Interactive Attention Transfer Network for Cross-domain Sentiment Classification

Kai Zhang[†], **Hefu Zhang**[†], **Qi Liu**[†§*], **Hongke Zhao**[†], **Hengshu Zhu**[‡], **Enhong Chen**[†§]
[†]Anhui Province Key Laboratory of Big Data Analysis and Application, University of Science and Technology of China
[§]School of Data Science, University of Science and Technology of China
[‡]Baidu Talent Intelligence Center, Baidu Inc
{sa517494, zhf2011, zhhk}@mail.ustc.edu.cn, {qiliuql, cheneh}@ustc.edu.cn, zhuhengshu@gmail.com

Abstract

Cross-domain sentiment classification refers to utilizing useful knowledge in the source domain to help sentiment classification in the target domain which has few or no labeled data. Most existing methods mainly concentrate on extracting common features between domains. Unfortunately, they cannot fully consider the effects of the aspect (e.g., the battery life in reviewing an electronic product) information of the sentences. In order to better solve this problem, we propose an Interactive Attention Transfer Network (IATN) for cross-domain sentiment classification. IATN provides an interactive attention transfer mechanism, which can better transfer sentiment across domains by incorporating information of both sentences and aspects. Specifically, IATN comprises two attention networks, one of them is to identify the common features between domains through domain classification, and the other aims to extract information from the aspects by using the common features as a bridge. Then, we conduct interactive attention learning for those two networks so that both the sentences and the aspects can influence the final sentiment representation. Extensive experiments on the Amazon reviews dataset and crowdfunding reviews dataset not only demonstrate the effectiveness and universality of our method, but also give an interpretable way to track the attention information for sentiment.

Introduction

Sentiment analysis, which aims to identify the overall emotional label (i.e., positive or negative) of the sentences, has attracted more and more research attention in recent years. Traditional sentiment classification methods usually perform well on label-rich data (Wang et al. 2014; Tripathy, Agrawal, and Rath 2016). However, in practice, there still exists a huge amount of insufficiently labeled data, where the traditional methods are hard to be utilized. For example, the reviews on the prevalent e-commerce platforms (e.g., Amazon) often contain score option, which can reflect the sentiment labels of the reviews. But in some specialized applications such as crowdfunding platforms (Zhao et al. 2017a), (e.g., Indiegogo.com), nearly all the project reviews do not have sentiment labels.

*Corresponding Author.

Figure 1: Aspects in different domains. For computer, there are "*appearance*" and "*battery life*" aspects; for cloth, there are "*appearance*" and "*fabric*" aspects.

To learn the emotion of the sentences from unlabeled data, cross-domain sentiment classification has been proposed as a promising direction. It uses effectual information in the *source domain* (with sufficient labeled data) to help sentiment classification in the *target domain* (with few or no labeled data). As it is crucial for reducing the reliance on the massive amount of labeled data and significant for domains which are lack of labels, much research attention has been attracted from both academia and industry. In the literature, many methods have been proposed to solve the cross-domain sentiment classification problem (Blitzer, McDonald, and Pereira 2006; Pan et al. 2010; Chen et al. 2012), especially various solutions for learning shared features have been designed. The shared features were usually assumed as words with high co-occurrence in both domains which are regarded as the good predictors of source domain labels.

Recently, with the rapid development of deep learning techniques, researchers proposed many neural network based methods to automatically capture the shared sentiment features across domains (Glorot, Bordes, and Bengio 2011; Chen et al. 2012; Li et al. 2017; 2018b). However, most of the previous efforts ignore the characteristics which do not express the sentiment directly, such as the individual modeling of the aspect. As Figure 1 shows, there are two reviews *"The appearance of the PC looks good, but the battery life is too short"* and *"The appearance of this dress looks nice,*

and the fabric is not bad". The aspect "appearance" appears in both reviews but has different importance in the two domains. Specifically, on electronics (e.g., PC) domain, the other aspect "battery life" affects the sentiment of the review much greater than "appearance". Although the first review is positive on the aspect of appearance, it is still negative overall because the battery life is not satisfying. From this example, we can conclude two main characters of aspects. First, different aspects in the same domain may have different importance. Aspect with a heavier weight affects the sentiment of sentences greater. Second, different domains may share the same aspects. These shared aspects can help us extract more common knowledge between the source and the target domain, and further help to determine the focus in the transfer process. Thus, it is necessary to specifically exploit the aspects of cross-domain sentiment classification.

With the above analysis, we propose an *Interactive Attention Transfer Network* (IATN) model based on Long-Short Term Memory networks (Hochreiter and Schmidhuber 1997; Huang, Xu, and Yu 2015). IATN models sentences and aspects independently and conducts an interactive learning to them. First, IATN makes use of the interactive information from sentences to supervising modeling aspects, which is helpful to judge the across-domain sentiment. Second, IATN utilizes the attention mechanism associated with the aspects to get important information from the sentences and compute ultimate representation for sentiment classification. Finally, by concatenating both the sentence representation and the aspect representation, we can improve the effectiveness of cross-domain sentiment classification. In summary, the main contributions of our work can be summarized as follows.

- For the first time, we propose to make cross-domain sentiment classification by integrating the sentence and aspect representation interactively.
- We propose a novel IATN method which associates with aspects. It utilizes the interactive attention mechanism to get important information from both the sentence and aspect for the sentiment classification task.
- We conduct extensive experiments on two real-world datasets. The experimental results clearly validate that our method outperforms other state-of-the-art methods.

