Combination of Multi-view Multi-source Language Classifiers for Cross-Lingual Sentiment Classification

Mohammad Sadegh Hajmohammadi, Roliana Ibrahim*,
Ali Selamat, and Alireza Yousefpour

Faculty of Computing, Universiti Teknologi Malaysia (UTM),
81310 Skudai, Johor, Malaysia
{shmohammad2,yalireza3}@live.utm.my, {roliana,aselamat}@utm.my

Abstract. Cross-lingual sentiment classification aims to conduct sentiment classification in a target language using labeled sentiment data in a source language. Most existing research works rely on machine translation to directly project information from one language to another. But cross-lingual classifiers always cannot learn all characteristics of target language data by using only translated data from one language. In this paper, we propose a new learning model that uses labeled sentiment data from more than one language to compensate some of the limitations of resource translation. In this model, we first create different views of sentiment data via machine translation, then train individual classifiers in every view and finally combine the classifiers for final decision. We have applied this model to the sentiment classification datasets in three different languages using different combination methods. The results show that the combination methods improve the performances obtained separately by each individual classifier.

Keywords: Cross-lingual, Sentiment classification, classifier combination, multi-view, multi-language.

1 Introduction

Together with the very rapid increasing of the internet access in the world, the volume of user generated contents have also been increased on the web. Due to the high quantity of user-generated contents, the task of summarizing their information into a useful format is a very hard and challenging problem. This challenge motivates the natural language processing (NLP) communities to design and develop computational methods to analyze these text documents.

Opinion mining or sentiment analysis is one of the most interesting fields in this area that analyzes people's opinions, attitudes and sentiments towards entities such as products, individuals, events, etc. [1]. Text document sentiment classification is the task of determining the sentiment polarity (e.g. positive or negative) of a given text document [2] and has received considerable attention due to its many useful application in product reviews classification [3] and opinion summarization [4].

* Corresponding Author.

N.T. Nguyen et al. (Eds.): ACIIDS 2014, Part I, LNAI 8397, pp. 21–30, 2014.

Up until now, different methods have been used in sentiment classification. These methods can be categorized into two main groups, namely; lexicon based and corpus based. The lexicon based methods classify text documents based on the polarity of words and phrases contained in the text [5, 6]. This group of methods needs a sentiment lexicon to distinguish between the positive and negative terms. In contrast, corpus based methods train a sentiment classifier based on labelled corpus using machine learning classification algorithms [7, 8]. The performance of these methods intensively depends on the quantity and the quality of labelled corpus as the training set.

Sentiment lexicons and annotated sentiment corpora are the most important resources for the sentiment classification. However, since most recent research studies in sentiment classification are in the English language, there are not enough labelled corpus and sentiment lexicons in other languages [9, 10]. Further, manually construction of reliable sentiment resources is a very hard and time-consuming task. Therefore, the challenge is how to utilize labelled sentiment resources in one language (i.e. English) for sentiment classification in another language and leads to an interesting research area called cross-lingual sentiment classification (CLSC). The most direct solution of this problem is the use of machine translation systems to directly project the information of data from one language into the other language[9-15]. The most existing research works develop a sentiment classifier based on the translated labelled data from the source language and use this classifier to determine the sentiment polarity of test data in the target language [12, 13]. Machine translation can be employed in the opposite direction by translating the test documents from the target language into the source language [9, 14, 15]. In this situation, the sentiment classifier is trained based on the original labelled data in the source language and then applied to the translated test data. A few number of research works used both direction of translation to create two different views of the training and the test data to compensate some of the translation limitations [10, 16]. But because the training set and the test set are from two different languages with different writing styles and from different cultures, these methods cannot reach the performance of monolingual sentiment classification methods in which the training and test samples are from the same language. The performance also can be influenced by the low quality of translation because machine translation is still far from satisfactory and therefore the translated text documents cannot cover all the vocabularies in the original text documents. Different term distribution between the original and the translated text documents is another important factor that can reduce the performance of CLSC. It means that a term may be frequently used in one language to express the opinion while the translation of that term is rarely used in the other language.

