YZU-NLP at EmoInt-2017: Determining Emotion Intensity Using a Bi-directional LSTM-CNN Model

Yuanye He[2,3,4], Liang-Chih Yu[1,3], K. Robert Lai[2,3] and Weiyi Liu[4]
[1]Department of Information Management, Yuan Ze University, Taoyuan, Taiwan
[2]Department of Computer Science & Engineering, Yuan Ze University, Taoyuan, Taiwan
[3]Innovation Center for Big Data and Digital Convergence, Yuan Ze University, Taoyuan, Taiwan
[4]School of Information Science and Engineering, Yunnan University, Kunming, P.R. China
Contact: lcyu@saturn.yzu.edu.tw

Abstract

The EmoInt-2017 task aims to determine a continuous numerical value representing the intensity to which an emotion is expressed in a tweet. Compared to classification tasks that identify 1 among *n* emotions for a tweet, the present task can provide more fine-grained (real-valued) sentiment analysis. This paper presents a system that uses a bi-directional LSTM-CNN model to complete the competition task. Combining bi-directional LSTM and CNN, the prediction process considers both global information in a tweet and local important information. The proposed method ranked sixth among twenty-one teams in terms of Pearson Correlation Coefficient.

1 Introduction

Categorical and dimensional representations are two major approaches to representing emotional states (Calvo and Kim, 2013; Gunes and Schuller, 2013). The categorical approach represents emotional states using several discrete classes such as positive and negative (binary) or Ekman's (1992) six basic emotions (anger, happiness, fear, sadness, disgust, and surprise), which have been successfully adopted in various sentiment applications (Pang and Lee 2008; Liu, 2012; Feldman, 2013). Based on this representation, application tasks focus on classification (i.e., identify 1 among *n* emotions for a given text). The dimensional approach provides a more fine-grained (real-valued) sentiment analysis. Knowing the intensity or degree to which an emotion is expressed in text is useful for more intelligent sentiment applications (Thelwall et al., 2012; Paltoglou et al., 2013; Malandrakis et al., 2013; Kiritchenko and Mohammad, 2016; Wang et al., 2016a; 2016b, Yu et al., 2016).

The EmoInt-2017 task (Mohammad and Bravo-Marquez, 2017b) seeks to automatically determine a continuous numerical value representing the intensity or degree to which an emotion is expressed in a tweet. That is, given a tweet and an emotion X, determine the intensity of emotion X felt by the speaker ranging from 0 (feeling the least amount of emotion X) to 1 (feeling the maximum amount of emotion X). The proposed system uses word embeddings (Mikolov et al., 2013a; 2013b) and a bi-directional LSTM-CNN model to complete the competition task.

Word embeddings can capture both semantic and syntactic information of selected words and provide a low dimensional and continuous vector representation for them. Convoluational neural network (CNN) (Kim, 2014; Kalchbrenner et al., 2014) is effective for extracting features in texts without considering the global information of that text. Long short-term memory (LSTM) (Tai et al., 2015) can capture long-distance dependencies by sequentially modeling texts across words. The proposed bi-directional LSTM-CNN model combines LSTM and CNN to model texts, encoding global information captured by LSTM in the most principal features extracted by CNN.

We first use word vectors to transform tweets into text matrices. The bi-directional LSTM is applied to these matrices to build new text matrices. CNN is applied to the output of the bi-directional LSTM to obtain text vectors for emotion intensity prediction. LSTM, CNN and their combination are described in detail in the following section.

Proceedings of the 8th Workshop on Computational Approaches to Subjectivity, Sentiment and Social Media Analysis, pages 238–242
Copenhagen, Denmark, September 7–11, 2017. ©2017 Association for Computational Linguistics

