Research on Named Entity Recognition Algorithm for Chinese Electronic Medical Records Based on Knowledge Embedding

Haotian Yang, Yanpeng Yang

Abstract—**In the era of increasing healthcare informatisation, a large number of electronic medical records (EMR) have emerged in China. Identifying medical named entities in EMRs helps medical decision-making. However, due to the specificity of medical terminology, word embeddings in the general domain cannot accurately express the semantics of the entities, leading to problems such as unclear entity boundaries. In this paper, we propose a knowledge embedding algorithm KGBERT-BiGRU-Att-CRF for the recognition of named entities in Chinese EMR, which constructs a medical knowledge graph, represents entities and their relationships with vectors, and combines BERT embeddings. BiGRU extracts the character context, computes the concern values, and integrates them into the textual representations. CRF models the label dependencies to output the best annotated sequence. Experiments show that knowledge embedding improves medical entity recognition. The algorithm reaches an F1 score of 86.97% on the merged CCKS dataset.**

Index Terms—**named entity recognition, knowledge embedding, attention, electronic medical records.**

I. Introduction

THE electronic medical records (EMR) [1] refer to information generated by healthcare professionals during medical procedures within healthcare information systems. Free-text unstructured data constitutes a significant component of electronic medical records. To promote more efficient extraction of medical knowledge embedded within electronic medical records, it is imperative to structure the EMR. Consequently, the accurate identification of medical named entities in the text of EMR is imperative for the effective structured processing of EMR.

Deep learning-based named entity recognition (NER) for Chinese EMR is currently the most stable method [2]. Bidirectional long short-term memory network with conditional random field (BiLSTM-CRF) is the most basic method to extract named entities [3]. However, due to the long Chinese e-reader text, long-range dependencies between words prevent the model from capturing the characteristics of e-reader texts in depth. In order to incorporate global semantic information into the character vectors and add a global attention mechanism into the model structure, the model's ability to process long sequential texts is improved by incorporating the attention mechanism [4]. However, this improvement makes it difficult to truly understand the specialised domain entities and inter-entity relationships contained in the EMR.

To solve this problem, this paper proposes a named entity recognition algorithm based on knowledge embedding that KGBERT-BiGRU-Att-CRF, which aims to enhance the linguistic representation with external knowledge in the knowledge graph. This approach is designed to refine the model's capacity to generalise word vectors in the healthcare domain, thereby boosting the model's recognition performance. The model presented in this paper achieves an F1 value of 86.97% on the merged CCKS dataset.

Manuscript received October 18, 2024; revised October 7, 2025

Haotian Yang is an engineer of Computer Science and Software Engineering, University of Science and Technology Liaoning, Anshan, CO 114051 China (e-mail: m18524337033@163.com).

Yanpeng Yang is a Senior Engineer of Network Information Centre, University of Science and Technology Liaoning, Anshan, CO 114051 China (e-mail: yyp@ustl.edu.cn).

II. Related Work

NER is the term used to denote the process of identifying entities with specific meanings in text. There are three common approaches to NER:

The rule-based and dictionary-based approaches rely on human rule building, with the former relying heavily on manual construction of rule templates and selection of features for matching. This results in low recognition recall [5].

In the domain of statistical machine learning, the NER problem is approached from the perspective of a sequence annotation problem. The annotation of each sentence constituent is achieved through the utilisation of an annotation model that has been trained on a substantial corpus. In practical research, generative models are commonly employed: HMM [6], CRF [7], etc. The CRF model solves the problem of dependencies between labels and has therefore become the standard NER model structure. However, the method is highly dependent on predefined artificial features.

In recent years, deep learning has been increasingly applied to named entity recognition (NER) problems and has achieved good results in general domains. Researchers have proposed using convolutional neural network (CNN) to extract features from input sentence vectors [8], which outperforms traditional machine learning methods that require manual feature extraction. However, the amount of contextual information this method can process at any given time depends on the fixed window size, making it incapable of handling long-range contextual information. As improved structures of recurrent neural network (RNN), long short-term memory network (LSTM) [9] and gated recurrent unit (GRU) [10] have become the mainstream methods in current text named entity recognition research due to their ability to address the aforementioned limitations. Additionally, considering that in Chinese electronic medical records, the semantic information of the current character not only exists within the character itself but also includes forward and backward information.

Researchers have proposed a recognition method based on BiLSTM [11], which can synthesize the contextual meaning of time series data and enhance data feature extraction and predictive performance by leveraging both forward and backward historical information. Therefore, the BiLSTM-CRF model is gradually becoming a typical structure in entity recognition [12]. By replacing the LSTM network with a simpler GRU network, a recognition model based on BiGRU-CRF was proposed [13], which achieved better results under the same training time and number of iterations.

