Integrating Grammatical Features into CNN Model for Emotion Classification

Thi-Thanh-Thuy Huynh
NLP-KD Lab, Faculty of Information Technology
Ton Duc Thang University
Ho Chi Minh city, Vietnam
httthuy.ktthntrang@khanhhoa.edu.vn

Anh-Cuong Le*
NLP-KD Lab, Faculty of Information Technology
Ton Duc Thang University
Ho Chi Minh city, Vietnam
leanhcuong@tdtu.edu.vn

***Abstract*—Emotion analysis is currently an attractive research topic in data mining and natural language processing. Along with the development of technology, people are also gradually evolving to post their emotional thinking on social media. Emotional information is useful for various aspects of business such as advertisement. Automatically classifying user emotions therefore becomes very important. In this paper we firstly formulate this problem under Convolutional Neural Network (CNN) framework. Actually language to express emotions is very diverse that make deep learning techniques such as CNN are ineffective in feature learning when the training data is not large enough. To solve this problem, we propose to use predefined grammatical patterns, which contain potential emotional information, to extract external features and integrate them into the CNN model. Our experiment are performed on two datasets, the ISEAR[1] (International Survey On Emotion Antecedents And Reactions) dataset and the Vietnamese emotion dataset. The experimental results show that the proposed model is very effective in comparison with previous studies.**

***Index Terms*—Emotion classification, convolutional neural networks (CNN), ISEAR dataset, grammatical features.**

I. INTRODUCTION

Human emotions can be seen through direct observation of behaviors, gestures, facial expressions, or speech and language. With the rise of social media on the Internet, people's information is coming from many directions with different effects on the reader's emotions. People express their emotions through social commentary, forums, etc. Access to opinions, public interest in a social issue has also attracted the attention of the business intelligence sector. Service industries increasingly want to grasp the user psychology to penetrate the market, refining their products to reach the consumer expectations. With this trend there is a lot of research on sentiment analysis from customer reviews, and with it, emotion analysis is becoming a hot topic in data mining and knowledge discovery research fields.

Actually, the root of the emotion classification derived from sentiment analysis. Emotions are the subject matter of many psychologists interested in research such as [4], [17]. Since the early years of the past decade, emotion classification has attracted the attention of a lot researches. Initial studies have proposed simple solutions to the classification of emotions using word feature, the semantic framework to classify emotions [2] or has applied the approach distant supervision, classifications of emotions that have multiple layers as a binary problem [22]. The combination of maximum entropy and word features, the topic-level feature is also used in subsequent studies [11], [18].

Recent studies have begun to use more complex research directions, including complex combinations such as incorporating traditional features based on the representation of the bag-of-word and based on new functional representations through the word emotions to create a new feature expression of the text, emotions are detected through the state analysis of the emotional expression or additional information representation of the traditional features, vocabulary using a combination of available information [1], [10], [21]. The lexical syntax has also been used in several studies associated with the ontology [5]. Hybrid Neural Network (HNN) combines the power of unattended learning methods and artificial neural networks to produce good results [13].

In recent years, deep learning models, especially Convolutional Neural Network (CNN) and Long Short-Term Memory (LSTM) have been applied in combination with vector expression and embedded words are increasingly effective in learning the semantic relevance of vocabulary in document. It is clear that vocabulary, semantics, and the relationship between them have been exploited more in recent years. In particular, the Word2Vec model and embedded words attract the attention of researchers, embedded words can capture both the semantic information and the syntax of the words [7], [13].

In this paper we firstly apply CNN model for the problem of emotion classification. We then exploit the external knowledge when integrating it into a deep learning model. In particular we will investigate the effectiveness of adding grammatical features into a CNN model for emotion classification problem. Our experiment will show that the proposed model's performance is much better than the original CNN model which does not included grammatical features. We evaluated the proposed model on the two datasets including the ISEAR dataset [19] and the Vietnamese emotion dataset constructed by ourself. The seven degrees of emotions are studied including joy,

*Corresponding author

[1]http://affective-sciences.org/home/research/materials-and-online-research/research-material/

fear, anger, sadness, disgust, shame and guilt. Notice that the ISEAR dataset has also been used by others studies such as [1], [11], [13], [15].

The remainder of this paper is organized as follows: Section 2 presents related studies. Section 3 present how to use CNN and Word Embeddings for the problem. Section 4 presents our proposed model that integrates grammatical features into a CNN model. Section 5 describes the experiments and the results of the proposed model. Finally the conclusion is presented in section.

