Received February 11, 2019, accepted March 11, 2019, date of publication March 21, 2019, date of current version April 11, 2019.

Digital Object Identifier 10.1109/ACCESS.2019.2906754

A Survey on Opinion Mining: From Stance to Product Aspect

RUI WANG, DEYU ZHOU, MINGMIN JIANG, JIASHENG SI, AND YANG YANG
School of Computer Science and Engineering, Key Laboratory of Computer Network and Information Integration, Ministry of Education, Southeast University, Nanjing 210096, China

Corresponding author: Deyu Zhou (d.zhou@seu.edu.cn)

This work was supported by the National Key Research and Development Program of China under Grant 2016YFC1306704, in part by the National Natural Science Foundation of China under Grant 61772132, and in part by the Natural Science Foundation of Jiangsu Province of China under Grant BK20161430.

ABSTRACT With the prevalence of social media and online forum, opinion mining, aiming at analyzing and discovering the latent opinion in user-generated reviews on the Internet, has become a hot research topic. This survey focuses on two important subtasks in this field, stance detection and product aspect mining, both of which can be formalized as the problem of the triple $\langle target, aspect, opinion\rangle$ extraction. In this paper, we first introduce the general framework of opinion mining and describe the evaluation metrics. Then, the methodologies for stance detection on different sources, such as online forum and social media are discussed. After that, approaches for product aspect mining are categorized into three main groups which are corpus level aspect extraction, corpus level aspect, and opinion mining, and document level aspect and opinion mining based on the processing units and tasks. And then we discuss the challenges and possible solutions. Finally, we summarize the evolving trend of the reviewed methodologies and conclude the survey.

INDEX TERMS Opinion mining, stance detection, product aspect mining, topic model, deep neural network.

I. INTRODUCTION

With the pervasiveness of online discussion forum and social media platform, user generated text containing opinions on some hot issues has increased significantly. Due to the large amount of such emotional reviews and posts on Internet, it is impossible for users to digest such information manually. Therefore, automatically mining opinion from online texts, aiming at discovering user concerned topics and the corresponding opinion, becomes essential. In general, opinion mining aims to extract a quintuple $< e, a, s, h, t >$ [1] from texts, where e is the entity or the target, a is the aspect of the entity e, h is the opinion holder, t is the time when the opinion holder expresses her opinion on the entity e, and s is the opinion which h holds to the aspect a of the entity e at t. For example, opinion mining processes the review text "*I bought a new iPhoneX today, the screen is great, but the voice quality is poor*" and outputs two quintuples < iPhoneX, screen, great, I, today> and < iPhoneX, voice quality, poor, I, today>. However, not all the opinion mining tasks need to extract all the five elements in quintuple. For example, sentiment analysis cares more about the sentiment polarity s of the text, stance detection aims to identify the opinion s to the specific target e, and product aspect mining focus on extracting the aspect a and corresponding opinion s from text.

In the last decade, stance detection [2] and product aspect mining [3] have attracted many scholars. Following the general opinion mining framework, stance detection can be formalized as the task of extracting tuple $< e, s >$ (e means target and s represents opinion) without considering other elements. Stance detection focuses on detecting the user stance (favor, against) on a particular debate topic or hot-debated event. It is similar to sentiment analysis [4], [5], but with big difference. In specific, sentiment analysis aims to identify the sentiment polarity (positive, negative) of the text while stance detection cares about the stance on the target. For example, the tweet "*Jeb bush is the only sane candidate in this republican lineup, I support him*" will be assigned positive by sentiment analysis [140], [143], but extracted with '*against*' stance to the topic "*Donald Trump as President*" by stance detection. Research on stance detection can be categorized into four groups based on debate settings, such as congressional floor debates [6]–[9], company-internal discussions [10], [11],

The associate editor coordinating the review of this manuscript and approving it for publication was Alexandros Iosifidis.

online forums ideological debates [12]–[27] and hot-event oriented debates on social media [28]–[50]. The latter two are open domain and flexible, therefore more challengeable. For example, the debate forum such as convinceMe.net has a wide range of debate topics from the playful (e.g. '*Cats vs. Dogs*', '*Mac vs. PC*', '*Superman vs. Batman*' and etc) to the ideological (e.g. '*Death penalty*', '*Exist God*', '*Gay marriage*', '*Healthcare*' and etc). Furthermore, the participants prefer to use colorful and emotional language to express their viewpoints, such as the tweet "*It looks like they like Hilary more... and that plain stupid*" related to the hashtag "*#2016US election#*". Therefore, we concentrate on the prevalent work of stance detection for online debate forums and social media in this survey.

Different from stance detection, product aspect mining aims at detecting relevant aspects and opinions. Following the general opinion mining framework, product aspect mining can be formalized as the task of extracting triple $< e, a, s >$ (e means target, a and s represent aspect and opinion respectively). Based on text granularity, it could be categorized into corpus level and document/sentence level mining. The corpus level mining could be further divided into two categories: aspect extraction, aspect and opinion mining. Corpus level aspect extraction aims to mine the aspect terms or aspect phrases in the corpus while ignoring where the aspects are discussed. Similarly, corpus level aspect and opinion mining extract both the aspects and the corresponding opinions without considering where they are expressed. Actually, corpus level mining pays attention to the aspects that most reviews are interested in and the corresponding opinions while document/sentence level mining focus on extracting aspect and opinion terms in a single review. After the first attempt [51] of corpus level extraction, numerous approaches have been proposed which could be categorized as rule-based and unsupervised learning based models. Early aspect mining systems employ frequency pattern mining technique [52]–[60] to extract aspect terms in the reviews. To overcome the disadvantages of missing low frequency aspects and ignoring semantic similarity of aspects, unsupervised learning based approaches [61]–[71] are proposed by casting the task into a clustering problem. Different from corpus level mining, document/sentence level aspect and opinion mining concentrates on detecting the aspects and opinions in each individual document or sentence. It can be viewed as a sequential tagging problem. Hidden Markov Model (HMM) [72], Condition Random Fields (CRFs) [73]–[78] and deep neural networks, such as Convolutional Neural Network (CNN) [79]–[82], Recurrent Neural Network (RNN) [83]–[86] and Long-Short Term Memory (LSTM) [87]–[90] based approaches have been proposed to tackle the problem.

Both stance detection and product aspect mining might benefit various downstream applications, such as public mood prediction regarding political movement, market intelligence [91] and movie sales prediction [92]. Also, they could benefit latent customers by providing smart purchase decision, manufactures by providing the measurement of customer satisfaction [93] to adjust their manufacturing process and sales strategy, and governmental organizations by informing public opinions of a political election [94].

Several surveys on opinion mining have been published. Pang [95] gives a good survey and introduction to the field of sentiment analysis. However, the coverage of the survey published in [95] is restricted mostly to document-level machine learning approaches. Likewise, Tang [96] presents a shorter survey mainly focusing on document-level machine learning approach as well. Besides, Liu [4] gives a survey, with an updated overview of the entire field of sentiment analysis.

In this paper, we focus on the methodologies for two subtasks of opinion mining which are online stance detection and product aspect mining. Thus, we provide a brief survey and categorization on the previous methodologies. Earlier work mainly relies on frequency, relation rules and feature engineering. Later, computational linguistic model and machine learning techniques are explored for the two tasks. With the popularity of deep learning in last five years, researchers have paid attention to employ representation learning and neural-based models [79], [97]–[99]. The remains of the paper will be organized as follows. Section 2 represents the evaluation metrics and the available datasets. Section 3 describes the methods applied in stance detection. A survey on product aspect mining will be presented in Section 4. Challenges and possible solutions are described in Section 5. In Section 6, we will conclude the survey. And the organization of the survey is shown in Fig. 1

II. EVALUATION METRICS AND AVAILABLE DATASETS

To evaluate the performance of opinion mining system, different evaluation metrics are employed.

For stance detection, Accuracy (Acc), Precision (Pre), Recall (Rec) and F_{score}, defined in Eq. 1-4, are used for evaluation [100]. Here, TP is the number of posts which support the debate and are predicted as favor, FP represents the number of posts which are against the debate and are predicted as favor. Similarly, TN is the number of posts which are against the debate and are predicted as against. FN is the number of posts which support the debate and are predicted as against.

$$Accuracy = \frac{TP + TN}{TP + FP + TN + FN} \quad (1)$$

$$Precision = \frac{TP}{TP + FP} \quad (2)$$

$$Recall = \frac{TP}{TP + FN} \quad (3)$$

$$F_{score} = \frac{2 \times Precision \times Recall}{Precision + Recall} \quad (4)$$

Precision, Recall and F_{score} could also be employed to evaluate the performance of product aspect mining (e.g. aspect extraction) and the calculation of precision and recall refers

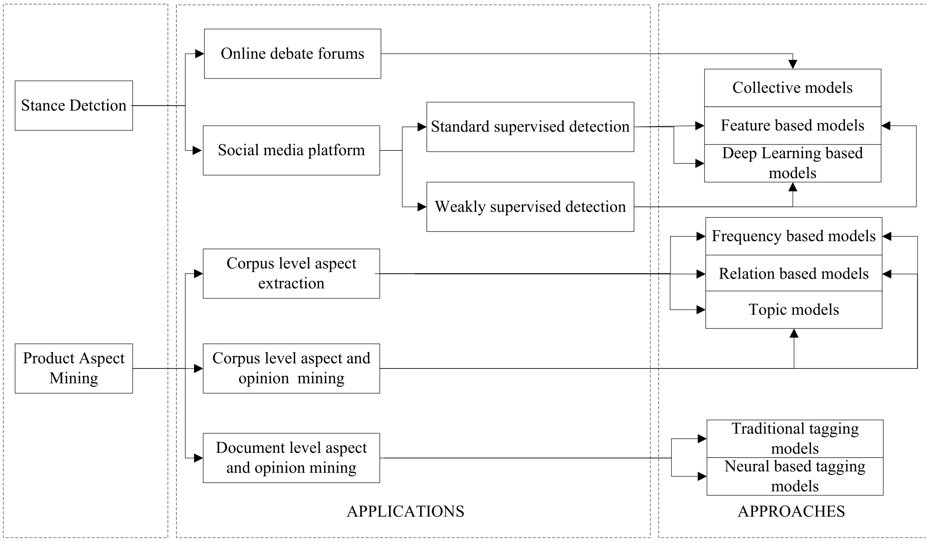


FIGURE 1. **Organization of the survey.**

to Eq. 5 and Eq. 6.

$$Precision = \frac{|extracted\ aspects \cap\ gold\ aspects|}{|extracted\ aspects|} \tag{5}$$

$$Recall = \frac{|extracted\ aspects \cap\ gold\ aspects|}{|gold\ aspects|} \tag{6}$$

Moreover, some other measures are also used in product aspect mining, such as Macro-averaged Mean Absolute Error (MAE^M), Ranking Loss, RandIndex, Precision@n and Kendall's tau coefficient. In these evaluation metrics, opinion labels are assumed to be an integer variable.

Macro-averaged Mean Absolute Error (MAE^M) [101], calculated by Eq. 7, is suitable for tackling highly imbalanced dataset.

$$MAE^M(\mathbf{y}, \hat{\mathbf{y}}) = \frac{1}{k}\sum_{j=1}^{k}\frac{1}{|\mathbf{y}_j|}\sum_{y_i \in \mathbf{y}_j}|y_i - \hat{y}_i| \tag{7}$$

where k is the number of opinion/sentiment label (e.g. $k = 2$ for binary classification), y is the gold label vector, $\hat{y}$ is the predicted label vector and y_j is the subset of review corpus constituted by the reviews whose gold label is j.

Mean Square Error (MSE), a widely used measure in regression problem, is employed to evaluate the performance of opinion/sentiment classification [102]. It is defined as Eq. 8, where n is the number of reviews in corpus, y_i and $\hat{y}_i$ are the gold label and predicted label of the $i-$th review individually.

$$MSE = \frac{1}{n}\sum_{i=1}^{n}(\hat{y}_i - y_i)^2 \tag{8}$$

Similar to MAE^M, Ranking Loss could measure the average distance between gold aspect rating and predicted rating. For a k-level rating problem, the average deviation between y and $\hat{y}$ can be calculated by Eq. 9 [103].

$$Ranking\ loss = \sum_{i=1}^{n}\frac{|y_i - \hat{y}_i|}{k \times n} \tag{9}$$

RandIndex, a measure generally employed in soft clustering algorithm [117], is used to evaluate the aspects detected by topic model based approaches [118]. It is defined as Eq. 10.

$$RandIndex(C_{model}, C_{manual}) = \frac{2(x+y)}{k \times (k-1)} \tag{10}$$

where C_{model} and C_{manual} are clusters which produced by the model and manual annotation, k is the number of aspects to be detected. And the agreement of the clusters generated by the model and annotation could be checked on $k \times (k-1)$ pairs. x is the number of pairs assigned to the same cluster in both partitions, and similarly, y is the number of pair assigned to different clusters.

Precision@n (Pre@n), a common used metric in information retrieval, is used in [120] to evaluate the ability of detecting aspect based opinion. It is defined as Eq.11. And k is the number of gold standard opinion words which appear in top n word set of an opinion topic.

$$Precision@n = \frac{k}{n} \tag{11}$$

Meanwhile, some hypothesis testing techniques are applied to evaluate the performance of product based aspect mining. Kendalls tau coefficient, $\tau_k = (|C| - |D|)/|T|$, is used in [119] to evaluate the quality of detected opinions. Where T denotes the ordered pairs in the gold standard, C and D represents the set of concordant pairs and discordant pairs respectively. It aims to show the percentage of pairs of ranked

TABLE 1. **Online available lexicons, corpus resources.**

	Description	Url
External Lexicons Tools:		
Bing Liu's Opinion Lexicon [51]	English. The latest version of this lexicon includes 4.7K negative words and 2K positive ones.	http://www.cs.uic.edu/˜liub/FBS/sentiment-analysis.html
Harvard General Inquirer [104]	English. It contains 182 categories including positive and negative indicators. 1.9K positive words and 2.3K negative words are marked.	http://www.wjh.harvard.edu/˜inquirer/
WordNet [105]	English. A large semantic lexical database where nouns, verbs, adjectives and adverbs are grouped into sets of cognitive synonyms or synsets.	http://wordnet.princeton.edu/
SentiWordNet [106]	English. It associates words to numerical scores ranging in [0.0, 1.0] which indicate the positivity, negativity an neutrality.	http://sentiwordnet.isti.cnr.it/
MPQA Subjectivity lexicon [107]	English. It includes 8.2K words with their subjective words, POS tags and polarities.	http://mpqa.cs.pitt.edu/lexicons/subj_lexicon/
NRC-emotion lexicon [108]	English. It, also called EmoLex, contains a list of words associate with eight basic emotions (anger, fear, anticipation, trust, surprise, sadness, joy and disgust).	http://saifmohammad.com/WebPages/NRC-Emotion-Lexicon.htm
Linguistic Inquiry and Word Counts (LIWC) [109]	English. It provides a lot of categorized regular expressions including some sentiment related categories.	http://liwc.wpengine.com
HowNet [110]	Chinese & English. It provides a vocabulary for sentiment analysis, including 8.9K Chinese items and 8.9K English items.	http://www.keenage.com/html/e_index.html
NTUSD [111]	Chinese. It provides 2.8K positive words and 8.2K negative word in Chinese.	http://academiasinicanlplab.github.io/
Available datasets:		
LARA data [112]	English. It contains Amazon product reviews under six categories: camera, mobile phone, TV, laptop, tablet and video surveillance system. Each review contains the attributes such as: reviewID (unique), author, title, content, overall rating, date.	http://sifaka.cs.uiuc.edu/ wang296/Data/
ASUM data [113]	English. It contains two sets of reviews. One dataset, named ELECTRONICS, is a collection electronic device reviews from Amazon; and the other, named RESTAURANT-S, is restaurant reviews from Yelp.	http://uilab.kaist.ac.kr/research/WSDM11
TripAdvisor	Spanish. It contains 18K customer reviews on hotels and restaurants from Hopinion.	http://clic.ub.edu/corpus/es/node/106
Multi-Domain dataset [4]	English. It contains multi-domain datasets (like customer reviews, debate dataset, etc.) for many opinion mining tasks, such as review opinion summarization and opinion detection.	http://www2.cs.uic.edu/liub/FBS/sentiment-analysis.html
SCL data [114]	English. It contains Amazon reviews on 4 domains (books, DVDs, electronics kitchen appliances) for opinion mining.	http://www.cs.jhu.edu/mdredze/datasets/sentiment/
SE-ABSA14 [115]	English. It contains two subsets, which consist of 3.8K restaurants and 3.8K laptop reviews, for aspect based opinion mining tasks.	http://metashare.ilsp.gr:8080/.
Stance detection data in twitter [116]	English. The annotated (stance and sentiment) collection is created for the task of stance detection in twitter. It contains six subsets of tweets for different hot topic in social media.	http://www.saifmohammad.com/WebPages/StanceDataset.htm

items that agree or disagree with the ordering in the gold standard.

