IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Received June 2, 2019, accepted June 21, 2019, date of publication June 26, 2019, date of current version July 15, 2019.

Digital Object Identifier 10.1109/ACCESS.2019.2925059

A Survey of Sentiment Analysis Based on Transfer Learning

RUIJUN LIU[1,3], YUQIAN SHI[1,3], CHANGJIANG JI[2], AND MING JIA[1]

[1]School of Computer and Information Engineering, Beijing Technology and Business University, Beijing 100048, China
[2]Beijing Moviebook Technology Corporation Ltd., Beijing 100020, China
[3]Beijing Key Laboratory of Big Data Technology for Food Safety, Beijing Technology and Business University, Beijing 100048, China

Corresponding author: Ming Jia (jiaming@th.btbu.edu.cn)

This work was supported in part by the National Natural Science Foundation of China under Grant 61877002, in part by the Beijing Municipal Commission of Education under Grant PXM2019_014213_000007, and in part by the General Project of Beijing Municipal Education Commission under Grant KM201710011007.

ABSTRACT With the rapid development of the Internet industry, sentiment analysis has grown into one of the popular areas of natural language processing (NLP). Through it, the implicit emotion in the text can be effectively mined, which can help enterprises or organizations to make an effective decision, and the explosive growth of data undoubtedly brings more opportunities and challenges to the sentiment analysis. At the same time, transfer learning has emerged as a new machine learning technique that uses the existing knowledge to solve different domain problems and produces state-of-the-art prediction results. Many scholars apply transfer learning to the field of the sentiment analysis. This survey summarizes the relevant research results of the sentiment analysis in recent years and focuses on the algorithms and applications of transfer learning in the sentiment analysis, and we look forward to the development trend of the sentiment analysis.

INDEX TERMS Sentiment analysis, transfer learning, natural language processing.

I. INTRODUCTION

Sentiment analysis is a subdivision of text mining, which uses natural language processing and related computer technology to automatically extract or classify emotions in text [1]. The research object includes various information on the web such as hot topics which attracted users on social media. Sentiment analysis has a very wide range of applications. For instance in the consumer industry, user preference for different commodities can be obtained through the analysis of product reviews to help companies adjust sales strategies and make decisions. In social media [2]–[4], sentiment analysis of event comments plays an important role in public opinion control and emergency detection.

Since the beginning of 2000, sentiment analysis has become one of the most active research fields in natural language processing (NLP) [5]. In recent years, there have been many studies on sentiment analysis [1], [6], [7]. Zhang *et al.* [1] studied the sentiment analysis based on deep learning, and discussed the recent papers from three aspects that include aspect level, sentence level and document level. Kharde and Sonawane [6] gave a research review of sentiment analysis techniques based on Twitter data, compared existing analysis techniques such as machine learning and lexicon-based methods, and discussed the challenges and applications of sentiment analysis on Twitter. Cambria [7] divided the basic tasks of sentiment analysis into sentiment recognition and polarity detection, pointing out that the former focuses on extracting a set of emotional tags. In the current surveys of sentiment analysis, most of the sentiment analysis methods are classified from two perspectives. One is according to the granularity of text analysis, which is divided into three aspects: document level, statement level and aspect level. The other is according to the principle of method, it is mainly divided into three types: rule-based, machine learning, and deep learning. However, with the increasing volume of information, the collected data that need to be analyzed are huge, chaotic and irregular. This has brought more difficulties for sentiment analysis such as large data labeling and high cost of computing. Although the traditional sentiment analysis methods have some advantages, they are more limited due to massive data in practical application. Therefore, researchers combined sentiment analysis with transfer learning. Transfer learning [8] is an important research direction in machine

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

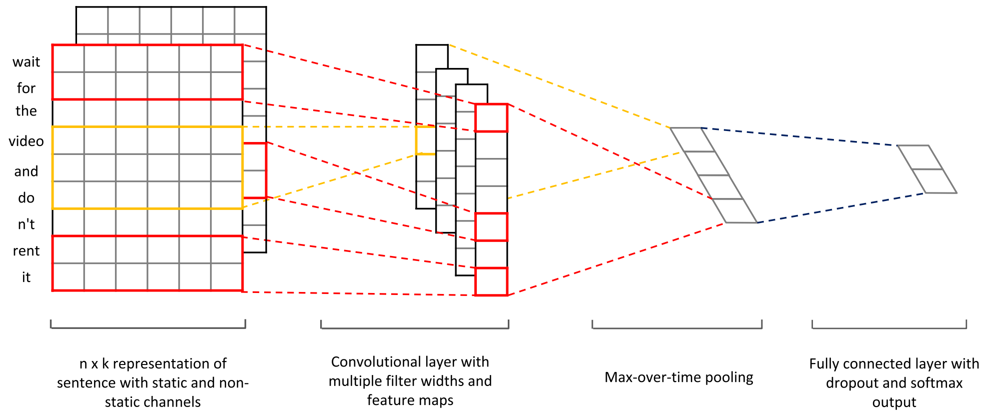

FIGURE 1. **CNN model architecture for processing a sentence [32].**

learning. It utilizes the similarity of data, data distribution or task. In this survey, we comprehensively summarize the sentiment analysis based on transfer learning, and hope to help people have a more all-sided and deep understanding of sentiment analysis. The main contributions of this survey are as follows: (1) The sentiment analysis methods are divided into traditional methods, deep learning methods, and transfer learning methods. And the point is on the summary of sentiment analysis methods based on transfer learning. (2) Summarize and compare the frequently used datasets of current sentiment analysis. (3) Discuss the challenges and opportunities of the current sentiment analysis.

II. THE RELATED METHODS

According to the different text types and application fields, sentiment analysis is also known as evaluation extraction, opinion mining and emotional polarity judgment [9], [10]. According to the different principles of sentiment analysis algorithms, this paper divides the methods into three categories, namely traditional methods, deep learning methods, and transfer learning methods to discuss and investigate recent related papers.

A. TRADITIONAL SENTIMENT ANALYSIS

Traditional sentiment analysis methods include lexicon-based approaches and non-neural network classifier approaches. The lexicon-based method [11]–[14] refers to judging the sentiment polarity by constructing an emotional dictionary, extracting emotional values, etc.

Turney and Littman [11] proposed the use of semantic polarity (ISA) algorithm to analyze text sentiment tendency. Hu and Liu [12] generated a dictionary containing positive and negative sentiment words through seed words in WordNet, and classify sentences according to the dictionary. Bravo-Marquez *et al.* [13] introduced a method of expanding the opinion lexicon in a supervised manner. An improved SO-PMI algorithm has been proposed by Yang *et al.* [14], which is more effective for emotional computing and emotional vocabulary modeling. The lexicon-based approach is easy to operate, but it relies heavily on the sentiment dictionary, and ignores the connection of the positional between words. The extraction of the text is simplistic and not comprehensive.

The non-neural network classifier belongs to the supervised machine learning method, which is realized by manual labeling and training model. The most common classifiers are Naive Bayes (NB) [15]–[17], Maximum Entropy [18]–[20], and SVM [21]–[23]. These classifiers perform well in different classification fields such as image classification [24] and text classification. With the increase of data volume and the need of diverse tasks, more scholars use the neural network model based on deep learning to apply to sentiment analysis tasks and achieve better results.

B. SENTIMENT ANALYSIS BASED ON DEEP LEARNING

Deep learning plays an important role in many fields such as Artificial Intelligence [25], Computer Vision [26] and Internet of Things [27], [28]. With the great progress in deep learning in the field of NLP, the application of deep learning to sentiment analysis has become more popular. The word representations of text is important for sentiment analysis, Bag of Words (BoW) and Word Embedding are models that commonly be used. The typical representative of the word embedding model is Word2Vec [29], [30]. In this paper, the neural network model for sentiment analysis is divided into three types that include CNN-based models, RNN-based models and hybrid neural network models to summarize recent related researches.

