J Inf Process Syst, Vol.18, No.4, pp.535~548, August 2022
https://doi.org/10.3745/JIPS.04.0249

ISSN 1976-913X (Print)
ISSN 2092-805X (Electronic)

JOURNAL OF INFORMATION PROCESSING SYSTEMS JIPS

Burmese Sentiment Analysis Based on Transfer Learning

Cunli Mao, Zhibo Man, Zhengtao Yu*, Xia Wu, and Haoyuan Liang

Abstract

Using a rich resource language to classify sentiments in a language with few resources is a popular subject of research in natural language processing. Burmese is a low-resource language. In light of the scarcity of labeled training data for sentiment classification in Burmese, in this study, we propose a method of transfer learning for sentiment analysis of a language that uses the feature transfer technique on sentiments in English. This method generates a cross-language word-embedding representation of Burmese vocabulary to map Burmese text to the semantic space of English text. A model to classify sentiments in English is then pre-trained using a convolutional neural network and an attention mechanism, where the network shares the model for sentiment analysis of English. The parameters of the network layer are used to learn the cross-language features of the sentiments, which are then transferred to the model to classify sentiments in Burmese. Finally, the model was tuned using the labeled Burmese data. The results of the experiments show that the proposed method can significantly improve the classification of sentiments in Burmese compared to a model trained using only a Burmese corpus.

Keywords

Burmese, Cross-Lingual, Sentiment Analysis, Transfer Learning

1. Introduction

Deep neural networks have delivered good results in the classification of sentiments in English because a large number of corpora with labelled data for English are available. However, for languages, such as Burmese ("Burmese" means the same as "Myanmar") [1], which are not widely spoken, only a small amount of labeled data can be obtained by collecting and manually labeling corpora in the language. Thus, the training data were insufficient in volume, which affected the accuracy of the classification of sentiments in Burmese.

Most currently available methods for analyzing the sentiments of texts are based on sentiment rules [2,3] and deep learning methods [4]. Sentiment rules are based on natural language processing technology that uses sentiment knowledge to classify tendencies toward sentiment in the textual content. For example, Ohana and Tierney [3] used the universal sentiment dictionary SentiWordNet to identify sentiment words in texts and calculate sentiment scores. Hatzivassiloglou and McKeown [4] claimed that junctions that connect adjectives can discriminate the sentiment-based tendencies of adjectives. Because the natural

Manuscript received November 19, 2020; first revision February 17, 2022; second revision March 21, 2022; accepted March 25, 2022.
* **Corresponding Author:** Zhengtao Yu (ztyu@hotmail.com)
College of Information Engineering and Automation & Yunnan Key Laboratory of Artificial Intelligence, Kunming University of Science and Technology, Kunming, China (maocunli@163.com, zhibo_man@yeah.net, ztyu@hotmail.com, 1421675565@qq.com, 379970083@qq.com)

language processing technology used in such methods is not sufficiently advanced, it is important to construct a knowledge base for sentiments and rules of sentiment classification in advance; however, this imposes significant limitations on the development of methods based on sentiment knowledge.

With the use of deep learning in major breakthroughs in many natural language processing problems in recent years, scholars have introduced deep learning models for sentiment analysis tasks. A growing number of researchers are using semi-supervised learning methods to reduce dependence on labeled samples. For example, co-training [5], label propagation (LP) [6], and deep learning [7] have achieved good results. The authors proposed a collaborative training algorithm based on the double-view bag-of-words representation that automatically constructs antisense comments and models the comment text through a pair of opposite-view bag-of-words representations [5]. Chen et al. [6] proposed a knowledge verification model that integrates knowledge transfer using sentiment knowledge from resource-rich languages to identify the polarity of sentiments in texts in low-resource languages. This study aimed to prevent erroneous knowledge by distinguishing highly trustworthy knowledge. Ruangkanokmas et al. [7] proposed a semi-supervised learning algorithm called a deep dependency network with feature selection to reduce noise and redundant vocabulary in sentiment classification algorithms.

Using labeled data from resource-rich languages to automatically detect the polarity of sentiments in unlabeled data in the target language can solve the problem of classifying sentiments in resource-scarce languages [8-12]. For example, Wan [8] used machine translation to translate training comment texts in English into text in Chinese, translated unlabeled comments into Chinese, and jointly trained two sentiment features for cross-language sentiment analysis. Lin et al. [9] proposed a cross-language joint aspect/sentiment (CLJAS) model to perform sentiment analysis of a target language using the knowledge learned from the source language. The CLJAS model detects the emotions of the two languages by integrating emotions into a cross-language topic model framework. Chen et al. [10] proposed an adversarial deep averaging network (ADAN1) to transfer knowledge learned from labeled data in a resource-rich source language to a resource-poor language for which only unlabeled data are available. Zhou et al. [11] proposed a representation-learning approach that simultaneously learns vector representations for texts in both source and target languages. With the rapid development of neural networks, increasing attention has been paid to other aspects of sentiment analysis tasks, such as graph-network-based sentiment analysis [12], fine-grained sentiment analysis [13], and triple-based sentiment analysis methods [14]. These tasks are fundamental monolingual sentiment analyses.