II. Related Work

In studies of emotion analysis, the text is labeled based on polarization (positive, negative, or neutral), emotions (types of emotions), and intensity (emotional intensity according to emotional rank). In this paper we prefer referring studies related to methods for detecting different types of emotions. Jared Suttles and Nancy Ide [22] have tackled the inherent multilayered problem of emotion classification as a binary problem, replacing explicitly crafted labels by manipulating emoticons and the tag starts with the hashtag. They used Naive Bayes and Maximum Entropy for classification methods. Plaban Kr. Bhowmick et al. [2] has used Multi-Label K Nearest Neighbor (MLkNN) classification technique with a combination of three feature kinds including word features, polarity features, semantic frame features. [18] used Topic-level Maximum Entropy (TME) model. [11] has introduced a Multi-label Maximum Entropy (MME) model for user emotion classification over short text. They has developed a co-training algorithm for MME and use the L-BFGS algorithm for the generalized MME model. Chengxin Li et al. [10] has developed an emotion lexicon, a new feature representation of text which is named emotion vector. Yan Sun et al. [21] has provided a method for detecting emotions through adverb analysis in emotional expressions. [1] explored emotions using a combination of n-grams, part-of-speech information and the lexicon based features extracted using the uni-gram mixture model (UMM) based lexicon. The English-based ontology based on WordNet was also used as a tool to improve the accuracy of the emotional quotient for complex sentences. The verb concept is verifiable if they are really emotional verbs and are in place in ontology, as introduced in [5]. Kaitlyn Mulcrone [15] presented a general overview of the various approaches for detecting and classifying emotions including Latent Sentiment Analysis (LSA), Probabilistic Latent Sentiment Analysis (PLSA), Non-negative Matrix Factorization (NMF) , Valence-Arousal-Dominance (VAD), Majority Class Baseline (MCB). Weiyuan Li and Hua Xu [12] tried to infer and extract emotional reasons by entering knowledge and theory from other fields. Based on the theory that an event is an integral part of the emotion, based on the combination of multidisciplinary knowledge and careful investigation of microblog data. They used SVM and Support Vector Regression (SVR) for emotion classification.

The development of deep learning models combined with the vector representation of words has been increasingly effective in learning the semantic relevance of the vocabulary in the text. They have achieved the effect of task categorization [6], [8], [9], [14], sentiment analysis [20] and emotion analysis. Xiangsheng Li et al. [13] has proposed a novel model of semantically rich HNN which leverages unsupervised teaching models (e.g., Biterm Topic Model (BTM), Replicated Softmax Machine (RSM), and Word2vec) to incorporate semantic domain knowledge into the neural network to bootstrap its inference power and interpretability. Yuanye He et al. [7] has developed a Bi-directional LSTM-CNN Model using GloVe [16] pre-trained word vectors for Twitter (glove.twitter.27B) with 200 dimensions.

In this study, we will propose a new architecture which combines the basic CNN model and grammatical features containing information about emotion. The difference of our study with previous studies is that we have integrated new features (i.e. the grammatical features) into the CNN model for the task of emotion classification.

III. Using CNN for Emotion Classification

This section first introduces word embeddings and CNN model. We then present how to formulate the emotion classification problem under CNN model with word embeddings as input.

A. Word Embedding

Semantic information is obviously a useful factor for Natural Language Processing (NLP). Very early on, vector space models have been used in the semantic distribution of words. Many word embedding models have been developed. Word embbding has two famous models, word2vec and GloVe. Word embeddings is the result of representing a word as a vector of real numbers. The word vectors represent a semantics and their dimension is usually low, this enables to compute the similarity between words. Alternatively, words can be represented as vectors in a semantic space. This approach is also called as distributed representations of words, or word embeddings. Each dimension of word vectors represents a latent semantic and their dimensionality is usually low (in comparison with one-hot-vector). This technique not only expresses each word by a vector of real numbers but also shows a semantic relationship between those vectors. Nearly identical vectors will express words with similar meanings, this allows the calculation of similarities between words.

