiguous reviews could improve performances of document level opinion mining. Li et al. [70] proposed a semi-supervised method for imbalanced sentiment classification. Under-sampling is employed to generate multiple sets of balanced initial training data. Then two feature subspaces are selected by random subspace generation, and two classifiers are trained on the subspaces, respectively. Most positive reviews and most negative reviews from unlabeled reviews are selected according to the classification result. Finally, the above steps are repeated until all reviews are labeled. In this work, random subspace generation rather than a static one avoids the negative influences caused by improper selection of subspaces. Recently, a semi-stacking framework of semi-supervised learning methods was proposed to integrate two or more semi-supervised learning algorithms in an ensemble learning manner [71].

Unsupervised learning approaches, such as Latent Dirichlet Allocation (LDA) and its variants, have been developed to further reduce the dependence on annotated corpora. Lin and He [72] presented a Joint Sentiment/Topic (JST) model based on LDA. It determines sentiments and topics simultaneously for each document. In JST, words appeared in a document are generated according to not only the topics but also the sentiment polarities indirectly. Thus, the mutual influences between sentiments and topics are captured. Li et al. [73] proposed a Dependency-Sentiment-LDA model, which assumes that the sentiments of words form a Markov chain, i.e., the sentiment of a word is dependent on previous ones. The transitions of sentiments are determined by two types of conjunctions: related conjunctions, e.g., "and", and adversative conjunctions, e.g., "but". Both of the above models incorporate sentiment lexicons as prior information which could improve performances.

Because of the occurrences of conflicting opinions, some work incorporates discourse structures to address this problem. Somasundaran et al. [74] presented two frameworks to infer global discourse-based polarities for document level opinion mining. The first is an iterative collective classification based supervised framework where discourse relations and the corresponding neighborhood information are incorporated as features. The second is an unsupervised framework that incorporates discourse structures as constraints of integer linear programming. Both frameworks train local classifiers with sentiment lexicon and unigram features, and then propagate classification results along discourse structures. The proposed discourse-based approaches and their combination obtain substantial improvements which indicate the effectiveness of incorporating discourse structures into opinion mining. Trivedi and Eisenstein [75] proposed a model to capture relevance of discourse connectors. The discourse connectors are words and phrases that indicate shifts or continuation in discourse structures, e.g., "*but*" indicates shift and "*moreover*" indicates continuation. The proposed model incorporates subjectivity transitions between sentences as transition features. The performance of the proposed model shows the effectiveness of such linguistic knowledge. Bhatia et al. [76] proposed a discourse depth reweighting approach to determine the document level sentiment polarities. This approach reweights discourse units according to their positions in the dependency representation of discourse structures. The underlying idea is that the positions of constituents in discourse structures represent their importance to the overall sentiment polarity through some weighting schemes.

Factual and opinionated constituents coexist in documents, and the existence of objective sentences hinder sentiment classification performances. It is natural to remove these noises for better performances. Subjectivity classification is proposed to address this problem at the sentence level which will be introduced in Section 4.2. Spam detection and usefulness measurement are document level applications of opinion mining. We will introduce them in Section 8.

4.2. Sentence Level Opinion Mining

Opinion mining at the sentence level is similar to that at the document level, since sentences can be regarded as short documents. Besides sentiment classification, subjectivity classification is another problem at the sentence level. Subjectivity classification aims at determining whether a sentence is subjective or objective.

Yu and Hatzivassiloglou [38] presented three sentence level subjectivity classification approaches which are based on sentence similarity, naïve Bayes classifier and multiple naïve Bayes classifiers, respectively. The similarity-based approach computes similarity scores of given sentences to opinionated and factual sentences, respectively, and determines subjectivities by comparing similarity scores to a predetermined threshold. The naïve Bayes approach trains a classifier on sentences from opinionated and factual documents. The feature space contains unigram, bigram, trigram, POS tag and counts of sentiment words. The approach of multiple naïve Bayes classifiers trains multiple classifers on different subsets of the whole feature space. Then the sentences that have different labels with the corresponding documents are removed from the training set. Training and removing are iterative until no more sentences in the training set will be removed. Finally, classifiers trained on the reduced training sets are used to determine sentences as subjective or objective. Pang and Lee [39] proposed a graph-based algorithm to extract subjective parts of documents. The association and individual information of sentences are used as the weights in a graph. The original problem is formulated as a minimum cost problem that can be solved by the minimum cut algorithm. Kim and Hovy [32] presented a sentence level sentiment classification approach to find opinion holders as well as the corresponding sentiments. Named entity recognition is adopted to identify potential opinion holders including persons and organizations. Various window sizes are defined for sentiment regions. The sentiment polarity of each sentiment region is determined by averaging polarities and strengths of words in this region.

Several approaches have been proposed to model the dependencies between adjacent sentences. Täckström and McDonald [77] presented a model which combines fully supervised document labels and less supervised sentence labels. The model is analogous to hidden-state random condition fields [78] where observation nodes represent features of sentences, and the label of each sentence is treated as a latent variable. As the flexibility of the model, relations between adjacent sentences can be captured. The optimal state of each latent variable can be determined by a dynamic planning algorithm. The optimal states are sentiment polarities of sentences. Yang and Cardie [79] incorporated the discourse structure to capture the context dependence between sentences and presented a CRF model with posterior regularization [80]. The lexical and discourse information are incorporated as constrains of posterior regularization.

Tang et al. [81] proposed a joint segmentation and classification framework which simultaneously generates useful segmentations and predicts the sentence level polarity based on the segmentation units. There are three components in the proposed framework: a candidate generation model, a segmentation ranking model, and a sentiment classification model. The candidate generation model produces segmentations by a beam search method with constraints from a phrase dictionary. The segmentation ranking model computes a score for each segmentation according to a log-linear segmentation model with segmentations' polarities. Finally, the sentiment classification model is a linear kernel SVM with various hand-crafted features.

4.3. Fine-Grained Opinion Mining

Document and sentence level opinion mining provide useful information in many application scenes. However, more detailed information is necessary for other advanced applications. Fine-grained opinion mining is proposed to discover details in opinionated texts, which receives great interests. Fine-grained opinion mining has several variations including aspect (also known as feature or attribute) and concept [82, 83] level opinion mining. We will mainly discuss aspect level opinion mining and the related work about comparative opinion mining will be discussed in Section 5.

With the definition of the quintuple in Section 1, aspect level opinion mining discovers the first three components: entity, aspect and sentiment. For an online product review, the entity is explicit and several aspects of this entity with the corresponding opinions would be mentioned. The objective of aspect level opinion mining is to discover the specific targets (aspects or entities) and the corresponding sentiment polarities. Thus it is divided into two sub-tasks: target extraction and sentiment classification. The difficulties in aspect level opinion mining are the complicated expressions of opinions and lack of annotated corpora at the fine-grained level. The formal requires sophisticated techniques and the latter resorts to semi-unsupervised and unsupervised approaches.

The earliest work was presented in [31] where a pipeline, including product aspect mining, opinion sentence identification, sentiment classification and opinion summarization, is designed to perform aspect level opinion mining. In this work, the authors focused on explicit aspects, i.e., the aspects which are expressed by nouns or noun phrases. The POS tag of each word is tagged in advance. Association mining [84] is adopted to detect frequently mentioned items in all reviews and pruning techniques are used to reduce unlikely aspects. The opinion sentences identification and sentiment classification are performed in a lexicon-based manner. After the above steps, review summaries are generated according to the extracted aspects and the corresponding sentiment polarities and strengths. This kind of summarization is called aspect-based summarization which is different from traditional summarization. We will discuss further in Section 7. Popescu and Etzioni [63] proposed an unsupervised information extraction method for aspects and opinions. In the aspect extraction part, PMI scores between the noun phrases and the meronymy discriminators associated with the product class (e.g., "*of scanner*", "*scanner has*", "*scanner comes with*", etc. for the *Scanner* class) are used to determine likely aspects.

Qiu et al. [85] presented a double propagation algorithm to discover pairs of opinion word and opinion target. The algorithm starts from a set of seed pairs. The opinion words are used to find more opinion targets, and in turn the opinion targets are used to find more opinion words. However, the double propagation algorithm does not perform well on a large or small dataset where it achieves a low precision and low recall [86]. To increase the recall, Zhang et al. [86] introduced part-whole and "no" patterns into the double propagation algorithm. The part-whole patterns (i.e., meronymy discriminators) are used to indicate whether an aspect is a part of an entity or not. The "no" patterns are used to extract some short patterns of expressions which extensively exist in product reviews and forum posts. An aspect ranking algorithm is applied to find the most likely aspects according to the aspect relevance and frequency.

Gindl et al. [64] proposed a rule-based entity and aspect extraction method. Several linguistic rules, which are represented as regular expressions of POS tags, are designed to extract adjectival noun, adverbial noun and extended noun phrases. For the sake of detecting opinion targets across sentences, the authors adopted a heuristic anaphora resolution technique to deal with the coreference problem of personal pronoun and sentiment polarity propagation. It connects the personal pronoun to the last noun in the previous sentence and propagate sentiment through the connection. Such rule-based approach obtains robust performance in practice as it is unsupervised and domain independent. However, the rules require manually design and selection. Recently, an automatic rule selection method was presented in [87] which can select an effective subset of all rules with performance even better than the full rule set.

In reviews and posts, aspects may be expressed by different words or phrases which leads to redundancy in the aspect extraction step. Brody and Elhadad [88] introduced a local topic model with a small number of topics to automatically extract aspects (i.e., topics) of each product and the representative words of aspects are selected by mutual information between words and aspects. In this work, each sentence is treated as a document. The experimental result shows that unsupervised methods can obtain robust performance when applied to different domains, such as different products. Zhai et al. [89] proposed a semi-supervised approach to clustering aspects which have multiple expressions. Unlike other unsupervised methods, such as [88], two kinds of pre-existing knowledge (i.e., sharing words and lexical similarity information) are exploited to generate labeled aspects which cast the unsupervised problem into semi-supervised. The sharing words are common words in different aspect expressions which can be used to initialize the clusters of aspects. The lexical similarity information is used to extend clusters. The evaluation results show the effectiveness of the introduced knowledge.

