International Journal of Latest Research in Engineering and Technology (IJLRET)
ISSN: 2454-5031
www.ijlret.com || Volume 05 - Issue 11 || November 2019 || PP. 43-51

Research on Integrated Algorithm for Travel Question Text Classification

Jian Wang[1]
[1]*(Department of control and computer engineering, North China Electric Power University, China)*

Abstract: Question classification has become an important task in the research of automatic question answering but remain challenging for complexity of problem, which can be analyzed from syntax and semantic subspace generally. Consequently, an integrated algorithm that mixes multiple deep learning models is proposed to capture the syntactic information and deep semantic subspace information of travel-related textsin this paper. On the one hand, a word-level Convolutional Neural Network (CNN) can acquires low-level subspace structure information from a travel-related text sequence to improve the characterizing ability of text-level spatial structure information.On the other hand, a sentence-level Bi-directional Long Short-Term Memory Network (Bi-LSTM)obtains the global deep semantic and syntactic information of the text from the text sentence, and complements the low-level spatial structure information acquired by the word-level CNN. Furthermore, the multi-headed self-attention mechanism is used to integrate the both networksabove, and assign weights of attention to the captured features. Finally, accurate classification of texts in the tourism field is achieved through the SoftMax classifier at the end of the model. The experimental results indicate that the proposed integrated algorithm improves the accuracy of travel text classification, compared with the traditional deep learning model.

Keywords: Bidirectional long short-term memory network, Convolutional neural network, Deep abstract semantic information, Low-level subspace structure information, Multi-headed attention

I. INTRODUCTION

With the continuous growth of the social economy, the tourism industry has also ushered in unprecedented development, but various tourism problems have also increased [1]. As one of the main tasks of the question and answer system, question classification directly affects the quality of the question and answer system. How to quickly and effectively classify various types of tourism problems has become a research hotspot.

In recent years, with the continuous development of deep learning technology, deep learning technology has been widely used in problem text classification tasks, and has made great achievements [2]. Compared with the traditional machine learning method, the deep learning method can capture the deep semantic information of the text deeper, and solve the error caused by the artificial participation feature design, and also improve the classification accuracy [3].However, most of them are based on a single structure deep learning model, or a simple series of multiple models, so that when mining deep features of text, it will lose a lot of grammar and syntax information, and increase the use of redundant information. Therefore, we combine a multi-head attention mechanism to propose a classification method for tourism question texts that integrates multiple deep learning models. First, the word-level convolutional neural network obtains the low-level spatial structure information of the words to construct the grammatical information of the text. Sentence-level two-way long-term and short-term memory networks model the global semantics of travel text and syntactic information. Secondly, the two methods are jointly studied through the multi-attention mechanism, and different weights are assigned to enhance the utilization of key features of the travel question text.

In summary, the main contributions of this paper are as follows:

(1) We propose a tourism text classification method integrating multiple deep learning models to make information complement each other to improve the accuracy of tourism text classification.

(2) We design a Word Level Convolution Neural Network (WL-CNN) to capture the hierarchical structure and subspace structure information of words in the travel text, as well as the grammatical information of the text.

(3) We design a Sentence Level Long Short-Term Memory Network (SL-Bi-LSTM) to capture the deep semantic information and syntactic information contained in the sentence sequence in the travel text.

(4) We integrate WL-CNN and SL-Bi-LSTM using a multi-headed self-attention mechanism model, and assign attention weights to various features to highlight valuable information.