In this paper, we propose a new cross-lingual sentiment classification model that use more than one language (for example two languages) as the source languages to compensate some of the aforementioned limitations of resource translation in CLSC. We use labelled corpus from two different languages as the training data and also use both translation directions to create three different view of data, one in the target language and two in the source languages. Accordingly, three different classifiers are trained based on these three views and finally the predictions of these classifiers are combined using ensemble method. The proposed model was applied to the book review datasets in three different languages and experiments showed that using multiple source language in multiple views obtains better performance in comparison with the methods that use only one language as the source language.

The reminder of this paper is organized as follows: The next section presents related works on cross-lingual sentiment classification. Section 3 describes the proposed model. The experimental setup is explained in Section 4, while results and discussion are given in Section 5. Finally, Section 6 concludes this paper.

2 Related Work

Cross-lingual sentiment classification has been extensively studied in recent years. These research studies are based on the use of annotated data in the source language (always English) to compensate for the lack of labelled data in the target language. Most approaches focus on resource adaptation from one language to another language with few sentiment resources. For example Mihalcea et al. [17] generate subjectivity analysis resources into a new language from English sentiment resources by using a bilingual dictionary. In other works [13, 18], automatic machine translation engines were used to translate the English resources for subjectivity analysis. Banea et al. [18], showed that automatic machine translation is a viable alternative for the construction of resources for subjectivity analysis in a new language. Wan [19] used unsupervised sentiment polarity classification in Chinese product reviews. He translated Chinese reviews into different English reviews using a variety of machine translation engines and then performed sentiment analysis for both Chinese and English reviews using a lexicon-based technique. Finally, he used ensemble methods to combine the results of analysis. Another approach is that of cross-lingual classification, that is translating the features extracted from labelled documents [20, 21]. The features, selected by a feature selection algorithm, are translated into different languages. Subsequently, based on those translated features; a new model is trained for every language. This approach only needs a bilingual dictionary to translate the selected features. It can, however, suffer from the inaccuracies of dictionary translation, in that, words may have different meanings in different contexts. In another work, Wan [10] used the co-training method to overcome the problem of cross-lingual sentiment classification. In this paper, he exploited a bilingual co-training approach to leverage annotated English resources to sentiment classification in Chinese reviews. In this work, firstly, machine translation services were used to translate English labelled documents (training documents) into Chinese and similarly, Chinese unlabeled documents into English. The author used two different views (English and Chinese) in order to exploit the co-training approach into the classification problem. In an early work, Martin-Validivia et al. [9] proposed a meta classifier system that integrated the corpus based and lexicon based methods in order to create a sentiment classifier in Spanish language. In the first place, they used Spanish corpus along with its translated version in English and create two individual models based on these two corpora and applied machine learning to train the models. Next, they integrated SentiWordNet into the translated data to generate a new lexicon based model. Lastly, they combined these systems by Meta classifiers. To the best of our knowledge, using multiple source languages in multiple views has not yet been investigated in the field of cross-lingual sentiment classification.

3 Proposed Model

In this section we present a new cross-lingual model for sentiment classification that uses multiple languages as source language in multiple views. In this model, after the construction of different views in the target and the source languages, a classifier is trained based on the labeled data in every view and is applied to the test data in corresponding view and finally, the prediction results of each individual classifiers are combined to form the final results.

