IEEE Access
Multidisciplinary : Rapid Review : Open Access Journal

Received August 16, 2018, accepted September 27, 2018, date of publication October 10, 2018, date of current version October 29, 2018.

Digital Object Identifier 10.1109/ACCESS.2018.2874623

Cross-Domain Sentiment Classification by Capsule Network With Semantic Rules

BOWEN ZHANG[1], XIAOFEI XU[1], MIN YANG[2], XIAOJUN CHEN[3], AND YUNMING YE[4]

[1]School of Computer Science and Technology, Harbin Institute of Technology, Harbin 150001, China
[2]Shenzhen Institutes of Advanced Technology, Chinese Academy of Sciences, Shenzhen 518055, China
[3]College of Computer Science and Software, Shenzhen University, Shenzhen 518060, China
[4]School of Computer Science and Technology, Harbin Institute of Technology, Shenzhen 518055, China

Corresponding author: Yunming Ye (yeyunming@hit.edu.cn)

This work was supported in part by NSFC under Grant 61832004, Grant 61572158, and Grant 61602132, and in part by the Shenzhen Science and Technology Program under Grant JCYJ20160330163900579.

ABSTRACT Sentiment analysis is an important but challenging task. Remarkable success has been achieved on domains where sufficient labeled training data is available. Nevertheless, annotating sufficient data is labor-intensive and time-consuming, establishing significant barriers for adapting the sentiment classification systems to new domains. In this paper, we introduce a Capsule network for sentiment analysis in domain adaptation scenario with semantic rules (CapsuleDAR). CapsuleDAR exploits capsule network to encode the intrinsic spatial part-whole relationship constituting domain invariant knowledge that bridges the knowledge gap between the source and target domains. Furthermore, we also propose a rule network to incorporate the semantic rules into the capsule network to enhance the comprehensive sentence representation learning. Extensive experiments are conducted to evaluate the effectiveness of the proposed CapsuleDAR model on a real world data set of four domains. Experimental results demonstrate that CapsuleDAR achieves substantially better performance than the strong competitors for the cross-domain sentiment classification task.

INDEX TERMS Cross-domain sentiment classification, capsule network, semantic rules, deep learning.

I. INTRODUCTION

With the growth of the large collection of opinion-rich resources, much attention has been given to sentiment classification, which aims to automatically predict the sentiment polarity of a piece of text. Sentiment classification has been one of the most popular research fields due to its broad applications in brand monitoring, customer services, market research, politics and social sciences. For example, companies can develop strategies by mining users' attitudes toward product reviews.

Existing sentiment classification methods can be divided into two categories based on the knowledge and information they use: lexicon-based methods [1]–[4] and corpus-based methods [5]–[8]. The lexicon-based approach counts negative and positive words in the text relying on the sentiment lexicon, and assigns the sentiment polarity of the text as positive if the number of positive words is larger than that of the negative words. In contrast, the corpus-based method utilizes machine learning algorithms to train a sentiment classifier. The performance of corpus-based methods is often more superior than that of lexicon-based methods when the labeled training data is sufficient. In this paper, we mainly concentrate on the corpus-based approach.

The main idea of the conventional corpus-based approaches is to employ machine learning approaches such as support vector machine (SVM) [9], naive Bayes [10], decision tree [11], and logistic regression [12] as text classier to predict the sentiment polarity of the given texts. The success of these machine learning algorithms generally relies heavily on feature engineering which is labor-intensive and highlights the weakness of the conventional corpus-based sentiment classification algorithms.

Inspired by the recent success of deep learning in computer vision and natural language processing [13], the deep neural networks (e.g., convolutional neural network and long short-term memory network) have become dominant in the literature. Remarkable success has been achieved by the previous studies on domains where a large number of labeled data is available. Nevertheless, such impressive performance relies on the assumption that the training data and the test data should come from the same underlying data distribution, establishing significant barriers for adapting the sentiment classification methods to new domains. To avoid annotating sufficient instances when dealing with the new domain data, it is important to explore cross-domain sentiment classification algorithms.

Recently, several techniques have been developed for cross-domain sentiment classification. Structural correspondence learning (SCL) [1], [14] is a prominent domain adaptation approach. Pan *et al.* [2] proposed Spectral Feature Alignment (SFA) to achieve cross-domain classification by aligning pivots with non-pivots. To take the advantage of deep neural networks, Glorot *et al.* [15] proposed a Stacked Denoising Auto-Encoder (SDA) that automatically learns common feature representations from large amounts of data in different domains. Zhou *et al.* [16] proposed a deep learning method for learning feature mapping between cross-domain heterogeneous features. Subsequently, Ziser and Reichart [17] proposed a automatic encoder-SCL model, which combines the SCL method with the autoencoder model. This model is proven to be superior to previous methods. Generally, the aforementioned cross-domain sentiment classification method not only extract features from the source domain but also learn the shared features that bridge the knowledge gap across domains.

Despite the effectiveness of prior work, cross-domain sentiment classification in real-world remains a challenge for several reasons. (1) First, most of the conventional cross-domain sentiment classification methods rely heavily on hand-crafted features, such as the pivot features. Such feature engineering process usually requires comprehensive human domain expertise, which is time-consuming and cost-intensive. In order to ease the applicability of cross-domain sentiment classification, it would be highly desirable to develop cross-domain algorithms that are less dependent on feature engineering, so that the sentiment classifier for new domains could be constructed faster. (2) In neural network approaches, spatial patterns aggregated as lower levels contribute to representing higher level concepts. Here, they form a recursive process to articulate what to be modeled. However, these models cannot encode the intrinsic spatial part-whole relationship constituting viewpoint invariant knowledge that can automatically generalize to novel viewpoints, which limits the transferability of deep neural networks from the source domain to the target domain.

To alleviate the aforementioned limitations, in this paper, we propose a Capsule network for sentiment analysis in Domain Adaptation scenario with semantic Rules (CapsuleDAR). Capsule network is originally introduced by [18]. Each capsule in capsule network is an aggregation of neurons, where each neuron represents various attributes of a particular feature presented in the text. These attributes can be different types of instantiation parameters, such as n-gram features, position information of words or phrases, syntactic features of sentences, etc. A metaphor (also as an argument) they made is that human visual system intelligently assigns parts to wholes at the inference time without hard-coding patterns to be perspective relevant. Thus, capsule network has much stronger transferring capability than conventional deep neural networks and traditional machine learning algorithms. The proposed CapsuleDAR model consists of two capsule networks, i.e., Base network and Rule network, which have a similar network structure.

The Base network works on the textual features of the sentence. Its goal is to leverage capsule network to encode the intrinsic spatial part-whole relationship constituting domain invariant knowledge that automatically generalizes to novel domains, bridging the knowledge gap between the source and target domains. Furthermore, we also propose a Rule network to incorporate the semantic rules (such as the structure of the sentence) into the capsule network to further improve the performance of the cross-domain sentiment classification. Specifically, the first rule is to leverage the pivot features between source and target domains to enhance the performance of the convolution filter. The second rule is to exploit the sentence structure information to enrich the comprehensive representation learning of sentences for cross-domain sentiment classification.