Related Work

The related works can be classified into three categories: domain adaptation, aspect extraction and attention mechanism.

Domain Adaptation. Domain adaptation such as cross-domain sentiment classification is a hot topic in natural language processing, which has been well studied over the decades. Among them, Blitzer et al. (Blitzer, McDonald, and Pereira 2006) proposed a representative study called Structural Correspondence Learning (SCL), which mainly utilized multiple shared features to predict tasks. Pan et al. (Pan et al. 2010) proposed Spectral Feature Alignment (SFA) algorithms to solve the feature mismatch problem by aligning domain-specific words with the help of domain-independent words. Glorot et al. (Glorot, Bordes, and Bengio 2011; Chen et al. 2012) proposed a Marginal Stacking Denoising Autoencoder (mSDA) model which aimed to improve the speed and scalability on high-dimensional data. However, all the methods mentioned above need to manually select some information such as shared or unshared features between the source domain and the target domain.

In recent years, many researchers have studied the neural network-based domain adaptation solutions. For example, Yu and Jiang (Yu and Jiang 2016) proposed two auxiliary tasks to learn sentence embedding based on Convolutional Neural Network (Kim 2014). Ganin et al. (Ganin et al. 2016) added adversarial mechanism into the training of deep neural networks, which called Domain-Adversarial Neural Network (DANN). Li et al. (Li et al. 2017) proposed an Adversarial Memory Network (AMN) that automatically identified common features by applying attention mechanisms and adversarial training. Hierarchical Attention Transfer Network (HATN) was proposed by Li et al. (Li et al. 2018b) which paid attention to word-level and sentence-level sentiment at the same time. The above works have recognized the importance of shared features and developed various methods to improve transfer efficiency through two main tasks like domain classification and sentiment classification. Unfortunately, they have not paid enough attention to the effects of "aspects" (Wang et al. 2016) in the cross-domain sentiment classification task.

Aspect Extraction. Aspect extraction is one of the key tasks in sentiment analysis and has been widely studied in recent years. It aims to extract entity aspects on which opinions have been expressed (Hu and Liu 2004). Currently, there are many outstanding works in the area of aspect extraction, but it has not been applied to improve the effect of cross-domain sentiment classification yet. Li et al. (Li et al. 2018a) proposed an aspect extraction framework by exploiting the opinion summary and the aspect detection history. We apply their model to process our data and extract all the aspects in the reviews.

Attention Mechanism. In the conventional neural networks, encoding the whole input into one feature vector without considering special words usually leads to unsatisfactory classification. In order to solve this problem, attention mechanism has been successfully exploited in various natural language processing tasks, such as machine translation (Tu et al. 2017), sentiment analysis (Ma et al. 2017; Ma, Peng, and Cambria 2018) and question prediction (Huang et al. 2017). Besides, the interactive attention mechanism has been proved to be better than the traditional mechanism because of its effectiveness in extracting powerful features at related domains (Zhang et al. 2017).

Interactive Attention Transfer Network

In this section, we first present the problem of cross-domain sentiment classification, followed by an overview of the model. Then we introduce the technical details of IATN.

Problem Definition

In this paper, we focus on cross-domain sentiment classification. We assume that there are two domains, D_s is the source domain and D_t is the target domain. We further as-

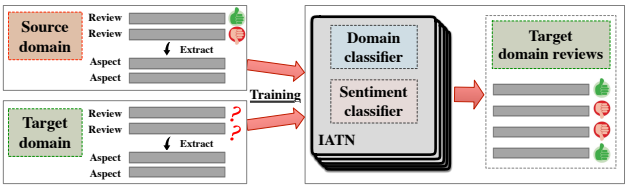

Figure 2: The flowchart overview of our work.

sume that we are given a set of labeled training data $\mathbf{X}_s^l = \{x_s^i, y_s^i\}_{i=1}^{N_s^l}$ and unlabeled training data $\mathbf{X}_s^u = \{x_s^j\}_{j=N_s^l+1}^{N_s}$ from the source domain, where N_s^l and N_s^u are the number of labeled data and all data, respectively. Also, we have a set of unlabeled data from the target domain, denoted by $\mathbf{X}_t = \{x_t^j\}_{j=1}^{N_t}$, where N_t is the number of unlabeled data. Note that each item (e.g., review) at both domains consists of *n* words marked as $s = \{w_s^1, w_s^2, w_s^3 ... w_s^n\}$ and their aspect sequence contains *m* words marked as $a = \{w_a^1, w_a^2, w_a^3 ... w_a^m\}$. The goal of cross-domain sentiment classification is to train a robust model based on labeled and unlabeled data in the source domain and adopt it to predict the unlabeled data in the target domain.

An Overview of IATN

As shown in Figure 2, our approach is modeled with information from sentences (i.e., reviews) and aspects simultaneously. In this study, we apply the existing work (Li et al. 2018a) to extract all aspects of the sentence. After obtaining the aspects, we utilize all data in the source domain and the target domain for training domain classifier. Meanwhile, we apply source labeled data for training sentiment classifier. Finally, with the shared features from both domains, we predict the sentiment label for the target domain data.