However, the above models divide annotated data and unlabeled data into different models, and most rely on machine translation without considering the resulting loss in the correlation between the models owing to errors occurring during translation. Accordingly, in this study, we propose a classification of sentiments for Burmese based on transfer learning. The proposed classification method uses rich labeled data from sentiment classification in the English language.

2. Method of Burmese Sentiment Analysis based on Transfer Learning

As shown in Fig. 1, training the sentiment classification model in Burmese can be divided into three steps.

(1) Pre-training the sentiment classification model in English.
(2) Training the sentiment classification model in Burmese using the parameters of the English sentiment classification model by incorporating the features of the former into the semantic space of the latter, obtaining the classification of sentiments of Burmese, and
(3) Using tagged data to tune the Burmese model.

Using the word vector conversion tool word2vec, the English sentence was expressed as word vectors, and the vector form corresponding to the input to the convolutional neural network (CNN) was thus established. The features of a sentence are extracted by the CNN to obtain an effective feature representation and a convolutional nerve. The features extracted by the network are max-pooled to obtain the most valuable part, and the softmax layer outputs the classification of sentiments according to probability. The classification of English sentiments was improved by pre-training the model. The Burmese word vector was integrated into the semantic space of English through mapping. For a merged Burmese sentence, the parameters between the corresponding pairs of network layers are shared in the same way, and a sentiment classifier for Burmese is obtained. Finally, we use Burmese data with sentiment markers to tune the model.

Li et al. [14] trained a neural network using a combination of data from a main language and an auxiliary language. In contrast to this method, all layers of our cross-language deep learning model were shared by building a mapping between the English and Burmese bilingual word vectors. This implies that we have incorporated the features of Burmese into English. The two languages already have a certain similarity because information can be shared to improve the accuracy of the classification of sentiments in Burmese.

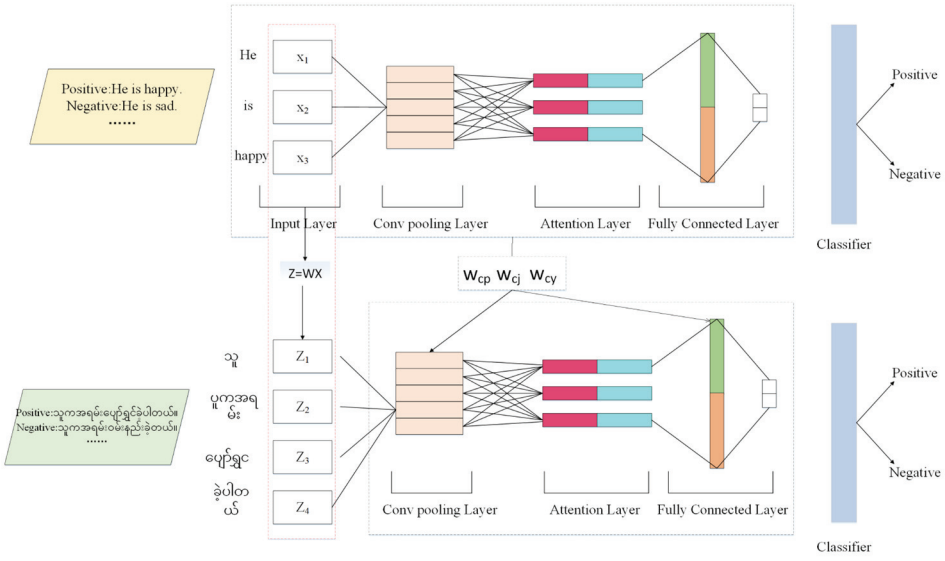


Fig. 1. Overall architecture of our model.

2.1 Pre-training English Sentiment Classification Model

2.1.1 Extracting features of the convolutional layer of English model

Because a CNN can obtain contextual features of vocabulary, it can help obtain effective feature representations. However, for natural language processing tasks, the input is not a pixel of an image but

a sentence represented by a matrix. The convolution operation of the target matrix obtains each local feature and combines the feature vectors to obtain the feature vector of the target matrix. In pre-training the English network, the input is an English sentence X characterized as a sentence vector matrix $[CW_1, CW_2, \dots, CW_n]$ consisting of the word vector of the sentence. Each row in the matrix represents an English word vector CW, and n represents the number of words in the sentence. The vector representation can be obtained by combining contextual information and new English sentences. As in the method proposed by Wang et al. [16], our convolution operation contains a filter W that causes the CW vectors to generate a new feature Z:

$$Z = W_j X_i \tag{1}$$

Here, W_j is the ith input matrix and ith instance, and W_j is the j-th filter in the convolution operation ($1 \le j \le 30$). When extracting Burmese features, all filters W share the extraction parameters, which significantly reduces the number of parameters in the learning process. After passing through the filter W, the corresponding characteristic output Z is obtained. To obtain the most useful information from eigenvector Z, we perform a max-pooling operation on Z:

