Ensemble of Binary Classification for the Emotion Detection in Code-Switching Text

Xinghua Zhang[1](✉), Chunyue Zhang[2](✉), and Huaxing Shi[2](✉)

[1] Harbin Institute of Technology, Harbin, China
xing_hua_zhang@126.com
[2] DeepIntell Co. Ltd., Harbin, China
{cyzhang,shihuaxing}@deepintell.cn

Abstract. This paper describes the methods for the DeepIntell who participated the task1 in the NLPCC2018. The task1 is to label the emotion in a code-switching text. Note that, there may be more than one emotion in a post in this task. Hence, the assessment task is a multi-label classification task. At the same time, the post contains more than one language, and the emotion can be expressed by either monolingual or bilingual form. In this paper, we propose a novel method of converting multi-label classification into binary classification task and ensemble learning for code-switching text with sampling and emotion lexicon. Experiments show that the proposed method has achieved better performance in the code-switching text task.

Keywords: Multi-label classification · Binary classification
Sampling · Emotion lexicon · Ensemble learning

1 Introduction

With the widespread popularity of social media, such as Weibo, WeChat, Twitter, etc., the analysis of the content of user articles has played a pivotal role in the field of natural language processing. Most of the previous emotion classification problems were conducted in monolingual corpora. However, with the diversification of culture, people usually publish some multilingual articles or comments. [E1-E4] are four examples of code-switching text on evaluation data that contain both Chinese and English words. In order to detect the emotions better in the text, it is necessary to consider the emotional expression in all of them. The assessment task mainly includes automatic classification of emotion (*happiness*, *sadness*, *anger*, *fear*, *surprise*) in code-switching text. Different from monolingual emotion detection, the emotion in code-switching text can be expressed in either monolingual or bilingual forms. In this task, we focus on Chinese and English mixed code-switching text. Although Chinese is the major language, it has been shown that English words are critical for emotion expression.

M. Zhang et al. (Eds.): NLPCC 2018, LNAI 11109, pp. 178–189, 2018.
https://doi.org/10.1007/978-3-319-99501-4_15

[E1] 这个 show 真**好看**, 今天感觉很 **happy**!
(*The show is really* ***nice****, I feel very* ***happy*** *today!*)
[E2] 上了一天的课了，嗓子 **hold 不住** 了啊。
(*I have been teaching the whole day, my throat* ***can't take it anymore****.*)
[E3] 我在 caffe in (中山南路店)。
(*I'm at caffe in (Zhongshan South Road).*)
[E4] 你今天真得**很棒**, **give me five**!
(*You are really* ***good*** *today,* ***give me five!***)

In fact, examples above are essentially different. E1 expresses the emotion by monolingual form, either Chinese or English. E2 has the bilingual form to express sadness. English words in E3 can't express any emotions. E4 expresses the emotion by English phrase. And the corpus data can be roughly divided into four categories above. However, if we want to fully uncover the emotions in the text, using an emotion lexicon as an aid seems like a good choice. Thus, we have collected the emotion lexicon in Chinese, English and mixture to explore the emotional information better in the code-switching text.

We adopt the multi-label classification method RCNN and the binary classification Fasttext [21], RCNN, and CNN [20] for the automatic classification of emotion. The experiment results show that the methods above have different advantages and disadvantages for different emotions, so ensemble learning technique may achieve better results. In addition, pre-processing (sampling) of the training data will get better scores when using the binary classification method for emotion recognition.

In fact, the evaluation is essentially a multi-label classification task. To detect the emotion better in the code-switching text, we propose a method of binary classification and ensemble learning with sampling and emotion lexicon(BCEL). For the five emotions of this evaluation: *happiness*, *sadness*, *anger*, *fear*, and *surprise*, we generated five data sets of the same size as the original training set. However, sentences in each sets are labelled differently. Take *happiness* as an example, if a sentence contains this emotion, it is marked as *happiness*, otherwise it is marked as *non*. Then we adopt the binary classification methods to train five training sets respectively to obtain the optimal model. At the end, we adopt ensemble learning technique and emotion lexicon for the five optimal models to obtain the final result.

The rest of the paper is organized as follows. In Sect. 2, we give an overview on the related work. In Sect. 3, we introduce our methods that we have tried. In Sect. 4, we describe our BCEL method. Experiment results and discussion are reported in Sect. 5. Finally, we draw some conclusions and give the future works.