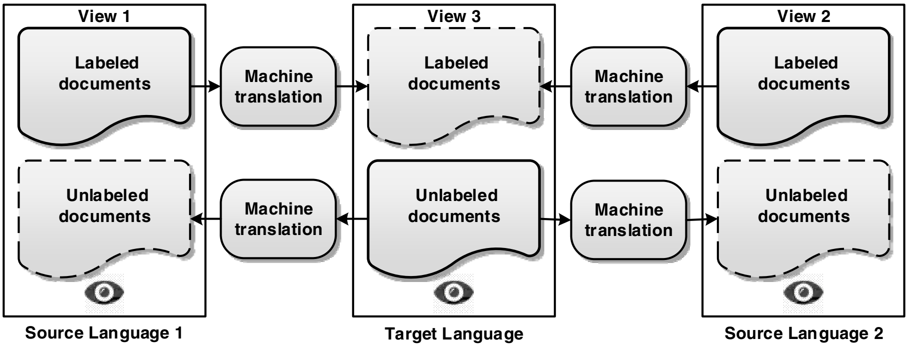

Fig. 1. Creation of multiple views of documents in the case of two source languages, using machine translation

3.1 Multi-view Data Creation

To create multiple views of labeled and unlabeled documents in the source and target languages, we perform machine translation in two different directions. At the first step, unlabeled document (test data) are translated from the target language into the source languages. Next, labeled documents (training data) are translated from the source languages into the target language. Fig. 1 diagrammatically shows the process of multi-view data creation for the situation that two source languages are used. As we can see in this figure, when two different languages are used as source language, we have training and test documents (labeled and unlabeled) in three different views. It means that each document is presented based on three different feature sets, one in the target language and two in the source languages. Therefore, classification process can be performed based on three different feature sets from three different languages.

3.2 Classification Combination in Different Views

Multiple classifiers combination is a well-known learning strategy when a set of classifiers is trained for a same classification. Combination of classifiers is the most reasonable solution when more than one single training set exist or different presentations of the training set are available [22]. Combining multiple classifiers can be advantageous since different classifiers would induce complimentary information for the classification.

In our proposed method, first, the training documents in every view are used to train a member classifier. Then, the trained classifiers are applied to the test set, represented based on the feature sets of corresponding views to determine the prediction label of each sample. After that, a combination rule is used to integrate the output predictions of the member classifiers to make the final classification decision. Several combination algorithms can be used to integrate the results of member classifiers. Definitely, we tried three groups of the most widely used methods, namely: majority voting, fixed rules and stacking, which are explained in the following subsections.

Majority Voting. Majority voting is the simplest method used for classifier combination. In this method, the final predicted class is selected by polling all the classifiers to see which class is the most popular. Whichever class that receives most votes is selected. Majority voting is always successful when the classifiers' output are binary.

Fixed Rules. Individual classifiers always provide not only the predicted label, but also one kind of confidence measure, such as posterior probability. The fixed rule combiner is used to combine these probabilities. Suppose that $p_c(w_j|x)$ is the posterior probability in predicting class w_j for instance x provided by member classifier c.

- Product rule integrates individual classifiers by multiplying the posterior possibilities and use the result for final output based on (1).where m is the number of the individual classifiers.

$$f_{Prod}(x) = \arg\max_{w_j} \prod_{c=1}^{m} p_c(w_j|x) \tag{1}$$

- Sum rule integrates individual classifiers by summing the posterior possibilities and use the results for final output based on (2).

$$f_{sum}(x) = \arg\max_{w_j} \sum_{c=1}^{m} p_c(w_j|x) \tag{2}$$

Stacking. The stacking combiner adds a new classifier (called combiner classifier) that uses the outputs of the member classifiers as input feature vector and learns the best method to integrate these classifiers. In this paper, we used the prediction confidence of every member classifier to form the input feature vector for the combiner classifier. At the first, the member classifiers are trained based on training set and then applied to a development set for prediction task. After that, the prediction confidences are used to form the input feature vector for the combiner classifier. Finally, the combiner classifier is trained based on these new training data to learn the best combination rule. We used four different machine learning algorithms to learn the combination rule: Support Vector Machine (SVM), Naïve Bayes (NB), Artificial Neural Network (ANN) and Linear Least Square (LLS).

4 Experiment

In this section, we evaluate our proposed approach in cross-lingual sentiment classification on three different languages in the book review.