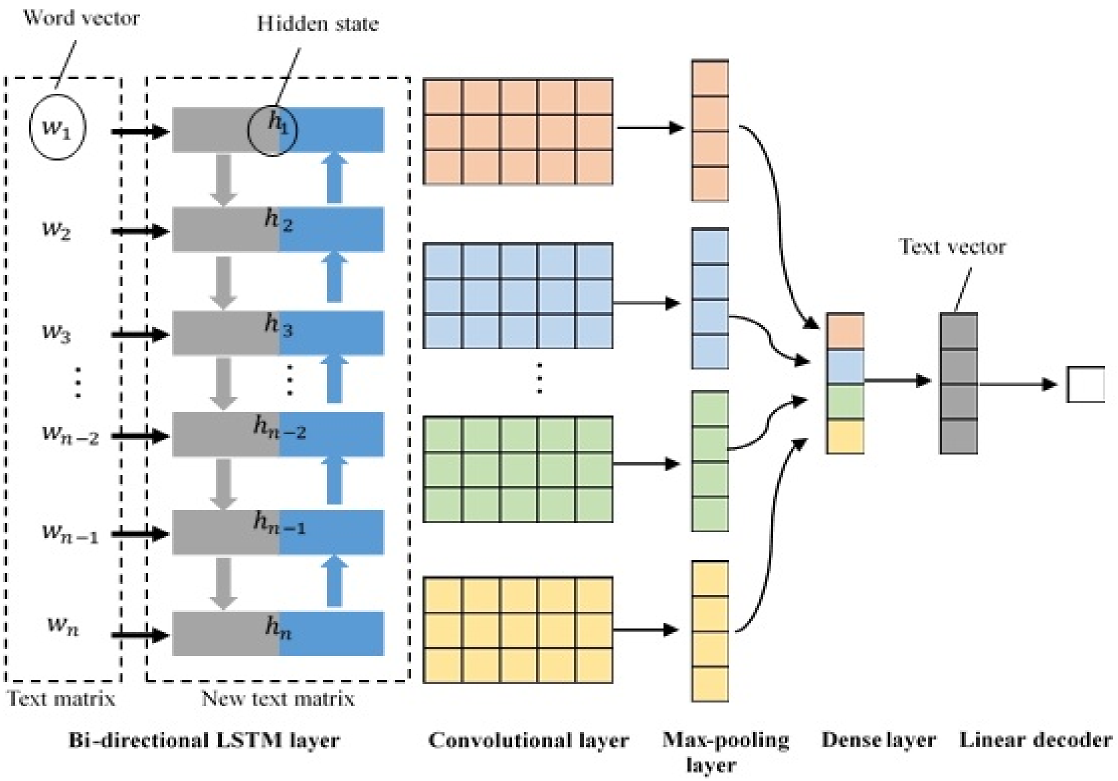

Figure 1: System architecture of the proposed Bi-directional LSTM-CNN model.

2 Bi-directional LSTM-CNN Model

Figure 1 shows the overall framework of the proposed Bi-directional LSTM-CNN model. For a given sentence, the system's input is a sentence matrix composed of the word vectors of all words and punctuation in the sentence. The sentence matrix is further transformed into a new sentence matrix by the Bi-directional LSTM model. The new sentence matrix is then sequentially passed through a convolutional layer and a max pooling layer for feature extraction. The extracted features are then passed through a dense layer to build a sentence vector for emotion intensity prediction.

2.1 Long Short-Term Memory (LSTM)

The LSTM (Hochreiter et al., 1997) uses a gating mechanism to track the state of sequences. There are three gates and a memory cell in the LSTM architecture. The LSTM transition functions are defined as follows:

$$\begin{aligned}
i_t &= \sigma(W^i x_t + U^i h_{t-1} + b^i) \\
f_t &= \sigma(W^f x_t + U^f h_{t-1} + b^f) \\
o_t &= \sigma(W^o x_t + U^o h_{t-1} + b^o) \\
g_t &= \tanh(W^g x_t + U^g h_{t-1} + b^g) \\
c_t &= f_t \circ c_{t-1} + i_t \circ g_t \\
h_t &= o_t \circ \tanh(c_t)
\end{aligned} \tag{1}$$

Here W^j, U^j, b^j for $j \in \{i, f, o, g\}$ are the parameters to be learned. h_t is the hidden state to be produced in time step t. The input vector x_t and the hidden state h_{t-1} are the input in time step t. $\sigma(\cdot)$ and $\tanh(\cdot)$ are the logistic sigmoid and hyperbolic tangent functions, $\circ$ is the element-wise multiplication operator, and i_t, f_t, o_t whose values are in (0, 1) are respectively called the input, forget and output gates. c_t is the internal memory cell. i_t controls how much new information will be stored in the current memory cell, f_t controls how much information from the old memory cell will be maintained and o_t controls how much information will be output as the hidden state in the current time step.

LSTM is theoretically powerful in language modelling due to its capability of representing a sentence or text with sequence order information. The last hidden state of the LSTM layer can be

	Anger	Fear	Joy	Sadness
Training set	857	1147	823	786
Development set	84	110	79	74
Test set	760	995	714	673
Max-length in training set	37	42	43	46

Table 1: Summary of data statistics.

regarded as the text representation containing the contextual information of the text. However, LSTM is a biased model, where the words in the tail of a text are more dominant than the words in the header. Thus, prediction performance could be reduced when it is used to capture the emotion intensity of a whole text, since the key components could appear anywhere in the text.

To avoid this problem, we maintain the hidden states of all time steps, and sequentially use the hidden state to replace the original word vector input. Then we build a new text matrix.

In addition, we replace the LSTM layer with a bi-directional LSTM layer consisting of two LSTMs running in parallel: one on the input sequence and the other on the reverse of the input sequence. At each time step, the hidden state of the bi-directional LSTM is the concatenation of the forward and backward hidden states. The hidden state can thus capture both past and future information.

