Cross-Domain Text Sentiment Analysis Based on Deep Recurrent Neural Networks: Postprint

Authors: Yu Chuanming

Date: 2023-08-26T00:00:00+00:00

Abstract

[Purpose/Significance] By learning from a source domain with abundant annotation resources and projecting documents from the target domain into the same feature space as the source domain, this approach addresses the problem of obtaining good classification models in the target domain due to its limited amount of annotated data. [Method/Process] Using Chinese reviews from Amazon's book, DVD, and music categories as experimental data and cross-domain sentiment analysis as the research task, we propose a Cross Domain Deep Recurrent Neural Network (CD-DRNN) model to achieve knowledge transfer across different domain environments. The CD-DRNN model achieved an average classification accuracy of 81.70% in cross-domain settings, outperforming traditional Stacked Long Short Term Memory (Stacked-LSTM) models (79.90%), Bidirectional Long Short Term Memory (Bi-LSTM) models (80.50%), Convolution Neural Network with Long Short Term Memory (CNN-LSTM) models (74.70%), and Merged Convolution Neural Network with Long Short Term Memory (Merged-CNN-LSTM) models (80.90%). [Results/Conclusion] Knowledge transfer between source and target domains can effectively solve the problem that supervised learning struggles to achieve good classification performance on small datasets. The CD-DRNN model can effectively filter features from unannotated data, thereby significantly reducing the workload associated with data annotation in the target domain.

Full Text

Preamble

ChinaXiv Partner Journal, Vol. 62, No. 11, June 2018
Cross-Domain Text Sentiment Analysis Based on Deep Recurrent Neural Networks
Yu Chuanming

School of Information and Safety Engineering, Zhongnan University of Economics and Law, Wuhan 430073

Abstract

[**Purpose/Significance**] This study addresses the challenge of obtaining effective classification models in target domains with limited annotated data by leveraging knowledge from source domains rich in labeled resources. It projects target domain documents into the same feature space as the source domain to enable knowledge transfer.

[**Method/Process**] Using Chinese reviews from Amazon's book, DVD, and music categories as experimental data, we propose a Cross-Domain Deep Recurrent Neural Network (CD-DRNN) model for cross-domain sentiment analysis. This model achieves knowledge transfer across different domain environments.

[**Result/Conclusion**] The CD-DRNN model attains an average classification accuracy of 81.70% in cross-domain settings, outperforming traditional Stacked Long Short-Term Memory (Stacked-LSTM) (79.90%), Bidirectional Long Short-Term Memory (Bi-LSTM) (80.50%), Convolutional Neural Network with Long Short-Term Memory (CNN-LSTM) (74.70%), and Merged Convolutional Neural Network with Long Short-Term Memory (Merged-CNN-LSTM) (80.90%) models. Knowledge transfer between source and target domains effectively solves the difficulty of supervised learning in achieving good classification performance on small datasets. The CD-DRNN model can effectively select features from unlabeled data, thereby significantly reducing the workload associated with data annotation in the target domain.

Keywords: cross-domain, transfer learning, deep recurrent neural network, sentiment analysis

With the development and proliferation of social networks, an increasing number of people tend to express and share their subjective opinions about products through these platforms. Meanwhile, consumers increasingly rely on online reviews as important references before making purchase decisions. From literary works (e.g., book, movie, and music reviews) to electronic products (e.g., mobile phones, digital cameras, and home appliances), and from lifestyle services (e.g., dining and travel reviews) to daily necessities (e.g., clothing and real estate reviews), the domains covered by user reviews are widely distributed and often imbalanced. In automated information analysis processes, building separate models for each domain would consume substantial time and human resources. Consequently, researchers aim to establish transfer learning or domain adaptation models across similar or related domains. For instance, by annotating data in the book review domain, the learned sentiment classification or topic distribution models can be transferred or adapted to the movie review domain, thereby saving significant annotation time.

In practical applications of transfer learning (domain adaptation), severe concept drift and feature drift often occur. For example, sensory words like "sour, sweet, bitter, spicy" appear far more frequently in food reviews than in home appliance reviews. The substantial differences in feature distribution across domains lead to degraded performance of sentiment classification models in cross-domain applications. Additionally, the same feature may reveal different sentiment polarities in different domains. For instance, "long time" in restaurant reviews tends to express negative sentiment about waiting for food, while in electronic product reviews, it more likely indicates positive sentiment about long battery life. This causes sentiment models trained on restaurant reviews to perform poorly when applied to electronic product reviews.

To address concept drift (feature drift) across domain boundaries, this study selects Chinese reviews from Amazon's book, DVD, and music categories as experimental data and proposes a Cross-Domain Deep Recurrent Neural Network (CD-DRNN) model for cross-domain sentiment analysis tasks, enabling knowledge transfer across different domain environments.

1 Related Research

1.1 Cross-Domain Text Sentiment Analysis from a Knowledge Transfer Perspective

From a knowledge transfer perspective, cross-domain text sentiment analysis primarily addresses feature distribution discrepancies and resulting concept drift (feature drift) across domains. Methods include feature transfer (feature adaptation), instance transfer (instance adaptation), and model transfer (model adaptation).

