Cross-lingual Sentiment Analysis via AAE and BiGRU

Jianghong Shen*
Key Laboratory of Optoelectronic Science and Technology for Medicine of Ministry of Education,
Fujian Provincial Engineering Technology Research Center of Photoelectric Sensing Application
College of Photonic and Electronic Engineering
Fujian Normal University
Fuzhou, China
jianghongshen@139.com

Xiaodong Liao
College of Photonic and Electronic Engineering
Fujian Normal University
Fuzhou, China
liaoxd@fjnu.edu.cn

Shuai Lei
College of Photonic and Electronic Engineering
Fujian Normal University
Fuzhou, China
eishuai915171505@163.com

***Abstract*—As a significant task in nature language processing (NLP), sentiment analysis has got more and more attention. Sufficient data and helpful tools are available when doing the studying. Therefore, it is meaningful to study techniques to leverage labeled data and models valid in rich datasets language when addressing any questions in rare language. As a result, cross-lingual sentiment analysis has become more and more popular. Compared with traditional translation method, transfer learning is the main trend and using deep learning to generate a cross-lingual word embeddings in the single vector space is more useful and stable. Hence getting the high quality of cross-lingual word embedding is the big problem which needs to be settled urgently. In this paper, we use the LSTM and Adversarial Auto Encoders (AAE) to generate contextual cross-lingual word embeddings for transfer learning, then the BiGRU is used for analyzing the sentiment. Experimental results prove that our method has the best performance.**

Keywords—contextual cross -lingual word embedding, transfer learning, sentiment analysis, AAE, BiGRU

I. Introduction

As the territory of research, sentiment analysis aims to analyze writers' minds, sentiments, attitudes and etc, which are showed inside the text [1][2]. Because of huge improvement of science and technology, people from all over the word are more likely to express their comments at any time, which has drawn much attention both in academic world and in business & society. For example, online shops in Amazon can sell their stuffs to all over the word. Meanwhile, they could receive comments with different languages, such as "This skirt is beautiful", "Este vestido es hermoso". Although they are in different languages, they show people's nature opinion. It means that mining the sentiment information among different languages is helpful for official website or merchants to recommend the right products for their customers. English resource is rich and has abundant annotated corpora, so researchers have done large study based on it. But it's not easy to analyze other sparse data languages, it is necessary to enable knowledge transfer from languages which has rich training data to scarce data languages. Therefore, cross-lingual sentiment analysis becomes more and more important.

For the past few years, many researchers have committed themselves to this studying and have acquired much progress. The traditional way to analyze cross-lingual sentiment is machine translation [3]. However, it is easy affected by the quality of machine translation. Deep learning gives a new way to solve the problem, namely mapping different language into a shared embedding space. Recent researchers are mainly focused on mapping. Obtaining high quality of cross-lingual embeddings, which designed to represent words in different languages in a shared vector space by catching semantic similarities across languages, is the core of the task. Hence there is the hypothesis that the embedding spaces of multifarious languages display a similar structure [4]. Former works aim at learning a linear map, which schedules independently to generate monolingual space into a single shared space by means of a seed translation dictionary [5].

The word embedding is the critical factor during transfer learning, because it influences the degree of accuracy. Adversarial Auto Encoders (AAE) proposed by Makhzani etc, is conductive to learning high quality of cross-lingual embeddings [6], so we carry out the work relied on deep learning. The work is separated into two parts, which is cross-lingual embedding and transfer classifier.

Below are the remaining parts of this paper: firstly, retrospect related work in Section 2; secondly, we express our model in Section 3; thirdly, experiments and results are showed in Section 4; finally, summarizing the paper and sketching research orientation in future are arranged in the Section 5.

II. Related Work

A. Cross-lingual Sentiment Analysis

Different from the monolingual sentiment analysis, cross-lingual Sentiment Analysis is more complex, which forecasts the sentiment classes of low-resource language according to the annotated rich-resource language. Since lack of labeled data is a key problem for the task of sentiment analysis, there are two prime means to solve the issue, namely machine translation and mapping method through transfer learning.

Machine translation method means we translate the source language into target language, or translate target language into source language to get the embedding in the same space, which is convenient to do the downstream task. However, Zhang find this method may sometimes change the polarity of source [7], which means the quality of translation is a key issue. For example, people complain that it's too fragrant to sleep, translate the sentence into Chinese by Google translation and the result is "shui de tai xiang le", which is positive sentiment.