From word embeddings, each sentence can be represented as a matrix that represents each word as a row. As described above, each word is represented by a vector and to create a word matrix representing the sentence, we match the vectors representing each word into a matrix, each vector represents each word as a row. If we have the dimension of the word vector represented by d and the maximum length of a given text is s, then the dimension of the text matrix is (s x d). By using these vectors as input to the Deep Learning model, the model will be able to learn better and the ability to identify will also increase. We can observe how to convert a sentence into a matrix representing sentences as shown in Figure 1.

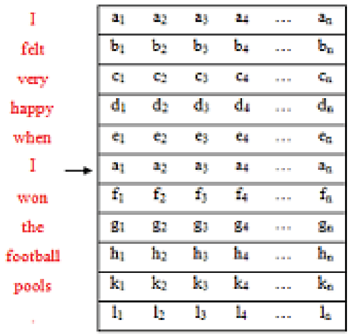

Fig. 1. Matrix representing sentence by word embedding with dimensions of the word vector represented d = n and the length of a sentence s = 11

In this study we used the vector set GloVe [16], it is pretrained word vectors for Common Crawl (glove.42B.300d) with 300 dimensions for word embeddings to use for English data. For Vietnamese emotion data, we used the GloVe v.1.2 toolkit[2] to train word embeddings also with 300 dimensions from 3GB of Vietnamese data.

B. Convolutional Neural Network (CNN) for Emotion Classification

CNN includes three layers are Convolutional layer, Pooling Layer, and Full-Connected Layer. The Convolution layer is the most important floor in the structure of CNN. The goal of the Convolution layer is to extract the characteristics of the input by using a slide filter and applies the scalar product. The output of Convolution is a called feature maps. We can use different kernels to form different feature maps as desired. The convolution operator will transform the input information into unique elements. The example can be seen in Figure 2.

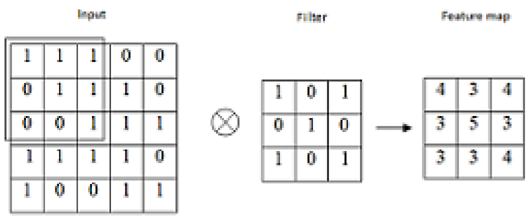

Fig. 2. Convolution layer with filter size = (3 x 3) and stride step = 1

We now formulate a CNN based text classification model. Suppose that we are given an input text which is a sequence of words $\{w_1, w_2, \ldots, w_n\}$. Each word w_i is represented by a vector of m dimension $(d_{i1}, d_{i2}, \ldots, d_{im})$. We then represent the input as a matrix $W = (a_{ij}) \in \mathbb{R}^{m \times n}$.

The filter matrix $E = (e_{ij})$ is usually designed with size of $k \times t$. If we use a stride step with size $s \times s$, then we obtain a new matrix via each filtering. Suppose that the filter starts at the cell $a_{rc} \in W$, then we calculate the convolutional operator to as follows:

$$b_{rc} = \sum_{i=1,j=1}^{i=k,j=t} a_{r+i-1,c+j-1} * e_{i,j} \tag{1}$$

to obtain value for the cell b_{rc} of the obtain new representation $B = b_{ij}$ of the input via the filter E.

From the matrices obtained by Convolutional operators we will transform each of them via an activation function. We use the Rectified Linear Unit (ReLU) as the activation function that wherever there is a negative value then the ReLU will convert this value to zero, as the follows:

$$ReLU(x) = \begin{cases} x & if \quad x > 0 \\ 0 & if \quad x \leq 0 \end{cases}$$

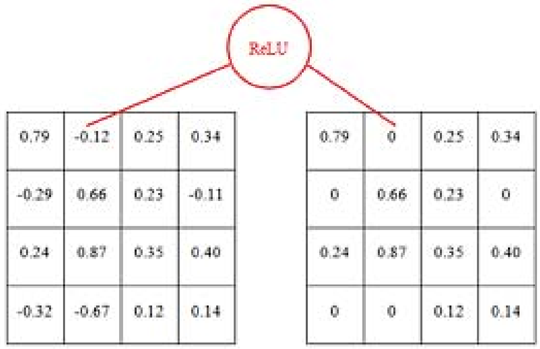

Fig. 3. ReLU

Next is Pooling layer. It reduces the size of the feature map by taking the average or largest value. Pooling works by sliding a window over the input and bringing the contents of the window into a composite function and the aggregate function by taking the largest value is still the most commonly used function in the Pooling layer. The output of the matrix after pooling can be seen in the example in Figure 4.