Feature transfer aims to resolve inconsistencies between source and target domain feature spaces by transforming target domain data features into the same vector space as the source domain. For example, Blitzer et al. proposed the Structural Correspondence Learning (SCL) model, which first screens common feature words (pivot words) across multiple domains, then quantifies their relationships with positive and negative sentiment expressions, and models these relationships using traditional machine learning approaches. Pan et al. introduced the Spectral Feature Alignment (SFA) method, which distinguishes between domain-specific and domain-independent features, using domain-independent features as anchors to cluster domain-specific features to their nearest domain-independent counterparts. These methods rely on the distribution of pivot words (or domain-independent words) across domains; if their distribution is too sparse, achieving good performance becomes difficult.

Instance transfer involves selecting annotated data from the source domain that is most similar to the target domain to augment the target domain training set. For example, Dai et al. proposed the TrAdaBoost method, which leverages a small amount of target domain labeled data along with large amounts of source domain labeled data to build high-quality text classification models.

Ma et al. proposed a dual-selection strategy that first selects samples based on source domain sample weights, then performs secondary selection based on classification confidence. Experimental results demonstrate that this approach improves overall classification accuracy. However, instance transfer methods depend heavily on the quantity of source domain labeled samples; if the quantity is insufficient, these methods become less applicable.

Model transfer involves sharing models or parameters between source and target domains to achieve text sentiment analysis in the target domain. For instance, Glorot et al. utilized a deep learning model called Stacked Denoising Autoencoder (SDAe) to pre-train on unlabeled data from multiple domains, then combined the pre-trained model with source domain labeled documents to train sentiment classification models. This approach outperformed SCL and SFA models across 22 domains. However, the SDAe algorithm suffers from slow speed and heavy reliance on parameter initialization. Researchers improved upon it with Marginalized SDAe, achieving faster runtime. Since model transfer methods reduce dependence on pivot words (or domain-independent words) and source domain labeled data, they generally outperform traditional feature transfer and instance transfer methods when combined with deep learning.

1.2 Cross-Domain Text Sentiment Analysis from a Machine Learning Perspective

From a machine learning perspective, cross-domain text sentiment analysis 主要包括 graph-based methods, traditional statistical learning methods, and deep learning-based methods.

Graph-based methods typically abstract words from source and target domains as nodes and relationships between words (e.g., co-occurrence) as edges, calculating similarity between nodes through specific algorithms to perform sentiment analysis in the target domain. The SimRank algorithm is the most typical graph-based method, based on the principle that a node is most similar to itself, and two nodes are more similar if they share more identical or similar neighbor nodes. Wei et al. improved SimRank with a weighted version and applied it to cross-domain sentiment analysis, validating the effectiveness of the weighted model.

Traditional statistical learning methods typically estimate the conditional probability of target domain text belonging to a certain sentiment category given source and target domain texts. For example, Zhang applied logistic regression (LR) to cross-domain sentiment analysis, demonstrating its effectiveness. Huang et al. compared Naive Bayes (NB), Support Vector Classification (SVC), and Expectation Maximization (EM) methods for cross-domain sentiment analysis, finding EM slightly outperformed NB and SVC. Xia et al. used combined features including part-of-speech tagging and word relationships to build ensemble models with Naive Bayes, maximum entropy, and support vector machine models, achieving better results than traditional single models. Deshmukh et al. combined an improved maximum entropy model with bipartite graph clustering, achieving relatively high accuracy for sentiment word classification.

Deep learning-based methods typically represent words and texts from source and target domains as lower-dimensional vectors (word vectors, sentence vectors, or document vectors) and apply them to deep learning algorithms (e.g., deep neural networks). Tang et al. early analyzed the application of deep learning methods in sentiment analysis, finding them superior to traditional methods in sentiment classification, opinion extraction, and sentiment lexicon construction. Yu et al. combined structural correspondence learning with convolutional neural networks for cross-domain sentiment analysis, showing deep learning models outperform traditional models on multiple metrics. Yu et al. also used deep learning methods for sentence modeling, demonstrating deep learning models outperform traditional structural correspondence learning models. Since deep learning models can better reveal and capture the intrinsic semantic representations of text information across domains, they enable researchers to move away from cumbersome traditional feature engineering and generally outperform traditional graph-based and statistical learning methods.

1.3 Cross-Domain Text Sentiment Analysis from a Sentiment Computation Perspective

From a sentiment computation perspective, cross-domain text sentiment analysis 主要包括 lexical sentiment computation, document-level sentiment computation, and sentence-level sentiment computation.

Cross-domain lexical sentiment computation involves determining whether words express positive or negative sentiment across different domains. For example, Wu et al. divided cross-domain sentiment word computation into three steps: extracting benchmark words (representative words with clear sentiment orientation), disambiguating benchmark words (judging sentiment ambiguity across domains), and determining target word sentiment orientation (by calculating correlation strength between target and benchmark words). Feng et al. proposed a semi-supervised sentiment orientation algorithm based on word vector similarity (SO-WV) on top of deep learning, showing word vector-based models outperform traditional Pointwise Mutual Information (PMI) and Label Propagation (LP) algorithms.