The state-of-the-art ways are mainly based on mapping method, namely learning a transformation to map different language into a shared space [8]. Subsequent research focused on improving the quality of mapping. Mikel Artetxe aimed to learn principled bilingual mappings through a series of linear transformations [9][10]. Learning the map is the significant issue in the task of cross-lingual sentiment analysis. Put source language and target language into a same space, and then get the cross-lingual word embeddings. Adversarial Auto Encoders (AAE) is conducive to improve the performance of mapping [6].

B. Deep learning

In comparison with traditional feature skills, deep learning technique are good at extracting features from sentence or text automatically, which is a wonderful way to save cost, such as time, labor and money. For the past few years, it attracted a large number of researchers and had acquired wonderful models in many tasks of NLP, especially in task of sentiment analysis.

As an extraordinary kind of RNN, Long Short Term Memory networks (LSTM) [11] is able to mitigate the trouble of long-term dependencies. LSTM is proposed by Hochreiter and Schmidhuber (1997). The cell states, which move through the top of the graph like a horizontal line, are the core of LSTM. It works like a conveyor belt, which moves straight down the total chain with only some minor linear interactions. Meanwhile, data flows among it easily. There are three gates for a LSTM, which aim to omit or add data to the cell state. Check below Fig. 1 to get more visual details.

As a more notable variation on the LSTM, the Gated Recurrent Unit (GRU), proposed by Cho, et al. (2014), combines the input gate and forget gate into a single update gate and also combines the cell state and hidden state. Moreover, some other kinds of characteristics have been improved. So compared with normal LSTM models, the ultimate model GRU is more concise. Below Fig. 2 shows the interior construction of GRU.

C. Transfer learning

Currently, the state-of-the-art models for cross-lingual sentiment analysis are mainly relied on supervised means. Train the model in rich-source language and then fine-tune the parameter to match the rare language. Transfer learning is used between two different languages. Kim presented a frame model for transfer learning, which utilized shared word embeddings between different languages and used adversarial training to surmount the discrepancy between different languages during training [12].

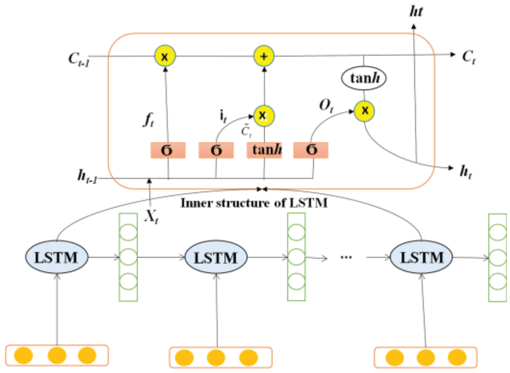


Fig. 1. Structure of LSTM

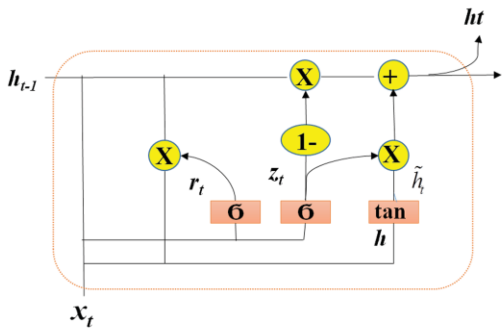
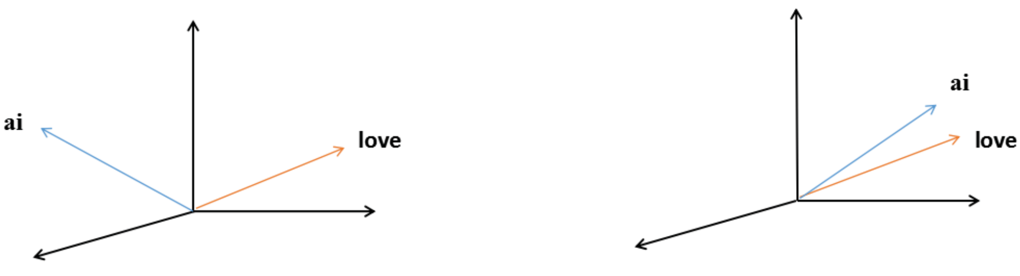


Fig. 2. Structure of GRU

III. METHODOLOGY

Our method will be expounded in details in this section. There are two modules in our method. Firstly, we use Word2Vec and LSTM to obtain the contextual representation of the sentence for both source and target language, then use the (Adversarial Auto Encoders) AAE to train the M (transfer matrix). Hence all languages can be mapped into a shared vector space. Check the Fig. 3 for the examples. Meanwhile, we calculate the cross-lingual word embeddings via average method. Secondly, the classifier are transferred from source language to the target languages. Below Fig. 4 is the framework diagram of our method.