When doing stance detection and product aspect mining, some systems and models use external tools. As no task-specific lexicons (e.g. stance lexicons) have been constructed, most researchers employ the sentiment lexicon and the semantic lexicons to improve the performance. We summarize the general resource, such as lexicons, dataset which are widely used for stance detection and product aspect mining in Table 1. First eight rows in the table are sentiment or semantic lexicons, from the 9-th row to the 15-th rows list seven online dataset for stance detection and aspect mining.

III. STANCE DETECTION

Stance detection aims at recognizing the holistic subjective disposition (favor, against) that the author/speaker holds by analyzing the author generated reviews or arguments. As a specific type of opinion mining, it has been studied extensively. Thus, we provide a brief summary of stance detection methodologies for debates in online forums and social media in this section.

TABLE 2. Approaches for online debate stance detection, N/A means that the corresponding paper doesn't provide the result.

	Data source	Performance
Textual Content based approaches:		
Somasundaran S and WiebeJ [22]	Debate forum, Createdebate.com	Acc:60.55%-70.59%
Anand P and Walker M [12], [13]	Debate forum, Convinceme.net	Acc:54.00%-69.00%
Hasan K S and Ng V [15]	Debate forum, Createdebate.com	Acc:70.60%-75.90%
Elfardy H and Diab M [18]	Ideological debate	F_{score} :58.10%-71.6%
T Kyaw, SS Aung [23]	Debate forum, Createdebate.com	Acc:55.95%-72.50%
S Ghosh, K Anand [24]	Debate forum, Createdebate.com	N/A
Collective models:		
Walker M and Anand P [14]	Debate forum, Convinceme.net	F_{score} :18.00%-82.00%
Hasan K S and Ng V [16]	Debate forum, Createdebate.com	Acc:71.10%-75.70%
Sridhar D and Getoor L [17]	Debate forum, 4forums.com	F_{score} : 74.00%
Sridhar D and Foulds J R [19]	Debate forums, 4forums.com and Createdebate.com	Acc:63.50%-80.50%
Qiu M and Sim Y [20]	Debate forums, Createdebate.com	Acc:71.20%
Li and Porco [25]	CreateDebate.com,4Forums.com	Acc:75.20%-81.10%
Trabelsi A and Zaiane OR [26]	CreateDebate.com, 4forums.com	Acc:67.80%-92.00%
Dong R and Sun Y [27]	Debate forums, 4forums.com	Acc:64.70%-75.60%

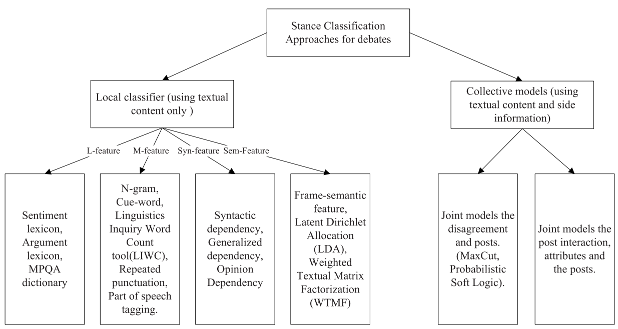


FIGURE 2. Taxonomy for forum debate stance detection approaches using the main characteristics of the published work.

A. STANCE DETECTION IN ONLINE DEBATES FORUMS

Identifying user stance (favor, against) in online debate forums attracts researchers in last years. In general, approaches for the task could be categorized as textual content based approaches and collective models. Detailed categorization could be described as Fig. 2, where L-feature, M-feature, Syn-feature and Sem-feature stand for lexicon-based feature, morphologic-based feature, syntactic-based feature and semantic-based feature respectively.

1) PROBLEM SETTINGS

Online debate forums, such as convinceme.com, createdebate.com and debatepedia.com, have many dual-side debate topics ranging from playful (e.g. '*Cat vs. Dogs*', '*Mac vs. PC*' and '*Superman vs. Batman*') to ideological (e.g. '*Death penalty*', '*Exist God*' and '*Gay marriage*'). And stance detection in online debate forums could be formalized as: Given a debate specific corpus C_t^d and the corresponding debate topic t_d, it aims to determine the stance $s_i^{t_d}$ of each post $p_i \in C_t^d$ in corpus. We use the debate topic '*death penalty*' and two different stance posts as an example, shown in Table 3.

TABLE 3. An example of stance detection in online debate forum.

target	stance	posts
death penalty	favor	"I value human life so much that if someone takes one than this should be taken. Also if someone is thinking about taking a life they are less likely to do so knowing that they might lose theirs"
death penalty	against	"Death penalty is only a costlier version of a lifetime prison sentence, bearing the exception that it offers euthanasia to criminals longing for an easy escape, as opposed to a real punishment".

Textual content based approaches regard the task as a typical classification problem and only use the textual information of the posts such as sentiment lexicons and syntactic patterns to capture the stance information. Another group of researchers claim that textual content could not provide sufficient information for detecting the stance precisely. Thus, collective models which also employ the relationship between posts (e.g. disagreement, argument) or users are proposed. And the related approaches, together with their reported performance can be found in Table 2.

2) TEXTUAL CONTENT BASED APPROACHES

Textual content based approaches view stance detection as a binary classification problem. And the key point of these models is feature engineering. As shown in Fig. 2, these features could be further categorized as lexicon-based feature, morphologic-based feature, syntactic feature and semantic feature. And $1st - 6th$ rows in Table 2 are six representative works.

To delve into the mechanism of these approaches, [12] is selected to illustrate the process of detecting the stance by using textual content only. To capture the stance related information from the post content, Anand and Walker [12] design a feature set containing n-grams, repeated punctuations, cue words, LIWC [109] and three variant syntactic dependency based features. The n-grams feature isc constituted by uni-grams, bi-grams and the basic counts of the post, such as post length. The statistical measures like word per sentence (WPS), pronominal forms (Pro), positive emotion words (PosE) and negative emotion words (NegE) are employed as LIWC feature. And the repeated punctuations are also considered as punctuation feature in [12] because Anand assumes that the repeated punctuations have a special meaning. Besides, three dependency features (Dependency feature, Generalized dependency feature and opinion dependency feature) are employed to capture the syntactic relations between terms in the posts. In detail, the Stanford parser [121] generates a set of grammatical relations represented as (rel_i, w_j, w_k) to capture the relations between words. The generalized dependency features are constructed by replacing the head term with its part-of-speech tag in the output of Stanford parser. Similarly, opinion dependency features are created by replacing the sentiment word with the corresponding polarity label (e.g. '$+$' or '$-$'). To make full use of rebuttal links and improve the detection performance, the rebuttal posts is combined with its parent post in [12]. And the proposed approach achieves a competitive accuracy (54%-69%).

Although, the work in [12] can achieve better performance compared with uni-grams baseline. However, it could not capture any semantic information from post content. Thus, Hasan adds both linguistic extension based on the semantic frame patterns [122] and extra-linguistic extension [15] into the approach. Similarly, Elfardy [18] uses the latent dirichlet allocation (LDA) and the weighted textual matrix factorization(WTMF) to extract the topic distribution of the posts as the semantic representation, and the two distributions are combined using word sense disambiguation (WSD) [123] strategy to achieve a better performance. Somasundaran and Wiebe [22] construct an arguing lexicon and obtain an argument feature which helps with task. Likewise, Kyaw [23] combines tf-idf weights with POS-tags to build the text representation for stance detection. Ghosh and Anand [24] propose a two stage method which firstly detect the argumentative posts from review corpus and then detect the stance for argumentative ones.

3) COLLECTIVE MODELS

Unlike the works discussed in previous subsection, collective models employ both the content and the auxiliary information jointly. These models share a same assumption that two posts with negative relationship may express the different stance and vice versa. Eight collective models for stance detection on debate forums are presented in $7th - 14th$ rows of Table 2.

To explain how to improve the performance of debate stance detection using the information about dialogic relations between posts, [14] is selected. In [14], Walker proposes a graph (V, E)-based approach, each node v in the graph represents a post and edge e indicates the relation, either agreement or disagreement, constructed by the relationship between the posts/authors. Two assumptions are made that all posts written by the same author share the same stance to a specific debate and the rebuttal links in the forum indicate the disagreement relation between corresponding posts. And the proposed approach obtains F_{score} 82%. However, it could not work well in some debate topics (such as Mac vs. PC) and get F_{score} :18% which may because that MaxCut algorithm actually divides the posts into clusters, but then assigns them to the wrong stance.

Sridhar and Getoor [17] propose a probabilistic soft logic (PSL) [124] based model to model the structural and linguistic feature of the posts collectively. By using the PSL rules and the specific corpus (agreement and disagreement relations between context posts has been manual annotated in the collection), the proposed approach infers the probabilistic relationship between posts and the stance labels with good performance, and other relational information could be incorporate into the framework easily. Following this way, to avoid annotating the degree of disagreement beforehand, an improved hinge-loss Markov Random Field [19] is proposed by using a disagreement classifier for determining the agreement polarity between posts. Besides, motivated by the observation that if a post in a post sequence is a reply to its parent, its stance should be depend on that of its parent, Hasan [16] views the task as a sequential labeling problem and devises a HMM-based approach which gets an accuracy 57.5%.

However, the previously discussed collective models do not perform well when facing the particular debate with a low participating rate. Inspired by the collaborative filtering, Qiu and Sim [20] integrate the auxiliary information (textual content, user interactions and user attributes provided in biographical information) into the regression-based latent factor model and employ the binomial matrix factorization for stance modeling, the experiment results prove that interaction and user preference provide rich information for estimating the stance of cold-start user. To further consider the relationship between different debate topics, Li and Porco [25] view stance detection as a representation learning task. Thus, they embed the text content and user interactions into a same space and obtain a more informative representation. Trabelsi and Zaïane [26] devise an unsupervised model based on the assumption that users with different viewpoint are prone to communicate frequently, and the proposed method improves the performance in both user level and post level stance detection.

B. STANCE DETECTION ON SOCIAL MEDIA

Stance detection on social media (e.g. Twitter, Facebook, Chinese micro-blog and etc) could be viewed as a classification task. However, it is often oriented by hot-event, such as '*USA election*' or '*Brexit*', and has its own characters. Firstly, its content contains noisy information (e.g. spelling error, grammar error and abbreviations). The texts on these platforms are often not well written because of the limitation of length and the instantaneity. For example, "*Seems FBI had evidence of Trump-Russia contacts during the campaign when Mosco was attacking US election.*", "*This is the SOB*

TABLE 4. Methodologies for stance detection on social media platform.

	Core Technique	Performance		
		F_{favor}	$F_{against}$	F_{avg}
Standard Supervised Learning:				
Zarrella [41]	RNN, word2vec, representation learning	59.32%	76.33%	67.82%
Wan [33]	CNN, word embeddings, vote scheme	61.98%	72.67%	67.33%
Boltuzic [36]	Word embeddings, ensemble learning, SVM	60.93%	72.73%	66.83%
Elfardy [34]	lexical & semantic feature, SVM	54.99%	72.21%	63.60%
Prashanth [40]	Character CNN, word embeddings	58.44%	68.65%	63.54%
Igarashi [37]	CNN, word embeddings	49.25%	75.18%	62.21%
Wojatzki [32]	Stacked classifier, syntactic feature	48.71%	74.75%	61.73%
Patra [31]	Lexical and syntactic feature, SVM	46.68%	74.53%	60.60%
Misra [35]	LIWC, dependency based feature, SVM	50.90%,	67.81%	59.36%
Wei and Mao [42]	CNN, word embedding, target embedding	65.52%	76.55%	N/A
Gadek [43]	Co-occurrence graph, SVM	N/A	N/A	65.00%
Benton and Dredze [44]	RNN, author embedding	N/A	N/A	53.00%
Sun and Wang [45]	LSTM, hierarchical attention	N/A	N/A	61.00%
Zhou and Cristea [46]	GRU, CNN, semantic attention	N/A	N/A	69.42%
Wei and Lin [47]	Memory Network, word embedding	N/A	N/A	56.73%
Du and Xu [48]	LSTM, attention mechanism	N/A	N/A	68.79%
Dey [49]	Lexicon, WordNet, SVM	69.53%	79.36%	74.44%
Weakly Supervised Learning:				
Wan [33]	CNN, automatic labeling	57.39%	5.17%	59.36%
Dias [38]	Weakly supervised learning, labeling rules	32.56%	50.09%	42.32%
Augenstein [39]	Stacked auto-encoder, word embeddings	10.93%	54.46%	32.70%
Wojatzki [32]	Stacked classifier, transfer learning	46.56%	5.71%	26.14%
Dey [49]	Lexicon, WordNet, SVM	59.72%	63.41%	61.57%
Xu and Paris [50]	LSTM, attention mechanism	N/A	N/A	43.80%

TABLE 5. An example of stance detection on social media.

target	stance	tweets
legalization of abortion	favor	'The pregnant are more than walking incubators. They have rights too!'
Hillary Clinton	neutral	'Hillary Clinton has strength and some weaknesses.'
Donald Trump	against	'Jeb Bush is the only sane candidate in this republican lineup!'

who said Trump was mentally ill!, *Flynn misleed them. He still had Trump's complete trust?* " and etc. Secondly, users of social media prefer to use symbols and emoji to express their opinion, like "*Vote Trump hoes!!!√ √ √ :)thumb up*" and "*Hilary Clinton number one supporter ♡♡♡*". Obviously, it is difficult to process such reviews using traditional text analysis tools.

1) PROBLEM SETTINGS

The stance detection on social media platform can be formalized as: given the target t_e and related corpus C_t^e, for each tweet $tweet_i \in C_t^e$, the model should determine the stance $s_i^{t_e} \in \{favor, neutral, against\}$. An example is given as Table 5.

Mohammad [94], [116] constructs an annotated corpus in which all 4163 tweets are assigned with a stance label except the tweets related to '*Donald Trump*'. It covers six hot topics (*'Atheism#733', 'Climate Change is a Real Concern#564', 'Feminist Movement#949', 'Hillary Clinton#984', 'Legalization of Abortion#933' and 'Donald Trump#707', and the number after '#' denotes the size of the corpus*).

Based on the problem setting and the corpus, dozens of work such as stance detection based on standard supervised learning [31]–[37], [40]–[49] and weakly supervised stance detection [32], [33], [38], [39], [49], [50], have been devised recently. Table 4 describes the core techniques and the statistical results of related models.

2) STANCE DETECTION BASED ON SUPERVISED LEARNING

The first seventeen rows of Table 4 list the results of standard supervised detection methods. Both traditional feature engineering (e.g. lexical feature, syntactic feature and semantic feature) and prevalent deep learning based techniques (e.g. CNN [79], RNN [97], LSTM [45] and Memory Network [42]) have been employed for the task.

In this part, the system designed by Zarrella [41] is selected as a representative method to show the process of stance detection based on supervised learning. Fig. 3 shows the four layer neural network employed in [41]. Firstly, the input terms are encoded in a one-hot fashion, and each term is represented by a sparse binary vector containing a single one-value at the

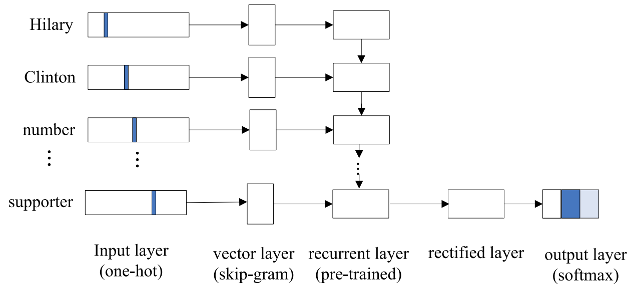


FIGURE 3. **A Long-Short Term Memory based stance detection system.**

index corresponding to the term's position in the vocabulary. Then, each term is represented by a 256-dimensional word vector using the embedding layer. The weights of embedding layer are pre-trained by the skip-gram [99] model on a corpus with 218M tweets. The third layer in network employs Long-Short Term Memory (LSTM) units to capture the context dependency of the text. Besides, to overcome the shortage of labeled training data (contains only 4K topic related tweets) and learn a suitable representation, the system collects a tweet corpus (constituted by 298K tweets corresponding to 197 different hashtags). And a 197-dimension rectified layer is used as the output of classifier to pre-train the LSTM layer by predicting the hashtag of each tweet in constructed corpus. Both word2vec model and hashtag prediction task are used to initialize weights for stance detection neural network. Finally, the whole network is trained on the annotated corpus provided in [116], both embedding layer and LSTM layer are fine-tuned in this process. By using these two pre-train strategies, the knowledge in external corpus is transferred to domain specific corpus and the stance detection system achieves a good performance.