Cross-domain document-level sentiment computation involves determining whether text documents express positive or negative sentiment across domains, using rule-based and statistical approaches. In rule-based approaches, Denecke et al. attempted to use domain-general sentiment lexicons (SentiWordNet) for cross-domain document sentiment analysis, finding that combining domain-general sentiment dictionaries with rules (e.g., accumulating sentiment polarity intensity) underperformed direct statistical learning methods. In statistical approaches, methods range from simple logistic regression and Naive Bayes to complex support vector classification and expectation maximization, from single models to ensemble models, and from traditional machine learning to deep learning. Statistical learning methods that bypass sentiment lexicons have become mainstream for cross-domain document sentiment computation.

Cross-domain sentence-level sentiment computation is a special case of document-level computation (limiting documents to single sentences), with similar principles and methods that need not be elaborated further. Across lexical, document, and sentence levels, deep learning methods generally outperform traditional dictionary-based and rule-based methods.

Overall, the difficulty of cross-domain versus non-cross-domain analysis lies in solving transfer learning problems; the difficulty of moving from non-sentiment to sentiment analysis lies in improving machine learning algorithms. Since deep learning models can better reveal and capture the intrinsic semantic representations of cross-domain text information, they offer significant advantages in solving transfer learning problems and improving machine learning algorithms. As an important representative of deep learning models, recurrent neural networks have gained increasing attention in recent years. Typical recurrent neural networks include Long Short-Term Memory (LSTM) and Bidirectional LSTM (Bi-LSTM) models. A literature search reveals no empirical studies applying recurrent neural networks to cross-domain sentiment analysis. Nevertheless, numerous experiments demonstrate that LSTM effectively improves machine translation, language modeling, multilingual information processing, and automatic image captioning, while Bi-LSTM achieves good results in Chinese word segmentation, syntactic parsing, and part-of-speech tagging. Therefore, this study introduces recurrent neural networks into empirical research on cross-domain text sentiment analysis to examine their effectiveness in solving transfer learning problems and improving machine learning algorithms, providing insights for related research.

2 Research Questions and Methods

2.1 Formal Definition of Research Questions

This study addresses the following question: Given large amounts of unlabeled data in both source and target domains, along with a small amount of labeled data in the source domain, how can we better leverage transfer learning and deep learning theories to solve concept drift (feature drift) problems across domain boundaries and achieve text sentiment analysis in the target domain at the document level?

Specifically, let subscript s denote the source domain and t denote the target domain; D represents the domain set; $Train$, $Test$, and U represent training sets, test sets, and unlabeled documents, respectively. The problem can be described as: Given labeled dataset $Train$ and unlabeled dataset U in source domain D , and test set $Test$ and unlabeled documents U in target domain D , use $Train$, U , U , and the proposed model to perform sentiment classification on $Test$.

For simplicity, we use $D \to D$ to denote domain transfer learning from source to target domain. For source and target domains, to facilitate comparison with previous research, this study selects three commonly used domains in cross-domain research: books, DVDs, and music, denoted as Book, DVD, and Music, respectively.

2.2 CD-DRNN Model Structure

Based on deep learning and transfer learning theories, this study proposes the CD-DRNN model. The fundamental idea is to first use deep learning models for cross-domain representation learning to build a domain-independent feature space and project features from different domains into this common space. Then, a stacked bidirectional LSTM network performs supervised learning on source domain labeled data, with each layer learning the previous layer's output sequence in both forward and reverse order. At any time step t, the model captures forward and backward context information, with features extracted through multiple layers ultimately fed into activation units for sentiment recognition. The model and parameters are then shared with the target domain to perform sentiment analysis on unlabeled target domain data and compare results with ground truth.

In terms of methodology, the model adopts model transfer for knowledge migration, sharing the stacked bidirectional LSTM model and parameters between source and target domains. For machine learning, it combines "deep representation learning + stacked bidirectional LSTM model," using unsupervised pre-training on unlabeled data from both domains for representation learning to extract common semantic features, which are then loaded into bidirectional LSTM units that accumulate layer by layer before being fed into the activation layer. For sentiment computation, the model leverages knowledge transfer and deep learning for document-level cross-domain text sentiment analysis.

The CD-DRNN model structure is shown in Figure 1 [Figure 1: see original paper]. The architecture consists of an Input Layer, Representation Learning Layer, Bidirectional Long Short-Term Memory Layer, and Activation Function Layer (output layer), discussed in detail below.

2.2.1 Input Layer The input layer simultaneously accepts documents from both source and target domains. For each batch of user reviews of size *batch_{size}*, the model takes length s as the uniform length for the current input document set. Documents shorter than s are padded with zeros.