However, the majority of such algorithms are only capable of extracting contextual information and are unable to address the long-distance dependencies between words. Consequently, some researchers have proposed BiGRU-Att-CRF [14], a NER method incorporating the attention mechanism. In comparison with the utilisation of BiGRU for the extraction of contextual information, BiGRU-Att demonstrates a superior capacity to integrate the semantic information of the text, thereby circumventing the information failure that arises from the length of the textual sequence. Consequently, the link between any two word vectors in the sequence is no longer disregarded as the sequence progresses through the propagation process.

The emergence of pre-trained large language models eliminates the need for large-scale corpus training for the word embedding process, such as the BERT model, which comprehensively captures character-level, word-level, and sentence-level relationships compared to static word vectors.The BERT-BiGRU-Att-CRF model [15] divides the task of recognising the named entities in Chinese EMR into two parts: word embedding and sequence annotation. In comparison with word embedding that does not employ the BERT model, which incorporates context and global semantics, the word vectors yielded by the BERT model encompass character-level, word-level, and sentence-level semantic attributes. These attributes are progressively integrated with contextual semantics and full-text information during the subsequent sequence annotation process. This integration renders the word vectors more precise in their expression and, in practice, enhances the recognition accuracy. [16].

However, extant NER methodologies are frequently predicated on large-scale generic domain prediction and Chinese EMR text itself for word embedding and feature extraction. This text contains specialised domain entities and inter-entity relationships that are difficult to be truly understood by the model. Consequently, this paper proposes a knowledge embedding-based NER algorithm, KGBERT-BiGRU-Att-CRF, to enhance the semantic representation of medical named entities with external medical domain knowledge, and to directionally improve the model's ability of word vector generalisation in the medical domain. This, in turn, enhances the model's recognition effect, a field which has not yet been the focus of research in the task of Chinese EMR named entity recognition.

III. Named Entity Recognition Framework

In order to achieve knowledge embedding, the fusion of textual information and additional knowledge information, heterogeneous information fusion is required.

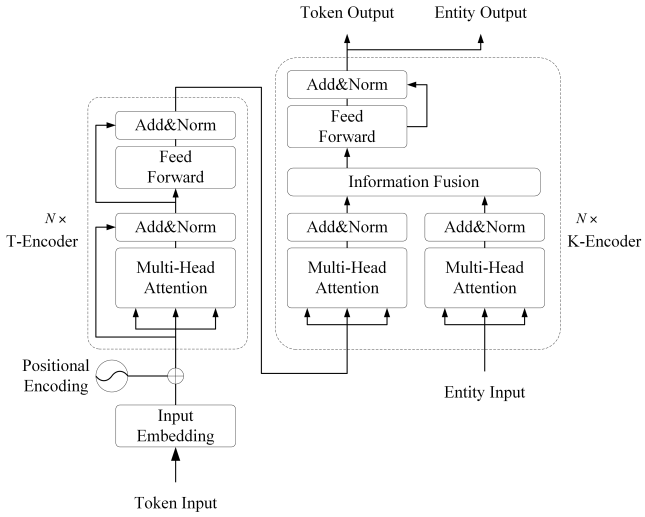

Fig. 1. Structure of KGBERT Model

A. Knowledge Embedding

The BERT model able to learn the connection of contextual characters, but not enough to constitute knowledge. Moreover, the space covered by different types of corpus will be different, and the trained model generalises better to the test set for the space covered by the training set, the BERT model is trained on a large number of general-purpose corpora, and the task of this paper belongs to the medical domain, so the BERT model generalises generally on the task of this paper [17].

In this paper, based on the BERT model, the knowledge graph was enbed into the BERT model and propose the KG-BERT model, which incorporates multi-information entities into the knowledge graph as external knowledge to enhance the linguistic representation of domain-specific vocabulary, so that the model can conform to human language at the knowledge level. The structure of the KGBERT model is shown in Figure 1.

The overall architecture of the KGBERT model consists of two main modules, the text encoder T-Encoder and the knowledge encoder K-Encoder. where the T-Encoder is aligned with the Transformer-Encoder in the BERT model. The K-Encoder is designed to achieve the fusion of heterogeneous information by integrating the additional knowledge information from the underlying textual information so that the heterogeneous information of tokens and entities can be characterised in a unified feature space.

Based on the alignment of text entities with the entities in the knowledge graph, K-Encoder fuses the entity representations of the knowledge module into the hidden layer of the semantic module. As shown inshown in Figure 2. $w_1, w_2, \ldots, w_n$ is used to denote the token sequence embedding, and $e_1, e_2, \ldots, e_n$ is used to denote the embedding of entities in the sequence.