We summarize our main contributions as follows:

1. To the best of our knowledge, this is the first work dealing with sentiment classification in the *domain adaptation scenario* using capsule network. Capsule network shows strong capability of transferring common knowledge from source domain to target domain.
2. We propose a Rule network to incorporate the semantic rules into the capsule network to capture the common knowledge across domains.
3. Extensive experiments have been conducted on a real-world data set of 4 domains to evaluate the effectiveness of CapsuleDAR model for cross-domain sentiment classification. The experimental results demonstrate that the proposed CapsuleDAR model achieves substantial improvements over the compared methods.

The remainder of this paper is organized as follows. Section II reviews and discusses the related work, including some traditional and recent methods of generic sentiment classification, cross-domain sentiment classification, and capsule networks. In Section III, we fully describe the proposed model. The experimental setup is introduced in Section IV, including the evaluation datasets, the compared methods, the automatic evaluation metrics, and the implementation details. Section V shows the quantitative evaluation results and analysis. Section VI makes the conclusions and discusses the future work.

II. RELATED WORK

A. CROSS-DOMAIN SENTIMENT CLASSIFICATION

Sentiment classification is a popular research area, which has gained much attention from both academia and industry [5]–[7], [19], [20] and [21] introduced most previous techniques and datasets for sentiment analysis. In online social networks, the large volume of data sources make it costly and difficult to build a robust and generalized sentiment classifier across domains. This motivates many studies that analyze the sentiments of cross-domain textual data.

Most previous work belongs to the feature-based transfer, requiring manual selection of the pivot or non-pivot features. Among them, the structural correspondence learning (SCL) introduced by Blitzer *et al.* [14] is the representative one, which tried to obtain the mapping matrix from non-pivot feature space to pivot feature space. Pan *et al.*, [2] introduced the SFA method, which aimed to establish a bridge between the source domain and the target domain by aligning the pivot features and the non-pivot features of different domains. Tan *et al.* [10] tried to select the generalizable features which occurred frequently in different domains and had similar probabilities. In [22], a Bayesian probabilistic model is proposed to deal with the data from multiple sources and domains. The above work all needs a large amount of unlabeled data in the target domain to help build the transfer procedure. In addition, these methods do not fully explore the semantics of the words and exploit the data as well as domain labels.

Recently, deep learning models are proposed to learn common features and shared parameters for sentiment classification in domain adaption scenario, and these models have yielded impressive results. Glorot *et al.* [15] introduced a stacked denoising auto-encoder (SDA) to learn uniform abstractive feature representations for the documents from both source and target domains. Chen *et al.*, [23] proposed the mSDA algorithm, which retained the powerful feature learning ability and solved the high computational cost and scalability issues of SDA. Subsequently, many extensions of the SDA method were proposed to further improve the performance of cross-domain sentiment classification [24]–[26]. However, the previous method lacks explanation ability, that is, it is impossible to demonstrate whether the network has sufficient ability to learn the pivot features.

In order to improve the interpretability of deep models, Ziser and Reichart, [17] proposed AE-SCL-SR method which combined the autoencoder and pivot-based method for cross-domain classification. Li *et al.*, [6] introduced AMN method(IJCAI-17) to employ the attention mechanism to automatically capture the pivot without manual intervention. In order to improve classification ability, Ziser and Reichart, [7] proposed a PBLMs method which combined LSTM and CNN with pivot features.

B. CAPSULE NETWORKS

Recently, a novel type of neural network is proposed using the concept of capsules to improve the representational limitations of CNN and RNN. Hinton *et al.* [18] firstly introduced the concept of "capsules" to address the representational limitations of CNNs and RNNs. Capsules with transformation matrices allowed networks to automatically learn part-whole relationships. Consequently, Sabour *et al.* [27] proposed capsule networks that replaced CNNs with vector-output capsules. In addition, the max-pooling operation in CNNs is also replaced with a novel routing-by-agreement algorithm. The capsule network has shown its potential by achieving a state-of-the-art result on MNIST data. Xi *et al.* [28] further tested out the application of capsule networks on CIFAR data with higher dimensionality. Hinton *et al.* [29] proposed a new iterative routing procedure between capsule layers based on EM algorithm, which achieves significantly better accuracy on the smallNORB data set. Wang *et al.* [30] proposed a capsule model based on Recurrent Neural Network (RNN) for sentiment analysis. Given an instance encoded in hidden vectors by a typical RNN, the representation module builds capsule representation by the attention mechanism. Zhao *et al.* [31] applied capsule network in text classification. They showed that capsule networks outperform strong baseline methods in text classification.

To date, no work investigates the performance of capsule networks in cross-domain sentiment classification. This study takes the lead in this topic.

III. OUR METHODOLOGY

Our model, depicted in Figure 1, is a variant of the capsule networks proposed in Sabour *et al.*, [27]. It consists of two networks: a basic network and a rule network, to improve the performance of cross-domain sentiment classification. In the rest of this section, we first give the problem definition and the overview of our model in Section III.A and Section III.B, respectively. Then, we elaborate the basic network and the rule network in detail in Section III.C. Finally, the training process of our model is introduced in Section III.D.

A. PROBLEM DEFINITION

We use $X^{sl} = \{x_i^{sl}\}_{i=1}^{N^{sl}}$ to denote the collection of labeled documents in the source domain, where N^{sl} is the number of samples in source domain. Each document $x^{sl} \in X^{sl}$ has a sentiment label y^{sl} which is a one-hot representation of the correct label. In the target domain, we are given an unlabeled dataset $X^{tu} = \{x_i^{tu}\}_{i=1}^{N^{tu}}$, where N^{tu} is the number of samples in target domain. The goal of our model is to predict the sentiment polarity of samples in target domain using the sentiment classifier pre-trained on the source domain data.

B. FRAMEWORK OVERVIEW

As shown in Figure 1, our overall architecture consists of two main components: *Base* network and *Rule* network. The *Base* network employs a capsule network trained with textual features to perform sentiment prediction. It contains an embedding layer, a convolutional layer for capturing n-gram features of the text, an Incaps and a Outcaps layers followed by a Classcaps layer. The *Rule* network leverages semantic rules into the capsule network, which uses the common knowledge to bridge the knowledge gap between source and target domains. The structure of Rule network is similar to that of the Base network, and the parameters of the convolutional layers are shared by these two networks. In the Rule network,