2.2.2 Representation Learning Layer In the representation learning layer, the model projects words from source and target domain documents into a common d-dimensional space, generating d-dimensional word vectors. Simultaneously, the model treats each review as a document and generates d-dimensional document vectors. During representation learning, the model uses word and document vectors (collectively referred to as Word2Vec) to predict the next word in sentences from both domains. In our experiments, we use averaging as the composition method to combine vectors and train on the aforementioned corpus to obtain relevant vectors and parameters. During model initialization, document and word vectors are randomly initialized. By defining a loss function in deep learning (quantifying the gap between predicted and actual values) and employing optimization methods (e.g., stochastic gradient descent), the model ultimately obtains document and word vectors as byproducts of the prediction task.

Assuming the corpus contains N reviews and the vocabulary contains M words, the model has a total of $N\times d + M\times d$ parameters. When N and M are large, the number of parameters may also be large, and parameter updates during training are typically sparse. Notably, word and document vectors are simultaneously learned from unlabeled source and target domain data, enabling the capture of common features across both domains. Feature acquisition through predicting words from context eliminates the need for laborious manual annotation. Additionally, during learning, the model considers word order within small contexts, similar to n-gram models, which preserve substantial paragraph information including word order. While traditional n-gram models require extremely high-dimensional representations, representation learning models create relatively low-dimensional representations, offering better generalization performance.

2.2.3 Bidirectional Long Short-Term Memory Layer Based on the common features learned for source and target domains, the model builds a multi-layer stacked LSTM network, with each layer connecting two LSTMs (forward LSTM and backward LSTM). The forward LSTM reads input sequences in the traditional order, while the backward LSTM reads them in reverse. Each LSTM network consists of input gates, output gates, forget gates, and memory cells. This combination of control gates and memory cells better represents and controls long-term, distant, and recent memories of input sequences, enhancing the model's ability to handle long-distance dependencies. At any time step t, the model can capture historical and future context information. The bidirectional LSTM is multi-layered: on top of the first bidirectional LSTM layer, another bidirectional LSTM layer is stacked, using the first layer's output as input to the second layer's corresponding nodes, and so on. Since adjacent layers have a one-to-one structural correspondence, stacking can be implemented effectively in practice.

2.2.4 Output Layer The output layer consists of activation units. Features extracted through multiple layers are ultimately fed into activation units for sentiment classification. The activation unit produces an N-dimensional vector where the first value represents the probability of the current review belonging to the first class, the second value represents the probability of belonging to the second class, and so on.

In this study, model parameters are set as shown in Table 1 .

2.3 Comparison Methods

This study compares the proposed model with baseline methods, selecting three representative single machine learning models (Support Vector Classification, Logistic Regression, and Decision Tree), three ensemble models (Random Forest, Bagging, and AdaBoost), and three recurrent neural network models (Stacked LSTM, Bidirectional LSTM, and CNN-LSTM combination models in both serial and parallel configurations). For feature extraction, we select the most representative traditional TF-IDF method and compare it with word vector methods from deep learning.

When using traditional feature engineering to extract vocabulary from source and target domains, we apply Principal Component Analysis (PCA) to reduce dimensions to 100 for consistency with CD-DRNN and other baselines. When using deep representation learning for vocabulary extraction, we also set dimensions to 100 for fair comparison. Specific settings are as follows:

1. **Cross-domain TF-IDF-based machine learning methods**: Using source and target domain training sets and unlabeled review sets, we build a dictionary (size: 93,000) and calculate TF-IDF weights. For the source domain training set, we apply PCA to reduce feature dimensions from 93,000 to 100. The reduced training data is fed into traditional machine learning models including SVC, LR, DT, RF, Bagging, and AdaBoost. PCA is also applied to the target domain test set before classification with trained models.

2. **Cross-domain word embedding-based machine learning methods**: Using source and target domain training sets and unlabeled review sets, we train word vectors (dimension: 100). Source domain reviews are converted to word vectors as training data for traditional machine learning models (SVC, LR, DT, RF, Bagging, AdaBoost). Target domain reviews are tested using trained models with word vector input.

3. **Stacked LSTM model**: Stacking LSTM layers in a multi-layer architecture where lower-layer LSTM hidden state outputs serve as inputs to higher-layer LSTMs. Source domain reviews are converted to word vectors as training data; target domain reviews are tested using the trained model.

4. **Bidirectional LSTM model**: Adding a backward learning process (processing input in reverse order) to the LSTM model. Source domain reviews are converted to word vectors as training data; target domain reviews are tested using the trained model.

5. **CNN-LSTM serial model**: Adding convolution and pooling processes before the LSTM model to extract more accurate features. Source domain

reviews are converted to word vectors as training data; target domain reviews are tested using the trained model.

6. **Merged CNN-LSTM model**: Adding convolution and pooling processes alongside the LSTM model, feeding jointly learned features to the activation function. Source domain reviews are converted to word vectors as training data; target domain reviews are tested using the trained model.

Baseline method parameters are set as shown in Table 2 .