$$m_s = \max(Z_s), 0 \le s \le j \tag{2}$$

The feature vector of an English sentence automatically synthesizes a linear vector. To learn more complex features, we designed a nonlinear layer and selected a rectified linear function (ReLU) as an activation function. In training the sentiment classification model of English, using the sigmoid function when the random initial network weight is too large causes network training to become unstable; however, using the ReLU activation function can effectively prevent the weight from being too large or too small. The activation function can be written as

$$g = \max(0, W_y, T) \tag{3}$$

Here, W_y is a linear transformation equation that maps the vector T to the hidden layer and uses the ReLU activation function to obtain g, which denotes a higher level of the characteristics of English speech syllables.

2.1.2 Attention mechanism

Following convolution, the attention mechanism is used to obtain the feature information of different important programs to improve classification accuracy [15]. The attention text was used here.

$$s_i = fun(x_{ij}, U_i) \tag{4}$$

$$a_i = \frac{\exp(s_i)}{\sum_{i=1}^{n} s_i}, 1 \le i \le n \tag{5}$$

$$att_i = \sum_{i=1}^{n} a_i \cdot x_i, \tag{6}$$

where x_{ij} represents a sentence, U_i represents the label corresponding to this sentence, fun represents a forward network with a hidden layer, and s_i and a_i represent the importance of the corresponding words in the text.

2.1.3 Output layer of English sentiment classification

To estimate the classification of the sentiment expressed in each input English sentence X, the predicted output of the softmax layer is:

$$o = W_p(g \otimes a_i), \tag{7}$$

where W_p is a linear transformation equation, vector g and attention text a_i are fully connected and mapped to the output layer, and $\otimes$ indicates a fully connected operation. Each output o is the sentiment score of the input English sentence vector matrix X, and there are two predictions of 0 and 1. If the score is 0, the relevant English sentence represents negative sentiment; if the score is 1, the relevant English sentence expresses positive sentiment.

2.1.4 Defining the English sentiment classification model losses to solve the model

Finally, the probabilities of the positive and negative terms are obtained through the softmax layer, and the highest probability is used as the label for English sentiment classification.

$$p(Xci) = \frac{e^{o_{ci}}}{\sum_{k=1}^{2} e^{o_{ck}}}. \tag{8}$$

The final label U_c is obtained according to the calculated probability. If the positive sentiment value is greater than the negative sentiment value, U_c is a positive sentiment and vice versa. As with English sentiment classification, cross-entropy is used as a loss function.

$$loss_{Ly} = CrossEntropy(U_c, \overline{U}_c), \tag{9}$$

where U_c is the sentiment rating of the model and $\overline{U}_c$ is the tag of the relevant sentence. By finding a loss in the model, all its parameters are updated in reverse such that they are closer in value to the data for the classification of English sentiments. The cross-entropy alone was used as the loss function. During the update, the values of the parameters may be too large or small. Therefore, the parameters of the L2 canonical constraint model were used here by increasing the regularity constraints. The parameters in the model include the English sentence vector x input to the model and the weight matrices W_j, W_y, W_p. The loss in classifying sentiments in English can then be expressed as

$$loss_{Ly} = CrossEntropy(Uc, \overline{U}c) + \left(W_j\right)^2 + \left(W_y\right)^2 + \left(W_p\right)^2. \tag{10}$$

A stochastic gradient descent algorithm was used to solve the model to obtain the minimum loss of the sentiment classification model in English. Once the model converges, its parameters W_j, W_y, W_p were obtained for sentiment analysis in English, and fixed to obtain W_{cj}, W_{cy}, W_{cp}. These parameters were also used in the model to classify the sentiments in Burmese.

2.2 Fusing Training of Burmese Classification Model using Features for English Sentiment Classification

The parameters are used as the initialization parameters for sentence sentiment classification in Burmese. Using the mapping relationship between English and Burmese, the latter language is mapped onto the space of the former, and the characteristics of sentiment classification in English are used to compensate for the lack of features in Burmese. Finally, the model parameters were updated using the loss.

2.2.1 Bilingual vectorization representation

The English-Burmese bilingual word vector map and bilingual sentence mapping between English and Burmese were used to establish the relationship between sentences in the languages. This reduces the difference between languages and avoids performance degradation during the feature transfer. The use of mapping can also complement information that is absent from the classification of sentiments in Burmese. The Burmese sentence input to the model consisted of words. The word vector of each Burmese word is M_W and the target sentence matrices are $[M_{W1}, M_{W2}, \dots, M_{WS}]$.

The crucial step in learning bilingual word-vector mapping is to establish a mapping relationship of words through a bilingual dictionary. Words that do not appear in the dictionary are then used to find the target words according to the constructed word-mapping relationship. In our model, Mikolov's method was used in a nested loop [17]. Each time a loop is executed, the dictionary is updated and used to train the mapping relationship, until the model converges.