1) CNN-BASED MODELS

Convolutional Neural Network (CNN) was used to process images at first and achieve excellent results in computer vision [31]. In recent years, CNN has also been proven to be equally effective for NLP and get great results in text analysis.

Kim [32] proposed using CNN combined with pre-trained word vectors to classify texts at sentence level, and proved that a simple CNN model with a small amount of hyperparameters combined with static word vectors can achieve great results on many benchmarks. Figure 1 is the structure of this model.

Conneau *et al.* [33] applied the deeper CNN to NLP and proposed a new architecture VDCNN for text processing. The model combines VGGNet (Visual Geometry Group) with deep residual network (ResNet). The authors proved that using 29 convolutional layers can achieve better.

Johnson and Zhang [34] used the one-dimensional structure of data for text classification and prediction. This method changed the practice of using low-dimensional word vectors as input, directly using high-dimensional text as the processing object of CNN, and also explored the expansion of combining multiple convolutional layers to get more accurate prediction.

CNN can model the combination of features when processing text, and it is fast in processing speed, but it also has shortcomings. In order to achieve better results, new features need to be added to the CNN model.

2) RNN-BASED MODELS

Compared with the CNN model, Recurrent Neural Network (RNN) is more widely used in the NLP field, because RNN can process sequence information, and it has advantages in positional relationships, dependencies, etc., so it is more suitable for processing text information.

Tang *et al.* [35] used a bottom-up method to learn vector-based document representation. First, use CNN or Long Short-Term Memory (LSTM) to implement single-sentence representation, and then use GRU (Gated Recurrent Unit) network to encode the intrinsic and semantic connections between sentences. By this method, semantic information between sentences can be captured easily and better analysis results can be achieved.

Day and Lin [36] introduced a method for text analysis of Chinese Google Play consumer reviews using the LSTM deep learning model, achieving better performance in comparison with the Naive Bayes method and the support vector machine method.

Johnson and Zhang [37] explored a more complex LSTM region embedding method. LSTM can embed variable-sized text regions and achieve optimal results by combining LSTM-style region embedding and convolutional layers.

RNN can effectively integrate the information of the adjacent position of the text. However, the principle of the RNN model itself has some limitations, for example, it cannot distinguish the importance of word context cues, so it is necessary to explore a hybrid system of various model mechanisms to make up for its own shortcomings.

3) HYBRID NEURAL NETWORK MODELS

In this paper, the hybrid model is defined as a combined system, which is a combination of network models or mechanisms other than CNN and RNN. The emergence of hybrid models is due to the fact that existing traditional models cannot fully meet the task requirements, and people expect to improve the accuracy of text analysis from other angles than just the improvement of a single model. In general, the hybrid model is dominated by a neural network model, and other mechanisms are added to achieve higher performance based on the shortcomings. With the diversified development of sentiment analysis needs, the hybrid models have occupied an increasingly important position. By combining the two models or citing other algorithms in a single model, the analysis accuracy of the single model can be further improved. The paper recently applied a hybrid model to sentiment analysis to deepen the people's understanding.

Akhtar *et al.* [38] proposed a hybrid deep learning architecture, first learning the sentiment embedding vector from the convolutional neural network (CNN), and then selecting a set of optimization features through the multi-objective optimization (MOO) framework, finally SVM is used to classify the sentiment augmented optimized vector. This is the first attempt to apply this deep learning model to sentiment analysis in languages with less resources.

Yang *et al.* [39] formulated a hierarchical attentional network for document classification. A two-level attention mechanism is adopted at the word and sentence level. The visualization of the attention layer indicates the specific words selected by the model. Naturally, the benefit of joining the attention mechanism is the ability to intuitively explain the importance of individual sentences and words to the classification category.

Traditional neural models such as LSTM capture context information implicitly, do not explicitly display important contextual cues for specific aspects of the text, and use the same operations for each context word, so it cannot explicitly show the importance of each contextual word. The ideal solution should be able to explicitly capture the importance of contextual words. Therefore, Tang *et al.* [40] proposed a sentiment analysis of aspect-level text using the Deep Memory Network, which explicitly captures the importance of each context word when inferring the level of emotional polarity.The effect is much better than LSTM and attention-based LSTM.

The hybrid system can combine the advantages of different models to maximize model performance. Practice has proved that sentiment analysis methods based on deep learning can continuously train and improve new models based on large amounts of data, and apply them to different sentiment analysis tasks. With the explosion of data today, for a new task, it is necessary to start a new training-process, which faces three serious challenges. First, a great model relies on an annotated corpus, but a corpus that is correctly labeled is difficult to obtain. Manually labeling data is a time-consuming and laborious process. Secondly, the training of the new model has higher performance of the computer, but most ordinary users do not have strong computing power, which limits people's research on higher-level models. Third, people want to develop a unique solution for different individual needs tasks, but according to the above, this process is time consuming and difficult.Therefore, it is more expected that existing models have better generalization ability and can be applied to more different fields. It is precisely because of the above problems that the sentiment analysis method combined with transfer learning has emerged.

C. SENTIMENT ANALYSIS BASED ON TRANSFER LEARNING

Transfer learning is the method that uses the similarity of data, data distribution, model, task and so on to apply the knowledge already learned in one domain to the new domain. Through transfer learning, the labeled data can be used to construct the model and be used in the target domain data to increase the annotation of the target data. Before we introduce different sentiment analysis methods based on transfer learning, we first describe some notations and categories that be used in the paper. Domain: Consists of data features and feature distributions, including the source domain and domain, where the source domain exists knowledge and the target domain needs to be learned. Task: Includes the objective function and learning outcomes for the purpose of learning.

In a review article made by Pan and Yang [8], a more authoritative definition of transfer learning has been made, and according to the learning method, the transfer learning is divided into four categories, namely instance-transfer approach, feature-representation-transfer approach, parameter-transfer approach and relational-knowledge-transfer approach. Zhang and Yang [42] also introduced Multi-Task Learning (MTL), analyzed the characteristics of different methods, discussed the combination of Multi-Task Learning (MTL) and other learning paradigms. Weiss *et al.* [41] conducted a detailed survey about the related methods and applications of transfer learning, divided the transfer learning methods into two categories: homogeneous and heterogeneous, and discussed the negative transfer problem. At the same time, many scholars have made surveys on transfer learning applications, such as research [43] on Visual Domain Adaptation and research [44] on Activity Recognition.

In the next section, the transfer learning methods applied in sentiment analysis will be introduced, and it will be discussed comprehensively in three categories: parameter-transfer methods, instance-transfer methods and feature-representation-transfer methods.

1) PARAMETER-TRANSFER METHODS

Parameter-transfer method leverages the parameter sharing model of the source and target domains. Using the parameter-transfer method, scholars can transfer the trained model parameters in a large number of datasets to the target task.

In the parameter-transfer method applied to the NLP field, word2vec [29] which is a simple transfer technology only for the first layer of the model has many applications, which has a great impact in practice and be used in many advanced technology.

Glorot *et al.* [45] studied the domain adaptive problem of sentiment classifiers, using Stacked Denoising Auto-encoder (SDA) to solve edge distribution differences. The first step of the algorithm is to find an invariant potential feature space by the input space that has been trained by the Stacked Denoising Auto-encoder, and then uses the potential features and the labeled source data to train the classifier. This method is excellent in comment sentiment analysis, but it is more dependent on parameter initialization.