II. RELATED WORK

The early question text classification mainly uses shallow machine learning or simple deep learning technology to classify and recognize different types of question texts. For example, in order to achieve accurate detection of sports texts, Dalal et al. [4] proposed a semi-supervised machine learning method for automatic text classification, and achieved 87% classification accuracy. Salles T et al. [5] proposed an automatic text classification method for temporal weighting function (TWF) based on the original automatic text classification method to further improve the classification accuracy and classification efficiency of texts. As a time-aware classifier, the final experimental results show that the classification accuracy is improved by 17% compared to the traditional machine learning method (SVM). Mohammed Elarnaoty et al. [6] proposed a vector space model for matching Arabic texts to a textual theme, formalizing the problems in the text, constraining the features, and realizing the classification of the Arabic text of the text. Mohamed Goudjil et al. [7] designed large-scale high-dimensional texts, combined with manual annotation and multi-class support vector machines to screen different texts, and designed an active learning classification method to achieve accurate classification of texts. And reduce the amount of manual labeling. Huang Xianying, XieJin, et al. [8] extended the problem text classification theme by combining the fusion word vector and the BTM model, and finally used SVM to classify the questions. Due to the simple machine learning method, it is difficult to capture the deep abstract features of the question text, as well as being involved by humans. The impact of feature design, therefore, makes simple machine learning methods that do not effectively identify and process relatively complex question texts. With the continuous maturity of deep learning technology, it provides a new way of thinking about the classification of question texts. For example, Yu Bengong, Xu Qingtang et al. [9] proposed a multi-level attention convolution long-term memory model (MAC-LSTM) question classification model in order to solve the problem of collinearity of question text information to achieve an accurate classification of questions. Chenbin Li et al. [10] proposed a Bi-LSTM-CNN news text classification and recognition method, and improved classification accuracy, in order to solve the problem of text data sparsity, dimensional explosion and other problems and solve the limitations caused by manual participation in annotation. Zhang Dong, Li Shoushan et al. [11] proposed the introduction of word vector model combined with convolutional neural network to classify problems and improve the utilization of unlabeled samples. Guolong Liu et al. [12] proposed a bilingual text classification method for mixed deep learning using Bayesian Support Vector Machine (NB-SVM), word vector and long- and short-term memory recurrent neural network (RNNS). The deep semantic information is extracted from the bilingual text, which effectively improves the classification and recognition effect of the text. Arman S et al. [13] classify the emotional texts of the film, combine Word2vec with the decision tree, and introduce deep learning techniques to achieve accurate classification of the text. Zhang Qing et al. [14] achieved an accurate classification of the text of the neighborhood problem and achieved good experimental results.

III. INTEGRATED ALGORITHM

The syntactic and semantic information of the travel text is mainly determined by the composition of the text and the order of the sequences. On the one hand, the grammar of the travel question consists of multiple question keywords and some popular online words, and the words in the text sequence are modeled to form the low-level subspace structure information of the text sequence. On the other hand, the semantic information and syntactic information of the travel text comes from the text sequence itself. Therefore, we use the One-hot method to encode and model the travel question text to further capture the semantic information and syntactic information of the text. In order to better represent the travel text, we designed word-level convolutional neural networks and sentence-level two-way long-term and short-term memory networks to learn word vector features and text sentence semantic information. These two deep learning models are integrated through a multi-attention mechanism model while maintaining a complementary relationship from the syntactic and semantic aspects of the language. Therefore, in this chapter, we introduce the integrated travel text classification algorithm proposed in this paper. The integrated algorithm is mainly composed of three parts, namely WL-CNN, Sentence Level Bidirectional Long-Short term Memory Independently Recurrent Neural Network, (SL-Bi-LSTM-Ind-RNN) and Multi-Head Attention Mechanical (MH-AM). Different types of deep learning models obtain different deep semantic features and subspace structure information of the travel text, as well as deep structure information and global semantic information of the travel text. And through the attention to their classification, and to achieve the complement of the information obtained by different deep learning models.

3.1 Word-level convolutional neural network

Convolutional Neural Networks (CNN) have made great achievements in the fields of scene classification, pattern recognition and image segmentation [15]. The convolutional layer and the pooling layer in the deep convolutional neural network are not only robust to the noise contained in the text, but also model the spectrum or semantics of the text due to parameter sharing and local connectivity[16]. Therefore, the convolutional neural network is suitable for modeling the words in the travel text and further mining the subspace structure information of the words in the travel text.