2.2 Convolutional Neural Network (CNN)

The CNN architecture consists of a convolutional layer and a max pooling layer. The convolutional layer's input is the bi-directional LSTM layer's output which is a new text matrix. Once the new text matrix sequentially passes through the convolutional layer, the local *n*-gram features can be extracted.

The max-pooling layer subsamples the output of the convolutional layer. Pooling is conducted by maintaining the max value of the result of each filter. The max-pooling layer can reduce the dimension of the extracted feature vector and extract the local dependency to maintain the most important information for prediction.

The obtained vectors are then fed to a dense layer to build a text representation. Since emotion intensity is a continuous value, a linear decoder layer uses a linear regression to transform the text representation into a real value.

	LSTM	CNN	BiLSTM-CNN
LSTM hidden state size	200	-	64
Filter windows	-	2	3
Feature maps	-	128	128
Convolutional layer activation	-	ReLU	ReLU
Dense layer size	-	64	64
Dense layer activation	-	ReLU	ReLU
Dense layer dropout	-	0.3	-
Loss function	MSE	MSE	MSE
Optimizer	adam	adam	adam
Mini-batch size	10	10	10

Table 2: Hyper-parameters Used

3 Experiments and Evaluation

This section evaluates the performance of the proposed bi-directional LSTM-CNN model by submitting the results to the EmoInt-2017 task.

Dataset. The statistics of the official dataset (Mohammad and Bravo-Marquez, 2017a) used in this competition are summarized in Table 1. Each tweet was rated with a real-value (emotion intensity) in the range of (0, 1). Training, development and test datasets are provided for four emotions: joy, sadness, fear, and anger. We trained four models corresponding to four emotions using their respective training sets without their development sets. The anger, joy and fear models used the architecture of the proposed bi-directional LSTM-CNN model. To improve results, the sadness model used the architecture of CNN model which excludes the bi-directional LSTM layer shown in Fig.1. For word embeddings, we used GloVe pre-trained word vectors for Twitter (glove.twitter.27B) with 200 dimensions (Pennington et al., 2014).

Implementation details. The hyper-parameters of the network are chosen based on the performance on the development set. In our experiments, we set the length of all tweets in the training set to be

	Anger	Fear	Joy	Sadness	**Avg**
Pearson	0.666	0.677	0.658	0.709	**0.677**
Rank	5	8	6	5	**6**
Spearman	0.641	0.655	0.652	0.713	**0.665**
Rank	5	8	7	5	**6**

Table 3: Results of the proposed BiLSTM-CNN model.

	Anger	Fear	Joy	Sadness	**Avg**
CNN	0.645	0.662	0.617	0.709	**0.658**
LSTM	0.503	0.590	0.585	0.567	**0.561**
BiLSTM-CNN	0.666	0.677	0.658	0.706	**0.677**

Table 5: Comparative results of different methods.

the maximum length in the training set. A linear activation function is used in the output layer. Other hyper-parameters are presented in Table 2.

Evaluation metrics. The EmoInt-2017 task published the results for all participants using the Pearson and Spearman correlation coefficient.

Results. A total of twenty-one teams participated in the task. Table 3 shows the results of the proposed bi-directional LSTM-CNN model. Table 4 shows the results over the subset of the test data with a gold emotion intensity score greater than or equal to 0.5. Table 5 shows the experimental results for CNN, LSTM and their combineations after the release of test set ratings. LSTM used the last hidden state as the text vector, which caused the worse performance than CNN and BiLSTM-CNN. In addition, BiLSTM-CNN performed a little better than CNN and performed well for the subset with higher emotion intensity scores (>=0.5).

4 Conclusions

This study presents a deep learning approach to determine the emotion intensity of tweets. The proposed model combines long short-term memory networks and the convolutional neural networks to encode the global information captured by LSTM among the most principal features extracted by CNN. Experimental results show that the proposed method archived good performance. Future work will focus on other deep learning approaches such as the attention-based model and tree-LSTM to improve performance, and adopt an additional sentiment corpus to allow the system to capture more sentiment information.

	Anger	Fear	Joy	Sadness	**Avg**
Pearson	0.544	0.552	0.471	0.495	**0.516**
Rank	3	5	5	7	**4**
Spearman	0.528	0.520	0.460	0.493	**0.500**
Rank	3	5	5	8	**4**

Table 4: Results of the proposed BiLSTM-CNN model over a subset of the test data with a gold emotion intensity score greater than or equal to 0.5.

Acknowledgments

This work was supported by the Ministry of Science and Technology, Taiwan, ROC, under Grant No. MOST 105-2221-E-155-059-MY2 and MOST 105-2218-E-006-028. The authors would like to thank the anonymous reviewers and the area chairs for their constructive comments.