In Fig. 2, *X* represents the distribution of the word vectors of Burmese in the Burmese semantic space, and *Z* represents the distribution of the word vectors of English in the English semantic space. Using an initial English-Burmese dictionary, the spatial distances between pairs of translated words in the dictionary were minimized, and a transformation matrix *W* was learned. The word vectors of Burmese were then mapped into the semantic space of English using *W*, and the dictionary was supplemented. Using this new dictionary, the spatial distances between the translated words in the dictionary were minimized again and the transformation matrix W was relearned to further expand the dictionary. This iteration stops when the dictionary does not expand in size during the successive iterations.

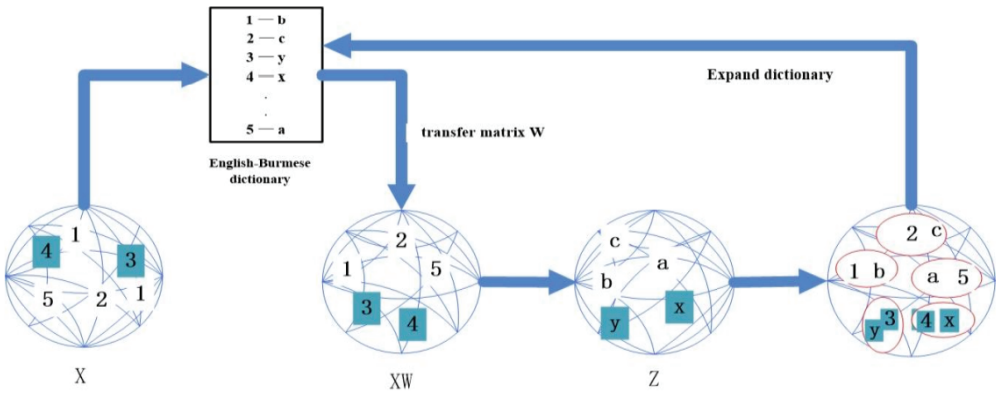


Fig. 2. Establish bilingual word vector mapping.

A self-learning English-Burmese training framework is proposed in this study.

Input: A single-word vector trained by two languages in their respective corpora. X is the original language, Z is the target language, and D is the bilingual dictionary. The process is as follows.

a) Iteration (through the iterative process, constantly expand the dictionary).

b) Spatial mapping matrix W obtained by (X, Z, D) training.

c) Expand dictionary D by (X, Z, W) until the model converges.

d) Until the model converges.

e) Evaluation dictionary D.

2.2.2 Extracting features of Burmese sentences through convolution layers

The convolution operation of the target matrix obtains each local feature and combines the feature vectors to obtain the feature vector of the target matrix. In pre-training the English network, the input is a Burmese sentence X characterized by a sentence vector matrix $[CW_1, CW_2, \dots, CW_n]$ consisting of the word vector of the sentence. Each row in the matrix represents an English word vector CW, and n represents the number of words in the sentence. The representation of the vector can be obtained by combining the contextual information of the syllables into new Burmese sentences. Using the same model as for English, the parameters of the convolution network filter CW were used to convolute the sentence vectors of Burmese, the features were extracted, and a new vector M_W was generated

$$Z_B = W_{Cj} X_{Bi}, \tag{11}$$

where X_{Bi} is the ith Burmese sentence input to the model, and W_{Cj} represents the well-trained parameters of the English model. Parameters that have been trained on the English model are not changed here, and the vectors of Burmese sentences are convoluted to extract their features. The same parameters were used in English ($1 \le j \le 30$). Once the sentences pass the filter W_{Cj}, a corresponding characteristic output Z_{Bs} is obtained. To obtain the most useful information from the feature vector Z_{Bs}, we perform the same max-pooling operation as in English:

$$m_{Bs} = \max(Z_{Bs}), 0 \le s \le j \tag{12}$$

For the most valuable information, m_{Bs}, the ReLU activation function used for sentiment classification in English was employed:

$$g_B = \max(0, W_{Cy}, T), \tag{13}$$

where W_{Cy} is the linear transformation equation. Vector T is mapped to the hidden layer, and the ReLU activation function is used to obtain g, which represents the characteristics of higher-level Burmese sentences.

2.2.3 Attention mechanism

Following the convolution operation, the attention mechanism is used to obtain feature-related information on important programs to improve the accuracy of classification. The attention text is used to indicate the following.

$$s_i = fun(x_{ij}, U_i) \tag{14}$$

$$a_i = \frac{\exp(s_i)}{\sum_{i=1}^{n} s_i}, 1 \le i \le n \tag{15}$$

$$att_i = \sum_{i=1}^{n} a_i \cdot x_i. \tag{16}$$

Here, x_{ij} represents a sentence, U_i represents the label corresponding to this sentence, fun represents a forward network with a hidden layer, and s_i and a_i represent the importance of the corresponding words in the text.

