Extracting Knowledge Entities from Sci-Tech Intelligence Resources Based on BiLSTM and Conditional Random Field

Weizhi LIÃO(a), Member, Mingtong HUANG(b), Pan MA(c), and Yu WANG(d), Nonmembers

SUMMARY There are many knowledge entities in sci-tech intelligence resources. Extracting these knowledge entities is of great importance for building knowledge networks, exploring the relationship between knowledge, and optimizing search engines. Many existing methods, which are mainly based on rules and traditional machine learning, require significant human involvement, but still suffer from unsatisfactory extraction accuracy. This paper proposes a novel approach for knowledge entity extraction based on BiLSTM and conditional random field (CRF). A BiLSTM neural network to obtain the context information of sentences, and CRF is then employed to integrate global label information to achieve optimal labels. This approach does not require the manual construction of features, and outperforms conventional methods. In the experiments presented in this paper, the titles and abstracts of 20,000 items in the existing sci-tech literature are processed, of which 50,243 items are used to build benchmark datasets. Based on these datasets, comparative experiments are conducted to evaluate the effectiveness of the proposed approach. Knowledge entities are extracted and corresponding knowledge networks are established with a further elaboration on the correlation of two different types of knowledge entities. The proposed research has the potential to improve the quality of sci-tech information services.

key words: sci-tech intelligence resources, knowledge entity, sequence labeling, BiLSTM-CRF

1. Introduction

Sci-tech intelligence services currently play a fundamental role in economic development, scientific research, and many design and production processes. In particular, acquiring the latest sci-tech information in a timely manner has become an imperative task for any successful enterprise in today’s rapidly growing knowledge economy. It is well recognized that such service highly depends on the exploration and utilization of sci-tech intelligence resources, which mainly exist in the form of sci-tech literature; the amount of such resources is facing an explosive growth as we enter the era of “Big Data” and “Internet plus concept”. As a result, traditional data processing models no longer suffice, and there is an urgent need for more automated and intelligent technology to mine, analyze, and study the massive amount of available sci-tech intelligence resources. The recent advances in machine learning and deep learning have facilitated the development of such technologies, among which knowledge entity extraction technology is of great importance. The research work of extracting knowledge from the sci-tech intelligence resources is relatively rare, so we choose this field and achieved very good results.

The main task of knowledge entity extraction is the identification of specific types of knowledge entities in sci-tech intelligence resources. Because there are many types of entities in the existing literature, and the title of many papers or patents is presented in the form of a study based on a certain point of knowledge, this paper classifies knowledge entities into two categories: knowledge point entities and research area entities.

Knowledge point entity: This term expresses a key knowledge point or technology point in sci-tech intelligence resources. For example, in the phrase “speech emotion recognition based on the RBF neural network”, “RBF neural network” is the knowledge point entity that needs to be extracted.

Research area entity: This term refers to the name of a specific application area of a knowledge point or technology point in sci-tech intelligence resources. For example, in the phrase “speech emotion recognition based on the RBF neural network”, “speech emotion recognition” is the research area entity.

The main task of this paper is to extract these two entities. These knowledge entities serve as the basis for various value-added activities, including developing knowledge networks, providing personalized intelligence services, exploring knowledge systems, and offering other sci-tech information services. Therefore, knowledge entity extraction has been a focus of research in sci-tech intelligence. Sci-tech intelligence resources knowledge entity extraction is a specific type of domain entity recognition in the field of named entity recognition (NER), and NER is mainly used to recognize persons, locations, and organizations.

The initial NER approach was based on some manually defined rules and dictionaries. Because they depend on domain experts and linguists, these methods cannot be adapted to other domains. In view of this problem, researchers have used machine learning methods with artificial features to identify named entities.

The traditional machine learning methods include hidden Markov models (HMMs) [1], [2], maximum entropy models (MEMs) [3], and conditional random fields (CRFs) [4], [5]. These methods require domain knowledge to manually design features. Therefore, the quality of feature templates directly affects the performance of
NER [6], [7]. It is very difficult to design a good feature template, as it requires good domain knowledge and language knowledge. With the development of neural networks and computing power, researchers are increasingly using neural networks to replace the artificial features [8], and have obtained great feedback.