3 Experiments and Analysis

3.1 Dataset

We crawled Chinese reviews from Amazon's book, music, and DVD categories. Each user review corresponds to a rating from one to five stars, expressing user satisfaction. Reviews with three or more stars are labeled as positive, those with fewer than three stars as negative; three-star reviews are removed. Each domain's reviews are divided into training sets, test sets, and unlabeled documents. The book, music, and DVD domains each include 3,000 reviews in the training set (1,500 positive and 1,500 negative), 1,000 reviews in the test set (500 positive and 500 negative), and approximately 130,000 unlabeled reviews.

Statistics on original review lengths across the three domains are shown in Table 3 . The maximum review length is 1,104 for books (average: 19), 741 for DVDs (average: 19), and 759 for music (average: 18). In terms of length variation, DVD reviews show the greatest variability, followed by books, with music showing the least.

The distribution of review lengths across the three domains is shown in Figure 2 [Figure 2: see original paper]. The three domains show strong consistency in length distribution proportions, with most reviews falling in the [0,10) and [10,30) intervals, followed by [30,50) and [50,100). Considering computational constraints and the need to preserve original review data, we set the review length to the larger value of 120 for all models (including baselines and CD-RNN).

3.2 Experimental Results

3.2.1 Cross-Domain TF-IDF-Based Machine Learning Methods We use Receiver Operating Characteristic (ROC) curves and Area Under the Curve (AUC) to measure method effectiveness. ROC curves and AUC values for each model are shown in Figure 3 [Figure 3: see original paper]. Across six cross-domain sentiment analysis experiments (Book→DVD, DVD→Book, Music→Book, Book→Music, Music→DVD, DVD→Music), the LR model achieves the best performance with AUC values of 0.84, 0.87, 0.91, 0.83, 0.90, and 0.84 (mean: 0.865). SVC ranks second with AUC values of 0.82, 0.86, 0.91, 0.81, 0.88, and 0.83 (mean: 0.852). AdaBoost ranks third with AUC values of 0.83, 0.82, 0.86, 0.81, 0.86, and 0.82 (mean: 0.833). RF and Bagging rank fourth and fifth with mean AUC values of 0.812 and 0.757, respectively. DT performs worst with AUC values of 0.67, 0.67, 0.74, 0.68, 0.71, and 0.68 (mean: 0.733).

Cross-domain adaptation effectiveness varies across different algorithm-group combinations (i.e., when source and target domains change). The Music→Book experiment achieves the best results across SVC, LR, DT, RF, Bagging, and AdaBoost algorithms, followed by Music→DVD, DVD→Book, DVD→Music, Book→DVD, and Book→Music. This indicates that source and target domain selection significantly impacts cross-domain text sentiment analysis.

3.2.2 Cross-Domain Word Embedding-Based Machine Learning Methods Using ROC and AUC to measure model effectiveness, results are shown in Figure 4 [Figure 4: see original paper]. Across the six cross-domain experiments, SVC and LR achieve the best performance with mean AUC values of 0.843. AdaBoost ranks second with AUC values of 0.78, 0.81, 0.85, 0.82, 0.81, and 0.78 (mean: 0.808). Bagging ranks fourth with AUC values of 0.72, 0.77, 0.80, 0.74, 0.76, and 0.71 (mean: 0.750). RF ranks fifth with AUC values of 0.72, 0.70, 0.78, 0.72, 0.74, and 0.69 (mean: 0.725). DT performs worst with AUC values of 0.61, 0.59, 0.64, 0.58, 0.64, and 0.59 (mean: 0.608).

Comparing Figures 3 and 4 reveals that cross-domain word embedding methods fail to effectively improve cross-domain machine learning performance compared to TF-IDF-based methods. In Music→Book, Book→DVD, and DVD→Music tests, word vector-based approaches even show decreased performance across SVC, LR, DT, RF, Bagging, and AdaBoost algorithms. These results indicate that cross-domain word vector learning alone cannot effectively capture common features across domains. It is necessary to introduce additional feature extraction mechanisms (e.g., multi-layer bidirectional LSTM) to enhance cross-domain knowledge transfer and feature mapping.

3.2.3 Recurrent Neural Network-Based Methods Since recurrent neural network models typically test in batch-wise fashion, AUC and ROC values depend on random batch partitioning of the test set. To ensure reproducibility, we use Accuracy (Acc) to measure experimental results for recurrent neural networks.

Table 4 shows results for various recurrent neural network models (Stacked-LSTM, Bi-LSTM, CNN-LSTM, Merged-CNN-LSTM, and CD-DRNN), cross-domain TF-IDF machine learning models (TF-IDF+PCA+SVC, TF-IDF+PCA+LR, TF-IDF+PCA+DT, TF-IDF+PCA+RF, TF-IDF+PCA+Bagging, TF-IDF+PCA+AdaBoost), and cross-domain word embedding machine learning models (WE+SVC, WE+LR, WE+DT, WE+RF, WE+Bagging, WE+AdaBoost) across six cross-domain experiments. Bolded values indicate maximum performance for each experiment; the rightmost column shows average performance across six experiments.