When using convolutional neural networks to extract the relevant features of words in tourist texts, we will replace each of the words in the travel text with the predefined 1-of-n coded numbers in the convolutional layer [17], and the largest coded number is 225. When replacing a coded number with a word in a text, if the word contained in each sentence is not enough to replace all the numbers, the number is filled with the number 0, that is, the input vector in the input layer in the convolutional neural network is (Nan, 225, 225), then input the initial convolutional layer to encode the words in the travel text. The specific calculation steps are as follows:

(1) The encoded input vector is x_i, the output number is z_{xy}^l, and l is the number of hidden layers of the convolutional neural network, the output vector of the convolutional layer is y_i^l, and the number of filters is k . The word W satisfies the formula below by the initial convolutional layer:

$$v = conv2(\omega, x, "padding") + b,$$

where ω denotes the weight matrix.

The output is expressed as:

$$y = \eta(v),$$

(2) The convolution vocabulary feature vector z_{xy}^l for obtaining the travel text can be calculated by the multi-layer convolutional layer to satisfy the formula as follow:

$$z_{xy}^l = \sum_i^{k-1} \sum_j^{k-1} w_{ij}\, y_{(x+i)(y+i)}^l,$$

(3) The CNN feature vector of all words in the travel text can be obtained by the maximum pooling layer operation. The specific calculation process is as shown:

$$p_{xy}^l = max\, z_{xy}^{l-1} = max \sum_i^{k-1} \sum_j^{k-1} w_{ij}\, y_{(x+i)(y+i)}^l,$$

wherep_{xy}^l represents the output eigenvector of the max-pooling layer.

The WL-CNN uses a 5-layer convolution and pooling layer when extracting word features in tourist texts, and the number of neurons in each layer is 32-64-128-256-512. The size of the convolution core and the pooling core are 3x3 and 2x2, respectively.In order to further improve the ability of the WL-CNN to model words in the texts, the 2-layer of fully connected layer is connected to the 5-layer convolution pooling layer after, where the number of neurons in the fully connected layer is 512-64.Consequently,word-level subspace structure information is extracted in tourist texts.

Since the WL-CNN is mainly used to capture low-level subspace structure information and deep space features of words in tourist texts, it is necessary to perform related preprocessing on the input tourist texts to improve the characterization ability of the proposed algorithm and the anti-noise ability of the algorithm.

3.2 Sentence level two-way long-term memory network

The WL-CNN mainly obtains the word-level subspace structure information and local deep semantic information of the tourism texts, but ignores the global semantic information of the tourism text. Therefore, in order to further explore the global semantic features of the travel texts, the two-dimensional long-term and short-term memory network is introduced to model the time series of the travel text sequence. The specific calculation steps are as follows:

First, the travel text is represented as a one-hot code vector by the word embedding[18], and the temporal features in the code vector sequence of the travel texts are modeled by the forward and backward long-term and short-term memory layers.

Secondly, the relationship between the input sequences x is mapped by the bidirectional LSTM [19] [20], and the output sequence y is obtained by the activation state calculation of the neurons at different times. The specific calculation process is as shown:

$$i_t = \sigma(w_{ix} x_t + w_{i\alpha} \alpha_{t-1} + w_{ic} c_{t-1} + b_i),$$

Text sequence needs to forget the unnecessarytextual informationafter the sequence through the gate i_t at time t,which is calculated as shown:

$$f_t = \sigma(w_{fx} x_t + w_{f\alpha} \alpha_{t-1} + w_{fc} c_{t-1} + b_f)$$

where $f_t$denotes the text at $t$moment through the forgotten gate.The calculation process of the text through the memory unit is as follow:

$$c_t = f_t * c_{t-1} + i_t * g(\omega_{cx} x_t + \omega_{ca} \alpha_{t-1} + b_f)$$

The text is output through the output gate as shown:

$$o_t = \sigma(w_{ox} x_t + w_{o\alpha} \alpha_{t-1} + w_{oc} c_{t-1} + b_o),$$

The text is obtained by the calculation of the above formula to obtain the output of the text encoding vector. The specific calculation process is as shown in equations (9) and (10).

$$\alpha_t = o_t * h(c_t),$$
$$y_t = \pi(\omega_{ya}\alpha_t + b_y),$$

where$w$ denotes the weight vector,b denotes a bias vector,σ denotes an activation function,i denotes the input,f denotes the forgotten,o denotes the output,c denotes the memory unit;π denotes the total output activation function,g, h denote the input and output activation function ofthe memory unit.