Firstly, the two sequences are made to pass through their respective Multi-Head Self-Attention layers. The tokens in the sequence are then aligned with the corresponding entities (entities are aligned with the corresponding first token), and for a token w_j and its entity e_k in the case of layer i (total N layers of K-Encoders).

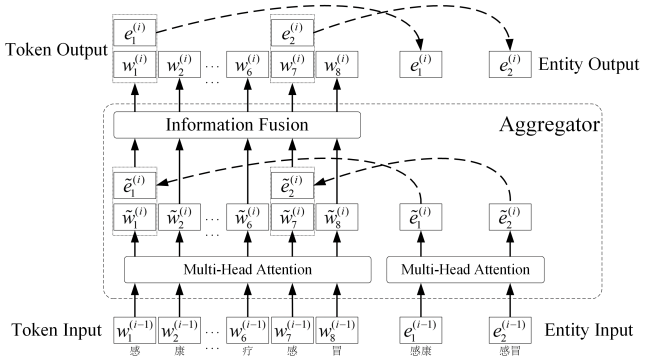

Fig. 2. Structure of K-Encoder

The information fusion are shown in Equation 1 to 3.

$$h_j = \sigma(\tilde{W}_t^{(i)} \tilde{w}_j^{(i)} + \tilde{W}_e^{(i)} \tilde{e}_k^{(i)} + \tilde{b}^{(i)}) \quad (1)$$

$$w_j^{(i)} = \sigma(W_t^{(i)} h_j + b_t^{(i)}) \quad (2)$$

$$e_k^{(i)} = \sigma(W_e^{(i)} h_j + b_e^{(i)}) \quad (3)$$

Where the intermediate hidden state h_j combines token and entity information, and σ is a nonlinear activation function, for some token in the sentence, which has no corresponding entity in the knowledge graph, only the second term is removed in the intermediate hidden state h_j computationally.

The token and entity embedding computed in the last information fusion layer (layer N) will be used as the final output embedding of the K-Encoder.

The KGBERT model constructs a language model that generates semantic representations containing domain knowledge by integrating external entity knowledge based on the knowledge graph.

B. Model Structure

In this paper, Chinese EMR named entity recognition is regarded as a sequence annotation task.

The specific process of named entity recognition is as follows: in the pre-training stage, firstly, the electronic medical records in the form of free text input at character level are preprocessed and annotated to obtain character vectors; and then the external knowledge graph is used for information fusion to obtain the character vector representation with knowledge. In the training phase, the text context information is firstly extracted by a bidirectional encoder; then the global attention is calculated to incorporate the full text information for the character vectors; and the above two steps are iterated until the end of the training according to the score as shown in Figure 3.

The model consists of two parts: the word embedding model and the sequence annotation model (Bidirectional coding layer, attention layer and decoding layer), as shown in Figure 4.

1) Word Embedding Model: First, the electronic medical record (EMR) text is converted into a vector representation suitable for deep learning, and the KGBERT model embedded in the knowledge graph is used as a word embedding tool to represent the input EMR text sequence in terms of words, phrases, sentences, and knowledge layer vectors.