The marginalized SDA (mSDA) [46] algorithm, which improved the shortcomings of the SDA algorithm: high computational cost and lack of scalability of high-dimensional features. The algorithm has powerful feature learning capabilities and does not require optimization algorithms to train parameters.

Sun [47] proposed a modified version of the mSDA ++ algorithm for the marginalized SDA (mSDA), which can learn low-dimensional features. The combination of mSDA and EASYADAPT algorithm can improve the accuracy of text classification. In addition, the mSDA ++ algorithm accelerates subsequent calculations and reduces data storage space.

Mccann *et al.* [48] and Peters *et al.* [49] used a method of combining embedding from other tasks with different levels of input. Mccann *et al.* [48] trained the neural network model through an English-German translation task to obtain an output called "context vectors (CoVe)". They used CoVe for many NLP tasks, including semantic sentiment analysis, problem classification, text implied, question and answer. Figure 2 is a schematic diagram of the process about combining other tasks with the target task.

Peters *et al.* [49] proposed EMLo (Embeddings from Language Models), which is a method for extracting word vectors of deep semantic features. It can train the bidirectional LSTM model (biLM) on the big corpus at first, then connect its own neural network directly on the basis of biLM, take the output of biLM as the input of its own network, and use LSTM to generate characterization of words. In this way, a richer word representation can be introduced and applied to many types of NLP tasks, but this method still trains the main task model from the beginning, and regards pre-training embedding as a fixed parameter, limiting its effect.

Dai and Le [50] first put forward a fine-tuning of the language model (LM) and proposed two methods for improving sequence learning with recurrent networks using unlabeled data.The function of the first method is to predict the subsequent content in the sequence; the second method can use the sequence-autoencoder to select the input sequence into the vector and predict the input again.Figure 3 is the model structure. The two algorithms can be exploited as pre-training algorithms for post-supervised sequence learning algorithms, enabling powerful performance in many classification tasks. However, applicability is limited severely due to the need for millions of intra-domain documents to achieve good performance.

Howard and Ruder [51] proposed the Universal Language Model Fine-tuning (ULMFiT), which can be exploited to NLP. ULMFit recommends training language models on very large corpus to contain more semantic information and use it as the backbone of the classifier. When applying it to different

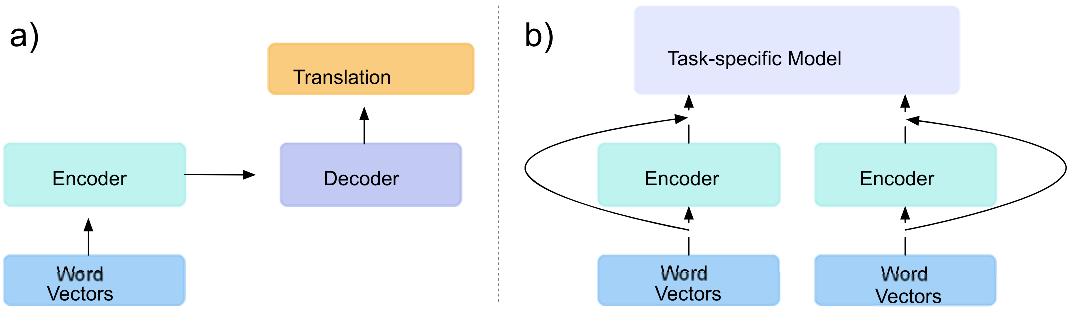

FIGURE 2. Schematic diagram of the process about combining other tasks with the target task [48].

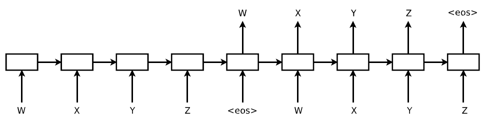

FIGURE 3. Self-encoder model structure [50].

datasets, you only need to fine-tune the parameters of the language model to take these differences into account, then add a classifier layer at the top of the language model and train only that layer. Howard J et al. have provided pre-trained language models. Figure 4 is a training procedure for ULMFiT.

2) INSTANCE-TRANSFER METHODS

Instance-transfer method refers to the data sharing of the source and target domain. It can be filtered from source domain by re-weight. Through it, data from the target domain can be augmented with the labeled source domain samples. Several classic algorithms for transfer learning will be introduced at first.

TrAdaBoost [52] is an algorithm for extracting instances from source data, which combines some of the available labeled source data with a small number of labeled target data to construct a more accurate model than simply using labeled target data. The algorithm works well when the source data and the target data have many similarities, but if the samples in the target data have more noise, it is possible to increase the difficulty of training the classifier.

Vieriu *et al.* [53] proposed a boosting-based transfer learning method, which can transfer knowledge to different target areas in the training field by learning training samples from multiple fields, and reduce the cost of labeling target areas.

Wang and Li [54] introduced a transfer learning method based on boosting framework and distributed measurement method. They believe that the TrAdaBoost algorithm forces the boosting mechanism to ignore valuable data while learning valuable data, but it is still not perfect.

CP-MDA [55] (conditional probability based multi-source domain adaptation) can correct the difference in conditional distribution between the two domains. The core idea is to use the source domain classifier to tag unlabeled data in the target domain. The first step is to assign a classifier to the source domain, and then find a weight value for each classifier as a function of the proximity of the two domains in the conditional distribution between the two domains. The weighted combination source domain classifier creates a learning task by accumulating a pseudo tag for the unlabeled target domain data. Finally, the target learner is trained based on the tagged target domain data and the target domain data of the pseudo tag.

2SW-MDA [55] (two stage weighing framework for multi-source domain adaptation), which is an upgraded version of CP-MDA, can handle the difference in edge distribution. The algorithm can calculate the weight for each source domain by the difference of the edge distribution between the two domains, and use the weight as a function of the difference of the conditional distribution obtained in the CP-MDA method. The innovation of the method is to calculate the weight of the source domain as a conditional probability function.

Next, we will introduce examples on the application of transfer learning algorithms for sentiment analysis.

Xu *et al.* [56] proposed the instance level transfer learning method applied to cross lingual opinion analysis, and translated other useful markup languages into target language as supplementary training data to improve the opinion classification in the target language. TrAdaBoost [52] algorithm is used to reduce the impact of low quality translation corpus. The algorithm effectively improves the current situation of opinion analysis of resource-scarce languages by using small target language training data and large cross lingual training data.

Transfer learning can use the source domain data to augment the target domain data by the above method, but in the process of transfer, misclassification may occur to cause negative transfer, that is, the knowledge learned in the source domain has a negative effect on learning on the target domain. In response to the above problems, Gui *et al.* [57] proposed a negative transfer detection method for cross lingual analysis, which examines and removes the noise of the training samples by detecting high quality samples. This method can effectively reduce the misclassification by iterative means.