Ding et al. [90] performed aspect level opinion mining on a dataset from a product discussion forum where multiple products and comparisons exist. They proposed two approaches to deal with the extraction of explicit and implicit entities respectively. The explicit entities are detected by an iterative expansion of the entity set starting with a seed set without any annotated corpus. The extraction of the implicit entity is based on the explicit entities and comparative sentences. Some rules are designed to deal with different cases including whether a sentence is comparative and whether explicit entities exist.

Conditional opinions exist in some reviews. For example, in the sentence "*This restaurant is fairly good for having lunch.*", "*for having lunch*" is condition of "*fairly good*". Nakayama and Fujii [91] proposed a method to extract conditional opinions from online reviews. The conditional opinions are useful for further analyses of reviewers' attributes, purposes and situations. The authors formulated this problem as a sequential labeling problem and adopted CRF with these labels: *BIO* (i.e., *Begin*, *Inside* and *Outside*), *Target*, *Aspect* and *OpinionWord*. In this work, *Target*, *Aspect*, *OpinionWord* are determined in advance which simplify the problem into three labels. Distance information towards *Target*, *Aspect* and *OpinionWord* and linguistic knowledge is adopted to construct the feature space. Experiment was performed on hotel reviews in Japanese.

Thet et al. [92] took clause structures into consideration and proposed a linguistic approach which exploits grammatical dependency structures of clauses. The grammatical dependency structure is obtained by a dependency syntactic tree, and then sub-trees which represent clauses are extracted (i.e., dividing a sentence into separate clauses which express opinions towards the corresponding aspects). The sub-tree generation can be regarded as a decomposition of a complex sentence into several simple sentences, which transforms the original problem to a sentence level task. The aspect extraction in this work is simplified to a semantic annotation process which gives a tag to each sentence by a set of predefined aspect indicating words. Lazaridou and Titov [93] introduced discourse structures into a generative model. The proposed model generates discourse cues, aspects, sentiment polarities and words sequentially.

Wang et al. [94] proposed an unsupervised model, based on restricted Boltzmann machines. The proposed model jointly extracts the sentiment and aspect. Similar to LDA-based models, this model incorporates aspect, sentiment, and background variables as latent variables. However, the latent variables in the proposed model have no conditional relationship, i.e., there are no edges connecting the latent variables in the graphical model.

4.4. Cross-Domain Opinion Mining

The sentiment is expressed differently in different domains, i.e., the distributions of words differ in different domains, which raises a problem on how to exploit the information from a domain with annotated corpora to perform opinion mining on other domains with few annotated resources. Manual annotation requires expensive costs, which is often not practical.

First, we give the description about cross-domain opinion mining. The source domain D_s is a set of annotated corpora while the target domain D_t is unannotated. Each domain has a probability distribution of words ($P_s(W_s)$ and $P_t(W_t)$, respectively). The opinion mining task is denoted as T_s and T_t. The problem is defined as follows: given a source domain D_s and opinion mining task T_s, a target domain D_t and opinion mining task T_t, the objective is to improve the performance of T_t by using the knowledge from D_s and T_s. Domain adaptation methods are adopted to address this problem and have become the commonly used techniques for cross-domain opinion mining tasks.

Blitzer et al. [23] proposed an algorithm based on Structural Correspondence Learning (SCL) [95] and a measure of domain similarity. The measure is correlated with the adaptation of a classifier from one domain to another. The SCL algorithm aligns sentiment words that express a similar polarity in different domains. The original SCL algorithm selects pivots by the frequency of words. In the proposed algorithm, mutual information between words and the sentiment label is adopted instead. The comparison of the above two pivot selecting methods shows that mutual information is more suitable for opinion mining tasks. The proposed domain similarity measure is based on Humber loss [96] and characterizes the domains by their distributions which can be used to calculate divergence of different domains. According to the similarity of different domains, one can determine which domain should be labeled to obtain the best performance on all available domains. The authors constructed a multi-domain sentiment dataset from *Amazon* product reviews including four different product types (books, DVDs, electronics and kitchen appliances). This dataset is commonly used in cross-domain opinion mining.

Liu and Zhao [24] presented a two-stage approach for the domain adaptation problem of sentiment classification. In the first stage, shared sentiment words are detected by bridging domain-specific words with common topics. If two sentiment words (e.g., *exquisite* and *noble*) in different domains express the same polarity (e.g., positive sentiment) in a common aspect (e.g., *appearance*), these words are treated similarly. In other words, the common aspects are the connection of domain-specific sentiment words. Based on Probabilistic Latent Semantic Analysis (PLSA) [97], a topic model named transfer-PLSA, which jointly models different domains, was proposed to find topic distributions for two different datasets. In the second stage, a scheme of retraining [98] is adopted. A classifier trained by source labeled data is used to select informative instances in D_t. These instances are appended to the training set. Then a new classifier is trained on the extended training set.

Pan et al. [25] proposed a framework, including graph construction and the Spectral Feature Alignment (SFA) algorithm, to construct a new representation of cross-domain sentiment data which reduces the gap between domains. The words are divided into two categories: domain-specific and domain-independent. Co-occurrence information between domain-specific words and domain-independent words is used to construct a bipartite graph. The spectral clustering algorithm is performed on the bipartite graph to co-align domain-specific and domain-independent words into word clusters which reduce the mismatches between domain-specific words and both domains. He et al. [99] proposed a domain adaptation method based on a modified JST model [72]. The modified JST model introduces a sentiment prior distribution for words to the JST model and extracts polarity-bearing topics which are used to augment the original feature space. Then feature selection is performed on the augmented feature space by the information gain. Finally, a classifier is trained on the resultant training set.

The above work [23, 24, 25, 99] utilizes the similarity of different domains. An active learning approach [100] was proposed to deal with the situation that the word distribution differs significantly in D_s and D_t. The proposed approach employs the manually annotated data in both the sample selection strategy and the classification algorithm. In the sample selection strategy, two classifiers are trained on source and target domain data, respectively, and both of them are exploited to select label-disagreed samples. Then the label-disagreed samples are labeled manually and appended to the training set. The sample selection iterates for several times. In the classification algorithm, a label propagation algorithm is adopted to predict labels of unlabeled data. Once all data are labeled, the original unlabeled data with prediction labels and the original labeled data are combined to construct a large training set for both source and target domains. Two classifiers are trained on the combined training set, respectively. Finally, the labels of the unlabeled data are determined by multiplication of posterior possibilities estimated by the above classifiers.

Additionally, some work focuses on sentiment lexicon creation and expansion in the cross-domain situation. Choi and Cardie [101] focused on domain-specific lexicon creation and proposed a lexicon adapting method based on integer linear programming that can adapt an existing lexicon into a new one which is more suitable for the target domain. The proposed method takes word level polarities, negations and expression (i.e., multiple words) level polarities as constrains of the optimization problem. The modifications of the original sentiment lexicon can be obtained through the optimization result. Bollegala et al. [102] proposed a multiple-domain sentiment classification method which uses an automatically created sentiment sensitive thesaurus. The authors formulated this problem as a sentiment lexicon expansion problem, where additional related sentiment words from the automatically created thesaurus are appended to feature vectors in both the training and test periods. The appended sentiment words reduce mismatches of sentiment polarities between the two domains. In the construction of the sentiment sensitive thesaurus, POS tags and lemmatization are employed to reduce sparseness in texts. Then the co-occurrence information of words and sentiment labels in the source domain are used to construct the thesaurus. In feature expansion, weighted correlations of candidates and existing sentiment words are calculated to determine which candidates would be selected. Unlike SCL and SFA which consider a single source domain, this method can be extended to multiple source domains.

Transfer learning plays an important role in cross-domain opinion mining. Pan and Yang [103] provided a review of transfer learning. In the cross-domain opinion mining, most of the situations are that D_s is labeled while D_t not. Thus domain adaptation methods are commonly used. Recently, Sun et al. [104] provided a comprehensive survey on multi-source domain adaptation including representative work in algorithms and theoretical results.

4.5. Cross-Lingual Opinion Mining

Sentiment resources (i.e., sentiment lexicons and annotated corpora) are crucial for opinion mining. However, most of the available resources are in English, which makes it difficult to perform accurate analysis for texts in other languages, such as Chinese and Japanese. It is expensive to manually label reliable sentiment corpora or create sentiment lexicons in another language. In this situation, the language which has abundant and reliable resources is called source language (e.g., English). The resource-lacking language is called target language (e.g., Chinese, Japanese). Cross-lingual opinion mining are proposed to identify sentiment polarities of texts in the target language by the resources in the source language. It is analogous to the situation of cross-domain opinion mining where the source domain plays the same role of the source language and the target domain plays the same role of the target language. The common idea in recent work is also transfer learning, and domain adaptation methods are adopted widely.

Mihalcea et al. [26] utilized a bilingual dictionary and parallel corpora to bridge the language gap. The bilingual dictionary is used to build sentiment lexicons in the target language by translating existing English lexicons. An annotation projecting technique is used to create sentiment annotated corpora in the target language from parallel corpora. Banea et al. [27] adopted the Machine Translation (MT) technique to produce pseudo-corpora in the target language instead of relying on the manually translated parallel corpora. Other multilingual resources, such as Wikipedia[6], can be exploited for cross-lingual analysis [29]. So far, MT becomes an important technique in cross-lingual opinion mining. Duh et al. [105] analyzed instance mismatches and labeling mismatches in MT-based methods and concluded that MT is ripe for cross-lingual opinion mining. Google translator[7], Bing translator[8], and Microsoft translator[9] are commonly used translation engines which provide APIs to access the services conveniently. However, the sentiment shifting and limited vocabulary coverage problems are unavoidable in MT, which hinders the performance of MT-based approaches.

In the above studies [26, 27], the approaches employ classifiers (e.g., naïve Bayes and SVM) directly on the translated training set. Wan [106] performed cross-lingual opinion mining for English and Chinese reviews. They translated Chinese reviews to English and determined sentiment polarities for reviews in both languages by individual resource, respectively, in a lexicon-based manner. Then ensemble methods are employed to combine the results from both languages. The experimental result shows that although MT brings about errors into corpora, the performance is better than that using Chinese resources alone. And the ensemble result is better than individual

[6]`https://www.wikimedia.org/`

[7]`http://translate.google.com`

[8]`https://www.bing.com/translator`

[9]`https://www.microsoft.com/en-us/translator/default.aspx`

results. A co-training approach was proposed by Wan [28] where English features and Chinese features are treated as two independent views. The proposed approach makes use of English and Chinese reviews together to improve classification performances. MT is also used to produce pseudo-reviews in both Chinese and English. In the training phase, the co-training algorithm learns two classifiers: C_{en} for English reviews and C_{cn} for Chinese reviews. The learned classifiers C_{en} and C_{cn} are used to classify unlabeled reviews. The predicted reviews with high confidence are appended to the training set. Then the above steps iterate for several times. Finally, unlabeled reviews are classified by averaging the results from the English classifier and Chinese classifier.