Components of IATN

In this subsection, we will introduce the framework of IATN in technical details. As shown in Figure 3, IATN mainly contains two parts, i.e., the sentence network referred as S-net which is mainly for domain classification and the aspect network referred as A-net which is mainly for aspect learning. S-net and A-net have similar structures for word embedding, hidden state learning and pooling operation. While, after the pooling layer, S-net and A-net adopt an interactive attention mechanism jointly. Finally, the output of S-net is sent to the domain classifier and both of the outputs are sent to the sentiment classifier. In the following, we introduce the components of IATN successively.

Word Embedding. In order to represent sentences, we need to map each word into a low-dimensional real-value vector. Word embedding (Bengio et al. 2003) is a kind of required mapping methods, which can be regarded as a functional part of a neural network or a language model pre-trained from proper corpus. Here we choose the pre-training method and take each word as input to get the sentences embedding vectors $e_s = \{e_s^1, e_s^2, e_s^3 ... e_s^n\}$. Similarly, the aspects are also embedded as vectors $e_a = \{e_a^1, e_a^2, e_a^3 ... e_a^m\}$.

Hidden State Learning. After word embedding, we adopt LSTM to learn hidden states because it performs well in learning long-term dependencies and can effectively solve gradient vanishing and expansion problems. Formally, given the word embedding of sentences $e_s = \{e_s^1, e_s^2, e_s^3 ... e_s^n\}$ as the input, LSTM updates the cell vector sequence $c = \{c_1, c_2, c_3 ... c_n\}$ and hidden state $h = \{h_1, h_2, h_3 ... h_n\}$ from t = 1 to n. After the initialization, at *t-th* interaction step, the hidden state h_t of each interaction is updated by the previous hidden state h_{t-1} and the current sentences embedding vector e_s^t as:

$$\begin{aligned} i_t &= \delta(W_{ei} e_s^t + W_{hi} h_{t-1} + \widehat{b}_i) , \\ f_t &= \delta(W_{ef} e_s^t + W_{hf} h_{t-1} + \widehat{b}_f) , \\ c_t &= f_t \cdot c_{t-1} + i_t \cdot \tau(W_{ec} e_s^t + W_{hc} h_{t-1} + \widehat{b}_c) , \\ o_t &= \delta(W_{eo} e_s^t + W_{ho} h_{t-1} + \widehat{b}_o) , \\ h_t &= o_t \cdot tanh(c_t) , \end{aligned} \tag{1}$$

where i_t, f_t and o_t are the input, forget and output gates at *t-th* step respectively. e_s^t is the embedding sentence vector. c_t is the cell memory and h_t is the output. $\delta(\cdot)$ is non-linear activation function which is stated as $sigmoid$ in this paper. Dot $\cdot$ denotes the element-wise multiplication between vectors. W_* denotes weight matrices, $\widehat{b}_*$ is the bias vectors. They are all optimized in the training by the network.

After this layer, the sentence representation is transformed from word embedding vectors to the semantic *sentence hidden states* (i.e., $h_s = \{h_s^1, h_s^2, h_s^3 ... h_s^n\}$), which is the final word representation for sentences. Similarly, we use the same method to obtain the *aspect hidden states* (i.e., $h_a = \{h_a^1, h_a^2, h_a^3 ... h_a^m\}$) for all aspects of the sentences.

Pooling Operation. After getting the hidden state representations of sentences and aspects, we need to make them interactive. We adopt the pooling method, which is a form of non-linear down-sampling to reduce the spatial size of the representation and retain important features. Thus, we can prepare for the interaction of sentence hidden features and aspect hidden features. There are several non-linear functions to implement pooling, among which *mean pooling* works better in practice. Here, we adopt it to calculate the sentence pooling vector (i.e., h_s^p) and the aspect pooling vector (i.e., h_a^p) through the following equations:

$$h_s^p = \sum_{i=1}^{n} h_s^i \,/\, n , \quad h_a^p = \sum_{i=1}^{m} h_a^i \,/\, m . \tag{2}$$

Interactive Word Attention. As we discussed before, each word has a different influence on the representation of the sentences. As shown in Figure 1, compared to the electronic product domain, "*appearance*" may play a more important role in dress domain. Even in the same review of PC, "*appearance*" and "*battery life*" have different effects on the ultimate sentiment classification goal. Therefore, it is necessary to qualify the contributions of each word and learn the special representation for it. Fortunately, attention mechanisms can highlight different parts of the input by assigning weights to encoding vectors in each step of text representation. As we mentioned above, we consider the impact of

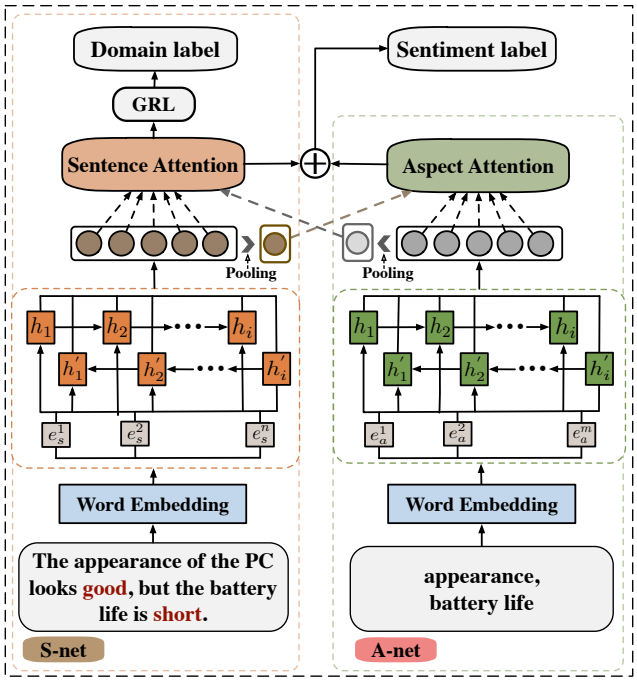

Figure 3: The framework of IATN.