2.2.4 Updating model parameters through Burmese sentiment classification loss

The extracted Burmese sentences pass through the softmax layer and are assigned a Burmese score for each category under the model.

$$o_B = W_{cp}(g_B \otimes a_i). \tag{17}$$

Finally, the probabilities of the positive and negative terms are obtained through the softmax layer, and the term with the highest probability is chosen as the label for sentiment classification in Burmese.

$$p(X_{Bi}) = \frac{e^{o_{Bi}}}{\sum_{k=1}^{2} e^{o_{Bk}}} \tag{18}$$

The final tag U_B was obtained based on probability. If a positive emotion is greater than a negative emotion, $\overline{U}_B$ is a positive sentiment, and vice versa. As with sentiment analysis in English, cross-entropy was used as a loss function.

In the training process of the classifier for Burmese, the parameters that were trained in the space of the model for English sentiments were used as the initial parameters. The loss in this model is inversely updated according to its value. This model was applied to the sentiment analysis of the Burmese people.

$$loss_{L_y} = CrossEntropy(U_B, \overline{U}_B) + \left(W_{Bj}\right)^2 + \left(W_{By}\right)^2 + \left(W_{Bp}\right)^2. \tag{19}$$

2.2.5 Model tuning

After mapping Burmese to English, although the result has the semantic features of English, there are deviations. Therefore, a small-scale Burmese training model was used to obtain the loss in the sentiment classification of Burmese, and was solved by minimizing the loss. In the Burmese sentiment classification update parameters W_{Bj}, W_{By}, W_{Bp}, through the formula (19) constraint model when fitting the Burmese sentiment analysis features, the model cannot be infinitely close to Burmese features. The set of annotations in Burmese is only a small part; thus, a constraint is implemented to avoid overfitting. The Burmese sentences were mapped into the semantic space of English, with similar semantic features as English, but with subtle differences. By learning this difference, the model could perform better in classifying sentiments in Burmese.

$$\left(W_{Bj} - W_{Cj}\right)^2 + \left(W_{By} - W_{Cy}\right)^2 + \left(W_{Bp} - W_{Cp}\right)^2. \tag{20}$$

Finally, the loss of sentiment classification in Burmese can be expressed as:

$$loss_{L_d} = CrossEntropy(U_B, \overline{U}_B) + \left(W_{Bj} - W_{Cj}\right)^2 + \left(W_{By} - W_{Cy}\right)^2 + \left(W_{Bp} - W_{Cp}\right)^2. \tag{21}$$

Negative transfer is an important factor that affects the performance of the model. In our study, we mainly artificially evaluated and corrected the negative data. In the constructed English-Burmese parallel data, we adopted professional manual data correction and annotation.

3. Experiment

3.1 Experimental Data

The labeled corpus used for sentiment classification in English was obtained from English sentiment analysis data. As shown in Table 1, 50,000 English sentences are used. Their polarity indicates semantic tendencies. The data for Burmese sentiment classification were obtained from an artificially constructed labeled dataset consisting of 15,000 English-Burmese sentences. Examples of partially constructed English-Burmese parallel sentence pairs are provided in Table 2.

Table 1. Details of dataset

	English	Burmese
Train data	25,000	10,500
Test data	2,000	3,000
Validation data	5,000	1,500

3.2 Experimental Methods and Evaluation Indicators

The Burmese language is resource-poor in terms of its labelled datasets. There is no public dictionary of sentiment-related words. This study used feature transfer (Att-CNN-Trans) [18] to exploit the advantages of an English corpus, specifically sentiment analysis in English, for sentiment classification in Burmese to compensate for the scarcity of Burmese corpora.

To verify the effectiveness of the proposed method, comparative experiments were designed:

(1) Traditional SVM [19] and linear regression (LR) [20] were used for comparison with the proposed method. CNN, LSTM, BILSTM, and fastText [21] were used to train 10,500 labeled Burmese sentences, which did not use an English labeled set to pre-train the model and did not map Burmese sentences to English.

(2) Att-BiLSTM [22] was used to train the 10,500 labeled Burmese sentences.

(3) CNN-Trans was used to train the 10,500 labeled Burmese sentences. A labeled dataset of English was used to pretrain the model, and Burmese sentences were mapped to English. However, an attention mechanism was not used.

(4) Att-CNN was used to train 10,500 labeled Burmese sentences. A labeled dataset of English was used to pretrain the model, and Burmese sentences were mapped to English. An attention mechanism was also used.

