

Cross-Lingual Sentiment Classification with Bilingual Document Representation Learning

Xinjie Zhou, Xianjun Wan and Jianguo Xiao

Institute of Computer Science and Technology, Peking University
The MOE Key Laboratory of Computational Linguistics, Peking University
{zhouxinjie, wanxiaojun, xiaojianguo}@pku.edu.cn

Abstract

Cross-lingual sentiment classification aims to adapt the sentiment resource in a resource-rich language to a resource-poor language. In this study, we propose a representation learning approach which simultaneously learns vector representations for the texts in both the source and the target languages. Different from previous research which only gets bilingual word embedding, our Bilingual Document Representation Learning model BiDRL directly learns document representations. Both semantic and sentiment correlations are utilized to map the bilingual texts into the same embedding space. The experiments are based on the multilingual multi-domain Amazon review dataset. We use English as the source language and use Japanese, German and French as the target languages. The experimental results show that BiDRL outperforms the state-of-the-art methods for all the target languages.

1 Introduction

Sentiment analysis for online user-generated contents has become a hot research topic during the last decades. Among all the sentiment analysis tasks, polarity classification is the most widely studied topic. It has been proved to be invaluable in many applications, such as opinion polling (Tang et al., 2012), customer feedback tracking (Gamon, 2004), election prediction (Tumasjan et al., 2010), stock market prediction (Bollen et al., 2011) and so on.

Most of the current sentiment classification systems are built on supervised machine learning algorithms which require manually labelled da-

ta. However, sentiment resources are usually unbalanced in different languages. Cross-lingual sentiment classification aims to leverage the resources in a resource-rich language (such as English) to classify the sentiment polarity of texts in a resource-poor language (such as Japanese). The biggest challenge for cross-lingual sentiment classification is the vocabulary gap between the source language and the target language. This problem is addressed with different strategies in different approaches. Wan (2009) use machine translation tools to translate the training data directly into the target language. Meng et al. (2012) and Lu et al. (2011) exploit parallel unlabeled data to bridge the language barrier. Prettenhofer and Stein (2010) use correspondence learning algorithm to learn a map between the source language and the target language. Recently, representation learning methods has been proposed to solve the cross-lingual classification problem (Xiao and Guo, 2013; Zhou et al., 2015). These methods aim to learn common feature representations for different languages. However, most of the current researches only focus on bilingual word embedding. In addition, these models only use the semantic correlations between aligned words or sentences in different languages while the sentiment correlations are ignored.

In this study, we propose a cross-lingual representation learning model BiDRL which simultaneously learns both the word and document representations in both languages. We propose a joint learning algorithm which exploits both monolingual and bilingual constraints. The monolingual constraints help to model words and documents in each individual language while the bilingual constraints help to build a consistent embedding space across languages.

For each individual language, we extend the paragraph vector model (Le and Mikolov, 2014)

to obtain word and document embeddings. The traditional paragraph vector model is fully unsupervised without using the valuable sentiment labels. We extend it into a semi-supervised manner by forcing the positive and negative documents to fall into different sides of a classification hyperplane. Learning task-specific embedding has been proved to be effective in previous research. To address the cross-language problem, different strategies are proposed to obtain a consistent embedding space across different languages. Both sentiment and semantic relatedness are exploited while previous studies only use the semantic connection between parallel sentences or documents.

The performance of BiDRL is evaluated on a multilingual multi-domain Amazon review dataset (Prettenhofer and Stein, 2010). By selecting English as the source language, a total of nine tasks are evaluated with different combinations of three different target languages and three different domains. The proposed method achieves the state-of-the-art performance on all the tasks.

The main contributions of this study are summarized as follows:

- 1) We propose a novel representation learning method BiDRL which directly learns bilingual document representations for cross-lingual sentiment classification. Different from previous studies which only obtain word embeddings, our model can learn vector representations for both words and documents in bilingual texts.

- 2) Our model leverages both the semantic and sentiment correlations between bilingual documents. Not only the parallel documents but also the documents with the same sentiment are required to get similar representations.

- 3) Our model achieves the state-of-the-art performances on nine benchmark cross-lingual sentiment classification tasks and it consistently outperforms the existing methods by a large margin.

2 Related Work

Sentiment analysis is the field of studying and analyzing people's opinions, sentiments, evaluations, appraisals, attitudes, and emotions (Liu, 2012). Most of the previous sentiment analysis researches focus on customer reviews and classifying the sentiment polarity is the most widely studied task (Pang et al., 2002).