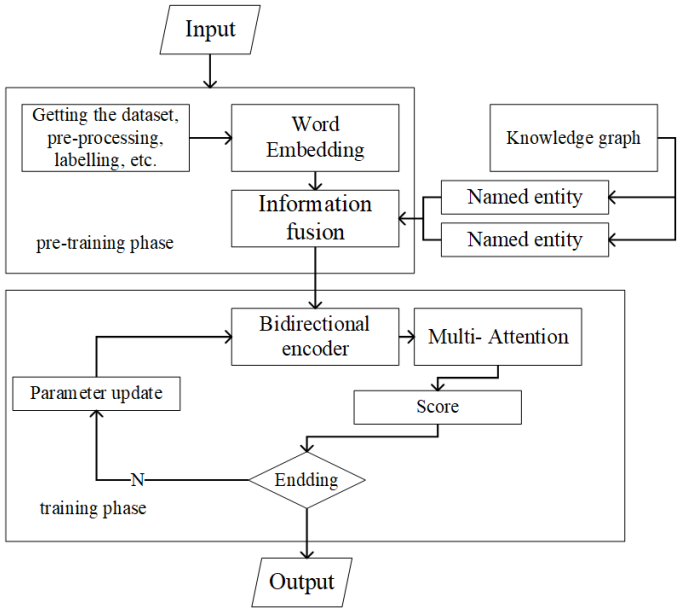

Fig. 3. Overall Process of Named Entity Recognition

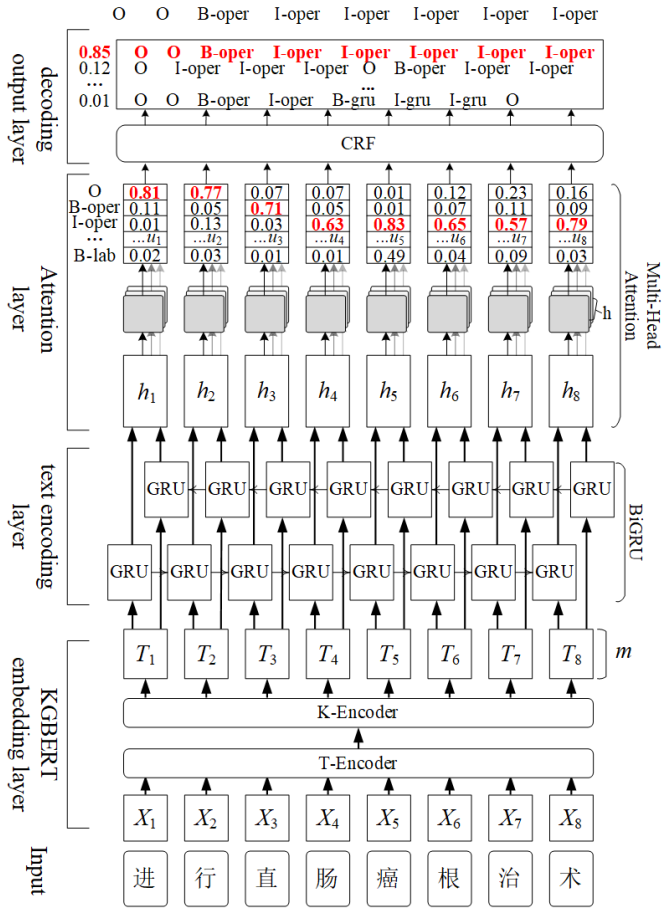

Fig. 4. Structure of KGBERT-BiGRU-Att-CRF

The underlying text encoder T-Encoder of KGBERT first combines two special tokens [CLS] and [SEP], retrieves the full-text character word vectors from the query word vector table, and then parameterizes and trains the sentence-level vectors and position vectors through pre-training to obtain the word vector sequence representing the text. Then, the upper-level knowledge encoder K-Encoder of KGBERT integrates additional text-oriented medical knowledge information into

the lower-level text information, thereby obtaining a text vector sequence containing medical knowledge. The text vector representation of the input text sequence based on the KGBERT model simultaneously contains the contextual semantic information of characters and external medical domain knowledge information, while also addressing the issue of polysemy in Chinese text and the differences between medical domain and general domain word vector representations.

2) Bidirectional Eoding Layer: Since the information to be predicted at a certain position of the text in the Chinese EMR is determined by both the feed-forward information and the posterior information, the text coding layer mainly relies on the BiGRU model to perform bidirectional text coding on the sequence of word vectors containing the knowledge information in order to extract the contextual information of the word vectors.

BiGRU uses the reset gate and update gate structure to update the information input from KGBERT in both directions and add the feature sequences in both directions to obtain the feature word vector sequence containing contextual information, which incorporates contextual information for the representation of word vectors.

3) Attention Layer: The core of the Attention mechanism is to filter the most critical information related to the current task goal. The Attention layer filters a sequence of word vectors containing contextual information, incorporates global semantic information into the word vectors, and assigns higher weights to word vectors related to medical entities.

4) Decoding Layer: The CRF model provides globally optimal and logically sound output sequences by using a transfer score matrix to capture interdependencies between labels and increase constraints between labels.

The multi-information entities in the knowledge graph are used as external knowledge to improve the linguistic representation, so that the model conforms to human language not only at the word, lexical, and syntactic levels, but also at the level of knowledge in the medical domain.The KGBERT-BiGRU-Att-CRF model obtains the word vectors containing the medical knowledge through the KGBERT model, extracts the contextual information through BiGRU, and performs the feature encoding, using the attention mechanism to obtain the global information into the feature vector and generate the labelled sequence, and finally, relying on the constraints of CRF at the decoding layer to solve the dependency problem between the labels to achieve the task of labelling the sequence.

IV. Experiment

A. Dataset

The datasets used for comparison tests in this paper are derived from more than 1,000 entity recognition datasets provided by Task I of CCKS2019 of the National Conference on Knowledge Graph and Semantic Computing and 1,050 medical record datasets with annotations provided by Task III of CCKS2020 [18]. The two parts of the dataset are subjected to the removal of the number of duplicated entities. There are six categories of Disease, Image, Laboratory, Operation, Drug, and Anatomy. The detailed statistics of the number of entities are shown in Table I.