References

Rafael A. Calvo and Sunghwan Mac Kim. 2013. Emotions in text: dimensional and categorical models. *Computational Intelligence*, 29(3):527-543.

Paul Ekman. 1992. An argument for basic emotions. *Cognition and Emotion*, 6:169-200.

Ronen Feldman. 2013. Techniques and applications for sentiment analysis. *Communications of the ACM*, 56(4):82-89.

Hatice Gunes and Björn Schuller. 2013. Categorical and dimensional affect analysis in continuous input: Current trends and future directions. *Image and Vision Computing*, 31(2):120-136.

Sepp Hochreiter and Jürgen Schmidhuber. 1997. Long short-term memory. *Neural computation*, 9(8), 1735-1780.

Yoon Kim. 2014. Convolutional neural networks for sentence classification. In *Proceedings of the 2014 Conference on Empirical Methods on Natural*

Language Processing (EMNLP-14), pages 1746-1751.

Nal Kalchbrenner, Edward Grefenstette, and Phil Blunsom. 2014. A convolutional neural network for modelling sentences. In *Proceedings of the 52nd Annual Meeting of the Association for Computational Linguistics (ACL-14)*, pages 655-665.

Svetlana Kiritchenko and Saif M. Mohammad. 2016. The effect of negators, modals, and degree adverbs on sentiment composition. In *Proceedings of the 7th Workshop on Computational Approaches to Subjectivity, Sentiment and Social Media Analysis at NAACL-HLT 2016*, pages 43-52.

Bing Liu. 2012. Sentiment Analysis and Opinion Mining. Morgan & Claypool, Chicago, IL.

Nikos Malandrakis, Alexandros Potamianos, Elias Iosif, Shrikanth Narayanan. 2013. Distributional semantic models for affective text analysis. *IEEE Trans. Audio, Speech, and Language Processing*, 21(11): 2379-2392.

Tomas Mikolov, Kai Chen, Greg Corrado, and Jeffrey Dean. 2013a. Efficient estimation of word representations in vector space. In *Proceedings of International Conference on Learning Representations (ICLR-13): Workshop Track.*

Tomas Mikolov, Ilya Sutskever, Kai Chen, Greg Corrado, and Jeffrey Dean. 2013b. Distributed representations of words and phrases and their compositionality. In *Advances in Neural Information Processing Systems 26*, pages 3111-3119.

Saif M. Mohammad and Felipe Bravo-Marquez. 2017a. Emotion Intensities in Tweets. In *Proceedings of the sixth joint conference on lexical and computational semantics (*Sem).*

Saif M. Mohammad and Felipe Bravo-Marquez. 2017b. WASSA-2017 Shared Task on Emotion Intensity. In *Proceedings of the EMNLP 2017 Workshop on Computational Approaches to Subjectivity, Sentiment, and Social Media (WASSA).*

Georgios Paltoglou, Mathias Theunis, Arvid Kappas, and Mike Thelwall. 2013. Predicting emotional responses to long informal text. *IEEE Trans. Affective Computing*, 4(1):106-115.

Bo Pang and Lillian Lee. 2008. Opinion mining and sentiment analysis. *Foundations and trends in information retrieval*, 2(1-2):1-135.

Jeffrey Pennington, Richard Socher, and Christopher D. Manning. 2014. Glove: Global Vectors for Word Representation. In *Proceedings of the 2014 Conference on Empirical Methods on Natural Language Processing (EMNLP-14)*, pages 1532-1543.

Kai Sheng Tai, Richard Socher, and Christopher D. Manning. 2015. Improved semantic representations from tree-structured long short-term memory networks. In *Proceedings of the 52nd Annual Meeting of the Association for Computational Linguistics (ACL-14)*, pages 1556–1566.

Mike Thelwall, Kevan Buckley, and Georgios Paltoglou. 2012. Sentiment strength detection for the social web. *Journal of the Association for Information Science and Technology*, 63(1):163–173.

Jin Wang, Liang-Chih Yu, K. Robert Lai, and Xuejie Zhang. 2016a. Community-based weighted graph model for valence-arousal prediction of affective words. *IEEE/ACM Trans. Audio, Speech and Language Processing*, 24(11):1957-1968.

Jin Wang, Liang-Chih Yu, K. Robert Lai, and Xuejie Zhang. 2016b. Dimensional sentiment analysis using a regional CNN-LSTM model," in *Proceedings of the 53th Annual Meeting of the Association for Computational Linguistics (ACL-16)*, pages 225-230, 2016.

Liang-Chih Yu, Lung-Hao Lee, and Kam-Fai Wong. 2016. Overview of the IALP 2016 shared task on dimensional sentiment analysis for Chinese words. In *Proceedings of the 20th International Conference on Asian Language Processing (IALP-16)*, pages 156-160.