Due to the strong representative ability of the deep neural networks, convolutional neural network based models [33], [37], [40], [42], [46] are also chosen for the supervised stance detection. The system, designed by Wei [33], employs the Google News corpus to train the word2vec and the learned word embeddings are taken as the input. Besides, the '*vote scheme*' (e.g. for each tweet in test set, ten candidate labels are employed to vote the final stance of the tweet) and the '*divide and conquer scheme*' (e.g. training set and test set are separated by specific topic and the models are trained respectively) are incorporated to improve the performance. Tohoku [37] uses the word embeddings trained by Continuous Bag-Of-Words (CBOW) model on Wikipedia article corpus. A comparison is also conducted with hybrid feature-based logistic regression model (e.g. Reply, Bag-Of-Words, Bag-Of-Dependency and SentiWordNet features). Experimental results show that CNN based approach performs better in validation set while hybrid feature based classifier achieves better performance in test set. To explore the impact of embedding granularity, Prashanth [40] proposes the word-level CNN and the character-level CNN based model with a novel data augmentation technique which could expand and diversify the training dataset. And the proposed approach reaches a relatively good performance (F_{avg} : 63.53%). To incorporate the target information, Wei and Mao [42] propose a model which could build tweet representation conditioned on the target and obtain a good performance ($F_{against}$: 76.55%). Likewise, Zhou [46] models the target under a GRU-CNN framework.

On the other hand, the RNN and its variants, such as LSTM and GRU, have also been employed for the task [44], [45], [48]. To use the meta information of the tweets, Benton [44] employs a semi-supervised approach to predict the author embedding and incorporate it into the detection process. However, the proposed model performs the worst in the all supervised models, this because the predicted author embedding may incorporate the noise information into the model and impact the performance. Sun and Wang [45] employ a hierarchical attention mechanism into LSTM framework to capture the various linguistic information which are helpful for detecting stance. Likewise, Du and Xu [48] incorporate the target information into LSTM framework using a target specific attention, this specific type of attention make the neural model more sensitive to the target specific stance information and obtain an improved performance. Compared with the Du and Xu's work, [47] obtains a relatively bad results which may because that Wei's model do not consider the target information when predicting the stance of the tweet.

Besides, there also exists some work focus on feature engineering and traditional classifier. Approaches in [31], [34]–[36], [43], [49] use SVM as classifier and explore various feature, such as text feature and dependency based features. Patra [31] uses the sentiment feature (e.g. built by SentiWordNet, NRC Emotion Lexicon and Hashtag Emotion Lexicon [125]) and the dependency relation which is created by Stanford parser (e.g. search the word pairs in dependency relations that consist of two component words, one is '*favor*' or '*against*', and the other should appear in SentiWordNet). Elfardy [34] employs the latent semantic feature which is obtained by latent dirichlet allocation or weighted textual matrix factorization to capture the semantic information in tweets. Misra [35] employs the LIWC feature and proposed a feature based approach. Gadek and Betsholtz [43] mine the contextonyms and contextosets from the co-occurrence graph to build context-based feature and obtain F_{score} : 65.0%. Boltuzic [36] designs a hybrid feature set containing word features, word embeddings, document statistical feature and hashtag features. Besides, he proposes an ensemble approach based on the genetic algorithm. Wojatzki [32] views the task as a multi-dimensional classification problem and employs a stacked classifier which firstly identifies the tweets that contain the stance from neutral tweets and then detects the specific stance. Likewise, Dey [49] devises a two stage approach by using various features (e.g. MPQA subjectivity lexicon, WordNet Adjective) and obtains the best overall performance (F_{avg} : 74.44%). The efficiency of the Dey's work may resort to the effectiveness of semantic dictionary (such as subjectivity lexicons, WordNet adjectives

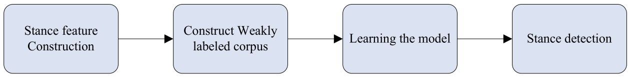


FIGURE 4. Flowchart of the weakly supervised stance detection system.

and sentiment lexicons) and elaborated handmade features (such as n-grams, sentiment feature, target, POS tags) [49].

From the statistics in Table 4, we could observe that the stance detection approaches on social media platform do not perform very well. And this may be caused by the following factors: i). The sparsity and the flexibility of the social media text, such as examples in Table 5, make it difficult to extract high-efficiency document representation for classification. ii). The datasets of social media stance detection [116] have limited labeled document. Within the labeled datasets, the corpus related to *'Hillary Clinton'* is the largest one which contains only 984 training instance and the limitation of the training corpus may also deteriorate the performance, especially for the deep learning based approaches.

3) WEAKLY SUPERVISED STANCE DETECTION

Weakly supervised stance detection aims to assign a stance to each tweet by using both labeled (source domain, e.g. '*Atheism*', '*Hillary Clinton*' and etc) and unlabeled collection (target domain, '*Donald Trump*' provided in [116]). The last six rows in Table 4 describe the approaches for weakly supervised stance detection.

To illustrate how the weakly supervised stance detection system works, we choose the [38] as example. The flowchart is shown in Fig. 4. The approach proposed in [38] firstly identifies the stance features (e.g. the n-grams which could indicate a stance or sentiment polarity) in the unlabeled corpus by inspecting the high frequent n-grams and employs the well-defined filtering rules to remove the irrelevant tweets. Then, three off-the-shelf sentiment analysis APIs (HP Haven On Demand [126], IBM Alchemy [127] and Vivekn [128]) are employed to automatically assign a stance for each tweet according to the summation of three candidate scores. Finally, the proposed approach labels the tweet as *favor* if the value is positive and vice versa. Specifically, the tweet will be labeled as *neutral* when the scores achieved from Haven and Alchemy are equal to zero. And a SVM classifier which is trained on the constructed noisy labeled training set is used to predict the stance for tweets in corpus about 'Donald Trump'.

Similarly, Wan [33] employs a two-step strategy to build the weakly labeled corpus automatically and predicts the stance of tweet in the unlabeled dataset. The corpus is built based on the assumption that some expressions and hashtags may reveal the tendency of the stance to a specific topic. For example, '*go trump*' and '*#MakeAmericaGreatAgain#*' reveal favor tendency, '*idiot*' and '*#BeatTrump#*' could induce against tendency to Trump. Wan employs the tendency expressions and hashtags (e.g. '*idiot*', '*go trump*' and etc) to retrieve the tweets in domain corpus provided in [116]. Thus, Wan obtains a training set with noisy label which contains 2K favor tweets and 3K against tweets. Then, the constructed corpus is used to train a three class CNN-based detector. Likewise, another two-stage approach is proposed in [32]. Firstly, a tweet is considered as neutral if no target related frequent nouns appear. Then, the topic specific classifiers are trained on five labeled datasets. For each opinionated tweet, a topic which is most similar to the tweet is chosen, and the corresponding classifier is selected to detect the stance.

Besides, deep learning based approaches have also been employed for the task. Augenstein [39] uses the stacked auto-encoder to learn the representation of the tweets. And the auto-encoder is trained on a collected corpus (contains 395K unlabeled tweets) which is related to all the six targets provided in [116] (e.g. '*Hillary Clinton*', '*Atheism*', '*Donald Trump*' and etc). Thus, the trained auto-encoder learns the mapping function and reduce the domain gap. Finally, Augenstein employs the representative vector (mapped by the learned encoder) of labeled tweets in [116] to train the stance detector and obtains a comparable result.

It could be observed from Table 4 that weakly supervised models performs worse than the standard supervised approaches. This may because: i). weakly supervised stance detection requires the models to transfer the knowledge contained in source domain (e.g 'Atheism', 'Hillary Clinton', 'Climate Change', 'Feminist Movement' and 'Legal. Abortion') into the target domain 'Donald Trump'. Compared with the standard supervised detection task, it is more challenging. ii). the reviewed approaches in this field employ the learned knowledge from source domain to label the training data in target domain which may feed noise in label space and such noise will also impact the performance. iii). As discussed in last subsection, the dataset contains only hundreds of train data in each domain, the shortage of the train data make it hard to learn an effective neural network to transfer the knowledge into the target domain which may also result in the bad performance of the weakly supervised models.

Although many deep learning based approaches [33], [41], [46], [48] obtain better results than the traditional feature based models [34]–[36], the traditional feature engineering based method [49] performs the best. This may because: i). Dey [49] employs so many elaborate features (such as sentiment lexicons, n-grams, POS tags) to capture the syntactic and semantic meaning of the tweets; ii) the small number of training data in this field limit the representation ability and the efficiency of the deep learning based model. Thus, we also believe that the deep learning based approaches will beat the traditional methods as long as the model has more labeled training data (such as thousands of tweets).

IV. PRODUCT ASPECT MINING

With the development of Internet and the usage of e-commerce, customers prefer to post reviews of products and show their opinions on shopping websites (e.g. Amazon, ebay and etc) which provides plentiful information for marketing intelligence.

TABLE 6. Methodologies for corpus level product aspect extraction.

	Core technique	Performance
Rule based Methods:		
Hu and Liu [129]	POS tagging, feature pruning	Pre:72.0%, Rec:80.0%
Liu and Hu [55]	POS tagging, association rules	Pre:84.0%, Rec:86.3%
Scaffidi. C [130]	POS tagging, lemma examining	Pre:85.0%-90.0%
Zhao [53], [146]	Syntactic similarity, kernel	Pre:71.77%, Rec:66.0%
Li [131]	Extraction patterns, adjective rule	F_{score}:74.0%
Zhang [132]	Dependency relation, HITS algorithm	Pre:72.3%, Rec:55.0%
Zhen [54]	Association rule, clustering	Pre:76.2%, Rec:72.7%, F_{score}:74.4%
Konjengbam [56]	POS tagging, ontology	Pre:56.00%, Rec:73.68%, F_{score}:63.63%
Vo [57]	Dependency parser, CRF	Pre:63.4%, Rec:81.8%, F_{score}:71.4%
Unsupervised Approaches:		
Raju [133]	Clustering, attribute extraction	Pre:92.4%, Rec:62.7%
Yu [134]	Aspect ranking, regression algorithm	F_{score}:71.0%-76.0%
Titov. I [61]	Topic model, local topic, global topic	N/A
He R [64]	Attention mechanism, word embedding	N/A
Yang Y [65]	Topic model	Pre:79.1%
Rosa A [66]	LDA, knowledge mining	Pre@10:75.0%, Pre@20:59.0%
Chen R [67]	Domain ontology, LDA	F_{score}:75.0%
Angelidis S [68]	Attention mechanism, word embedding	F_{score}:49.1%

Product aspect mining, aiming at extracting the aspects and corresponding opinions from the product review, will benefit customers and merchants by helping them making smart purchase decision and efficient marketing strategy. Therefore, This research topic attracts many researchers and has been extensively explored in last decade. Previous works could be generally divided into three groups which are corpus level aspect extraction [53]–[57], [61], [64]–[68], [129]–[134], corpus level aspect and opinion mining [51], [58]–[60], [62], [63], [69]–[71], [101]–[103], [113], [118]–[120], [135]–[137] and document/sentence level aspect and opinion tagging [72], [74]–[78], [80]–[82], [84]–[86], [88]–[90], [138], [139], [144], [145]. For the former two categories, rule-based methods and unsupervised based methods are commonly used, and various supervised learning models (e.g. HMMs, CRFs and Deep Neural Networks) are proposed for the third category which is commonly viewed as a sequential tagging problem.

A. CORPUS LEVEL ASPECT EXTRACTION

1) PROBLEM SETTINGS

Given the specific product review corpus, corpus level aspect extraction aims to extract global $< target, aspect >$ pairs. And it could be formalized as:

Given the target t_{ca} and target specific review corpus C_t^{ca}, the model will output the $< target, aspect >$ pairs in corpus level and it does not need to identify which aspect is reviewed in the specific review. To directly illustrate the task, we use the '*iPhoneX*' as an example. When given the specific target '*iPhoneX*' and a review corpus about '*iPhoneX*', the mining approach should output the pairs as:

Example:
Inputs: An iPhoneX review corpus generated by consumer.
Target: iPhoneX
Outputs: $<$ iPhoneX, screen$>$,
$<$ iPhoneX, battery$>$,
$<$ iPhoneX,camera$>$,
$<$ iPhoneX,weight$>$,
$<$ iPhoneX,portability$>$,
$<$ iPhoneX,appearance$>$,
$<$ iPhoneX,size$>$
and ...

Each aspect in the output tuple represents a particular component (e.g. *screen, battery*) or an attribute (e.g. *portability, usability*) of the product that many consumers care about.

To extract high-quality and representative aspects of the reviewed product, various approaches have been proposed and could be categorized into rule-based and unsupervised models. Table.6 presents the main approaches for aspect extraction together with their core techniques and the performance.

2) RULE-BASED ASPECT EXTRACTION

Rule-based aspect extraction approaches [53]–[57], [129]–[132] dominate the field at the early stage. In detail, it could be further categorized into frequency based methods, relation based methods and hybrid methods.

These approaches share the same assumption that an aspect term should appear frequently in the corpus, and most works believe that the aspects are nouns or noun phrases. Based on the two mentioned criterions, appropriate rules, such as POS patterns, syntactic dependency patterns and associa-

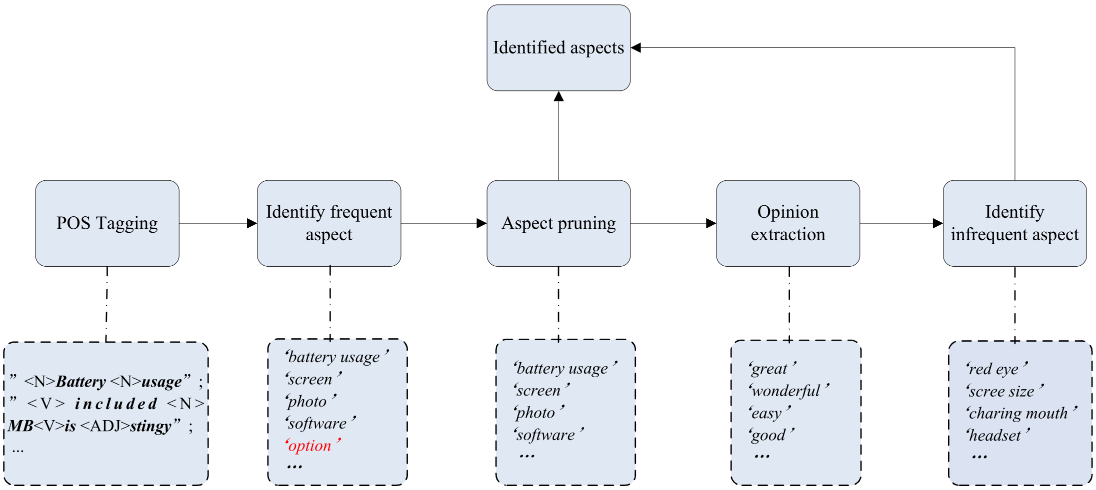


FIGURE 5. A framework of frequency and relation based aspect extraction system.

tion mining rules, are employed for aspect extraction. The $1st - 9th$ rows in Table 6 list the rule-based approaches for corpus level aspect extraction.

To delve into the mechanism of rule-based aspect extraction system, the method reported in [129] (Fig. 5 depicts the framework) is employed to illustrate the extraction process. As shown in Fig. 5, the proposed approach contains two stages: frequent aspect extraction and infrequent aspect extraction. When the product-specific review corpus (e.g. related to Digital camera, Cellular phone, Mp3 player or DVD player) has been crawled from e-commerce sites, the mining system firstly conducts POS tagging on the corpus. Then, aspect generation module which is based on the apriori algorithm is employed to extract the frequent noun terms and noun phrases as candidate set. Obviously, not all terms in candidate set are aspects. To improve the precision of the extraction system, two pruning strategies (compactness pruning and redundancy pruning) are designed to filter the uninterested and redundant ones in the candidate set. For example, '*life*' is not an appropriate aspect while '*battery life*' is a good aspect candidate in an electronic device related topic. After pruning the extracted aspect set, the system will turn to opinion word extraction phase, which also helps to extract the infrequent aspects. Based on the assumption that adjective which appears near the aspect term is prone to express an opinion, the adjectives nearby the aspect word are identified to construct the opinion word set. Finally, the infrequent aspects (discussed by a small group of customers but attractive to potential users, e.g. '*red eye*' in a camera review) are extracted by identifying the noun terms or noun phrases nearby the opinion words in each sentence. Experimental results show that the proposed approach achieves the performance of $Recall$: 80% and $Precision$: 72%.