the aspects of the sentences, which can provide more information to represent the final sentiment features. Thus, we take a pair of sentences (i.e., $h_s = \{h_s^1, h_s^2, h_s^3 ... h_s^n\}$) and its aspects (i.e., h_a^p) to describe the attention process. With the representation of those two feature vectors, we can simply connect them to one vector (i.e., $h_s^i = \{h_s^1, h_s^2, h_s^3 ... h_s^n, h_a^p\}$, here we denote $h_s^{n+1} = h_a^p$). Then with the following formula, we can generate the attention vector α_i:

$$\alpha_i = \frac{exp(\gamma(h_s^i))}{\sum_{j=1}^{n+1} exp(\gamma(h_s^i))}, \quad (3)$$

where γ is score function which is defined as:

$$\gamma(h_s^i) = tanh(h_s^i \cdot W_s + \widehat{b}_s), \quad (4)$$

W_s and $\widehat{b}_s$ are weight matrix and bias matrix respectively. $tanh$ is a non-linear function. Particularly, the attention weight α_i greatly enhances the explanatory ability of IATN. It enables us to extract words with high sentiment scores, which is helpful for sentiment cross-domain transfer. In the experiments, we will conduct a deep analysis on attention results to specific words.

After computing the word attention weights, we can get the sentence representation with its auxiliary aspects. Then the final expression is:

$$S_r = \sum_{i=1}^{n+1} \alpha_i h_s^i . \quad (5)$$

Similarly, for the aspects, we use the same method to generate the final representation with its auxiliary sentence hidden pooling vector h_s^p. The attention vector β_i is calculated by:

$$\beta_i = \frac{exp(\gamma(h_a^i, h_s^p))}{\sum_{j=1}^{m} exp(\gamma(h_a^i, h_s^p))}, \quad (6)$$

where γ is the score function which is different from Eq.(5). We adopt spot multiplication ($\cdot$) rather than simple connection. Because we assume that the pooling feature h_s^p contains emotional features and domain shared features. In this way, we can develop sentiment tendencies of different aspects. The score function γ and ultimate aspects representation A_r are formalized as follows:

$$\gamma(h_a^i, h_s^p) = tanh(h_a^i \cdot h_s^p \cdot W_a + \widehat{b}_a), \quad (7)$$

$$A_r = \sum_{i=1}^{m} \beta_i h_a^i . \quad (8)$$

Until now, we get the sentence representation and the aspect representation by interacting with them. Then we will take them into the following tasks.

Domain Classifier. The domain classifier aims to learn cross-domain sentiment feature representations, where the inputs are the source domain data and target domain data. In detail, we utilize all the data $\mathbf{X}_s$ and $\mathbf{X}_t$ to do domain classification, which predicts the domain labels of the samples. Meantime, we use the labeled data $\mathbf{X}_s^l$ in the source domain to do sentiment classification. However, the goal of the common training process is to minimize the classification error, i.e., to distinguish the two domains as accurately as possible. Differently, our intention is to learn common features which the domain classifier cannot discriminate between domains. To solve this problem, we add the Gradient Reversal Layer (GRL) (Ganin and Lempitsky 2014; Ganin et al. 2016) to reverse the gradient direction in the training process. Through the domain classifier, we can get invariance features which are domain shared and sentiment sensitive.

Mathematically, we can formally treat the gradient reversal layer as a "*pseudo-function*", which is defined by two incompatible equations describing its forward and back-propagation behaviors:

$$G(x) = x, \quad \frac{\partial G(x)}{\partial x} = -\lambda I . \quad (9)$$

Then we deal with the sentence representation S_r through the GRL as $G(S_r) = \widetilde{S}_r$ and then feed it to the softmax layer as domain classification:

$$y_d^{'} = softmax(W_d \widetilde{S}_r + \widehat{b}_d) . \quad (10)$$

Sentiment Classifier. The sentiment classifier focuses on mining the information of aspects in sentences. Many efforts have proved that aspects are crucial for classification tasks (Wang et al. 2016). We presume that the coordination of aspects and sentences can enhance the performance of sentiment classification. For example, as shown in Figure 1, not only the aspect "*appearance*" has specific effects on these two reviews, but also "*appearance*" and "*battery life*" make unique contribution to the same sentence. Thus, we

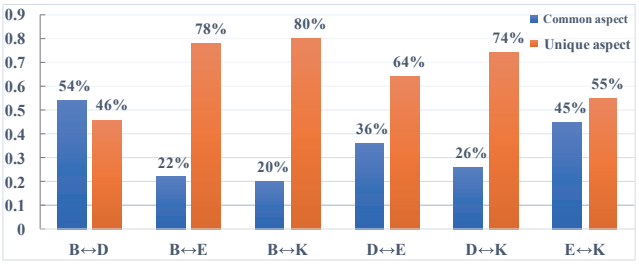

Figure 4: Top-100 aspects analysis between domains.

propose that the weights of aspects should be computed in order to capture their sentiment tendency. Finally, we combine sentence representation and aspect representation for sentiment classification:

$$y_s^{'} = softmax(W_s \cdot [S_r \oplus A_r] + \widehat{b}_s) . \quad (11)$$

where the symbol $\oplus$ represent the connection of vectors.