Serikawa and Lu [58] introduced a fast joint trigonometric filtering dehazing algorithm that can detect and reduce the

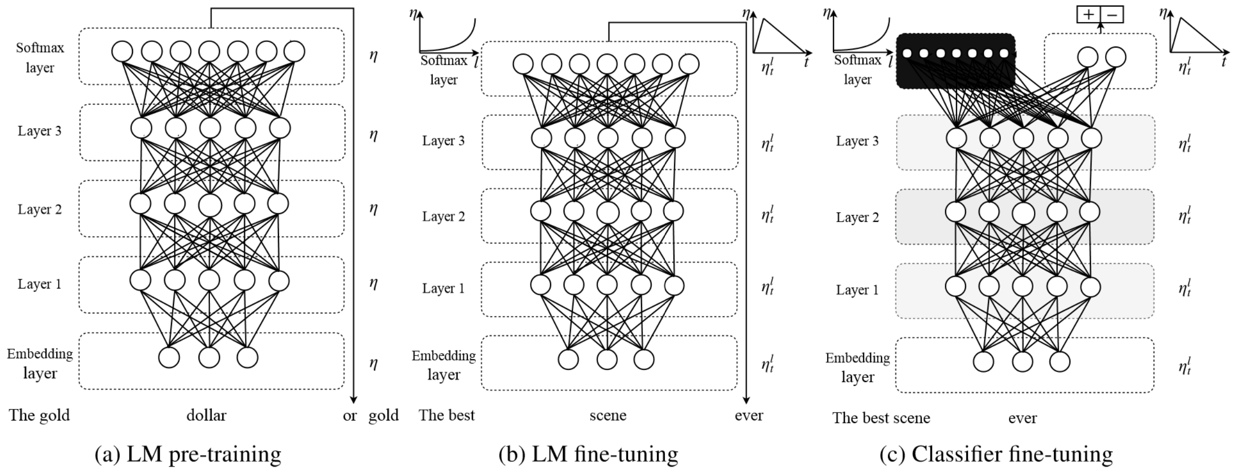

FIGURE 4. **ULMFiT training steps: (a) universal domain LM pre-training ; (b) target task LM fine-tuning; (c) target task classifier fine-tuning [51].**

image noise, it can be used as a reference for noise detection of cross lingual analysis.

Wang *et al.* [59] described an adaptive training data selection method, which can effectively avoid the noisy and fuzzy of source domain training data. The iterative method is used to combine the samples from the source domain and the target domain training data by using the informativeness measures, to evaluate the performance of the classifier of the target domain, and to update the informativeness measures of the next iteration.

3) FEATURE-REPRESENTATION-TRANSFER METHODS

The condition of feature-representation-transfer method is that the source domain and the target domain have part crossover features. It is necessary to transform the data of the two domains into the same feature space through feature transformation, and then perform traditional machine learning. Because this method has lower requirements for similarity between the two domains, it is more widely used and performs well in various tasks of NLP. The classic algorithms and applications for sentiment analysis are introduced below.

SCL [60] (Structure Correspondence Learning): SCL uses the pivot and non-pivot features to construct a correlation model in order to complete the transfer learning. The algorithm is very dependent on the quality of the latent space and the number of auxiliary learning tasks, which has significant limitations.

SFA [61] (Spectral Feature Alignment): The algorithm uses domain-independent words as a bridge to construct common latent space of domain-specific words in different fields and reduces differences in different fields for transfer learning. SFA assumes that there are rich source domain data and some labeled target domain data, and the bilateral graph model is built by identifying domain-independent words. If the graph shows that the domain-specific words of the two domains are connected to a common domain-independent word, then there is a higher chance that the two domain-specific word features are aligned.

SS-FE [62] (Feature Ensemble Plus Example Selection): A tag adaptation algorithm based on part-of-speech(POS) feature ensemble, which combines its proposed Feature Ensemble(FE) model with a Principal Component Analysis-based Sample Selection(PCA-SS) algorithm to achieve transfer learning.

TCT [63] (topical correspondence transfer): To reduce the distribution difference by establishing the correspondence between other topics in different topics, and finally discover the topic and classify the emotions in the process of optimizing the objective function. The core element of TCT is the optimization problem, which is expressed as a joint non-negative matrix factorization. When the features of the two domains are completely different, they belong to a heterogeneous environment. Zhou *et al.* [64] mentioned a hybrid heterogeneous transfer learning (HTTL) method to solve the sentiment classification of a heterogeneous environment with rich labeled source domain data and unlabeled target domain data.This method can simplify the problem to the isomorphic adaptation problem by learning the asymmetric transformation from the target domain to the source domain. The key to solving the isomorphic adaptation problem is to reduce the distribution deviation between the target domain and the labeled source domain of the transition, that is, to find a common potential feature space. Finally, the classifier of the method is jointly trained by a common potential feature space from the marked source domain data.

4) SUMMARY

The above is an overview of sentiment analysis algorithms and applications based on transfer learning. People can have a more comprehensive understanding about the application scenes of the three types of methods. For the parameter-transfer method, it can utilize the trained parameters and the rich semantic features of large corpus, and at the same time reduce the computational overhead of model training so that users can reduce the computational cost. However, the method has the disadvantage that the parameters are

TABLE 1. Approaches for sentiment analysis.

Research	Basic model/core idea	Transfer learning methods	Datasets	Accuracy(%)
Kim Y, CNN-static /CNN-multichannel [32]	CNN		MPQA/SST-2	89.60/88.10
Alexis Conneau et al., VDCNN [33]	CNN		Yelp	95.72
Johnson R et al., seq-CNN [34]	CNN		IMDB	91.26
Tang et al., LSTM-GRNN [35]	LSTM		Yelp	67.60
Day M Y et al. [36]	LSTM		Google Play Consumer Review	94.00
Johnson R et al., oh-2LSTMp [37]	LSTM		IMDB	94.06
Akhtar M S et al., CNN-SVM(W+X)(resource-poor languages) [38]	CNN		Online reviews for aspect based sentiment analysis in Hindi	65.96
Yang Z et al., HN-ATT [39]	GRU(Gated Recurrent Unit)		Yelp	71.00
Sun M et al., EA+mSDA [47]	SDA(Stacked Denoising Autoencoder)	Parameter-transfer	Amazon Product Reviews	81.08
McCann et al., CoVe [48]	LSTM	Parameter-transfer	IMDB/SST-5	91.80/53.70
Peters et al., EMLo [49]	biLM(bidirectional LSTM)	Parameter-transfer	SST-5	54.70
Dai and Le, SA-LSTM [50]	LSTM	Parameter-transfer	IMDB	92.80
Howard J et al., ULMFiT [51]	AWD-LSTM	Parameter-transfer	IMDB	95.40
Dai W et al., TrAdaBoost [52]	Boosting, SVM	Instance-transfer	20 Newsgroups	93.15
Xu R et al., TrStr [56]	Transfer Self-training algorithm	Instance-transfer	NTCIR-7 MOAT Corpora	73.04
Lin Gui et al. [57]	Rademacher distribution(noise reduction)	Instance-transfer	NLP&CC 2013 cross-lingual opinion analysis dataset	80.84
Wang W et al., CDS [59]	adaptive source-domain training instance selection	Instance-transfer	Blog, Diary, Exp, Fairy	67.03
Blitzer J et al.,SCL [60]	Identify correspondences among features from different domains	Feature-representation-transfer	Source Domain: WSJ,Target Domain: Biomedical Text	96.10
Pan S J et al., SFA [61]	construct a bipartite graph to model the co-occurrence relationship	Feature-representation-transfer	Amazon Product Reviews	72.86/86.75
Xia R et al., SS-FE [62]	feature ensemble(FE), PCA-based sample selection(PCA-SS)	Feature-representation-transfer	Amazon Product Reviews	72.50/84.87
Zhou G et al., TCT [63]	a joint non-negative matrix factorization	Feature-representation-transfer	Amazon Product Reviews	74.83/86.33

not easy to converge. For the instance-transfer method, its application scope is affected by the similarity between the two domains. The analysis result is poor when the two domains have large differences. However, this method is easy to apply in an unlabeled dataset. For the feature-representation-transfer method, the application scope is more extensive than the instance-transfer method. When the two domains have partial cross features, the transformation can be used to obtain a great analysis effect, but it is difficult to solve the optimization problem. According to the above analysis, transfer learning methods can solve some problems such as data labeling, computational overhead and personalized demand brought by big data. In order to better compare the differences between different algorithms, we will show the experimental performance between the different methods. As shown in Table 1, it shows the main information of different algorithms, including four aspects: research content, basic model/core idea, transfer learning methods, datasets and accuracy. By comparing the basic models and experimental performance of different methods, users can be provided with more detailed sentiment analysis resources to compare different methods.

