Sub-word attention mechanism and ensemble learning-based semantic annotation for heterogeneous networks

Liang Zhang*, Zhaobin Liu, Jinxiang Li, Gang Liu, Yuanfeng Yang, Yi Jin and Xu Zhang

School of Computer Engineering,
Suzhou Vocational University,
No. 106, Zhineng Str., International Education Park,
Suzhou 215104, China
Email: zhangl@jssvc.edu.cn
Email: liuzhaobin@jssvc.edu.cn
Email: ljx@jssvc.edu.cn
Email: liugang@jssvc.edu.cn
Email: yangyuanfeng@jssvc.edu.cn
Email: jinyi@jssvc.edu.cn
Email: zhangxu@jssvc.edu.cn
*Corresponding author

Abstract: The sensing device and wireless sensor networks (WSN) can provide information to the application of Internet of Things (IoT), but data from different types of devices present significant polyphyly and heterogeneity, which poses challenges to the collaboration and interaction of information resources in IoT application and service. Especially when the device uses Chinese characters for information representation and there is an Out-of-Vocabulary (OOV) problem, it will make this work more challenging. This paper introduces an ensemble learning model based on sub-word attention mechanism and bidirectional long short-term memory model (SWAT-Bi-LSTM) which can provide an internal structural attention ability of Chinese characters. The boosting strategy and gradient boosting decision tree (GDBT) integration scheme is adopted to complete the final integrated output. The experimental results show that the proposed method can effectively improve the accuracy of sentiment analysis, and the integrated learning model can further improve the accuracy and stability.

Keywords: sentiment analysis; sub-word units; Bi-LSTM; ensemble learning; WSN; wireless sensor networks.

Reference to this paper should be made as follows: Zhang, L., Liu, Z., Li, J., Liu, G., Yang, Y., Jin, Y. and Zhang, X. (2020) 'Sub-word attention mechanism and ensemble learning-based semantic annotation for heterogeneous networks', *Int. J. Wireless and Mobile Computing*, Vol. 18, No. 1, pp.51–58.

Biographical notes: Liang Zhang received his BS and MS degrees in Computer Science and Technology both from Soochow University, Suzhou, China, in 2005 and 2008, respectively. He is currently a Lecturer of Suzhou Vocational University. He is a Member of the CCF and the ACM. His research interests include artificial intelligence, wireless sensor networks and natural language processing.

Zhaobin Liu received his MS degree of Computer Science and Technology from Xian Jiaotong University. He is currently a Professor in the School of Computer Engineering, Suzhou Vocational University. He is a Member of the CCF and the ACM. His research interests include computer application and pervasive computing.

Jinxiang Li received his BS and MS degrees from Shaanxi University of Technology and Beijing University of Posts and Telecommunications, respectively. He is currently a Professor in the School of Computer Engineering, Suzhou Vocational University. His research interests are in computer application and artificial intelligence.

Gang Liu received his BS and PhD degrees in Computer Science and Technology both from Nanjing University of Science and Technology, Nanjing, China, in 2008 and 2013, respectively. He is currently a Lecturer of Suzhou Vocational University. His research interest includes network information security and wireless sensor networks.

Yuanfeng Yang received his PhD degree in Computer Science and Technology from Soochow University, Suzhou, China. He is currently an Associate Professor in the School of Computer Engineering, Suzhou Vocational University. His research interest includes artificial intelligence and image processing.

Yi Jin received her BS and MS degrees of Computer Science and Technology from Beijing Institute of Technology and Soochow University in 2003 and 2007, respectively. She is currently an Associate Professor in the School of Computer Engineering, Suzhou Vocational University. Her research interests include computer application and artificial intelligence.

Xu Zhang received his Doctor degree in School of Computer and Information, Hefei University of Technology in 2016. He is currently a Lecturer of Suzhou Vocational University. His research interests include artificial intelligence, computer vision and image processing.

1 Introduction

Adding semantic information to the framework of Internet of Things (IoT) or Wireless Sensor Networks (WSN) will make the interactions among machines, devices and people more convenient, because it can transform sensor information or users' requests in natural language to commands which can be recognised by computers. In the field of IoT, it can also improve machine interpretability by adding semantic information (Selbage et al., 2006). Researchers have already proposed some impressive work in the field of Semantic Web of Things. Terziyan et al. (2010) designed a semantic middleware that combines the agent and semantic technology, through which it can integrate a variety of applications and solve the interoperability problem among various heterogeneous IoT devices in intelligent transportation. Wang et al. (2013) designed a lightweight semantic description model for knowledge representation in the field of IoT. Zhang and Hu (2018) used the semantic information for lightweight description of sensors to help finding and managing the sensor data in complicated networks. Liu et al. (2017) leveraged semantic technology to solve the IoT information interoperability problem among scattered, hierarchical and heterogeneous devices, and proposed an automatic labelling method for equipment. In their method, algorithms such as information extraction, text classification, semantic label selection and information fusion etc. are used to achieving automatic labelling of equipment and devices in IoT and WSNs.