Training Strategy

Different from the traditional methods, IATN has two tasks, i.e., domain classification and sentiment classification. Thus, our training process comprises two parts.

- Individual Attention Learning. The cross-domain sentiment classifier needs to learn the domain-shared feature representations which contribute to sentiment classification. In order to achieve this objective, we design the two tasks, i.e., domain classification and sentiment classification. We introduce cross-entropy loss functions for training these two classifiers respectively:

$$L_{sen} = -\frac{1}{N_s^l}\sum_{i=1}^{N_s^l} y_s^{'} ln y_s + (1-y_s^{'}) ln(1-y_s) , \quad (12)$$

$$L_{dom} = -\frac{1}{N_d}\sum_{i=1}^{N_d} y_d^{'} ln y_d + (1-y_d^{'}) ln(1-y_d) , \quad (13)$$

  where y_s and y_d denote the ground truth, N_s^l denotes the data which come from labeled source domain, N_d denotes all data from source and target domain.

- Interactive Attention Learning. Based on the above individual attention learning results, we also conduct interactive attention learning for them to optimize the parameters of both tasks simultaneously. In order to avoid overfitting, we add the squared regularization and combine them into an entire objective function:

$$L = L_{sen} + L_{dom} + \rho L_{reg} , \quad (14)$$

  where L_{reg} is the regularization which can avoid overfitting, and ρ is the regularization parameter. The training goal is to minimize L with respect to the model parameters except the GRL training part which will be maximized. Additionally, all the parameters are optimized by the standard back-propagation algorithm (LeCun, Bengio, and Hinton 2015).

Table 1: Statistics of datasets after pre-processing.

Domains	Testing set percentage		
	# Train	# Test	# Unlabel
Books	5,000	1,000	8,000
DVD	5,000	1,000	8,000
Electronics	5,000	1,000	8,000
Kitchen	5,000	1,000	8,000

Experiments

Dataset Preparation

For the reliability and authority of experimental results, we use the Amazon reviews dataset, which has been widely used for cross-domain sentiment classification. Meanwhile, we make the necessary pre-processing as follows. First, we choose the reviews data from four domains: Book (B), DVD (D), Electronics (E) and Kitchen appliances (K). Each of the domains contains 6,000 labeled data, in which there are 3,000 positive reviews (*higher than 3 stars*) and 3,000 negative reviews (*lower than 3 stars*). Additionally, the dataset also contains lots of unlabeled data. Here we randomly select 8,000 unlabeled reviews from each domain as training data. Table 1 summarizes the statistics of dataset after pre-processing.

Hyperparameters Setting

In our experiments, all word embeddings from sentences and aspects are initialized as 200-dimension vectors by word2vec (Goldberg and Levy 2014). The dimensions of word embeddings, attention vectors and LSTM hidden states are set to 200, 64 and 64 respectively. All weight matrices are randomly initialized by a uniform distribution $\mathcal{U}(-0.01, 0.01)$, and all biases are set to zeros. For the performance of IATN, we finally set the coefficient of l_2 normalization, the learning rate and the dropout rate as 10^{-4}, 10^{-3} and 0.25.

The aspect words are not given in the training data directly. Therefore, we need to extract every aspect word of the sentences so that we can make full use of the aspect information. For each sentence S (i.e., $w_s^1, w_s^2 ... w_s^n$), we extract their aspect sequence with *m* words as A (i.e., $w_a^1, w_a^2 ... w_a^m$). As shown in Figure 4, we also extract the top 100 most frequently occurring aspects in each domain and analyze the similar proportions of them between domains. After all the data is processed, we conduct the cross-domain experiments between every two domains, which means we have 12 classification tasks: B→D, B→E, B→K, D→B, D→E, D→K, E→B, E→D, E→K, K→B, K→D, K→E. For example, the notation "B→D" represents the task which transfers from the source domain B to the target domain D.

Benchmark Methods

- **Naive** is a non-domain-adaptive method which is trained in the source domain and predicts in target domain directly. It is designed based on LSTM (Hochreiter and Schmidhuber 1997).

Table 2: Sentiment classification accuracy on the Amazon reviews dataset.