Similarly, Liu and Hu [55] propose an aspect extraction and visualization approach, named OpinionObserver. In [55], Liu proposes a novel extraction approach for review collections with special format (Each review contains Pros, Cons and the detailed review. Namely, consumers are asked to express Pros and Cons briefly together with a detailed review. For example, reviewer could write *'heavy, bad picture quality, battery life too short, and keyboard easy to use'* to express Cons opinion). Liu employs a POS-tagger to preprocess the corpus and generate a set of language patterns for aspect words extraction. For example, the pattern '[*aspect*] easy to $<v>$' will match the aspect '*keyboard*' from '*keyboard easy to use*'. Based on the assumption that domain specific term prone to appear more frequent in domain corpus than the general corpus, Scaffidi proposes an aspect scoring method in [130] to evaluate the aspect quality. The proposed approach firstly detects the frequent uni-gram nouns and noun phrases. Then, a 100 million-word corpus is employed as the general corpus to evaluate aspect candidateshigh score candidates are considered as aspects. Finally, this ranking approach improves the quality of extracted aspects ($Precision$: $85\% \sim 90\%$). Likewise, Zhang [132] proposes a two-stage approach to rank the extracted aspects according to the aspect importance and the score is estimated Hyperlink-induced Topic Search (HITS) algorithm. To incorporate the syntactic structure information into extraction system, Zhao [53] employs a tree kernel based approach to capture the relations between aspect and opinion words which helps to obtain high quality aspects.

The approaches discussed above follow the assumption that product aspects are nouns or noun phrases and could not extract the aspects such as '*operating system*' which matches '*Verb+Noun*' pattern. To deal with such problem, Li [131] employs four POS patterns (e.g. '*Noun*', '*Verb*', '*Noun+Noun*', '*Verb+Noun*') to extract the aspect candidates and creates a noisy aspect set. Further, a syntax adjective rule (most customers use the syntax '*aspect+adjective*' to express their opinion to product, e.g. "*The operating is simple, and the reaction is quick*") is employed to filter the

MP3 reviews	**sound quality** : sound quality headphones volume bass earphones good settings ear rock
	appearance : case pocket silver screen plastic clip easily small blue black
	controls : button play track menu song buttons volume album tracks artist
	battery : battery hours life batteries charge aaa rechargeable time power lasts
	tech. problems: reset noise blacklight show freeze turn remove playing icon creates
	radio/recoding: radio fm voice recording record recorder audio mp3 microphone wma
Hotel reviews	**bathroom** : shower water bathroom hot towels toilet tub bath sink pressure
	parking : parking car park lot valet garage free street parked rental
	staff : staff friendly helpful very desk extremely help directions courteous concierge
	comfort : room bed beds bathroom comfortable large size tv king small
	location : walk warking restaurants distance street away close location shopping shops
	pricing : $ night rate price paid worth pay cost charge extra

FIGURE 6. Extracted aspects for MP3 and Hotel reviews using unsupervised models [61].

noisy aspect. These four elaborate patterns and adjective rules help to improve the quality of the extracted aspects.

Differs from the discussed explicit aspect extraction approaches, Zhen [54] focus on the implicit aspect extraction task. A novel approach based on co-occurrence association rule mining is proposed in [54]. At the first stage, association rules (in the form of [*opinion, explicit aspect*]) are generated from the co-occurrence matrix. Afterwards, aspect clusters could be constructed by clustering the rules according to their semantic similarity. When given a review without any explicit aspects, the system could select the cluster with the highest frequency weight and choose one representative word as the implicit aspect of the review. Likewise, Konjengbam [56] employs POS-tags and dependency tree to mine the frequent nouns as aspect terms. Based on the extracted aspect terms, Konjengbam also devises an ontology based opinion summarization method in [56].

3) UNSUPERVISED APPROACHES FOR ASPECT EXTRACTION

Rule-based approaches develop rapidly at the early stage due to simplicity and effectiveness. However, they have several limitations: 1) too many non-aspects are produced and low-frequency aspects are missed; 2) it is hard to adapt manual constructed patterns to another domain. Moreover, such models often ignore the semantic similarity between aspects and generate redundant aspects. For example, '*price*', '*cost*' and '*fee*' are three semantic relatedness aspect terms extracted by the rule-based approaches, and '*cost*' and '*fee*' are the redundant aspects.

In this subsection, unsupervised approaches such as the topic model, clustering algorithms which could extract the product aspect from unlabeled review corpus are discussed. Compared with the rule-based approaches, unsupervised models have several advantages: 1) The extracted aspects can be grouped according to their semantic similarity; 2) The proposed approaches are domain independent and could be transferred to new domain easily. The last eight rows of Table 6 present eight representative unsupervised approaches, which will be discussed in this subsection in details.

As pointed in [61] that standard topic models, such as LDA and PLSA, prefer to generate global properties (e.g. product type or brand) rather than the local aspects as they consider word co-occurrences at document level. Therefore, Multi-Grain Topic model [61] is proposed for corpus level aspect extraction. Titov distinguishes the global topics and local topics in MG-LDA. Local topics are used to capture the local aspects and the global properties of the reviewed product will be captured by the global topics. In MG-LDA, a document d is generated from a mixture of global and local topics. Titov employs a set of sliding windows (each window has its particular local topic distribution and constituted by T adjacent sentences) to represent the document. Also, MG-LDA assumes that document d has a fixed global topic distribution and a varies local topic distribution during the generative process. In other words, a sentence s should be generated from the global topic mixture and windows mixture which covers sentence s. Further, the overlap between sliding windows provide a large co-occurrence domain. And the extraction results on two different review corpus (*MP3 reviews and Hotel reviews*) are shown in Fig. 6. Bold terms, at the beginning of each line, are the aspects extracted from the review collections, italic words behind the colon are the corresponding representative terms of each aspect. And the tokens in each line are sorted by the probability of each term according to the word distribution of specific topic/aspect(e.g. *battery, location* and *pricing*) in descend order.

Along this line, scholars have devised some similar topic models [65]–[67]. To model the aspect relationship between different products, Yang [65] assumes that child categories will inherit aspects from parent categories, and the proposed approach incorporates the category hierarchy information into modeling process to enhance the extraction ability. To enhance the coherence of the generated aspect topics, Shams [66] mines knowledge automatically from pre-extracted topics and proposes an ELDA to inject the knowledge into modeling process.

On the other hand, clustering algorithms are also employed for aspect extraction. For example, group average agglomerative clustering (GAAC) algorithm is employed by Raju [133]. Raju considers both semantic similarity of aspects and the domain specialty. Raju firstly employs the term frequency and KL divergence of term probabilities in two corpus (a domain specific corpus and a general English corpus) to detect the domain aspect candidates. Subsequently, the similarity matrix based on the Dice's Coefficient similarity is built to group the aspect candidates into clusters according to their semantic similarity. Finally, the approach selects the high ranked phrases in each cluster as aspects and gets a competitive *Recall* : 62%. Inspired by the fact that important aspect are usually commented by a large number of consumers and one's opinion on important aspect could greatly influence their overall opinion on the product, Yu [134] develops an aspect ranking approach which considers both term frequency and the importance of the detected aspect to enhance the performance of aspect extraction.

Besides, several work based on unsupervised neural network have also been proposed [64], [68]. To solve the incoherent aspect of the traditional extraction model, He [64] represents document with word embedding and proposes a neural-based aspect extraction model which combines the attention mechanism with reconstruction loss. Likewise, Angelidis [68] employs the limited annotated data to mine the aspect-specific seed words as the prior knowledge of aspect embedding. And with the help of this initialization mechanism, the proposed the model obtains more coherent aspect.

B. CORPUS LEVEL ASPECT AND OPINION MINING

Differs from the aspect extraction discussed above, corpus level aspect and opinion mining aims to extract the target related triples < *target, aspect, opinion*> from the review corpus. In this subsection, we will briefly survey the methodologies for corpus level aspect and opinion mining.

1) PROBLEM SETTINGS

Given a specific target product t_{cao} and a homologous review collection C_t^{cao}, corpus level aspect and opinion mining extracts the aspects A related to t_{cao} and the opinions O to each aspect a_i in corpus level. Commonly, the opinion expressed on the aspect (the third element in the triple) could be described in three types: 1) emotional adjectives are used (e.g.'*cool appearance*'); 2) two or five star rating (e.g. high rating score means positive opinion); 3). an ordered set of non-numerical labels (e.g. *poor, average, good, very good, excellent*) is used. Take '*Mac*' as an example. The model uses the target '*Mac*' and the specific corpus as input and should output the corpus level triples:

Example: (Here, we use two level rating 'good' and 'bad')

Inputs: A review corpus of Mac generated by users

Outputs: < Mac, sound, good>,
< Mac, CPU, good>,
< Mac, price, bad>,
< Mac, screen,good>,
< Mac, appearance, good>,
< Mac, weight, bad>,
and ...

It can be observed from the example that aspect and opinion mining approach outputs plentiful useful information about public opinion on '*Mac*', such as consumers care much more about the '*sound*', '*CPU*', '*price*', '*screen*' and '*appearance*' of Mac. Moreover, lots of reviewers think Mac is too expensive.

Methodologies for corpus level aspect and opinion mining could be categorized as frequency and relation based approaches [51], [58]–[60], [101], [103], [135], [137] and topic model based approaches [62], [63], [69]–[71], [102], [113], [118]–[120], [136]. Table 7 describes the approaches together with their core techniques and performances.

2) FREQUENCY AND RELATION BASED APPROACHES

Frequency and relation based approaches share the assumption that aspect words and opinion expressions should appear in the review corpus frequently. In general, these mining systems follow the procedure (*aspect extraction→ opinion extraction → opinion word orientation identification*). And the representative methods are list in $1st-8th$ rows of Table 7. More detailed discussion will be presented below.

As far as we know, the first mining system in this subdomain is the feature-based summarization (FBS), developed by Hu [51]. Hu firstly employs the frequency based approach [129] to extract the product aspects and opinions words. Then, an orientation identification algorithm based on a pre-defined seed set (e.g. a small set of opinion words) and the semantic structure of WordNet are employed to automatically identify the opinion orientation. Similarly, OPINE [137] employs an aspect assessment method based on pointwise mutation information (PMI) and the syntactic dependency rules to improve the quality of extracted aspect terms and opinion expressions. Then, the relaxation labeling is employed by OPINE to detect the orientation of opinion expression in the specific context. To incorporate the external sentiment information, Jiang [135] defines four tree kernels and proposes a tree kernel based approach to incorporate the syntactic dependency and sentiment information.

TABLE 7. Approaches for corpus level aspect and opinion mining.

	Core technique	subtask	Performance
Frequency and Relation based Approaches:			
Hu [51]	POS tagging, pruning, WordNet	aspect extraction, opinion extraction	Pre:72.0%, Rec:80.0% Pre:64.2%, Rec:69.3%
Baccianella [101]	Ordinal regression [147], General Inquirer [104]	aspect rating	MAE^M : 1.032
Popescu [137]	Relaxation Labeling	aspect extraction, opinion extraction	Pre:94.0%, Rec:77.0% Pre:71.0%, Rec:78.0%
Moghaddam [103]	syntax pattern, WordNet	aspect extraction, aspect rating	Pre:80.0%, Rec:87.0% Ranking loss:0.49
jiang [135]	Syntactic parser, tree kernel	opinion extraction	Pre:89.8%
Das [58]	POS tagging, Fuzzy logic	aspect rating	N/A
Shafie [59]	POS tagging, Dependency parser	aspect extraction	F_{score}:35.98%
Srividya [60]	POS tagging, Chunk parser	aspect extraction	Pre:89.82%, Rec:94.19%
Topic Model based Approaches:			
Titov [63]	Topic model, sliding window	aspect extraction	N/A
Brody [119]	Topic model, adjective graph	aspect extraction	N/A
Zhao [120]	Topic model, MaxEnt-LDA model	aspect identification, opinion identification	F_{score}:70.5% Pre@5:82.5%
Jo [113]	Topic model, ASUM model,	opinion identification	Acc:84%-86%
Wang [102]	generative model, LARAM model	aspect extraction, aspect rating	KL-div:11.1 MSE:1.23
Zhan [136]	Topic model, word pair dependent model	aspect extraction	N/A (graph only)
moghaddam [118]	Topic model, Interdependent LDA	aspect extraction, opinion identification	RandIndex:83%, RandIndex:73.0%
Lakkaraju [62]	Topic model, CFACTs model	opinion identification	Acc:81.3%
Ekinci [69]	Topic model	aspect and opinion pair extraction	N/A(pair list)
Xiao [70]	Matrix Factorization, Topic model	aspect rating	MSE:0.42
Zuo [71]	Topic model, Dirichlet Process	aspect identification	Pre:88.9%, Rec:82.6%, F_{score}:85.6%

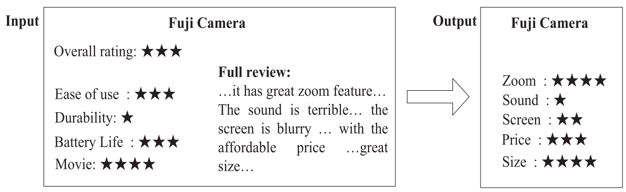


FIGURE 7. An example of aspect extraction and aspect rating.

On the other hand, Baccianella [101] views aspect rating as a regression problem and propose an ordinal regression based approach using various features, such as n-grams, POS pattern, General Inquirer [104].

To explain how the frequency and relation based system works, we choose Opinion Digger [103] as an example. Instead of solely using review corpus, auxiliary information is employed in Opinion Digger. As shown in Fig. 7, the collection contains a set of pre-defined aspect set (e.g. '*Ease of use*', '*Durability*' and '*Battery life*' in Fig. 7) and a rating guideline (e.g. '*terrible*: 1', '*poor: 2*', '*average: 3*', '*good: 4*' and '*excellent: 5*') which infers the relationship between opinion words and corresponding numerical ratings. Opinion Digger aims to extract the aspect and opinion terms together with predicting the aspect-specific rating (range from 1 to 5) using corpus and additional information as input. In general, the whole mining process should be divided as aspect extraction phase, opinion extraction phase and the rating prediction phase. In aspect extraction phase, all the noun terms are firstly extracted using the apriori algorithm. Then, a set of POS patterns in review corpus are identified based on the pre-defined aspects. For example, the pattern [*_JJ _ASP*] (e.g. adjective+aspect) is extracted from the aspect phrase '*long battery life*'. And the constructed patterns are used to refine the candidate aspect set. After the aspect extraction phase, aspect specific opinion words are identified based on the nearest rule (e.g. if opinion word modifies the aspect, they should co-occurrence with a small distance). To estimate the aspect specific rating, the system should firstly detect the aspect specific opinion words. For each opinion word, Opinion Digger employs the WordNet synonymy hierarchy to find the two nearest opinion words in rating guideline (e.g.'*terrible*', '*poor*', '*average*', '*good*' and '*excellent*'). For example, '*terrible*' and '*poor*' are two nearest opinion words of '*defective*' and their corresponding ratings could be used to predict the rating score of '*defective*'. Further, Opinion Digger predicts the corpus level aspect based rating by aggregating the ratings of the relevant opinion words to the specific aspect.

To identify multiple word aspect, POS tags and word embedding are employed in [58], and Das checks the spelling using Fuzzy Logic tools to improve the extraction performance. Besides, it is a challenge work to extract aspect and opinion correctly in the review that consists of multiple aspect with various opinion. Shafie [59] employs a dependency parser to capture the relation between aspect and opinion words and solves the problem.

Hotel reviews	Room Condition	**aspect** : room bathroom bed air tv conditioning water rooms beds bath
		opinion : shower small clean comfortable hot large nice safe double well
	Ambience	**aspect** : room floor hotel noise street view night breakfast room terrace
		opinion : quiet open small noisy nice top lovely hear overlooking beautiful
	Meal	**aspect** : breakfast coffee fruit buffet eggs pastries cheese room tea cereal
		opinion : good fresh continental included hot cold nice great delicious adequate
	Serice	**aspect** : staff desk hotel english reception help service concierge room restaurant
		opinion : helpful friendly front polite courteous pleasant asked good excellent rude
Electronics reviews	battery	**positive** : battery long great lithium huge convenient capacity serviceable
		negative : battery charge volume frequent broken damage color help
	price	**positive** : worth monei extra well everi price dollar spend save cost hundr
		negative : monei save notwast wast yourself notbui spend notworth time favor
	screen	**positive** : screen color bright clear video displai crisp great quality picture
		negative : screen glossi display keboard bright glare view color light lcd reflect

FIGURE 8. Example of two forms of extracted aspects and opinions by topic model based Approaches [113], [120].