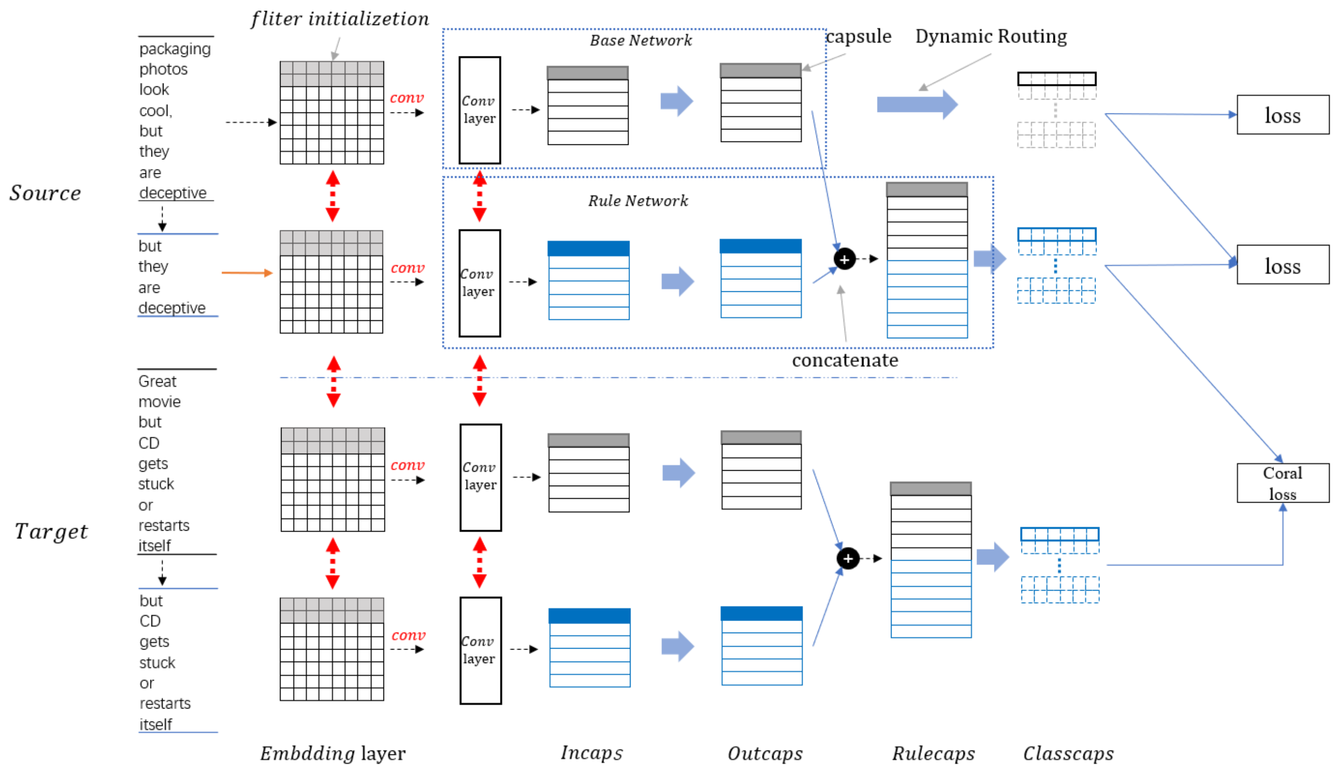


FIGURE 1. The Architecture of the proposed CapsuleDAR model for cross-domain sentiment classification.

we have an additional Rulecaps layer between Outcaps and Classcaps layers to fully exploit the semantic rules.

C. OUR METHODOLOGY

Our model consists of two components: Base network and Rule network. Next, we will elaborate each component of our model in detail.

1) BASE NETWORK

The Base network is employed to model the text content features.

a: Embedding Layer

The first layer is the embedding layer. Given input sequence x, we first convert the i-th word into a low-dimensional vector representation $e_i \in \mathbb{R}^d$ by embedding layer, where d donates the dimension of embedded vectors. We donate the embedding of the sentence as $e \in \mathbb{R}^{m \times d}$, where the length as m, is the concatenation of each word vectors:

$$e_{1:m} = e_1 \oplus e_2 \oplus \ldots \oplus e_m \tag{1}$$

where $\oplus$ is the concatenation operator.

b: Conv Layer

The second layer is the convolutional layer which consists of one convolutional operation to extract n-gram features of the input sequence through the convolutional operations. For sentiment analysis, it is essential for the network to learn the n-gram features, such as "not bad" for the data from both source and target domains.

Suppose the vector $W \in \mathbb{R}^{k \times d}$ is the filter of the convolution, where k is the filter width. A filter with k width enables the convolution layer to slide over the input sequence and obtain a new feature. Formally, a feature z_i is learned from a local window of word sequence $e_{i:i+k-1}$ by

$$z_i = \sigma(W \odot e_{i:i+k-1} + b) \tag{2}$$

where b is the bias vector, and $\odot$ is a convolutional operator. σ represents a non-linear hyperbolic tangent function. This convolution filter is applied to every possible window of words in the input sequence $\{e_{1:k}, e_{2:k+1}, \ldots, e_{m-k+1:m}\}$ to produce a feature map $\mathbf{z} \in \mathbb{R}^{n-k+1}$, computed by:

$$\mathbf{z} = [z_1, z_2, \ldots, z_{m-k+1}] \tag{3}$$

Here, the filter weights and bias terms of each filter are shared among all locations in the input, preserving spatial locality. We then apply a max-pooling operation over the feature map and obtain the maximum value $\hat{z} = max\{\mathbf{z}\}$ as the learned feature with respect to the k-length filter. The main idea behind this max operation is to extract the most salient n-gram feature for each feature map. In addition, the max pooling operation naturally handle the problem of variable sentence lengths. We use multiple filters (with different window sizes) to learn multiple features, thus the final can be represented as $Z = [\hat{z}_1, \ldots, \hat{z}_l]$, where l is the number of filters.

We believe that there are two types of features that are important for cross-domain sentiment classification: (i) the

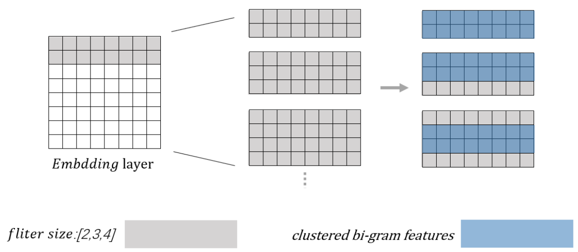


FIGURE 2. The process of filter initialization.

n-gram features that are essential for the sentiment classification in source domain; (ii) and the pivot features across the source domain and the target domain, which can be utilized to bridge the gap of knowledge between the source domain and the target domain. Thus, we introduced a **pivot-based filter initialization method** to improve the ability of text expression of our method instead of using random initialization filters for the convolutional layer.

In particular, we first use the tf-idf [32] method to select the n-grams with the highest contribution for each sentiment class in source domain. Then, we employ the SCL method [7] to select pivot features. Following the strategy in [6], given a collection of the extracted tf-idf features and pivot features, we use k-means method to cluster these features, and the centroid vectors of the clusters are used to initialize the filter weights. Figure 2 shows the example of the bi-gram filter embedding process, where the gray area shows the filters of width {2,3,4} with random parameters and blue area donates the centroid vectors of the clusters.

c: Incaps and Ourcaps Layers

One capsule is a collection of neurons that represent instantiation parameters for a particular type of object. One advantage of capsules is that they provide a effective way to resemble the human perception system, which identifies the part-whole relationship.

Incaps is the first capsule layer, which is constructed by the data transformation from the Conv layer. This transformation allows the output of the Conv layer to be directly used as input to the dynamic routing method. Outcaps is the second capsule layer, which is produced by Incaps layer via a routing method.