The results show that CD-DRNN achieves the highest Acc values in DVD→Book, Music→Book, Music→DVD, and DVD→Music experiments (81.50%, 86.10%, 85.30%, and 78.20%, respectively). In Book→DVD and Book→Music experiments, Merged-CNN-LSTM and Stacked-LSTM achieve the highest Acc values (82.50% and 78.50%, respectively). Ranked by average performance across six experiments, the models are: CD-DRNN (81.70%), Merged-CNN-LSTM (80.90%), Bi-LSTM (80.50%), Stacked-LSTM (79.90%), and CNN-LSTM (74.70%). This demonstrates that the proposed model better performs cross-domain text sentiment analysis, validating its effectiveness.

Comparing recurrent neural network models with cross-domain TF-IDF and word embedding machine learning models reveals that recurrent neural networks generally outperform both TF-IDF and word embedding methods. CD-DRNN achieves the best average performance (81.70%), surpassing the best cross-domain TF-IDF model (TF-IDF+PCA+LR, 77.20%) by 4.50% and the best cross-domain word embedding model (WE+LR, 76.80%) by 4.90%.

3.3 Discussion

Comparing the three experimental groups demonstrates that the proposed deep recurrent neural network approach outperforms all baseline methods for cross-domain text sentiment analysis. Compared to traditional TF-IDF methods, the model achieves a maximum improvement of approximately 28.6% in mean performance across six experiments. Compared to word embedding methods, it achieves a maximum improvement of about 20.9%. Compared to existing recurrent neural networks, it achieves a maximum improvement of approximately 7%. Unlike word embedding methods that fail to effectively improve cross-domain machine learning, CD-DRNN significantly enhances cross-domain sentiment recognition effectiveness, indicating that stacking multiple LSTM layers with bidirectional learning effectively captures useful features across domains, addressing concept drift and substantially improving classification performance.

From the model's internal mechanisms, this study involves knowledge transfer, machine learning, and sentiment computation perspectives. For knowledge transfer, CD-DRNN adopts model transfer, sharing the stacked bidirectional LSTM model and parameters between source and target domains. This reduces dependence on pivot words (domain-independent words) in traditional feature transfer methods and on source domain labeled data in instance transfer methods, yielding a more concise model. For machine learning, CD-DRNN combines "deep representation learning + stacked bidirectional LSTM," where deep representation learning better reveals intrinsic semantic representations across domains. Compared to traditional RNNs, the inclusion of memory cells better represents and controls long-term, distant, and recent memories of input sequences, enhancing long-distance dependency handling and effectively addressing gradient vanishing/exploding problems in traditional RNNs when processing long sequences. The bidirectional approach resolves reverse dependency issues, while multi-layer stacking extracts domain-related sentiment polarity features at multiple levels, collectively improving cross-domain sentiment analysis accuracy. For sentiment computation, the model performs document-level cross-domain text sentiment analysis through knowledge transfer and machine learning, eliminating reliance on cross-domain sentiment dictionaries and facilitating broader application.

Compared to similar research (e.g., [21]), this study shares some connections but also has essential differences. Methodologically, [21] uses deep convolutional neural networks, while this study uses deep recurrent neural networks—fundamentally different approaches that extract features through convolution/pooling versus multi-layer bidirectional LSTM, belonging to different deep learning research categories. Experimentally, due to different theoretical foundations, we employ different baseline methods and evaluation metrics. [21] compares with structural correspondence learning methods using F1 scores, while we compare with nine baseline methods (three single models, three ensemble models, three RNN models) using AUC and Acc metrics. In terms of internal mechanisms, we adopt different knowledge transfer methods: [21] relies on feature transfer (domain-independent words) using pivot words to represent document terms, while we use model transfer, sharing the stacked bidirectional LSTM model and parameters between domains, reducing dependence on pivot words and source domain labeled data. Finally, our experimental results demonstrate more significant advantages over most baseline methods, more powerfully proving the effectiveness of deep recurrent neural networks for cross-domain text sentiment analysis.

From a data usage perspective, this study uses only 4,000 labeled reviews per domain (2,000 positive and 2,000 negative), far fewer than the 130,000 unlabeled reviews. Despite limited labeled data, the model still outperforms baseline methods, indicating that CD-DRNN can effectively select features from unlabeled data, substantially reducing annotation workload in both source and target domains. Notably, CD-DRNN simultaneously learns from unlabeled source and target domain documents, acquiring features through context-based word prediction as an entry point, enabling application without heavy manual annotation.

From group experimental results, different algorithm-group combinations show varying adaptation effectiveness. For example, Music→Book achieves the best results across almost all algorithms, followed by Music→DVD, Book→DVD, DVD→Book, DVD→Music, and Book→Music. When source and target domains change, model performance varies significantly, indicating that domain selection impacts cross-domain text sentiment analysis. While domain similarity is generally believed to positively correlate with adaptation effectiveness, this study does not formally test this relationship. However, comparing Book→Music and Music→Book results reveals significant differences across algorithms, suggesting that measuring domain similarity should consider potential asymmetry between domains.