2 Related Work

In this section, we discuss related works on emotion classification and code-switching text.

2.1 Emotion Classification

With the increasing number of texts with subjective assessment on the Internet, text emotion classification has gradually become a research hotspot in the field of natural language processing. Mostafa [3] used a pre-defined vocabulary of approximately 6,800 adjectives to perform emotion analysis of microblogs from multiple company consumers and found that consumers had positive emotions for well-known brands. Pang Lei et al. (2012) used emotional words and facial expressions as learning knowledge, using support vector machines, naive Bayes and maximum entropy to propose an unsupervised emotion classification method based on Sina Weibo.

A major related research in the emotion classification task is the creation of emotional resources, such as the creation of emotion lexicons. Xu et al. [8] apply a graph-based algorithm and multiple types of resources to create a Chinese emotion lexicon. Volkova et al. [9] introduced a dictionary for exploring language colors, concepts, and emotions. Moreover, most of the relevant studies have focused on supervised learning methods. Alm et al. [10] achieved text-based emotion prediction using machine learning methods. Aman and Szpakowicz [11] implemented sentence-level fine-grained emotion recognition through a knowledge-based approach. Chen et al. [7] detected emotion-induced events by analyzing the language architecture. Purver and Battersby [12] trained a fully supervised classifier using auto-labeled data to achieve multiple types of emotion prediction without manual intervention. Lin et al. [6] first described the emotion classification tasks of readers in news texts, and then applied some standard machine learning methods to train a classifier that recognizes readers' emotions.

2.2 Code-Switching Text

Code-switching text has received considerable attention in the NLP community. Several studies have focused on identification and analysis. Ling et al. [13] presented a novel method for translation in code-switched documents. Solorio and Liu [14] predicted potential code-switching points in Spanish-English. Lignos and Marcus [15] tried to identify code-switched tokens and Li et al. [16] added code-switched support to language models. Peng et al. [17] learned poly-lingual topic models from code-switching text. Lee et al. [1] proposed a multiple-classifier-based automatic detection approach to detect emotion in the code-switching corpus for evaluating the effectiveness of both Chinese and English texts. Wang et al. [2] proposed a Bilingual Attention Network (BAN) model to aggregate the monolingual and bilingual informative words to form vectors from the document representation, and integrate the attention vectors to predict the emotion.

3 Methods

3.1 Fasttext

Fasttext [21] is a text classification tool developed by Facebook. It provides a simple but efficient method for text representation and text classification. For the text classification part, it only has one hidden layer in the architecture so that the classification process is relatively fast. Fasttext classification function is similar to word2vec's continue bag of words (CBOW) algorithm. Firstly, the feature vector combined with word sequence is linearly projected to middle hidden layer. Secondly, there is a non-linear activation function which projects middle hidden layer to the categorization label. The difference between Fasttext and CBOW is that Fasttext predicts labels while CBOW predicts middle terms.

We adopt Fasttext to train one classifier for each emotion. And then we will get five classifiers and finally perform five different emotion predictions.

3.2 RCNN(Multi-label)

The structure of RCNN model is shown in Fig. 1. RCNN has a two-layer bidirectional LSTM to extract the context information of each word. It also uses Embedding to obtain the word information directly. Then the output of LSTM and Embedding is concatenated, and the concatenating result is performed with two-layer convolution operation after K-MaxPooling that extracts the local features further. Finally, the features are continuously input into a classification network composed of two-layer fully connected networks.

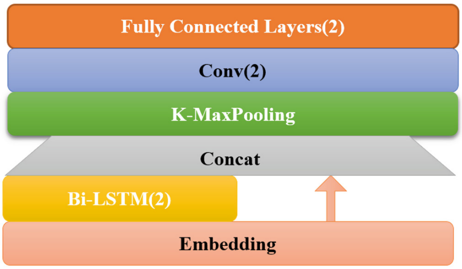

Fig. 1. The structure of Recurrent Convolutional Neural Networks model

3.3 RCNN(Binary)

The RCNN model adopted in this method is the same as the model used in the second method, except that the output of model is changed to the binary classification (idea of the first method). That is, using the RCNN to train five classifiers according to the emotion category, namely *happiness*, *sadness*, *anger*, *fear*, and *surprise*.

3.4 CNN(Binary)

The Convolutional Neural Networks(CNN) [20] model is shown in Fig. 2. It uses the word vectors (Word2Vec embedding) to obtain the word information directly.