The main idea of dynamic routing [27] algorithm is to find $\mathcal{A}$ denoting a set of coupling coefficients. Each item $\alpha_{i,j}$ determines a mapping relationship between two capsule layers: u_i and v_j. The coupling coefficient determines the correlation between the j and i layers. Inspired by [33], we denote the routing process as a minimized agglomerative fuzzy K-Means like clustering loss function, defined as:

$$min_{\mathcal{A},\mathcal{V}} \left\{ \mathcal{L}(\mathcal{A},\mathcal{V}) := -\sum_i \sum_j \alpha_{ij} < W_j T_i u_i, v_j > + \lambda \sum_i \sum_j \alpha_{ij} log \alpha_{ij} \right\},$$
$$s.t.\ \alpha_{ij} > 0, \sum_j \alpha_{ij} = 1, ||v_j|| \leqslant 1 \tag{4}$$

where $<,>$ represents the inner product, $\mathcal{V}$ donates the collection of higher layer capsules $\mathcal{V} = \{v_1, v_2, v_3 \ldots, v_c\}$. An effective way to solve this problem is to use an optimized coordinate descent to optimize $\mathcal{A}$ and $\mathcal{V}$.

The length of capsules can represent the probability of the presence of the corresponding feature. After obtaining the high-level capsule feature, we perform a nonlinear squash operation on the high-level capsule ensuring that the direction of the vector is constant.

$$squash(v_j) = \frac{\|v_j\|^2}{\xi + \|v_j\|^2} \frac{v_j}{\|v_j\|} \tag{5}$$

where ξ is the soft parameter. We summarize the dynamic routing method in Algorithm 1.

Algorithm 1 Dynamic Routing

```
for all capsule i in layer l and j in l+1:
initial: s_j ← Σ_{i=1}^n W_j T_i u_i,
for iteration t do
    s_j ← s_j/||s_j||,
    b_ij ←< W_j T_i u_i, s_j >,
    c_ij ← softmax(b_ij),
    s_j ← Σ_i c_ij W_j T_i u_i,
for all capsule j: v_j = squash(s_j),
return v_j;
```

d: Classcaps

The last layer for Base network is the Classcaps layer. Classcaps layer is constructed from Outcaps using dynamic routing. The number of capsules in this layer is dependent on the category of the sentiment classification task, where the length of the capsule represents the probability of the existence of each class.

2) RULE NETWORK

Each neuron in a capsule not only represents the features in sentence level (words, entities, lengths, etc.,), but also represents whether this instance contains particular structure (e.g., sentence structure) and how each element in this structure contributes to the sentiment classification. In this section, we designed a **Rule network** that allows the network to learn from the sentence structures.

Different from the Base network, the Rule network is employed to model the sentence structure features. The basic structure of Rule network is similar to that of the Base network. In particular, the first and second layers are embedding and conv layers respectively, followed by two capsule layers: Incap and Outcaps layers. In the Rule network, we have an additional Rulecaps layer between Outcaps and Classcaps layers, which is different from the Base network.

In Rule network, the input of Embedding layer is the clause behind the transition words. So in the Outcaps layer, the learned information is a textual representation that includes the particular structure of sentences. Since the embedding layer, conv layer, Outcaps and Classcaps in Rule network are similar to that of the Base network, we will mainly explain the unique Rulecaps layer of Rule network in this section.

Rulecaps: As illustrated in Figure 1, the input of Rulecaps is the output of the Outcaps layers from both Base and Rule networks. Specifically, the output of Outcaps layers from both Base and Rule networks are connected in parallel, and we then send it into a dynamic routing algorithm to obtain the final Classcaps layer for Rule network. Thus, the capsules of the final output layer will consist of the feature information of the entire sentence and clause.

However, in the process of constructing Rulecaps with dynamic routing algorithm (as illustrated in Algorithm 1), the large iteration value t will lead to over-fitting of sentence structure information. Therefore, we add a temperature parameter into the softmax operations in Algorithm 1, which increases the contribution of these clauses to higher-level capsule networks.

$$p_i = \frac{e^{(b_i/T)}}{\sum_j e^{(b_j/T)}} \tag{6}$$

where b_i represents the output of softmax layer, representing the probability of each class. p_i is the standard softmax function when T is set to 1. Setting a higher value of T can generate a smoother probability distribution on each class.

D. TRAINING OBJECTIVE

The training objective of our method consists of two parts. One is for the sentiment classifier, another is for minimizing the distance between the source and target features. We use the margin loss function for sentiment classification objective and use deep CORAL loss for minimizing the difference between source and target domains.

1) SENTIMENT CLASSIFICATION LOSS

Inspired by CapsNet proposed by Sabour *et al.*, [27], we design a two-branch object functions that integrate the knowledge of sentence structures into the networks. For sentiment classification, the goal of the first branch is to maximize the active probability of correct sentiment for the Base Network. We calculate this objective (denoted as $\mathcal{L}^s$) by hinge loss:

$$\mathcal{L}^s(\theta) = \sum_{p=1}^{N} max(0, 1 - T_p C_p^s) \tag{7}$$

where C_p^s is the first branch output layer of capsules index by p. T_p is a indicator such that $p = 1$ if the capsule p is active, and otherwise $p = 0$. Similar to the first objective, the aim of the second objective $\mathcal{L}^c$ is to maximize the probability of the output layer capsule C_p^c for Rule Network. In particular, this output includes information learned by the structured knowledge.

$$\mathcal{L}^c(\theta) = \sum_{p=1}^{N} max(0, 1 - T_p C_p^c) \tag{8}$$

TABLE 1. Transition words list.

yet	and yet
nevertheless	after all
nonetheless	but
however	otherwise
on the contrary	in contrast
on the other hand	though
notwithstanding	

The final sentiment classification Objective function $\mathcal{L}$ is obtained by adding the above two parts:

$$\mathcal{L}(\theta) = \mathcal{L}^s(\theta) + \mathcal{L}^c(\theta) \tag{9}$$

2) CORAL LOSS

In this paper, we use CORAL [34] loss to minimize the feature differences between source and target domains. CORAL is an effective method for measuring the distances of distributions between two domains. Recall that we denote labeled training instances as $X^{sl} = \{x_i^{sl}\}$ (with N^{sl} instances), and unlabeled target data as $X^{tu} = \{x_i^{tu}\}$ (with N^{tu} instances).

Suppose that M_s and M_t indicate the feature covariance matrices. The distance between the second-order statistic of the source and target features are treated as CORAL loss:

$$\mathcal{L}^{coral} = \frac{||M_s - M_t||^2_{\mathcal{F}}}{4d^2} \tag{10}$$

where $||\ldots||_{\mathcal{F}}$ represents the Frobenius norm and the covariance matrix of two domains data are computed by:

$$M_s = \frac{(X^{sl\top}X^{sl} - (1/N^{sl})(I^\top X^{sl})^\top(I^\top X^{sl}))}{N^{sl} - 1} \tag{11}$$

$$M_t = \frac{(X^{tu\top}X^{tu} - (1/N_{tu})(I^\top X^{tu})^\top(I^\top X^{tu}))}{N_{tu} - 1} \tag{12}$$

where I indicates the column vector that all equal to 1.

