

Cross-Lingual Non-Ferrous Metals Related News Recognition Method Based on CNN with A Limited Bi-Lingual Dictionary

Xudong Hong¹, Xiao Zheng^{1,*}, Jinyuan Xia¹, Linna Wei¹ and Wei Xue¹

Abstract: To acquire non-ferrous metals related news from different countries' internet, we proposed a cross-lingual non-ferrous metals related news recognition method based on CNN with a limited bilingual dictionary. Firstly, considering the lack of related language resources of non-ferrous metals, we use a limited bilingual dictionary and CCA to learn cross-lingual word vector and to represent news in different languages uniformly. Then, to improve the effect of recognition, we use a variant of the CNN to learn recognition features and construct the recognition model. The experimental results show that our proposed method acquires better results.

Keywords: Non-ferrous metal, CNN, cross-lingual, text classification, word vector.

1 Introduction

With the increasing internationalization of non-ferrous metals industry, more and more enterprises pay attention to other countries' marketing environment, industrial development, policies, and regulations, etc. which, if could be timely acquired, can effectively support enterprises' business decision. A great deal of such information could be easily found in online news on different countries' websites, providing us with a convenient way to acquire them. However, considering the various topics of the news and different languages they are using, how to accurately recognize non-ferrous metals related news on different countries' Internet is a key issue to be solved.

The recognition of news related to non-ferrous metals is essentially a text classification problem, i.e. to determine whether a piece of news is non-ferrous metal related or not. Due to the news on different countries' Internet are in different languages, if monolingual text classification method is adopted directly, the training corpus and classification model need to be constructed for different languages respectively, which will consume a large amount of manual effort and lingual resources. We would like to adopt a method that uses only one classifier and monolingual training corpus, to complete the classification of news in different languages. That is cross-lingual text classification method.

There are many studies on cross-lingual text classification. Most of them focus on how to construct a cross-lingual feature space to uniformly represent news in different languages. Methods proposed include the use of machine translation, bilingual dictionaries, multi-lingual knowledge base, parallel corpus, comparable corpus and other language resources.

¹ Anhui University of Technology, Maanshan, 243002, China.

* Corresponding Author: Xiao Zheng. Email: xzheng@ahut.edu.cn.

And on this base, the monolingual text classification method will be used. However, based on the fact that machine translation are not available for many languages yet, non-ferrous metals domain is short of language resources, and currently existing methods classify texts on cross-lingual feature space directly, which do not take into account the unique features of non-ferrous metals related news, there is no doubt that these methods will not work well for cross-lingual non-ferrous metals related news recognition.

This paper, taking the news of Chinese and Vietnamese as the research object, proposed a cross-lingual non-ferrous metals related news recognition method. Firstly, a limited bilingual dictionary and CCA (canonical correlation analysis) were selected to conduct cross-lingual word vectors training for all different languages words. Then the acquired cross-lingual word vectors were adopted to uniformly represent news in different languages and a variant CNN was modified to learn the unique features of non-ferrous metals related news, to construct the recognition model and to recognize the emerging news of different languages.

2 Related work

The direct way to achieve cross-lingual text classification is to use machine translation tools to translate different language texts into the same language and then to classify. However, machine translation tools are not yet available to some languages and further research on machine translation in specific domains need to be carried out. Most current methods firstly use cross-lingual resources, including bilingual dictionaries, multi-language knowledge bases, parallel corpora, comparable corpora, etc. to construct a cross-lingual feature space and uniformly represent different language texts. Secondly, they use traditional monolingual text classification methods to classify, e.g., K-nearest Neighbor [Shin, Abraham and Han (2006)], Naive Bayes [Kim, Han, Rim et al. (2006)], Support Vector Machines [Martens, Huysmans, Setiono et al. (2008)] and so on. The main difference between different methods is the construction of cross-lingual feature space.