Then, the model adopts four-layer convolution and two pooling layers (MaxPooling, AveragePooling) to extract features, in which the MaxPooling is used after the first two convolution layers, AveragePooling is adopted at the end. After extracting features using AveragePooling, the features are input into the fully connected layer continuously for classification.

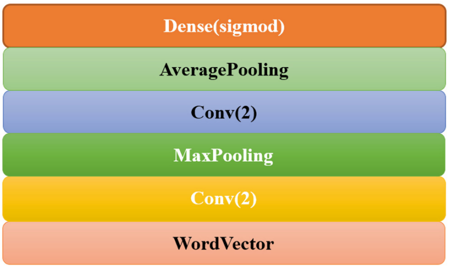

Fig. 2. The structure of Convolutional Neural Networks model

4 Binary Classification and Ensemble Learning with Sampling and Emotion Lexicon

Although research on text classification has been carried out for many years, most of the current studies assume a balanced distribution of samples of various categories. However, the reality is often not the case. In the actual collection of corpus, whether it is the product review text or microblogging text, Twitter text, etc., multi-class classification or binary classification, the distribution of samples in each category is often very unbalanced. What's more, the unbalanced distribution of samples will make the classification results obtained by applying the neural network classification method heavily biased towards classes with a large number of samples, thereby greatly reducing the classification performance. For the data of this evaluation, regardless of multi-label classification or binary classification, the distribution of samples in the classification is unbalanced. If the neural network model is adopted only, the performance will be affected.

To address the above challenges, we propose a binary classification and ensemble learning with sampling and emotion lexicon(BCEL) method. We firstly take sample on training set and perform ensemble learning with emotion lexicon finally.

4.1 Sampling

The current mainstream imbalance classification method is based on undersampling and oversampling machine learning classification methods. The main idea of the method is to use the undersampling or oversampling technique to obtain a balanced sample, and then classify the sample by a machine learning classification method. However, only a part of samples in training set can be used when adopting the undersampling technique, thus losing many samples that may be helpful for classification. Therefore, if the training set is too small, the undersampling method is not advocated. In order to make full use of the existing samples, we adopt the oversampling method (based on the undersampling technique) so that the number of positive and negative samples is basically the same. The basic model is shown in Fig. 3.

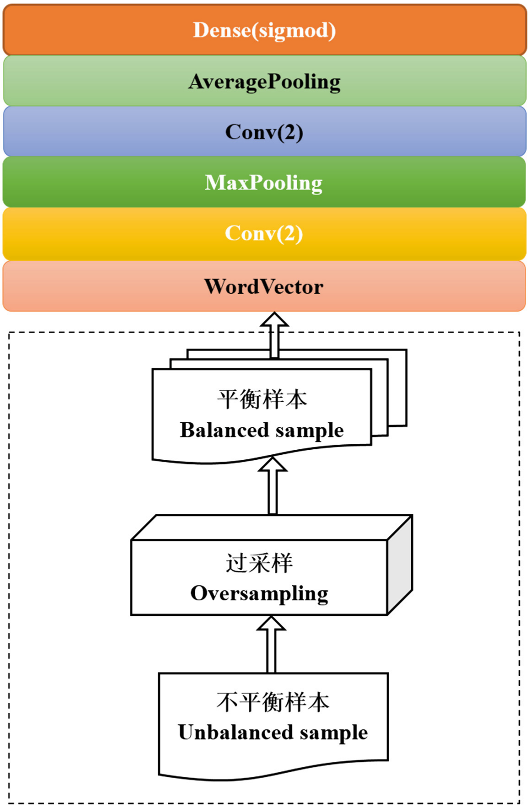

Fig. 3. The structure of Convolutional Neural Networks model with sampling

The oversampling technique we adopt is essentially based on undersampling, and the specific process is as follows: We first use the random undersampling method to perform undersampling n times for each category of samples (positive and negative samples for binary classifications). The number of samples we choose is equal to category which has minimum quantity (In this evaluation, it is the number of samples with a certain type of emotion). In the end, we get n sets of balanced samples, and then merge the n sets of balanced samples into a training set sample for training, as shown in Fig. 4.