4.1 Experimental Datasets

To create an evaluation dataset, we selected 2000 book reviews (1000 positive and 1000 negative) from Prettenhofer and Stein dataset [23] in three different languages: English, French and German. By combining these three languages, we obtained three different dataset for evaluation. In each dataset, one language is considered as target language and two other languages are the source languages. Documents in each language are translated into two other languages using Google translate service (http://translate.google.com/) to create different views of data. Table 1 shows the characteristics of evaluation datasets. In the pre-processing step, all English reviews are converted into lowercase. Special symbols and other unnecessary characters are eliminated from every review document.

Table 1. Different datasets used for evaluation

Data Set	Languages		
	Source 1	**Source 2**	**Target**
EF-G	English	French	Germany
EG-F	English	Germany	French
FG-E	French	Germany	English

To reduce computational complexity, we performed feature selection using the information gain (IG) technique. We selected 5000 high score unigrams and bi-grams as final features. Every document is represented by a feature vector. Each entry of a feature vector contains a feature weight. We used term presence as feature weights because this method has been confirmed as the most effective feature weighting method in sentiment classification [7, 24].

4.2 Experimental Setup

In all experiments, we used the support vector machine classifier (SVM) as the member classifiers in every view. SVM^{light} (http://svmlight.joachims.org/) is used as the SVM classifier in the experiments with all parameters set to their default values.

However, SVM^{light} does not directly output the posterior probabilities of predicted labels. Therefore we use a strategy that introduced in [25] to compute these probabilities. Labeled documents in every view are randomly divided into training and development sets with the portion of 80% and 20% respectively. The development set is used to train the classifier combiner in stacking methods. For the classifier combiner in stacking methods we used the original MATLAB implementation of machine learning algorithms. For the ANN we used one hidden layer with 20 neurons. Other algorithms were used with default parameters.

5 Results and Discussion

We conducted several experiments with different combination methods in three different languages. In this section, we compare the accuracy of each combination technique with other techniques and also with the accuracy of the member classifiers as base classifiers. The main goal of the classifier combination is to correct the errors of the member classifiers. In our experiment, we can approve the improvement achieved by using this approach since all the combination methods improve the final classification results in compare with individual classifiers.

Table 2. Accuracy of the member classifiers and the combining methods

Data Sets			EF-G	EG-F	FG-E	Average
Member classifiers		View 1	77.57%	77.10%	74.99%	
		View 2	76.41%	74.85%	74.99%	
		View 3	78.62%	78.40%	80.64%	
Combination Methods	Majority Voting		80.07%	80.45%	80.54%	80.35%
	Fixed Rule	Product	80.82%	80.35%	81.19%	80.79%
		sum	80.47%	80.25%	80.44%	80.39%
	Stacking	SVM	**81.27%**	80.45%	81.29%	81.00%
		NB	80.87%	80.15%	**81.39%**	80.80%
		ANN	**81.27%**	**80.65%**	81.19%	**81.04%**
		LLS	80.97%	79.70%	80.89%	80.52%

As we can see in Table 2, all the classifiers combination methods outperform the member classifiers. This means that the information of multiple source languages can cover each other in predicting the sentiment labels of the target language documents. This table also shows that the accuracy of target language view (View 3) is grater then two other views. Because in this view, features are extracted from the training data that translated from two source languages and therefore cover more vocabularies from target language documents so documents in target language are presented much better with this feature set. Figure 1 also shows the comparison results in graphical format. In this figure, we can also see that stacking methods shows better performance for classifiers combination in comparison with other combination. This is due to the fact that stacking method tries to learn the best combination rule through machine learning.