To construct a cross-lingual feature space, Rigutini et al. [Rigutini, Maggini and Liu (2005)] uses bilingual dictionaries to translate words from one language to another and the translated bilingual word pairs as the cross-lingual features. Gliozzo et al. [Gliozzo (2006); Balamurali (2012)] uses the multilingual concepts of WordNet as the cross-lingual features. Due to the ambiguity of the words and the limited coverage of dictionaries or knowledge bases, the effect of these methods is poor. Littman et al. [Littman, Dumais and Landaue (1998)] represents the parallel corpus as a matrix of document-bilingual words, conduct LSI on the matrix to acquire the cross-lingual representations of the different language documents. Xiao et al. [Xiao and Guo (2014)] puts parallel corpus, labeled documents of source language, and documents to be classified of target language together into the matrix of document-bilingual words: NMF is first used to fill the matrix, then conduct LSI on the matrix in the same way as mentioned. Li et al. [Li and Shawe-Taylor (2006)] uses KCCA to analyze the correlation between words of different languages on parallel corpus and find a language-independent cross-lingual feature space to represent different language documents. These methods perform better. However, parallel corpus requires strict mutual translation between different languages documents. The construction of parallel corpus is effort costly.

Recently many researchers use comparable corpus to construct cross-lingual feature space. Since comparable corpus just requires different languages documents describe the same entities or events, the construction of that is relatively easier. Ni et al. [Ni, Sun, Hu et al. (2011)] uses the different languages documents that explain the same concept in Wikipedia as the comparable corpus and uses the bi-lingual topic model to do the bi-lingual topic mining. The bi-lingual topics acquired will be used as cross-lingual features. To achieve the same effect, the methods based on comparable corpus require a larger corpus than the methods based on parallel corpus.

As lacking language resources related to non-ferrous metals, these methods are difficult to be directly applied to cross-lingual non-ferrous metals related news recognition. In addition, these methods do not take into account the unique features of the non-ferrous metal related news, and it is difficult to achieve good results. In recent years, neural networks, due to its capacity of features learning, have achieved good results on vector representation of words' semantic [Yoshua (2003); Mikolov (2013)] and document classification [Collobert (2011); Kim (2014)]. In this paper, we will study how to use fewer languages resources to construct cross-lingual word vector to overcome languages barrier, and how to use CNN, neural networks, to recognize non-ferrous metals related news.

3 Cross-lingual word vector training

Cross-lingual word vector represents different languages words' semantics in the same space, measures the semantic similarity between different language words, and is the foundation of our method. Considering non-ferrous metals domain lack of bilingual resources, firstly we train monolingual word vector for different languages words respectively; then use a limited bi-lingual dictionary and CCA to conduct cross-lingual word vector training.

Table 1: Bi-lingual lexicon of non-ferrous metals domain

Type	Vietnamese	Chinese
Organization	Tập đoàn Đồng Lãng	铜陵有色金属集团
Organization	Tập đoàn Công nghiệp than – khoáng sản VN	越南煤炭与矿产工业集团
Person	Ông Vương Hồng Tiền	王宏前
Person	ông Nguyễn Văn Biên	阮文编
Product	máy cấp liệu mâm tròn	圆盘给料机
Term	tuyển từ	磁选

We train the monolingual word vectors respectively, according to whether the words belong to the non-ferrous metals domain or not. For the words that do not belong to the non-ferrous metals domain, different languages news crawled from the different countries Internet are used as training corpus; NLPiR² and JvnTextPro³ are used for Chinese and Vietnamese Word Segmentation respectively; CBOW of word2vec⁴ is used for word

² NLPiR: <http://ictclas.nlp.ir.org>.

³ JvnTextPro: <http://jvntextpro.sourceforge.net/>.

⁴ word2vec: <https://code.google.com/archive/p/word2vec/>.

vectors training. For the words that belong to the non-ferrous metals domain, considering they seldom appear in this news, a pre-collected bi-lingual lexicon of non-ferrous metals domain is used, as shown in Tab. 1. We search the different language words of the lexicon in Google. The recalled sentences containing those words are used as the training corpus. The words of non-ferrous metals domain in the sentences are recognized by means of Maximum Matching. In the word vectors training of these words, the word vectors of the words that are not non-ferrous metals related are fixed to the previous training results.

Let $CW \in R^{n_1 \times d_1}$ and $VW \in R^{n_2 \times d_2}$ be the acquired n_1 and n_2 monolingual word vectors of Chinese and Vietnamese words. d_1 and d_2 are the vectors' dimension. $M \in R^{d_1 \times d}$ and $N \in R^{d_2 \times d}$ are two mapping matrices, can map different languages words vectors to a new feature space of d dimension, $d = \min\{\text{rank}(CW^*), \text{rank}(VW^*)\}$. Thus we can obtain $CW' \in R^{n_1 \times d}$ and $VW' \in R^{n_2 \times d}$, by Eq. (1). CW' and VW' are the new representation of different languages words in the d -dimensional language-independent feature space.

$$CW' = CW * M, \quad VW' = VW * N \quad (1)$$

Let $CW^* \subseteq CW$, $VW^* \subseteq VW$. CW^*UVW^* be a limited dictionary which contain n Chinese-Vietnamese words transliteration pairs, $CW^* \in R^{n \times d_1}$, $VW^* \in R^{n \times d_1}$. The Chinese and Vietnamese words in them are one-to-one correspondence. Let x and y be the monolingual word vectors of a translation pair in CW^* and VW^* , thus $x' = xM$, $y' = yN$ will be the new representation of the words in the new space.

