A Multi-emotion Classification Method Based on BLSTM-MC in Code-Switching Text

Tingwei Wang[1], Xiaohua Yang[1](✉), Chunping Ouyang[1], Aodong Guo[2], Yongbin Liu[1], and Zhixing Li[3]

[1] School of Computer, University of South China, Hengyang 421001, China
xiaohua1963@usc.edu.cn
[2] College of Information Engineering, Xinhua University, Hefei 230031, China
[3] Chongqing University of Posts and Telecommunications, Chongqing 400065, China

Abstract. Most of the previous emotion classifications are based on binary or ternary classifications, and the final emotion classification results contain only one type of emotion. There is little research on multi-emotional coexistence, which has certain limitations on the restoration of human's true emotions. Aiming at these deficiencies, this paper proposes a Bidirectional Long-Short Term Memory Multiple Classifiers (BLSTM-MC) model to study the five classification problems in code-switching text, and obtains text contextual relations through BLSTM-MC model. It fully considers the relationship between different emotions in a single post, at the same time, the Attention mechanism is introduced to find the importance of different features and predict all emotions expressed by each post. The model achieved third place in all submissions in the conference NLP&&CC_task1 2018.

Keywords: Multiple emotion classification · Code-switching texts
Attention mechanism · BLSTM multiple classifiers

1 Introduction

Emotion classification refers to mapping the information to be classified into a predefined emotional category system. Which is widely used in recommendation and public opinion analysis. According to different emotional granularity, emotion classification can be divided into binary classification (subjective, objective), ternary classification (allegative, derogatory, neutral), or multivariate classification. Among them, multivariate classification can classify the emotions more close to human real emotions. In 2001, Parrott divided human social psychology into Happiness, Sadness, Anger, Fear, and Surprise in the results of research on human social psychological and emotional expression [1].

The code-switching text contains five kinds of emotions (happiness, sadness, anger, fear, surprise).Each post contains both English and Chinese. Emotions can be expressed individually or mixed both Chinese and English. Therefore, there are four forms for expression of emotions: none, Chinese, English and both. None means this post does not contain any corresponding emotions (E1). Chinese or English means that

M. Zhang et al. (Eds.): NLPCC 2018, LNAI 11109, pp. 190–199, 2018.
https://doi.org/10.1007/978-3-319-99501-4_16

emotions are expressed only by Chinese or English (E2, E3). Both mean that emotions are expressed in Chinese and English (E4). A single post may also contain multiple emotions (E5), so it is very different from monolingual and bilingual texts.

E1. 年底party 比较多啊，我在泰晤士小镇“美食玩家”大聚会正在准备中 ......

(At the end of the year, there are more parties, and I'm preparing for the "Gastro Gamer" party in Thames Small Town…)

E2. 心情极度不爽啊,为什么会这样, why? why? why?
(Extremely bad mood, why does this happen, why? why? why?)

E3. I'm so happy!虽然天空还飘着小雨~
(I'm so happy! Although the sky is still floating ~)

E4. 开学以来,浮躁的情绪。不安稳的心态。确实该自己检讨一下了。。。Sig

(I have been grumpy and emotional since the first day of school, unstable mindset too. It's really time to self-evaluate…sigh.)

E5. 做了四张英语周报后发现还有两张真崩溃 i hate english。
(After making four English Weekly Reports, I discovered that there are two other real collapses. i hate english)

In recent years, deep learning technology has gradually replaced traditional machine learning methods and has become the mainstream model of emotional classification [2]; Socher et al. [3] used the recursive neural network (RNN) model to perform emotional classification on the Film review data. Kim [4] uses convolution neural network (CNN) model to classify emotion. Literature [5] uses Long-Short Term Memory (LSTM) model to comment sentences into word sequences for emotion classification; Cheng Yu et al. [6] used emotion-based LSTM model to classify Chinese product reviews.