Cumulative noise in transfer learning hinders performances when large training datasets are used [107]. Several methods have been developed for reducing the noise introduced by transfers. Wei and Pal [108] regarded cross-lingual opinion mining as a special case of the domain adaptation problem. An SCL based approach was proposed to learn a low-dimensional representation of texts that captures similarities of sentiment words between both languages by pivot words (i.e., words which represent similar sentiment polarities in both languages). The pivot words are selected by mutual information of words in different languages. The translated texts are used only through the pivot words. The non-pivot words are removed to reduce the influence of sentiment shifting. Chen et al. [109] proposed a credible boosting framework to perform cross-lingual opinion mining. A knowledge validation method, including label consistency checking and linguistic distribution checking, is adopted to prevent the negative influence from wrong knowledge and distinguish knowledge with high confidence. It is different from the co-training algorithm which transfers knowledge between each other. The proposed framework transfers knowledge from the source language to the target language mono-directionally. The knowledge is represented in the semantic space with low-dimensional document vectors [110], which makes it feasible to measure distances of reviews. Knowledge validation functions exploit the distance information of reviews and their labels in both languages. The proposed model obtains promising results, which shows the effectiveness of knowledge validation.

MT-based methods suffer from the limited coverage of vocabulary in the MT results. Meng et al. [111] proposed a generative Cross-Lingual Mixture Model (CLMM) to model unlabeled bilingual parallel corpora and discover unseen sentiment words, which improves the coverage of vocabulary. CLMM defines a generative mixture model for generating a parallel corpus. The unobserved polarities of documents are introduced as hidden variable (i.e., C_s and C_t, here s and t represent "source language" and "target language", respectively) and the observed words are produced by a set of generation distributions conditioned on the hidden variables (i.e., $P(w_s|C_s)$ and $P(w_t|C_t)$). The conditional probability $P(w_t|C_t)$ can be used to determine the polarities of texts in the target language. Standard MT techniques may rearrange sequences of words, which introduce errors and leave some parts without the corresponding segments in the source language [112]. To overcome this shortcoming, Lambert [112] proposed an aspect level cross-lingual sentiment classification method based on constrained MT which preserves opinion structures. Moses statistical MT [113] is adopted which allows customizations of translation results.

Additionally, a sentiment lexicon in the cross-lingual situation is also important. Gao [114] treated sentiment lexicon creation as word-level sentiment classification. First, the authors constructed a bilingual word graph where words in different languages are represented as nodes. The edges between words in the same language are constructed by synonym and antonym dictionaries. Large parallel corpora and word alignment information are used to construct edges between words in the different language. A bilingual word graph label propagation algorithm was developed to identify polarities of unlabeled words in both languages.

4.6. Summary of Reviewed Techniques

In Table 2 we provide a summary for the reviewed techniques in this section.

Table 2: Summary of Reviewed Techniques

Level	Technique	Description
Document	Supervised approaches	Traditional classifiers in machine learning, such as naïve Bayes classifier, SVM and maximum entropy classifier, are used for document level sentiment classification on various kinds of features, including unigram, bigram, POS tags, position information [35], semantic features [40] and discourse features [74, 75]. Combined classifiers based on SVM and naïve Bayes classifier have also been proposed [34].

	Probabilistic generative model based approaches	Inspired by the LDA topic model, some generative models were proposed, including joint sentiment topic model [72] and dependency-sentiment-LDA model [73], which model the transitions between sentiments of words with a Markov chain.
	Unsupervised lexicon-based approaches	Sentiment orientations of words provide sentiment information of documents and averaged sentiment orientation of occurred words was used to indicate the overall sentiment [67]. Some reweighting schemes of sentiment orientations were proposed to improve the performances, such as intensification and negation indicators [41] and discourse structure-based reweighting scheme [76].
	Semi-supervised approaches	To reduce the dependence on annotated corpora, some semi-supervised approaches have been proposed, including active learning based method [69] which manually labels sentimentally ambiguous documents and co-training method [70] for imbalanced sentiment classification. Recently, an ensemble learning framework which combines semi-supervised methods was proposed [71].
Sentence	Supervised approaches	Similarly to the situation of document level opinion mining, supervised classifiers are adopted. Naïve Bayes classifier and ensemble of Naïve Bayes classifiers were adopted for detecting subjectivity of sentences [38]. CRFs were adopted for exploiting the dependencies of sentences [79]. Recently, a joint segmentation and classification framework was proposed [81].
	Unsupervised approaches	A graph-based approach which exploits association and individual information from sentences was proposed for sentence level subjectivity classification [39]. Similar to [67], a lexicon-based approach was proposed for sentence level sentiment classification [32].
	Semi-supervised approaches	Due to the lack of sentence labels, a sequential model that combines fully supervised document labels and less supervised sentence labels was proposed to perform semi-supervised classification [77].

Fine-grained	Unsupervised approaches	Because of the less annotation, unsupervised approaches play an important role on fine-grained level opinion mining. For aspect detection, association mining algorithm has been adopted [31]. And linguistic knowledge, such as meronymy discriminators [63] and part-whole patterns [86], were also taken into account. Double propagation algorithm was proposed for joint opinion words and aspects extraction [85]. Additionally, rule-based methods are also effective for detection of explicit entities and aspects [64]. Comparative sentences are used for detection of implicit aspects [90]. Clause structures are exploited for splitting documents into sentences which is helpful for aspect detection [92].
	Probabilistic generative model based approaches	LDA topic model and its variants are adopted for aspects detection [88] and jointly aspect and sentiment detection [93, 94].
	Semi-supervised approaches	An aspect clustering method which incorporates external knowledge has been proposed for semi-supervised aspect detection [89].
Cross-domain	Domain adaption based approaches	Domain adaption methods have been widely applied in cross-domain opinion mining, such as SCL [23] and SFA [25] where the words are aligned according sentiment polarities expressed in different domains. Some probabilistic models were also adopted, including transfer-PLSA [24] and modified JST [99]. An active learning method was proposed to reduce label-disagreed samples from different domains [100].
	Cross-domain lexicon based approaches	These approaches aim at adapting original sentiment lexicons to be domain-specific which makes them appropriate for target domains [101, 102].

Cross-lingual	Combination of single lingual approaches	The main idea of these methods are bridging the gap among different languages. Bilingual dictionaries and existing parallel corpora are exploited to create annotated pseudo-corpora [26]. Machine translation techniques also achieve this end, which are less restricted by parallel corpora [27, 112]. Ensemble of classifiers trained on different language improves performances [106]. Some methods were proposed to reduce the noises which are introduced by machine translation, including SCL based approach [108] and credible boosting approach [109].
	Cross-lingual lexicon based approaches	To improve the coverage of vocabulary in machine translation based approaches, a label propagation algorithm was developed to identify polarities of unlabeled words in both languages [114]. To the same end, a generative cross-lingual mixture model was proposed to model unlabeled bilingual parallel corpora and discover unseen sentiment words [111].
	Semi-supervised approaches	A co-training approach was proposed, in which English features and Chinese features are treated as two independent views [28].

4.7. Available Resources

Sentiment resources, i.e., corpora and lexicons, are important for opinion mining. Now we provide an overview of available resources which are commonly used.

4.7.1. Corpora for Opinion Mining

Annotated corpora are necessary for supervised methods. Table 3 presents an overview of annotated corpora for opinion mining.

Table 3: Annotated Corpora for Opinion Mining

Corpora	Language	Description

MPQA Opinion Corpora [115]	English	This corpus contains news articles manually annotated using an annotation scheme for opinions. Several versions annotated in different levels are provided. http://mpqa.cs.pitt.edu/corpora/mpqa_corpus
Movie Review Polarity Dataset [35]	English	The latest version of this dataset contains 1000 positive and 1000 negative processed reviews. http://www.cs.cornell.edu/people/pabo/movie
Movie Review Subjectivity Dataset [39]	English	This dataset includes 5000 subjective and 5000 objective processed sentences. http://www.cs.cornell.edu/people/pabo/movie
Multi-Domain Sentiment Dataset [23]	English	The dataset is constructed by Amazon product reviews for books, DVDs, electronics and kitchen appliances. Two kinds of datasets are available, one with the number of stars, the other with positive or negative labels. https://www.cs.jhu.edu/~mdredze/datasets/se

Besides the above manually annotated corpora, Pak and Paroubek [116] proposed a method to collect texts with the corresponding polarities from abundant opinionated posts in twitter, which makes it possible to construct large corpora without manual annotation.

4.7.2. Lexicon Resources for Opinion Mining

Lexicons are important for both lexicon and machine learning approaches. In Table 4, we provide a brief overview of some widely used lexicons.

Table 4: Sentiment Lexicons for Opinion Mining

Lexicon	Language	Description
Bing Liu's Opinion Lexicon [31]	English	The latest version of this lexicon includes 4,783 negative words and 2,006 positive ones. http://www.cs.uic.edu/~liub/FBS/sentiment-a

MPQA Subjectivity Lexicon [117]	English	This lexicon includes 8,222 words with their subjectivities (strong or weak), POS tags and polarities. `http://mpqa.cs.pitt.edu/lexicons/subj_lexic`
SentiWordNet [118]	English	SentiWordNet associates words to numerical scores ranging in $[0.0, 1.0]$ which indicate the positivity, negativity and neutrality. For each word, the three scores sum up to 1.0. `http://sentiwordnet.isti.cnr.it/`
Harvard General Inquirer [119]	English	Harvard General Inquirer contains 182 categories including positive and negative indicators. 1,915 positive words and 2,291 negative words are marked. `http://www.wjh.harvard.edu/~inquirer/`
LIWC [120]	English	Linguistic Inquiry and Word Counts (LIWC) provides a lot of categorized regular expressions including some sentiment related categories such as "Negate" and "Anger". `http://liwc.wpengine.com`
HowNet [121]	Chinese & English	HowNet provides a Chinese/English vocabulary for sentiment analysis, including 8,942 Chinese entries and 8,945 English entries. `http://www.keenage.com/html/e_index.html`
NTUSD [122]	Chinese	NTU Sentiment Dictionary provides 2,812 positive words and 8,276 negative words in both simplified and traditional Chinese. `http://academiasinicanlplab.github.io/`

Some methods have been proposed to create lexicons from some seed words through synonyms and antonyms [31, 32, 33]. However, the above methods are domain independent. As mentioned in Section 4.4, sentiment orientations of words may be inconsistent in different domains, which is the shortcoming of unsupervised methods. Learning lexicons from annotated corpora in a particular domain would obtain a better performance [123, 124, 125].