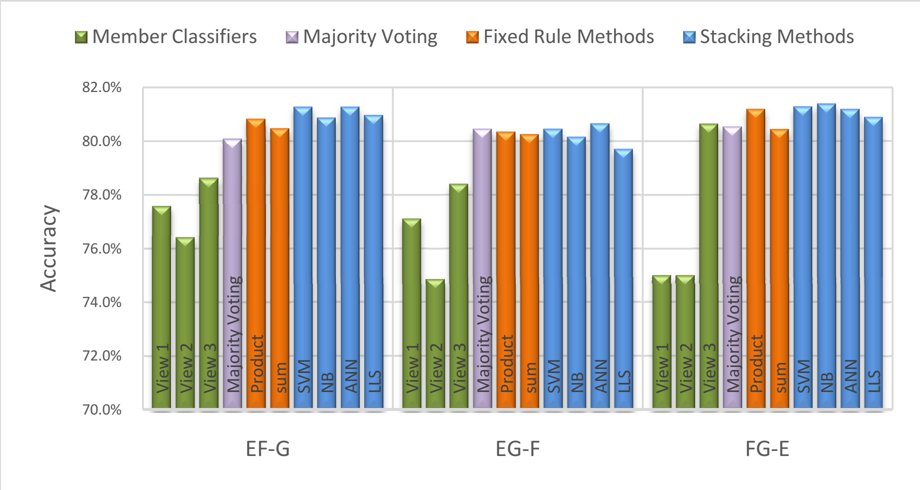

Fig. 2. Comparison results obtained by different combination methods

6 Conclusion and Further Work

In this paper, we have proposed a classifier combination model for using multiple source languages in different views to improve the performance of cross-lingual sentiment classification. In the proposed model, automatic machine translation was used to project the information of target language documents into the source languages and also to translate the training data from the source language into the target language. This bi-direction translation creates different views of classification data and sentiment classification can be performed in every view. Finally, the member classifiers were integrated using different aggregation methods such as fixed rules, majority voting and stacking algorithm. We applied this model to cross-lingual sentiment classification datasets in three different languages and we have shown that the combination methods improve the performances obtained separately by every single classifier. These results shows that using the information of multiple source languages can cover more characteristics of target language in sentiment classification and therefore can improve the performance of sentiment classification in the target language.

In addition, we would like to exploit unlabeled data from target language in this multi-view framework and use semi-supervised multi-view learning algorithms to improve the performance of cross-lingual sentiment classification.

Acknowledgement. This work is supported by the Ministry of Higher Education (MOHE) and Research Management Centre (RMC) at the Universiti Teknologi Malaysia (UTM) under Research University Grant Scheme (Vote No. Q.J130000.2628.07J52)