a. Before transformation **b. After transformation**

Fig. 3. 3-D projection of Chinese and English vectors

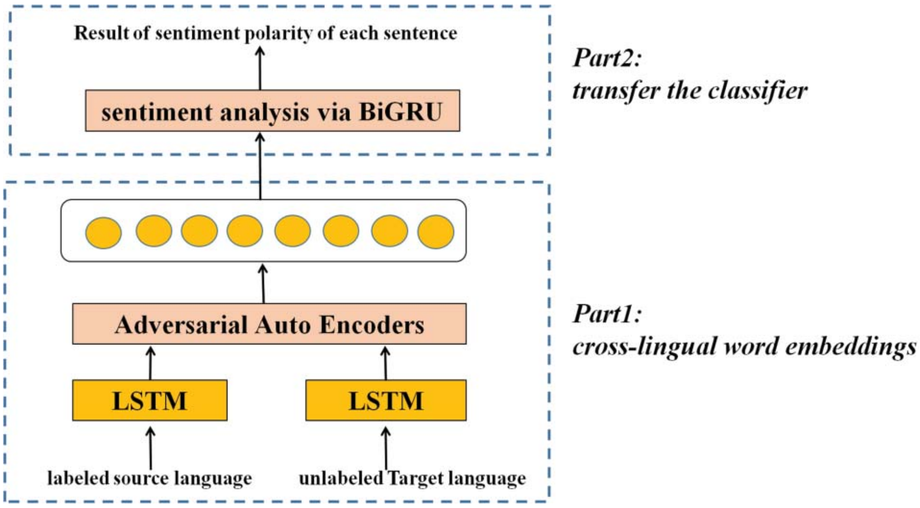


Fig. 4. Framework diagram of cross-lingual sentiment analysis

A. Contextual Cross-lingual word embeddings

The contextual embeddings are trained via Word2Vec and LSTM. After that we use transformation matrix ***M*** to transfer the embeddings of different languages into a shared vector space. ***M*** is obtained by Adversarial Auto Encoders, which were proposed by Makhzani etc [13]. After the transformation matrix ***M*** is got, we multiply the target word vector ***X*** with the transformation matrix ***M***, then acquire the transformed vector ***Z***.

$$Z = M \times X \tag{1}$$

After all the languages have been mapped into a single vector space, we calculate the cross-lingual embeddings via average method.

$$D = \frac{\sum_{i=1}^{N} W_i}{N} \tag{2}$$

B. Transfer sentiment classification model

Act as a peculiar kind of LSTM, Gated Recurrent Unit (GRU) networks have the ability to remember data for a long period of time. BiGRU (Bidirectional Gated Recurrent Unit) utilizes two GRU to capture each token of the sequence grounded on both the previous and the rear context token. We make use of BiGRU as the classification model.

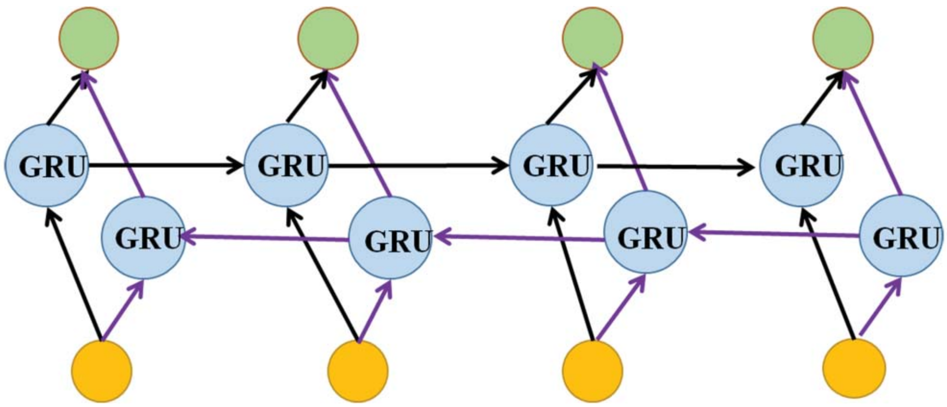


Fig. 5. Structure of BiGRU

As is demonstrated in Fig.5, one GRU controls the sequence from left to right, the other GRU do it from right to left. The current hidden state is formed via three parts, namely the current input X_t, the previous forward hidden state $\overrightarrow{H_{t-1}}$ and the backward hidden state $\overleftarrow{H_{t-1}}$. The rightward context representations denoted by $\overrightarrow{H_t}$ and leftward one denoted by $\overleftarrow{H_t}$, are linked into a long vector. The predictions of polarity are showed via the conjunctive output results.