5. Comparative Opinion Mining

Comparison is a common way to express opinions. When reviewers post their opinions towards a target product, it is natural to compare with related products from different aspects and the compared products are usually potential rival products. The information contained in comparative reviews can help enterprises discover potential risks and further improve products or marketing strategies [4].

A comparative opinion expresses a relation of similarities or differences between two targets [2]. The mentioned targets and preferences of opinion holders are valuable information. To this end, comparative opinion mining was proposed. Compared entities, comparative words and aspects can be extracted from comparative sentences. For instance, in the sentence "*Phone X's screen is better than phone Y.*", "*phone X*" and "*phone Y*" are the compared entities, "*better*" is the comparative word and "*screen*" is the compared aspect. Obviously, "*phone X*" is preferred because the comparative word "*better*" expresses preference clearly. As shown above, words of comparative or superlative form, e.g., "*better*", "*worse*", and "*best*", determine the preference expressed in sentences. However, many comparative words, e.g., "*longer*", "*smaller*", are not opinionated or express different sentiment orientations (i.e., positive or negative) in different contexts. Opinions that depend on contexts are called implicit opinions. For instance, the word "*longer*" expresses positive orientation for entity feature "*battery life*" but negative for "*response time*".

Jindal and Liu [126] proposed a rule-based method for comparative sentence analysis. The authors decomposed this problem into two sub-tasks: comparative sentence identification and comparative relation extraction. Two types of rules, Class Sequential Rules (CSRs) [127] and Label Sequential Rules (LSRs) are designed to deal with the two sub-tasks, respectively. The CSRs are sequential rules with class labels (i.e., "*comparative*" or "*non-comprative*") which can be extracted by sequential pattern mining. A naïve Bayes classifier with CSR features is trained to identify comparative sentences. Then comparative sentences are classified into four types: *non-equal gradable*, *equative*, *superlative* and *non-gradable*. The first three types of comparative sentences are targets which are used for further analysis. The entities, comparative words, and entity aspects are extracted by applying LSRs on the three types of comparative sentences, respectively. However, the proposed method does not find preferred entities.

Ganapathibhotla and Liu [128] analyzed different forms of comparative sentences and proposed a method to identify preferred entities for different types of comparative sentences. The authors divided comparative sentences into two categories: *opinionated comparatives* and *comparatives with context-dependent opinions*. In the former case, preferred aspects could be identified according to the comparative words. In the latter case, external information is required to determine the preferences. Pros and Cons in reviews, which are key words that briefly express reviewers' opinions, are used by the proposed method. The Pros and Cons contain important information about reviewers' preferences. According to whether C and A are opinionated or not, various methods are developed to determine the preferred aspects for different cases. When C and A are not opinionated, a measure, called One-Side Association (OSA), is proposed to measure the association of comparative word C and aspect A,

$$\mathrm{OSA}\,(A, C) = \log \frac{\Pr(A, C)\Pr(C|A)}{\Pr(A)\Pr(C)}, \tag{2}$$

which is used to identify preferences. Given comparative word C and aspect A, the corresponding OSA values are computed for both positive and negative cases, denoted by $\mathrm{OSA}_P(A, C)$ and $\mathrm{OSA}_N(A, C)$ respectively. The $\mathrm{OSA}_P(A, C)$ ($\mathrm{OSA}_N(A, C)$) reflects co-occurrence of C and A in Pros (Cons). Synonyms and antonyms of C and A are also adopted in computation. If $\mathrm{OSA}_P(A, C) > \mathrm{OSA}_N(A, C)$, the former entity is preferred and vice versa. For other cases, the preferences can be identified by other rules.

Xu et al. [4] proposed a CRF-based model to extract comparative relations between products from customer reviews. Generally, one comparative word represents one comparative relation. Some sentences that contain more than one comparative words bring along difficulties for analysis. The proposed model takes interdependencies among comparative relations into consideration, which is useful for capturing comparative information in more complex situations than the above work [126, 128]. The proposed model is a two-level CRF with unfixed interdependencies. In the graphical representation of the proposed model, the highest level represents relations, the middle level represents entities and the bottom level represents words.

6. Deep Learning for Opinion Mining

Deep learning is a kind of approach with multiple levels of representation learning, which has become popular in applications of computer vision, speech recognition and natural language processing. In this section, we introduce some successful deep learning algorithms for natural language processing.

With the rapid growth of deep learning, many recent studies expect to build low-dimensional, dense, and real-valued vector as text features for opinion mining without any feature engineering. The task of opinion expression extraction is formulated as a token-level sequential labeling task. In order to address such problem, a lot of studies use CRF or semi-CRF with manually designed discrete features such as word features, phrase features, and syntactic features [129, 130]. Recurrent Neural Networks (RNNs) are popular models that have shown great promise in many NLP tasks. An RNN is an extension of a conventional feedforward neural network, which is able to handle variable length input sequences. Thus, RNNs are naturally applicable for language modeling and other related tasks. Irsoy and Cardie [131] applied Deep Recurrent Neural Networks (DRNNs) to extract opinion expressions from sentences and showed that DRNNs outperform CRFs. The method is constructed by stacking Elman-type RNNs on top of each other. Every layer of the DRNN treats the memory sequence from the previous layer as the input sequence, and computes its own memory representation. In the field of NLP, syntactic parsing is also a central task because of its importance in mediating between linguistic expressions and meanings. Socher et al. [132, 133] introduced a Compositional Vector Grammar (CVG), which combines Probabilistic Context-Free Grammars (PCFGs) with a syntactically untied RNN that learns syntactic-semantic, compositional vector representations. Socher et al. [55] proposed a model called Recursive Neural Tensor Network (RNTN). They represented a phrase through word vectors and a parsing tree and then computed the vectors for higher nodes in the tree by the same tensor-based composition function. Paulus et al. [134] investigated an analogous tree structured RNN for fine-grained sentiment analysis.

Over the years researchers have developed more sophisticated types of RNNs to deal with some of the shortcomings of the vanilla RNN model. Bidirectional RNNs are based on the idea that the output at time t may depend on not only previous elements in the sequence, but also future elements. For example, to predict a missing word in a sequence one would look at both the left and the right context. Bidirectional RNNs are quite simple. They are just two RNNs stacked on top of each other. The output is then computed based on the hidden states of both RNNs. Deep bidirectional RNNs are similar to bidirectional RNNs, though we now have multiple layers per time step. In practice this gives us a higher learning capacity. Mikolov et al. [135, 136] presented several modifications of the original RNN language model.

Long Short-Term Memory (LSTM) [137] is specifically designed to model long-term dependencies in RNNs. LSTMs do not have a fundamentally different architecture from RNNs, but they use a different function to compute the hidden states. The memories in LSTMs are called cells which take the previous state h_{t-1} and current observation x_t as inputs. Internally these cells decide what to keep in and what to erase from memory. They then combine the previous state, current memory, and current observation. It turns out that these types of units are very efficient at capturing long-term dependencies. LSTMs have obtained strong results on a variety of sequence modeling tasks. Sequential models like RNNs and LSTMs are also verified as powerful approaches for semantic composition [138]. Liu el al. [139] proposed a general class of discriminative models based on RNNs and word embeddings that can be successfully applied to fine-grained opinion mining without any task-specific feature engineering effort.

Another powerful neural network for semantic composition is Convolutional Neural Networks (CNNs). Kalchbrenner el at. [140] described a convolutional architecture called Dynamic Convolutional Neural Networks (DCNNs) for semantically modeling of sentences. The network uses dynamic k-max pooling, a global pooling operation over linear sequences. The network handles input sentences with variable lengths, and induces a feature graph over the sentences. The feature graph is capable of explicitly capturing short and long-range relations.

In the meantime, the advances in word representation using neural networks have contributed to the advances in opinion mining by using deep learning methods. A pioneering work in this field is given by Bengio et al. [141]. The authors introduced a neural probabilistic language model that learns a continuous representation for words and a probability function for word sequences based on the word representations. Mikolov et al. [142, 143] introduced Continuous Bag-of-Words (CBOW) and skip-gram language models, and released the popular word2vec[10] toolkit. The CBOW model predicts the current word based on the embeddings of its context words, and the skip-gram model predicts surrounding words according to the embedding of the

[10] https://code.google.com/p/word2vec/

current word. Pennington et al. [144] introduced Global Vectors for Word Representation (GloVe), an unsupervised learning algorithm for obtaining vector representations of words. Training is performed on aggregated global word-word co-occurrence statistics from a corpus, and the resultant representations show interesting linear substructures of the word vector space.

7. Opinion Summarization

We have introduced the classification and analysis techniques for reviews or other opinionated texts. Now we turn to introduce opinion summarization. Opinion summarization can be regarded as multi-document summarization. However, it is quite different from traditional text summarization, since the opinion summarization focuses on the opinionated parts and the corresponding sentiment orientations while traditional summarization focuses on extracting informative parts and removing redundancy [1]. As we mentioned in Section 4.3, aspect level opinion mining extracts entities and aspects with sentiment polarities which can be used to provide summaries by ranking them according to sentiment strengths [31]. More discussions can be found in [145]. This form of summarization is called aspect-based opinion summarization and is commonly adopted in industrial communities. In this section, we mainly discuss the summarization techniques for opinion mining, including both abstractive and extractive approaches.

Carenini [146] presented and compared two traditional summarization approaches for opinion summarization. The proposed sentence extraction approach is based on MEAD [147], which is a framework for multi-document summarization. The author indicated that sentence extraction based summarization cannot provide quantitative information, such as the frequency of each aspect and ratio of each sentiment orientation. The proposed language generation approach performs aspect level opinion mining firstly. Then it aggregates the extracted information according to the importance of aspects and the strength of sentiments. Finally, the summary is produced by a generator of evaluative arguments [148]. Quantitative information is involved in the generated summary.