4 Conclusion

This study selects Chinese reviews from Amazon's book, DVD, and music categories as experimental data, proposes a deep recurrent neural network model for cross-domain sentiment analysis, and achieves an average classification accuracy of 81.70% in cross-domain environments. This outperforms traditional Stacked-LSTM, Bi-LSTM, CNN-LSTM serial, and CNN-LSTM parallel models, demonstrating the model's effectiveness.

A limitation of this study is that the corpus is primarily Chinese. Future research will validate the model on multilingual corpora and test it in environments with larger domain gaps (e.g., electronics vs. restaurant reviews). Additionally, we will explore more deep learning models (e.g., generative adversarial networks) to further enrich cross-domain text sentiment analysis research.

References

[1] Weiss K, Khoshgoftaar TM, Wang DD. A survey of transfer learning[J]. Journal of big data, 2016, 3(1): 1-40.

[2] Tahmoresnezhad J, Hashemi S. Visual domain adaptation via transfer feature learning[J]. Knowledge and information systems, 2017, 50(2): 1-21.

[3] Bifet A, Pechenizkiy M, Bouchachia A. A survey on concept drift adaptation[J]. ACM computing surveys, 2014, 46(4): 1-44.

[4] Barddal JP, Gomes HM, Enembreck F, et al. A survey on feature drift adaptation: definition, benchmark, challenges and future directions[J]. Journal of systems and software, 2016, 127: 278-294.

[5] Vapnik V, Izmailov R. Knowledge transfer in SVM and neural networks[J]. Annals of mathematics and artificial intelligence, 2017, 80(1-2): 3-19.

[6] Xu J, Ding Y, Wang X. Using machine learning methods for automatic news sentiment classification[J]. Chinese information processing, 2007, 21(6): 95-100.

[7] Wu Y, Huang Y, Wang X. A review of online Chinese user comment research: from a sentiment computation perspective[J]. Information science, 2017(6): 159-163.

[8] Blitzer J, McDonald R, Pereira F. Domain adaptation with structural correspondence learning[C]//Conference on empirical methods in natural language processing. Stroudsburg, PA, USA: Association for Computational Linguistics, 2006: 120-128.

[9] Pan SJ, Ni X, Sun JT, et al. Cross-domain sentiment classification via spectral feature alignment[C]//International conference on World Wide Web, WWW 2010, Raleigh, North Carolina, U.S.A., April. DBLP, 2010: 751-760.

[10] Dai W, Yang Q, Xue GR, et al. Boosting for transfer learning[C]//International conference on machine learning. New York, NY, USA: ACM, 2007: 193-200.

[11] Ma F, Wu J, Yang G. Cross-domain sentiment orientation analysis based on dual selection strategy[J]. Journal of the China Society for Scientific and Technical Information, 2012, 31(11): 1202-1209.

[12] Glorot X, Bordes A, Bengio Y. Domain adaptation for large-scale sentiment classification: a deep learning approach[C]//International conference on machine learning. Madison, WI, USA: Omnipress, 2011: 1-9.

[13] Sun M, Tan Q, Ding R, et al. Cross-domain sentiment classification using deep learning approach[C]//International conference on cloud computing and intelligence systems. New York, NY, USA: IEEE, 2015: 60-64.

[14] Lü S, Yang L, Lin H. Research on cross-domain sentiment orientation analysis algorithm based on SimRank[J]. Chinese information processing, 2012, 26(6): 38-44.

[15] Wei X, Zhang S, Yang L, et al. Cross-domain text sentiment orientation analysis based on weighted SimRank[J]. Pattern recognition and artificial intelligence, 2013, 26(11): 1004-1009.

[16] Zhang Z. Cross-domain transfer learning for product review sentiment analysis[J]. New technology of library and information service, 2013(6): 49-54.

[17] Huang R, Kang S. An improved EM algorithm for cross-domain sentiment classification[J]. Application research of computers, 2017, 34(9): 2696-2699.

[18] Xia R, Zong C, Hu X, et al. Feature ensemble plus sample selection: domain adaptation for sentiment classification[J]. IEEE intelligent systems, 2013, 28(3): 10-18.

[19] Deshmukh JS, Tripathy AK. Entropy based classifier for cross-domain opinion mining[J]. Applied computing and informatics, 2017, 14(1): 55-64.

[20] Tang D, Qin B, Liu T. Deep learning for sentiment analysis: successful approaches and future challenges[J]. Wiley interdisciplinary reviews data mining and knowledge discovery, 2015, 5(6): 292-303.

[21] Yu C, Feng B, An L. Cross-domain sentiment analysis based on deep representation learning[J]. Data analysis and knowledge discovery, 2017(7): 73-81.

[22] Yu J, Jiang J. Learning sentence embeddings with auxiliary tasks for cross-domain sentiment classification[C]//Conference on empirical methods in natural language processing. Stroudsburg, PA, USA: ACL, 2016: 236-246.

[23] Wu F, Zhang Y, Hu X. Cross-domain lexical sentiment orientation discrimination method for review information[J]. Computer science, 2015, 42(6): 220-222.

[24] Feng C, Liang X, Li Y, et al. Cross-domain Chinese sentiment dictionary construction method based on word vectors[J]. Data acquisition and processing, 2017, 32(3): 579-587.