Cross-lingual sentiment classification is a popular topic in the sentiment analysis community

which aims to solve the sentiment classification task from a cross-language view. It is of great importance for the area since it can exploit the existing labeled information in a source language to build a sentiment classification system in any other target language. It saves us from manually labeling data for all the languages in the world which is expensive and time-consuming. Cross-lingual sentiment classification has been extensively studied in the very recent years. Mihalcea et al. (2007) translate English subjectivity words and phrases into the target language to build a lexicon-based classifier. Banea et al. (2010) also use the machine translation service to obtain parallel corpus. It investigates several questions based on the parallel corpus including both the monolingual sentiment classification and cross-lingual sentiment classification. Wan (2009) translates both the training data (English to Chinese) and the test data (Chinese to English) to train different models in both the source and target languages. The co-training algorithm (Blum and Mitchell, 1998) is used to combine the bilingual models together and improve the performance. In addition to the translation-based methods, several studies utilize parallel corpus or existing resources to bridge the language barrier. Balamurali (2012) use WordNet senses as features for supervised sentiment classification. They use the linked WordNets of two languages to bridge the language gap. Lu et al. (2011) consider the multilingual scenario where small amount of labeled data is available in the target language. They attempted to jointly classify the sentiment for both source language and target language. Meng et al. (2012) propose a generative cross-lingual mixture model to leverage unlabeled bilingual parallel data. Prettenhofer and Stein (2010) use the structural correspondence learning algorithm to learn a map between the source language and the target language. Xiao and Guo (2014) treat the bilingual feature learning problem as a matrix completion task.

This work is also related to bilingual representation learning. Zou et al. (2013) propose to use word alignment as the constraints in bilingual word embedding. Each word in one language should be similar to the aligned words in another language. Gouws et al. (2015) propose a similar algorithm but only use sentence-level alignment. It tries to minimize a sampled L2-loss between the bag-of-words sentence vectors of the parallel cor-

pus. Xiao and Guo (2013) learn different representations for words in different languages. Part of the word vector is shared among different languages and the rest is language-dependent. Klementiev et al. (2012) treat the task as a multi-task learning problem where each task corresponds to a single word, and task relatedness is derived from co-occurrence statistics in bilingual parallel data. Hermann and Blunsom (2015) propose the bilingual CVM model which directly minimizes the representation of a pair of parallel documents. The document representation is calculated with a composition function based on words. Chandar A P et al. (2014) and Zhou et al. (2015) use the autoencoder to model the connections between bilingual sentences. It aims to minimize the reconstruction error between the bag-of-words representations of two parallel sentences. Luong et al. (2015) propose the bilingual skip-gram model which leverages the word alignment between parallel sentences. Pham et al. (2015) extend the paragraph vector model to force bilingual sentences to share the same sentence vector.

This study differs with the existing works in the following three aspects, 1) we exploit both the semantic and sentiment correlations of the bilingual texts. Existing bilingual embedding algorithms only use the semantic connection between parallel sentences or documents. 2) Our algorithm learns both the word and document representations. Most of the previous studies simply compute the average of the word vectors in a document. 3) Sentiment labels are used in our embedding algorithm by introducing a classification hyperplane. It not only helps to achieve better embedding performance in each individual language but also helps to bridge the language barrier.

3 Framework

Firstly we introduce several notations used in BiDRL. Let S and S_u denote the documents from the training dataset and the documents from the unlabeled dataset in the source language respectively. For each document $d \in S$, it has a sentiment label $y \in \{1, -1\}$. We denote the sentiment label set of all the documents in S as Y . Let T and T_u denote the documents from the test dataset and the documents from the unlabeled dataset in the target language. The documents in the training and test datasets in the source and target languages are translated into the other language using the on-

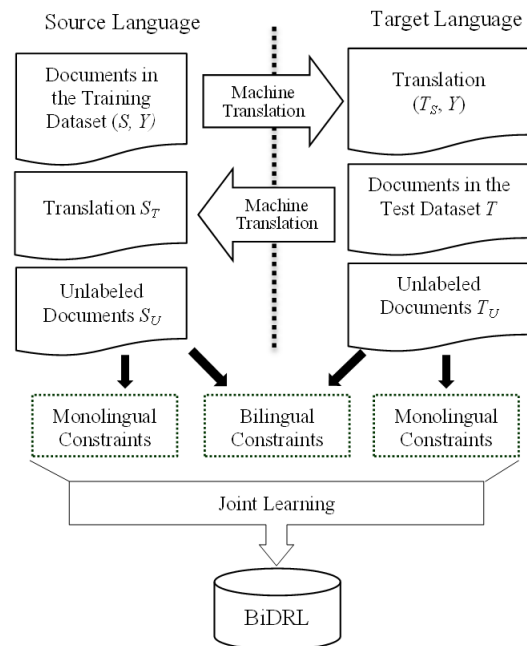


Figure 1: Framework of BiDRL

line machine translation service - Google Translate¹. We denote them as T_s (the translation of S) and S_t (the translation of T). We wish to learn a D -dimensional vector representation for all the documents in the dataset.