3.3 Bullet attention mechanism

The Attention Mechanism [21][22] refers to obtaining the assigned weight by calculating the similarity $f(Q, K)$between Query and each Key. Then the weights of the assignments are weighted and summed with the corresponding Value, and the corresponding attention value is obtained, where Query, Key and Value are vectors, and Key=Value. Compared to the attention mechanism of a single structure, the bullish attention mechanism performs multiple linear mappings on each dimension of Queries, Keys, and Values, which stitches them together to obtain the final attention weight.

In this paper, we use the multi-head attention mechanism [23][24] to not only assign different weights to the low-level information of WL-CNN and the global semantic features of SL-Bi-LSTM, but also to further capture different sub-space structure information at different locations and deep global semantic information, and further portrays the internal structure information of travel text words and sentences. The calculation process of the bullish attention weight is as shown in equation (11).

$$Head_i = Attention(QW_i^Q, KW_i^K, VW_i^V),$$

where$W$ represents a linear variation parameter, and Queries, Keys, and Values correspond to different W value. Consequently, we obtain the weight value of the multi-head attention, and the specific calculation process is shownas:

$$MultiHead(Q, K, V) = Concat(Head_1, \dots, Head_l)W^o,$$

where l represents the number of parallel layers of multi-attention.

In summary, we use theWL-CNN and the SL-Bi-LSTM to learn the word sequence subspace vector and the deep semantic information of the sentence sequence.And both deep learning models are integrated through a multi-attention mechanism to achieve the complementary relationship between subspace information and sentence semantic information. Our network structure is shown as:

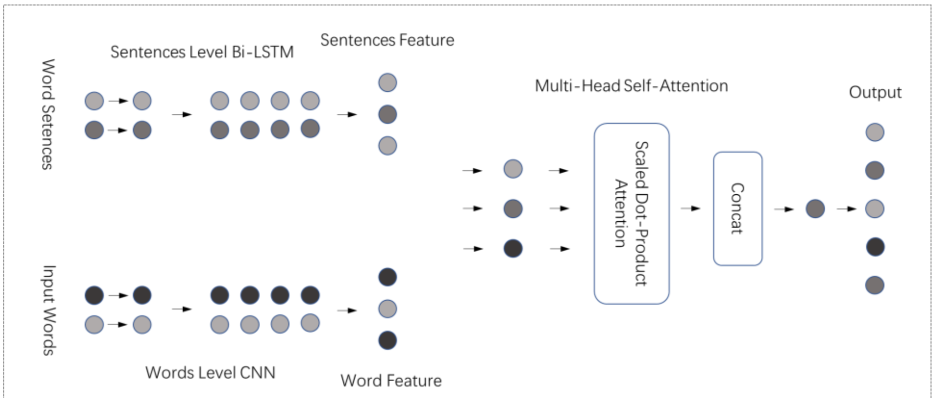

Figure 1 Network integration structure

As seen inFigure 1, our steps as follows:

(1) The travel text isinput into the SL-Bi-LSTM usinga form of sentences, and the global semantic information of the travel text and the deep semantic information of the texts are obtained bythe neural network, which is expressed as $x^{sentence}$ .The travel text is input into the WL-CNN using a form of words, and the low-level subspace structure information and the local deep semantic information of the words in the text are obtained through the calculation of the neural network, that is, expressed as x^{word} .

(2) In the above-mentioned feature input multi-head attention mechanism model, the multi-head attention mechanism is used to realize the weight matching of different hierarchical features obtained by different deep learning models.Finally, the attention weights obtained by each attention mechanism are spliced together.

(3) We use SoftMax for accurate classification.The labelvector is expressed as $z^{1\sim m}$, where $m$denotes the number of categories of classification.