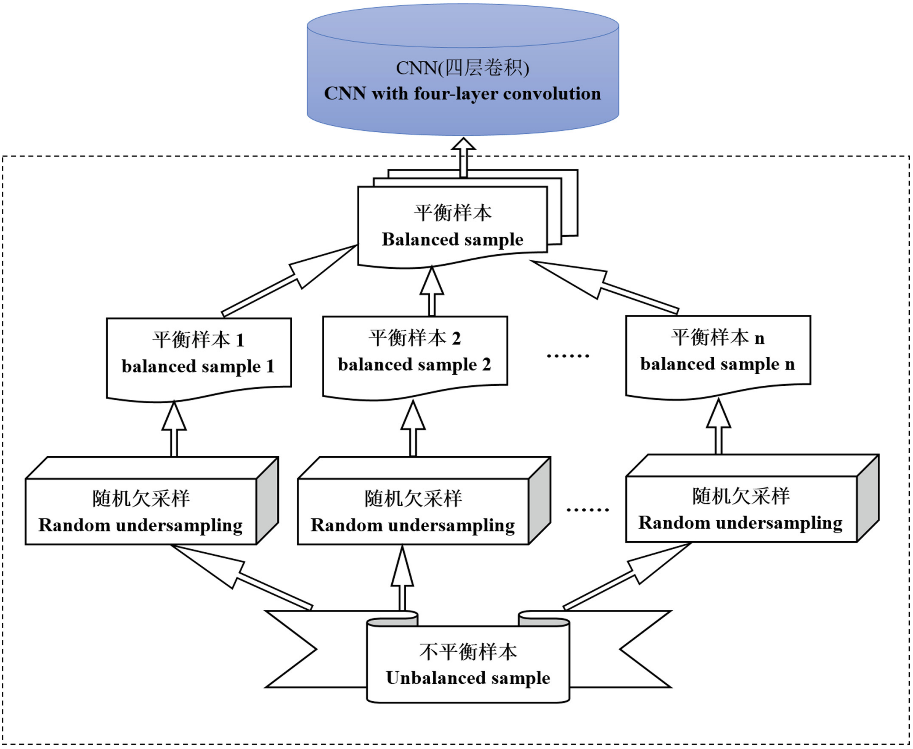

Fig. 4. The oversampling technique based on undersampling

4.2 Ensemble Learning

The idea of ensemble learning is to integrate several individual classifiers when classifying new instances, and to determine the final classification through a combination of the classification results of multiple classifiers to obtain better performance than a single classifier. If a single classifier is compared to a decision maker, the ensemble learning approach is equivalent to multiple decision makers making a common decision. Figure 5 shows the basic idea of integrative learning.

After getting multiple classifiers, the last step which is also the most important is to adopt combination strategies. Since there are few training data sets, we use a simple voting method. The specific process is as follows:

1. Using the classifier model with the highest accuracy (P-value) (we'll call it H) as the standard, whose prediction is right we think.
2. If the prediction of classifier model H is *non* (that is, there is no emotion), we look at the prediction results of other classifier models above, and take the emotion with the largest number of predictions (that is, other classifiers except H adopt the principle of the minority obeying the majority).
3. Because there are more posts containing *happiness* or *sadness*, therefore we introduce emotion lexicon for *happiness* and *sadness* classification.

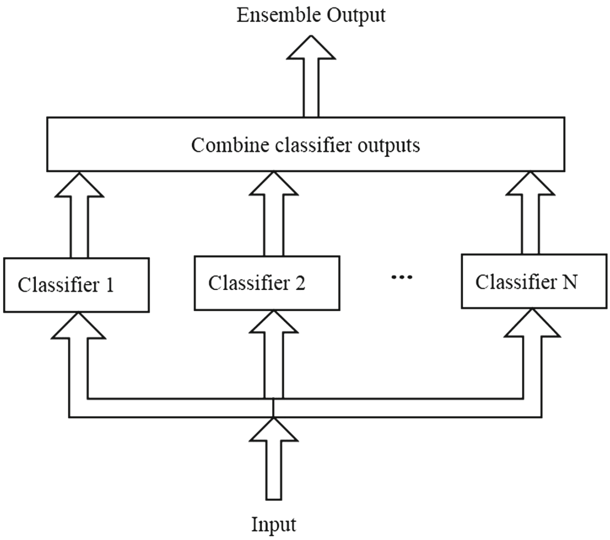

Fig. 5. The basic idea of ensemble learning

5 Experiments

5.1 Datasets

The number of training set is 6000, development set is 728, and the test set has 1200 posts in this shared task. There may be more than one emotion in a post. For each post, five basic emotions were annotated, namely *happiness*, *sadness*, *fear*, *anger* and *surprise*. The distribution of different emotions in training, development and test data is shown in Table 1.