The general framework of BiDRL is shown in Figure 1. After we obtain the data in the source and target languages, we propose both the monolingual and bilingual constraints to learn the model. The monolingual constraints help to model words and documents in each individual language. The bilingual constraints help to build a consistent embedding space across different languages. The joint learning framework is semi-supervised which uses the sentiment labels Y of the training documents.

3.1 Monolingual Constraints

In this subsection, we describe the representation learning algorithm for the source and target languages. We start from the paragraph vector model (Le and Mikolov, 2014) which has been proved to be one of the state-of-the-art methods for document modeling. In the paragraph vector framework, both documents and words are mapped to unique vectors. Each document is treated as a unique token which is the context of all the words in the document. Therefore, each word in the doc-

¹<http://translate.google.com/>

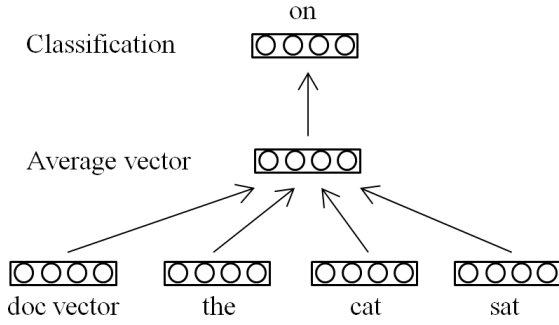


Figure 2: Paragraph vector

ument can be predicted by the context tokens and the document as shown in Figure 2. The idea leads to the following function,

$$\arg \max \sum_{d \in D} \sum_{(w, C) \in d} \log p(w | C, d) \quad (1)$$

where D is the document set, w and C are the word and its context in a document d . The only difference between paragraph vector algorithm and the well-known `word2vec` algorithm (Mikolov et al., 2013) is the additional document vector.

The conditional probability of predicting a word from its context is modeled via softmax which is very expensive to compute. It is usually approximately solved via negative sampling. A log-bilinear model is used instead to predict whether two words are in the same context. For a word and context pair (w, C) in a document d , the objective function becomes,

$$L_1 = -\log \sigma(v_w^T \cdot \frac{1}{k+1} \cdot (v_d + \sum_{c \in C} v_c)) - \sum_{i=1}^n E_{w' \in p_n(w)} (\log \sigma(-v_{w'}^T \cdot \frac{1}{k+1} \cdot (v_d + \sum_{c \in C} v_c))) \quad (2)$$

where $\sigma(\cdot)$ is the sigmoid function, c is a context word in C , k is the window size, v_w and v_c are the vectors for words and context words, v_d is the vector for the document d , w' is the negative sample from the noise distribution $P_n(w)$. Mikolov et al. (2013) set $P_n(w)$ as the $\frac{3}{4}$ power of the unigram distribution which outperforms the unigram and the uniform distribution significantly.

The paragraph vector model captures the semantic relations between words and documents. In addition, we hope that the vector representation

of a document can directly reflect its sentiment. The document vectors associated with different sentiments should fall into different positions in the embedding space. We introduce a logistic regression classifier into the embedding algorithm. For each labeled document d with the label y , we use it to classify the sentiment based on the cross entropy loss function,

$$L_2 = -y \log \sigma((x^T v_d + b)) - (1 - y) \log \sigma(-x^T v_d - b) \quad (3)$$

where x is the weight vector for the features and b is the bias, v_d is the document vector.

Combining the above two loss functions, we obtain the monolingual embedding algorithm. For the source language, we get the word and document representations via

$$\arg \min L_s = \sum_{d \in S \cup S_u} \sum_{(w, C) \in d} L_1 + \sum_{d \in S} L_2 \quad (4)$$

The embedding for the target language is obtained similarly,

$$\arg \min L_t = \sum_{d \in T \cup T_u} \sum_{(w, C) \in d} L_1 + \sum_{d \in T_s} L_2 \quad (5)$$

where the labeled dataset T_s is translated from the source language.

3.2 Bilingual Constraints

The key problem of bilingual representation learning is to obtain a consistent embedding space across the source and the target languages. We propose three strategies for BiDRL to bridge the language gap.