Overall, the objective function for cross-domain sentiment classification is:

$$\mathcal{J}(\theta) = \mathcal{L}(\theta) + \mathcal{L}^{coral} \tag{13}$$

Finally, we utilize ADADELTA optimization algorithm with the minibatch strategy to update the parameters of our model.

IV. EXPERIMENTAL SETUP

A. EXPERIMENTAL DATA

In this study, we conduct extensive experiments on the Amazon Review dataset [14] to illustrate the effectiveness of the proposed CapsuleDAR model for cross-domain sentiment classification task. Following Blitzer *et al.*, [14], we use the reviews from four product domains, including DVD (D), Kitchen (K), Books (B) and Electronics (E), that have been widely applied to evaluate the performance of sentiment classification in domain adaptation scenario. In total, there are 2000 labeled reviews (1000 positive and 1000 negative) given for each domain. Furthermore, a large-scale unlabeled dataset

TABLE 2. The classification results for cross domain sentiment analysis.

	Source-Target	B-E	D-E	K-E	B-D	E-D	K-D	B-K	E-K	D-K	E-B	K-B	D-B	ALL
		Our method												
1	CapsuleDAR	**88.3**	**89.3**	**91.4**	**90.1**	**88.0**	**86.5**	**89.4**	**93.2**	**90.0**	**84.4**	**83.6**	**89.6**	**88.6**
		State-of-the-art domain adaptation models												
2	PBLM-CNN	77.6	79.6	87.1	84.2	75.0	79.8	82.5	87.8	83.2	71.4	74.2	82.5	80.4
3	PBLM-LSTM	74.5	80.4	85.4	82.6	77.6	78.6	80.9	87.1	83.3	70.8	73.5	80.5	79.6
4	AMN	81.2	80.5	82.5	81.3	76.7	80.1	82.5	90.0	83.6	77.5	78.9	82.8	80.7
5	AE-SCL-SR	77.9	78.1	84.0	81.1	74.5	76.3	80.1	84.6	80.3	71.2	73.0	77.3	78.2
6	DACNN	79.5	77.0	81.5	77.5	75.9	73.1	78.9	82.8	79.5	73.7	73.5	77.1	77.5
7	DANN	68.0	74.5	82.1	73.7	71.0	71.4	78.8	84.5	77.6	70.0	71.2	75.0	74.8
8	DAmSDA	74.9	75.0	82.4	79.7	73.1	73.8	75.4	85.0	77.4	71.9	70.0	76.1	76.2
9	SS-FE	74.2	77.1	82.9	79.1	74.6	75.7	78.1	84.9	77.8	72.9	72.9	80.4	77.5
10	SVMmSDA	76.6	73.9	86.1	83.0	77.0	78.8	82.1	84.7	84.2	76.2	76.9	82.6	80.2
11	SCL-MI	71.9	71.5	82.2	78.8	70.4	72.2	77.2	82.9	74.0	68.5	69.3	73.2	74.3
		Baseline methods without domain adaptation												
12	CAPSULENoDA	84.0	86.5	86.9	87.6	86.6	84.7	84.7	90.2	87.3	81.9	81.2	87.9	85.8
13	CNN	72.1	69.7	79.9	73.6	67.1	70.8	72.7	80.6	72.6	65.6	66.5	71.2	71.9
14	LSTM	65.9	68.3	78.2	72.8	68.1	66.2	72.1	80.6	70.5	67.9	67.5	69.2	70.6

is also provided (34741 reviews for D, 6000 reviews for B, 13153 reviews for E and 16785 reviews for K).

Following the common settings in previous cross-domain sentiment classification studies [2], we constructed 12 cross-domain sentiment classification tasks ($K \rightarrow E, K \rightarrow D, K \rightarrow B, D \rightarrow E, D \rightarrow B, D \rightarrow K, E \rightarrow D, E \rightarrow K, E \rightarrow B, B \rightarrow E, B \rightarrow D, B \rightarrow K$). Where the left side of the arrow corresponds to the source domain and right side of the arrow is the target domain. For each source domain for train, we randomly selected 800 samples of positive and negative respectively. For the test set, we randomly select 400 reviews(200 positive and 200 negative) from target domain, The remaining 1600 reviews are used to minimize the difference between the source domain and the target domain during training process.

B. MODEL CONFIGURATIONS

We summarize the implementation details of our model as follows:

1) **sentence structure**: Before training the model, we preprocess the data through the sentence transition structure, and extracted the clauses after the transition words as part of the training sample. Transition words, for example "however", are strong indicators, which imply that the words following the transition words generally carry the whole sentiment of the review. For instance, suppose a sentence contains two clauses, connected by the transition word "however", the clauses behind "however" represent the sentiment of the whole sentence. The common transition words list is given in Table 1.
2) **hyper-parameter adjustment**: In all the experiments, we set the kernel size for filters of the convolutional layer as {1, 2, 3, 4, 5} respectively and the number of filters is set to 100. For filters initialization, 500 pivots feature were chosen. We utilize 300-dimensional word2vec trained on English Google News corpus to initialize the word embeddings. Other weight parameters are initialized by randomly sampling the values from the uniform distribution U(−0.01, 0.01). We select the number of iterations for dynamic routing between 1-5. The model is optimized with the ADADELTA optimization algorithm with batch-size 64 and decay rate 0.95 [35].

C. BASELINE METHODS

To fully estimate the effectiveness of our model, we compare CapsuleDAR model with several strong baselines for sentiment classification in domain adaptation scenario:

- **SCL-MI** [14]: It is a structural correspondence learning (SCL) for learning distributed pivot feature representations from both the source domain and the target domain. Following [14], the number of pivot features are chosen from the range of 500 and 1000, and we set the SVD size as one of the value in (50, 100, 150).
- **SS-FE** [36]: This method uses the principal component analysis (PCA) to select important features and sent them to a sample Naive Bayes method. We choose the number of principal components with the percentage of variance contribution that is larger than 99.5 percent.
- **DANN** [5]: It applies a generative adversarial network (GAN) framework to the cross-domain sentiment analysis with neural network. Following [5], the adaptation parameter is selected between 10^{-2} and 1 on a logarithmic scale. We set size of the hidden layer l to either 50 or 100. the learning rate is 10^{-3}.
- **SVM** [5]: It uses support vector machine (SVM) with a linear kernel based on mSDA representations. For SVM, the hyper-parameter C is selected between 10^{-5} and 1.
- **DACNN** [6]: It is a variant of DANN, which uses convolutional neural network to replace the full connect neural networks. The parameters are the same with [37].
- **DAmSDA** [5]: It is also an extension of DANN method. The feature representation obtained from Marginalized Stacked Denoising Autoencoders (**mSDA**) are