However, the algorithm designed for Latin hasn't considered the characteristics of Chinese characters, which cannot effectively use the unique morphological features of Chinese characters, such as strokes, radicals and other sub-word level structural information. In the complex and giant environment of IoT, it is almost impossible to label and train many different types of devices. This raises new requirements and poses great challenges for the identification of Chinese characters under semi-supervised condition, especially for the identification of Out-of-Vocabulary (OOV) words.

This paper proposes a sub-word attention mechanism (SWAT-Bi-LSTM) and an ensemble learning-based semantic annotation method for heterogeneous networks. By leveraging the unique morphological characteristics of Chines characters, we combine the internal sub-word structural information (such as strokes and radicals) which does not contain Latin to help understand the meaning of Chinese text. Furthermore, an integrated learning model based on the bagging method was adopted, which can further improve the precision and stability of the algorithm.

The rest of this paper is organised as follows: Section 2 outlines the sentiment analysis, LSTM, Bi-LSTM and attention mechanism which are the basis of our work. An integrated learning model based on the morphology of the Chinese characters and Bi-LSTM is proposed in Section 3. It consists of three main parts which is a word embedding that integrates the morphological information, a highway network and Bi-LSTM model and a scaled dot-product attention mechanism. Simulation results are provided in Section 4 to evaluate the performance of the proposed algorithm. Section 5 concludes the paper.

2 Background

2.1 *Sentiment analysis*

Sentiment analysis (or sentiment mining) refers to the process to analyse and process the text with sentiment, which was initially proposed by Nasukawa (2003). Its standard definition is: sentiment analysis is the computation and research of the opinion, sentiment, emotions and attitude of an entity (Guan et al., 2017). Sentiment analysis has very important significance in various application fields, such as evaluation of filtering and classification, opinion mining, user classification and clustering, and network public opinion prediction.

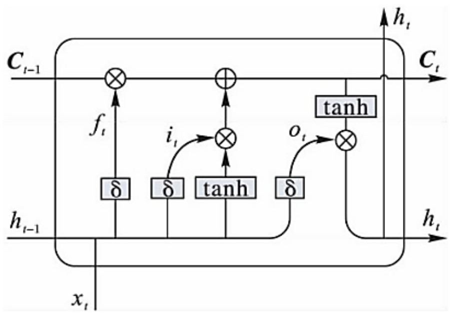
Sentiment analysis based on text (natural language) can generally be divided into three types, the sentiment dictionary-based method, the machine learning-based method and the deep learning-based method. The "sentiment dictionary-based method" is a classic unsupervised learning method, such as the method proposed by Kim and Hovy (2006, 2004), but dependency on sentiment dictionary has become the barrier in the application and development of such methods. The machine learning-based method treats text sentiment analysis as text classification problem, which was initially proposed by Pang et al. (2002). In the film criticism data set, they utilise the three classification methods of Naive Bayes (NB), Support

Vector Machine (SVM) and Maximum Entropy (ME) to determine the sentiment polarity of text. Later, many researchers studied the sentiment analysis problem based on the traditional machine learning method. For example, Wan et al. proposed the text classification method based on KNN and SVM (Wan et al., 2012); through many comparison experiments, Wang and Manning (2012) find that the SVM-based method (or its simple variation) can provide great results in text sentiment analysis. However, the traditional machine learning method only has good performance in a specific field, which has the main defects of poor generalisation ability and low fitting precision. The "deep learning-based method" can well prevent the weakness of the two methods discussed above. At the beginning, it was used in machine vision and speech recognition, and significant success was achieved; in recent years, it has also attracted more attention in the research work of Natural Language Processing (NLP) and sentiment analysis (Bengio et al., 2003). Mikolov et al. (2013), from Google Inc. (Mikolov et al., 2013a; 2013b) proposed Word2vec (Skip-gram and CBOW model training method) and achieved great results. This has promoted the eruption of deep learning technology in the fields of natural language processing and sentiment analysis. The subsequent works include: Kim (2014) used Convolutional Neural Network (CNN) to process the sentiment classification task (i.e., TEXTCNN); Irsoy and Cardie (2014) and Cho et al. (2014) used Recurrent Neural Networks (RNNs) for modelling of sentences to conduct text classification; etc.