The semantics of translated words should be similar. To make the acquired words vectors in the new space be computable, the correlation between word vectors of translated words in the new space should be higher. Based on that, we use CCA to maximize the correlation $p(x', y')$ for all x' and y' to get M and N . $p(x', y')$ is the Pearson correlation coefficient between x' and y' .

$$p(x', y') = \frac{E(x'y')}{\sqrt{E(x'^2)E(y'^2)}} \quad (2)$$

$E(x'y')$ is expectations of the inner product of x' and y' . The principle of CCA can be expressed as Eq. (3). We used a Matlab toolkit to get the solution procedure of CCA⁵.

$$M, N = \text{CCA}(x, y) = \underset{M, N}{\text{arg max}} p(xM, yN) \quad (3)$$

After obtaining M and N , CW' and VW' can be obtained by Eq. (1). The feature space acquired will be the cross lingual feature space. CW' and VW' will be the cross lingual word vectors of the Chinese and Vietnamese words.

4 Recognition model construction

We construct the recognition model based on a variant CNN 6 model. The model architecture shown in Fig. 1, has 4 layers: word vectors layer, convolutional layer, pooling layer, fully connected layer.

⁵ CCA: <http://www.mathworks.com/help/stats/canocorr.html>.

⁶ CNN: <https://github.com/fchollet/keras.git>.

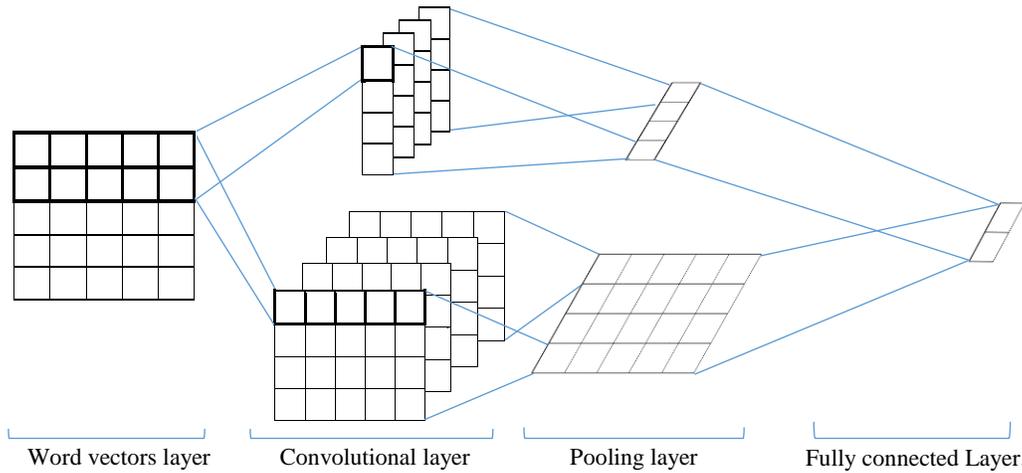


Figure 1: Recognition model based on CNN

Word vectors layer represents different languages news by cross lingual word vectors of pre-training. Let $x_i \in \mathbb{R}^d$ be the d -dimensional cross-lingual word vector of the i -th word of the news. Thus the news of length s is represented as:

$$c = \{x_1 \oplus x_2 \oplus \dots \oplus x_s\} \tag{4}$$

where \oplus is the concatenation operator. In general, let $x_{i:j}$ refer to the concatenation of words $x_i, x_{i+1}, \dots, x_{i+j}$.