The problems of extracting sci-tech knowledge entities as a specific type of domain entity recognition in NER remain largely unexplored. Most existing efforts involve the identification of named entities in specific fields. For example, for social media, Ritter et al. [9] attempted to identify named entities in Twitter texts. Peng et al. [10] used the CRF approach to study the recognition of named entities in Weibo texts. Identifying named entities is also a hot topic in biomedicine. Settles et al. [11] used simple orthogonal features in CRF to carry out pertinent biomedical entity recognition. Similar research has also been conducted in several other fields, such as chemical entities [12], E-mail [13], tourism domain entities [14], crime entities [15], and clinical entities [16].

For sci-tech literature, Jiang et al. [17] used ontology, vocabulary, and writing rules to extract knowledge entities from patent abstracts. Singh et al. [18] proposed a rule-based information extraction framework that automatically extracts knowledge entities from sci-tech literature. However, these methods require the arduous manual construction of rules, and the extraction performance highly depends on the quality of such rules. Wen et al. [19] attempted to obtain knowledge entities of professional literature by building the feature template and adopting the CRF model. Chen et al. [20] employed the CRF model in combination with extra knowledge, e.g., the selection of speech components and HowNet semantics, to extract theoretical knowledge entities from 1,822 articles. With the support of the manual construction of complex templates and the use of extra knowledge bases for feature building, their method exhibits good results, but suffers from high energy consumption and the failure to achieve end-to-end learning. Basaldella [21] used the BiLSTM neural network model for the identification of term entities from papers. The bidirectional LSTM (BiLSTM) is a two-direction, i.e., forward and backward, LSTM. In most cases, BiLSTM is used to model contextual information and capture two-way semantic dependencies. This model does not require the building of complex extraction rules nor the establishment of feature templates, and enables the learning of context information, which is useful for predicting the final entities. However, because this method uses the softmax layer to output the entity label directly, it does not consider the global label information to obtain the optimal prediction of the sequence label, and it may not always yield satisfactory performance.

To address these challenges, this paper proposes a feasible sequence labeling approach based on BiLSTM-CRF to extract knowledge entities from sci-tech intelligence resources. The proposed approach first uses the BiLSTM neural network to obtain sentence contexts, and then presents the optimal label sequence by integrating the global label information through CRF, hence obviating the need for manual feature construction and extra expert knowledge.

The main contributions of this paper lie in the following three aspects.

1) When dealing with sci-tech literature, the authors have provided a clear definition and analysis of the related knowledge entities, and have advanced a BiLSTM-CRF-based approach for the extraction of knowledge entities that does not require additional artificial features or expert knowledge.

2) The BiLSTM-CRF-based approach was proven to be effective via experimental verification and the analysis of 20,000 works of sci-tech literature.

3) In terms of the correlation between knowledge points and the names of research areas in sci-tech literature, knowledge networks have been drawn, and corresponding elaborations have been made. The efforts of the authors may contribute to the improvement of sci-tech intelligence services.

The remainder of this paper is organized as follows. Section 2 provides the definition and analysis of knowledge entities in sci-tech literature, as well as a description of the overall research framework. Section 3 details the proposed approach. Section 4 presents experimental results for performance evaluation, and Sect. 5 concludes the work.

2. Knowledge Entity Extraction Method Based on the BiLSTM-CRF Model

In this work, the BiLSTM-CRF-based sequence labeling method is employed to deal with the knowledge entities existing in sci-tech resources. First, through word vector technology, the unstructured text data in sci-tech resources can be converted into word vectors that can be processed by the neural network. These word vectors are then input into the layer of the BiLSTM neural network, and the BiLSTM hidden layer is output for splicing. After a fully connected layer is formed, the dimension is converted into the number of labeled categories that need to be predicted, and finally the global label information is integrated by CRF to obtain an optimal labeling sequence. The structure of the BiLSTM-CRF model is illustrated as follows.