Since target data is low in scale, we try to utilize some rich corpus as a countermeasure to settle the difficulty of resource scarcity. Therefore transfer learning is employed for this objective [14][15]. The sentiment classification model is trained on the corpus of source language, and then we transfer the model to target languages.

IV. Experiment And Results

A. Datasets

We use the cross-lingual sentiment dataset provided by Prettenhofer and Stein [16], which contains product comments on Amazon for the three product categories, namely books, DVD, and music in Chinese, English and German. The contents in the data are composed by product category and language. There are three harmonious disjoint datasets of training, test, and unlabeled documents for each pair of language category. The respective sizes are 27815, 2100, and 80,000.

English is set as the SL(source language), then Chinese and German are the TL (target language).

B. Basic Experiments

- **Based on Machine Translation and SVM (MT-SVM):** translate the target language into source language. After that, make use of SVM as the classifier.
- **Based on Machine Translation and BiGRU (MT-BiGRU):** translate the target language into source language. After that, make use of BiGRU as the classifier.
- **Based on transfer learning and SVM (TL-BiGRU):** get the normal word embeddings through deep learning, and then utilize BiGRU as the classifier.

C. Experimental setup

We train the contextual word embedding via Word2Vec and LSTM on both source and target languages. In order to map them into a shared vector space, we use Adversarial Auto Encoders (AAE) to train the transformation matrix ***M.*** After that, we use the average method to calculate the cross-lingual word embeddings. Following embedding layer is BiGRU model, which we train on the English data, and transfer it for Chinese and German data. We also train MT-SVM, MT-BiGRU and TL-SVM models as base experiments.

Experiments are trained on the deep learning frameworks of Keras and TensorFlow. Each kind of corpus is separated into two portions, which are training data and test data in accordance with 9:1. Cross entropy loss [17] are used as the loss function of model. Adam optimization method is the optimization method of model training [18]. The initial learning rate is set to 0.0005 and train 20 epoches. The evaluation of the model is the F1-Score, which is showed in table I.

TABLE I. F1- SCORE OF DIFFERENT METHODS. THE BEST RESULT IS SHOWN IN BOLDFACE.

METHOD	BOOKS	DVD	MUSIC	AVERAGE
MT-SVM	0.6799	0.6926	0.6614	0.6780
MT-BiGRU	0.6956	0.6895	0.6527	0.6793
TL-BiGRU	0.7325	0.7658	0.7524	0.7502
TL-AAE-BIGRU (Ours)	**0.7813**	**0.7826**	**0.7891**	**0.7857**

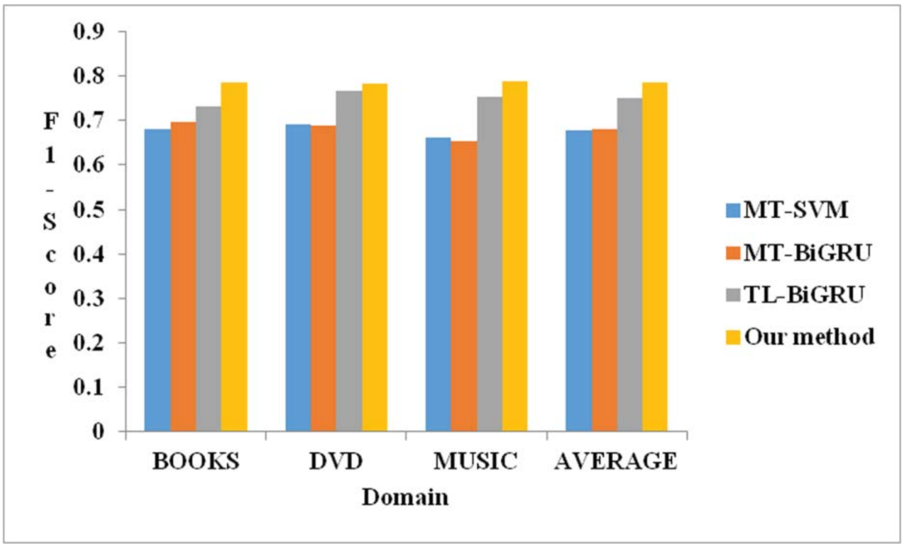


Fig. 6. F1-Score of different methods-Bar Chart

D. Results

According to the Fig. 6, we can learn that the performance improves over the machine translation method by using transfer learning strategy. Particularly getting the contextual cross-lingual embeddings could effectively promote the capability of transfer learning and the average F1-Score was added to 0.7813 from 0.7502. Training the contextual cross-lingual embeddings with the help of Adversarial Auto Encoders is the critical step. Hence our method acquires the best result compared with other base experiments. Meanwhile, our method has the strong robustness among different domain.