2.2 LSTM and Bi-LSTM

The Long Short-Term Memory (LSTM) model was proposed by Hochreiter and Schmidhube (1997). As a special RNN model, it can capture the long-term dependency in sentence, so that the sentiment of text can be better understood in an integral way. Its memory unit has the structure as shown in Figure 1.

Figure 1 The LSTM memory unit structure



The memory unit is set with three gate structures: the forget gate f_t , the input gate i_t and the output gate o_t , which are used to record and update the information of memory unit. At moment t, the updated states of various gates are as follows:

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

$$i_t = \sigma\left(W_i \cdot \left[h_{t-1}, x_t\right] + b_i\right) \tag{2}$$

$$O_t = \sigma\left(W_o \cdot \left[h_{t-1}, x_t\right] + b_o\right) \tag{3}$$

$$C_t = f_t \cdot C_{t-1} + i_t \cdot \tanh\left(W_c \cdot \left[h_{t-1}, x_t\right] + b_c\right) \tag{4}$$

$$h_t = O_t \cdot \tanh\left(C_t\right) \tag{5}$$

where, x represents the input data, h refers to the unit output of LSTM, C is the memory unit value, σ is the sigmoid function, W_f, W_i, W_o, b_f, b_i and b_o represent the weights and biases of 3 gates respectively. Formula (4) shows the update of memory unit, how much information is forgotten, and which part of current input information requires being updated to current memory unit. Formula (5) generates current output result, and the output gate decides which information is finally output. The experiment and research of Li and Qian (2016) shows the LSTM model outperforms the standard RNN model in many sentiment analysis tasks.

Based on that, Graves and Schmidhuber (2005) proposed the Bidirectional Long Short-Term Memory (Bi-LSTM) network. This method employs two LSTM networks for simultaneous training from front and behind, which are connected to the same output layer to obtain the ability to capture the past and future information. It can avoid the shortage of traditional LSTM method that it cannot capture the context information due to serialisation processing.

2.3 Attention mechanism

The attention model can distinguish the importance of various target units, which can help find important features. Mnih et al. (2014) were the first to use the RNN model in image classification; later, Bahdanau et al. (2014) used it in the field of natural language processing, which was used in machine translation (Google™ Neural Network Translation System). Then, the attention model has been gradually expanded to the sentiment analysis tasks. For example, Zhang et al. (2018) proposed a weibo sentiment analysis method based on the double-attention model; Zhao et al. (2018) proposed a double-attention sentiment analysis model by combining the characteristics of word. The traditional self-attention mechanism can be represented with Formulas (6), (7) and (8), in which, h_i is the input hidden vector; *tanh* is the activation function; t is the sequence length of original sentence.

$$Y_i = tanh\left(Wh_i\right) \tag{6}$$

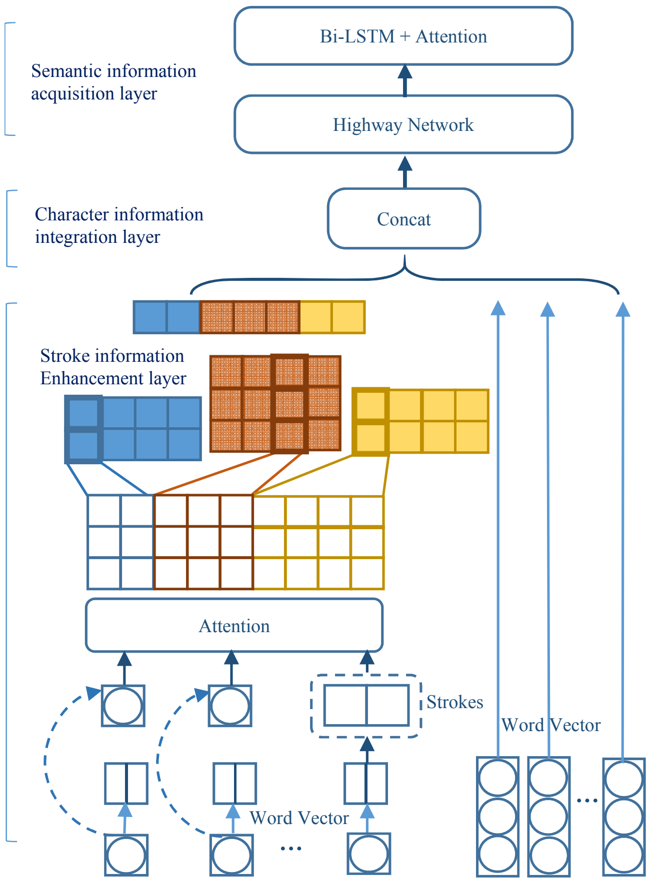
$$\alpha_i = softmax\left(Y_i\right) = \frac{exp\left(Y_i\right)}{\sum_{i=1}^{t} exp\left(Y_i\right)} \tag{7}$$