2.1 Word Embedding Layer

In general, there are two kinds of word vector technology, word-based and character-based word vector, used on Chinese entity recognition. Word-based word vector technology first segments the Chinese text, and then employs the word vector technique. The character-based word vector technology applies the word vector technology immediately, without the step of word segmentation. The word vectors thus obtained are character-based. Character-based models abandon the semantic information that words have, as well as the pre-trained vector ecosystem that is directly available. The advantage of character-based models is that they alleviate the vocabulary problems we encounter in model input...
and avoid computational bottlenecks on model output. In terms of input, character-based word vector can greatly improve the vocabulary that our model can handle. In terms of output, the vocabulary of character-based models is small, so computational costs are lower. This nature allows for the use of certain training techniques (e.g., joint training of a language model) and faster training speeds when budgets are limited. Technically, the word vector used in this paper is character-based, the adopted word vector matrix is randomly initialized, and the relevant matrix parameters are constantly updated in the process of model training.

Through the word embedding layer, a given input sequence converts into where \( n \) refers to the time step of the input sequence, and \( d_{x} \) is the dimension of a word vector.

\[
X = [x_1^o, x_2^o, \ldots, x_n^o], \quad x_i^o \in R^{d_{x}}, \quad i \in 1, 2, \cdots, n \tag{1}
\]

### 2.2 BiLSTM Layer

The recurrent neural network (RNN) has enjoyed wide application in tasks such as speech recognition, machine translation, and part-of-speech annotation since it was proposed. However, one of the biggest shortcomings of RNN is its long-term dependence. When dealing with long-term problems, and for a long text, the language meaning may exist on the opening and ending words, but the usual recurrent neural network cannot establish a corresponding semantic connection due to the vanishing gradient problem. To solve the long-term dependence of RNN, Hochreiter [22] et al. developed a special gate structure to prevent vanishing gradient, and proposed an updated RNN called long short-term memory (LSTM).

By the vehicle of the designed “gate” (input gate, forget gate, and output gate), LSTM passes or forgets information. The LSTM transition functions are defined as follows.

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

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

\[
C_t = \text{tanh} (W_c \cdot [h_{t-1}, x_t] + b_c) \tag{5}
\]

\[
C_t = f_t \times C_{t-1} + i_t \times \tilde{C}_t \tag{6}
\]

\[
h_t = o_t \times \text{tanh} (C_t) \tag{7}
\]

In these equations, \( \sigma \) is the sigmoid function. For input gate \( i_t \), output gate \( o_t \), and forget gate \( f_t \), \( W_i \), \( W_o \), and \( W_f \) respectively represent the parameter matrix in the three gate structures, and \( b_i \), \( b_o \), \( b_c \) indicate the bias of the three gate structures. \( h_{t-1} \) is the output of the hidden layer of the previous time step, while \( x_t \) represents the input of the current time step. \( C_{t-1} \) indicates the cell state in the previous time step, and \( C_t \) expresses the cell state at the current time step. These “gate” structures equip LSTM with strong control capabilities to forget and retain information.

When inputting the output \( X \) of the word embedding layer to the BiLSTM layer, \( \tilde{h} \) and \( \overline{h} \) are obtained from memory cells \( C_t \) through the output door \( o_t \).

\[
\tilde{h} = [\tilde{h}_1, \tilde{h}_2, \cdots, \tilde{h}_n], \quad \tilde{h}_i \in R^r, \quad i \in 1, 2, \cdots, n \tag{8}
\]

\[
\overline{h} = [\overline{h}_1, \overline{h}_2, \cdots, \overline{h}_n], \quad \overline{h}_i \in R^r, \quad i \in 1, 2, \cdots, n \tag{9}
\]

After splicing the outputs, the results are as follows.

\[
h = [h_1, h_2, \cdots, h_n], \quad h_i = [\tilde{h}_i \oplus \overline{h}_i] \tag{10}
\]

Where \( \oplus \) represents the operation of concat the vector in both directions.

### 2.3 Fully Connected Layer

The fully connected layer is designed to convert the output dimension of the BiLSTM layer into the number of label categories that need to be predicted. The equation for the output \( h_t \) in the time step \( t \) of the output sequence \( h \) of the BiLSTM layer is given below.

\[
z^{T}_t = h^{T}_tM, \quad M \in R^{2d_{xm}} \tag{11}
\]

\[
Z = [z_1, z_2, \cdots, z_n], \quad z_i \in R^m, \quad i \in 1, 2, \cdots, n \tag{12}
\]

In the above equation, \( M \) represents the parameter matrix, and \( m \) is the number of labels that need to be predicted. In this work, there are seven types of labels that need to be predicted.