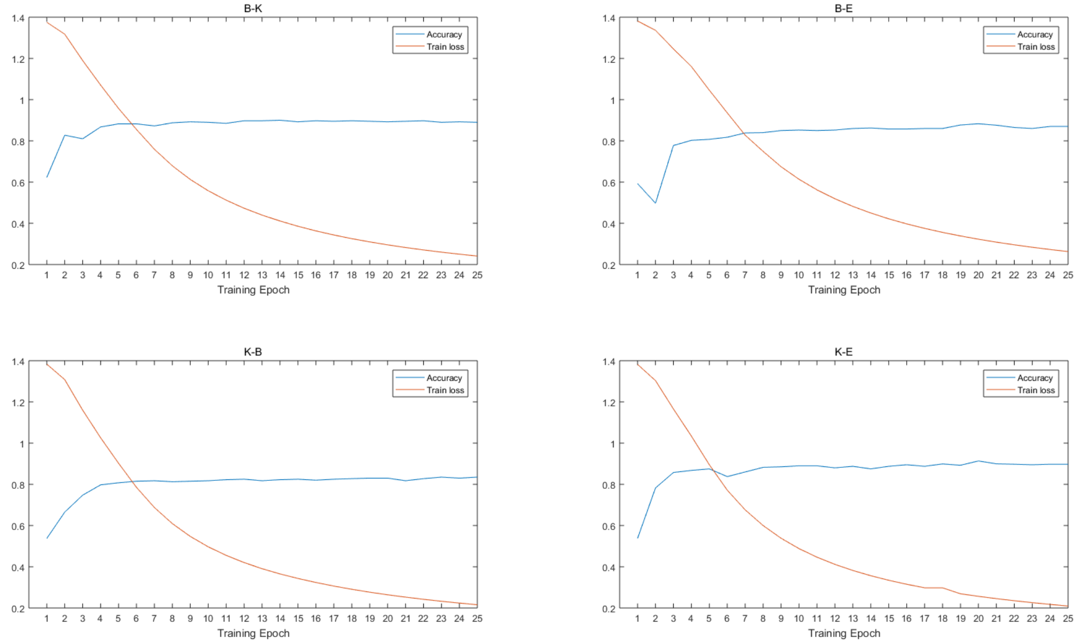


FIGURE 3. Loss vs. Accuracy.

TABLE 3. Ablation test results of our model on three kinds of domain setups.

Method	B-E	E-D	K-B
CapsuleDAR	88.3	88.0	83.6
w/o sentence structure	88.1	87.9	83.2
w/o weight initialization	87.5	87.3	83.1
w/o both rules	86.7	86.9	82.1
w/o capsule	81.5	80.7	79.0

exploited. Each instance is encoded into a condense vector of 30000 dimensions.

- **AE-SCL-SR** [17]: This model combines the benefits of autoencoder and structural correspondence learning to boost the performance of cross-domain sentiment classification. The learning rate for stochastic gradient descent(SGD) is set to 0.1. The decay weight is set to 10^{-5}. The number of pivot features is selected from (100,200,300,400,500).
- **PBLM** [7]: It utilizes a DNNs framework (CNN and LSTM) for cross-domain sentiment classification based on the pivot features. The hyperparameters are set to the same values as in [17]. Especially, the kernel of the size of convelutional layer is 3 for CNN.
- **CapsuleNoDA**: This model is a variant of our model without performing domain adaptation. Which is only training on source domain and test on target domain directly.

V. EXPERIMENTAL RESULTS

In this section, we report the model comparison from both quantitative and qualitative perspectives.

TABLE 4. The experimental results with respect to varying iteration numbers.

iteration No.	B-E	E-D	K-B
1	**88.3**	85.9	83.1
2	87.9	86.5	83.3
3	87.8	87.2	**83.6**
4	87.8	87.3	82.9
5	87.2	**88.0**	81.5
6	87.1	87.7	81.8
7	86.9	87.1	81.3
8	86.5	86.8	81.3
9	86.5	86.6	81.1
10	86.4	86.5	81.0

A. CROSS-DOMAIN SENTIMENT CLASSIFICATION RESULTS

In our experiments, the automatic evaluation metric is classification accuracy, which is widely adopted in sentiment classification. We summarize the experimental results in Table 2. From the results, we can observe that our model consistently and substantially outperforms the compared baseline methods, and achieves the state-of-the-art results on all the 12 setups. For example, our model gains 7.9% improvement over AMN (the best competitor) on average. The standard CNN and LSTM without performing domain adaptation perform poorly since they do not consider the difference between source and target domains and exploit the benefit of large-scale source domain data. The pivot-based neural networks, such as PBLM-CNN, AMN, AE-SCL-SR, DACNN, stably exceed the standard LSTM and CNN methods by a significant margin, which verifies the effectiveness of the pivot-based features in cross-domain sentiment classification.

TABLE 5. Visualization of pivot features from source and target domains.

Kitchen Domain	Books Domain
Negative	
dreadful terrible ergonomics! it was so difficult and uncomfortable to keep the trigger engaged and carve at the same time my hand was aching by the time i'd carved the brisket i've never used a more awkward or uncomfortable appliance	horrible book, horrible this book was horrible if it was possible to rate it lower than one star i would have i am an avid reader and picked this book up after my mom had gotten it from a friend i read half of it, suffering from a headache the entire time , and then got to the part about the relationship the 13 year old boy had with a 33 year old man and i lit this book on fire one less copy in the world don't waste your money i wish i had the time spent reading this book back so i could use it for better purposes this book wasted my life
they rust ! big disappointment i have probably seen hundreds of tea kettles by now and still love the look of this one the best only problem is that they rust! i had mine a few weeks when i noticed the rust spots on the interior i don't imagine the enamel could have been chipped on the inside of the pot, so it must have happened simply because i leave water in my tea kettle at all times the directions that come with this kettle say not to so, but any serious tea drinker like me knows's i threw mine and am very disapointed that something that cost so much money didn't last long at thought le creuset was supposed to really high quality but i'm disillusioned thanks to this tea pot.	this book is only for the apprentice lover i gave this book two stars because the book was a little boring and dragged on in some areas the best information was in the margins and the quotes not really any of the information in the book i have read carolyn, bill, and amy's books and i want to give this one a try this book is truly for the apprentice tv lover but if you are looking for business advice this is not the book for you
unhappy dad i picked one of these up for my husband for father's day it looked like a great idea , avoiding the mess of ice and rock salt we followed the directions exactly and were very disappointed when chocolate soup emerged from the maker we put it in a container in the freezer and ended up with a chocolate brick ! needless to say , we are not at all happy with this product.	drop the religion the biblical and creator references in this book are over-whelming i thought i was buying a book about knitting , but found the references too distracting to be able to focus on her knitting instructions
Positive	
awesome great product i am using this because my kitchen is under remodelling works great with smaller vessels heats up evenly i give 4.00000 points because i am annoyed by the bulb on the stove it is on as long as the burner is connected to power i would rather have a burner which has bulb to show if the burner is on off	1776 i read 1776 shortly sfter having read mccullough's john adams so some of the same ground was covered but from a different perspective, i found 1776 to be an interesting and smoothly flowing read anyone who has enjoyed and appreciated david mccullough's other works will enjoy 1776
electric wok unit works great , heats up fast , easy to clean , and much better built than the typical discount store models it looks like this willlast a long time	a perfect whodunnit the twist here is that the reader knows while the characters do not in a matter of minutes, nine people's fates are determined by three men sitting in a jail cell simply playing a game to pass the time, it's up to aidan shields to protect her while the book is decidedly light on the traditional sex part of a romantic suspense, it is implied i liked both characters
bodum beer milk glasses while these are described as beer glasses , they are fantastic for keeping milk cold through a meal i actually pour the milk and place the glasses in the freezer while i start dinner these glasses are easy to clean and retain their sunny shine , even though we have very hard water no spots ! great product	wonderful and most fun i just finished my second reading of it i really cannot add much to the great reviews already written i've always loved dickens' way with words, and this book is a treasure in that department the english, the era , the characters, the countryside it got to the point where i couldn't wait for the evening to come so i could continue reading this

The CapsuleNoDA, which is trained only on the source domain and directly used to predict the sentiment polarity of the reviews from target domain, still outperform the baseline methods by a large margin. This may be because that the capsule network is capable of preserving the instantiation parameters of the sentiment categories and has the potential of encoding the intrinsic spatial relationship between a part and a whole constituting viewpoint invariant knowledge that automatically generalizes to new domains.