$$V = \sum_{i=1}^{t} \alpha_i h_i \tag{8}$$

3 Integrated learning model based on the morphology of Chinese characters and Bi-LSTM

Because the internal structure of Chinese characters, such as strokes and radicals, contains semantic information, more valuable information can be obtained by exploring their correlation, which can effectively solve the OOV problem. By combining character, word and stroke-radical level information with the Highway Network and Bi-LSTM model, this paper proposes the Sub-word Attention-based Bi-LSTM (SWAT-Bi-LSTM) model that integrates the morphology of Chinese characters (e.g., stroke and radical). The model consists of the stroke information enhancement layer, character information integration layer and semantic information acquisition layer, as shown in Figure 2.

Figure 2 The structure of stroke-radical attention combined with Bi-LSTM



3.1 *Word embedding that integrates the morphological information*

The character and word vector modelling (word embedding) is a process is to convert the word in sentence to dense vector that can be understood by the computer. In the meantime, due to extraction of word vector information by LSTM, long sequence length will result in attenuation of semantic information.

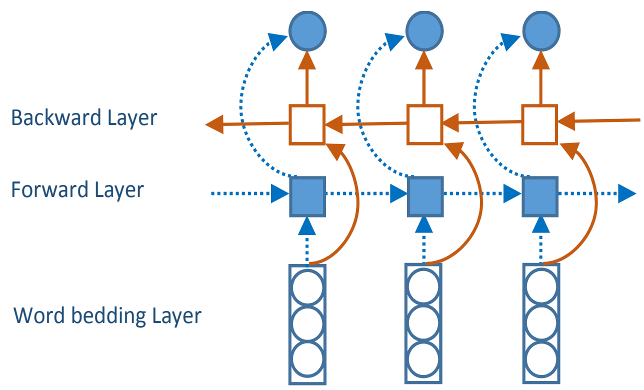
We find that for the Chinese text, both the Chinese character and its radical contain abundant semantic information. In their work conducted in Zhang et al. (2015) from New York University proved that great results can be obtained in text classification and sentiment analysis tasks by using word-level characteristics. Based on the word vector, this paper integrates the character-radical level features to enhance the semantic information of semantic analysis and address the OOV and information attenuation problem.

In this paper, firstly, we use the Jieba word vector (https://pypi.org/project/jieba/) tool to obtain the words from each sentence; then, obtain the basic stroke structure and radical data of Chinese character from (http://www.zdic.net/), so that the Chinese character can divide to the stroke and radical level. Initialise the character and word vector through random initialisation of word vector. $\mathrm{W} = \{w_1, w_2, \ldots, w_n\}$ is the word vector set, $\mathrm{C} = \{c_1, c_2, \ldots, c_n\}$ is the character vector set. E^w is an $n \times d_w$ -dimension word vector matrix; E^c is an $n \times d_c$ -dimension character vector matrix, in which, n and m are the numbers of words and characters in the sentence, d_w and d_c are the dimensions of word and characters respectively. In set *C*, the character element c_i can be divided into a group of stroke structure sequences $S = \{s_1, s_2, \ldots, s_p\}$ according to the stroke structure and radical data of Chinese characters, where, p is the number of strokes. In order to utilise the morphological information of Chinese characters, in this paper, we use the cw2vec model proposed by Cao et al. (2018) from Ant Financial Services Group at AAAI (2018), embed the stroke and radical level "sub-word" information into the word vector, and finally obtain the character vector set *C*' that contains the stroke and radical information.

3.2 *Highway network and Bi-LSTM model*

The Bi-LSTM model consists of two LSTMs: a forward LSTM and a backward LSTM, which output the sequential and reversal features of text respectively, and its structure is as shown in Figure 3. The model proposed in this paper uses CNN to extract the radical information to obtain the character-level semantic information; then, the Bi-LSTM is combined with the context word information for further information extraction; finally, the Highway Network is used to integrate deep semantic information.

Figure 3 The Bi-LSTM model structure



The Highway Network (Srivastava et al., 2015; Kim et al., 2016) is a network structure proposed by Srivastava et al. (2015) inspired by the LSTM gate mechanism. This network allows information to pass through various layers of deep neural network in a high speed without any barrier, which can effectively alleviate the vanishing gradient problem and further enhance the fine-grained information. Assume a classic neural network consists of L layers, x represents the input of previous layer, y represents the output, and W is the random initialisation weight matrix. Its output can be represented with Formula (9). Define T as a non-linear transformation function, and the Highway Network can be represented with Formula (10).

$$y = \mathrm{H}\left(x, W_H\right) \tag{9}$$

$$y = H\left(x, W_H\right) \cdot T\left(x, W_T\right) + x \cdot \left(1 - T\left(x, W_T\right)\right) \tag{10}$$

3.3 Scaled dot-product attention mechanism

The scaled dot-product attention model was proposed by Vaswani et al. (2017) at NIPS, which has higher efficiency and can also prevent the vanishing gradient problem caused by flow of high-dimensional information. On the stroke information enhancement layer, this paper uses this mechanism to efficiently obtain the combinational relation between the internal stroke and radical information of character. On the semantic information extraction layer, this mechanism is used to enhance the ability to extract the semantic sequence information.