Lerman et al. [149] analyzed the result of human evaluation for various sentiment summarizers. The authors found that summarizers with sentiment information are favored, which indicates the necessity of aspect level opinion mining. The authors proposed an improved opinion summarizer that exploits multiple summarizers. The human evaluation data are adopted to train a preference ranking classifier which is used to determine the most appropriate summary from all summarizers.

Nishikawa et al. [150] formulated opinion summarization as an optimization problem in a graph. In the graph, each node represents a sentence. The proposed algorithm is an extractive method which constructs a path that passes through several sentence nodes. The objective function of the optimization problem is the sum of content scores which measures the importance of sentences and coherence scores which measures coherence degrees between sentences according to various sentence features including content words, POS tags of content words, named entity tags and conjunctions. With a constraint on the summary length, a path that has a maximum function value is constructed. The sentences in this path constitute the summary.

Gerani et al. [151] took discourse structures into account and proposed an abstractive summarization approach for product reviews. The first step is parsing reviews to discourse trees and modifying the trees to ensure that leaf nodes only contain aspect words. Then the aspect discourse trees are aggregated to generate a directed graph where each node represents an aspect with an importance measure, and each edge represents the relation between two aspect nodes with the corresponding strength of the relation. The relation represents the variation of the importance measure from one aspect to another. The most important aspects with the corresponding relations between them are selected by the PageRank algorithm. Finally, based on the selected aspects and relations, a template-based natural language generation framework is employed to generate summaries. The importance measures, relation types, and relation strengths are used to determine the polarity words, connective words, quantifiers and so on.

8. Advanced Topics

As the increase of user generated contents, such as online product reviews, a large amount of information is available as references to make purchase decisions for individuals and develop strategies for enterprises [3, 5, 4]. However, online reviews are not always credible as the existence of fake reviews, and the quality of reviews is uneven, which makes it difficult to obtain useful information quickly. Opinion spam detection and usefulness measurement of reviews are proposed to deal with the above problems, respectively. Both tasks aim to improve the effectiveness of opinion mining but in different manners: opinion spam detection reduces the suspect reviews while usefulness measurement emphasizes the credible ones. It is worth noting that these tasks are not quite the same as previous opinion mining tasks: much non-textual information is integrated into the feature space, such as reviewer information, reviewers' activities.

8.1. Spam Detection

Opinion spam is quite different from other spam, such as email spam, since opinion spam is a kind of deception which is hard to determine even for humans [152]. Additional non-textual information plays an important role for opinion mining.

Jindal and Liu [152] showed the particularity of opinion spam and proposed a supervised approach to determine whether reviews are fake or not. The review content, reviewer information and product information are adopted to construct the feature space. Instead of detecting the fake reviews, some approaches are proposed to find the spammers who post fake reviews frequently. Mukherjee et al. [18] proposed an unsupervised generative model which treats the spamicity of a reviewer as a latent variable. This model assumes that behavioral distributions of spammers and non-spammers are different, and clusters reviewers into two classes. Various reviewer and review features are extracted, such as review content similarities, the number of reviews, reviewing burstiness, the ratio of the first reviews, duplicate/near duplicate reviews, extreme rating, rating deviations, early time frames and rating abuses.

Recently, Li et al. [17] presented the first large-scale analysis of restaurant reviews filtered by *Dianping*[11]. The dataset from *Dianping* contains richer information than previous ones, including users' IP addresses and users' profile. The temporal and spatial patterns of spammers were studied.

8.2. Usefulness Measurement

Typically, three factors are involved in reviews: reviewer, product, and rater. The usefulness of a review is commonly reflected by the raters, which avoids the unsupervised situation confronted in opinion spam detection [2]. Usefulness measurement of reviews aims to rank the reviews according to the degree of usefulness, which is usually formulated as a regression problem with the features of review lengths, review rating scores, POS tags, sentiment words, tf-idf weighting scores and so on [153, 154]. Other features are

[11] `https://www.dianping.com/`

also taken into account, such as the subjectivity of reviews, reviewers' expertises, the timeliness of reviews, and review styles and social contexts of reviewers [19, 20, 21, 22]. The above approaches assume that the helpfulness of a review is independent of the raters. However, the raters evaluate the helpfulness according to their backgrounds, and the corresponding influences should be taken into consideration. To this end, Moghaddam et al. [155] proposed tensor factorization-based models to predict the personalized review quality. The factorization is performed on the three-dimensional tensor *rater* $\times$ *reviewer* $\times$ *product*. The proposed models outperform regression approaches.

9. Challenges and Open Problems

We now turn to discuss some challenges and open problems for opinion mining.

9.1. Annotated Corpora

Semi-supervised and unsupervised approaches have been employed to deal with the lack of annotated corpora [109, 28, 108, 23, 94]. However, the annotated corpora are still crucial for improving the performances of opinion mining systems. Most of the existing annotated corpora are in English, which is the main obstacle for opinion mining in other languages. In cross-lingual opinion mining, the existing methods, such as transfer learning and machine translation based methods, have their respective drawbacks (i.e., the cumulative error in transfer learning, and the sentiment shifting and limited vocabulary coverage problem in machine translation). Moreover, the annotations are most at the document or sentence level. For fine-grained opinion mining, manual annotation is unavoidable to some extent [91]. In terms of machine learning, large training sets make sense for obtaining better performances [116, 123, 124, 125]. As a whole, sophisticated approaches are required to deal with this situation and more efforts on annotation are expected.

9.2. Cumulative Errors from Preprocessing

Opinion mining is a high-level NLP task which depends on several subtasks, including word segmentation, POS tagging and parsing. Word segmentation is important for some languages, such as Chinese, Japanese. POS tagging and parsing are used to producing useful information for sentiment classification [35, 38, 60, 61, 62, 92] and entity/aspect extraction [31, 63, 64]. Although such sub-tasks have achieved promising performance for news or other canonical documents, it is unsatisfactory when applied to reviews which are usually not grammatical. This problem is worse for Chinese as it is more flexible than English. To deal with this problem, more accurate approaches for free form documents are required, and end-to-end approaches are expected to bypass the preprocessing steps.

9.3. Deep Learning

Deep learning is an emerging field of machine learning for both research and industrial communities. It has been applied to sentiment classification, aspect extraction, lexicon creation and so on [131, 134, 139]. And some work has incorporated sentiment information into word representation learning. However, other subfields of opinion mining, such as opinion summarization, have not been studied much. In addition, deep learning has greatly improved the state-of-the-art of computer vision and speech recognition. However, it did not obtain such a remarkable improvement for NLP, which indicates that more efforts are needed on developing deep learning approaches for NLP. In summary, how to improve the existing work and discover new applications of deep learning for opinion mining are the problems that need further studies.

10. Conclusion

In this paper, we introduced the problem of opinion mining and gave a brief illustration for information fusion in opinion mining. We investigated various approaches to opinion mining for different levels and situations. Opinion mining for comparative sentences was introduced separately as the particularity of comparative sentences. We presented some representative work of deep learning for NLP and the progress of deep learning for opinion mining. Then two forms of opinion summarization were introduced. After that, we investigated opinion spam detection and usefulness measurement which have great significance for practical applications. Finally, we discussed some challenges and open problems in opinion mining, which require further research.

Acknowledgements

This work is supported by the National Natural Science Foundation of China under Projects 61673179 and 61370175, Shanghai Knowledge Service

Platform Project (No. ZF1213), and the Fundamental Research Funds for the Central Universities.

References

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

[2] B. Liu, Sentiment analysis and opinion mining, Synthesis Lectures on Human Language Technologies 5 (2012) 1–167.

[3] N. Archak, A. Ghose, P. G. Ipeirotis, Show me the money!: Deriving the pricing power of product features by mining consumer reviews, in: Proceedings of the 13th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, 2007, pp. 56–65.

[4] K. Xu, S. S. Liao, J. Li, Y. Song, Mining comparative opinions from customer reviews for competitive intelligence, Decision Support Systems 50 (2011) 743–754.

[5] Y. Liu, X. Huang, A. An, X. Yu, ARSA: A sentiment-aware model for predicting sales performance using blogs, in: Proceedings of the 30th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval, 2007, pp. 607–614.

[6] W. Zhang, S. Skiena, Trading strategies to exploit blog and news sentiment, in: Proceedings of the 4th International AAAI Conference on Weblogs and Social Media, 2010, pp. 375–378.

[7] B. O'Connor, R. Balasubramanyan, B. Routledge, N. Smith, From tweets to polls: Linking text sentiment to public opinion time series, in: Proceedings of the 4th International AAAI Conference on Web and Social Media, 2010, pp. 122–129.

[8] A. Tumasjan, T. O. Sprenger, P. G. Sandner, I. M. Welpe, Predicting elections with Twitter: What 140 characters reveal about political sentiment, in: Proceedings of the 4th International AAAI Conference on Weblogs and Social Media, 2010, pp. 178–185.

[9] N. Chambers, V. Bowen, E. Genco, X. Tian, E. Young, G. Harihara, E. Yang, Identifying political sentiment between nation states with social media, in: Proceedings of the Conference on Empirical Methods in Natural Language Processing, 2015, pp. 65–75.

[10] C. Nopp, A. Hanbury, Detecting risks in the banking system by sentiment analysis, in: Proceedings of the Conference on Empirical Methods in Natural Language Processing, 2015, pp. 591–600.

[11] J. Shawe-Taylor, S. Sun, A review of optimization methodologies in support vector machines, Neurocomputing 74 (2011) 3609–3618.

[12] D. M. Blei, A. Y. Ng, M. I. Jordan, Latent Dirichlet allocation, Journal of Machine Learning research 3 (2003) 993–1022.

[13] B. Khaleghi, A. Khamis, F. O. Karray, S. N. Razavi, Multisensor data fusion: A review of the state-of-the-art, Information Fusion 14 (2013) 28–44.

[14] J. A. Balazs, J. D. Velásquez, Opinion mining and information fusion: A survey, Information Fusion 27 (2016) 95–110.

[15] D. Tang, B. Qin, T. Liu, Learning semantic representations of users and products for document level sentiment classification, in: Proceedings of the 53rd Annual Meeting of the Association for Computational Linguistics and the 7th International Joint Conference on Natural Language Processing, 2015, pp. 1014–1023.

[16] E.-P. Lim, V.-A. Nguyen, N. Jindal, B. Liu, H. W. Lauw, Detecting product review spammers using rating behaviors, in: Proceedings of the 19th ACM International Conference on Information and Knowledge Management, 2010, pp. 939–948.