Benchmarks	(a) **Book** →			(b) **DVD** →			(c) **Electronics** →			(d) **Kitchen** →			**Avg**
	Tasks			Tasks			Tasks			Tasks			
	B→D	B→E	B→K	D→B	D→E	D→K	E→B	E→D	E→K	K→B	K→D	K→E	
Naive	0.786	0.752	0.737	0.756	0.734	0.767	0.696	0.722	0.787	0.686	0.723	0.807	**0.746**
SCL	0.807	0.763	0.771	0.782	0.754	0.779	0.716	0.745	0.817	0.713	0.752	0.818	**0.768**
SFA	0.813	0.776	0.785	0.788	0.758	0.786	0.724	0.754	0.825	0.724	0.758	0.825	**0.776**
mSDA	0.819	0.783	0.789	0.783	0.770	0.793	0.738	0.761	0.837	0.730	0.755	0.831	**0.782**
DANN	0.832	0.764	0.790	0.805	0.796	0.814	0.735	0.786	0.841	0.752	0.776	0.843	**0.794**
CNN-a	0.843	0.810	0.813	0.829	0.803	0.819	0.749	0.793	0.843	0.779	0.803	0.855	**0.811**
AMN	0.855	0.824	0.811	0.846	0.812	0.827	0.766	0.827	0.857	0.805	0.812	0.867	**0.825**
HATN	0.858	0.853	0.849	0.858	0.849	0.853	0.808	0.838	0.868	0.824	0.841	0.868	**0.847**
HATN^h	0.861	0.857	0.852	0.863	0.856	**0.862**	0.810	0.840	0.879	0.833	**0.845**	0.870	**0.851**
IATN^n	0.854	0.849	0.838	0.848	0.855	0.839	0.768	0.825	0.859	0.828	0.835	0.864	**0.837**
IATN	**0.868**	**0.865**	**0.859**	**0.870**	**0.869**	0.858	**0.818**	**0.841**	**0.887**	**0.847**	0.844	**0.876**	**0.859**

- **SCL** (Blitzer, McDonald, and Pereira 2006) is a linear method, which aims to solve feature mismatch problem by aligning domain common and unique features.
- **SFA** (Pan et al. 2010) is a method which aims to build a bridge between the source and the target domains by aligning common and unique features.
- **mSDA** (Chen et al. 2012) is proposed to automatically learn a unified feature representation for sentences from a large amount of data in all the domains.
- **DANN** (Chen et al. 2012) is based on the adversarial training. DANN performs domain adaptation with the representation encoded in a 5000-dimension feature vector.
- **CNN-aux** (Yu and Jiang 2016) is based on Convolutional Neural Network (Kim 2014) and makes use of two auxiliary tasks to help inducing sentence embedding.
- **AMN** (Li et al. 2017) is a method which learns domain-shared representations based on memory networks and adversarial training.
- **HATN&HATN**^{h} (Li et al. 2018b) are hierarchical attention networks to focus on both the word level and the sentence level sentiment. The former one does not contain the hierarchical positional encoding and the latter one does.

Experimental Results

To demonstrate the effectiveness of our proposed model, we compare IATN with other state-of-the-art methods on the cross-domain sentiment classification task. Meanwhile, we use classification accuracy (Stehman 1997) to evaluate the models because our training and testing data are balanced. The results of all methods on Amazon dataset are shown in Table 2. From the comprehensive views, IATN model has achieved the best performances on most tasks of this dataset.

Specifically, the Naive method performs badly at every task because it does not use the data of the target domain when training. The performance of traditional methods (i.e., *SCL*, *SFA*, *mSDA*) has reached to 78.2% on average, which are still not accurate enough because they are all based on manually selecting common features, which means the features learned by them are limited. The neural network-based methods have made great improvements compared with the traditional ones, which come to 85.1%. IATN outperforms the state-of-art methods by reaching 85.9% because we extract and make full use of the aspect and sentence information. In order to demonstrate the effects of aspect more intuitively, we also compare IATN with a variant without aspects information, denoted by IATN^n. The average accuracy of IATN^n is 83.7%, which is 2.2% lower than IATN. Through the comparison between IATN^n and IATN, we can conclude that IATN with aspect information does improve the accuracy of the sentiment classification. Specifically, from the

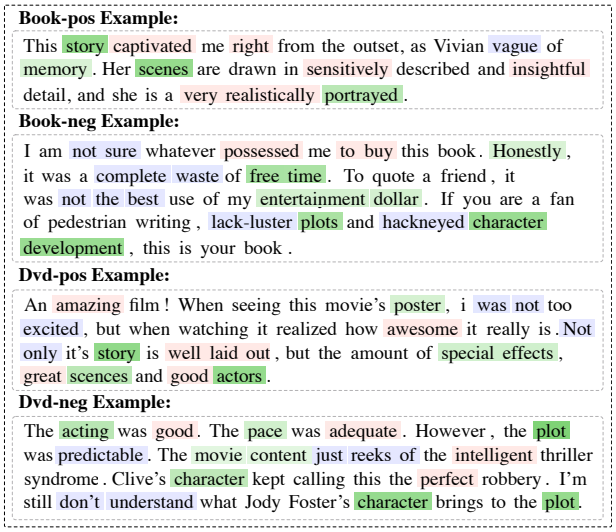

Figure 5: Attention visualization of the aspect words for sentiment classification in B→ D task.

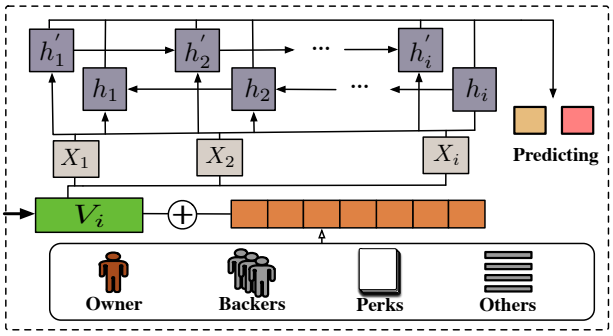

Figure 6: The prediction method of crowdfunding project. V_i is the review representation vector.

experimental results, we can also observe that the classification accuracy between similar domains will be higher than different domain. For example, "B↔D" task is more accurate than "B↔K" task because they have more similar aspects as shown in Figure 4. Finally, compared with the benchmark methods, our model shows great improvement for cross-domain sentiment classification.