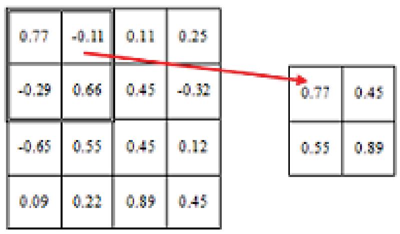

Fig. 4. Max Pooling with window size = (2 x 2) and stride step = 2

Finally, Fully Connected layer filters high-level representation of the input and converts them into votes. Fully connected is a way to connect neural layers in two layers together in which the back-connected layer pulses sufficiently with the neural layers in the previous layer. The last layer will use the

[2]https://nlp.stanford.edu/projects/glove/ - Prepared by Russell Stewart and Christopher Manning

Softmax function to classify the object based on the calculated vector of the previous layer. CNN is basically a multi-layered Convolution network. Each layer after the activation of the activation function will generate more abstract information for the next layer.

Note that like other NLP problems using CNN, the kernels usually slip through the lines of the word vector matrix. Therefore the kernel's width will be equal to the width of the matrix, the length may vary. Figure 5 shows how CNN works in this case.

We have a set of data to be classified. Each text in the dataset has been labeled specifically corresponding to the emotional label among the emotion set including *joy, fear, anger, sadness, disgust, shame, guilt*. Text will be processed into input for the CNN model. The output will be the predicted label of the input matrix.

IV. CNN Integrated with Grammatical Features

A. Build grammatical features

We propose a "lasting emotions" model, which is to add a basic pattern to some of the possible emotional signs, thereby increasing the emotional signature in the text, making the model predictably increased accuracy.

The input of the proposed model is still the vector matrix representing the word, but we add grammatical features to increase the sign of emotion. We find that some words appear more in some emotional layers, they can increase the sense of emotion in the text, making the emotion in the text more visible, more recognizable, such as verbs, adjectives, or adverbs. Table 1 shows a number of words that often appear in different emotion layers (meaning that the frequency of words in this emotion layer is higher than that of other emotion layers).

TABLE I
SOME WORDS OR PHRASES APPEAR IN THE EMOTION LAYER

Emotion layers	Words
Joy	long, first, love, study, receive, announced, pass, invited, joy, etc.
Fear	involved, serious, driving, knocked, alone, attack, afraid, fear, etc.
Anger	several, refused, regard, angry, unjustly, anger, etc.
Sadness	caused, close, short, break, decided, leave, sad, heard, sadness, etc.
Disgust	speak, drunk, torture, smell, disgust, etc.
Shame	asked, remember, relate, discovered, found, wet, laugh, prepare, ashamed, etc.
Guilt	hard, hurt, keep, responsible, late, promise, guilt, etc.

First we proceeded to parse and part-of-speech tagging (POS tagging) for the dataset. For English data sets, we used the Stanford Parser version 3.9.1 toolkit[3] [3], for Vietnamese emotion data, we used the JH-POS-TAG toolkit[4]. We have built a template for extracting grammatical features in the text using regular expressions. A regular expression is a special character segment - a template - used to match strings or sets of strings. The template we build will extract the grammatical characteristics that can cause the text to feel as we have described above. The template has been developed for English data[5] as follows:

```
\(RB [\w\d\s\-] + \) + |\(RBR [\w\d\s\-] + \) + |\(RBS [\w\d\s\-] + \) + |\(VB [\w\d\s\-] + \) + |\(VBN [\w\d\s\-] + \) + |\(VBD [\w\d\s\-] + \) + |\(VBG [\w\d\s\-] + \) + |\(VBZ [\w\d\s\-] + \) + |\(VBP [\w\d\s\-] + \) + |\(JJ [\w\d\s\-] + \) + |\(JJR [\w\d\s\-] + \) + |\(JJS [\w\d\s\-] + \)
```

And the template has been developed for Vietnamese emotion data[6] as follows:

```
[\w\d\s\-] + \/A + | [\w\d\s\-] + \/V + | [\w\d\s\-] + \/R
```