The first strategy is to share the classification hyperplane in Equation 3. The logistics regression parameter x and b are the same in the source and the target languages. By sharing the same classification parameter, bilingual documents with the same sentiment will fall into similar areas in the embedding space. Therefore, introducing the logistic regression classifier in our embedding algorithm not only helps to obtain task-specific embedding but also helps to narrow the language barrier.

The second strategy is to minimize the difference between a pair of parallel documents, i.e., the

original documents and the translated documents. Such word or document based regularizer is widely used in previous works (Gouws et al., 2015; Zou et al., 2013). We simply measure the differences of two documents via the Euclidean Distance. The parallel documents refer to the documents in S and T and their corresponding translations (S_t and T_s). It leads to the following loss function,

$$L_3 = \sum_{(d_s, d_t) \in (S, T_s)} \|v_{d_s} - v_{d_t}\|^2 + \sum_{(d_s, d_t) \in (S_t, T)} \|v_{d_s} - v_{d_t}\|^2 \quad (6)$$

where (d_s, d_t) is a pair of parallel documents and v_* is their vector representations.

Our third strategy aims to generate similar representations for texts with the same sentiment. The traditional paragraph vector model focuses on modeling the semantic relationship between words and documents while our method aims to preserve the sentiment relationship as well. For each document $d \in S$ in the source language, we find K least similar documents with the same sentiment. The document similarity is measured by cosine similarity using TF-IDF features. We hope that these K documents should have similar representation with document d despite of their textual difference. We denote the K documents as Q_s and their parallel documents in the target language as Q_t . It leads to the following loss function,

$$L_4 = \sum_{d \in S} \left(\sum_{d_s \in Q_s} \|v_{d_s} - v_d\|^2 + \sum_{d_t \in Q_t} \|v_{d_t} - v_d\|^2 \right) \quad (7)$$

where v_* denotes the vector representation of a document.

Combining the monolingual embedding algorithm and the cross-lingual correlations, we have the overall objective function as follows,

$$\arg \min L = L_s + L_t + L_3 + L_4 \quad (8)$$

which can be solved using stochastic gradient descent (SGD).

After learning BiDRL, we represent each document in the training and test dataset by the concatenation of its vector representation in both the source and the target languages. In particular, for each training document $d \in S$, we represent it as $[v_d v_{d'}]$ where d' is its corresponding translation and v_* is the learned document representation in BiDRL. Similarly, for each test document $d \in T$, we represent it as $[v_{d'} v_d]$. Afterwards, a logistic regression classifier is trained using the concatenated feature vectors of S . The polarity of the reviews in T can be predicted by applying the classifier on the concatenated feature vectors of T .

4 Experiments

4.1 Dataset

We use the multilingual multi-domain Amazon review dataset² created by (Prettenhofer and Stein, 2010). It contains three different domains book, DVD and music. Each domain has reviews in four different languages English, German, French and Japanese. In our experiments, we use English as the source language and the rest three as target languages. Therefore, we have a total of nine tasks with different combinations of three domains and three target languages. For each task, the training and test datasets have 1000 positive reviews and 1000 negative reviews. There are also several thousand of unlabeled reviews but the quantity of them varies significantly for different tasks. Following (Prettenhofer and Stein, 2010), when there are more than 50000 unlabeled reviews we randomly selected 50000 of them, otherwise we use all the unlabeled reviews. The detailed statistics of the dataset are shown in Table 1.

We translated the 2000 training reviews and 2000 test reviews into the other languages using Google Translate. Prettenhofer and Stein (2010) has already provided the translation of the test data. We only need to translate the English training data into the three target languages. All the review texts are tokenized and converted into lowercase. We use Mecab³ to segment the Japanese reviews.

4.2 Implementation

In the bilingual representation learning algorithm, we set the vector size as 200 and the context windows as 10. The learning rate is set to 0.025 fol-

²<https://www.uni-weimar.de/medien/webis/corpora/corpus-webis-cls-10/>

³<http://taku910.github.io/mecab/>

| Target Language | Domain | MT-BOW | MT-PV | CL-SCL | BSE | CR-RL | Bi-PV | BiDRL |
|------------------|--------|--------|-------|--------|-------|-------|-------|-------|
| German | book | 79.68 | 79.90 | 79.50 | 80.27 | 79.89 | 79.51 | 84.14 |
| | DVD | 77.92 | 80.09 | 76.92 | 77.16 | 77.14 | 78.60 | 84.05 |
| | music | 77.22 | 80.71 | 77.79 | 77.98 | 77.27 | 82.45 | 84.67 |
| French | book | 80.76 | 80.14 | 78.49 | - | 78.25 | 84.25 | 84.39 |
| | DVD | 78.83 | 81.49 | 78.80 | - | 74.83 | 79.60 | 83.60 |
| | music | 75.78 | 81.92 | 77.92 | - | 78.71 | 80.09 | 82.52 |
| Japanese | book | 70.22 | 67.45 | 73.09 | 70.75 | 71.11 | 71.75 | 73.15 |
| | DVD | 71.30 | 68.86 | 71.07 | 74.96 | 73.12 | 75.40 | 76.78 |
| | music | 72.02 | 74.53 | 75.11 | 77.06 | 74.38 | 75.45 | 78.77 |
| Average Accuracy | | 75.97 | 77.23 | 76.52 | 74.26 | 76.08 | 78.57 | 81.34 |