3) TOPIC MODEL BASED APPROACHES

Although the extraction process is simplified by assuming that aspects and opinions are noun and adjective terms, the infrequent or non-noun aspects terms (e.g. '*red eye*' for product camera) may be omitted. Moreover, the errors in early stage of the pipeline system will propagate and heavily affect the final extraction performance. To overcome such disadvantages, topic model based approaches have been proposed. And many various of LDA [148] have the ability to perform corpus level aspect extraction and opinion mining in parallel. In Table 7, last eleven rows describe the topic model based approaches [62], [63], [69]–[71], [102], [113], [118]–[120], [136] for corpus level aspect and opinion mining.

Based on the different way of aspect and opinion modeling, topic model based approaches could be further categorized into two classes which are separated extraction and joint extraction. By dividing the topics into two categories, aspect topics and opinion topics, separated extraction models output aspects and opinions separately as shown in '*Hotel reviews*' in Fig. 8. While jointly extraction model assume that aspect specific terms, both aspect and opinion words, are generated from one aspect related topic, and the example of the output is shown as '*Electronics reviews*' in Fig. 8.

To interpret the topic model based mining approaches and their mechanism, we employ the MaxEnt-LDA [120] (e.g. an integration of generative topic model and a discriminative maximum entropy module) as an example which extracts the aspects and opinions in the form of '*Hotel review*' shown in Fig. 8. With the usage of a small amount of training corpus, it could extract the aspect words and aspect specific opinion words separately. The MaxEnt-LDA categorizes the topics as background topic t^B, global aspect topic $t^{A,g}$, global opinion topic $t^{O,g}$, aspect topics $t^{A,i}, i \in \{1, 2, 3, \ldots, T\}$ and aspect specific opinion topics $t^{O,i}, i \in \{1, 2, 3, \ldots, T\}$. Differs from the Latent Dirichlet Allocation, MaxEnt-LDA makes a further assumption that each sentence is assigned to one specific aspect. Namely, each review sentence s is generated from a mixture of background topic (t^B), global topics (both $t^{A,g}$ and $t^{O,g}$) and the assigned aspect specific topics ($t^{A,i}$ and $t^{O,i}$). In detail, two switch variables $u \in \{0, 1\}$ and $y \in \{0, 1, 2\}$, draw from binomial and multinomial distribution, are designed to cooperatively determine the source topic of current word w. The variable u determines the category (background, general or aspect) of w. Due to the deficiency of symmetric dirichlet prior (induced in [149] that fully unsupervised model could not separate opinion words and aspect words efficiently), Zhao uses the part-of-speech information of the context words which is helpful for distinguishing the aspect word and opinion word as features and designs a maximum entropy classifier to determine y for the current word w. And the model parameters could be learned from a small domain independent training corpus (annotated with background, aspect and opinion). Combing the advantages of generative model and supervised information of pre-trained ME-classifier, MaxEnt-LDA naturally encodes the external information (e.g. lexical features and POS tags) into the model. And the side information really contributes to the aspect and aspect-specific opinion mining.

Besides, scholars have attempt various research directions. Moghaddam [118] points out the existence of interdependence between aspect and opinions (For example, the word '*low*' describes totally different opinion in '*low LCD resolution*' and '*low price*') will improve the extraction performance. Thus, Moghddam models this interdependence and proposes the ILDA which follows the assumption of bag of opinions phrases (e.g.$<$ *screen, bright*$>$, $<$ *battery life, long*$>$) to enhance the extraction ability. To combine the label information with the unsupervised extraction model, Brody [119] employs the seed set automatic construction and sentiment propagation mechanisms to identify the aspect based opinion. Furthermore, to capture the syntax and semantics relations between words, Lakkaraju [62] assumes that word category is conditioned on previous word and employs the hidden markov model to capture the short-range syntactic structure and long range semantic dependencies.

Differs from mining approaches above, some efforts have been made to model *aspect and opinion* pair jointly, and the extracted topics are in the form of '*Electronic reviews*'. Ekinci [69] incorporates a ranking mechanism into LDA and proposes an approach for aspect-sentiment pair extraction in Turkish language corpus. To use complementary information provided by different type online media, Zuo [71] propose a cross-collection topic modeling approach for aspect and opinion mining. In [113], Jo firstly proposes a SLDA which assumes that all words in a sentence are generated from a single aspect due to the observation that one sentence in a review tends to express opinion on one aspect. Moreover, Jo extends the SLDA and proposes the Aspect and Sentiment Unification Model (ASUM) which views one topic as an aspect coupled with an opinion (called senti-aspect topic). Likewise, ASUM also follows the constrain that all words in a specific sentence should be generated from one topic. With the extracted corpus-level senti-aspect topics, the topic distribution could be employed to predict the sentiment polarity of a review.

In most e-commerce sites, customers are asked to rate the pre-defined aspects. And this side information are helpful for aspect and opinion mining. Thus, instead of solely using the textual content [61], Titov [63] incorporates the rating information provided in review corpus and proposes a novel multi-aspect sentiment model (MAS) to extract the senti-aspect topics. Considering that matrix factorization is an effective tool for rating prediction and topic modeling is widely used for review processing. Xiao [70] combines the matrix factorization with topic modeling and proposes a model for aspect rating prediction.

Apart from aspect and opinion mining, some other work focus on the aspect ratings. Wang [102] observes that when writing a review the relative weight placed by a reviewer on each aspect could be useful for aspect rating (e.g. when booking a hotel, five stars on '*value*' and two stars on '*room*', and if the aspects weights are 0.5 and 0.2, then, the reviewer will care more about the price and give a relatively high overall rating). Thus, he exploits the relationship between aspect based ratings and the overall rating and develops a latent aspect rating analysis model. The proposed approach could not only extract the latent aspects and aspect specific opinions jointly but also provide the aspect weight for each user generated review without any supervision.

C. DOCUMENT/SENTENCE LEVEL ASPECT AND OPINION MINING

In last two subsections, previous work of corpus level aspect mining have been discussed. Most existing approaches tackle the tasks using linguistic analysis or unsupervised learning (e.g. topic models and clustering algorithms). Due to the arbitrariness of syntactic rules and the lack of necessary supervision in unsupervised model, those approaches have drawbacks such as aspect redundancy and poor performance. Thus, another group of researchers focus on the methodologies in document/sentence level mining and view this task as a sequential tagging problem. This subsection will survey the corresponding supervised methodologies.

1) PROBLEM SETTINGS

Document/sentence level aspect and opinion mining could be viewed as a sequential tagging problem. Like the part-of-speech tagging and the semantic role labeling, tagging approaches usually assign a tag from pre-defined tag set (e.g. [BOA, MOA, EOA, PO, NO, OT], first three tags mean begin/middle/end of the aspect, PO and NO represent positive opinion and negative opinion, OT means other token) for each token in document/sentence and then uses the corresponding tag sequence to extract the aspect and opinion expressions.

The formal definition of the task could be described as: Given the target product t_{st} and the target-related review collection C_t^{st}, how to extract the aspect and the opinion terms for each review $r_i^{t_{st}} \in C_t^{st}$. Here, we use '*iPhoneX*' as an example.

Example:

Target: iPhoneX

Review: iPhoneX has a great sound quality, I like it.

Outputs: <OT>iPhoneX</OT><OT>has</OT><OT>a </OT><PO>great</PO><BOA>sound</BOA> <EOA>quality</EOA>,<OT>I</OT><OT>like </OT><OT>it</OT>

Extracted triple of the review:

$<$ iPhoneX, sound quality, great$>$.

In general, traditional sequential models like HMMs [72], and CRFs [74]–[77] have been employed by community for document/sentence level aspect and opinion mining. Meanwhile, many prevalent techniques (e.g. word embeddings [99], representation learning [98] and etc) related to the deep learning [80]–[82], [84]–[86], [88]–[90], [138], [139], [144], [145] have also been explored for the task. And all the related approaches are presented in Table 8 together with their core techniques and performance. Detailed discussion of these approaches will be described below.

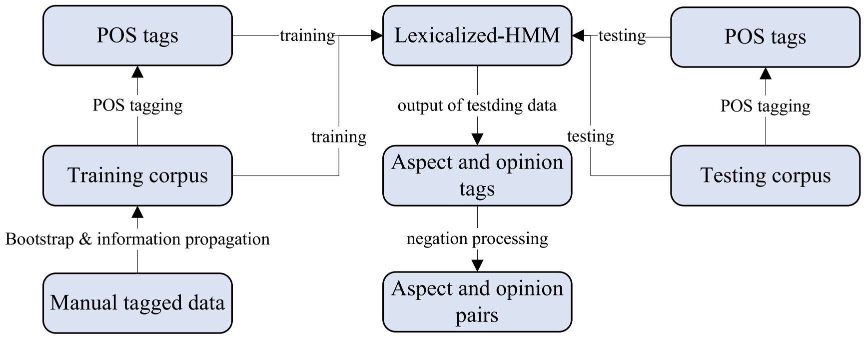


FIGURE 9. The general framework of Lexicalized HMM model [72].

TABLE 8. Approaches for document/sentence level aspect and opinion tagging.

	Core technique	Performance
Traditional Sequential Models:		
Jin [72]	Lexicalized HMM, POS tagging	Pre:80.3%, Rec:86.0%, F_{score}:82.7%
Li [74]	Skip-treeCRF, syntactic dependency	Pre:86.6%, Rec:76.2%, F_{score}:79.3%
Jakob [75]	CRF, dependency parser tree	Pre:79.2%, Rec:66.1%, F_{score}:70.2%
Wang [138]	CRF, lexical feature, dependency tree-RNN	F_{score}:78.4%-84.9%
Shu [76]	CRF, Lifelong learning	Pre:81.3%, Rec:76.0%, F_{score}:78.6%
Xiang [77]	CRF, Multi-Feature Embedding	Pre:87.11%, Rec:70.09%, F_{score}:80.43%
Deep Learning based Approaches:		
Poria [80]	CNNs, word embeddings, linguistic patterns	Pre:92.8%, Rec:88.3%, F_{score}:90.4%
Ye [139]	Dependency tree based CNN, linguistic feature, word embeddings	F_{score}:78.48%
Liu [84]	LSTM, word embeddings, linguistic feature	Pre:82.0%, Rec:75.54%, F_{score}:78.53%
Ding [144]	RNN, auxiliary label, word embedding	F_{score}:50.4%
Wang [145]	Memory network, word embeddings	F_{score}:67.29%
Wang [88]	GRU, coupled attention	F_{score}:74.48%
Xu [82]	CNN, general embedding, domain embedding	F_{score}:77.98%
Yu [89]	LSTM, Multi-task, integer linear programming	F_{score}:78.43%
Wang [85]	GRU, CRF, synthetic pattern	F_{score}:56.08%
Li [90]	LSTM, history attention	F_{score}:77.55%
Wang [86]	Tree-RNN, domain adaptation	F_{score}:51.49%

2) TRADITIONAL SEQUENTIAL MODEL BASED APPROACHES

Traditional sequential models are suitable for sequential labeling task. Thus, several published works have explored the field by incorporating task specific strategies with HMMs and CRFs [62], [72], [74]–[77], [138]. The $1st - 6th$ rows in Table 8 list the corresponding models.

Here, we use Jin's model [72] as an example to illustrate the extraction process of document/sentence level the aspect and opinion mining. To incorporate the extra linguistic lexical features into HMM, Jin [72] employs the POS tags and lexical patterns and proposes the Lexicalized HMM which is shown in Fig. 9. The objective of the proposed system could be described as: given a review (word sequence) $R = r_1, r_2, \ldots, r_n$ and the corresponding POS tag sequence $S = s_1, s_2, \ldots, s_n$, the task is to predict the suitable tag sequence $\hat{T} = t_1, t_2, \ldots, t_n$ that could maximize the conditional probability $P(T|R, S)$. To simplify the approach and make it computable, three assumptions have been made: i) the current tag t_i only depends on the previous tag t_{i-1} and the word w_{i-1}; ii) the probability of current word w_i only depends on the current POS tag s_i and the previous word w_{i-1}; iii) the probability of current POS s_i only depends on the current tag t_i and the previous word w_{i-1}. Based on these approximations, when given an annotated training corpus, model parameters could be estimated by maximum likelihood estimation. To further reduce the human labor in training corpus construction phase, Jin employs a bootstrap program with information propagation mechanism to automatically generate the annotated reviews by using a small annotated corpus as seed corpus. Finally, the system employs the viterbi algorithm to generate the optimal tag sequence $\hat{T}$ for each test review. For example, the sentence "*I love the ease of transferring the pictures* *to my computer.*” could be tagged as “< BG> *I*< /BG> < OPINION_POS_EXP> *love*</OPINION_POS_EXP> < BG> *the*< /BG> < PROD_FEAT>*ease of transferring the pictures*< /PROD_FEAT> < BG> *to*< /BG> < BG> *my* < /BG> < BG > *computer*< /BG>.” Thus, the aspect ‘*ease of transferring the picture*’ and opinion ‘*love*’ could be extracted.

On the other hand, CRF-based approaches have also been proposed. Due to the simplicity, the linear chain CRF is commonly used. However, the relation between neighbor context will not be sufficient for the sophisticated aspect and opinion mining task. Thus, Skip-chain CRFs [74] is employed for the task, and it could model the long distance dependency between conjunctions (e.g. only consider ‘*adjective*’, ‘*noun*’ and ‘*verb*’). For example, in the sentence “*iPhoneX has a great camera and a cool appearance”*, two long distance dependencies (*dep(great, cool) and dep(camera, appearance)*) could be captured by Skip-chain CRFs. Furthermore, Li also proposes the Skip-tree CRFs [74] which incorporates the syntactic tree structure into the CRF framework and outperforms traditional CRF. And the proposed Skip-tree CRF has good expansibility that external features (e.g. semantic dependency tree and sentiment lexicons) could be naturally encoded. Likewise, Jakob [75] proposes a CRF-based approach with various features (e.g. token, part-of-speech, short dependency path, word distance and etc) to perform opinion target extraction task in both single-domain and cross-domain setting. Moreover, scholars claim that if the system has performed aspect extraction from many past domain and stores the result as knowledge it will be helpful for improving the extraction performance on other domain corpus. Based on the assumption, Shu [76] incorporates the lifelong learning into CRF and the proposed approach performs markedly better than the traditional CRF. To capture the semantic relatedness from multiple source embedding, Xiang [77] employs multi-feature embedding as additional position feature to train a CRF-based tagger, and the MFE-CRF also outperforms the traditional CRF.

Besides, a novel attempt of combining the traditional CRF with deep neural network has been proposed in [138], named Recursive Neural CRFs. The RNCRF contains a dependency-tree RNN and a Condition Random Field as output layer. Combining the superiority of representative presentation learnt by neural network and the CRF, the proposed RNCRF outperforms the traditional CRFs and needs less human intervention for designing appropriate features.

3) DEEP LEARNING BASED APPROACHES

Despite the success of traditional sequential models for document level aspect and opinion mining, they could be easily affected by the selection of features and external relations.

Thus, many researchers expect that neural-based model could provide a more suitable way for the task. Fortunately, deep learning approaches [80]–[82], [84]–[86], [88]–[90], [139], [144], [145] (e.g. CNN, RNN and LSTM) have also contributed to the advances in document level aspect and opinion mining task. Due to the strong representation ability of deep learning and related techniques (e.g. word embeddings), a mass of deep learning based approaches employ word embedding as the text feature to capture the semantic meaning of token in this specific task. As summarized in Table 8, last twelve rows list the deep learning based approaches for aspect and opinion tagging. The detailed discussion will be given in this subsection.

Recurrent Neural Networks (RNN) and Long-Short-Term Memory (LSTM) which could tackle variable length input sequence have shown great promise in many sequential labeling tasks. To illustrate the tagging process of the deep learning based approach, we use the model proposed in [84] as example. And Fig. 10 depicts the structure of the devised network. For encoding the semantic relatedness of tokens, Liu trains a task specific embeddings using CBOW model [98] on an amazon review corpus which contains 34M reviews. And the proposed approach also employs a window-based mechanism to capture the local dependency at the input layer. For example, if the window size is set to 3, the ‘*disk*’ shown in Fig. 10 should be represented by the vector which is the concatenation of the embeddings corresponding to ‘*hard*’, ‘*disk*’, ‘*is*’. Besides, Liu finds that the future information may be crucial for predicting the tag current word(e.g. when given the bi-gram ‘*hard disk*’, the observation of ‘*disk*’ will be helpful for assigning a ‘*B-TARG*’ for ‘*hard*’). To capture the long distance dependencies of two directions, both forward and backward, in text sequence, the proposed approach employs a bi-directional LSTM to learn the high level distributed representation of the input word sequence and outperforms the CRF-based approaches which employs sophisticated hand-crafted feature. Furthermore, it has a strong expandability and the linguistic features (e.g. part-of-speech tags) could be incorporated into the model by concatenating it with the output of LSTM unit (shown in Fig. 10, and f_i means the POS tag of x_i and the learned weights are helpful for tagging.