### 2.4 CRF Layer

The labels that need to be predicted are highly dependent on the sequence; for example, “E-POI” can only appear after “I-POI”, etc. The application of CRF can provide desirable corresponding solutions, as it displays a good performance in the modeling of grammar rules.

For a possible sequence of prediction result is \( y \).

\[
y = (y_1, y_2, \cdots, y_n) \tag{13}
\]

then the score of the predicted sequence is as follows:

\[
f_t = \sigma (W_f \cdot [h_{t-1}, x_t] + b_f) \tag{3}
\]

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

\[
C_t = \text{tanh} (W_c \cdot [h_{t-1}, x_t] + b_c) \tag{5}
\]

\[
C_t = f_t \times C_{t-1} + i_t \times \tilde{C}_t \tag{6}
\]

\[
h_t = o_t \times \text{tanh} (C_t) \tag{7}
\]
s(Z, y) = \sum_{i=0}^{n-1} A_{y_{i}, y_{i+1}} + \sum_{i=1}^{n-1} P_{y_{i}} \tag{14}

where \( A \) and \( P \) respectively correspond to the transfer score matrix and the State score matrix.

The probability of sequence can be calculated by the softmax function.

\[
p(y|Z) = \frac{e^{s(Z, y)}}{\sum_{\tilde{y} \in Y_Z} e^{s(Z, \tilde{y})}} \tag{15}
\]

In the following equation, \( Y_Z \) represents a collection of all possible observation sequences for \( Z \). For the correct prediction sequence \( y \), it is necessary to continuously maximize the corresponding logarithmic probability during the training process.

\[
\log(p(y|Z)) = s(Z, y) - \log(\sum_{\tilde{y} \in Y_Z} e^{s(Z, \tilde{y})}) \tag{16}
\]

In the process of prediction, the prediction sequence \( y^* \) is chosen to be that with the highest score among all possible prediction sequences, and the final output layer obtains the one-hot code of the corresponding label.

\[
y^* = \text{argmax}(s(Z, y)), \ y \in Y_Z \tag{17}
\]

3. Experiment and Analysis

This section illustrates the details of data processing, as well as the overall research framework and experimental verification of the proposed approach. The processing and details of the dataset are elaborated upon in the second paragraph. Additionally, in the proposed model, the character-based word vectors that are employed are randomly initialized and constantly updated during the training of the model.

3.1 Experimental Data Description

The research object of the present study is Chinese sci-tech literature based on professional sci-tech information resources. These works of literature have large quantities of knowledge point and research area entities. To fully explore the connections between these knowledge entities and certain research areas, these two types of knowledge entities in the literature were extracted. According to the previous definitions and corresponding examples, it is evident that these two kinds of knowledge entities are usually specific and recognized technical terms. In general, they are characterized by a fixed structure and clear definition, and therefore methods based on sequence labeling can promise good results.

The application of the BiLSTM-CRF neural network does not require the manual construction of complex feature templates, and the powerful extraction capabilities of BiLSTM enable the context features to be built automatically, which also saves extra labor in the process of feature building.

Figure 2 illustrates the text processing details and the overall research framework. First, 20,000 Chinese “neural network”-themed works of sci-tech literature were processed, including journals, conferences, and dissertations. Specific operations in the pre-processing phase included the segmentation of sentences, the elimination of sentences of inappropriate length, and so on. The pre-processed sentences were then annotated, as presented in Table 1. In total, 50,243 pieces of annotated data were obtained. The data was then segmented; 45,243 pieces were taken as the training set, while the remaining 5,000 were designated as the test set. The numbers of the two types of knowledge entities in the training set and the test set were respectively counted, and the details are presented in Table 2. The BiLSTM-CRF neural network model was then introduced to the training set. After training, the trained model was tested on the test set, and the effects of the model with
Table 3  Parameters of the BiLSTM-CRF model

| Parameter          | Value |
|--------------------|-------|
| character embedding| 100   |
| batch size         | 64    |
| hidden size        | 300   |
| dropout            | 0.5   |
| optimizer          | Adam  |
| learning rate      | 0.001 |