Following [7], we also illustrate the adequacy of the CapsuleDAR for domain adaptation in Figure 3. As we can see from Figure 3 that for our model, there is a strong correlation between the cross-entropy loss value and the classification accuracy of the proposed model.

B. ABLATION STUDY

In order to investigate the impact of each part of our model, we perform the ablation test of CapsuleDAR in terms of discarding the sentence structure rule (denoted as w/o sentence structure), the weight initialization rule (denoted as w/o weight initialization), both sentence structure and weight initialization rules (denoted as w/o both rules), and capsule network (denoted as w/o capsule). Note that for the model without capsule network, we remove the capsule network from CapsuleDAR, and a CNN version of our method is applied, which combines pivot based filter initialization method and semantic sentence structure rule.

The ablation results are summarized in Table 4. From the results, we can observe that all the proposed factors contribute great improvement to CapsuleDAR. In particular, the accuracy scores decrease sharply when discarding the capsule network. This is within our expectation since capsule network is able to capture the part-whole relationship that constitutes domain invariant knowledge automatically generalizing to novel domains. In addition, the proposed two semantic rules also contribute to the effectiveness of CapsuleDAR. Not surprisingly, combining all factors achieves the best performance for all the experiments.

C. NUMBER OF ROUTING ITERATIONS

The number of routing iterations is one of the most important hyperparameters of CapsuleDAR, which is unique to the capsule network and has a big influence on the performance and runtime of our model. In this experiment, we investigate what the best number should be on all three dataset setups. We run the experiments with the number of routing iterations as one of (1, 2, 3, 4, 5, 6, 7, 8, 9, 10). The experimental results show that we can achieve the best results with the number of iterations between 1 to 5. After 5 iterations, there will be overfitting phenomenon leading to a decrease in accuracy.

D. VISUALIZATION OF ATTENTION WEIGHTS

CapsuleDAR also provides an intuitive way to demonstrate the important pivot features from source and target domains by visualizing the connection strength between capsule layers. The connection strength shows the importance of each primary capsule for text categories, acting like a parallel attention mechanism. Due to the space limitation, we take three negative samples and three positive samples from each of Kitchen and Books domains as examples, and highlight the pivot features in red (for negative samples) or blue color (for positive samples). The results are reported in Table 5. Our model can effectively learn important features in both source and target domains. For instance, ‘‘terrible’’, ‘‘horrible’’, ‘‘disappointment’’, ‘‘unhappy’’, ‘‘boring’’ are selected for negative reviews and ‘‘awesome’’, ‘‘great’’, ‘‘perfect’’, ‘‘wonderful’’ are selected for positive reviews.

VI. CONCLUSION AND FUTURE WORK

In this paper, we proposed a capsule network framework for sentiment classification in domain adaptation scenario. Capsule network could preserve the instantiation parameters of the training text categories and encode the intrinsic spatial part-whole relationship constituting viewpoint invariant knowledge that could generalize to new domains, bridging the knowledge gap between the source domain and the target domain. In addition, we also incorporate two typical semantic rules, e.g., weight initialization rule and sentence structure information, into the capsule network to further improve the performance of the cross-domain sentiment classification. Extensive experiments have been conducted on a real-world dataset from 4 domains. The experimental results show that the proposed CapsuleDAR significantly outperforms the state-of-the-art methods for cross-domain sentiment classification.

In the future, we plan to incorporate the sentiment resources (e.g., sentiment lexicon, intensity words, negation words) into the capsule networks, which can provide more comprehensive information for sentiment classification. Furthermore, we may also devote our effort to exploit the human reading cognitive process in sentiment analysis, which helps comprehend and understand the text in depth.

REFERENCES

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

[2] 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*, 2010, pp. 751–760.

[3] M. Yang, D. Zhu, and K.-P. Chow, ‘‘A topic model for building fine-grained domain-specific emotion lexicon,’’ in *Proc. 52nd Annu. Meeting Assoc. Comput. Linguistics*, vol. 2, 2014, pp. 421–426.

[4] M. Yang, D. Zhu, M. Rashed, and K. P. Chow, ‘‘Learning domain-specific sentiment lexicon with supervised sentiment-aware lda,’’ *Frontiers Artif. Intell. Appl.*, vol. 263, pp. 927–932, Aug. 2014.

[5] Y. Ganin *et al.*, ‘‘Domain-adversarial training of neural networks,’’ *J. Mach. Learn. Res.*, vol. 17, no. 1, pp. 2030–2096, 2016.

[6] Z. Li, Y. Zhang, Y. Wei, Y. Wu, and Q. Yang, ‘‘End-to-end adversarial memory network for cross-domain sentiment classification,’’ in *Proc. Int. Joint Conf. Artif. Intell. (IJCAI)*, 2017, pp. 2237–2243.

[7] Y. Ziser and R. Reichart, ‘‘Pivot based language modeling for improved neural domain adaptation,’’ in *Proc. Conf. North Amer. Chapter Assoc. Comput. Linguistics, Hum. Lang. Technol.*, vol. 1, 2018, pp. 1241–1251.

[8] M. Yang, W. Tu, Z. Lu, W. Yin, and K.-P. Chow, ‘‘LCCT: A semisupervised model for sentiment classification,’’ in *Proc. Hum. Lang. Technol. Annu. Conf. North Amer. Chapter ACL Assoc. Comput. Linguistics (ACL)*, 2015, pp. 546–555.

[9] T. Mullen and N. Collier, ‘‘Sentiment analysis using support vector machines with diverse information sources,’’ in *Proc. Conf. Empirical Methods Natural Lang. Process.*, 2004, pp. 412–418.

[10] S. Tan, X. Cheng, Y. Wang, and H. Xu, ‘‘Adapting naive Bayes to domain adaptation for sentiment analysis,’’ in *Proc. Eur. Conf. Inf. Retr.* Springer, 2009, pp. 337–349.

[11] S. V. Wawre and S. N. Deshmukh, ‘‘Sentiment classification using machine learning techniques,’’ *Int. J. Sci. Res.*, vol. 5, no. 4, pp. 819–821, 2016.

[12] 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, pp. 142–150.

[13] I. Goodfellow, Y. Bengio, A. Courville, and Y. Bengio, *Deep Learning*, vol. 1. Cambridge, MA, USA: MIT Press, 2016.