[17] H. Li, Z. Chen, A. Mukherjee, B. Liu, J. Shao, Analyzing and detecting opinion spam on a large-scale dataset via temporal and spatial patterns, in: Proceedings of the 9th International AAAI Conference on Web and Social Media, 2015, pp. 634–637.

[18] A. Mukherjee, A. Kumar, B. Liu, J. Wang, M. Hsu, M. Castellanos, R. Ghosh, Spotting opinion spammers using behavioral footprints, in:

Proceedings of the 19th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, 2013, pp. 632–640.

[19] A. Ghose, P. G. Ipeirotis, Designing novel review ranking systems: Predicting the usefulness and impact of reviews, in: Proceedings of the 9th International Conference on Electronic Commerce, 2007, pp. 303–310.

[20] Y. Liu, X. Huang, A. An, X. Yu, Modeling and predicting the helpfulness of online reviews, in: Proceedings of the 8th IEEE International Conference on Data Mining, 2008, pp. 443–452.

[21] Y. Lu, P. Tsaparas, A. Ntoulas, L. Polanyi, Exploiting social context for review quality prediction, in: Proceedings of the 19th International Conference on World Wide Web, 2010, pp. 691–700.

[22] A. Ghose, P. G. Ipeirotis, Estimating the helpfulness and economic impact of product reviews: Mining text and reviewer characteristics, IEEE Transactions on Knowledge and Data Engineering 23 (2011) 1498–1512.

[23] J. Blitzer, M. Dredze, F. Pereira, Biographies, bollywood, boom-boxes and blenders: Domain adaptation for sentiment classification, in: Proceedings of the 45th Annual Meeting of the Association of Computational Linguistics, 2007, pp. 440–447.

[24] K. Liu, J. Zhao, Cross-domain sentiment classification using a two-stage method, in: Proceedings of the 18th ACM Conference on Information and Knowledge Management, 2009, pp. 1717–1720.

[25] S. J. Pan, X. Ni, J.-T. Sun, Q. Yang, Z. Chen, Cross-domain sentiment classification via spectral feature alignment, in: Proceedings of the 19th International Conference on World Wide Web, 2010, pp. 751–760.

[26] R. Mihalcea, C. Banea, J. Wiebe, Learning multilingual subjective language via cross-lingual projections, in: Proceedings of the 45th Annual Meeting of the Association for Computational Linguistics, 2007, pp. 976–983.

[27] C. Banea, R. Mihalcea, J. Wiebe, S. Hassan, Multilingual subjectivity analysis using machine translation, in: Proceedings of the Conference

on Empirical Methods in Natural Language Processing, 2008, pp. 127–135.

[28] X. Wan, Co-training for cross-lingual sentiment classification, in: Proceedings of Joint Conference of the 47th Annual Meeting of the Association for Computational Linguistics and the 4th International Joint Conference on Natural Language Processing of the Asian Federation of Natural Language Processing, 2009, pp. 235–243.

[29] S. Hassan, R. Mihalcea, Cross-lingual semantic relatedness using encyclopedic knowledge, in: Proceedings of the Conference on Empirical Methods in Natural Language Processing, 2009, pp. 1192–1201.

[30] H. Cho, S. Kim, J. Lee, J.-S. Lee, Data-driven integration of multiple sentiment dictionaries for lexicon-based sentiment classification of product reviews, Knowledge-Based Systems 71 (2014) 61–71.

[31] M. Hu, B. Liu, Mining and summarizing customer reviews, in: Proceedings of the 10th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, 2004, pp. 168–177.

[32] S.-M. Kim, E. Hovy, Determining the sentiment of opinions, in: Proceedings of the 20th International Conference on Computational Linguistics, 2004, pp. 1367–1373.

[33] D. Rao, D. Ravichandran, Semi-supervised polarity lexicon induction, in: Proceedings of the 12th Conference of the European Chapter of the Association for Computational Linguistics, 2009, pp. 675–682.

[34] S. Wang, C. D. Manning, Baselines and bigrams: Simple, good sentiment and topic classification, in: Proceedings of the 50th Annual Meeting of the Association for Computational Linguistics, 2012, pp. 90–94.

[35] B. Pang, L. Lee, S. Vaithyanathan, Thumbs up?: Sentiment classification using machine learning techniques, in: Proceedings of the Conference on Empirical Methods in Natural Language Processing, 2002, pp. 79–86.

[36] J. Martineau, T. Finin, Delta TFIDF: An improved feature space for sentiment analysis, in: Proceedings of the 3rd International AAAI Conference on Weblogs and Social Media, 2009, pp. 258–261.

[37] G. Paltoglou, M. Thelwall, A study of information retrieval weighting schemes for sentiment analysis, in: Proceedings of the 48th Annual Meeting of the Association for Computational Linguistics, 2010, pp. 1386–1395.

[38] H. Yu, V. Hatzivassiloglou, Towards answering opinion questions: Separating facts from opinions and identifying the polarity of opinion sentences, in: Proceedings of the Conference on Empirical Methods in Natural Language Processing, 2003, pp. 129–136.

[39] B. Pang, L. Lee, 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, pp. 271–278.

[40] A. Kennedy, D. Inkpen, Sentiment classification of movie reviews using contextual valence shifters, Computational Intelligence 22 (2006) 110–125.

[41] M. Taboada, J. Brooke, M. Tofiloski, K. Voll, M. Stede, Lexicon-based methods for sentiment analysis, Computational Linguistics 37 (2011) 267–307.

[42] E. Cambria, B. White, Jumping nlp curves: A review of natural language processing research [review article], IEEE Computational Intelligence Magazine 9 (2014) 48–57.

[43] J. D. Lafferty, A. McCallum, F. C. N. Pereira, Conditional random fields: Probabilistic models for segmenting and labeling sequence data, in: Proceedings of the 8th International Conference on Machine Learning, 2001, pp. 282–289.

[44] F. Peng, F. Feng, A. McCallum, Chinese segmentation and new word detection using conditional random fields, in: Proceedings of the 20th International Conference on Computational Linguistics, 2004, pp. 562–568.

[45] T. Kudo, K. Yamamoto, Y. Matsumoto, Applying conditional random fields to Japanese morphological analysis, in: Proceedings of the Conference on Empirical Methods in Natural Language Processing, 2004, pp. 230–237.

[46] H. Tseng, A conditional random field word segmenter, in: Proceedings of the 4th SIGHAN Workshop on Chinese Language Processing, 2005, pp. 168–171.

[47] J. Ma, E. Hinrichs, Accurate linear-time Chinese word segmentation via embedding matching, in: Proceedings of the 53rd Annual Meeting of the Association for Computational Linguistics and the 7th International Joint Conference on Natural Language Processing, 2015, pp. 1733–1743.

[48] X. Chen, X. Qiu, C. Zhu, X. Huang, Gated recursive neural network for Chinese word segmentation, in: Proceedings of the 53rd Annual Meeting of the Association for Computational Linguistics and the 7th International Joint Conference on Natural Language Processing, 2015, pp. 1744–1753.

[49] C. Sutton, A. McCallum, An introduction to conditional random fields, Foundations and Trends in Machine Learning 4 (2012) 267–373.

[50] Y. Zhang, S. Clark, Joint word segmentation and pos tagging using a single perceptron, in: Proceedings of the 46th Annual Meeting of the Association for Computational Linguistics: Human Language Technologies, 2008, pp. 888–896.

[51] X. Qian, Y. Liu, Joint Chinese word segmentation, pos tagging and parsing, in: Proceedings of the Joint Conference on Empirical Methods in Natural Language Processing and Computational Natural Language Learning, 2012, pp. 501–511.

[52] X. Zheng, H. Chen, T. Xu, Deep learning for Chinese word segmentation and pos tagging, in: Proceedings of the Conference on Empirical Methods in Natural Language Processing, 2013, pp. 647–657.

[53] S. Bird, E. Klein, E. Loper, Natural Language Processing with Python, O'Reilly Media, 2009.

[54] C. D. Manning, M. Surdeanu, J. Bauer, J. Finkel, S. J. Bethard, D. McClosky, The Stanford CoreNLP natural language processing toolkit, in: Proceedings of the 52nd Annual Meeting of the Association for Computational Linguistics: System Demonstrations, 2014, pp. 55–60.

[55] R. Socher, A. Perelygin, J. Wu, J. Chuang, C. D. Manning, A. Y. Ng, C. Potts, Recursive deep models for semantic compositionality over a sentiment treebank, in: Proceedings of the Conference on Empirical Methods in Natural Language Processing, 2013, pp. 1631–1642.

[56] R. Řehůřek, P. Sojka, Software framework for topic modelling with large corpora, in: Proceedings of the LREC 2010 Workshop on New Challenges for NLP Frameworks, 2010, pp. 45–50.

[57] X. Qiu, Q. Zhang, X. Huang, Fudannlp: A toolkit for Chinese natural language processing, in: Proceedings of the 51st Annual Meeting of the Association for Computational Linguistics, 2013, pp. 49–54.

[58] W. Che, Z. Li, T. Liu, LTP: A Chinese language technology platform, in: Proceedings of the 23rd International Conference on Computational Linguistics: Demonstrations, 2010, pp. 13–16.

[59] J. Zhu, M. Zhu, Q. Wang, T. Xiao, Niuparser: A Chinese syntactic and semantic parsing toolkit, in: Proceedings of the 53rd Annual Meeting of the Association for Computational Linguistics and the 7th International Joint Conference on Natural Language Processing: System Demonstrations, 2015, pp. 145–150.

[60] M. Gamon, Sentiment classification on customer feedback data: Noisy data, large feature vectors, and the role of linguistic analysis, in: Proceedings of the 20th International Conference on Computational Linguistics, 2004, pp. 841–847.

[61] V. Ng, S. Dasgupta, S. M. N. Arifin, Examining the role of linguistic knowledge sources in the automatic identification and classification of reviews, in: Proceedings of the COLING/ACL on Main Conference Poster Sessions, 2006, pp. 611–618.

[62] M. Joshi, C. Penstein-Rosé, Generalizing dependency features for opinion mining, in: Proceedings of the 47th Annual Meeting of the Asso-

ciation for Computational Linguistics and the 4th International Joint Conference on Natural Language Processing, 2009, pp. 313–316.