The above literature has studied the classification of emotion carefully, but there are two shortcomings: (1) most of the studies are based on the binary or ternary classification; (2) there is no consideration for the existence of a variety of emotions at the same time in a single post. Therefore, there is a certain limitation on the restoration of human true feelings. Aiming at these two points, this paper proposes a BLSTM-MC (Multiple classifiers) model for multiple emotional classification of code-switching texts. By creating multiple classifiers of BLSTM, the model is associated with different emotional semantic information, At the same time, the Attention mechanism is introduced to different words with different text weights. The experiments use the dataset provided evaluation by the conference NLP&&CC_task1 2018. the results show that, the proposed model get third place in all submission results.

2 Related Work

High-quality Word Embedding is one of the important factors for the deep learning model. The traditional document representation are mostly based on the Bag of Words (BOW) method. It discards the word order information and the resulting text has sparseness and high dimensionality. Mikolov, Benkiv, et al. [7] expressed the text through the neural network training word vector, which solved the above problems well. Using word2vec to represent text and combining deep learning models such as convolution neural networks (CNN) [4, 8, 9], recurrent neural networks (RNN) [10, 11] and so on, emotion classification can be achieved better results than traditional Machine learning methods.

When sentence-level semantic features are modeled by word vectors, sequence models such as RNNs are widely used in emotion classification because of sequence structures in sentences or documents. In 1990, Elman [12] proposed a recurrent neural network that keeps the nodes in the hidden layers connected, but RNNs have long-distance dependence and gradient disappearance problems. In 1999, the LSTM proposed by Gers and Schmidhuber et al. [13] solved these problems by interacting with the information of the memory cells through the design of three sophisticated gating units. Literature [5] proposed using LSTM to model comment sentences into word sequences for emotion classification. However, the LSTM training process will lead to the deviation of weights; the bidirectional LSTM integrates the context information through the convolutional layer, and connects the two LSTM networks with opposite timings to the same output at the same time to improve the accuracy of the model. Cheng Lu [6] used the bidirectional LSTM model based on the attention mechanism to do emotion classify Chinese of product reviews and achieved good results in both the two-category and three-category tasks. However, the emotional classification result contains only one kind of emotion, and the emotional semantic information is lose, resulting in a single emotional outcome.

To solve the above problems, this paper proposes the LSTM-MC model, constructing five BLSTM classifiers to integrate different emotional semantic information, fully excavate the phenomenon of user multiple emotion coexistence and introduce the Attention mechanism to express the importance of different features.

3 BLSTM-MC Model

The BLSTM-MC model (see Fig. 1). First, the model enhances context semantic information by creating five BLSTM classifiers and introducing Attention mechanisms, and gets deeper features, then returns all the emotional predictions of all posts by Softmax.

3.1 Word Embedding

This paper uses the Skip-gram model to predict the words in its context window using the current word. First use the training document to construct a vocabulary, and then perform a one-hot encoding on the word. The value of each dimension in the one-hot

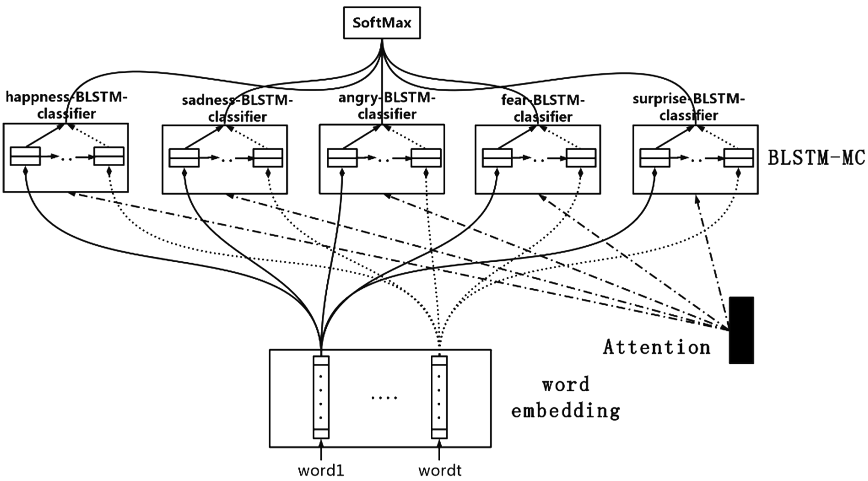

Fig. 1. BLSTM-MC model

encoding is only 0 or 1. The t-th word in the text is expressed as a vector of words: $w_t \in R^d$, where d is the dimension of the word vector. If the text length is T, the input text is represented as:

$$S = [w_1; w_2; \cdots; w_T] \in R^{T*d} \tag{1}$$

3.2 BLSTM Model

LSTM is applied to the processing of time series tasks (see Fig. 2). The BLSTM is constructed using the LSTM described by Zaremba et al. [12], and then the BLSTM model is used to integrate the context information to obtain text features.

Among them, f_t, i_t, o_t and c represent three kinds of gate mechanisms, the forget gate, the input gate, and the output gate, respectively, which control the read, write and lose operations of memory cell.

The input of this LSTM is a phrase representing the sequence $F = (F_1, \ldots, F_{l-w+1})$, the mechanism of which can be described by the following mapping.

Three gates of information flow input:

$$f_t = \sigma(w_f \cdot [h_{t-1}, x_t] + b_f) \tag{2}$$

$$i_t = \sigma(w_i \cdot [h_{t-1}, x_t] + b_i) \tag{3}$$

$$o_t = \sigma(w_o \cdot [h_{t-1}, x_t] + b_o) \tag{4}$$

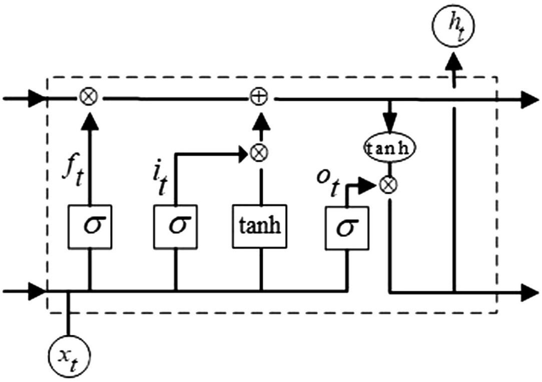

Fig. 2. LSTM structure diagram

Memory unit update:

$$\overline{c_t} = \tanh(w_c \cdot [h_{t-1}, x_t] + bc) \tag{5}$$

$$c_t = f_t * c_{t-1} + i_t * \overline{c_t} \tag{6}$$

Hidden layer unit update:

$$h_t = o_t * \tanh(c_t) \tag{7}$$

The timing t takes values $\{1, \ldots, l - w + 1\}, h_t, c_t \in R^n$, which are the hidden state and memory state at the time t, σ and tanh are the sigmoid and the hyperbolic tangent activation functions, respectively; i_t, f_t, o_t, $\overline{c_t}$ are input gates, forgetting gates, output gates, and new candidate memory states at time t, whose dimensions are equal to the hidden state dimension; * represents one by one element.

BLSTM includes forward $\overrightarrow{LSTM}$ and backward $\overrightarrow{LSTM}$ and these two parts share parameters. The forward LSTM reads F_1 to F_{l-w+1} sequentially from the phrase representation sequence, and the backward $\overrightarrow{LSTM}$ reads F_{l-w+1} to F_1 in turn, and its functions are shown in Eqs. (8) and (9).

$$\overrightarrow{h_t}, \overrightarrow{c_t} = \overrightarrow{LSTM}(F_t, \overrightarrow{h_{t-1}}, \overrightarrow{c_{t-1}}), t \in \{1, \cdots, l - w + 1\} \tag{8}$$

$$\overleftarrow{h_t}, \overleftarrow{c_t} = \overleftarrow{LSTM}(F_t, \overleftarrow{h_{t+1}}, \overleftarrow{c_{t+1}}), t \in \{l - w + 1, \cdots, 1\} \tag{9}$$