Table 2. Examples of English-Burmese sentences

ID	English	Burmese
1	The parents welcome the prodigal son with open arms.	အပြင်တွင်သုံးဖြုန်းပျော်ပါးပြီးအိမ်ပြန်လာသော သားကို မိဘတို့က နွေးထွေးစွာ လက်ကမ်း ကြိုဆိုကြသည် ။
2	I would be happy to sell you oil, said the false merchant.	ခင်ဗျားကို ဆီ ရောင်း ဖို့ ပျော်တာပေါ့ ဟု ကုန်သည် အတု က ပြောတယ် ။
3	I could only imagine that my life would be full of darkness, sadness and hopelessness.	ကျွန်တော့် ဘဝတစ်ခုလုံး ဝမ်းနည်းမှုတွေ မျှော်လင့်ချက်ကင်းမဲ့မှုတွေ နဲ့ မှောင်အတိကျသွား လိမ့်မယ် လို့ ကျွန်တော် ထင်လိုက်မိပါတော့တယ် ။
4	It was a great sadness to him that he never had children.	သူ ဘယ်တုန်းကမှ သားသမီး မရရှိနိုင်ခဲ့တာ ဟာ သူ့အတွက်တော့ အကြီးအကျယ် ဝမ်းနည်းစိတ်မကောင်းစရာ ဖြစ်တယ် ။
5	Scientists don't really know why we cry when we're unhappy or hurt or sometimes even joyful.	ကျွန်တော်တို့ စိတ်မချမ်းသာ တဲ့အခါ သို့မဟုတ် ထိခိုက်နာကျင် တဲ့အခါ သို့မဟုတ် ပျော် တဲ့အခါ ဘာကြောင့် ငို သလဲဆိုတာ သိပ္ပံပညာရှင်တွေက တကယ်ကို မ သိ သေးပါဘူး ။

In our experiments, we followed the standard evaluation indicators. The accuracy was calculated as follows:

$$A = \frac{\text{The number of sentences correct classified under this sentiment}}{\text{The number of all the sentences in the sentiment}} \times 100\%. \tag{22}$$

3.3 Hyperparameter Setting and Training

For the network structure used here, ReLU was used as the activation function, and multiple sets of convolution kernels were used for training. Filter windows of sizes three, four, and five were used, and the number of convolution units per filter was 100. The number of units in the hidden layer was 300, the output layer was classified by softmax, and dropout was used in the training process to prevent overfitting. Finally, the random gradient descent algorithm was used to update the weights.

3.4 Experimental Results and Analysis

Experiment 1: The results of five-fold cross-validation.

To evaluate the effect of the proposed model, all the data in the experiment were equally divided into five parts: one part was selected as the test corpus, and the other four parts were used as the training corpus. The evaluation results of the model and the experimental results are listed in Table 3, where the experimental average accuracy is 73.72%, which is the experimental effect of the proposed model.

Table 3. Results of five-fold cross-validation

ID	Corpus allocation	Accuracy (%)
1	The first is the test, and the other four are the training	73.12
2	The second is the test, and the other four are the training	74.05
3	The third is the test, and the other four are the training	73.45
4	The fourth is the test, and the other four are the training	73.89
5	The fifth is the test, and the other four are the training	74.10
Average		73.72

Experiment 2: Traditional machine learning methods and deep learning models for sentiment classification on the same test set.

As shown in Table 4, BiLSTM neural networks were more accurate than single-layer LSTM, indicating that the use of contextual information and consideration of time series can yield better solutions to the problem of classifying sentiments in text.

The CNN neural network model was not as effective as the LSTM neural network on sentiment analysis. A comparison of the results shows that CNN can be used to analyze text information, in addition to its benefits in image processing. The fastText had the lowest accuracy, but its model was simple and training was fast.

Table 4. Experimental results of traditional machine learning methods and deep learning models on the Burmese sentiment classification

Model	Accuracy (%)
SVM	52.14
LR	54.76
Fasttext	56.33
CNN	64.11
LSTM	64.45
BiLSTM	67.44
Att-CNN-Trans	73.72

Experiment 3: Ablation study

In the ablation experiment, we used CNN, attention, and BiLSTM mechanisms to perform sentiment analysis on Burmese. The specific comparison results are shown in Fig. 3.

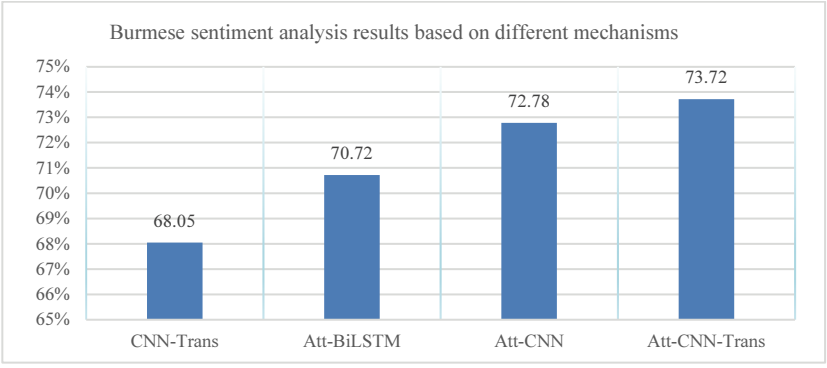


Fig. 3. Ablation study.