We will make a detailed analysis of the methods in Table 1. It can be seen in Table 1 that the models of [33] and [37] have achieved higher accuracy without using transfer learning. Although, LSTM and its derivative models are widely used in transfer learning as the basic model, [33] uses the deeper CNN in which max-pooling performs better than other pooling types and the result shows higher performance, which indicate that in transfer learning deeper CNN is also worth further exploration. The better effect of LSTM combined with one-hot CNN in [37] proves the advantages of the regional embedding method, and further proves that CNN has a good application in text processing. Most of the parameter-transfer methods are based on LSTM, and [50] and [51] get better results. In [50], the parameters obtained by the model trained by the unlabeled data are used as the initialization parameters for the training of the next stage model. This pre-training approach enhances the stability of training about LSTM and effectively combines the advantages of both unsupervised and supervised learning. What is worth learning in [51] is the versatility of the model, which can be used for various NLP tasks and achieve great results. Its fine-tuning tips are also used for reference. One is discriminative fine-tuning: different layers use different learning rates. The other is slanted triangular learning rates: a new way to correct learning rates. The core of the instance-transfer method is weight reuse. The current method does not use the neural network model, mainly in the improvement and optimization of machine learning algorithms. It can be seen from Table 1 that the datasets selected by experiments of different algorithms have certain differences. The disadvantage of this is that they cannot be compared with other methods of the same type. On the other hand, this type of method is also mainly applied to text tasks in specific fields such as cross-language sentiment analysis, small-field text analysis, etc., which may be convenient to apply in practice. However, due to the limitations of its own application scenarios (higher similarity between source and target domains), most scholars have studied more on model-transfer methods and feature-representation transfer methods. Most feature-representation transfer methods use Amazon Product Reviews as an experimental dataset, which makes it easier for users to compare different algorithms and visualize the pros and cons of them. The core of this type of method is the optimization of the feature space, and the neural network model is rarely used. There will be some deviations in the accuracy of the experimental results due to different experimental settings or different selection of source and target domains. In Table 1, we show the optimal and worst results for the three algorithms in [61]–[63]. It can be seen that the accuracy of the three algorithms in the experiment is not much different, and the highest accuracy is [61], but from the worst results, [63] is more stable. It is worth considering that the domain of experimental data is same in the best performance of the three methods. The source domain is all the Electronics (E) dataset, and the target domain is all the Kitchen (K) dataset, which means that the text data features have the same impact on different algorithms. In addition, the combination of different source and target domains can produce different results with large differences in accuracy, and also proves the impact of the similarity between source and target domains on transfer learning.

From Table 1, we can see that not all transfer learning methods are more accurate than normal methods. This is mainly because of two reasons. First, the difference between multi-task and single tasks. Due to the complexity of the text tasks, when the transfer learning method is applied to NLP, different target tasks need to adopt different fine-tuning methods, so many transfer learning methods are difficult to apply to multiple NLP tasks. Therefore, in terms of text sentiment analysis tasks, the performance of some transfer methods may be slightly worse than the single model for sentiment analysis tasks. Second, the limitations of the transfer learning itself and the lack of fine-tuning. In the process of transfer, due to the difference in data or tasks between the target domain and the source domain, it is necessary to pay attention to the negative effects that negative transfer may have and the catastrophic information forgetting or over-fitting when fine-tuning. Although the performance of transfer learning is not best, it can solve the problems encountered in current sentiment analysis such as label consumption, time consumption and computer performance problems. Therefore, the application of transfer learning is more extensive, and it can solve multi-angle text tasks, which is a hot research topic in the future.

III. SENTIMENT ANALYSIS DATASETS

Sentiment analysis relies on textual data, which means the validity and accuracy of the model or algorithm can be validated effectively only by selecting the appropriate datasets. How to select the best dataset in the shortest time for experiment is a problem worthy of attention. Therefore, this paper summarizes and compares the databases commonly used in sentiment analysis, so as to help users choose more suitable dataset for method validation.

A. IMDB

The large film review dataset IMDB [65] was proposed by Andrew Maas et al. in the 2011 ACL paper, which is a data set for binary emotion classification, each sample is a txt file, including training sets, test sets, and no tagged datasets. There are 25000 training sets, 12500 positive and negative, and 25000 test sets, 12500 positive and negative.

B. STANFORD SENTIMENT TREEBANK

Stanford Sentiment Treebank [66] is the semantic lexical dataset annotated by Stanford University, including a fine-grained emotional label of 215154 phrases in a parse tree of 11855 sentences. It is divided into two tasks, one is a two-category task, including 6920 training sets, 872 validation sets, and 1821 test sets; one is a five-category task, which contains 11855 sentences and 215154 phrases (category 5).

C. YELP

The Yelp dataset is an internal dataset published by Yelp, the largest review site in the United States. This dataset is a subset of the merchants, reviews, and users' data covered by Yelp and can be used for personal, educational, and academic purposes.

Yelp hopes that more scholars will use this dataset to make more innovative research. The dataset can be used in the following three aspects. The first is image classification. The second is natural language processing. There are a lot of mining metadata in user evaluation dataset, which can be used to infer semantics, business attributes and emotions. The third is image mining. For example, mining the relationship between users to find usage rules.

D. MULTI-DOMAIN SENTIMENT DATASET (AMAZON PRODUCT REVIEWS)

Multi-Domain Sentiment Dataset [67] was proposed by John Blitzer in the 2007 ACL paper, which contains product reviews from a lot of different domains acquired from Amazon.com. Comments contain levels (1 to 5 levels) and can be converted to binary tags. The data set contains more than 100000 sentences, which can be divided into positive and

TABLE 2. Comparison of datasets related to sentiment analysis.

Dataset	Number	Classification	Description
IMDB	100000 sentences	two-category	Movie review
Stanford Sentiment Treebank	11855 sentences divided into 215154 phrases	two-category and five-category	
Yelp	More than 163 million pieces of data		User reviews, business information, etc.
Multi-Domain Sentiment Dataset	More than 100000 sentences	two-category and five-category	Multi-domain product review
Sentiment140	1600000 tweets	three-category	Twitter information

negative categories or strong positive, weak positive, neutral, weak negative, strong negative five categories.

E. SENTIMENT140 (STS)

Sentiment140 [68] has 1600000 tweets extracted from the twitter api. These tweets have been annotated (Zero, two and four respectively represent negative, neutral and positive.). The tweets contain the following six fields: target, tweet polarity ;ID, tweet ID; date, tweet date; flag, query; user, the user who sent the tweet; text, tweet content. Table 2 compares the datasets related to sentiment analysis.

IV. DISCUSSION AND FUTURE PROSPECTS

In recent years, scholars' research has made the types of sentiment analysis more abundant and personalized. However, with the explosive growth of Internet users, the form and content of network data are developing towards diversification. The current methods still have certain limitations. People still expect to have more intelligent methods in the field of sentiment analysis. The following is a detailed introduction to the problems existing in the current sentiment analysis and the future development direction.

A. CROSS-DOMAIN SENTIMENT ANALYSIS

Network text has the characteristics of diversity and multi-domain, so people hope to apply sentiment analysis algorithms to further exploration of text in different domains. Sentiment analysis has certain limitations on its corpus, and it is difficult for texts in different domains to train sentiment analysis models through the same corpus. For example, when the model trained by the Film Review Corpus is used to analyze the emotion of restaurant reviews, the performance of the model tends to be poor due to excessive differences between the two domains, and the restaurant evaluation cannot be analyzed accurately. Based on this, we must use specific corpus according to different text fields such as film reviews and restaurant reviews, but in reality, there are not enough data in many fields to support scholars' research on sentiment analysis, so cross-domain sentiment analysis has gradually become a hot topic of sentiment analysis.