IV. EXPERIMENTAL RESULTS AND ANALYSIS

In order to test the reliability and correctness of the designed integrated algorithm, this chapter will explain the experimental data set and algorithm related parameters in detail. To further verify the feasibility of the algorithm, five cross-validations were performed.

4.1 Experimental data

We use a custom benchmark dataset namely Tourism text, which mainly consist ofdata from major travel websites such as Ctrip, Tuniu, Mafengwo and ly.com, including 10,000 pieces of data of 6 types including travel location, time et al. Before the experiment, we performed pre-processing operations on the data set, namely screening, cleaning, etc., to reduce the corresponding error.

In order to verify the effectiveness of proposed algorithm, 60% of dataset were randomly selected as the training dataset, and the remaining were test datasets, and 10% of the training dataset was used to make cross-validation. The summary of dataset as shown:

Table I: Dataset statistics

Category	*Training*	*Test*	*Valid*	*AgaveLength*	*All Doc*
Location	1053	700	106	25	1753
Time	1200	795	120	35	1995
Object	1140	758	114	61	1898
Number	1042	694	105	74	1736
Description	1422	994	143	22	2416
Figures	123	81	13	2453	204

where the Stop Words of data have been removed in our experiments.

4.2 Related experimental methods

In order to verify the feasibility of the proposed travel text classification algorithm, we compared it with the most popular problem text classification methods.

(1) Convolutional Neural Network (CNN) [25]: This method mainly uses a random initialization embedding method for words, which are input into the convolutional neural network for classification. In order to obtain the ideal classification result in the question text classification, the same random initial word embedding method is used to obtain the word vector feature in text for comparation.

(2) Word2vec+LR[26]: This method uses word vector embedding to represent each type of text, which maps it to a low-dimensional space vector, and classifies it by logistic regression.

(3) Word Embeddings+SVM: This method[27] is similar toWord2vec+LR, which also embeds the problem text into a low-dimensional space vector and uses a shallow machine learning method to achieve classification.

(4) Long-Short Term Memory networks (LSTM) and Bidirectional Long-Short Term Memory networks (Bi-LSTM) [28]: Above two algorithms are to input the word text vector of the training into the network model according to the time series, to sequence the question text and to capture the deep semantic information, and to classify the Chinese question textaccurately.

(5) Self-Attention Networks (SAN): This algorithm combines word vector. It maps question text into a low-dimensional space vector and assigns weights to key features through the attention network to achieve accurate classification.

(6) Recurrent Neural Networks (RNN) and Independently Recurrent Neural Network (Ind-RNN): These approaches solve the problem of global semantic information and gradient disappearance neglected by the convolutional neural network to further improve the accuracy of the classification.

(7) Independently Recurrent Neural Networks (Ind-RNN): This method is mainly to solve the problem of gradient disappearance in the RNN algorithm and to effectively capture long-time modeling problems.

(8) CNN-LSTM: This method mainly connects CNN and LSTM in series. After using CNN to capture the space of the question text, itmodels time series to improve the accuracy of the text classification.

4.3 Algorithm parameters and experimental environment

To ensure the consistency of the model algorithm, we set the initialization parameters for the algorithm.The embedding vector size of the WL-CNN is set to 512, which issame as size of the SL-Bi-LSTM. Both algorithms have two convolution layers. The convolution kernel size is 3×3 and 1×1,the number of neurons is set to 512 and 128, the learning rate is set to 0.0001, the learning rate is set to 0.5,the number of bidirectional LSTM neurons is 256 and 64 respectively, and the learning rate is 0.02. When we integrate by Self-Attention, the $Head$ is set to 10, different weight ratios are assigned to them, and probability calculations are used to calculate the probability of occurrence of different categories to achieve the correct classification of travel problem texts. The specific parameters are shown:

Table II: Initialization parameters and experimental environment

Model	*Parameters*	
WL-CNN	Layers	2
	Filters size	1 and 3
SL-Bi-LSTM	Filters	512 and 128
	Learning rate	1e-4
	Dropout	0.5
	Filters	256 and 64
	Learning rate	2e-2
Self-Attention	K	10
Software	Python	3.6
	Framewrok	Keras, Numpy et al.
Hardware	GPU	GTX1060

We set the word vector dimension to 100 dimensions temporarily. In subsequent experiments, the final number of dimensionswill be determined as the precision changes.