The convolutional layer contains several convolution kernels, A convolution kernel is applied to a window of h words $x_{i:i+h-1}$, to produce a new feature. A convolution operation can be represented as:

$$c_i = f(W * x_{i:i+h-1} + b) \tag{5}$$

In our work, $f(*)$ is sigmoid function, $W \in \mathbb{R}^{h \times d}$. We use 8 convolution kernels, for different convolution kernel, the values of W , b , h and the calculation of $W * x_{i:i+h-1}$ are different. No. 1 to 4 convolution kernels, as usual, calculate $W * x_{i:i+h-1}$ as:

$$W * x_{i:i+h-1} = \sum_{n=1}^d \sum_{m=0}^{h-1} w_{m,n} x_{i+m,n} \tag{6}$$

$b \in \mathbb{R}$, $c_i \in \mathbb{R}$, and the values of h are 2, 3, 4, 5 respectively. 5 to 8 convolution kernels are newly added by us, use a new way to calculate $W * x_{i:i+h-1}$. The calculation is shown as:

$$W * x_{i:i+h-1} = \sum_{m=0}^{h-1} \begin{bmatrix} w_{m 1} x_{i+m 1} \\ w_{m 2} x_{i+m 2} \\ \vdots \\ w_{m d} x_{i+m d} \end{bmatrix} \tag{7}$$

$b \in \mathbb{R}^d$, $c_i \in \mathbb{R}^d$, the value of h is 2, 3, 4, 5 respectively. These 8 convolution kernels are applied to each possible window of words in the news $\{x_{i:i+h-1} | 1 < i < s - h + 1\}$ to produce 8 feature maps. Each feature map is $C = \{c_i\}_{1 \times (s-h+1)}$. The features map acquired from No. 1 to 4 is consist of coded h -gram features of the news. The features maps acquired by No. 5-8 can be seen as the 4 different sematic representations of the

news' d topics. The W , b of 8 convolution kernels are $32 \times d + 4$ parameters to be solved.

In pooling layer, we apply k-max pooling operation over the feature maps to choose features and pass the features acquired to the fully connected layer. For the feature map acquired by No. 1-4 convolution kernel, the k-max pooling operation can be represented as:

$$k\text{-max}(c_1, c_2, \dots, c_{s-h+1}) \quad (8)$$

The result can be seen as the k most meaningful n-gram features of the new. For the feature map acquired by No. 1-4 convolution kernel, the operation can be represented as:

$$k\text{-max}(C) = \begin{bmatrix} k\text{-max}(c_{11}, c_{21}, \dots, c_{s-h+1\ 1}) \\ k\text{-max}(c_{12}, c_{22}, \dots, c_{s-h+1\ 2}) \\ \vdots \\ k\text{-max}(c_{1d}, c_{2d}, \dots, c_{s-h+1\ d}) \end{bmatrix} \quad (9)$$

The result is a matrix of size $d \times k$. The i-th row of the matrix can be seen as the most meaningful semantic representation of i-th topic.

Fully connected layer, whose output is the probability of the news related to Non-ferrous Metals, whose input is the $(8 \times k + 1) \times d$ features acquired from the pooling layer. Let φ be the concatenation of input features, θ be the weights of $(8 \times k + 1) \times d$ connections. The probability calculation is shown as:

$$y = h_{\theta}(\varphi) = \frac{1}{1 + e^{-\theta^T \varphi}} \quad (10)$$

If $y > 0.5$ the news is Non-ferrous Metals related. θ are the parameters to be solved.

5 Model learning

We use a supervised approach to learn the model. Let $\{y^{(1)}, y^{(2)}, \dots, y^{(n)}\}$ be the labels of n news for training. If the i-th news is Non-ferrous Metals related then $y^{(i)} = 1$ else $y^{(i)} = 0$. The loss function of the recognition model is shown as Eq. (11). $\{y_1, y_2, \dots, y_n\}$ are the outputs of the recognition model on the n news.

$$-\frac{1}{m} \log \left(\sum_{i=1}^m y_i y^{(i)} (1 - y_i)^{1 - y^{(i)}} \right) \quad (11)$$

We minimize the loss function to get the values of $\{W, b, \theta\}$. Different values of $\{W, b, \theta\}$ represent different encoding methods of features.