[25] Denecke K. Are SentiWordNet scores suited for multi-domain sentiment classification?[C]//International conference on digital information management. New York, NY, USA: IEEE, 2009: 1-6.

[26] Graves A. Supervised sequence labeling with recurrent neural networks[M]. Berlin, Heidelberg: Springer, 2012.

[27] Sutskever I, Vinyals O, Le QV. Sequence to sequence learning with neural networks[EB/OL]. [2017-06-30]. https://arxiv.org/pdf/1409.3215.

[28] Jozefowicz R, Vinyals O, Schuster M, et al. Exploring the limits of language modeling[EB/OL]. [2017-06-30]. https://arxiv.org/pdf/1602.02410.

[29] Dan G, Brunk C, Vinyals O, et al. Multilingual language processing from bytes[EB/OL]. [2017-06-30]. https://arxiv.org/pdf/1512.00103.

[30] Vinyals O, Toshev A, Bengio S, et al. Show and tell: a neural image caption generator[C]//IEEE Conference on Computer Vision and Pattern Recognition. New York, NY, USA: IEEE, 2015: 3156-3164.

[31] Yao Y, Huang Z. Bidirectional LSTM recurrent neural network for Chinese word segmentation[C]//International Conference on Neural Information Processing. Cham, Switzerland: Springer, 2016: 345-353.

[32] Cross J, Huang L. Incremental parsing with minimal features using Bi-Directional LSTM[C]//Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics. Stroudsburg, PA, USA: ACL, 2016: 32-37.

[33] Plank B, Søgaard A, Goldberg Y. Multilingual part-of-speech tagging with bidirectional long short-term memory models and auxiliary loss[C]//Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics. Stroudsburg, PA, USA: ACL, 2016: 1-7.

[34] Abdi H, Williams LJ. Principal component analysis[J]. Wiley interdisciplinary reviews computational statistics, 2010, 2(4): 433-459.

[35] Ullrich C. Support vector classification[M]//Forecasting and hedging in the foreign exchange markets. Berlin Heidelberg: Springer, 2009: 345-356.

[36] Menard S. Applied logistic regression analysis[J]. Technometrics, 2002, 38(2): 184-186.

[37] Landgrebe D. A survey of decision tree classifier methodology[J]. IEEE transactions on systems man and cybernetics, 2002, 21(3): 660-674.

[38] Breiman L. Random Forest[J]. Machine learning, 2001, 45(1): 5-32.

[39] Jordan S, Viviana A, Wohar ME. Forecasting market returns: bagging or combining?[J]. International journal of forecasting, 2017, 33(1): 102-120.

[40] Wang J, Gao L, Zhang H, et al. Adaboost with SVM-based classifier for the classification of brain motor imagery tasks[C]//International conference on uni-

versal access in human-computer interaction: user diversity. Berlin Heidelberg: Springer-Verlag, 2011: 629-634.

[41] Lai S, Liu K, He S, et al. How to generate a good word embedding[J]. IEEE intelligent systems, 2016, 31(6): 5-14.

[42] Dyer C, Ballesteros M, Wang L, et al. Transition-based dependency parsing with stack long short-term memory[J]. Computer science, 2015, 37(2): 321-332.

[43] Wang J, Yu LC, Lai KR, et al. Dimensional sentiment analysis using a regional CNN-LSTM model[C]//Meeting of the Association for Computational Linguistics. Stroudsburg, PA, USA: ACL, 2016: 225-230.

A Cross-Domain Text Sentiment Analysis Based on Deep Recurrent Neural Network

Yu Chuanming
School of Information and Safety Engineering, Zhongnan University of Economics and Law, Wuhan 430073

Abstract: [Purpose/significance] In order to solve the problem of classification model in target domain caused by the lack of data, this study firstly trains the model of source domain that includes rich labeling/tagging data, and then, projects source and target domain documents in the same feature space. [Method/process] The reviews of three product categories, i.e. books, DVD and music, from Amazon, which are written in Chinese, are taken as the experimental data, and the cross-domain sentiment analysis is considered as the research task. A novel model, i.e. the Cross-Domain Deep Recurrent Neural Network (CD-DRNN), is proposed to achieve knowledge transfer among domains. The average accuracy value of CD-DRNN achieves 81.70%, which excels the values of Stacked Long Short-Term Memory (79.90%), Bidirectional Long Short-Term Memory (80.50%), Convolutional Neural Network with Long Short-Term Memory (74.70%) and Merged Convolutional Neural Network with Long Short-Term Memory (80.90%). [Result/conclusion] Knowledge transfer in source domain and target domain could effectively solve the difficulties of achieving good classification performances on small datasets. The proposed method can be leveraged to effectively select features from unlabeled data, thereby greatly reducing the workload related to data annotation in the target domain.

Keywords: cross-domain; transfer learning; deep recurrent neural network; sentiment analysis

Note: Figure translations are in progress. See original paper for figures.

Source: ChinaXiv — Machine translation. Verify with original.