4.4 Experimental result

In order to ensure the feasibility of the experimental results, we used three types of evaluation indicators to evaluate the experimental results, namely Accuracy, Loss, Time and F-Score. The F-Score is defined as:

$$F - score = \frac{2PR}{P + R}$$

where $P$is accuracy rate and $R$is recall rate.

4.4.1 The effect of mapping dimensions on algorithm performance

The mapping dimension of the word vector plays a vital role in the accuracy of the algorithm. Therefore, we change the number of different dimensions to test the classification of the text of the travel question. The experimental resultsare shown as:

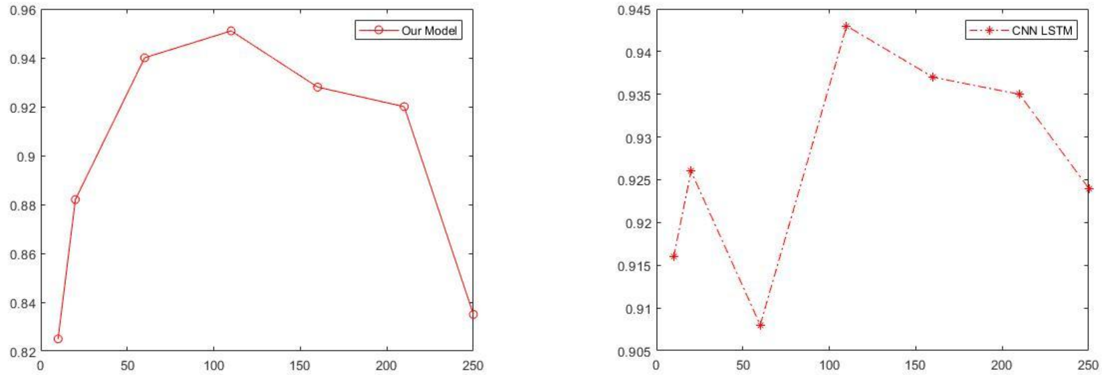

Figure 2. Experimental results of different dimension

Figure 2 shows test accuracy of Tourism text above and below the different word embedding dimensions. We can see that as the dimension increases, the test accuracy first increases, while the average precision stops increasingwhen the word embedding dimension is greater than 100. It is obvious that too low dimensional embedding may not map text well into low-dimensional space, while high-dimensional embedding may result in vector representations that are too sparse and do not improve classification performanceeffectively.Meanwhile, high-dimensional embedding may spend more training time.

4.4.2 Comparison of different classification algorithms

In order to further verify the effectiveness of the hybrid algorithm proposed in this paper, we compare the experimental results with different deep learning models and give the experimental results. The experimental results are shown as:

Table III: Results of multiple models

Models	*Accuracy*	*Loss*	*F-Score*	*Time*
Word2vec+LR[26]	0.9337	0.3867	0.9001	**15.13**
Embeddings[27]	0.9411	0.3719	0.9133	19.44
CNN[25]	0.9498	0.3697	0.9255	15.71
LSTM	0.9482	0.3154	0.9315	15.73
Bi-LSTM	0.9540	0.2867	0.9347	25.28
Self-Attention	0.9527	0.2836	0.9311	24.21
RNN	0.9472	0.2916	0.9160	40.06
Ind-RNN	0.9502	0.2231	0.3929	22.44
CNN-LSTM	0.9613	0.2948	0.9502	17.38
Our Models	**0.9766**	**0.2056**	**0.9605**	20.36

we also list the time required for the model algorithm to train once inTable 3.