B. The Proposed Model's Architecture

Our proposed model is intuitively described in Figures 6 and 7. The idea for "lasting emotions" is expressed by the addition of words and signs that can increase the ability to sense emotions. If a word is likely to cause emotion to the text, its representative vector will be doubled, i.e. append a representative vector immediately after its representative vector. If not, a vector 0 with a dimension number equal to the dimension number of the representative vector will be appended after the representative vector of the word. As mentioned above, they can be verbs, adjectives, or adverbs, so we are using regular expressions to create a pattern for extracting words with features that are considered important in this model. At this point the number of dimensions of the vector will double (i.e. 300 x 2 = 600). For example, the text is "I felt very happy when I won the football pools." After parsing, we will obtain words that can cause emotions in the text such as "felt", "very", "happy", "won". Then proceed word embedding and integrate the extracted words into the word vector matrix as shown in Figure 6. Where vec_w is a vector representation of a word, $vec_{integrated}$ is the vector after the integration operation, where $[vec_w, vec_w]$ is a concatenate of two vectors. We have a built-in grammatical feature (GraFea) for data input as follows:

$$vec_{integrated} = \begin{cases} [vec_w, vec_w] & if \quad word \quad is \quad GraFea \\ [vec_w, 0] & if \quad word \quad isn't \quad GraFea \end{cases}$$

[3]https://nlp.stanford.edu/software/lex-parser.html

[4]Phan Van Hung, Le Anh Cuong, NLP-KD lab at Ton Duc Thang University

[5]http://www.surdeanu.info/mihai/teaching/ista555-fall13/readings/PennTreebankConstituents.html

[6]Labeling Guide - Nguyen Phuong Thai, Vu Luong, Nguyen Thi Minh Huyen, and Data Group SP 7.3 - VLSP

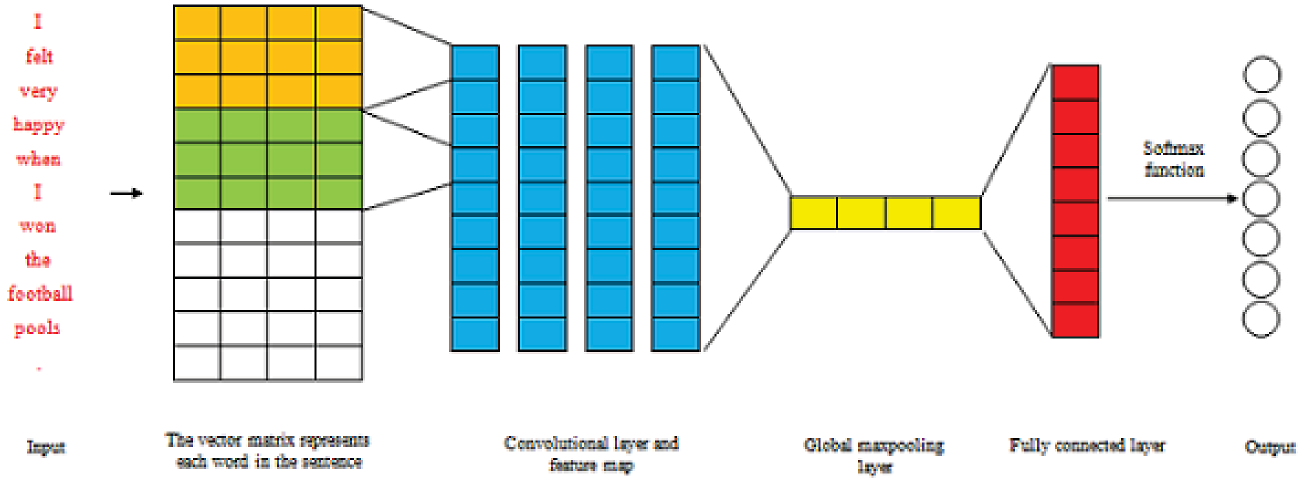

Fig. 5. CNN for Sentence classification

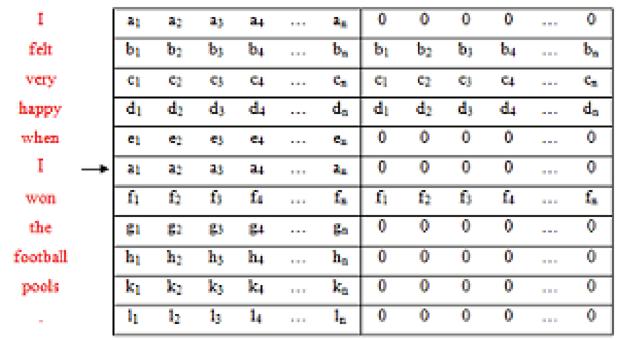

Fig. 6. Built-in grammatical feature in input data

V. Experiments

A. Datasets

First, the English dataset is an ISEAR dataset containing 7666 texts (one or more sentences) collected by a large group of psychologists around the world, collected by 3000 respondents in 37 countries on all 5 continents with multiple backgrounds, directed by Klaus R. Scherer and Harald Wallbott [19]. Students with psychology or other specializations are required to report situations, personal experiences in life, and respond to a number of events that they have experienced all 7 major emotions (joy, fear, anger, sadness, disgust, shame and guilt).