Among them, $\overrightarrow{h_0}, \overrightarrow{c_0}$ and $\overleftarrow{h_{l-w+2}}, \overleftarrow{c_{l-w+2}}$ are initialized to zero vectors. $\overrightarrow{h_t}$ is the phrase feature F_t fused with the above information representation, $\overleftarrow{h_t}$ is the phrase feature F_t fusion representation of the following information, and $h_t = \left[\overrightarrow{h_t}; \overleftarrow{h_t}\right]$ obtained by concatenating the two is the phrase representation of the fusion context information. Through the BLSTM layer, the resulting phrase sequence of fusion context information is represented by Eq. (10).

$$H = (h_1, \cdots, \mathrm{h}_{l-w+1}) \in R^{2\mathrm{n}*(l-w+1)} \tag{10}$$

3.3 Attention Mechanism

The core idea of the attention mechanism is to learn the weights of words in a word sequence to assign different attention to different content. The Attention mechanism has been widely used in image recognition [14], image annotation [15] and natural language processing [16]. In the Attention mechanism:

$$u_t = \tanh(w_w H_t + b_w) \tag{11}$$

$$a_t = soft\max(u_t^T, u_w) \tag{12}$$

$$v = \sum\nolimits_t a_t H_t \tag{13}$$

Among them, u_t is the hidden unit of $H_{t,}$ a_t is the attention vector, v is the output vector after processing by the Attention mechanism. u_w is the context vector, initialize randomly at the beginning of the experiment, and continues to improve during the learning process.

3.4 BLSTM-MC Emotion Classification

As shown in the BLSTM-MC layer in Fig. 1, the BLSTM classifier is constructed for five categories, respectively: Happiness classifier, Sadness classifier, Anger classifier, Fear classifier and Surprise classifier. In the Happiness classifier, the code-switching text expressed by the word vector is used as the input of the classifier, and the BLSTM will combine contextual contexts to capture textual features. The Attention mechanism gives the text feature weight to the deep feature vector v, and uses Softmax to regress the final distribution of the emotion prediction probability and P_i to express the text. The probability of an emotional i.

$$P_i = soft\max(w_c v + b) \tag{14}$$

For example, According to the probability, we can see whether the post belongs to Happiness. The other four classifiers have the same principle as Happiness. The final emotion prediction will fuse all the emotion prediction results of BLSTM-MC and correlate different emotion semantic information.

4 Experiment

4.1 Date Set

In order to verify the validity of the model, this paper uses the data set provided by the conference NLP&&CC_task1 2018, and the labels of data sets are divided into five categories, and is the form of Code-Switching Text. The training data set is divided into training set and validation set according to the 8:2 ratio. Of which there are 1824 posts in the happiness class, 1086 in sadness, 570 in anger, 648 in fear, 651 in surprise, a

Table 1. Training data set sample.

Posts	Happiness	Sadness	Anger	Fear	Surprise
翻译真的差很多耶,是这样吗? so lovely～ (Translation is really bad, is this? so lovely～)	T	F	F	F	T
做了四张英语周报后发现还有两张真崩溃 i hate english。 (After making four English Weekly Reports, I discovered that there are two other real collapses. i hate english.)	F	T	T	F	F

total of 4104 sentences with emotions. However, 675 of the 4104 Posts contain a variety of emotions, so the total emotion contained in the corpus is 4779, and the remaining 1896 unmarked emotions. From the data, it is found that the posting of sentimental coexistence accounts for 16.4% of the total number of emotional corpus, indicating that the situation of feeling coexistence occupies a large proportion in the task of emotional classification. The sample of the data set (see Table 1).

4.2 Data Preprocessing and Parameter Setting

Since the data set is a form of code-switching text, this paper uses Google translate API to translate the data set into Chinese text, and then preprocesses the data set. Then use the Skip-gram model of the word2vec tool to train the word vector, and the word vector parameter setting (see Table 2), finally gets a word vector list, the words that do not appear in the word vector list, random initialization of the word vector, and the dynamic update of the word vector during the training process.