Table 4  Results of various models

| Metrics | Algorithms | Knowledge Point Entity | Research Area Entity |
|---------|------------|------------------------|----------------------|
|         |            | Precision               | Recall               |
|         |            | RNN  | 72.93 | 55.42 |
|         |            | LSTM | 71.83 | 57.82 |
|         |            | BiRNN | 90.68 | 77.72 |
|         |            | BiLSTM | 91.82 | 85.37 |
|         |            | RNN-CRF | 89.58 | 89.11 |
|         |            | LSTM-CRF | 90.73 | 86.65 |
|         |            | BiRNN-CRF | 91.8 | 87.35 |
|         |            | BiLSTM-CRF | 93.31 | 90.9 |

Precision scores of five models on knowledge point entities

Recall scores of five models on knowledge point entities

F1 scores of five models on knowledge point entities

regard to precision and recall, and to the F1 values, were evaluated.

3.2 Experimental Results and Analysis

The proposed neural network was built with the Python language on Tensorflow, the Google deep learning development framework. The specific parameter settings are listed in Table 3.

A series of comparative experiments were conducted to verify the effectiveness of the proposed sci-tech knowledge entity extraction method based on BiLSTM-CRF. The related comparative results are presented in Table 4 and Figs. 3–8.

It is evident from Table 4 that the BiLSTM-CRF model exhibited the best performance. In terms of the extraction of knowledge point entities, its F1 score reached 93.31%, and the F1 score for the extraction of the research area entities registered 90.77%. In other baseline experiments, the worst performance was exhibited by the one-way RNN, whose F1 scores of the knowledge point and research area entities were only 71.11% and 41.35%, respectively. This may be because the research area entities in sentences are not as clear as knowledge point entities, and have relatively long lengths and involve many complex words. Compared with RNN, the one-way LSTM revealed a slightly better performance, but the improvement was not distinct. By contrast, the two-way language modeling BiRNN and BiLSTM manifested better results than their respective one-way models. The extraction results of RNN-CRF and LSTM-CRF were also greatly improved after considering the constraint relationship between the labels. Therefore, it is clear that context information is very helpful for the extraction of knowledge entities.

It is apparent from Figs. 3 and 4 that the precision and recall scores of the extraction of knowledge point entities
did not reach a stable state, and the performance was poor, only about 70%; in contrast, the results of other models achieved greater than 90%. The LSTM-CRF presented more remarkable improvement in performance than LSTM. This is because the CRF layer learned the grammar rules between labels, thus making more accurate predictions. Considerable progress can also be found in the BiLSTM. The reason for this is that, compared with LSTM, the BiLSTM can well model context information and capture two-way semantic dependencies, which has a great impact on the extraction effect. As presented in Fig. 5, most of the models reached a plateau after 20 epochs. The BiLSTM-CRF model exhibited the best performance, and its F1 score curve was consistently higher than those of the other models.

Research area and knowledge point entities are strikingly different. Knowledge point entities generally refer to a recognized terminology with a relatively fixed structure. Comparatively, research area entities are long in length and complicated in structure, with a large number of words involved. Due to these differences, the corresponding extraction effect of the research area entities was distinguished from that of the knowledge point entities. In Figs. 6–8, it can be seen that the precisions of the single language modeling RNN and LSTM fluctuated between 50% and 70%, their recall rates shifted from 20% to 50%, and their F1 scores fluctuated between 40% and 52%. However, these indicators all reached about 70% for the knowledge point entities. Therefore, it can be concluded that the one-way neural network model illustrated less satisfactory performance than the other models. Additionally, the BiLSTM in the other models manifested a significant improvement over the one-way LSTM, presenting an increase of nearly 40%. This is because context information has an important role in the extraction of research area entities with long sequence predictions and other similar long entities. According to these results, the LSTM-CRF model also achieved good effects, even surpassing the performance of the BiLSTM model. The argument can be made that, for entities with relatively long lengths and complex structures, the CRF layer can be a favorable effective enhancer. This is because the CRF layer can integrate all global label information to obtain the optimal label sequence. Overall, the BiLSTM-CRF was equipped with both advantages and exhibited the best results.