Secondly, the Vietnamese emotion dataset was not available, so we conducted a questionnaire survey of a group of nearly 500 high school students. We also relied on 7 major emotions similar to the ISEAR project, namely joy, fear, anger, sadness, disgust, shame and guilt, and collected 3437 texts (one or more sentences). Emotional texts are collected through the following questionnaire in Table 2. We split two datasets randomly, 80 percent for the training set and 20 percent for the test set. Some statistics of the data are shown in Table 3.

TABLE II
Question List

List memories or jobs that contain the following feelings	
Joy	________________
Fear	________________
Anger	________________
Sadness	________________
Disgust	________________
Shame	________________
Guilt	________________

TABLE III
Summary statistics of the datasets

Emotion layers	**Number of text**	
	ISEAR dataset	**Vietnamese dataset**
Joy	1094	491
Fear	1095	491
Anger	1096	491
Sadness	1096	491
Disgust	1096	491
Shame	1096	491
Guilt	1093	491
Total	***7666***	***3437***

B. Experimental implementation

We experimented with the basic CNN model and the improved CNN model on the datasets. First, data is preprocessed by word segmentation and vector representation of words. It

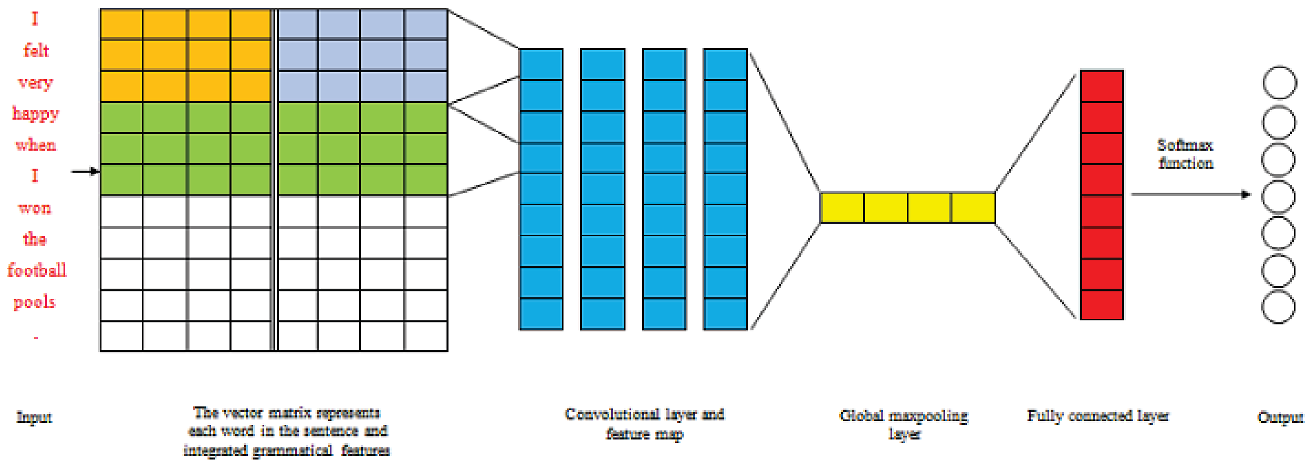

Fig. 7. ”Lasting emotions” model extends the built-in grammatical feature

then uses the syntax parsing tree tool and POS tagging to assign vocabulary labels to the text. Based on the parsed text and the lexical label, and based on the template generated from the regular expression (expressions that express the words that may have caused the emotion to the text), we have extracting words can cause emotion to the text. From there we continue to concate the vectors as mentioned in the previous section to create new inputs for the ”lasting emotions” model we have stated earlier.

In the basic CNN model, the input text is taken taking the vector representing each word, the input data of the model will be the magnetic vector matrix. For the improved model, the input data will be the matrix has been integrated grammatical feature. The size of each row in the input matrix of the improved CNN model will be double the size of each row in the input matrix of the basic CNN model. The length of the text will be padding for equal. The parameters of the models we tested are presented in Table 4.