Likewise, other variant RNN-based tagging approaches have been proposed [88]–[90]. As we know, the syntactic relation between aspect terms and opinion terms is useful for extracting aspect and opinion. To model such relations, Wang [88] devises a multi-layer attention based network with coupled attention, one attention is for extracting aspect terms and the other is for opinion terms. And the proposed model could further exploit indirect relations between terms for more precise information extraction through multiple layers attention mechanism. Yu [89] views the aspect and opinion extraction in a multi-task manner. To directly model the inter-task constrain, intra-task constrain and lexicon constrain, Yu formulate the problem as an Integer Liner Programming and devises a global inference approach to extraction aspect and opinion terms jointly. Li [90] exploit two useful clues (opinion summary and aspect detection history) and proposes an approach based on the history attention.

On the other hand, scholars have also proposed several CNN-based approaches to extract aspect and opinion terms in document level. Poria [80] proposes a CNN based approach

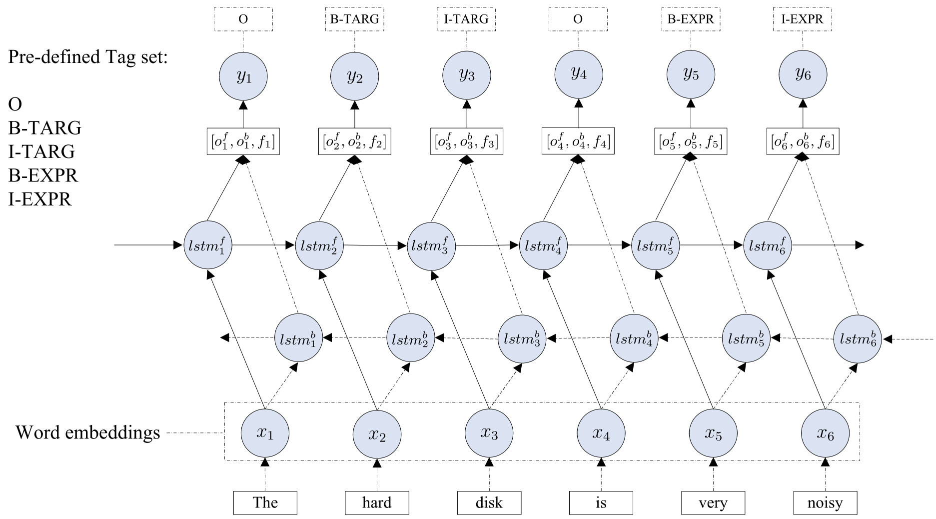


FIGURE 10. The bidirectional LSTM based approach for aspect and opinion tagging [84].

which incorporates linguistic patterns into neural network. Firstly, the approach uses a 7-layer deep CNN which combines with word embeddings and window mechanism to label each term in review with task specific tags. Then, five linguistic patterns are also employed to tag aspect term. Finally, the model uses a voting mechanism which considers the tags generated by CNN and patterns to judge if a term is an aspect or not. To capture the dependency information (which is critical and has been widely used in the field) in reviews, a dependency-tree based convolutional stacked neural network is proposed in [139]. The proposed approach has strong representation ability and could effectively exploit the dependency relations resort to its elaborated structure. Furthermore, Xu [82] claims that aspect extraction is a complex task that also requires fine-grained domain knowledge. To make full use of domain knowledge, Xu employs two form of embedding (general-purpose embeddings and domain-specific embeddings) to capture more useful information and proposes the DE-CNN for aspect extraction.

Besides, another group of approaches which focus on the cross-domain and cross-lingual scenarios have also been proposed. Inspired by the idea of learning a suitable representation for both source and target domains could be helpful for domain adaptation, Ding [144] firstly uses the auxiliary label sequences generated by syntactic rules to train a hidden recurrent neural network using both domain corpus. Then, the hidden layer is trained on source domain corpus using aspect and opinion annotations as supervision. By taking advantage of neural based supervised models and relation-based rules, the proposed approach achieves F_{score} : 50.2% which outperforms the traditional sequential models. Likewise, Wang [85] employs a recursive neural structural network to capture the syntactic relation between words, and the proposed SNSCN-GRU could efficiently reduce the domain shift and obtains a good performance in cross-domain extraction. To tackle the multi-lingual scenarios, Wang [86] utilizes transition-based mechanism that reads a word each time and forms a series of configurations (represented as a continuous feature vector) that represent the status of the whole sentence. And the configurations from different languages are aligned into a shared space through an adversarial network to transfer the knowledge to target language. The model, proposed in [86], obtains F_{score} : 51.49%.

Although it's not fair to directly compare the performance of the models in Table.8 due to the difference of the used datasets, we also have some findings. i). Models that focus on cross-domain and cross-lingual scenarios often obtain worse results compared with the in-domain tagging task, this is caused by the knowledge gap between different domains and languages; ii). Deep learning based tagging models that employ attention mechanism (such as history attention and coupled attention) often get better performance. This may because that attention mechanism could let the model care more about aspect and opinion terms which could further improve the extraction performance. iii). There is another trend that scholars prefer to combine traditional sequential models (such as condition random field) and syntactic patterns with neural based models. As deep neural networks have strong representative ability and traditional models could provide handcrafted features and syntactic features, this combination will make the model benefit from both strategies and promote the performance. iv). Word embedding is a critical technique for neural based extraction models. Under the help of word relatedness which retained in word embedding, neural based approaches perform better than traditional models.

V. CHALLENGES AND POSSIBLE SOLUTIONS

The increasing usage of online activities and social media platforms lead to a large expansion of unstructured data (e.g. user generated reviews). These emotional reviews or posts could be used for mining the opinions of the general public and consumers on social events, political movements and product preferences. Moreover, the mentioned tasks about opinion mining will benefit citizens, consumers, manufacturers and the government. To accomplish these tasks and make full use of online resources, researchers have proposed vast stance detection systems and product aspect mining approaches.

Early work mainly focus on frequency, relation rules and feature-based methods, the researchers endeavor to seek effective features and patterns for stance detection and product aspect mining. Later, computational linguistic models and statistical machine learning techniques are explored for the tasks and promote the performance. In last five years, scholars in opinion mining field turn their orientation to representation learning and neural based approaches to seek the breakthrough for stance detection and aspect based mining tasks. The reviewed work, in section 3 and section 4, also illustrates the trends of the domain and the increasing complexity of the models. However, there still are several obstacles in this area to overcome as listed below.

A. LACK OF STANCE LEXICON

Due to the similarity of stance detection and sentiment analysis. Previous work of stance detection usually employs the sentiment lexicons(e.g. Bing Liu's Lexicon [51], SentiWordNet [106], NRC-emotion Lexicon [108] and etc) to extract senti-feature. And the detection performance could be slightly improved by feeding the senti-feature into the classifiers. However, stance detection aims to identify the stance of each reviews on the specific target (e.g. "*The pregnant are more than walking incubators.*" express the favor stance for the target '*legalization of abortion*'). Thus, the existing lexicons which just collect positive words and negative words are not suitable for the task. And the stance lexicon which could capture the relation between target (e.g. *nouns and noun phrases*) and opinion (e.g. *positive words and negative words*) is urgently needed to further improve the performance of stance detection.

B. LACK OF LARGE SCALE ANNOTATED CORPUS FOR STANCE DETECTION

The stance detection, both in online debate forums and on social media platforms, could be viewed as a classification problem. And the annotated corpus, used in reviewed works, commonly contains hundreds reviews. For example, the tweets corpus, built by Mohammad [116], contains only 4870 tweets for six target (e.g. '*Atheism#733*', '*Climate Change Concern# 564*', '*Donald Trump# 707*', '*Feminist Movement# 949*', '*Hillary Clinton# 984*', '*Legalization of Abortion# 933*'). The limitation of corpus scale impedes the representation learning of task specific tweets and it also obstructs the construction of task specific tools, such as domain-specific word embedding. Thus, it is necessary to build a large scale annotated corpus to promote the development of stance detection systems and the corresponding techniques.

C. STRUCTURED ASPECT MINING

Previous product aspect extraction approaches (both rule-based and topic models based approaches) view the aspects are independent and omit the inter-relationship between extracted aspects. In other words, these approaches could only extract the flatten aspects and could not capture the aspect structure for the specific product. While in practice, the aspects of a specific product may follow a tree structure. (e.g. '*screen*' is an aspect of '*iPhoneX*', and '*screen size*', '*screen resolution*' and '*screen saturation*' are three sub-aspects of '*screen*'.) To handle this challenge, approaches like Self Organization Mapping [150] (SOM), hierarchical Latent Dirichlet Allocation(hLDA) [151], [152] and Latent Tree Analysis(LTA/LTM) [142], [153], [154] should be proposed to capture the structured aspects tree from product specific review collection.

D. INCORPORATE THE EXTERNAL KNOWLEDGE INTO THE MODEL

The reviewed works in section 3 grows from the simple rule and relation based pattern matcher to sophisticated, hybrid models which employ machine learning and statistic learning (e.g. Topic Models and Deep Neural Networks). Likewise, many aspect mining approaches extract the aspects and aspect-specific opinions using unsupervised framework. However, many easy obtained external information such as POS tagging, syntactic dependency tree and semantic dependency tree will helpful for the task. Thus, how to incorporate the syntactic and semantic level information into the mining approaches is another challenge. To tackle the challenge, the MaxEnt-LDA [120] provides a solution to separate the aspect word and opinion word using POS tags and Max-Entropy classifier. In future, other attempts should be explored to encode the external knowledge into topic models and deep learning approaches to enhance the performance of product aspect mining.

VI. CONCLUSION

From the overview of the state-of-the-art in stance detection and product aspect mining in this survey, it is clear that the field is transcending its early stage. We could observe another fact that the mining methodologies have evolved through time together with the increasing model complexity. Besides, we want to stress that transparency and standardization is needed in terms of evaluation methodology and datasets in order to draw firm conclusions about the current state-of-the-art. Benchmark initiatives like Sem-Eval provide a way to solve this problem.

Considering the future of stance detection, we foresee a move from relation rule and syntactic pattern based approaches towards target related neural based models, especially target related attention based networks. For example, in "*Religion has destroyed the ability for some to say* *know", (target: Atheism)*, it may be assigned an against stance by relation and pattern based approaches due to the negative sentiment of the content. However, it conveys an favor stance to *'atheism'* target. In [48], a target based attention model is presented that explicitly models the relationship between target and text content will be helpful for stance detection. Likewise, the methodologies of product aspect mining also have a trend which evolves from the traditional frequency and relation based approaches to the knowledge based topic models and neural based models. For example, [120] incorporates the pos tags into topic modeling and presents a MaxEnt-LDA which improves the extraction performance, and [82] uses two different word embedding to help with aspect and opinion term extraction. While knowledge based approaches have only recently begun to emerge and the knowledge used are outdated, recently proposed representation techniques like ELMo [155] and Bert [156] might also help mining approaches to enhance the extraction ability.

Combining the external knowledge with the power of machine learning will give rise to models which are able to reason with expert instruction, and it could reduce the influence of insufficient data. The knowledge based approaches will allow future applications to deal with complex language structures and to employ the available man-made knowledge bases. Moreover, this will enable many application domains to benefit from the knowledge obtained from stance detection and product aspect mining.

ACKNOWLEDGMENT

We would like to thank the reviewers for their valuable comments and helpful suggestions.

REFERENCES

[1] B. Liu, "Sentiment analysis and opinion mining," *Synth. Lectures Hum. Lang. Technol.*, vol. 5, no. 1, pp. 1–167, 2012.
[2] K. S. Hasan and V. Ng, "Extra-linguistic constraints on stance recognition in ideological debates," in *Proc. 51st Annu. Meet. Assoc. Comput. Linguistics*, vol. 2, 2013, pp. 816–821.
[3] S. Moghaddam and M. Ester, "Aspect-based opinion mining from product reviews," in *Proc. 35th Int. ACM SIGIR Conf. Res. Develop. Inf. Retr.*, 2012, p. 1184.
[4] B. Liu and L. Zhang, "A survey of opinion mining and sentiment analysis," in *Mining Text Data*. Boston, MA, USA: Springer, 2012, pp. 415–463.
[5] W. Medhat, A. Hassan, and H. Korashy, "Sentiment analysis algorithms and applications: A survey," *Ain Shams Eng. J.*, vol. 5, no. 4, pp. 1093–1113, 2014.
[6] M. Bansal, C. Cardie,and L. Lee, "The power of negative thinking: Exploiting label disagreement in the min-cut classification framework," in *Proc. COLING*, 2008, pp. 15–18.
[7] A. Balahur, Z. Kozareva, and A. Montoyo, "Determining the polarity and source of opinions expressed in political debates," in *Proc. 10th Int. Conf. Comput. Linguistics Intell. Text Process.*, 2009, pp. 468–480.
[8] O. Biran and O. Rambow, "Identifying justifications in written dialogs," in *Proc. IEEE 5th Int. Conf. Semantic Comput.*, Sep. 2011, pp. 162–168.
[9] M. Thomas, B. Pang, and L. Lee, "Get out the vote: Determining support or opposition from congressional floor-debate transcripts," in *Proc. Conf. Empirical Methods Natural Lang. Process.*, 2006, pp. 327–335.
[10] A. Murakami and R. Raymond, "Support or oppose?: Classifying positions in online debates from reply activities and opinion expressions," in *Proc. 23rd Int. Conf. Comput. Linguistics: Posters*, 2010, pp. 869–875.
[11] R. Agrawal, S. Rajagopalan, R. Srikant, and Y. Xu, "Mining newsgroups using networks arising from social behavior," in *Proc. 12th Int. Conf. World Wide Web*, 2003, pp. 529–535.
[12] P. Anand, M. Walker, R. Abbott, J. E. F. Tree, R. Bowmani, and M. Minor, "Cats rule and dogs drool!: Classifying stance in online debate," in *Proc. 2nd Workshop Comput. Approaches Subjectivity Sentiment Anal.*, 2011, pp. 1–9.
[13] M. A. Walker, P. Anand, R. Abbott, J. E. F. Tree, C. Martell, and J. King, "That is your evidence?: Classifying stance in online political debate," *Decis. Support Syst.*, vol. 53, no. 4, pp. 719–729, Nov. 2012.
[14] M. A. Walker, P. Anand, R. Abbott, and R. Grant, "Stance classification using dialogic properties of persuasion," in *Proc. Conf. North Amer. Chapter Assoc. Comput. Linguistics, Hum. Lang. Technol. Assoc. Comput. Linguistics*, 2012, pp. 592–596.
[15] K. S. Hasan and V. Ng, "Frame semantics for stance classification," in *Proc. 7th Conf. Comput. Natural Lang. Learn. (CoNLL)*, 2013, pp. 124–132.
[16] K. S. Hasan and V. Ng, "Stance classification of ideological debates: Data, models, features, and constraints," in *Proc. 6th Int. Joint Conf. Natural Lang. Process. (IJCNLP)*, 2013, pp. 1348–1356.
[17] D. Sridhar, L. Getoor, and M. Walker, "Collective stance classification of posts in online debate forums," in *Proc. Joint Workshop Social Dyn. Pers. Attributes Social Media*, 2014, pp. 109–117.
[18] H. Elfardy, M. T. Diab, and C. Callison-Burch, "Ideological perspective detection using semantic features," in *Proc. 4th Joint Conf. Lexical Comput. Semantics*, 2015, pp. 137–146.
[19] D. Sridhar, J. R. Foulds, B. Huang, L. Getoor, and M. A. Walker, "Joint models of disagreement and stance in online debate," in *Proc. 53rd Annu. Meet. Assoc. Comput. Linguistics*, 2015, pp. 116–125.
[20] M. Qiu, Y. Sim, N. A. Smith, and J. Jiang, "Modeling user arguments, interactions, and attributes for stance prediction in online debate forums," in *Proc. SIAM Int. Conf. Data Mining*, 2015, pp. 855–863.
[21] L. Li, Z. Wu, M. Xu, H. Meng, and L. Cai, "Recognizing stances in mandarin social ideological debates with text and acoustic features," in *Proc. IEEE Int. Conf. Multimedia Expo Workshops (ICMEW)*, Jul. 2016, pp. 1–6.
[22] S. Somasundaran and J. Wiebe, "Recognizing stances in ideological on-line debates," in *Proc. NAACL HLT Workshop Comput. Approaches Anal. Gener. Emotion Text*, 2010, pp. 116–124.
[23] T. Kyaw and S. S. Aung, "Stance mining for online debate posts using part-of-speech (pos) tags frequency," in *Proc. IEEE 16th Int. Conf. Softw. Eng. Res. Manage. Appl.(SERA)*, Jun. 2018, pp. 102–107.
[24] S. Ghosh, K. Anand, S. Rajanala, A. B. Reddy, and M. Singh, "Unsupervised stance classification in online debates," in *Proc. ACM India Joint Int. Conf. Data Sci. Manage. Data*, 2018, pp. 30–36.
[25] C. Li, A. Porco and D. Goldwasser, "Structured representation learning for online debate stance prediction," in *Proc. 27th Int. Conf. Comput. Linguistics*, 2018, pp. 3728–3739.
[26] A. Trabelsi and O. R. Zaïane, "Unsupervised model for topic viewpoint discovery in online debates leveraging author interactions," in *Proc. 12th Int. AAAI Conf. Web Social Media*, 2018, pp. 425–433.
[27] R. Dong, Y. Sun, L. Wang, Y. Gu, and Y. Zhong, "Weakly-guided user stance prediction via joint modeling of content and social interaction," in *Proc. ACM Conf. Inf. Knowl. Manage.*, 2017, pp. 1249–1258.
[28] Q. Sun, Z. Wang, Q. Zhu, and G. Zhou, "Exploring various linguistic features for stance detection," in *Proc. Int. Conf. Comput. Process. Oriental Lang.* Cham, Switzerland: Springer, 2016, pp. 840–847.
[29] M. Lukasik, K. Bontcheva, T. Cohn, A. Zubiaga, M. Liakata, and R. Procter. (Sep. 7, 2016). "Using Gaussian processes for rumour stance classification in social media." [Online]. Available: https://arxiv.org/abs/1609.01962
[30] J. Ebrahimi, D. Dou, and D. Lowd, "Weakly supervised Tweet stance classification by relational bootstrapping," in *Proc. Conf. Empirical Methods Natural Lang. Process. (EMNLP)*, 2016, pp. 1012–1017.
[31] B. G. Patra, D. Das, and S. Bandyopadhyay, "JU_NLP at semeval-2016 task 6: Detecting stance in Tweets using support vector machines," in *Proc. SemEval@ NAACL-HLT*, 2016, pp. 440–444.
[32] M. Wojatzki and T. Zesch, "ltl.uni-due at SemEval-2016 task 6: Stance detection in social media using stacked classifiers," in *Proc. 10th Int. Workshop Semantic Eval. (SemEval)*, 2016, pp. 428–433.
[33] W. Wei, X. Zhang, X. Liu, W. Chen, and T. Wang, "Pkudblab at SemEval-2016 task 6 : A specific convolutional neural network system for effective stance detection," in *Proc. 10th Int. Workshop Semantic Eval. (SemEval)*, 2016, pp. 384–388.
[34] H. Elfardy and M. T. Diab, "CU-GWU perspective at SemEval-2016 task 6: Ideological stance detection in informal text," in *Proc. 10th Int. Workshop Semantic Eval. (SemEval)*, 2016, pp. 434–439.