From Table 3, the following conclusions can be drawn:

(1) Compared with other text classification algorithms, the proposed algorithm achieves the best accuracy and loss values, namely 0.9866 and 0.1277 respectively. Compared with CNN-LSTM in series, it increased by 0.0253 and decreased by 0.1671. Because this method not only captures the local and global semantic information of the travel text, but also assigns different weights to various features and achieves accurate classification. However, the algorithm is slightly insufficient in time efficiency. Because the integrated neural network is integrated, the total parameter amount of the algorithm increases, so it takes a long time.

(2) The loss value of the independent cyclic neural network (Ind-RNN) is 0.2231, because the internal neurons of the independent cyclic neural network are independent of each other, and cross-layer connections are realized between the layers, so compared with RNN, independent cyclic neural network solves the gradient disappearance problem very well.

(3) Shallow machine learning methods, namely LR and SVM. Although better classification results are obtained, they are suitable for small-scale data levels. When the data size increases, it is slightly insufficient. Because the shallow machine learning method does not capture the hidden information in the question text, the accuracy of the classification recognition is lower than that of the deep learning method.

(4) Bi-LSTM's Accuracy and Loss values are better than LSTM because Bi-LSTM uses both forward and backward LSTM to encode text, which better captures the context information of the travel question text.

V. CONCLUSION

The travel text classification algorithm proposed in this paper integrates multiple deep learning techniques, and captures the local part of the travel text through word-level convolutional neural network (WL-CNN) and sentence-level bidirectional long-term and short-term memory network (SL-Bi-LSTM). The global features also effectively describe the semantic information and context level relationship of the travel text, and finally use the self-attention mechanism to achieve the integration of the two, to further assign different probability weights to achieve the correct classification of travel texts.

We will proceed from the fine-grained classification direction to further characterize the keywords contained in the travel text to enhance the characterizing ability of the feature on the tourism problem text and improve the accuracy of classification recognition.

VI. ACKNOWLEDGEMENTS

First, I would like to extend my sincere gratitude to my supervisor, for his instructive advice and useful suggestion on my paper. I am deeply grateful of his help in the completion of this paper.I am also deeply indebted to all the other tutors and teachers in studies for their direct and indirect help to me.

Special thanks should go to my friends who have put considerable time and effort into their comments on the draft.Finally, I am indebted to my parents for their continuous support and encouragement.

REFERENCES

[1] X.S.Liu, and L.I.Kun. Design and Implementation of Tourism Events Text Classification System Based on BP-NN,*Computer & Modernization*, 2011, 7(7):192-194.

[2] J.Z. Liu, W.C.Chang, Y.Wu, Y.X. Wu, and Y.M. Yang. Deep Learning for Extreme Multi-label Text Classification, *International AcmSigir Conference on Research & Development in Information Retrieval. ACM*, 2017.

[3] B.Zhao, J.S.Feng, X.Wu, and S.C. Yan. A survey on deep learning-based fine-grained object classification and semantic segmentation,*International Journal of Automation and Computing*, 2017, 14(2):119-135.

[4] M.K. Dalal, and M.A. Zaveri. Automatic Text Classification of sports blog data, Computing, Communications&*Applications Conference. IEEE*, 2012. doi:10.1109/comcomap.2012.615480.

[5] T. Salles,L. Rocha,F.Mourão,M. Gonçalves,F.Viegas, and W.Meira. A Two-Stage Machine learning approach fortemporally-robust text classification[J]. *Information Systems*,2017,69.

[6] M.Elarnaoty, and A.Farghaly. Machine Learning Implementations in Arabic Text Classification,*Intelligent Natural Language Processing: Trend s and Applications*, 2018.

[7] M.Goudjil, M.Koudil, and M. Bedda. A NovelActive Learning Method Using SVM for TextClassification,*International Journal of Automation and Computing*, 2018, v.15(3):44-52.

[8] Z.Cheng, W. Wenxiu, and D. Ning. Improved short text classification method based on BTM topic features,*Computer Engineering and Applications*, 2016.

[9] B.Yu, Q.Xu, and P.Zhang. Question Classification Based on MAC-LSTM, 2*018 IEEE Third International Conference on Data Science in Cyberspace (DSC). IEEE Computer Society*, 2018.