Table 2: Cross-lingual sentiment classification accuracy for the nine tasks. For all the methods, we get ten different runs of the algorithm and calculate the mean accuracy.

| Target Language | Domain | $ S_U $ | $ T_U $ |
|-----------------|--------|---------|---------|
| German | book | 50000 | 50000 |
| | DVD | 30000 | 50000 |
| | music | 25000 | 50000 |
| French | book | 50000 | 32000 |
| | DVD | 30000 | 9000 |
| | music | 25000 | 16000 |
| Japanese | book | 50000 | 50000 |
| | DVD | 30000 | 50000 |
| | music | 25000 | 50000 |

Table 1: The amount of unlabeled reviews used in the experiments. There are also 1000 positive and 1000 negative reviews both for training and test in each task, i.e. $|S| = |T| = 2000$.

lowing `word2vec` and it declines with the training procedure. K is empirically chosen as 10. The algorithm runs 10 iterations on the dataset.

4.3 Baseline

We introduce several state-of-the-art methods used for comparison in our experiment as follows.

MT-BOW: It learns a classifier in the source language using bag-of-words features and the test data is translated into the source language via Google Translate. We directly use the experimental results reported in (Prettenhofer and Stein, 2010).

MT-PV: We translate the training data into the target language and also translate the test data into the source language. In both the source and target languages, we use the paragraph vector model to learn the vector representation of the documents. Therefore, each document can be represented by

the concatenation of the vector in two languages. A logistic regression classifier is trained using the concatenated feature vectors similarly to BiDRL. MT-PV can be regarded as a simplified version of BiDRL without the L_2 , L_3 and L_4 regularizers.

CL-SCL: It is the cross-lingual structural correspondence learning algorithm proposed by (Prettenhofer and Stein, 2010). It learns a map between the bag-of-words representations in the source and the target languages. It also leverages Google Translate to obtain the word translation oracle.

BSE: It is the bilingual embedding method of (Tang et al., 2012). It aims to learn two different mapping matrices for the source and target languages. The two matrices map the bag-of-words representations in the source and the target languages into the same feature space. Tang et al. (2012) only report their results on 6 of the 9 tasks.

CR-RL: It is the bilingual word representation learning method of (Xiao and Guo, 2013). It learns different representations for words in different languages. Part of the word vector is shared among different languages and the rest is language-dependent. The document representation is calculated by taking average over all words in the document.

Bi-PV: Pham et al. (2015) extended the paragraph vector model into bilingual setting by sharing the document representation of a pair of parallel documents. Their method requires large amounts of parallel data and does not need the machine translation service during test phase. In our setting, there are not enough parallel data to train the model and it will lead to an unfair comparison without using the machine translated text. We implement a variant of their method which learns the

vector representation for the training and test data using both the original and the translated texts. Each pair of parallel documents shares the same document representation.

We also implement the method of (Zhou et al., 2015) which is originally designed for the English-Chinese cross-lingual sentiment classification task. We find that it is not very adaptable in our case because the negation pattern and sentiment words are hard to choose for our target languages. The results of our replication do not achieve comparable results with the rest methods and are not listed here to avoid misleading the readers.

4.4 Results and Analysis

Table 2 shows the experimental results for all the baselines and our method. For all the nine tasks, our bilingual document representation learning method achieves the state-of-the-art performance. The two most simple approaches MT-BOW implemented by (Prettenhofer and Stein, 2010) and MT-PV implemented by us are both strong baselines. They achieve comparable results with the more complex baselines on many tasks. MT-PV performs better than MT-BOW on most tasks which proves that the representation learning method is more useful than the traditional bag-of-words features.

The three word-based representation learning methods CL-SCL, BSE and CR-RL achieve similar results with the simple model MT-BOW and only outperform it on some tasks. However, the document representation learning methods MT-PV, Bi-PV and BiDRL performs much better. It shows that capturing the compositionality of words is important for sentiment classification. The isolated word representations are not enough to model the whole document. The Bi-PV model outperforms MT-PV on most tasks and shows that the authors idea of learning a single representation for a pair of parallel documents is more useful than learning them separately.