Table 2. Word2vec parameter setting.

Model	Skip-gram
Window_size	7
Word vector dimension	100
Sampling	Negative sampling
Word frequency threshold	10

This model selects sigmoid as an activation function, the learning rate is 0.01, Adam is used as an optimizer, Dropout is used to prevent over fitting, and set value to 0.5, the cross entropy is used as the loss function, the batch size (batch_size) is 32, and the number of training times (n_epochs) is 1000.

4.3 Experiment and Analysis

F1-Score is compute calculated for each emotion separately, and compute the macro averaged F1 with all emotions, we use the conference NLP&&CC_task1 2018 scoring metric Marco-F1 as the evaluation standard. The BLSTM-MC model is compared with the model previously tested on the dataset, including BLSTM and the BLSTM based on Attention mechanism model, The results (see Table 3)

Table 3. Multi-model classification F1 value.

Model	Happiness	Sadness	Anger	Fear	Surprise	Marco-F1
BLSTM	0.691	0.413	0.543	0.164	0.256	0.413
BLSTM based on Attention mechanism	0.695	0.634	0.528	0.289	0.136	0.456
BLSTM-MC	**0.710**	**0.652**	**0.540**	**0.292**	**0.139**	**0.467**

1. BLSTM: Improved model of RNN proposed by Schmidhuber et al. [13].
2. BLSTM based on Attention mechanism: The emotion analysis model proposed in the literature [6] for Chinese product reviews.

It can be seen from Table 3 that: Compared with the BLSTM model based on Attention mechanism, the Marco-F1 value of BLSTM model is reduced by 43%, which indicates that after joining the Attention mechanism, it can not only capture the impact of input nodes on output nodes, but also enrich the semantic information and reduce the information loss in the process of feature extraction.

The Marco-F1 value of the BLSTM-MC model proposed in this paper is higher 54% than that of the BLSTM model's Marco-F1.which is higher 11% than the Macro-F1 value of the BLSTM based on the Attention mechanism. It is proved that the proposed BLSTM-MC model considers the relationship between text emotions well, solves the problem of the loss of emotional semantic information and improves the Marco-F1 value of the model.

The results of the BLSTM-MC model experiment were compared with those of two terms DeepIntell and DUTIR_938. DeepIntell is a team name, which achieved the best result in the conference NLP&&CC_Task1 2018, DUTIR_938 is a term name, which achieved second place, we also compared with the median results, which is named baseline in the table. The results of the experiment (see Table 4).

It can be seen from Table 4 that: The Marco-F1 value of the BLSTM-MC model used in this article is 0.467, only lower 48% than that of DeepIntell, lower 1% than that of DUTIR_938, but the values of Sadness and Fear are higher than DeepIntell and DUTIR_938, the values of each class are also significantly higher than those of the

Table 4. Participation in NLP&CC2018_Task1 teams performance comparison.

Team	Happiness	Sadness	Anger	Fear	Surprise	Marco-F1
DeepIntell	0.734	0.616	0.543	0.264	0.418	0.515
DUTIR_938	0.715	0.521	0.541	0.166	0.396	0.468
Our Team	**0.710**	**0.652**	**0.540**	**0.292**	**0.139**	**0.467**
Baseline	0.587	0.500	0.390	0.108	0.128	0.342

Baseline. Among them, the F1 values of Happiness, Anger and Surprise corresponding to emotion are slightly lower than those of the other two teams. This paper calls Google translation to translate them into monolingual texts, because the texts in social media are relatively informal and require high quality of translation.

5 Conclusions

This paper proposes a BLSTM-MC model for the multi-emotion classification of code-switching texts. The code-switching text is converted to the word vector with the Skip-gram model, and the context information is fused by the multi classifiers of BLSTM. Take full account of the fact that a single post has multiple emotion s at the same time, mine the importance of different features, and to predict all emotions expressed by each posts, In code-switching text, each post contains a variety of languages, such as Chinese and English, and Chinese also contains Cantonese and other forms of language, which is more challenging than monolingual or bilingual texts.