Assume the stroke and radical level word embedding matrix is $\boldsymbol{E}$. Map it to the Queries(Q), Key (K) and Value (V) matrices according to Formulas (11), (12) and (13), where, P is the relative position matrix of random initialisation; W is the weight matrix of random initialisation; d is the dimension of matrix K. Then, the output matrix can be represented with Formula (14).

$$\boldsymbol{Q} = Relu\left(\boldsymbol{E} + P, \boldsymbol{W}_Q^T\right) \tag{11}$$

$$\boldsymbol{K} = Relu\left(\boldsymbol{E} + P, \boldsymbol{W}_K^T\right) \tag{12}$$

$$\boldsymbol{V} = Relu\left(\boldsymbol{E} + P, \boldsymbol{W}_V^T\right) \tag{13}$$

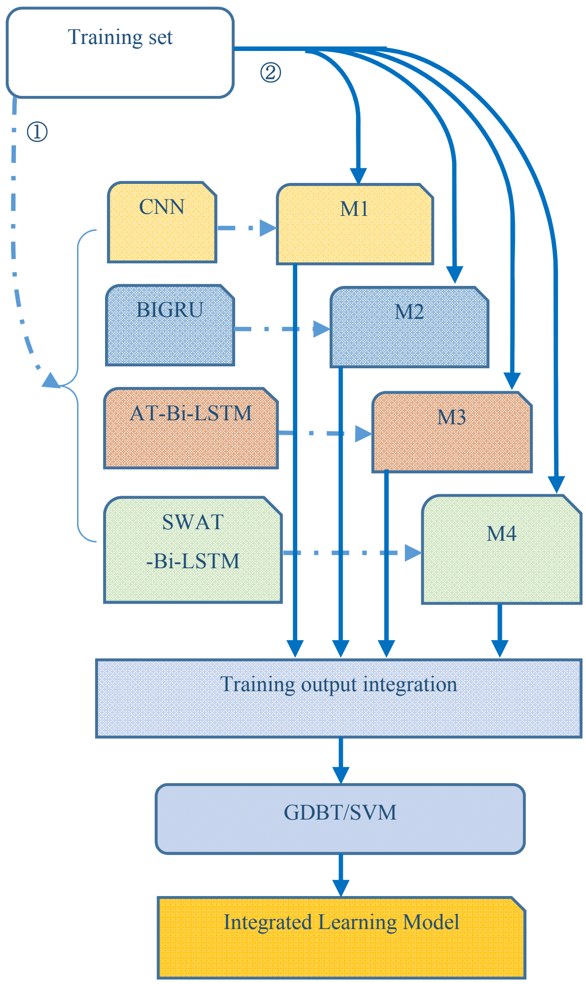
$$\text{Attention}\left(\boldsymbol{K}, \boldsymbol{Q}, \boldsymbol{V}\right) = \text{softmax}\left(\boldsymbol{Q}\boldsymbol{K}^T \big/ \sqrt{d\boldsymbol{K}}\right)\boldsymbol{V} \tag{14}$$

3.4 Ensemble learning model based on the bagging method

According to the concept of integrated learning, some weak learners can combine to form a strong learner to complete the learning task. In this paper, we first trained four base sub-learner (SRAT-Bi-LSTM, CNN, BIGRU and Bi-LSTM + Attention) on the training data set and got model M1~M4, as shown in Figure 4. Then, we leveraged boosting strategy GDBT to implement an integrated learning process to get the final integrated learning model, as shown in Figure 4. In particular, in the experiment, we also use an SVM-based integration strategy as a comparison of ensemble learning.

The results show that the performance of the integrated learning-based method is further improved compared with the single-model method. In the meantime, the GBDT-based strategy is simpler than the SVM learning strategy completely based on maximum geometric interval, which also has faster fitting speed and smaller deviation. The final experiment in this paper adopts the GDBT-based integration scheme. The framework of integration model proposed in this paper is as shown in Figure 4.