For all the baselines and our method, the performance of the English-Japanese tasks is lower than that of the English-German and English-French tasks. It is reasonable because the English language is much closer to German and French than Japanese. The machine translation tool also performs better when translating between the Western languages.

Our BiDRL model outperforms all the existing methods on all the tasks. The accuracy is over 80% on all the six tasks for the two European target languages. The mean accuracy of the nine tasks shows a significant gap between BiDRL and the existing models. It achieves an improvement about 3% compared to the previous state-of-the-art methods.

4.5 Parameter Sensitivity Study

In this subsection, we investigate the influence of the vector size of our representation learning algorithm. We conduct the experiments by changing the vector size from 50 to 400. For each parameter setting, we run the algorithm for ten times and get the mean accuracy.

The results of MT-PV and BiDRL on all the nine tasks are shown in Figure 3. For almost all the tasks, we can observe that our model BiDRL steadily outperforms the strong baseline MT-PV. It proves the efficacy of our bilingual embedding constraints.

For most of the nine tasks including DE-DVD, DE-MUSIC, FR-DVD, FR-MUSIC and JP-MUSIC, the performance of BiDRL increases with the growth of the vector size at the beginning and remains stable afterwards. For the rest tasks, our model responds less sensitively to the change of the vector size and the prediction accuracy keeps steady. However, the results of MT-PV show no regular patterns with the change of the vector size which makes it hard to choose a satisfying parameter value.

The parameter K is empirically chosen as 10 because we find that its value has little influence to our model when it is chosen between 10 and 50. Selecting a small K will help to accelerate the training procedure.

4.6 Analysis of the Sentiment Information

The traditional paragraph vector model only models the semantic relatedness between texts via the word co-occurrence statistics. In this study, we propose to learn the bilingual representation utilizing the sentiment information. Firstly, we introduce a classification hyperplane to separate the embedding of texts with different polarities, i.e the loss function L_2 . Secondly, we consider the texts with the same sentiment but has largely different textual expressions. They are forced to have similar representations, i.e. L_4 . Table 3 shows the