For the solution we use mini-batch gradient descent algorithm. For regularization, we employ dropout on the pooling layer with a constraint on l_2 -norms of the weight vectors [Hinton, Srivastava and Krizhevsky (2012)]. Dropout prevents co-adaptation of hidden units by randomly dropping out a proportion p of the chosen features during forward and back propagation. At training time, during forward propagation, dropout uses $\theta^T (\varphi \circ r)$ to replace the $\theta^T \varphi$ in Eq. (10), where \circ is the element-wise multiplication operator and r is a vector of Bernoulli random variables with probability p of being 1. r has the same length with φ . During forward Gradients are back propagated only through the units of being 1. At test time, dropout uses p to scale the learned weights θ^T , i.e. use $p\theta^T \varphi$ to replace the $\theta^T \varphi$. We additionally constrain l_2 -norms of the weights θ^T by rescaling θ^T to

have $\|\theta^T\|_2 = 0$ whenever $\|\theta^T\|_2 > 0$ after a gradient descent step.

6 Experiments

6.1 Data and evaluations

Since there is no available corpus for cross-lingual non-ferrous metal related news recognition currently, to the best of our knowledge, we construct a corpus by ourselves. We crawled a number of non-ferrous metal related news from “smm”, “cnmn”, etc. Chinese websites and “Vinanet”, etc. Vietnamese websites, as positive data, And we crawled a number of news from the political, military, etc. non-ferrous metal unrelated channels on some comprehensive news website, as negative data, at the same times. The number of training and test data is shown in Tab. 2.

Table 2: Data distribution of corpus

Data and its language	Positive	Negative
Training data (Chinese)	3000	6000
Test data (Chinese)	800	1600
Test data (Vietnamese)	600	1200

We use Precision, Recall, and F-measure to evaluate the effect of our method, which are calculated as follow.

$$\text{Precision} = \frac{TP}{TP + FP} \times 100\%$$

$$\text{Recall} = \frac{TP}{P} \times 100\% \quad (12)$$

$$\text{F-measure} = \frac{2 * \text{Precision} * \text{Recall}}{\text{Precision} + \text{Recall}}$$

TP is the number of true positive of recognition results. FP is the number of false positive of it P is the number of positive of test data.

4.2 Results and analysis

Experiment 1: Compare with different methods of cross-lingual news classification.

In this paper, we implement another two cross-lingual text classification methods presented in Gliozzo et al. [Gliozzo and Strapparava (2006)] and [Ni, Sun, Hu et al. (2011)]. The former bases on the bi-lingual dictionary. The latter bases on the comparable corpus. These two methods have the best performance in the methods that use the same bilingual resources.

For the methods presented in Ni et al. [Ni, Sun, Hu et al. (2011)], we collected 3000 Chinese-Vietnamese document-pairs from Wikipedia by a web crawler. Each pair is a different language description of the same event, person, etc. We use these document-pairs as comparable corpus and apply bi-lingual topic model on it. According to Ni et al. [Ni, Sun, Hu et al. (2011)], we use topics number of 400, hyper-parameters α of 0.5/400 and β of 0.1, use SVM as a classifier.

For the method presented in Gliozzo et al. [Gliozzo and Strapparava (2006)], since

MultiWordNet does not contain Vietnamese and contains very few non-ferrous metal related words, we experiment with the bi-lingual lexicon that we pre-constructed as shown in Tab. 1. The classifier is SVM too.

For our method, according to Hinton et al. [Hinton, Srivastava and Krizhevsky (2012)], we set dimension of cross-lingual word vectors $d = 400$. The k of pooling layer is set to 10 after several trials. The performance is the best like that

We train all these 3 methods on the same training data in Chinese and test them on the test data in Chinese and Vietnamese respectively and compare the results of 3 methods. The comparison results of these methods in monolingual environment are shown in Fig. 2(a). The results in cross-lingual environment are shown in Fig. 2(b). In the figure “A” is the results of our method, “B” is the results of Ni et al. [Ni, Sun, Hu et al. (2011)], “C” is the results of Gliozzo et al. [Gliozzo and Strapparava (2006)].