Cross-domain transfer learning aims to use source data from other domains to help target learning tasks [8], which is used by more scholars in cross-domain sentiment analysis. Different domains of data have different feature spaces and data distributions. In this case, the research purpose is to extract the features that are useful in both target and source domains, and to obtain the shared feature space of cross-domain transfer learning problem. Some scholars select the best features of the source and target domains and re-weight them to improve the final classification [69]. This process can be described as two phases. The first of which is to extract the common features of the two domains, and then the classifier is trained according to the useful features of the target domain [70]. However, the two-stage operation may be inconsistent due to the difference in feature space. To overcome this shortcoming, Maracini *et al.* [71] proposed CD-ALMA (Cross-Domain Aspect Label Propagation through Heterogeneous Networks), which has been proved to have a great effect by using heterogeneous network-based representations to combine different features into nodes in the network and then exploring language representations through tag propagation algorithms for heterogeneous networks. However, cross-domain transfer learning still has certain problems, such as the selection of source domain, the noise interference in the combination of source domain and target domain, the optimization of the combination of source domain and target domain combined with feature space, etc., which brings challenges to cross-domain sentiment analysis.

B. SENTIMENT ANALYSIS AND NEGATIVE TRANSFER

When using transfer learning for sentiment analysis, it is critical to select the source domain. The core problem of transfer learning is to find the similarity between the two domains, and complete the learning process from the source domain to the target domain through the similarity between the domains. But the similarities between the domains are very subjective, and there is no scientific and unanimous standard. If the two domains are not similar or even completely different, a negative transfer occurs, that is, the knowledge learned by the transfer learning has a negative impact on the task of the target domain.

In practical applications, it is necessary to find a reasonable similarity between the two domains, and choose a reasonable transfer method to avoid negative transfer. Recently, the paper published by Yang Qiang team [72] used face data to identify aircraft. The face is completely dissimilar to the plane but information can be transferred through transfer learning. It can be regarded as an effective method to solve negative transfer. Therefore, how to avoid and solve negative transfer effectively in sentiment analysis based on transfer learning is a hot research direction in the future.

C. FINE-GRAINED RESEARCH IN SENTIMENT ANALYSIS

Fine-grained sentiment analysis is based on aspect-level, also known as feature-based opinion mining. At present, the network text is becoming more and more diversified. While fine-grained sentiment analysis can extract the comment objects, object attributes and their corresponding emotional words in the comment at the data feature level, which can better meet the needs of people's sentiment analysis. So many scholars are committed to it [73]–[77]. But fine-grained analysis is still difficult for short texts, which have less information and are more difficult to extract semantic relations between words. Therefore, aspect-level sentiment analysis that centered on short texts of semantic richness is considered to be one of the most promising research directions in the future.

V. CONCLUSION

This review makes a comprehensive investigation and discussion on sentiment analysis in different directions: (1) Introduces traditional sentiment analysis methods and clarifies their advantages and disadvantages; (2) Analyzes the sentiment analysis method based on transfer learning, and summarizes the recent research; (3) Analyzes and summarizes of commonly used datasets. This paper also discusses the sentiment analysis and puts forward the prospect of it. (1) The application of cross-domain transfer learning in aspect extraction has not been fully explored. (2) How to solve the negative transfer problem in transfer learning becomes the difficulty of using transfer learning for text analysis. (3) Aspect-level sentiment analysis centered on short texts of semantic richness is considered to be one of the most promising research directions in the future.

REFERENCES

[1] L. Zhang, S. Wang, and B. Liu, "Deep learning for sentiment analysis: A survey," *Wiley Interdiscipl. Rev., Data Mining Knowl. Discovery*, vol. 8, no. 4, p. e1253, Jul./Aug. 2018.

[2] M. del P. Salas-Zárate, J. Medina-Moreira, P. J. Álvarez-Sagubay, K. Lagos-Ortiz, M. A. Paredes-Valverde, and R. Valencia-García, "Sentiment analysis and trend detection in Twitter," in *Proc. Int. Conf. Technol. Innov.* Cham, Switzerland: Springer, 2016, pp. 63–76.

[3] S. Alashri, S. S. Kandala, V. Bajaj, R. Ravi, K. L. Smith, and K. C. Desouza, "An analysis of sentiments on Facebook during the 2016 U.S. presidential election," in *Proc. IEEE/ACM Int. Conf. Adv. Social Netw. Anal. Mining (ASONAM)*, Aug. 2016, pp. 795–802.

[4] M. Mertiya and A. Singh, "Combining naive bayes and adjective analysis for sentiment detection on Twitter," in *Proc. Int. Conf. Inventive Comput. Technol. (ICICT)*, Aug. 2016, pp. 1–6.

[5] D. M. El-Din Mohamed Hussein, "A survey on sentiment analysis challenges," *J. King Saud Univ.-Eng. Sci.*, vol. 30, no. 4, pp. 330–338, 2018.

[6] V. A. Kharde and P. S. Sonawane, "Sentiment analysis of Twitter data: A survey of techniques," Apr. 2016, *arXiv:1601.06971*. [Online]. Available: https://arxiv.org/abs/1601.06971

[7] E. Cambria, "Affective computing and sentiment analysis," *IEEE Intell. Syst.*, vol. 31, no. 2, pp. 102–107, Mar./Apr. 2016.

[8] S. J. Pan and Q. Yang, "A survey on transfer learning," *IEEE Trans. Knowl. Data Eng.*, vol. 22, no. 10, pp. 1345–1359, Oct. 2010.

[9] E. Cambria, B. Schuller, Y. Xia, and C. Havasi, "New avenues in opinion mining and sentiment analysis," *IEEE Intell. Syst.*, vol. 28, no. 2, pp. 15–21, Mar. 2013.

[10] R. Xia, F. Xu, C. Zong, Q. Li, Y. Qi, and T. Li, "Dual sentiment analysis: Considering two sides of one review," *IEEE Trans. Knowl. Data Eng.*, vol. 27, no. 8, pp. 2120–2133, Aug. 2015.

[11] P. D. Turney and M. L. Littman, "Measuring praise and criticism: Inference of semantic orientation from association," *ACM Trans. Inf. Syst.*, vol. 21, no. 4, pp. 315–346, 2003.

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

[13] F. Bravo-Marquez, E. Frank, and B. Pfahringer, "Building a Twitter opinion lexicon from automatically-annotated tweets," *Knowl.-Based Syst.*, vol. 108, pp. 65–78, Sep. 2016.

[14] A. M. Yang, J. H. Lin, Y. M. Zhou, and J. Chen, "Research on building a Chinese sentiment lexicon based on SO-PMI," *Appl. Mech. Mater.*, vol. 263, pp. 1688–1693, Dec. 2013.

[15] S. Yoshida, J. Kitazono, S. Ozawa, T. Sugawara, T. Haga, and S. Nakamura, "Sentiment analysis for various SNS media using Naïve Bayes classifier and its application to flaming detection," in *Proc. IEEE Symp. Comput. Intell. Big Data (CIBD)*, Dec. 2014, pp. 1–6.

[16] M. Bilal, H. Israr, M. Shahid, and A. Khan, "Sentiment classification of Roman-Urdu opinions using Naïve Bayesian, decision tree and KNN classification techniques," *J. King Saud Univ.-Comput. Inf. Sci.*, vol. 28, no. 3, pp. 330–344, 2016.