Figure 4 Integrated learning model based on GDBT/SVM



4 Experiments and results

4.1 Data sets and experimental setting

In this section, details of our experimental settings and results are illustrated, including the processing of data sets, standard benchmarks and baseline algorithms. We use a data set about the hotel evaluation collected by Tan and Zhang (2008) as the experiment data, which is called "ChnSentiCorp-Hou" in this paper. It consists of 10,000 pieces of data in total, including 7000 positive evaluations and 3000 negative evaluations.

In the meantime, the character and word vectors in the experiment have the dimension of 300. The hidden layers of the character, word and Bi-LSTM all have 64 neurons. Set the learning rate of Adam as 0.001 and set dropout as 0.5. In the meantime, set the layers of bidirectional LSTM as 128 layers.

The following ten methods are used to conduct experiment on the ChnSentiCorp-Hou Chinese data set:

1 *SVM*: By referring to the work of Zhao et al. (2018) and Yang and Xia (2017), we use the SVM method to conduct experiment on the data set mentioned above and obtain the result.

2 *CNN*: By referring to the work of Zhao et al. (2018) and Yang and Xia (2017), we use the CNN method to conduct experiment on the data set mentioned above and obtain the result.

3 *LSTM*: By referring to the work of Hochreiter and Schmidhube (1997), we use the LSTM network to conduct experiment on the data set mentioned above and obtain the result.

4 *Bi-LSTM*: By referring to the work of Graves and Schmidhuber (2005), we use the Bi-LSTM network to conduct experiment on the data set mentioned above and obtain the result.

5 *Bi-GRU*: GRU can be regarded as a variant of LSTM, and it combines the forget gate and input gate into a single update gate. The final model is simpler than the standard LSTM, but they have close performance in many tasks. In this paper, it is used as a method for comparison in the experiment, which is also used in the integration model.

6 *AT-LSTM*: Use the LSTM method based on word attention mechanism to conduct experiment on the data set mentioned above and obtain the result.

7 *AT-Bi-LSTM*: Use the Bi-LSTM method based on word attention mechanism to conduct experiment on the data set mentioned above and obtain the result.

8 *CWAT-Bi-LSTM*: By referring to the work of Zhao et al. (2018), we use the character and word attention model based on Bi-LSTM (CWAT-Bi-LSTM) to conduct experiment on above data set and obtain the result.

9 *SWAT-Bi-LSTM*: This is the method proposed in this paper, and the SRAT-Bi-LSTM single model is used to conduct experiment on above data set and obtain the result.

10 *SWAT-Bi-LSTM-G*: This is the method proposed in this paper, and the integrated model introduced in Section 3, part D is used to conduct experiment on the data set mentioned above and obtain the result.

4.2 *Results and discussion*

In the experiment, we calculate and compare the precision, recall rate and F1 value of the models mentioned above, and the results are listed in Table 1.

According to the experimental results in Table 1, it can be seen that first of all, the simple model method (such as LSTM) can also achieve great results in small-scale test data set in this paper. By comparing the performance of SVM and CNN, we can see that the CNN method has higher accuracy than SVM. In general, the performance of deep learning-based method is greatly improved compared to the traditional machine learning method. Because the LSTM method can obtain the long-term dependency in sentence, its performance is significantly improved in sentiment analysis tasks compared to SVM and CNN. Through lateral comparison of LSTM, Bi-LSTM and Bi-GRU, it can be seen that the bilateral strategy-based model (Bi-GRU and Bi-LSTM) has better performance in accuracy, recall rate and F1 value than the standard LSTM model, which indicates that collection of context information plays a certain role in sentiment analysis. According to lateral comparison of Bi-GRU and Bi-LSTM, we find that these two algorithms do not present significance difference in performance.

Table 1 Sentiment analysis results on Chnsenti Crop-Hou data set

Model	*Precision*	*Recall*	*F1-Score*
SVM	0.8300	0.8765	0.8530
CNN	0.8840	0.9162	0.9128
LSTM	0.9015	0.9591	0.9246
Bi-LSTM	0.9125	0.9641	0.9388
Bi-GRU	0.9127	0.9655	0.9392
AT-LSTM	0.9105	0.9029	0.9189
AT-Bi-LSTM	0.8986	0.9138	0.9298
CWAT-Bi-LSTM	0.9080	0.9328	0.9281
SWAT-Bi-LSTM	**0.9135**	**0.9368**	**0.9331**
SWAT-Bi-LSTM-G	**0.9465**	**0.9662**	**0.9742**

Through lateral comparison of the attention mechanism-based models (AT-LSTM, CWAT-Bi-LSTM and SWAT-Bi-LSTM), we can see that the model which has adopted the character-word combinational attention mechanism has better performance than the method which only uses the word attention mechanism. In the meantime, on this data set, among the methods based on word attention mechanism, AT-LSTM outperforms AT-Bi-LSTM. The reason could be that for small data set, the advantages of complicated model cannot be fully reflected.