[35] A. Misra, B. Ecker, T. Handleman, N. Hahn, and M. A. Walker, "NLDS-UCSC at SemEval-2016 task 6: A semi-supervised approach to detecting stance in Tweets," in *Proc. 10th Int. Workshop Semantic Eval. (SemEval)*, 2016, pp. 420–427.

[36] M. Tutek *et al.*, "TakeLab at SemEval-2016 task 6: Stance classification in Tweets using a genetic algorithm based ensemble," in *Proc. 10th Int. Workshop Semantic Eval. (SemEval)*, 2016, pp. 464–468.

[37] Y. Igarashi, H. Komatsu, S. Kobayashi, N. Okazaki, and K. Inui, "Tohoku at SemEval-2016 task 6: Feature-based model versus convolutional neural network for stance detection," in *Proc. 10th Int. Workshop Semantic Eval. (SemEval)*, 2016, pp. 401–407.

[38] M. Dias and K. Becker, "INF-UFRGS-OPINION-MINING at SemEval-2016 task 6: Automatic generation of a training corpus for unsupervised identification of stance," in *Proc. 10th Int. Workshop Semantic Eval.*, 2016, pp. 378–383.

[39] I. Augenstein, A. Vlachos, and K. Bontcheva, "USFD at SemEval-2016 task 6: Any-target stance detection on Twitter with autoencoders," in *Proc. 10th Int. Workshop Semantic Eval. (SemEval)*, 2016, pp. 389–393.

[40] P. Vijayaraghavan, I. Sysoev, S. Vosoughi, and D. Roy. (Jun. 17, 2016). "DeepStance at SemEval-2016 task 6: Detecting stance in Tweets using character and word-level CNNs." [Online]. Available: https://arxiv.org/abs/1606.05694

[41] G. Zarrella and A. Marsh. (Jun. 13, 2016). "MITRE at SemEval-2016 Task 6: Transfer learning for stance detection." [Online]. Available: https://arxiv.org/abs/1606.03784

[42] P. Wei, W. Mao, and D. Zeng, "A target-guided neural memory model for stance detection in Twitter," in *Proc. Int. Joint Conf. Neural Netw. (IJCNN)*, 2018, pp. 1–8.

[43] G. Gadek, J. Betsholtz, A. Pauchet, S. Brunessaux, N. Malandain, and L. Vercouter, "Extracting contextonyms from Twitter for stance detection," in *Proc. 9th Int. Conf. Agents Artif. Intell.*, 2017, pp. 132–141.

[44] A. Benton and M. Dredze, "Using author embeddings to improve Tweet stance classification," in *Proc. EMNLP Workshop W-NUT, 4th Workshop Noisy User-Generated Text*, 2018, pp. 184–194.

[45] Q. Sun, Z. Wang, Q. Zhu, and G. Zhou, "Stance detection with hierarchical attention network," in *Proc. 27th Int. Conf. Comput. Linguistics*, 2018, pp. 2399–2409.

[46] Y. Zhou, A. I. Cristea, and L. Shi, "Connecting targets to Tweets: Semantic attention-based model for target-specific stance detection," in *Proc. Int. Conf. Web Inf. Syst. Eng.* Cham, Switzerland: Springer, 2017, pp. 18–32.

[47] P. Wei, J. Lin, and W. Mao, "Multi-target stance detection via a dynamic memory-augmented network," in *Proc. 41st Int. ACM SIGIR Conf. Res. Develop. Inf. Retr.*, 2018, pp. 1229–1232.

[48] J. Du, R. Xu, Y. He, and L. Gui, "Stance classification with target-specific neural attention networks," in *Proc. Int. Joint Conf. Artif. Intell.*, 2017, pp. 3988–3994.

[49] K. Dey, R. Shrivastava, and S. Kaushik, "Twitter stance detection—A subjectivity and sentiment polarity inspired two-phase approach," in *Proc. IEEE Int. Conf. Data Mining Workshops (ICDMW)*, Nov. 2017, pp. 365–372.

[50] C. Xu, C. Paris, S. Nepal, and R. Sparks. (May 17, 2018). "Cross-target stance classification with self-attention networks." [Online]. Available: https://arxiv.org/abs/1805.06593

[51] M. Hu and B. Liu, "Mining and summarizing customer reviews," in *Proc. 10th ACM SIGKDD Int. Conf. Knowl. Discovery Data Mining*, 2004, pp. 168–177.

[52] J. Han, J. Pei, and Y. Yin, "Mining frequent patterns without candidate generation," *ACM SIGMOD Rec.*, vol. 29, pp. 1–12, Jun. 2000.

[53] Y. Zhao, B. Qin, S. Hu, and T. Liu, "Generalizing syntactic structures for product attribute candidate extraction," in *Proc. Hum. Lang. Technol., Annu. Conf. North Amer. Chapter Assoc. Comput. Linguistics*, 2010, pp. 377–380.

[54] Z. Hai, K. Chang, and J.-J. Kim, "Implicit feature identification via co-occurrence association rule mining," in *Proc. Comput. Linguistics .Intell. Text Process.*, 2011, pp. 393–404.

[55] B. Liu, M. Hu, and J. Cheng, "Opinion observer: Analyzing and comparing opinions on the web," in *Proc. 14th Int. Conf. World Wide Web*, 2005, pp. 342–351.

[56] A. Konjengbam, N. Dewangan, and N. Kumar, M. Singh, "Aspect ontology based review exploration," *Electron. Commerce Res. Appl.*, vol. 30, pp. 62–71, Jul./Aug. 2018.

[57] A.-D. Vo, Q.-P. Nguyen, and C.-Y. Ock, "Opinion–aspect relations in cognizing customer feelings via reviews," *IEEE Access*, vol. 6, pp. 5415–5426, 2018.

[58] S. J. Das and B. Chakraborty, "Aspect aware optimized opinion analysis of online product reviews," in *Proc. 9th Int. Conf. Awareness Sci. Technol. (iCAST)*, Sep. 2018, pp. 144–149.

[59] A. S. Shafie, N. M. Sharef, M. A. A. Murad, and A. Azman, "Aspect extraction performance with pos tag pattern of dependency relation in aspect-based sentiment analysis," in *Proc. 4th Int. Conf. Inf. Retr. Knowl. Manage. (CAMP)*, Mar. 2018, pp. 1–6.

[60] K. Srividya, K. Mariyababu, and A. M. Sowjanya, "Mining interesting aspects of a product using aspect-based opinion mining from product reviews," *Int. J. Eng., Trans. B, Appl.*, vol. 30, no. 11, pp. 1707–1713, 2017.

[61] I. Titov and R. McDonald, "Modeling online reviews with multi-grain topic models," in *Proc. 17th Int. Conf. World Wide Web*, 2008, pp. 111–120.

[62] H. Lakkaraju, C. Bhattacharyya, I. Bhattacharya, and S. Merugu, "Exploiting coherence for the simultaneous discovery of latent facets and associated sentiments," in *Proc. SIAM Int. Conf. Data Mining*, 2011, pp. 498–509.

[63] I. Titov and R. McDonald, "A joint model of text and aspect ratings for sentiment summarization," *Proc. ACL, HLT*, vol. 8, pp. 308–316, Jun. 2008.

[64] R. He, W. S. Lee, H. T. Ng, and D. Dahlmeier, "An unsupervised neural attention model for aspect extraction," in *Proc. 55th Annu. Meet. Assoc. Comput. Linguistics*, vol. 1, 2017, pp. 388–397.

[65] Y. Yang, C. Chen, M. Qiu, and F. Bao, "Aspect extraction from product reviews using category hierarchy information," in *Proc. 15th Conf. Eur. Chapter Assoc. Comput. Linguistics*, vol. 2, Short Papers, 2017, pp. 675–680.

[66] M. Shams and A. Baraani-Dastjerdi, "Enriched LDA (ELDA): Combination of latent dirichlet allocation with word co-occurrence analysis for aspect extraction," *Expert Syst. Appl.*, vol. 80, pp. 136–146, Sep. 2017.

[67] R. Chen, Y. Zheng, W. Xu, M. Liu, and J. Wang, "Secondhand seller reputation in online markets: A text analytics framework," *Decision Support Syst.*, vol. 108, pp. 96–106, Apr. 2018.

[68] S. Angelidis and M. Lapata. (Aug. 27, 2018). "Summarizing opinions: Aspect extraction meets sentiment prediction and they are both weakly supervised." [Online]. Available: https://arxiv.org/abs/1808.08858

[69] E. Ekinci and S. I. Omurca, "An aspect-sentiment pair extraction approach based on latent dirichlet allocation," *Int. J. Intell. Syst. Appl. Eng.*, vol. 6, no. 3, pp. 209–213, 2018.

[70] D. Xiao, Y. Ji, Y. Li, F. Zhuang, and C. Shi, "Coupled matrix factorization and topic modeling for aspect mining," *Inf. Process. Manage.*, vol. 54, no. 6, pp. 861–873, 2018.

[71] Y. Zuo, J. Wu, H. Zhang, D. Wang, and K. Xu, "Complementary aspect-based opinion mining," *IEEE Trans. Knowl. Data Eng.*, vol. 30, no. 2, pp. 249–262, Feb. 2018.

[72] W. Jin, H. H. Ho, and R. K. Srihari, "Opinionminer: A novel machine learning system for web opinion mining and extraction," in *Proc. 15th ACM SIGKDD Int. Conf. Knowl. Discovery Data Mining*, 2009, pp. 1195–1204.

[73] B. Yang and C. Cardie, "Extracting opinion expressions with semi-Markov conditional random fields," in *Proc. Joint Conf. Empirical Methods Natural Lang. Process. Comput. Natural Lang. Learn.*, 2012, pp. 1335–1345.

[74] F. Li *et al.*, "Structure-aware review mining and summarization," in *Proc. 23rd Int. Conf. Comput. Linguistics*, 2010, pp. 653–661.

[75] N. Jakob and I. Gurevych, "Extracting opinion targets in a single- and cross-domain setting with conditional random fields," in *Proc. Conf. Empirical Methods Natural Lang. Process.*, 2010, pp. 1035–1045.

[76] L. Shu, H. Xu, and B. Liu. (Apr. 29, 2017). "Lifelong learning CRF for supervised aspect extraction." [Online]. Available: https://arxiv.org/abs/1705.00251

[77] Y. Xiang, H. He, and J. Zheng, "Aspect term extraction based on MFE-CRF," *Inf.*, vol. 9, no. 8, p. 198, 2018.

[78] Y. Ji, J. Li, and Y. Yu, "Linguistic attention-based model for aspect extraction," in *Proc. Int. Conf. Image Video Process., Artif. Intell.*, vol. 10836, 2018, p. 1083619.

[79] N. Kalchbrenner, E. Grefenstette, and P. Blunsom. (Apr. 8, 2014). "A convolutional neural network for modelling sentences." [Online]. Available: https://arxiv.org/abs/1404.2188

[80] S. Poria, E. Cambria, and A. Gelbukh, "Aspect extraction for opinion mining with a deep convolutional neural network," *Knowl.-Based Syst.*, vol. 108, pp. 42–49, Sep. 2016.

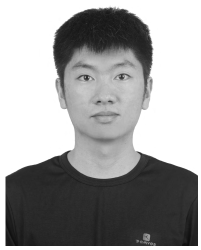
[81] H. Wu, Y. Gu, S. Sun, and X. Gu, "Aspect-based opinion summarization with convolutional neural networks," in *Proc. Int. Joint Conf. Neural Netw. (IJCNN)*, Jul. 2016, pp. 3157–3163.