[14] J. Blitzer, M. Dredze, and F. Pereira, ‘‘Biographies, bollywood, boomboxes and blenders: Domain adaptation for sentiment classification,’’ in *Proc. 45th Annu. Meeting Assoc. Comput. Linguistics*, 2007, pp. 440–447.

[15] 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. (ICML)*, 2011, pp. 513–520.

[16] G. Zhou, T. He, W. Wu, and X. T. Hu, ‘‘Linking heterogeneous input features with pivots for domain adaptation,’’ in *Proc. IJCAI*, 2015, pp. 1419–1425.

[17] Y. Ziser and R. Reichart, ‘‘Neural structural correspondence learning for domain adaptation,’’ in *Proc. 21st Conf. Comput. Natural Lang. Learn. (CoNLL)*, 2017, pp. 400–410.

[18] G. E. Hinton, A. Krizhevsky, and S. D. Wang, ‘‘Transforming auto-encoders,’’ in *Proc. Int. Conf. Artif. Neural Netw.* Springer, 2011, pp. 44–51.

[19] W. Tu, D. W. Cheung, N. Mamoulis, M. Yang, and Z. Lu, ‘‘Investment recommendation using investor opinions in social media,’’ in *Proc. 39th Int. ACM SIGIR Conf. Res. Develop. Inf. Retr.*, 2016, pp. 881–884.

[20] B. Pang and L. Lee, ‘‘Opinion mining and sentiment analysis,’’ *Found. Trends Inf. Retr.*, vol. 2, nos. 1–2, pp. 1–135, 2008.

[21] B. Liu, ‘‘Sentiment analysis and opinion mining,’’ *Synth. Lectures Hum. Lang. Technol.*, vol. 5, no. 1, pp. 1–167, 2012.

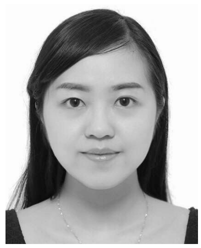
[22] Y. Yoshida, T. Hirao, T. Iwata, M. Nagata, and Y. Matsumoto, ‘‘Transfer learning for multiple-domain sentiment analysis-identifying domain dependent/independent word polarity,’’ in *Proc. AAAI*, 2011, pp. 1286–1291.

[23] M. Chen, Z. Xu, K. Weinberger, and F. Sha. (2012). ‘‘Marginalized denoising autoencoders for domain adaptation.’’ [Online]. Available: https://arxiv.org/abs/1206.4683

[24] Y. Yang and J. Eisenstein, "Fast easy unsupervised domain adaptation with marginalized structured dropout," in *Proc. 52nd Annu. Meeting Assoc. Comput. Linguistics*, vol. 2, 2014, pp. 538–544.
[25] S. Clinchant, G. Csurka, and B. Chidlovskii, "A domain adaptation regularization for denoising autoencoders," in *Proc. 54th Annu. Meeting Assoc. Comput. Linguistics*, vol. 2, 2016, pp. 26–31.
[26] D. P. Kingma and M. Welling, "Auto-encoding variational Bayes," in *Proc. Int. Conf. Learn. Represent. (ICLR)*, 2014, pp. 1–14.
[27] S. Sabour, N. Frosst, and G. E. Hinton, "Dynamic routing between capsules," in *Proc. Adv. Neural Inf. Process. Syst.*, 2017, pp. 3859–3869.
[28] E. Xi, S. Bing, and Y. Jin, "Capsule network performance on complex data," *CoRR*, vol. abs/1712.03480, pp. 1–7, Dec. 2017.
[29] G. E. Hinton, S. Sabour, and N. Frosst, "Matrix capsules with EM routing," in *Proc. 6th Int. Conf. Learn. Represent. (ICLR)*, 2018, pp. 1–12.
[30] Y. Wang, A. Sun, J. Han, Y. Liu, and X. Zhu, "Sentiment analysis by capsules," in *Proc. World Wide Web Conf.*, 2018, pp. 1165–1174.
[31] W. Zhao, J. Ye, M. Yang, Z. Lei, S. Zhang, and Z. Zhao. (2018). "Investigating capsule networks with dynamic routing for text classification." [Online]. Available: https://arxiv.org/abs/1804.00538
[32] Z. Yun-Tao, G. Ling, and W. Yong-Cheng, "An improved TF-IDF approach for text classification," *J. Zhejiang Univ. Sci. A*, vol. 6, no. 1, pp. 49–55, 2005.
[33] D. Wang and Q. Liu, "An optimization view on dynamic routing between capsules," in *Proc. 6th Int. Conf. Learn. Represent. (ICLR)*, 2018, pp. 1–4.
[34] B. Sun and K. Saenko, "Deep CORAL: Correlation alignment for deep domain adaptation," in *Proc. Eur. Conf. Comput. Vis.* Springer, 2016, pp. 443–450.
[35] M. D. Zeiler, "ADADELTA: An adaptive learning rate method," *CoRR*, vol. abs/1212.5701, pp. 1–6, Dec. 2012. [Online]. Available: http://arxiv.org/abs/1212.5701
[36] 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.
[37] Y. Kim, "Convolutional neural networks for sentence classification," in *Proc. 42nd Annu. Meeting Assoc. Comput. Linguistics*, 2014, pp. 1746–1751.

BOWEN ZHANG received the B.Sc. and M.Sc. degrees in computer science and application from the Macau University of Science and Technology, Taipa, China. He is currently pursuing the Ph.D. degree at the Harbin Institute of Technology. His research interests are in natural language processing, big data analysis, and data mining.

XIAOFEI XU received the Ph.D. degree in computer science from the Harbin Institute of Technology (HIT) in 1988. He is currently a Professor in computer science with HIT, and the Vice President at HIT, and the President of HIT, Weihai. His research interests include service computing and service engineering, cloud services and big services, enterprise computing and enterprise interoperability, supply chain management, software engineering, databases and data mining, business intelligence, smart city services, smart healthcare, and elder-care. He has authored/co-authored over 300 academic journal/conference papers, and seven academic books. He has achieved 10 Ministry/Province-Level Awards on Science and Technology, and over 30 software copyrights. He has supervised and cultivated 33 doctors and 83 masters on computer science.



MIN YANG received the Ph.D. degree from the University of Hong Kong in 2017. She is currently an Assistant Professor at the Shenzhen Institutes of Advanced Technology, Chinese Academy of Sciences. Her current research interests include machine learning and natural language processing.

XIAOJUN CHEN received the Ph.D. degree from the Harbin Institute of Technology in 2011. He is currently an Assistant Professor at the College of Computer Science and Software, Shenzhen University. His research interests include subspace clustering, topic model, feature selection, and massive data mining.

YUNMING YE received the Ph.D. degree from Shanghai Jiao Tong University in 2004. He is currently a Professor with the Department of Computer Science, Shenzhen Graduate School, Harbin Institute of Technology (HIT). He is also the Director of the Shenzhen Key Laboratory of Internet Information Collaboration, and the Director of the Intelligent Enterprise Computing Research Lab, HIT, Shenzhen. His research areas are data mining, intelligent Web search, and business intelligence.