References

1. Liu, B.: Sentiment Analysis and Opinion Mining. Morgan & Claypool Publishers (2012)
2. Hajmohammadi, M.S., Ibrahim, R., Ali Othman, Z.: Opinion Mining and Sentiment Analysis: A Survey. International Journal of Computers & Technology 2(3), 171–178 (2012)
3. Zhou, S., Chen, Q., Wang, X.: Active deep learning method for semi-supervised sentiment classification. Neurocomputing 120, 536–546 (2013)
4. Ku, L.W., Liang, Y.T., Chen, H.H.: Opinion extraction, summarization and tracking in news and blog corpora. In: Proceedings of AAAI-2006 Spring Symposium on Computational Approaches to Analyzing Weblogs (2006)
5. Taboada, M., Brooke, J., Tofiloski, M., Voll, K., Stede, M.: Lexicon-based methods for sentiment analysis. Comput. Linguist. 37(2), 267–307 (2011)
6. Turney, P.D.: Thumbs up or thumbs down?: Semantic orientation applied to unsupervised classification of reviews. In: Proceedings of the 40th Annual Meeting on Association for Computational Linguistics, pp. 417–424. Association for Computational Linguistics, Philadelphia (2002)
7. Pang, B., Lee, L., Vaithyanathan, S.: Thumbs up?: Sentiment classification using machine learning techniques. In: Proceedings of the ACL 2002 Conference on Empirical Methods in Natural Language Processing, vol. 10, pp. 79–86. Association for Computational Linguistics (2002)
8. Moraes, R., Valiati, J.F., Gavião Neto, W.P.: Document-level sentiment classification: An empirical comparison between SVM and ANN. Expert Systems with Applications 40(2), 621–633 (2013)
9. Martín-Valdivia, M.-T., Martínez-Cámara, E., Perea-Ortega, J.-M., Ureña-López, L.A.: Sentiment polarity detection in Spanish reviews combining supervised and unsupervised approaches. Expert Systems with Applications 40(10), 3934–3942 (2013)
10. Wan, X.: Bilingual co-training for sentiment classification of Chinese product reviews. Comput. Linguist. 37(3), 587–616 (2011)
11. Wan, X.: 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, pp. 235–243. Association for Computational Linguistics, Suntec (2009)
12. Balahur, A., Turchi, M.: Comparative experiments using supervised learning and machine translation for multilingual sentiment analysis. Computer Speech & Language 28(1), 56–75 (2014)
13. Banea, C., Mihalcea, R., Wiebe, J.: Multilingual subjectivity: are more languages better? In: Proceedings of the 23rd International Conference on Computational Linguistics, pp. 28–36. Association for Computational Linguistics, Beijing (2010)
14. Prettenhofer, P., Stein, B.: Cross-Lingual Adaptation Using Structural Correspondence Learning. ACM Trans. Intell. Syst. Technol. 3(1), 1–22 (2011)
15. Hajmohammadi, M.S., Ibrahim, R., Selamat, A.: Density Based Active Self-training for Cross-Lingual Sentiment Classification. In: Jeong, H.Y., Yen, N.Y., Park, J.J. (eds.) Advanced in Computer Science and its Applications. LNEE, vol. 279, pp. 1053–1059. Springer, Heidelberg (2014)
16. Pan, J., Xue, G.-R., Yu, Y., Wang, Y.: Cross-Lingual Sentiment Classification via Bi-view Non-negative Matrix Tri-Factorization. In: Huang, J.Z., Cao, L., Srivastava, J. (eds.) PAKDD 2011, Part I. LNCS (LNAI), vol. 6634, pp. 289–300. Springer, Heidelberg (2011)

17. Mihalcea, R., Banea, C., Wiebe, J.: Learning multilingual subjective language via cross-lingual projections. In: Proceedings of the 45th Annual Meeting of the Association of Computational Linguistics, pp. 976–983 (2007)
18. Banea, C., Mihalcea, R., Wiebe, J., Hassan, S.: Multilingual subjectivity analysis using machine translation. In: Proceedings of the Conference on Empirical Methods in Natural Language Processing, pp. 127–135. Association for Computational Linguistics, Honolulu (2008)
19. Wan, X.: Using bilingual knowledge and ensemble techniques for unsupervised Chinese sentiment analysis. In: Proceedings of the Conference on Empirical Methods in Natural Language Processing, pp. 553–561. Association for Computational Linguistics, Honolulu (2008)
20. Moh, T.-S., Zhang, Z.: Cross-lingual text classification with model translation and document translation. In: Proceedings of the 50th Annual Southeast Regional Conference, pp. 71–76. ACM, Tuscaloosa (2012)
21. Shi, L., Mihalcea, R., Tian, M.: Cross language text classification by model translation and semi-supervised learning. In: Proceedings of the 2010 Conference on Empirical Methods in Natural Language Processing, pp. 1057–1067. Cambridge, Massachusetts (2010)
22. Jain, A.K., Duin, R.P.W., Jianchang, M.: Statistical pattern recognition: A review. IEEE Transactions on Pattern Analysis and Machine Intelligence 22(1), 4–37 (2000)
23. Prettenhofer, P., Stein, B.: Cross-language text classification using structural correspondence learning. In: Proceedings of the 48th Annual Meeting of the Association for Computational Linguistics, pp. 1118–1127. Association for Computational Linguistics, Uppsala (2010)
24. Xia, R., Zong, C., Li, S.: Ensemble of feature sets and classification algorithms for sentiment classification. Information Sciences 181(6), 1138–1152 (2011)
25. Brefeld, U., Scheffer, T.: Co-EM support vector learning. In: Proceedings of the Twenty-First International Conference on Machine Learning, p. 16. ACM, Banff (2004)