Acknowledgements. This research work is supported by National Natural Science Foundation of China (No. 61402220, No. 61502221), the Philosophy and Social Science Foundation of Hunan Province (No. 16YBA323), the Double First Class Construct Program of USC (2017SYL16), scientific and technological research program of Chongqing municipal education commission (No. KJ1500438), basic and frontier research project of Chongqing, China (No. cstc2015jcyjA40018).

References

1. Parrott, W.: Emotions in Social Psychology: Essential Readings, pp. 1–392. Psychology Press, Philadelphia (2001)
2. Tang, D., Qin, B., Liu, T.: Deep Learning for sentiment analysis: successful approaches and future challenges. Wiley Interdisc. Rev. Data Min. Knowl. Discov. **5**(6), 292–303 (2015)
3. Socher, R., Huval, B., Manning, C.D., et al.: Semantic compositionality through recursive matrix-vector spaces. In: Proceedings of the Joint Conference on Empirical Methods in Natural Language Processing and Computational Natural Language Learning, pp. 1201–1211. MIT Press, Cambridge (2012)
4. Kim, Y.: Convolutional neural networks for sentence classification (2014). arXiv:1408.5882
5. Zhu, X., Sobihani, P., Guo, H.: Long short-term memory over recursive structures. In: Proceedings of the International Conference on Machine Learning, pp. 1604–1612. ACM, New York (2015)

6. Cheng, J.: Research on two-dimensional LSTM model based on attention mechanism in sentiment classification of Chinese commodity reviews. Softw. Eng. **20**(11), 4–6 (2017)
7. Mikolov, T., Sutskever, I., Chen, K., et al.: Distributed representations of words and phrases and their compositionality. In: Proceedings of NIPS, pp. 3111–3119 (2013)
8. Collobert, R., Weston, J., Bottou, L., et al.: Natural language processing (Almost) from scratch. J. Mach. Learn. Res. **12**(1), 2493–2537 (2011)
9. Kalchbrenner, N., Grefenstette, E., Blunsom, P.: A convolutional neural network for modelling sentences, p. 1. Eprint Arxiv (2014)
10. Lai, S., Xu, L., Liu, K., et al.: Recurrent convolutional neural network for text classification. In: Proceedings of the Twenty-Ninth AAAI Conference on Artificial Intelligence, pp. 2267–2273 (2015)
11. Teng, Z., Vo, D.T., Zhang, Y.: Context-sensitive lexicon features for neural sentiment analysis. In: Proceedings of the 2016 Conference on Empirical Methods in Natural Language Processing, pp. 1629–1638 (2016)
12. Elman, J.L.: Finding structure in time. Cogn. Sci. **14**(2), 179–211 (1990)
13. Gers, F.A., Schmidhuber, J., Cummins, F.: Learning to forget: continual prediction with LSTM. Neural Comput. **12**(10), 2451–2471 (2000)
14. Mnih, V., Heess, N., Graves, A., et al.: Recurrent models of visual attention. In: Advances in Neural Information Processing Systems 27 (NIPS), pp. 2204–2212 (2014)
15. Xu, K., Ba, J., Kiros, R., et al.: Show, attend and tell: neural image caption generational with visual attention. In: Proceedings of the 32nd International Conference on Machine Learning (ICML), pp. 2048–2057 (2015)
16. Zheng, X., Ding, L., Wan, R.: Hierarchical BGRU model based on user and product attention mechanism. Comput. Eng. Appl., 27 May 2017
17. Zhang, Y., Jiang, Q.: Textual sentiment analysis based on two LSTM structures. Software **1**, 116–120 (2018)
18. Liu, Y., Ouyang, C., Li, J.: Ensemble method to joint inference for knowledge extraction. Expert Syst. Appl. **83**, 114–121 (2017)