By comparing them with the method proposed in this paper, we can see that after combining the internal structure information of Chinese character, its performance in accuracy, recall rate and F1 value is generally improved, and it has the best performance among all methods for comparison. Therefore, it can be seen that the results of sentiment analysis can be improved by combining the internal structure information of Chinese character. The experimental results also show the integrated model-based method discussed in this paper can further improve the classification accuracy.

5 Conclusion

This paper introduces an ensemble learning model based on sub-word attention mechanism and bidirectional long short-term memory model (SWAT-Bi-LSTM) which can provide an internal structural attention ability of Chinese characters.

This model can be used to deal with OOV problems in an emotionally sensitive WSN scenario, and no manual labelling is required. In the sentiment analysis task, based on the context information mining of character and word with Bi-LSTM, the scaled dot-product attention mechanism is integrated to extract the stroke and radical information of Chinese characters. The results of comparison experiments show that our method can improve the performance in sentiment classification. In the meantime, the GDBT integration scheme based on the boosting strategy can further improve the precision of result. Above all, the proposed algorithm can utilise the internal structural information (such as radicals and strokes) of Chinese characters to help identify unlabelled words, and the obtained sentiment analysis results can provide effective semantic information for the IoT and WSN, which is very helpful for the identification and classification of sensors and other devices in the network. At the same time, the obtained semantic information has also laid down the basis for the interaction among machines, devices and people in the IoT. On larger-scale data sets, the potential of the sentiment analysis method based on the internal structure of Chinese characters could be further explored in the future.

Acknowledgements

This work is supported in part by the NSFC program under Grant No. 61672372, 61472211, Natural science fund for colleges and universities in Jiangsu Province in 2019, Science and Technology Project of Suzhou under Grant No. SGZ2013130, Special Projects of Industrial Technological Innovation in Suzhou under Grant No. SYG201710, Youth Fund Project of Anhui Natural Science Foundation under Grant No.1608085QF144, Research and innovation project of Suzhou Vocational University under Grant No. SVU2018CX10 and Outstanding Science-technology Innovation Team Program of Colleges and Universities in Jiangsu.

References

Bahdanau, D., Cho, K. and Bengio, Y. (2014) 'Neural machine translation by jointly learning to align and translate', *Computer Science*, pp.1–15.

Bengio, Y., Ducharme, R. and Vincent, P. et al. (2003) 'A neural probabilistic language model', *Journal of Machine Learning Research*, No. 3, pp.1137–1155.

Cao, S. et al. (2018) 'cw2vec: learning chinese word embeddings with stroke n-gram information', *Proceedings of the 32nd Association for the Advancement of Artificial Intelligence (AAAI'18)*, New Orleans, LA, USA, pp.5053–5061.

Cho, K., Van Merrienboer, B. and Gulcehre, C. et al. (2014) 'Learning phrase representations using rnn encoder-decoder for statistical machine translation', *Proceedings of the Conference on Empirical Methods in Natural Language Processing*, pp.1724–1734.

Graves, A. and Schmidhuber, J. (2005) 'Framewise phoneme classification with bidirectional LSTM and other neural network architectures', *Neural Networks*, Vol. 18, Nos. 5/6, pp.602–610.

Graves, A. and Schmidhuber, J. (2005) 'Framewise phoneme classification with bidirectional LSTM and other neural network architectures', *IEEE International Joint Conference on Neural Networks*, IEEE, Canada, Vol. 18, Nos. 5/6, pp.602–610.

Guan, Z.Y. and He, J.H. et al. (2017) 'A survey on sentiment classification', *Journal of Computer Research and Development*, Vol. 54, No. 6, pp.1150–1170.

Hochreiter, S. and Schmidhube, J. (1997) 'Long short-term memory', *Neural Computation*, Vol. 9, No. 8, pp.1735–1780.

Irsoy, O. and Cardie, C. (2014) 'Opinion mining with deep recurrent neural networks', *Proceedings of Conference on Empirical Methods in Natural Language Processing*, Association for Computational Linguistics, USA, pp.720–728.

Kim, S.M. and Hovy, E. (2004) 'Determining the sentiment of opinions', *Proceedings of the 20th International Conference on Computational Linguistics*, Association for Computational Linguistics, USA, pp.1367–1374.

Kim, S.M. and Hovy, E. (2006) 'Extracting opinions, opinion holders, and topics expressed in online news media text', *Proceedings of Workshop on Sentiment and Subjectivity in Text*, pp.1–8.

Kim, Y. (2014) 'Convolutional neural networks for sentence classification', *Proceedings of the Conference on Empirical Methods on Natural Language Processing*, pp.1746–1751.