[17] L. Wikarsa and S. N. Thahir, "A text mining application of emotion classifications of Twitter's users using Naïve Bayes method," in *Proc. 1st Int. Conf. Wireless Telematics (ICWT)*, Nov. 2015, pp. 1–6.

[18] X. Fei, H. Wang, and J. Zhu, "Sentiment word identification using the maximum entropy model," in *Proc. 6th Int. Conf. Natural Lang. Process. Knowl. Eng.*, Aug. 2010, pp. 1–4.

[19] H. Y. Lee and H. Renganathan, "Chinese sentiment analysis using maximum entropy," in *Proc. Workshop Sentiment Anal. AI Meets Psychol.*, 2011, pp. 89–93.

[20] M. Murata, K. Uchimoto, M. Utiyama, Q. Ma, R. Nishimura, Y. Watanabe, and K. Doi, "Using the maximum entropy method for natural language processing: Category estimation, feature extraction, and error correction," *Cogn. Comput.*, vol. 2, no. 4, pp. 272–279, 2010.

[21] J. Smailović, M. Grčar, N. Lavrač, and M. Žnidaršič, "Stream-based active learning for sentiment analysis in the financial domain," *Inf. Sci.*, vol. 285, pp. 181–203, Nov. 2014.

[22] T. T. Li and D. H. Ji, "Sentiment analysis of micro-blog based on SVM and CRF using various combinations of features," *Appl. Res. Comput.*, vol. 32, pp. 978–981, Sep. 2015.

[23] H. Jin, R. Tong, and J. Ruiquan, "Sentiment analysis in financial domain based on SVM with dependency syntax," *Comput. Eng. Appl.*, vol. 51, no. 23, pp. 230–235, 2015.

[24] R. Liu, Y. Chen, X. Zhu, and K. Hou, "Image classification using label constrained sparse coding," *Multimedia Tools Appl.*, vol. 75, no. 23, pp. 15619–15633, 2016.

[25] H. Lu, Y. Li, M. Chen, H. Kim, and S. Serikawa, "Brain intelligence: Go beyond artificial intelligence," *Mobile Netw. Appl.*, vol. 23, no. 2, pp. 368–375, Apr. 2018.

[26] R. Liu, R. Yang, S. Li, Yuqian Shi, and X. Jin, "Painting completion with generative translation models," *Multimedia Tools Appl.*, pp. 1–14, Oct. 2018.

[27] H. Lu, D. Wang, Y. Li, J. Li, X. Li, H. Kim, S. Serikawa, and I. Humar, "CONet: A cognitive ocean network," Jan. 2019, *arXiv:1901.06253*. [Online]. Available: https://arxiv.org/abs/1901.06253

[28] H. Lu, Y. Li, S. Mu, D. Wang, H. Kim, and S. Serikawa, "Motor anomaly detection for unmanned aerial vehicles using reinforcement learning," *IEEE Internet Things J.*, vol. 5, no. 4, pp. 2315–2322, Aug. 2018.

[29] T. Mikolov, K. Chen, G. Corrado, and J. Dean, "Efficient estimation of word representations in vector space," Sep. 2013, *arXiv:1301.3781*. [Online]. Available: https://arxiv.org/abs/1301.3781

[30] 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.

[31] H. Lu, Y. Li, T. Uemura, H. Kim, and S. Serikawa, "Low illumination underwater light field images reconstruction using deep convolutional neural networks," *Future Gener. Comput. Syst.*, vol. 82, pp. 142–148, May 2018.

[32] Y. Kim, "Convolutional neural networks for sentence classification," Sep. 2014, *arXiv:1408.5882*. [Online]. Available: https://arxiv.org/abs/1408.5882

[33] A. Conneau, H. Schwenk, L. Barrault, and Y. Lecun, "Very deep convolutional networks for text classification," Jun. 2016, *arXiv:1606.01781*. [Online]. Available: https://arxiv.org/abs/1606.01781

[34] R. Johnson and T. Zhang, "Effective use of word order for text categorization with convolutional neural networks," Dec. 2014, *arXiv:1412.1058*. [Online]. Available: https://arxiv.org/abs/1412.1058

[35] D. Tang, B. Qin, and T. Liu, "Document modeling with gated recurrent neural network for sentiment classification," in *Proc. Conf. Empirical Methods Natural Lang. Process.*, 2015, pp. 1422–1432.

[36] M.-Y. Day and Y.-D. Lin, "Deep learning for sentiment analysis on Google play consumer review," in *Proc. IEEE Int. Conf. Inf. Reuse Integr. (IRI)*, Aug. 2017, pp. 382–388.

[37] R. Johnson and T. Zhang, "Supervised and semi-supervised text categorization using LSTM for region embeddings," May 2016, *arXiv:1602.02373*. [Online]. Available: https://arxiv.org/abs/1602.02373

[38] M. S. Akhtar, A. Kumar, A. Ekbal, and P. Bhattacharyya, "A hybrid deep learning architecture for sentiment analysis," in *Proc. COLING 26th Int. Conf. Comput. Linguistics: Tech. Papers*, Dec. 2016, pp. 482–493.

[39] Z. Yang, D. Yang, C. Dyer, X. He, A. Smola, and E. Hovy, "Hierarchical attention networks for document classification," in *Proc. Conf. North Amer. Chapter Assoc. Comput. Linguistics: Hum. Lang. Technol.*, 2016, pp. 1480–1489.

[40] D. Tang, B. Qin, and T. Liu, "Aspect level sentiment classification with deep memory network," Sep. 2016, *arXiv:1605.08900*. [Online]. Available: https://arxiv.org/abs/1605.08900

[41] K. Weiss, T. M. Khoshgoftaar, and D. D. Wang, "A survey of transfer learning," *J. Big Data*, vol. 3, no. 1, p. 9, 2016.

[42] Y. Zhang and Q. Yang, "A survey on multi-task learning," Jul. 2017, *arXiv:1707.08114*. [Online]. Available: https://arxiv.org/abs/1707.08114

[43] V. M. Patel, R. Gopalan, R. Li, and R. Chellappa, "Visual domain adaptation: An overview of recent advances," submitted for publication.

[44] D. Cook, K. D. Feuz, and N. C. Krishnan, "Transfer learning for activity recognition: A survey," *Knowl. Inf. Syst.*, vol. 36, no. 3, pp. 537–556, Sep. 2013.

[45] X. Glorot, A. Bordes, and Y. Bengio, "Domain adaptation for large-scale sentiment classification: A deep learning approach," in *Proc. 28th Int. Conf. Mach. Learn.*, 2011, pp. 513–520.

[46] M. Chen, Z. Xu, K. Weinberger, and F. Sha, "Marginalized denoising autoencoders for domain adaptation," Jun. 2012, *arXiv:1206.4683*. [Online]. Available: https://arxiv.org/abs/1206.4683

[47] M. Sun, Q. Tan, R. Ding, and H. Liu, "Cross-domain sentiment classification using deep learning approach," in *Proc. IEEE 3rd Int. Conf. Cloud Comput. Intell. Syst.*, 2014.

[48] B. McCann, J. Bradbury, C. Xiong, and R. Socher, "Learned in translation: Contextualized word vectors," in *Proc. Adv. Neural Inf. Process. Syst.*, 2017, pp. 6294–6305.

[49] M. E. Peters, M. Neumann, M. Iyyer, M. Gardner, C. Clark, K. Lee, L. Zettlemoyer, "Deep contextualized word representations," Mar. 2018, *arXiv:1802.05365*. [Online]. Available: https://arxiv.org/abs/1802.05365

[50] A. M. Dai and Q. V. Le, "Semi-supervised sequence learning," in *Proc. Adv. Neural Inf. Process. Syst.*, 2015, pp. 3079–3087.