[63] A.-M. Popescu, O. Etzioni, Extracting product features and opinions from reviews, in: Proceedings of the Conference on Human Language Technology and Empirical Methods in Natural Language Processing, 2005, pp. 339–346.

[64] S. Gindl, A. Weichselbraun, A. Scharl, Rule-based opinion target and aspect extraction to acquire affective knowledge, in: Proceedings of the 22nd International Conference on World Wide Web, 2013, pp. 557–564.

[65] Y. Choi, C. Cardie, Learning with compositional semantics as structural inference for subsentential sentiment analysis, in: Proceedings of the Conference on Empirical Methods in Natural Language Processing, 2008, pp. 793–801.

[66] B. Zou, Q. Zhu, G. Zhou, Negation and speculation identification in Chinese language, in: Proceedings of the 53rd Annual Meeting of the Association for Computational Linguistics and the 7th International Joint Conference on Natural Language Processing, 2015, pp. 656–665.

[67] P. D. Turney, 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, pp. 417–424.

[68] P. D. Turney, Mining the web for synonyms: PMI-IR versus LSA on TOEFL, in: Proceedings of the 12th European Conference on Machine Learning, 2001, pp. 491–502.

[69] S. Dasgupta, V. Ng, Mine the easy, classify the hard: A semi-supervised approach to automatic 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, 2009, pp. 701–709.

[70] S. Li, Z. Wang, G. Zhou, S. Y. M. Lee, Semi-supervised learning for imbalanced sentiment classification, in: Proceedings of the 21nd International Joint Conference on Artificial Intelligence, 2011, pp. 1826–1831.

[71] S. Li, L. Huang, J. Wang, G. Zhou, Semi-stacking for semi-supervised sentiment classification, in: Proceedings of the 53rd Annual Meeting of the Association for Computational Linguistics and the 7th International Joint Conference on Natural Language Processing, 2015, pp. 27–31.

[72] C. Lin, Y. He, Joint sentiment/topic model for sentiment analysis, in: Proceedings of the 18th ACM Conference on Information and Knowledge Management, 2009, pp. 375–384.

[73] F. Li, M. Huang, X. Zhu, Sentiment analysis with global topics and local dependency, in: Proceedings of the 24th AAAI Conference on Artificial Intelligence, 2010, pp. 1371–1376.

[74] S. Somasundaran, G. Namata, J. Wiebe, L. Getoor, Supervised and unsupervised methods in employing discourse relations for improving opinion polarity classification, in: Proceedings of the Conference on Empirical Methods in Natural Language Processing, 2009, pp. 170–179.

[75] R. Trivedi, J. Eisenstein, Discourse connectors for latent subjectivity in sentiment analysis, in: Proceedings of the North American Chapter of the Association for Computational Linguistics, 2013, pp. 808–813.

[76] P. Bhatia, Y. Ji, J. Eisenstein, Better document-level sentiment analysis from RST discourse parsing, in: Proceedings of the Conference on Empirical Methods in Natural Language Processing, 2015, pp. 2212–2218.

[77] O. Täckström, R. McDonald, Semi-supervised latent variable models for sentence-level sentiment analysis, in: Proceedings of the 49th Annual Meeting of the Association for Computational Linguistics: Human Language Technologies, 2011, pp. 569–574.

[78] A. Quattoni, S. Wang, L.-P. Morency, M. Collins, T. Darrell, Hidden conditional random fields, IEEE Transactions on Pattern Analysis and Machine Intelligence 29 (2007) 1848–1852.

[79] B. Yang, C. Cardie, Context-aware learning for sentence-level sentiment analysis with posterior regularization, in: Proceedings of the

52nd Annual Meeting of the Association for Computational Linguistics, 2014, pp. 325–335.

[80] K. Ganchev, J. a. Graça, J. Gillenwater, B. Taskar, Posterior regularization for structured latent variable models, Journal of Machine Learning Research 11 (2010) 2001–2049.

[81] D. Tang, B. Qin, F. Wei, L. Dong, T. Liu, M. Zhou, A joint segmentation and classification framework for sentence level sentiment classification, IEEE/ACM Transactions on Audio, Speech, and Language Processing 23 (2015) 1750–1761.

[82] E. Cambria, B. W. Schuller, B. Liu, H. Wang, C. Havasi, Statistical approaches to concept-level sentiment analysis, IEEE Intelligent Systems 28 (2013) 6–9.

[83] E. Cambria, B. W. Schuller, B. Liu, H. Wang, C. Havasi, Knowledge-based approaches to concept-level sentiment analysis, IEEE Intelligent Systems 28 (2013) 12–14.

[84] R. Agrawal, R. Srikant, Fast algorithms for mining association rules in large databases, in: Proceedings of the 20th International Conference on Very Large Data Bases, 1994, pp. 487–499.

[85] G. Qiu, B. Liu, J. Bu, C. Chen, Expanding domain sentiment lexicon through double propagation, in: Proceedings of the 21st International Jont Conference on Artifical Intelligence, 2009, pp. 1199–1204.

[86] L. Zhang, B. Liu, S. H. Lim, E. O'Brien-Strain, Extracting and ranking product features in opinion documents, in: Proceedings of the 23rd International Conference on Computational Linguistics, 2010, pp. 1462–1470.

[87] Q. Liu, Z. Gao, B. Liu, Y. Zhang, Automated rule selection for aspect extraction in opinion mining, in: Proceedings of the 24th International Joint Conference on Artificial Intelligence, 2015, pp. 1291–1297.

[88] S. Brody, N. Elhadad, An unsupervised aspect-sentiment model for online reviews, in: Human Language Technologies: The 2010 Annual Conference of the North American Chapter of the Association for Computational Linguistics, 2010, pp. 804–812.

[89] Z. Zhai, B. Liu, H. Xu, P. Jia, Clustering product features for opinion mining, in: Proceedings of the 4th ACM International Conference on Web Search and Data Mining, 2011, pp. 347–354.

[90] X. Ding, B. Liu, L. Zhang, Entity discovery and assignment for opinion mining applications, in: Proceedings of the 15th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, 2009, pp. 1125–1134.

[91] Y. Nakayama, A. Fujii, Extracting condition-opinion relations toward fine-grained opinion mining, in: Proceedings of the Conference on Empirical Methods in Natural Language Processing, 2015, pp. 622–631.

[92] T. T. Thet, J.-C. Na, C. S. Khoo, Aspect-based sentiment analysis of movie reviews on discussion boards, Journal of Information Science 36 (2010) 823–848.

[93] A. Lazaridou, I. Titov, C. Sporleder, A Bayesian model for joint unsupervised induction of sentiment, aspect and discourse representations, in: Proceedings of the 51st Annual Meeting of the Association for Computational Linguistics, 2013, pp. 1630–1639.

[94] L. Wang, K. Liu, Z. Cao, J. Zhao, G. de Melo, Sentiment-aspect extraction based on restricted boltzmann machines, in: Proceedings of the 53rd Annual Meeting of the Association for Computational Linguistics and the 7th International Joint Conference on Natural Language Processing, 2015, pp. 616–625.

[95] J. Blitzer, R. Mcdonald, F. Pereira, Domain adaptation with structural correspondence learning, in: Proceedings of the Conference on Empirical Methods in Natural Language Processing, 2006, pp. 120–128.

[96] S. Ben-David, J. Blitzer, K. Crammer, F. Pereira, Analysis of representations for domain adaptation, Advances in Neural Information Processing Systems 19 (2007) 137–144.

[97] T. Hofmann, Probabilistic latent semantic indexing, in: Proceedings of the 22nd Annual International ACM SIGIR Conference on Research and Development in Information Retrieval, 1999, pp. 50–57.

[98] S. Tan, G. Wu, H. Tang, X. Cheng, A novel scheme for domain-transfer problem in the context of sentiment analysis, in: Proceedings of the 16th ACM Conference on Information and Knowledge Management, 2007, pp. 979–982.

[99] Y. He, C. Lin, H. Alani, Automatically extracting polarity-bearing topics for cross-domain sentiment classification, in: Proceedings of the 49th Annual Meeting of the Association for Computational Linguistics: Human Language Technologies, 2011, pp. 123–131.

[100] S. Li, Y. Xue, Z. Wang, G. Zhou, Active learning for cross-domain sentiment classification, in: Proceedings of the 23rd International Joint Conference on Artificial Intelligence, 2013, pp. 2127–2133.

[101] Y. Choi, C. Cardie, Adapting a polarity lexicon using integer linear programming for domain-specific sentiment classification, in: Proceedings of the Conference on Empirical Methods in Natural Language Processing, 2009, pp. 590–598.

[102] D. Bollegala, D. J. Weir, J. A. Carroll, Using multiple sources to construct a sentiment sensitive thesaurus for cross-domain sentiment classification, in: Proceedings of the 49th Annual Meeting of the Association for Computational Linguistics, 2011, pp. 132–141.

[103] S. J. Pan, Q. Yang, A survey on transfer learning, IEEE Transactions on Knowledge and Data Engineering 22 (2010) 1345–1359.

[104] S. Sun, H. Shi, Y. Wu, A survey of multi-source domain adaptation, Information Fusion 24 (2015) 84–92.

[105] K. Duh, A. Fujino, M. Nagata, Is machine translation ripe for cross-lingual sentiment classification?, in: Proceedings of the 49th Annual Meeting of the Association for Computational Linguistics: Human Language Technologies, 2011, pp. 429–433.

[106] X. Wan, Using bilingual knowledge and ensemble techniques for unsupervised Chinese sentiment analysis, in: Proceedings of the Conference on Empirical Methods in Natural Language Processing, 2008, pp. 553–561.

[107] L. Gui, R. Xu, Q. Lu, J. Xu, J. Xu, B. Liu, X. Wang, Cross-lingual opinion analysis via negative transfer detection, in: Proceedings of the 52nd Annual Meeting of the Association for Computational Linguistics, 2014, pp. 860–865.

[108] B. Wei, C. Pal, Cross lingual adaptation: An experiment on sentiment classifications, in: Proceedings of the 48th Annual Meeting of the Association for Computational Linguistics, 2010, pp. 258–262.

[109] Q. Chen, W. Li, Y. Lei, X. Liu, Y. He, Learning to adapt credible knowledge in cross-lingual sentiment analysis, in: Proceedings of the 53rd Annual Meeting of the Association for Computational Linguistics and the 7th International Joint Conference on Natural Language Processing, 2015, pp. 419–429.