As shown in Fig. 3, the Att-BiLSTM neural network model was not as effective as the Att-CNN neural network when applied to sentiment analysis. A comparison of the results shows that CNN is better at capturing local features than BiLSTM. Local features are important for sentiment analysis of sentences. The Att-CNN-Trans model used here introduced the attention mechanism of the convolutional neural network, whereas the CNN-Trans model did not. The attention mechanism can be used to target sentiment-related features in the extracted textual features, which is why the model yielded better classification performance than the CNN on the validation and test sets. The Att-CNN-Trans mode used

transfer learning in the CNN-based attention mechanism, whereas the Att-CNN model did not use an English labeled set to pre-train the model and did not map Burmese sentences to English; the accuracy was also improved because the classification of sentiments in Burmese lacks a labeled corpus. For transfer learning, cross-lingual sentiment features are learned by sharing the neural network layer parameters of the English sentiment analysis model, which can assist in the classification of sentiments in Burmese.

4. Conclusion

To solve the problem whereby the current classification of sentiments in Burmese is hindered by a lack of labeled corpora, this study proposes a method for the analysis of sentiments in Burmese text based on transfer learning. Cross-lingual sentiment-related features were learned by sharing the parameters of the neural network layer of an English sentiment analysis model to classify sentiments in Burmese. The results of the experiments show that the proposed method can significantly improve sentiment classification in Burmese compared with the prevalent methods, achieving an accuracy of 73.72% in Burmese sentiment classification. In addition, we constructed 15,000 English-Burmese parallel sentence pairs, which provided data for related research on Burmese. In future work, we will consider more word vector methods to transfer Burmese words.

Acknowledgement

This work was supported by the Key Program of the National Natural Science Foundation of China (No. U21B2027 and 61732005); the National Natural Science Foundation of China (No. 61866019, 62166023, and 61972186), Yunnan Provincial Major Science and Technology Special Plan Projects (No. 202103AA080015 and 202002AD080001), the Program for Applied & Basic Research of Yunnan Province in Key Areas (No. 2019FA023), and the Candidates of the Young and Middle Aged Academic and Technical Leaders of Yunnan Province (No. 2019HB006).

References

[1] J. W. Watkins, "Burmese," *Journal of the International Phonetic Association*, vol. 31, no. 2, pp. 291-295, 2001.

[2] R. Mihalcea, C. Corley, and C. Strapparava, "Corpus-based and knowledge-based measures of text semantic similarity," in *Proceedings of the 21st National Conference of Artificial Intelligence (AAAI)*, Boston, MA, 2006, pp. 775-780.

[3] B. Ohana and B. Tierney, "Sentiment classification of reviews using SentiWordNet," *Proceedings of IT&T*, vol. 11, no. 6, pp. 75-78, 2009.

[4] V. Hatzivassiloglou and K. McKeown, "Predicting the semantic orientation of adjectives. In *35th annual meeting of the association for computational linguistics and 8th Conference of the European Chapter of the Association for Computational Linguistics*, Madrid, Spain, 1997, pp. 174-181.

[5] R. Xia, C. Wang, X. Dai, and T. Li, "Co-training for semi-supervised sentiment classification based on dual-view bags-of-words representation," in *Proceedings of the 53rd Annual Meeting of the Association for Computational Linguistics and the 7th International Joint Conference on Natural Language Processing (Volume 1: Long Papers)*, Beijing, China, 2015, pp. 1054-1063.

[6] Q. Chen, W. Li, Y. Lei, X. Liu, and Y. He, "Learning to adapt credible knowledge in cross-lingual sentiment analysis," in *Proceedings of the 53rd Annual Meeting of the Association for Computational Linguistics and the 7th International Joint Conference on Natural Language Processing (Volume 1: Long Papers)*, Beijing, China, 2015, pp. 419-429.

[7] P. Ruangkanokmas, T. Achalakul, and K. Akkarajitsakul, "Deep belief networks with feature selection for sentiment classification," in *Proceedings of 2016 7th International Conference on Intelligent Systems, Modelling and Simulation (ISMS)*, Bangkok, Thailand, 2016, pp. 9-14.

[8] X. Wan, "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*, Singapore, 2009, pp. 235-243.

[9] Z. Lin, X. Jin, X. Xu, W. Wang, X. Cheng, and Y. Wang, "A cross-lingual joint aspect/sentiment model for sentiment analysis," in *Proceedings of the 23rd ACM International Conference on Conference on Information and Knowledge Management*, Shanghai, China, 2014, pp. 1089-1098.

[10] X. Chen, Y. Sun, B. Athiwaratkun, C. Cardie, and K. Weinberger, "Adversarial deep averaging networks for cross-lingual sentiment classification," *Transactions of the Association for Computational Linguistics*, vol. 6, pp. 557-570, 2018.

[11] X. Zhou, X. Wan, and J. Xiao, "Cross-lingual sentiment classification with bilingual document representation learning," in *Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, Berlin, Germany, 2016, pp. 1403-1412.

[12] H. Yan, J. Dai, X. Qiu, and Z. Zhang, "A unified generative framework for aspect-based sentiment analysis," in *Proceedings of the 59th Annual Meeting of the Association for Computational Linguistics and the 11th International Joint Conference on Natural Language Processing (ACL/IJCNLP) (Volume 1: Long Papers)*, Virtual Event, 2021, pp. 2416-2429.