[51] J. Howard and S. Ruder, "Universal language model fine-tuning for text classification," 2018, *arXiv:1801.06146*. [Online]. Available: https://arxiv.org/abs/1801.06146

[52] W. Dai, Q. Yang, G.-R. Xue, and Y. Yu, "Boosting for transfer learning," in *Proc. 24th Int. Conf. Mach. Learn.*, Jun. 2007, pp. 193–200.

[53] R. L. Vieriu, A. K. Rajagopal, R. Subramanian, O. Lanz, E. Ricci, N. Sebe, and K. Ramakrishnan, "Boosting-based transfer learning for multi-view head-pose classification from surveillance videos," in *Proc. 20th Eur. Signal Process. Conf. (EUSIPCO)*, Aug. 2012, pp. 649–653.

[54] S. Wang and Z. Li, "A new transfer learning boosting approach based on distribution measure with an application on facial expression recognition," in *Proc. Int. Joint Conf. Neural Netw. (IJCNN)*, Jul. 2014, pp. 432–439.

[55] R. Chattopadhyay, Q. Sun, W. Fan, I. Davidson, S. Panchanathan, and J. Ye, "Multisource domain adaptation and its application to early detection of fatigue," *ACM Trans. Knowl. Discovery Data*, vol. 6, no. 4, p. 18, Dec. 2012.

[56] R. Xu, J. Xu, and X. Wang, "Instance level transfer learning for cross lingual opinion analysis," in *Proc. 2nd Workshop Comput. Approaches Subjectivity Sentiment Anal.*, Jun. 2011, pp. 182–188.

[57] L. Gui, Q. Lu, R. Xu, Q. Wei, and Y. Cao, "Improving transfer learning in cross lingual opinion analysis through negative transfer detection," in *Proc. Int. Conf. Knowl. Sci., Eng. Manage.* Cham, Switzerland: Springer, Nov. 2015, pp. 394–406.

[58] S. Serikawa and H. Lu, "Underwater image dehazing using joint trilateral filter," *Comput. Elect. Eng.*, vol. 40, no. 1, pp. 41–50, 2014.

[59] W. Wang, L. Chen, K. Chen, K. Thirunarayan, and A. P. Sheth, "Adaptive training instance selection for cross-domain emotion identification," in *Proc. Int. Conf. Web Intell.*, Aug. 2017, pp. 525–532.

[60] J. Blitzer, R. McDonald, and F. Pereira, "Domain adaptation with structural correspondence learning," in *Proc. Conf. Empirical Methods Natural Lang. Process.*, Jul. 2006, pp. 120–128.

[61] S. J. Pan, X. Ni, J.-T. Sun, Q. Yang, and Z. Chen, "Cross-domain sentiment classification via spectral feature alignment," in *Proc. 19th Int. Conf. World Wide Web*, Apr. 2010, pp. 751–760.

[62] R. Xia, C. Zong, X. Hu, and E. Cambria, "Feature ensemble plus sample selection: Domain adaptation for sentiment classification," *IEEE Intell. Syst.*, vol. 28, no. 3, pp. 10–18, May 2013.

[63] G. Zhou, Y. Zhou, X. Guo, X. Tu, and T. He, "Cross-domain sentiment classification via topical correspondence transfer," *Neurocomputing*, vol. 159, pp. 298–305, Jul. 2015.

[64] J. T. Zhou, I. W. Tsang, S. J. Pan, and M. Tan, "Heterogeneous domain adaptation for multiple classes," in *Proc. 17th Int. Conf. Artif. Intell. Statist.*, Apr. 2014, pp. 1095–1103.

[65] A. L. Maas, R. E. Daly, P. T. Pham, D. Huang, A. Y. Ng, and C. Potts, "Learning word vectors for sentiment analysis," in *Proc. 49th Annu. Meeting Assoc. Comput. Linguistics, Hum. Lang. Technol.*, 2011.

[66] R. Socher, A. Perelygin, J. Wu, J. Chuang, C. D. Manning, A. Ng, and C. Potts, "Recursive deep models for semantic compositionality over a sentiment treebank," in *Proc. Conf. Empirical Methods Natural Lang. Process.*, 2013, pp. 1631–1642.

[67] J. Blitzer, M. Dredze, and F. Pereira, "Domain adaptation for sentiment classification," in *Proc. 45th Ann. Meeting Assoc. Comput. Linguistics*, 2007.

[68] A. Go, R. Bhayani, and L. Huang, "Twitter sentiment classification using distant supervision," Stanford, CA, USA, Tech. Rep. CS224N, 2009, vol. 1, no. 12.

[69] Z. Chen, A. Mukherjee, and B. Liu, "Aspect extraction with automated prior knowledge learning," in *Proc. 52nd Annu. Meeting Assoc. Comput. Linguistics*, vol. 1, 2014, pp. 347–358.

[70] T. A. Rana and Y.-N. Cheah, "A two-fold rule-based model for aspect extraction," *Expert Syst. Appl.*, vol. 89, pp. 273–285, Dec. 2017.

[71] R. M. Marcacini, R. G. Rossi, I. P. Matsuno, and S. O. Rezende, "Cross-domain aspect extraction for sentiment analysis: A transductive learning approach," *Decis. Support Syst.*, vol. 114, pp. 70–80, Oct. 2018.

[72] B. Tan, Y. Zhang, S. J. Pan, and Q. Yang, "Distant domain transfer learning," in *Proc. 31st AAAI Conf. Artif. Intell.*, Feb. 201, pp. 2604–2610.

[73] M. Van de Kauter, D. Breesch, and V. Hoste, "Fine-grained analysis of explicit and implicit sentiment in financial news articles," *Expert Syst. Appl.*, vol. 42, no. 11, pp. 4999–5010, 2015.

[74] K. Schouten and F. Frasincar, "Survey on aspect-level sentiment analysis," *IEEE Trans. Knowl. Data Eng.*, vol. 28, no. 3, pp. 813–830, Mar. 2016.

[75] 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.

[76] K. Ravi and V. Ravi, "A survey on opinion mining and sentiment analysis: Tasks, approaches and applications," *Knowl.-Based Syst.*, vol. 89, pp. 14–46, Nov. 2015.

[77] R. Y. K. Lau, C. Li, and S. S. Y. Liao, "Social analytics: Learning fuzzy product ontologies for aspect-oriented sentiment analysis," *Decis. Support Syst.*, vol. 65, pp. 80–94, Sep. 2014.

RUIJUN LIU received the M.S. degree from Beihang University, in 2009, and the Ph.D. degree from the Ecole Centrale de Nantes, France, in 2013. He is currently with Beijing Technology and Business University. His current research interests include machine learning, virtual reality, and 3D reconstruction.

YUQIAN SHI received the bachelor's degree in computer science and technology from Shenyang University, in 2017. She is currently pursuing the master's degree with the School of Computer and Information Engineering, Beijing Technology and Business University (BTBU), Beijing, China. Her current research interests involve natural language processing.

CHANGJIANG JI received the bachelor's degree from the School of the Gifted Young, University of Science and Technology of China. He currently serves as the Deputy General Manager with the Innovation, Research and Development Center, Beijing Moviebook Technology Corporation Ltd. His current research interests include mobile game and application development and Internet software design and development in the field of videos.

MING JIA received the master's degree in information technology from Edith Cowan University, WA, in 2005. He is currently with Beijing Technology and Business University. His current research interests include information management and virtual simulation.