Table 1. The distribution of different emotions.

	Train	Dev	Test
Happiness	0.304	0.302	0.408
Sadness	0.181	0.165	0.247
Anger	0.095	0.115	0.093
Fear	0.108	0.117	**0.031**
Surprise	0.109	0.126	**0.053**

In this shared task, we strictly use the given training dataset and construct the emotion lexicon by the emotion ontology of Dalian University of Technology (DLUT) and some English lexicons from github[1].

[1] https://github.com/timjurka/sentiment/tree/master/sentiment.

5.2 Experiment Settings

In our experiments, the input is a 300-dimensional vector denotes word embedding. We use the longest word-level (60) length of post in the data to unfold the BiLSTM networks. The learning rate is initialized to 0.001, and decay rate per 128 training steps is 0.9. We use a fixed number of epochs and always save the model with the best F1 score of the development set. The specific model parameters are set as shown in Table 2.

Table 2. Parameter configurations of our model.

Parameters	Configurations
Word embedding dimension	300
Learning rate	0.001
Loss function	Softmax,binary-crossentropy
PretrainedVectors	Yes
Number of epochs	20,40
Number of negatives sampled	3,20
Number of convolution	64,128,256
Size of kernel	3
Pooling Size	3
Decay rate	0.9
Batch size	128
Vocabulary size	213543

5.3 Results

According to official evaluation requirements, we compute precision (P), recall (R), and F1-Score for each emotion separately, and calculate the macro averaged P, R and F1 with all emotions. The official scoring metric is macro-averaged F1-Score.

We experimented with the mentioned methods above to get a comparison of the performance of the model on the development set, as shown in Table 3.

Table 3. The F1-Score of each model.

$F1 - Score$ Model / Emotion	$Fasttext$	$RCNN(B)$	$RCNN(M)$	CNN	$CNN(S)$
happiness	0.600	0.573	0.611	0.564	**0.615**
sadness	**0.491**	0.411	0.364	0.487	0.487
anger	0.489	0.454	0.397	0.482	**0.606**
fear	0.350	0.395	0.206	0.368	**0.429**
surprise	**0.383**	0.322	0.100	0.263	0.378

As shown in Table 3, Fasttext and CNN model are binary classification. RCNN(B) is binary while RCNN(M) is multi-label classification. CNN(S) is binary classification and adopts sampling technique.

To achieve better scores, we introduce ensemble learning with emotion lexicon to determine the final result of the whole development samples by sub-models voting. The precision (P), recall (R), and F1-Score for each emotion on the development set, are shown in Table 4. Table 5 shows the test set scores.

Table 4. The final result by BCEL method on the development set.

	P	R	F1
Happiness	0.609756	0.681818	0.643776
Sadness	0.460122	0.616666	0.527016
Anger	0.501905	0.741379	0.598579
Fear	0.445455	0.502941	0.472456
Surprise	0.289412	0.603478	0.391210

Table 5. The final result by BCEL method on the test set.

	P	R	F1
Happiness	0.776765	0.695918	0.734123
Sadness	0.683128	0.560811	0.615955
Anger	0.613636	0.486486	0.542714
Fear	0.222222	0.324324	0.263736
Surprise	0.366667	0.485294	0.417722

5.4 Discussion

For emotion *happiness* and *sadness*, the difference between the positive and negative examples in quantity is not disparate. Therefore, the data sampling preprocessing does not significantly improve their performance. In order to improve the performance of the model for happiness and sadness, we tried to merge the training results of other models and the regular methods of the lexicon (Chinese and English). We fused the ensemble learning results of the above five models with the lexicon rule method. And the lexicon rule threshold k is set to 2, that is, if the number of certain emotion words in a sentence is greater than or equal to 2 or the ensemble learning result contains the emotion, then the post has that emotion. After that, the final results of development and test data are shown in Tables 4 and 5.