V. CONCLUSION AND PROSPECT

In this paper, we propose a patent model for cross-lingual sentiment analysis, which is a hot issue in nature language processing. Our method train the contextual cross-lingual embeddings based on Adversarial Auto Encoders. We find that contextual cross-lingual embeddings can improve the property when using transfer learning between different languages. High quality of word embeddings boosts the result of sentiment classification.

In future, on one hand, we'll focus on exploiting new methods for transfer learning and expanding the target data to more languages. On the other hand, we'll try to debug our models to conduct a range of sentiments because there are more emotions except negative, neural and positive.

REFERENCES

[1] B. Liu, "Sentiment analysis: Mining opinions, sentiments, and emotions," Cambridge University Press J. Clerk Maxwell, A Treatise on Electricity and Magnetism, 3rd ed., vol. 2. Oxford: Clarendon, 1892, pp.68–73, 2015.

[2] T. Munkhdalai and H. Yu, "Neural tree indexers for text understanding," Proc. EACL 2017.

[3] X. J. Wan, "Co-training for cross-lingual sentiment classification," 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, pp.235-243, 2009.

[4] T. Mikolov, I. Sutskever, K. Chen, G. Corrado, and J. Dean. "Distributed representations of words and phrases and their compositionality," CoRR, abs/1310.4546, 2013.

[5] M. Faruqui and C. Dyer, "Improving vector space word representations using multilingual correlation," In Proceedings of the 14th Conference of the European Chapter of the Association for Computational Linguistics, Association for Computational Linguistics, pp. 462–471, 2014.

[6] M. Zhang, Y. Liu, H. Luan, M. Sun, "Adversarial training for unsupervised bilingual lexicon induction," Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics. Stroudsburg: ACL 2017.

[7] G. Zhou, Z. Zhu, T. He, et al., "Cross-lingual sentiment classification with stacked auto encoders," Knowledge & Information Systems, vol.47, no.1, pp.27-44, 2016.

[8] T. Mikolov, K. Chen, G. Corrado, and J. Dean, "Efficient estimation of word representations in vector space," CoRR, abs/1301.3781, 2013.

[9] M. Artetxe, G. Labaka, and E. Agirre, "Learning principled bilingual mappings of word embeddings while preserving monolingual invariance," pp.2289–2294, 2016.

[10] M. Artetxe, G. Labaka, and E. Agirre, "Generalizing and improving bilingual word embedding mappings with a multi-step framework of linear transformations.," In Proceedings of the Thirty Second AAAI Conference on Artificial Intelligence, pp. 5012–5019, 2018.

[11] S. Hochreiter, & J. Schmidhuber, "Long short-term memory," Neural Computation, vol.9, no.8, pp.1735–1780, 1997.

[12] J. K. Kim, Y.B. Kim, R. Sarikaya, & E. Fosler-Lussier, "Cross-lingual transfer learning for POS tagging without cross-lingual resources.," In Proceedings of the 2017 conference on empirical methods in natural language processing, pp. 2832–2838, 2017.

[13] A. Makhzani, J. Shlens, N. Jaitly, et al. "Adversarial auto encoders," 2018.

[14] S. J. Pan, Q. Yang, et al., "A survey on transfer learning," IEEE Transactions on Knowledge and Data Engineering, vol.22, no.10, pp.1345–1359, 2010.

[15] Z. Yang, R. Salakhutdinov, & W.W. Cohen, "Transfer learning for sequence tagging with hierarchical recurrent networks," 5th international conference on learning representations, Toulon, France, April 24-26, 2017, conference track proceedings, ICLR, 2017.

[16] P. Prettenhofer, B. Stein, "Cross-lingual adaptation using structural correspondence learning," ACM Transactions on Intelligent Systems & Technology, vol.3, no.1, pp.1-13, 2011.

[17] J. Shore, R. A. Johnson, "Derivation of the Principle of Maximum Entropy and the Principle of Minimum Cross-entropy," IEEE Transactions on Information Theory, vol. 26, no.1, pp.26-37, 1980.

[18] D. Kingma, J. A. Ba, "A Method for Stochastic Optimization," Computer Science, vol.12, pp.69, 2014.