Kim, Y., Jernite, Y. and Sontag, D. et al. (2016) 'Character-aware neural language models', *Proceedings of the 30th AAAI Conference on Artificial Intelligence (AAAI'16*), Phoenix, Arizona USA, pp.2741–2479.

Li, D. and Qian, J. (2016) 'Text sentiment analysis based on long short-term memory', *IEEE International Conference on Computer Communication and the Internet*, IEEE, China, pp.471–475.

Liu, F., Ping, L. and Dacheng, D. (2017) 'Device-oriented automatic semantic annotation in IoT', *Journal of Sensors*, No. 5, pp.1–14.

Mikolov, T., Chen, K. and Corrado, G. et al. (2013) 'Efficient estimation of word representations in vector space', *Proceedings of the International Conference on Learning Representations (ICLR)*, pp.1–12.

Mikolov, T., Sutskever, I. and Chen, K. et al. (2013) 'Distributed representations of words and phrases and their compositionality', *Advances in Neural Information Processing Systems*, No. 2, pp.3111–3119.

Mnih, V., Heess, N. and Graves, A. et al. (2014) 'Recurrent models of visual attention', *Advances in Neural Information Processing Systems*, Vol. 3, pp.2204–2212.

Nasukawa, T. (2003) 'Sentiment analysis: capturing favorability using natural language processing', *Proceedings of the 2nd International Conference on Knowledge Capture*, ACM, New York, USA, pp.70–77.

Pang, B., Lee, L. and Vaithyanathan, S. (2002) 'Thumbs up: sentiment classification using machine learning techniques', *Proceedings of the ACL Conference on Empirical Methods in Natural Language Processing*, Association for Computational Linguistics, USA, pp.79–86.

Selbage, M., Wolfson, D., Zurek, B. et al. (2006) *Achieve Semantic Interoperability in a SOA: Patterns and Best Practices*, IBM Technical library.

Srivastava, R.K., Greff, K. and Schmidhuber, J. (2015) 'Highway networks', *Computer Science*, pp.1–6.

Tan, S. and Zhang, J. (2008) 'An empirical study of sentiment analysis for Chinese documents', *Expert Systems with Applications*, Vol. 34, No. 4, pp.2622–2629.

Terziyan, V., Kaykova, O. and Zhovtobryukh, D. (2010) 'Ubi road: semantic middleware for context-aware smart road environments', *Proceedings of the 5th International Conference on Internet and Web Applications and Services*, IEEE, Spain, Barcelona, pp.295–302.

Vaswani, A., Shazeer, N. and Parmar, N. et al. (2017) 'Attention is all you need', *Proceedings of the 31st Conference on Neural Information Processing Systems (NIPS'17)*, Long Beach, CA, USA, pp.1–11.

Wan, C.H., Lee, L.H. and Rajkumar, R. et al. (2012) 'A hybrid text classification approach with low dependency on parameter by integrating K-nearest neighbor and support vector machine', *Expert Systems with Applications*, Vol. 39, No. 15, pp.11880–11888.

Wang, S. and Manning, C.D. (2012) 'Baselines and bigrams: simple, good sentiment and topic classification', *Proceedings of the 50th Annual Meeting of the Association for Computational Linguistics: Short Papers - Volume 2 Association for Computational Linguistics*, No. 2, pp.90–94.

Wang, W. et al. (2013) 'Knowledge representation in the internet of things: semantic modelling and its applications', *Journal for Control, Measurement, Electronics, Computing and Communications*, Vol. 54, No. 4, pp.388–400.

Yang, S. and Xia, Z. (2017) 'A convolutional neural network method for Chinese document sentiment analyzing', *Proceedings of the IEEE International Conference on Computer and Communications*, IEEE, China, pp.308–312.

Yangsen, Z., Jia, Z., Gaijuan, H. and Yuru, J. (2018) 'Microblog sentiment analysis method based on a double attention model', *Journal of Tsinghua University (Science and Technology)*, Vol. 58, No. 2, pp.122–130.

Zhang, P. and Hu, Y. (2018) 'Lightweight semantic model design for internet of things', *Digital Communication World*, Vol. 168, No. 12, pp.98–100.

Zhang, X., Zhao, J. and Lecun, Y. (2015) 'Character-level convolutional networks for text classification', *Proceedings of the Neural Information Processing Systems (NIPS'15)*, Montreal, Quebec, Canada, Vol. 215, No. 28, pp.1–12.

Zhao, F., Yang, Y., Jiang, R., Zhang, L. and Ren, X. (2018) 'Sentiment analysis based on double-attention Bi-LSTM using part-of-speech', *Journal of Computer Applications*, Vol. 38, No. S2, pp.103–106+147.