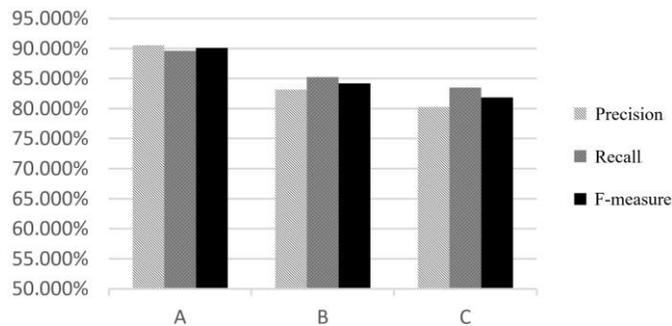


Figure 2 (a): Comparison results in monolingual environment

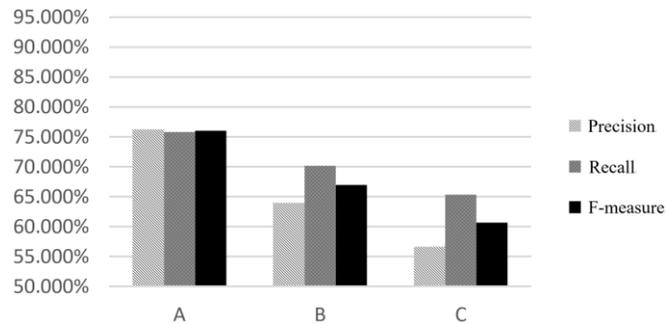


Figure 2 (b): Comparison results in cross-lingual environment

Experiment results show that our method has significantly improved the effect in both monolingual and cross-lingual environments compared with the other two methods. And in cross-language environments, our method improves the effect more obvious. It illustrates that using CNN to construct recognition model can improve the effect of recognition. Cross-lingual word vectors can better represent different language news and overcome languages barrier, so that improve the effect in cross-lingual environment.

Experiment 2: Compare the experiment results of whether non-ferrous metals related words are recognized

Before the training of cross-lingual word vector and the representation of news in different language, we both need to word segment. In our work, the words of non-ferrous metals related in the sentences are recognized in the segmentation by means of Maximum Matching. In order to explore the effect of that, we conduct this experiment on the whole test data include Chinese and Vietnamese test data. The results are shown in Fig. 3. “A” are the results of non-ferrous metals related words are recognized, “B” are the results of those are not recognized. Recognition or not will result in different representations of the news. For example, if the Chinese word “有色金属” (Non-ferrous metal) in a news have not been recognized, it will be divided as “有/色/金属” (have/colors/metal), then the meaning of the word and the representations of the news will be changed. That will affect the recognition result.

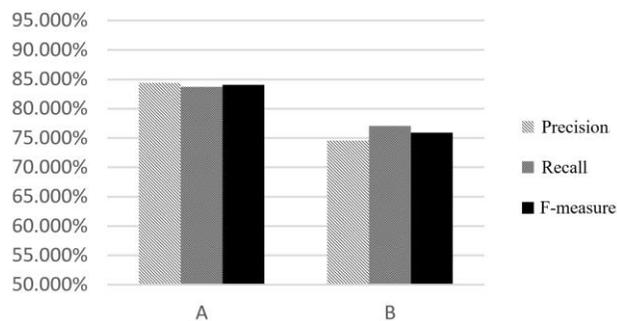


Figure 3: Results of whether non-ferrous metals related words are identified or not

Experiment results show that after non-ferrous metals related words are recognized the effect of recognition is improved obviously. It illustrates that recognizing the non-ferrous metals related words of the news is necessary for non-ferrous metals related news recognition. At the same time, it also illustrates that the non-ferrous metals related words of a piece of news are important to determine whether the news is non-ferrous metals related.

Experiment 3: Explorer the effect of the newly added convolution kernels

In the convolution layer of the CNN model, the No. 5-8 convolution kernels are newly added by us. They use a new calculation way. In order to explore the effect of the newly added kernels, we conduct this experiment on the whole test data. The results are shown in Fig. 4. “A” are the results of the model that uses the new kernels. “B” are the results of the model that doesn’t use.

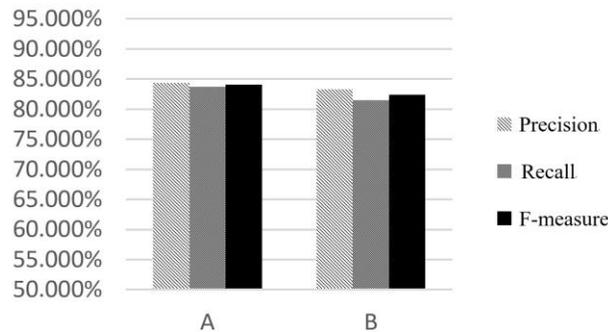


Figure 4: Results of the model that uses newly added kernels or not

Experiment results show that the model that uses newly added kernels performs better. It illustrates that the convolution kernels newly added are useful for improving the recognition effect.