3.3 Knowledge Network

We found there is a strong dependency relation between two types of entities when a knowledge point entity and a research area entity are in the a sentence, especially for title of sci-tech literature. For example, in the phrase “study on the method of stratigraphic section identification based on synergetic neural network algorithm”, there is a correlative knowledge point entity and research area entity pair, i.e., “synergetic neural network” and “stratigraphic section identification”, respectively.

Figure 9 presents a small portion of the knowledge networks that were drawn in the present study; it is a network centered on the knowledge point entity of “genetic algorithm”. The intuitive network explicitly reveals the research for which researchers have utilized a “genetic algorithm”. The knowledge network not only provides a faster channel for us to learn the application fields of the knowledge point entity, but also facilitates the comprehension of these knowledge points and the exploration of the potential relations between these fields.

A knowledge network centered on the research area entity is also presented. As indicated in Fig. 10, the re-
search area of “network intrusion detection” is associated with many knowledge point entities, and the knowledge points used by researchers in the field of “network intrusion detection” can be conspicuously observed. This is of great significance for the understanding of a research field and to follow up on a research work. When some descriptions are added to each associated knowledge point, e.g., the date, the most relevant and newest knowledge points of the current researchers in this field can be determined at first glance, and conclusions that could normally be drawn only by reading heaps of literature become readily available in knowledge networks. In this way, the time spent can be greatly reduced. Additionally, if links to relevant sci-tech literature are made in a knowledge network, the time taken to find the literature can be greatly reduced. As a result, service providers would be able to send quick and accurate sci-tech information that users are interested in, and enterprises and researchers can therefore master the latest research progress in the interested field. In short, many meaningful advancements can be made based on such knowledge networks.

4. Conclusion

For the better acquisition of intelligence information from sci-tech literature, this paper first defined and analyzed the knowledge entities that must be extracted from sci-tech intelligence resources, and then proposed an effective approach based on the BiLSTM-CRF neural network. The proposed approach does not require the manual construction of features or additional expert knowledge. Instead, the neural network can realize automatic feature extraction by transforming words into character vectors. Furthermore, according to experimental verification and relevant analysis, the approach is found to achieve good performance in the extraction of knowledge entities from sci-tech literature. Finally, knowledge networks centered on the knowledge point entity and the research area entity were respectively drawn on the basis of the experimental results.

Acknowledgments

This work was supported by the National Key R & D Program of China (2017YFB1400301 & 2017YFB1400300).

References

[1] L.R. Rabiner, “A tutorial on hidden Markov models and selected applications in speech recognition,” Proc IEEE, vol.77, no.2, pp.257–286, 1989.
[2] H.K. Yu, H.P. Zhang, and Q. Liu, “Chinese named entity identification using cascaded hidden Markov model,” J. Commun, vol.27, no.2, pp.87–94, 2006 (in Chinese).
[3] R. Koeling, “Chunking with maximum entropy models,” Proc 2nd Workshop on Learning Language in Logic and the 4th Conf on Computational Natural Language Learning, vol.7, pp.139–141, 2000.
[4] J.D. Lafferty, A. McCallum, and F.C.N. Pereira, “Conditional random fields: probabilistic models for segmenting and labeling sequence data,” 18th Int Conf on Machine Learning, pp.282–289, 2001.
[5] G. Luo, X. Huang, C.-Y. Lin, and Z. Nie, “Joint Entity Recognition and Disambiguation,” Proc Conf on Empirical Methods in Natural Language Processing, pp.879–888, 2015.
[6] A. Joshi, R. Lal, T. Finin, and A. Joshi, “Extracting cybersecurity related linked data from text,” IEEE 7th Int Conf on Semantic Computing, pp.252–259, 2013.
[7] Q.Q. Qiu, D.Q. Miao, and Z.F. Zhang, “Named entity recognition on Chinese microblog,” Comput Sci, vol.40, no.6, pp.196–198, 2013, (in Chinese).
[8] B. Tang, H. Cao, X. Wang, Q. Chen, and H. Xu, “Evaluating word representation features in biomedical named entity recognition tasks,” Biomed Res Int, vol.2014, 2014.
[9] A. Ritter, S. Clark, and O. Etzioni, “Named entity recognition in tweets: an experimental study[C],” // Proceedings of the Conference on Empirical Methods in Natural Language Processing, Stroudsburg: Association for Computational Linguistics, 1524-1534, 2011.
[10] N. Peng and M. Dredze, “Named entity recognition for Chinese social media with jointly trained embeddings,” [C]// Proceedings of the 2015 Conference on Empirical Methods in Natural Language Processing, Stroudsburg: Association for Computational Linguistics, pp.548–554, 2015.