Visualization of Attention

In order to validate that our model is able to identify the impact of aspect on sentiment representation, we visualize the aspect attention layer in Figure 5. Here we choose four reviews from the Book domain and DVD domain. Each domain has a positive review and a negative one. The green-colored words in the sentences refer to the aspects. An aspect colored in deeper green means that it gains the heavier weight through the aspect attention layer than the others.

Figure 5 shows that IATN pays higher word attention to the domain-shared aspects, such as "story", "scene", "plot" and "character". Specifically, in the Dvd-neg example, IATN has assigned heavier weights to "character" and "plot" than "acting" and "pace". Here we color the positive words in red and the negative words in blue by artificial judgment to highlight the relationship of the aspect to the emotion of the whole sentence. Although "acting" and "pace" both appear with positive emotions in this review, but the more effective aspects "character" and "plot" have received negative emotions. Affected by these more powerful aspects, this review is finally labeled as negative. In conclusion, the visualization of attention proves that IATN well models sentences and aspects together, and the concatenated representation of them is helpful for the cross-domain sentiment classification.

Application Verification in Crowdfunding

As we mentioned above, the further mission of cross-domain sentiment classification aims to solve the problem of unlabeled domain. However, the training and testing data of Amazon actually have labels so that the sentiment classification results can be evaluated. Thus, to further verify the effectiveness of IATN, we design an application on an unlabeled dataset, i.e., the crowdfunding (Zhao et al. 2017b; Liu et al. 2017) reviews dataset. Specifically, we use Amazon's labeled reviews data as the source domain, and the crowdfunding reviews data as the target domain to predict the emotional tendency (V_i). To evaluate the sentiment classification on unlabeled crowdfunding domain, the review sentiment is combined with multiple features to do the classification of the final states (i.e., *succeed* or *failed*). In order to achieve this goal, we combine review's emotional feature representation (e.g., $V_1, V_2 ... V_i$) with other represented features of projects as a new feature vector (e.g., $X_1, X_2 ... X_i$). Then we use this new feature vector as the input of LSTM to predict the final states of projects, the details are shown in Figure 6. For the benchmarks LSTM^{one}, LSTM^{hatn} and LSTM^{iatn}, V_i is the one-hot vector, the output of HATN and the output of IATN respectively.

Table 3: Results of crowdfunding project prediction.

Benchmarks	Metrics			
	Accu.	Prec.	Rec.	F1-score.
SVM	0.7671	0.4567	0.6971	0.5536
LSTM^{one}	0.7862	0.4813	0.6732	0.5689
LSTM^{hatn}	0.7940	0.4843	0.6805	0.5674
LSTM^{iatn}	**0.8182**	**0.4977**	**0.6733**	**0.5743**

In the crowdfunding dataset[1], there are 12,328 projects with almost 11,2560 reviews and 20 other kinds of features (e.g., *goal, duration and information of owner*). Note that those project's final states, which are unbalanced, consist of more failed but less successful projects. Therefore, we adopt various metrics to better evaluate the methods, i.e., accuracy, precision, recall and F1-score. As Table 3 shows, the SVM method without review information performs the worst and the accuracy result is 5.11% lower than LSTM^{iatn}. LSTM^{iatn} improves the accuracy by 3.20% and 2.42% than LSTM^{one} and LSTM^{hatn} respectively. Furthermore, the comparisons between LSTM^{one}, LSTM^{hatn} and LSTM^{iatn} not only show the excellent performance of LSTM^{iatn} in the project success rate prediction, but also prove that IATN has a good sentiment classification of crowdfunding reviews.

Conclusions

In this paper, we studied the problem of cross-domain in sentiment classification and proposed IATN model which considered the information from both sentences and aspects. Specifically, we aimed to find common features across domains and then extracted information from the aspects with the help of common features. Additionally, we adopted an interactive attention learning mechanism for the sentiment classification, which combined sentences and aspects together. Experiments on Amazon review dataset and crowdfunding dataset clearly verified the effectiveness of IATN. Since we proposed to study the aspect information in the cross-domain sentiment classification for the first time, we hope this work could lead to more researches in the future.

[1] A specific website for crowdfunding platform, https://www.indiegogo.com/

Acknowledgements

This research was partially supported by grants from the National Key Research and Development Program of China (No. 2016YFB1000904), and the National Natural Science Foundation of China (Grants No. 61672483, U1605251 and 91546103). Qi Liu gratefully acknowledges the support of the Young Elite Scientist Sponsorship Program of CAST.

References

Bengio, Y.; Ducharme, R.; Vincent, P.; and Jauvin, C. 2003. A neural probabilistic language model. *Journal of machine learning research* 3(Feb):1137–1155.

Blitzer, J.; McDonald, R.; and Pereira, F. 2006. Domain adaptation with structural correspondence learning. In *Proceedings of the 2006 conference on empirical methods in natural language processing*, 120–128. Association for Computational Linguistics.

Chen, M.; Xu, Z.; Weinberger, K.; and Sha, F. 2012. Marginalized denoising autoencoders for domain adaptation. *arXiv preprint arXiv:1206.4683*.