6 Conclusion

This paper presents an effective approach for code-switching text based on binary classification and ensemble learning. Firstly, we adopt the method of converting multi-label classification into binary classification task. At the end, the output of multiple classifiers is combined to form the final prediction result. However, the proportion of positive and negative examples in training data is very different, especially for the three emotions of *anger*, *fear*, and *surprise*. Therefore, we use the oversampling technique to balance the data and obtain relatively balanced training samples. However, for emotion *happiness* and *sadness*, the difference between the positive and negative examples in quantity is not disparate, so we introduced the English and Chinese emotion lexicon to fully explore the emotion of *happiness* and *sadness*. Next we will focus on the interaction between emotions and the construction of the Chinese-English hybrid dictionary will be an important aspect in the future.

References

1. Lee, S., Wang, Z.: Emotion in code-switching texts: corpus construction and analysis. In: Proceeding of SIGHAN-2015 (2015)
2. Wang, Z., Zhang, Y., Lee, S., Li, S., Zhou, G.: A bilingual attention network for code-switched emotion prediction. In: Proceeding of COLING-2016 (2016)
3. Mostafa, M.M.: More than words: social networks' text mining for consumer brand sentiments. Expert Syst. Appl. **40**(10), 4241–4251 (2013)
4. Vapnik, V.N.: The Nature of Statistical Learning Theory. Springer, New York (1995). https://doi.org/10.1007/978-1-4757-3264-1
5. Gao, W., Li, S., Lee, S.Y.M., Zhou, G., Huang, C.R.: Joint learning on sentiment and emotion classification. In: Proceedings of CIKM 2013 (2013)
6. Lin, K., Yang, C., Chen, H.: Emotion classification of online news articles from the reader's perspective. In: Proceeding of the International Conference on Web Intelligence and Intelligent Agent Technology, pp. 220–226 (2008)
7. Chen, Y., Lee, S., Li, S., Huang, C.: Emotion cause detection with linguistic constructions. In: Proceedings of COLING-10, pp. 179–187 (2010)
8. Xu, G., Meng, X., Wang, H.: Build Chinese emotion lexicons using a graph-based algorithm and multiple resources. In: Proceeding of COLING-2010, pp. 1209–1217 (2010)
9. Volkova, S., Dolan, W., Wilson, T.: CLex: a lexicon for exploring color, concept and emotion associations in language. In: Proceedings of EACL 2012, pp. 306–314 (2012)
10. Alm, C., Roth, D., Sproat, R.: Emotions from text: machine learning for text-based emotion prediction. In: Proceedings of EMNLP, pp. 579–586 (2005)
11. Aman, S., Szpakowicz, S.: Identifying expressions of emotion in text. In: Matoušek, V., Mautner, P. (eds.) TSD 2007. LNCS (LNAI), vol. 4629, pp. 196–205. Springer, Heidelberg (2007). https://doi.org/10.1007/978-3-540-74628-7_27
12. Purver, M., Lee, S., Li, S., Huang, C.: Emotion cause detection with linguistic constructions. In: Proceedings of COLING-2010, pp. 179–187 (2010)
13. Ling, W., Xiang, G., Dyer, C., Black, A., Trancoso, I.: Microblogs as parallel corpora. In: Proceedings of ACL-2013 (2013)

14. Solorio, T., Liu, Y.: Learning to predict code-switching points. In: Proceedings of EMNLP 2008 (2008)
15. Lignos, C., Marcus, M.: Toward web-scale analysis of codeswitching. In: Proceedings of Annual Meeting of the Linguistic Society of America (2013)
16. Li, Y., Fung, P.: Code-switch language model with inversion constraints for mixed language speech recognition. In: Proceedings of COLING-2012 (2012)
17. Peng N., Wang, Y., Dredze, M.: Learning polylingual topic models from code-switched social media documents. In: Proceedings of ACL14 (2014)
18. Yan, Y., Liu, Y., Shyu, M.L., et al.: Utilizing concept correlations for effective imbalanced data classification. In: IEEE International Conference on Information Reuse and Integration, pp. 561–568. IEEE (2014)
19. Wang, Z., Lee, S.Y.M., Li, S., Zhou, G.: Emotion detection in code-switching texts via bilingual and sentimental information. In: Proceeding of ACL-2015, short paper, pp. 763–768 (2015)
20. Kim, Y.: Convolutional Neural Networks for Sentence Classification[J]. Eprint Arxiv (2014)
21. Joulin, A., Grave, E., Bojanowski, P., et al.: Bag of tricks for efficient text classification, pp. 427–431 (2016)