[110] Q. Le, T. Mikolov, Distributed representations of sentences and documents, in: Proceedings of the 31st International Conference on Machine Learning, 2014, pp. 1188–1196.

[111] X. Meng, F. Wei, X. Liu, M. Zhou, G. Xu, Cross-lingual mixture model for sentiment classification, in: Proceedings of the 50th Annual Meeting of the Association for Computational Linguistics, 2012, pp. 572–581.

[112] P. Lambert, Aspect-level cross-lingual sentiment classification with constrained SMT, in: Proceedings of the 53rd Annual Meeting of the Association for Computational Linguistics and the 7th International Joint Conference on Natural Language Processing, 2015, pp. 781–787.

[113] P. Koehn, H. Hoang, A. Birch, C. Callison-Burch, M. Federico, N. Bertoldi, B. Cowan, W. Shen, C. Moran, R. Zens, C. Dyer, O. Bojar, A. Constantin, E. Herbst, Moses: Open source toolkit for statistical machine translation, in: Proceedings of the 45th Annual Meeting of the Association for Computational Linguistics on Interactive Poster and Demonstration Sessions, 2007, pp. 177–180.

[114] D. Gao, F. Wei, W. Li, X. Liu, M. Zhou, Cross-lingual sentiment lexicon learning with bilingual word graph label propagation, Computational Linguistics 41 (2015) 21–40.

[115] J. Wiebe, T. Wilson, C. Cardie, Annotating expressions of opinions and emotions in language, in: Proceedings of the 2nd International Conference on Language Resources and Evaluation, 2005, pp. 164–210.

[116] A. Pak, P. Paroubek, Twitter as a corpus for sentiment analysis and opinion mining, in: Proceedings of the 7th International Conference on Language Resources and Evaluation, 2010, pp. 1320–1326.

[117] T. Wilson, J. Wiebe, P. Hoffmann, Recognizing contextual polarity in phrase-level sentiment analysis, in: Proceedings of the Conference on Human Language Technology and Empirical Methods in Natural Language Processing, 2005, pp. 347–354.

[118] S. Baccianella, A. Esuli, F. Sebastiani, Sentiwordnet 3.0: An enhanced lexical resource for sentiment analysis and opinion mining, in: Proceedings of the 7th International Conference on Language Resources and Evaluation, 2010, pp. 2200–2204.

[119] P. J. Stone, D. C. Dunphy, M. S. Smith, D. M. Ogilvie, The General Inquirer: A Computer Approach to Content Analysis, The MIT Press, 1966.

[120] J. W. Pennebaker, R. L. Boyd, K. Jordan, K. Blackburn, The development and psychometric properties of LIWC2015, Austin, TX: University of Texas at Austin.

[121] W.-T. Chen, S.-C. Lin, S.-L. Huang, Y.-S. Chung, K.-J. Chen, E-hownet and automatic construction of a lexical ontology, in: Proceedings of the 23rd International Conference on Computational Linguistics: Demonstrations, 2010, pp. 45–48.

[122] L.-W. Ku, H.-H. Chen, Mining opinions from the web: Beyond relevance retrieval, Journal of the American Society for Information Science and Technology 58 (2007) 1838–1850.

[123] L. Velikovich, S. Blair-goldensohn, K. Hannan, R. Mcdonald, The viability of web-derived polarity lexicons, in: Human Language Technologies: The Annual Conference of the North American Chapter of the Association for Computational Linguistics, 2010, pp. 777–785.

[124] A. Hassan, D. Radev, Identifying text polarity using random walks, in: Proceedings of the 48th Annual Meeting of the Association for Computational Linguistics, 2010, pp. 395–403.

[125] D. Tang, F. Wei, B. Qin, M. Zhou, T. Liu, Building large-scale Twitter-specific sentiment lexicon: A representation learning approach, in: Proceedings of the 25th International Conference on Computational Linguistics, 2014, pp. 172–182.

[126] N. Jindal, B. Liu, Mining comparative sentences and relations, in: Proceedings of the 21st National Conference on Artificial Intelligence, 2006, pp. 1331–1336.

[127] N. Jindal, B. Liu, Identifying comparative sentences in text documents, in: Proceedings of the 29th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval, 2006, pp. 244–251.

[128] M. Ganapathibhotla, B. Liu, Mining opinions in comparative sentences, in: Proceedings of the 22nd International Conference on Computational Linguistics, 2008, pp. 241–248.

[129] Y. Choi, C. Cardie, E. Riloff, S. Patwardhan, Identifying sources of opinions with conditional random fields and extraction patterns, in: Proceedings of the Conference on Human Language Technology and Empirical Methods in Natural Language Processing, 2005, pp. 355–362.

[130] E. Breck, Y. Choi, C. Cardie, Identifying expressions of opinion in context, in: Proceedings of the 20th International Joint Conference on Artificial Intelligence, 2007, pp. 2683–2688.

[131] O. Irsoy, C. Cardie, Opinion mining with deep recurrent neural networks, in: Proceedings of the Conference on Empirical Methods in Natural Language Processing, 2014, pp. 720–728.

[132] R. Socher, J. Bauer, C. D. Manning, A. Y. Ng, Parsing with compositional vector grammars., in: Proceedings of the 51st Annual Meeting on Association for Computational Linguistics, 2013, pp. 455–465.

[133] R. Socher, C. C. Lin, C. Manning, A. Y. Ng, Parsing natural scenes and natural language with recursive neural networks, in: Proceedings of the 28th International Conference on Machine Learning, 2011, pp. 129–136.

[134] R. S. Romain Paulus, C. D. Manning, Global belief recursive neural networks, Advances in Neural Information Processing Systems 4 (2014) 2888–2896.

[135] T. Mikolov, M. Karafiát, L. Burget, J. Cernockỳ, S. Khudanpur, Recurrent neural network based language model, in: Proceedings of the 11th Annual Conference of the International Speech Communication Association, 2010, pp. 1045–1048.

[136] T. Mikolov, S. Kombrink, L. Burget, J. Černockỳ, S. Khudanpur, Extensions of recurrent neural network language model, in: Proceedings of the 2011 IEEE International Conference on Acoustics, Speech and Signal Processing, 2011, pp. 5528–5531.

[137] S. Hochreiter, J. Schmidhuber, Long short-term memory, Neural computation 9 (1997) 1735–1780.

[138] K. S. Tai, R. Socher, C. D. Manning, Improved semantic representations from tree-structured long short-term memory networks, in: Proceedings of the 53rd Annual Meeting of the Association for Computational Linguistics and the 7th International Joint Conference on Natural Language Processing of the Asian Federation of Natural Language Processing, 2015, pp. 1556–1566.

[139] P. Liu, S. Joty, H. Meng, Fine-grained opinion mining with recurrent neural networks and word embeddings, in: Proceedings of the Conference on Empirical Methods in Natural Language Processing, 2015, pp. 1433–1443.

[140] N. Kalchbrenner, E. Grefenstette, P. Blunsom, A convolutional neural network for modelling sentences, in: Proceedings of the 52nd Annual Meeting of the Association for Computational Linguistics, 2014, pp. 655–665.

[141] Y. Bengio, R. Ducharme, P. Vincent, C. Jauvin, A neural probabilistic language model, Journal of Machine Learning Research 3 (2003) 1137–1155.

[142] T. Mikolov, K. Chen, G. Corrado, J. Dean, Efficient estimation of word representations in vector space, CoRR abs/1301.3781 (2013) 1–12.

[143] T. Mikolov, I. Sutskever, K. Chen, G. S. Corrado, J. Dean, Distributed representations of words and phrases and their compositionality, Advances in Neural Information Processing Systems 26 (2013) 3111–3119.

[144] J. Pennington, R. Socher, C. D. Manning, Glove: Global vectors for word representation, in: Proceedings of the 2014 Empirical Methods in Natural Language Processing, 2014, pp. 1532–1543.

[145] M. Hu, B. Liu, Opinion extraction and summarization on the web, in: Proceedings, The 21st National Conference on Artificial Intelligence and the 8th Innovative Applications of Artificial Intelligence Conference, 2006, pp. 1621–1624.

[146] G. Carenini, R. Ng, A. Pauls, Multi-document summarization of evaluative text, in: Proceedings of the 11th Conference of the European Chapter of the Association for Computational Linguistics, 2006, pp. 305–312.

[147] D. R. Radev, S. Teufel, H. Saggion, W. Lam, J. Blitzer, H. Qi, A. Çelebi, D. Liu, E. Drabek, Evaluation challenges in large-scale document summarization, in: Proceedings of the 41st Annual Meeting on Association for Computational Linguistics, 2003, pp. 375–382.

[148] G. Carenini, J. D. Moore, Generating and evaluating evaluative arguments, Artificial Intelligence 170 (2006) 925–952.

[149] K. Lerman, S. Blair-Goldensohn, R. Mcdonald, Sentiment summarization: Evaluating and learning user preferences, in: Proceedings of the 12th Conference of the European Chapter of the Association for Computational Linguistics, 2009, pp. 514–522.

[150] H. Nishikawa, T. Hasegawa, Y. Matsuo, G. Kikui, Opinion summarization with integer linear programming formulation for sentence extraction and ordering, in: Proceedings of the 23rd International Conference on Computational Linguistics, 2010, pp. 910–918.

[151] S. Gerani, Y. Mehdad, G. Carenini, R. T. Ng, B. Nejat, Abstractive summarization of product reviews using discourse structure, in: Proceedings of the Conference on Empirical Methods in Natural Language Processing, 2014, pp. 1602–1613.

[152] N. Jindal, B. Liu, Opinion spam and analysis, in: Proceedings of the 3rd International Conference on Web Search and Data Mining, 2008, pp. 219–230.

[153] S.-M. Kim, P. Pantel, T. Chklovski, M. Pennacchiotti, Automatically assessing review helpfulness, in: Proceedings of the Conference on Empirical Methods in Natural Language Processing, 2006, pp. 423–430.

[154] Z. Zhang, B. Varadarajan, Utility scoring of product reviews, in: Proceedings of the 15th ACM International Conference on Information and Knowledge Management, 2006, pp. 51–57.

[155] S. Moghaddam, M. Jamali, M. Ester, ETF: Extended tensor factorization model for personalizing prediction of review helpfulness, in: Proceedings of the 5th ACM International Conference on Web Search and Data Mining, 2012, pp. 163–172.