Ganin, Y., and Lempitsky, V. 2014. Unsupervised domain adaptation by backpropagation. *arXiv preprint arXiv:1409.7495*.

Ganin, Y.; Ustinova, E.; Ajakan, H.; Germain, P.; Larochelle, H.; Laviolette, F.; Marchand, M.; and Lempitsky, V. 2016. Domain-adversarial training of neural networks. *The Journal of Machine Learning Research* 17(1):2096–2030.

Glorot, X.; Bordes, A.; and Bengio, Y. 2011. Domain adaptation for large-scale sentiment classification: A deep learning approach. In *Proceedings of the 28th international conference on machine learning (ICML-11)*, 513–520.

Goldberg, Y., and Levy, O. 2014. word2vec explained: deriving mikolov et al.'s negative-sampling word-embedding method. *arXiv preprint arXiv:1402.3722*.

Hochreiter, S., and Schmidhuber, J. 1997. Long short-term memory. *Neural computation* 9(8):1735–1780.

Hu, M., and Liu, B. 2004. Mining and summarizing customer reviews. In *Proceedings of the tenth ACM SIGKDD international conference on Knowledge discovery and data mining*, 168–177. ACM.

Huang, Z.; Liu, Q.; Chen, E.; Zhao, H.; Gao, M.; Wei, S.; Su, Y.; and Hu, G. 2017. Question difficulty prediction for reading problems in standard tests. In *AAAI*, 1352–1359.

Huang, Z.; Xu, W.; and Yu, K. 2015. Bidirectional lstm-crf models for sequence tagging. *arXiv preprint arXiv:1508.01991*.

Kim, Y. 2014. Convolutional neural networks for sentence classification. *arXiv preprint arXiv:1408.5882*.

LeCun, Y.; Bengio, Y.; and Hinton, G. 2015. Deep learning. *nature* 521(7553):436.

Li, Z.; Zhang, Y.; Wei, Y.; Wu, Y.; and Yang, Q. 2017. End-to-end adversarial memory network for cross-domain sentiment classification. In *Proceedings of the International Joint Conference on Artificial Intelligence (IJCAI 2017)*.

Li, X.; Bing, L.; Li, P.; Lam, W.; and Yang, Z. 2018a. Aspect term extraction with history attention and selective transformation. *arXiv preprint arXiv:1805.00760*.

Li, Z.; Wei, Y.; Zhang, Y.; and Yang, Q. 2018b. Hierarchical attention transfer network for cross-domain sentiment classification. In *Proceedings of the Thirty-Second AAAI Conference on Artificial Intelligence, AAAI 2018, New Orleans, Lousiana, USA, February 2–7, 2018*.

Liu, Q.; Wang, G.; Zhao, H.; Liu, C.; Xu, T.; and Chen, E. 2017. Enhancing campaign design in crowdfunding: a product supply optimization perspective. In *Proceedings of the 26th International Joint Conference on Artificial Intelligence*, 695–702. AAAI Press.

Ma, D.; Li, S.; Zhang, X.; and Wang, H. 2017. Interactive attention networks for aspect-level sentiment classification. *arXiv preprint arXiv:1709.00893*.

Ma, Y.; Peng, H.; and Cambria, E. 2018. Targeted aspect-based sentiment analysis via embedding commonsense knowledge into an attentive lstm. In *Proceedings of AAAI*.

Pan, S. J.; Ni, X.; Sun, J.-T.; Yang, Q.; and Chen, Z. 2010. Cross-domain sentiment classification via spectral feature alignment. In *Proceedings of the 19th international conference on World wide web*, 751–760. ACM.

Stehman, S. V. 1997. Selecting and interpreting measures of thematic classification accuracy. *Remote sensing of Environment* 62(1):77–89.

Tripathy, A.; Agrawal, A.; and Rath, S. K. 2016. Classification of sentiment reviews using n-gram machine learning approach. *Expert Systems with Applications* 57:117–126.

Tu, Z.; Liu, Y.; Shang, L.; Liu, X.; and Li, H. 2017. Neural machine translation with reconstruction. In *AAAI*, 3097–3103.

Wang, G.; Sun, J.; Ma, J.; Xu, K.; and Gu, J. 2014. Sentiment classification: The contribution of ensemble learning. *Decision support systems* 57:77–93.

Wang, Y.; Huang, M.; Zhao, L.; et al. 2016. Attention-based lstm for aspect-level sentiment classification. In *Proceedings of the 2016 conference on empirical methods in natural language processing*, 606–615.

Yu, J., and Jiang, J. 2016. Learning sentence embeddings with auxiliary tasks for cross-domain sentiment classification. In *Proceedings of the 2016 Conference on Empirical Methods in Natural Language Processing*, 236–246.

Zhang, X.; Li, S.; Sha, L.; and Wang, H. 2017. Attentive interactive neural networks for answer selection in community question answering. In *AAAI*, 3525–3531.

Zhao, H.; Ge, Y.; Liu, Q.; Wang, G.; Chen, E.; and Zhang, H. 2017a. P2p lending survey: platforms, recent advances and prospects. *ACM Transactions on Intelligent Systems and Technology (TIST)* 8(6):72.

Zhao, H.; Zhang, H.; Ge, Y.; Liu, Q.; Chen, E.; Li, H.; and Wu, L. 2017b. Tracking the dynamics in crowdfunding. In *Proceedings of the 23rd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, 625–634. ACM.