7 Conclusion

To timely and accurately acquire non-ferrous metals related news from different countries' internet. In this paper, we proposed cross-lingual non-ferrous metals related news recognition method based on CNN with a limited dictionary. We use CCA and a limited dictionary to train cross-lingual word vectors; use the acquired cross-lingual word vectors represent news in different languages, overcome languages barrier; use CNN to learn the features of non-ferrous metals related news and construct recognition model, add new convolution kernels to the CNN based model. Experiments results show these can improve the cross-lingual recognition results obviously. Considering that sentences or words in different places of a piece of news have different ability to represent news semantics and in a cross-lingual situation, a word in one language can be translated into multiple words in another language. All these factors can affect the recognition result. In our future study, we will take these factors into consideration.

Acknowledgement: The Major Technologies R & D Special Program of Anhui, China (Grant No. 16030901060). The National Natural Science Foundation of China (Grant No. 61502010). The Natural Science Foundation of Anhui Province (Grant No. 1608085QF146). The Natural Science Foundation of China (Grant No. 61806004).

References

- Balamurali, A. R.; Joshi, A.; Bhattacharyya, P.** (2012): Cross-lingual sentiment analysis for Indian languages using Linked WordNets. *Proceedings of COLING 2012: Posters*, pp. 73-82.
- Collobert, R.; Weston, J.; Bottou, L.** (2011): Natural language processing (almost) from scratch. *Journal of Machine Learning Research*, vol. 12, no. 1, pp. 2493-2537.
- Faruqui, M.; Dyer, C.** (2014): Improving vector space word representations using multilingual correlation. *Proceedings of 14th Conference of the European Chapter of the Association for Computational Linguistics*, pp. 462-471.

- Giozzo, A.; Strapparava, C.** (2006): Exploiting comparable corpora and bilingual dictionaries for cross-language text categorization. *Proceedings of the 21st International Conference on Computational Linguistics and the 44th Annual Meeting of the Association for Computational Linguistics*, pp. 53-560.
- Hinton, G. E.; Srivastava, N.; Krizhevsky, A.** (2012): Improving neural networks by preventing co-adaptation of feature detectors. *Computer Science*, vol. 3, no. 4, pp. 212-223.
- Kim, S. B.; Han, K. S.; Rim, H. C.; Myaeng, S. H.** (2006): Some effective techniques for naive bayes text classification. *IEEE Transactions on Knowledge & Data Engineering*, vol.18, no. 11, pp. 1457-1466.
- Kim, Y.** (2014): Convolutional neural networks for sentence classification. *Proceedings of 2014 Conference on Empirical Methods in Natural Language Processing*, pp. 1746-1751.
- Littman, M. L.; Dumais, S. T.; Landauer, T. K.** (1998): *Automatic Cross-Language Information Retrieval Using Latent Semantic Indexing*. Springer Press.
- Li, Y.; Shawe-Taylor, J.** (2006): Using KCCA for Japanese-English cross-language information retrieval and document classification. *Journal of Intelligent Information Systems*, vol. 27, no. 2, pp. 117-133.
- Martens, D.; Huysmans, J.; Setiono, R.; Vanthienen, J.; Baesens, B.** (2008): *Rule Extraction from Support Vector Machines: An Overview of Issues and Application in Credit Scoring*. Springer Press.
- Mikolov, T.; Chen, K.; Corrado, G.; Dean, J.** (2013): Efficient estimation of word representations in vector space. *Computer Science*.
- Ni, X.; Sun, J.; Hu, J.; Chen, Z.** (2011): Cross lingual text classification by mining multilingual topics from wikipedia. *Proceedings of 4th International Conference on Web Search and Web Data Mining*, pp. 375-384.
- Rigutini, L.; Maggini, M.; Liu, B.** (2005): An EM based training algorithm for cross-language text categorization. *IEEE/WIC/ACM International Conference on Web Intelligence*, pp. 529-535.
- Shin, K.; Abraham, A.; Han, S. Y.** (2006): Improving kNN text categorization by removing outliers from training set. *Proceedings of the 2006 Conference on Computational Linguistics and Intelligent Text Processing*, pp. 563-566.
- Xiao, M.; Guo, Y.** (2014): Semi-supervised matrix completion for cross-lingual text classification. *Proceedings of 28th AAAI Conference on Artificial Intelligence*, pp. 1607-1613.
- Yoshua, B.; Schwenk, H.; Senécal, J.; Morin, F.; Gauvain, J. L.** (2003): Neural probabilistic language models. *Journal of Machine Learning Research*, vol. 3, no. 6, pp. 1137-1155.