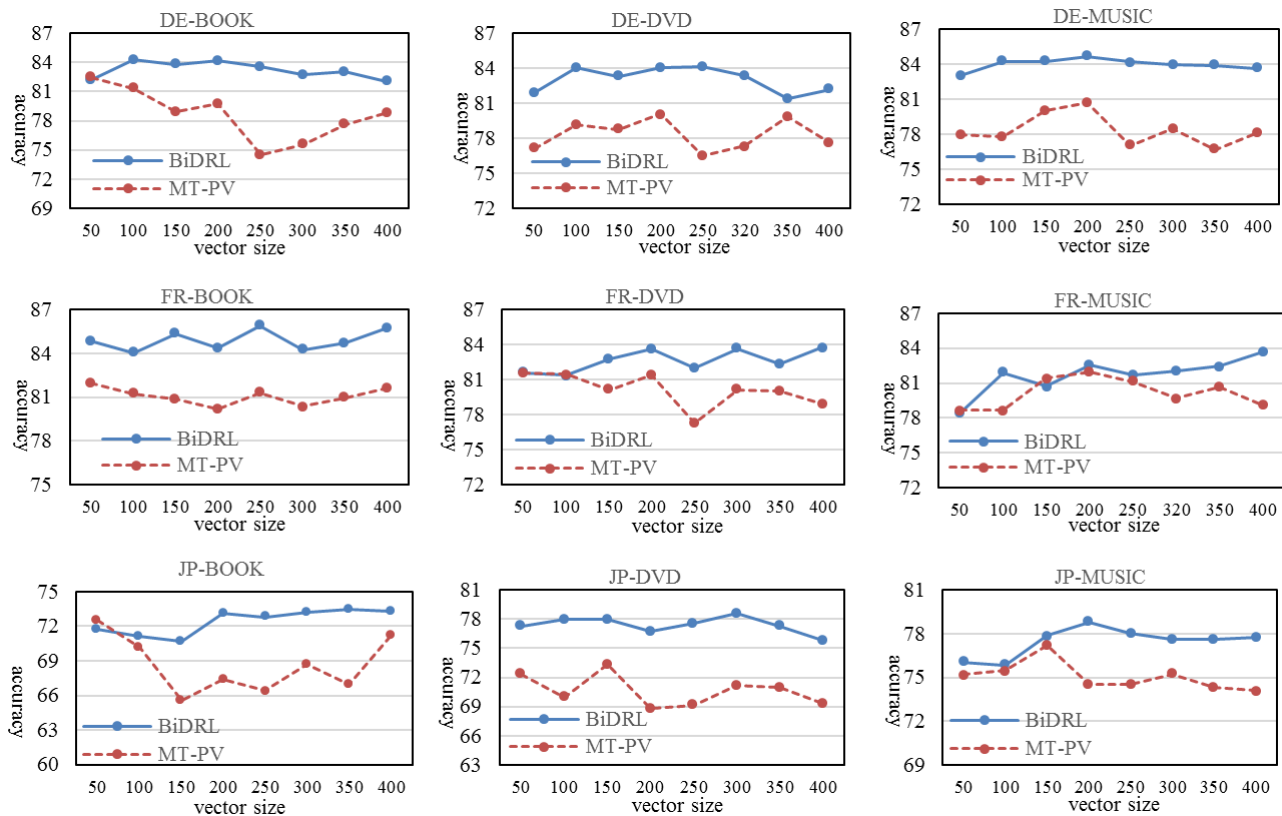


Figure 3: Influence of vector size for the nine cross-lingual sentiment classification tasks

results of our model without using these sentiment information.

| Model | MT-PV | BiDRL- L_2 | BiDRL- L_4 | BiDRL |
|----------|-------|--------------|--------------|-------|
| Accuracy | 77.23 | 79.51 | 80.43 | 81.34 |

Table 3: Influence of the sentiment information. We only show the mean accuracy of the nine tasks due to space limit.

MT-PV can be regarded as BiDRL without all the sentiment information. It achieves lower results than the other three methods. We can also observe that removing L_2 or L_4 both decreases the accuracy. It proves that the sentiment information helps BiDRL to achieve better results.

5 Conclusion and Future Work

In this study, we propose a bilingual document representation learning method for cross-lingual sentiment classification. Different from previous studies which only get bilingual word embeddings, we directly learn the vector representation for documents in different languages. We propose three strategies to achieve a consistent embedding space for the source and target languages. Both

sentiment and semantic correlations are exploited in our algorithm while previous works only use the semantic relatedness between parallel documents. Our model is evaluated on a benchmarking dataset which contains three different target languages and three different domains. Several state-of-the-art methods including several bilingual representation learning models are used for comparison. Our algorithm outperforms all the baseline methods on all the nine tasks in the experiment.

Our future work will focus on extending the bilingual document representation model into the multilingual scenario. We will try to learn a single embedding space for a source language and multiple target languages simultaneously. In addition, we will also explore the possibility of using more complex neural network models such as convolutional neural network and recurrent neural network to build bilingual document representation system.

Acknowledgments

The work was supported by National Natural Science Foundation of China (61331011), National Hi-Tech Research and Development Pro-

gram (863 Program) of China (2015AA015403, 2014AA015102) and IBM Global Faculty Award Program. We thank the anonymous reviewers for their helpful comments. Xiaojun Wan is the corresponding author.