TABLE IV
PARAMETERS USED IN THE STUDY

	ISEAR dataset		Vietnamese dataset	
	CNN	**CNN-grammatical feature**	**CNN**	**CNN-grammatical feature**
Max text length	60	60	25	25
Vector dimensions	300	600	300	600
Filter window	2	2	2	2
Feature map	256	256	256	256
Activation function	ReLU	ReLU	ReLU	ReLU
Dense layer size	128	128	128	128
Batch size	128	128	128	128

Figures 5 and 7 also show the basic CNN model and the improved CNN model we used in the study.

C. Results

We evaluated the models by using the Micro-Averaged F1 method. The results of our experiment are presented in Table 5 and the best result of our study and of previous studies using the ISEAR dataset are presented in Table 6. It can be observed that our models have outperformed the results of earlier studies on the ISEAR dataset.

The basic CNN model is better than previous studies, meanwhile the CNN model integrated with grammatical features gives the best results. The model is better than the basic CNN model, increasing by about 1.95% for the ISEAR dataset, and increasing by about 1.91% for the Vietnamese emotion dataset. Since then, we’ve found that our improved model is far better, delivering better results than the basic CNN model, and is far more effective than the approaches and methods before.

TABLE V
THE RESULTS OF OUR STUDY USING THE MICRO-AVERAGED F1

Models	**Datasets**	
	ISEAR	***Vietnamese emotion***
CNN	58.28%	67.65%
CNN-grammatical feature	**60.23%**	**69.56%**

TABLE VI
THE BEST RESULT OF OUR STUDY AND OF PREVIOUS STUDIES USING THE ISEAR DATASET

Models/Methods	**ISEAR Dataset**
CNN	58.28%
CNN-grammatical feature	**60.23%**
HNN-BTM [13]	51.21%
UMM based lexicon [1]	39.48%
Generalized MME [11]	54.86%
Valence-Arousal-Dominance [15]	37.2%

VI. CONCLUSION AND FUTURE WORK

Emotion classification is a research trend that is receiving much interest from market researchers and online entrepreneurs. One method, an accurate and effective classification model, is of great interest to researchers. In this study has done the work: first, learn new features can cause emotions in the text; second, extract these new features through tools such as parse tree or POS tagging, combining them with input data for the training model, to emotion classification models enhance the ability to identify and classify emotions. From that work we have a better model than previous research models. The subject has proposed an improved model in which significant grammatical features can enhance the sense of embedded text as a feature that supports the underlying CNN model. Our proposed model has achieved much higher accuracy on the classification of emotions in both the ISEAR dataset and the Vietnamese emotion dataset as compared to the underlying CNN model as well as previous studies.

The Vietnamese emotion dataset used by the subject is still limited. It only collects data on a particular type of object (high school students). Upcoming work may continue to collect for Vietnamese emotion dataset on different types of objects to enrich the content of the classification, using different ways of combining data between the original features and grammar to create new inputs for the training model such as the use of Vietnamese dictionaries for additional information. Emotional problems do not stop at deep learning models like CNN, but they can be combined with other deep learning models, such as LSTM.

ACKNOWLEDGEMENTS

This paper is supported by The Vietnam National Foundation for Science and Technology Development (NAFOSTED) under grant number 102.01-2014.22.

REFERENCES

[1] Anil Bandhakavi, Nirmalie Wiratunga, Deepak Padmanabhan and Stewart Massie, ”Lexicon based feature extraction for emotion text classification”, in Pattern Recognition Letters 93 (2017) 133-142, 2017.

[2] Plaban Kr. Bhowmick, Anupam Basu, Pabitra Mitra, and Abhishek Prasad, ”Multi-label Text Classification Approach for Sentence Level News Emotion Analysis”, in S. Chaudhury et al. (Eds.): PReMI 2009, LNCS 5909, pp. 261-266, 2009.

[3] Danqi Chen and Christopher D Manning, ”A Fast and Accurate Dependency Parser using Neural Networks”, in Proceedings of EMNLP 2014.

[4] Paul Ekman, ”An Argument for Basic Emotions”, COGNITION AND EMOTION, Vol 6, Issue 3-4, 169–200, 1992.