[11] B. Settles, “Biomedical named entity recognition using conditional random fields and rich feature sets,” [C]// Proceedings of the International Joint Workshop on Natural Language Processing in Biomedicine and its Applications, Stroudsburg: Association for Computational Linguistics, pp.104–107, 2004.

[12] R. Leaman, C.-H. Wei, and Z. Lu, “tmChem: a high performance approach for chemical named entity recognition and normalization [J],” Journal of Cheminformatics, vol.7, S3, 2015.

[13] E. Minkov, R.C. Wang, and W.W. Cohen, “Extracting personal names from email: Applying named entity recognition to informal text,” [C]// Proceedings of the Conference on Human Language Technology and Empirical Methods in Natural Language Processing, Stroudsburg: Association for Computational Linguistics, pp.443–450, 2005.

[14] J. Guo, Z. Xue, and Z. Yu, “Named entity recognition for the tourism domain based on cascaded conditional random fields,” Journal of Chinese Information Processing, vol.23, no.5, pp.47–52, 2009.

[15] S. Karur and P.S. Thilagam, “Crime base: Towards building a knowledge base for crime entities and their relationships from online news papers,” Information Processing & Management, vol.56, no.6, 2019.

[16] Y. Li, G. Du, Y. Xiang, S. Li, L. Ma, D. Shao, X. Wang, and H. Chen, “Towards Chinese clinical named entity recognition by dynamic embedding using domain-specific knowledge,” Journal of Biomedical Informatics, vol.106, 2020.

[17] J. Caihong, Q. Xiaodong, and Z. Lijun, “Ontology-based patent abstracts’ knowledge extraction,” Data Analysis and Knowledge Discovery, vol.2, pp.23–28, 2009.

[18] M. Singh, S. Dan, S. Agarwal, P. Goyal, and A. Mukherjee, “App Tech Miner: Mining Applications and Techniques from Scientific Articles,” [C]//Proceedings of the 6th International Workshop on Mining Scientific Publications, pp.1–8, 2017.

[19] W. Wen, W. Sijie, R. Cai, and H. Zhifeng, “Type extraction and labelling of knowledge entities in the field of professional literature,” Journal of Chinese Information Processing, vol.32, no.1, pp.102–115, 2018.

[20] C. Feng, Z. Yujia, and W. Fang, “Automatic theory recognition in academic journals based on CRF,” Library and Information Service, vol.60, no.2, pp.122–128, 2016.

[21] M. Basaldella, E. Antolli, G. Serra, and C. Tasso, “Bidirectional LSTM Recurrent Neural Network for Keyphrase Extraction,” [C]//Proceedings of the Italian Research Conference on Digital Libraries, vol.806, pp.180–187, 2018.

[22] S. Hochreiter and J. Schmidhuber, “Long short-term memory,” Neural computation, vol.9, no.8, pp.1735–1780, 1997.

Weizhi Liao received Ph.D. degree in the School of Information Science and Technology from Southwest Jiaotong University in 2010. She is an Associate Professor in the Industrial Engineering Department of the School of Mechanical and Electrical Engineering at the University of Electronic Science and Technology of China. She has a core background on algorithms and data structures and her main research line is focused on large-scale data driven and service.

Mingtong Huang is currently a postgraduate student in the School of Mechanical and Electrical Engineering, University of Electronic Science and Technology of China, ChengDu, China. His current research interests include deep learning technology and Natural Language Processing.

Pan Ma received the M.S. degree in the School of Mechanical and Electrical Engineering, University of Electronic Science and Technology of China, ChengDu, China. His current research interests include deep learning technology and Natural Language Processing.

Yu Wang received the M.S. degree in the School of Mechanical and Electrical Engineering, University of Electronic Science and Technology of China, ChengDu, China. His current research interests include deep learning technology and Natural Language Processing.