References

- AR Balamurali. 2012. Cross-lingual sentiment analysis for indian languages using linked wordnets. In *Proceedings of 2012 International Conference on Computational Linguistics (COLING)*, pages 73–82.
- Carmen Banea, Rada Mihalcea, and Janyce Wiebe. 2010. Multilingual subjectivity: Are more languages better? In *Proceedings of the 23rd international conference on computational linguistics*, pages 28–36. Association for Computational Linguistics.
- Avrim Blum and Tom Mitchell. 1998. Combining labeled and unlabeled data with co-training. In *Proceedings of the eleventh annual conference on Computational learning theory*, pages 92–100. ACM.
- Johan Bollen, Huina Mao, and Xiaojun Zeng. 2011. Twitter mood predicts the stock market. *Journal of Computational Science*, 2(1):1–8.
- Sarath Chandar A P, Stanislas Lauly, Hugo Larochelle, Mitesh Khapra, Balaraman Ravindran, Vikas C. Raykar, and Amrita Saha. 2014. An autoencoder approach to learning bilingual word representations. In *Advances in Neural Information Processing Systems*, pages 1853–1861.
- Michael Gamon. 2004. Sentiment classification on customer feedback data: noisy data, large feature vectors, and the role of linguistic analysis. In *Proceedings of the 20th international conference on Computational Linguistics*, page 841. Association for Computational Linguistics.
- Stephan Gouws, Yoshua Bengio, and Greg Corrado. 2015. Bilbowa: Fast bilingual distributed representations without word alignments. In *Proceedings of The 32nd International Conference on Machine Learning*, pages 748–756.
- Karl Moritz Hermann and Phil Blunsom. 2015. Multilingual models for compositional distributed semantics. In *Proceedings of 52rd Annual Meeting of the Association for Computational Linguistic*, pages 58–68.
- Alexandre Klementiev, Ivan Titov, and Binod Bhattarai. 2012. Inducing crosslingual distributed representations of words. In *Proceedings of COLING-2012*, pages 1459–1474.
- Quoc Le and Tomas Mikolov. 2014. Distributed representations of sentences and documents. In *Proceedings of the 31st International Conference on Machine Learning (ICML-14)*, pages 1188–1196.
- Bing Liu. 2012. Sentiment analysis and opinion mining. *Synthesis lectures on human language technologies*, 5(1):1–167.
- Bin Lu, Chenhao Tan, Claire Cardie, and Benjamin K. Tsou. 2011. Joint bilingual sentiment classification with unlabeled parallel corpora. In *Proceedings of the 49th Annual Meeting of the Association for Computational Linguistics: Human Language Technologies-Volume 1*, pages 320–330. Association for Computational Linguistics.
- Thang Luong, Hieu Pham, and Christopher D. Manning. 2015. Bilingual word representations with monolingual quality in mind. In *Proceedings of the 1st Workshop on Vector Space Modeling for Natural Language Processing*, pages 151–159.
- Xinfan Meng, Furu Wei, Xiaohua Liu, Ming Zhou, Ge Xu, and Houfeng Wang. 2012. Cross-lingual mixture model for sentiment classification. In *Proceedings of the 50th Annual Meeting of the Association for Computational Linguistics: Long Papers-Volume 1*, pages 572–581. Association for Computational Linguistics.
- Rada Mihalcea, Carmen Banea, and Janyce Wiebe. 2007. Learning multilingual subjective language via a cross-lingual projections. In *Proceedings of ACL-2007*, pages 976–983.
- Tomas Mikolov, Ilya Sutskever, Kai Chen, Greg S. Corrado, and Jeff Dean. 2013. Distributed representations of words and phrases and their compositionality. In *Advances in neural information processing systems*, pages 3111–3119.
- Bo Pang, Lillian Lee, and Shivakumar Vaithyanathan. 2002. Thumbs up?: sentiment classification using machine learning techniques. In *Proceedings of the ACL-02 conference on Empirical methods in natural language processing-Volume 10*, pages 79–86. Association for Computational Linguistics.
- Hieu Pham, Minh-Thang Luong, and Christopher D. Manning. 2015. Learning distributed representations for multilingual text sequences. In *Proceedings of NAACL-HLT*, pages 88–94.
- Peter Prettenhofer and Benno Stein. 2010. Cross-language text classification using structural correspondence learning. In *Proceedings of the 48th Annual Meeting of the Association for Computational Linguistics*, pages 1118–1127. Association for Computational Linguistics.
- Jie Tang, Yuan Zhang, Jimeng Sun, Jinghai Rao, Wenjing Yu, Yiran Chen, and Alvis Cheuk M. Fong. 2012. Quantitative study of individual emotional states in social networks. *Affective Computing, IEEE Transactions on*, 3(2):132–144.
- Andranik Tumasjan, Timm Oliver Sprenger, Philipp G. Sandner, and Isabell M. Welpe. 2010. Predicting elections with twitter: What 140 characters reveal about political sentiment. *ICWSM*, 10:178–185.

- Xiaojun Wan. 2009. Co-training for cross-lingual sentiment classification. In *Proceedings of the Joint Conference of the 47th Annual Meeting of the ACL and the 4th International Joint Conference on Natural Language Processing of the AFNLP: Volume 1-Volume 1*, pages 235–243. Association for Computational Linguistics.
- Min Xiao and Yuhong Guo. 2013. Semi-supervised representation learning for cross-lingual text classification. In *Proceedings of EMNLP-2013*, pages 1465–1475. Association for Computational Linguistics.
- Min Xiao and Yuhong Guo. 2014. Semi-supervised matrix completion for cross-lingual text classification. In *Twenty-Eighth AAAI Conference on Artificial Intelligence*, pages 1607–1614.
- Huiwei Zhou, Long Chen, Fulin Shi, and Degen Huang. 2015. Learning bilingual sentiment word embeddings for cross-language sentiment classification. In *Proceedings of the 53rd Annual Meeting of the Association for Computational Linguistics*, pages 433–440. Association for Computational Linguistics.
- Will Y. Zou, Richard Socher, Daniel M. Cer, and Christopher D. Manning. 2013. Bilingual word embeddings for phrase-based machine translation. In *EMNLP*, pages 1393–1398.