[5] R.Ezhilarasi and R.I.Minu, ”Automatic Emotion Recognition and Classification”, in Procedia Engineering 38 (2012) 21-26, 2012.

[6] Chinnappa Guggilla, Tristan Miller and Iryna Gurevych, ”CNN- and LSTM-based Claim Classification in Online User Comments”, in Proceedings of COLING 2016, the 26th International Conference on Computational Linguistics: Technical Papers, pages 2740-2751, Osaka, Japan, December 11-17 2016.

[7] Yuanye He, Liang-Chih Yu, K. Robert Lai and Weiyi Liu, ”YZU-NLP at EmoInt-2017: Determining Emotion Intensity Using a Bi-directional LSTM-CNN Model”, in Proceedings of the 8th Workshop on Computational Approaches to Subjectivity, Sentiment and Social Media Analysis, pages 238-242, Copenhagen, Denmark, September 7-11, 2017.

[8] Shiou Tian Hsu, Changsung Moon, Paul Jones and Nagiza F. Samatova, ”A Hybrid CNN-RNN Alignment Model for Phrase-Aware Sentence Classification”, in Proceedings of the 15th Conference of the European Chapter of the Association for Computational Linguistics: Volume 2, Short Papers, pages 443-449, Valencia, Spain, April 3-7, 2017.

[9] Yoon Kim, ”Convolutional Neural Networks for Sentence Classification”, in Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP), pages 1746-1751, Doha, Qatar, October 25-29, 2014.

[10] Chengxin Li, Huimin Wu and Qin Jin, ”Emotion Classification of Chinese Microblog Text via Fusion of BoW and eVector Feature Representations”, in C. Zong et al. (Eds.): NLPCC 2014, CCIS 496, pp. 217-228, 2014.

[11] Jun Li, Yanghui Rao, Fengmei Jin, Huijun Chen and Xiyun Xiang, ”Multi-label maximum entropy model for social emotion classification over short text”, in Neurocomputing 210 (2016) 247-256, 2016.

[12] Weiyuan Li and Hua Xu, ”Text-based emotion classification using emotion cause extraction”, in Expert Systems with Applications 41 (2014) 1742-1749, 2014.

[13] Xiangsheng Li, Yanghui Rao, Haoran Xie, Raymond Yiu Keung Lau, Senior Member, IEEE, Jian Yin and Fu Lee Wang, ”Bootstrapping Social Emotion Classification with Semantically Rich Hybrid Neural Networks”, in IEEE TRANSACTIONS ON AFFECTIVE COMPUTING, VOL. 8, NO. 4, OCTOBER-DECEMBER 2017.

[14] Mingbo Ma, Liang Huang, Bing Xiang and Bowen Zhou, ”Group Sparse CNNs for Question Classification with Answer Sets”, in Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics (Short Papers), pages 335-340, Vancouver, Canada, July 30 - August 4, 2017.

[15] Kaitlyn Mulcrone, ”Detecting Emotion in Text”, in UMM CSci Senior Seminar Conference, April 2012.

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

[17] Robert Plutchik and Henry Kellerman, ”Emotion: Theory, Research and Experience”, In: Theories of Emotion, vol. 1, Academic Press, New York, 1980.

[18] Yanghui Rao, Haoran Xie, Jun Li, Fengmei Jin, Fu Lee Wang and Qing Li, ”Social emotion classification of short text via topic-level maximum entropy model”, in Information & Management 53 (2016) 978-986, 2016.

[19] Klaus Scherer and Harald G. Wallbott, ”Evidence for universality and cultural variation of differential emotion response patterning”, in J. Personality Social Psychology, vol. 66, no. 2, pp. 310-328, 1994.

[20] Uladzimir Sidarenka, ”PotTS at SemEval-2016 Task 4: Sentiment Analysis of Twitter Using Character-level Convolutional Neural Networks.”, in Proceedings of SemEval-2016, pages 230-237, San Diego, California, June 16-17, 2016.

[21] Yan Sun, Changqin Quan, Xin Kang, Zuopeng Zhang and Fuji Ren, ”Customer emotion detection by emotion expression analysis on adverbs”, in Inf Technol Manag (2015) 16:303-311, 2015.

[22] Jared Suttles and Nancy Ide, ”Distant Supervision for Emotion Classification with Discrete Binary Values”, in A. Gelbukh (Ed.): CICLing 2013, Part II, LNCS 7817, pp. 121-136, 2013.