[13] L. Xu, Y. K. Chia, and L. Bing, "Learning span-level interactions for aspect sentiment triplet extraction," in *Proceedings of the 59th Annual Meeting of the Association for Computational Linguistics and the 11th International Joint Conference on Natural Language Processing (ACL/IJCNLP) (Volume 1: Long Papers)*, Virtual Event, 2021, pp. 4755-5766.

[14] R. Li, H. Chen, F. Feng, Z. Ma, X. Wang, and E. Hovy, "Dual graph convolutional networks for aspect-based sentiment analysis," in *Proceedings of the 59th Annual Meeting of the Association for Computational Linguistics and the 11th International Joint Conference on Natural Language Processing (Volume 1: Long Papers)*, Virtual Event, 2021, pp. 6319-6329.

[15] E. Kuriyozov, Y. Doval, and C. Gomez-Rodriguez, "Cross-lingual word embeddings for Turkic languages," in *Proceedings of the 12th Language Resources and Evaluation Conference (LREC)*, Marseille, France, 2020, pp. 4054-4062.

[16] D. Wang, B. Jing, C. Lu, J. Wu, G. Liu, C. Du, and F. Zhuang, "Coarse alignment of topic and sentiment: a unified model for cross-lingual sentiment classification," *IEEE Transactions on Neural Networks and Learning Systems*, vol. 32, no. 2, pp. 736-747, 2020.

[17] T. Mikolov, I. Sutskever, K. Chen, G. S. Corrado, and J. Dean, "Distributed representations of words and phrases and their compositionality," *Advances in Neural Information Processing Systems*, vol. 26, pp. 3111-3119, 2013.

[18] S. Li, L. Huang, J. Wang, and G. Zhou, "Semi-stacking for semi-supervised sentiment classification," in *Proceedings of the 53rd Annual Meeting of the Association for Computational Linguistics and the 7th International Joint Conference on Natural Language Processing (Volume 2: Short Papers)*, Beijing, China, 2015, pp. 27-31.

[19] D. Zeng, K. Liu, S. Lai, G. Zhou, and J. Zhao, "Relation classification via convolutional deep neural network," in *Proceedings of the 25th International Conference on Computational Linguistics: Technical Papers (COLING)*, Dublin, Ireland, 2014, pp. 2335-2344.

[20] C. Zhang, H. Zhang, J. Qiao, D. Yuan, and M. Zhang, "Deep transfer learning for intelligent cellular traffic prediction based on cross-domain big data," *IEEE Journal on Selected Areas in Communications*, vol. 37, no. 6, pp. 1389-1401, 2019.

[21] A. Joulin, E. Grave, P. Bojanowski, and T. Mikolov, "Bag of tricks for efficient text classification," in *Proceedings of the 15th Conference of the European Chapter of the Association for Computational Linguistics (EACL)*, Valencia, Spain, 2017, pp. 427-431.

[22] B. Rink and S. Harabagiu, "UTD: classifying semantic relations by combining lexical and semantic resources," in *Proceedings of the 5th International Workshop on Semantic Evaluation*, Uppsala, Sweden, 2010, pp. 256-259.

Cunli Mao **https://orcid.org/0000-0002-8289-6036**

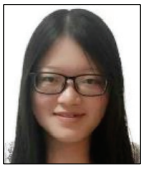
He received his Ph.D. degree from Kunming University of Science and Technology in 2014. He is currently a professor with the School of Information Engineering and Automation, Kunming University of Science and Technology, China. He has been an CCF member since 2011. His research interests include nature language processing, machine learning and information retrieval.

Zhibo Man **https://orcid.org/0000-0002-5843-3090**

He received B.S. degree in the School of Computer Science and Technology of Northeast Petroleum University in 2018 and is pursuing a master's degree at the School of Information Engineering and Automation of Kunming University of Science and Technology since September 2018. His current research interests include natural language processing and machine translation.

Zhengtao Yu **https://orcid.org/0000-0001-8952-8984**

In June 2017, he graduated from Beijing Institute of Technology with Ph.D. degree in computer science and Technology. He is currently a professor in the Department of Automation, Kunming University of Science and Technology, Yunnan. His research interests include multisource natural language processing, machine translation, etc.



Xia Wu **https://orcid.org/0000-0002-7646-0092**

She received B.S. degree in the School of Electrical, Energy and Power Engineering of Yangzhou University in 2017 and is pursuing a master's degree at the School of Information Engineering and Automation of Kunming University of Science and Technology since September 2017. Her current research interests include natural language processing and machine translation.

Haoyuan Liang **https://orcid.org/0000-0002-2000-1698**

He received B.S. degree in the School of Computer Science and Technology of Kunming University of Science and Technology in 2018 and is pursuing a master's degree at the School of Information Engineering and Automation of Kunming University of Science and Technology since September 2018. His current research interests include natural language processing.