[10] C.Li,G.Zhan,and Z. Li. News Text Classification Basedon Improved Bi-LSTM-CNN,*2018 9th International Conference on Information Technology in Medicine and Education (ITME). IEEE Computer Society*, 2018.

[11] S.Sun, C.Zhang, and N.Lu. A semi-supervised classification method based on transduction of labeled data,*IEEE Conference on Cybernetics & Intelligent Systems. IEEE*, 2004.

[12] G.Liu, X.Xu, and B.Deng. A hybrid method forbilingual text sentiment classification based on deeplearning,*2016 17th IEEE/ACIS International Conference on Software Engineering, Artificial Intelligence, Networking and Parallel/Distributed Computing (SNPD). IEEE*, 2016.

[13] A.S.Zharmagambetov, and A.A.Pak. Sentiment analysisof a document using deep learning approach anddecision trees,*Twelve International Conference on Electronics Computer & Computation. IEEE*,2016.

[14] L.Xiangdong, C.Huan, and D.Cong. Short-text Classification Based on HowN et and Domain Keyword Set Extension,*New Technology of Library and Information Service*, 2015.

[15] A.B.Duque, L.L.J. Santos, andD.Macêdo. SqueezedVery Deep Convolutional Neural Networks for TextClassification, 2019.

[16] C.Wigington, S.Stewart, and B.Davis. DataAugmentation for Recognition of HandwrittenWords and Lines Using a CNN-LSTM Network[C],*Iapr International Conference on Document Analysis & Recognition*, 2018.

[17] C.Adak, B.B.Chaudhuri, and M.Blumenstein. OfflineCursive Bengali Word Recognition Using CNNswith a Recurrent Model[C] *International Conference on Frontiers in Handwriting Recognition*.2017.

[18] J.Niu, M.Sun, S.Mo. Sentiment Analysis ofChinese Words Using Word Embedding andSentiment Morpheme Matching,*International Conference on Collaborative Computing: Networking, Applications and Worksharing. Springer, Cham*, 2017.

[19] Z.Yangsen, Z.Jia, and J.Yuru. A Text SentimentClassification Modeling Method Based onCoordinated CNN-LSTM-Attention Model,*Chinese Journal of Electronics*, 2019,28(1):124-130.

[20] L.Li, and Y.Jiang. Biomedical named entity recognitionbased on the two channels and sentence-levelreading control conditioned LSTM-CRF,*IEEE International Conference on Bioinformatics & Biomedicine*. 2017.

[21] L.Chen, G.Zeng, Q.Zhang, et al. Question answering over knowledgebase with attention-based LSTM networks and knowledge embeddings,*2017 IEEE 16th International Conference on Cognitive Informatics & Cognitive Computing (ICCI*CC). IEEE*, 2017.

[22] W.Lin, W.Yang, and L.Xue. Deep Attention-BasedSpatially Recursive Networks for Fine-GrainedVisual Recognition, *IEEE Transactions on Cybernetics*, 2018, PP (99):1-12.

[23] T.Alkhouli, G.Bretschner, and H.Ney. On the AlignmentProblem in Multi-Head Attention-Based NeuralMachine Translation, 2018.

[24] A.Vaswani, N.Shazeer, and N.Parmar. Attention IsAll You Need, 2017.

[25] C.Gu, M.Wu, and C.Zhang. Chinese SentenceClassification Based on Convolutional NeuralNetwork,*IOP Conference Series: Materials Science and Engineering,* 2017, 261:012008.

[26] T.Pranckevičius, and V.Marcinkevičius. Application ofLogistic Regression with part-of-the-speech taggingfor multi-class text classification,*Advances in Information, Electronic & Electrical Engineering*,2017.

[27] M.Goudjil, M.Koudil, and M.Bedda. A NovelActive Learning Method Using SVM for TextClassification,*International Journal of Automation and Computing,* 2018, v.15(3):1-9.

[28] C.H.Shih, B.C.Yan, and S.H.Liu. InvestigatingSiamese LSTM networks for text categorization,*Asia-pacific Signal & Information Processing Association Summit & Conference*,2018.