[82] H. Xu, B. Liu, L. Shu, and P. S. Yu. (May 11, 2018). "Double embeddings and CNN-based sequence labeling for aspect extraction." [Online]. Available: https://arxiv.org/abs/1805.04601
[83] O. Irsoy and C. Cardie, "Opinion mining with deep recurrent neural networks," in *Proc. Conf. Empirical Methods Natural Lang. Process.*, 2014, pp. 720–728.
[84] P. Liu, S. R. Joty, and H. M. Meng, "Fine-grained opinion mining with recurrent neural networks and word embeddings," in *Proc. Conf. Empirical Methods Natural Lang. Process.*, 2015, pp. 1433–1443.
[85] W. Wang and S. J. Pan, "Recursive neural structural correspondence network for cross-domain aspect and opinion co-extraction," in *Proc. 56th Annu. Meet. Assoc. Comput. Linguistics*, vol. 1, 2018, pp. 2171–2181.
[86] W. Wang and S. J. Pan, "Transition-based adversarial network for cross-lingual aspect extraction," in *Proc. 27th Int. Joint Conf. Artif. Intell.*, 2018, pp. 4475–4481.
[87] S. Hochreiter and J. Schmidhuber, "Long short-term memory," *Neural Comput.*, vol. 9, no. 8, pp. 1735–1780, 1997.
[88] W. Wang, S. J. Pan, D. Dahlmeier, and X. Xiao, "Coupled multi-layer attentions for co-extraction of aspect and opinion terms," in *Proc. 31st AAAI Conf. Artif. Intell.*, 2017, pp. 3316–3322.
[89] J. Yu, J. Jiang, and R. Xia, "Global inference for aspect and opinion terms co-extraction based on multi-task neural networks," *IEEE/ACM Trans. Audio, Speech, Lang. Process.*, vol. 27, no. 1, pp. 168–177, Jan. 2019.
[90] X. Li, L. Bing, P. Li, W. Lam, and Z. Yang. (May 2, 2018). "Aspect term extraction with history attention and selective transformation." [Online]. Available: https://arxiv.org/abs/1805.00760
[91] Y.-M. Li and T.-Y. Li, "Deriving market intelligence from microblogs," *Decision Support Syst.*, vol. 55, no. 1, pp. 206–217, 2013.
[92] H. Rui, Y. Liu, and A. Whinston, "Whose and what chatter matters? The effect of Tweets on movie sales," *Decision Support Syst.*, vol. 55, no. 4, pp. 863–870, 2013.
[93] D. Kang and Y. Park, "Review-based measurement of customer satisfaction in mobile service: Sentiment analysis and VIKOR approach," *Expert Syst. Appl.*, vol. 41, no. 4, pp. 1041–1050, 2014.
[94] S. M. Mohammad, P. Sobhani and S. Kiritchenko, "Stance and sentiment in Tweets," *ACM Trans. Internet Technol.*, vol. 17, no. 3, p. 6, 2017.
[95] B. Pang and L. Lee, "Opinion mining and sentiment analysis," *Found. Trends Inf. Retr.*, vol. 2, nos. 1–2, pp. 1–135, 2008.
[96] H. Tang, S. Tan, and X. Cheng, "A survey on sentiment detection of reviews," *Expert Syst. Appl.*, vol. 36, no. 7, pp. 10760–10773, 2009.
[97] T. Mikolov, M. Karafiát, L. Burget, J. Cernocký, and S. Khudanpur, "Recurrent neural network based language model," *Interspeech*, vol. 2, no. 3, pp. 1045–1048, 2010.
[98] T. Mikolov, I. Sutskever, K. Chen, G. S. Corrado, and J. Dean, "Distributed representations of words and phrases and their compositionality," in *Proc. Adv. Neural Inf. Process. Syst.*, 2013, pp. 3111–3119.
[99] T. Mikolov, K. Chen, G. Corrado, and J. Dean. (Jan. 16, 2013). "Efficient estimation of word representations in vector space." [Online]. Available: https://arxiv.org/abs/1301.3781
[100] M. Junker, R. Hoch, and A. Dengel, "On the evaluation of document analysis components by recall, precision, and accuracy," in *Proc. 5th Int. Conf. Document Anal. Recognit.*, Sep. 1999, pp. 713–716.
[101] S. Baccianella, A. Esuli, and F. Sebastiani, "Multi-facet rating of product reviews," in *Proc. Eur. Conf. Inf. Retr. (ECIR)*, vol. 9. Berlin, Germany: Springer, 2009, pp. 461–472.
[102] H. Wang, Y. Lu, and C. Zhai, "Latent aspect rating analysis without aspect keyword supervision," in *Proc. 17th ACM SIGKDD Int. Conf. Knowl. Discovery Data Mining*, 2011, pp. 618–626.
[103] S. Moghaddam and M. Ester, "Opinion digger: An unsupervised opinion miner from unstructured product reviews," in *Proc. 19th ACM Int. Conf. Inf. Knowl. Manage.,* 2010, pp. 1825–1828.
[104] P. J. Stone, D. C. Dunphy, and M. S. Smith, *The General Inquirer: A Computer Approach to Content Analysis*. Oxford, U.K.: MIT Press, 1966.
[105] G. A. Miller, "WordNet: A lexical database for english," *Commun. ACM* vol. 38, no. 11, pp. 39–41, 1995.
[106] S. Baccianella, A. Esuli, and F. Sebastiani, "Sentiwordnet 3.0: An enhanced lexical resource for sentiment analysis and opinion mining," in *Proc. Int. Conf. Lang. Resour. Eval. (LREC)*, vol. 10, 2010, pp. 2200–2204.
[107] T. Wilson, J. Wiebe, P. Hoffmann, "Recognizing contextual polarity in phrase-level sentiment analysis," in *Proc. Conf. Human Lang. Technol. Empirical Methods Natural Lang. Process.*, 2005, pp. 347–354.
[108] S. M. Mohammad and P. D. Turney, "Emotions evoked by common words and phrases: Using mechanical turk to create an emotion lexicon," in *Proc. NAACL HLT Workshop Comput. Approaches Anal. Gener. Emotion Text*, 2010, pp. 26–34.
[109] J. W. Pennebaker, R. L. Boyd, K. Jordan, and K. Blackburn, "The development and psychometric properties of LIWC2015," Univ. Texas Austin, Austin, TX, USA, Tech. Rep., 2015.
[110] W.-T. Chen, S.-C. Lin, S.-L. Huang, Y.-S. Chung, and K.-J. Chen, "E-HowNet and automatic construction of a lexical ontology," in *Proc. 23rd Int. Conf. Comput. Linguistics, Demonstrations*, 2010, pp. 45–48.
[111] L.-W. Ku and H.-H. Chen, "Mining opinions from the Web: Beyond relevance retrieval," *J. Assoc. Inf. Sci. Technol.*, vol. 58, no. 12, pp. 1838–1850, 2007.
[112] H. Wang, Y. Lu, and C. Zhai, "Latent aspect rating analysis on review text data: A rating regression approach," in *Proc. 16th ACM SIGKDD Int. Conf. Knowl. Discovery Data Mining*, 2010, pp. 783–792.
[113] Y. Jo and A. H. Oh, "Aspect and sentiment unification model for online review analysis," in *Proc. 4th ACM Int. Conf. Web Search Data Mining,* 2011, pp. 815–824.
[114] J. Blitzer, M. Dredze, and F. Pereira, "Biographies, bollywood, boomboxes and blenders: Domain adaptation for sentiment classification," in *Proc. 45th Annu. Meet. Assoc. Comput. Linguistics,* vol. 7, 2007, pp. 440–447.
[115] M. Pontiki *et al.*, "Semeval-2016 task 5: Aspect based sentiment analysis," in *Proc. 10th Int. Workshop Semantic Eval. (SemEval)*, 2016, pp. 19–30.
[116] S. M. Mohammad, S. Kiritchenko, P. Sobhani, X.-D. Zhu, and C. Cherry, "A dataset for detecting stance in Tweets," in *Proc. LREC*, 2016, pp. 1–8.
[117] W. M. Rand, "Objective criteria for the evaluation of clustering methods," *J. Amer. Stat. Assoc.*, vol. 66, no. 336, pp. 846–850, 1971.
[118] S. Moghaddam and M. Ester, "Ilda: Interdependent lda model for learning latent aspects and their ratings from online product reviews," in *Proc. 34th Int. ACM SIGIR Conf. Res. Develop. Inf. Retr.*, 2011, pp. 665–674.
[119] S. Brody and N. Elhadad, "An unsupervised aspect-sentiment model for online reviews," in *Hum. Lang. Technol., Annu. Conf. North Amer. Chapter Assoc. Comput. Linguistics*, 2010, pp. 804–812.
[120] W. X. Zhao, J. Jiang, H. Yan, and X. Li, "Jointly modeling aspects and opinions with a maxent-lda hybrid," in *Proc. Conf. Empirical Methods Natural Lang. Process.*, 2010, pp. 56–65.
[121] M.-C. De Marneffe, B. MacCartney, and C. D. Manning, "Generating typed dependency parses from phrase structure parses," in *Proc. 5th Int. Conf. Lang. Resour. Eval. (LREC)*, Genoa, Italy, vol. 6, 2006, pp. 449–454.
[122] C. F. Baker, C. J. Fillmore, and J. B. Lowe, "The berkeley framenet project," in *Proc. 36th Annu. Meet. Assoc. Comput. Linguistics 17th Int. Conf. Comput. Linguistics*, vol. 1, 1998, pp. 86–90.
[123] M. Stevenson and Y. Wilks, "Word sense disambiguation," in *The Oxford Handbook of Computational Linguistics*. London, U.K.: Oxford Univ. Press, 2003, pp. 249–265.
[124] A. Kimmig, S. H. Bach, M. Broecheler, B. Huang, and L. Getoor, "A short introduction to probabilistic soft logic," in *Proc. NIPS Workshop Probab. Program., Found. Appl.*, 2012, pp. 1–4.
[125] S. M. Mohammad, "# emotional Tweets", in *Proc. 1st Joint Conf. Lexical Comput. Semantics-Main Conf. Shared Task, 6th Int. Workshop Semantic Eval.*, 2012, pp. 246–255.
[126] HP. *Hp Haven on Demand*. Accessed: Jul. 28,2017. [Online]. Available: https://www.havenondemand.com
[127] IBM. *IBM Alchemy*. Accessed: Jul. 28,2017. [Online]. Available: http://www.alchemyapi.com
[128] Vivekn. *Vivekn*. Accessed: Jul. 28,2017. [Online]. Available: http://sentiment.vivekn.com/docs/api
[129] M. Hu and B. Liu, "Mining opinion features in customer reviews," in *Proc. 19th Nat. Conf. Artif. Intell. (AAAI)*, vol. 4, 2004, pp. 755–760.
[130] C. Scaffidi, K. Bierhoff, E. Chang, M. Felker, H. Ng, and C. Jin, "Red opal: Product-feature scoring from reviews," in *Proc. 8th ACM Conf. Electron. Commerce*, 2007, pp. 182–191.
[131] Z. Li, M. Zhang, S. Ma, B. Zhou, and Y. Sun, "Automatic extraction for product feature words from comments on the Web," *Inf. Retr. Technol.* pp. 112–123, 2009.
[132] L. Zhang, B. Liu, S. H. Lim, and E. O'Brien-Strain, "Extracting and ranking product features in opinion documents," in *Proc. 23rd Int. Conf. Comput. Linguistics, Posters*, 2010, pp. 1462–1470.
[133] S. Raju, P. Pingali, and V. Varma, "An unsupervised approach to product attribute extraction," in *Proc. Eur. Conf. Inf. Retr.*, vol. 9. Berlin, Germany: Springer, 2009, pp. 796–800.
[134] J. Yu, Z.-J. Zha, M. Wang, and T.-S. Chua, "Aspect ranking: Identifying important product aspects from online consumer reviews," in *Proc. 49th Annu. Meet. Assoc. Comput. Linguistics: Hum. Lang. Technol.*, (Association for Computational Linguistics), vol. 1, 2011, pp. 1496–1505.

[135] P. Jiang, C. Zhang, H. Fu, Z. Niu, and Q. Yang, "An approach based on tree kernels for opinion mining of online product reviews," in *Proc. IEEE Int. Conf. Data Mining*, Dec. 2010, pp. 256–265.
[136] T.-J. Zhan and C.-H. Li, "Semantic dependent word pairs generative model for fine-grained product feature mining," in *Proc. Pacific-Asia Conf. Knowl. Discovery Data Mining*. Berlin, Germany: Springer, 2011, pp. 460–475.
[137] A.-M. Popescu and O. Etzioni, "Extracting product features and opinions from reviews," in *Natural Language Processing and Text Mining*. London, U.K.: Springer, 2007, pp. 9–28.
[138] W. Wang, S. J. Pan, D. Dahlmeier, and X. Xiao. (Mar. 22, 2016). "Recursive neural conditional random fields for aspect-based sentiment analysis." [Online]. Available: https://arxiv.org/abs/1603.06679
[139] H. Ye, Z. Yan, Z. Luo, and W. Chao, "Dependency-tree based convolutional neural networks for aspect term extraction," in *Proc. Pacific-Asia Conf. Knowl. Discovery Data Mining*. Cham, Switzerland: Springer, 2017, pp. 350–362.
[140] H. Saif, Y. He, and H. Alani, "Semantic sentiment analysis of Twitter," in *Proc. Int. Semantic Web Conf.* Berlin, Germany: Springer, 2012, pp. 508–524.
[141] Y. Yang, Z. Deyu, and Y. He, "An interpretable neural network with topical information for relevant emotion ranking," in *Proc. Conf. Empirical Methods Natural Lang. Process.*, 2018, pp. 3423–3432.
[142] L. Zhu, Y. He, and D. Zhou, "Hierarchical viewpoint discovery from Tweets using bayesian modelling," *Expert Syst. Appl.* vol. 116, pp. 430–438, Feb. 2019.
[143] D. Zhou, Y. Yang, and Y. He, "Relevant emotion ranking from text constrained with emotion relationships," in *Proc. Conf. North Amer. Chapter Assoc. Comput. Linguistics: Hum. Lang. Technol.*, vol. 1, 2018, pp. 561–571.
[144] D. Ying, J. Yu, and J. Jiang, "Recurrent neural networks with auxiliary labels for cross-domain opinion target extraction," in *Proc. 31st AAAI Conf. Artif. Intell.* 2017, pp. 1–11.
[145] W. Wang, S. J. Pan, and D. Dahlmeier. (Feb. 6,2017). "Multi-task coupled attentions for category-specific aspect and opinion terms co-extraction." [Online]. Available: https://arxiv.org/abs/1702.01776
[146] J. Zhao, H. Xu, X. Huang, S. Tan, K. Liu, and Q. Zhang, "Overview of chinese opinion analysis evaluation 2008," in *Proc. First Chin. Opinion Anal. Eval.*, 2008, pp. 1–21.
[147] F. Harrell, *Regression Modeling Strategies: With Applications to Linear Models, Logistic and Ordinal Regression, and Survival Analysis*. Cham, Switzerland: Springer, 2015.
[148] D. M. Blei, A. Y. Ng, and M. I. Jordan, "Latent Dirichlet allocation," *J. Mach. Learn. Res.*, vol. 3, pp. 993–1022, Jan. 2003.
[149] C. Lin and Y. He, "Joint sentiment/topic model for sentiment analysis," in *Proc. 18th ACM Conf. Inf. Knowledge Manage.*, 2009, pp. 375–384.
[150] H.-C. Yang, C.-H. Lee, and H.-W. Hsiao, "Incorporating self-organizing map with text mining techniques for text hierarchy generation," *Appl. Soft Comput.* vol. 34, pp. 251–259, Sep. 2015.
[151] T. L. Griffiths, M. I. Jordan, J. B. Tenenbaum, and D. M. Blei, "Hierarchical topic models and the nested chinese restaurant process," in *Proc. 16th Int. Conf. Neural Inf. Process. Syst.*, 2004, pp. 17–24.
[152] D. M. Blei, T. L. Griffiths, and M. I. Jordan, "The nested chinese restaurant process and bayesian nonparametric inference of topic hierarchies," *Journal of the ACM (JACM)*, vol. 57, no. 2, Art. no. 7, 2010.
[153] T. Liu, N. L. Zhang, and P. Chen, "Hierarchical latent tree analysis for topic detection," in *Proc. Joint Eur. Conf. Mach. Lear. Knowledge Discovery Databases*. Berlin, Germany: Springer, 2014, pp. 256–272.
[154] N. L. Zhang and L. K. Poon, "Latent tree analysis," in *Proc. 31st AAAI Conf. Artif. Intell.*, 2017, pp. 4891–4898.
[155] M. E. Peters *et al.* (Feb. 15, 2018). "Deep contextualized word representations." [Online]. Available: https://arxiv.org/abs/1802.05365
[156] J. Devlin, M.-W. Chang, K. Lee, and K. Toutanova. (Oct. 11, 2018). "BERT: Pre-training of deep bidirectional transformers for language understanding." [Online]. Available: https://arxiv.org/abs/1810.04805

RUI WANG received the M.S. degree in computer science and technology from Guangxi University for Nationalities, Nanning, China, in 2016. He is currently pursuing the Ph.D. degree in computer science with Southeast University. His research interests include natural language processing, sentiment analysis, unsupervised machine learning, neural generative models, and evolutionary computation.

DEYU ZHOU received the M.S. degree in computer science and technology from Nanjing University, Nanjing, China, in 2003, and the Ph.D. degree in computer science and technology from the University of Reading, U.K., in 2008. He is currently a Full Professor with Southeast University, Nanjing. His research interests include natural language processing, opinion mining, event extraction, sentiment analysis, and bioinformatics.



MINGMIN JIANG received the M.S degree in computer science and technology from Jinan University, Jinan, China, 2017. He is currently pursuing the Ph.D. degree in computer science with Southeast University. His research interests include natural language processing, computer vision, and machine learning.

JIASHENG SI received the M.S. degree in science of automation from the Hefei University of Technology, Hefei, China, in 2018. He is currently pursuing the Ph.D. degree in computer science with Southeast University, Nanjing, China. His research interests include natural language processing and machine learning.

YANG YANG received the B.S. degree in computer science and technology from Nanjing Normal University, Nanjing, China, in 2016. She is currently pursuing the M.S. degree in computer science with Southeast University, Nanjing. Her research interests include natural language processing and sentiment analysis.