TABLE I
Statistical Results For Different Entity Types

Type	Disease	Image	Laboratory	Operation	Druy	Anatomy
Train set	4314	1168	634	1485	1057	2933
Test set	682	222	193	140	263	447

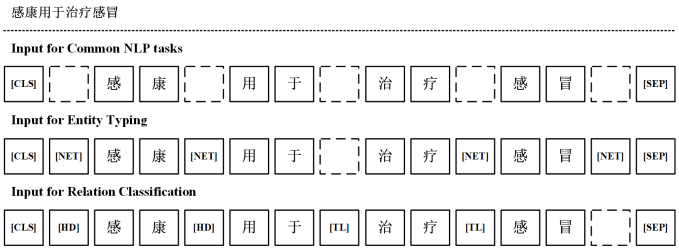

Fig. 5. Different Tokens in Pre-training

B. Pre-training

In order to better incorporate the knowledge information, a new design of the training task is required. In addition to BERT's original masked language model (MLM) and the next sentence prediction (NSP) as pre-training tasks, the model also introduces a new training task called denoising entity auto-encoder (dEA), which does the job of randomly masking out entities and asking the model to predict the most probable entity from a given sequence of entities based on entity-aligned tokens, with the predicted probability distribution computed by the equation4.

$$p(e_j|w_i) = \frac{exp(linear(w_i^o) \cdot e_i)}{\sum_{k=1}^{m} exp(linear(w_i^o) \cdot e_k)} \tag{4}$$

Where linear() represents the linear layer and the loss function for dEA prediction can be calculated by cross entropy.

In practice, the token-entity alignment sometimes has some errors, so this dEA task also introduces some errors artificially. (1) The model will transpose the entity aligned with the token to another random entity with a probability of 5%, which is done in order to be able to correct some cases of incorrect token-entity correspondences. (2) The model will have a probability of 15% to MASK out the correspondence between the token and entity, this is done so that the model can correct some cases where the token-entity pair is not extracted. (3) The remaining probability that the token-entity pairs will not change is to train the model to integrate the entity information into the token representation for better language understanding.

For the fine-tuning process of downstream tasks, the model designs different tag tokens to adapt to different tasks, which means that the input training varies depending on the task. Among them, [ENT], [HD], and [TL] are specific tag tokens. In order to align different types of inputs, placeholders (dashed boxes) are used for corresponding ordinary tasks.Different tokens in pre-training are shown in Figure 5.

The initialisation parameters of the model are adopted from the BERT model due to the high cost of training the model from scratch. As for the training data, the model can't just use unsupervised plain text data anymore, it also needs to extract and align the entities in the text data. The model in the paper adopts the domestic open source diabetes-related

electronic medical record data as the corpus, and aligns the text and entities therein, and then uses the TransE algorithm to transform the entities into vectors, and finally the aligned tokens and entity vectors can be input into the model for training.

For the BERT model, its masking strategy is based on words, and such a strategy is not conducive to the learning of knowledge information, especially for the Chinese language model, it masks off only some words, so it learns more relationships between words during training, while for the model in this paper, it also masks off some consecutive tokens, so that the model learns the relationships between entities besides the relationships above, it also learns the relationships between entities, and this is called knowledge information.

C. Comparative Experiment

In order to verify the effectiveness of the proposed model in the Chinese EMR named entity recognition task, two comparison experiments are used in this section, one of which compares the current mainstream named entity recognition models, such as the LSTM-CRF model, BiLSTM-CRF model, BiGRU-CRF model, etc., to the KGBERT-BiGRU-Att-CRF model proposed in this paper. model; and its second word embedding model BERT is compared with the KGBERT model proposed in this paper [19]. Experiment one based on the same EMR dataset for experiments, Experiment two is based on the same generic corpus dataset that does not contain medical domain knowledgeand. The combined precision (P), recall (R) and F1 values are used to evaluate the experimental results. The experimental results of each model on the CCKS dataset are shown in Table II and Table III.

D. Experimental Results and Analysis

1) Experiment one: (1) Compare the results of experiments of BiLSTM-CRF model with LSTM-CRF model.BiLSTM-CRF model improves 0.96%, 1.21%, and 1.09% compared to LSTM-CRF model, respectively.Compared with unidirectional structure, BiLSTM can merge the forward and backward information and extract contextual features from text, which improves the recognition ability of the model.

TABLE II
RESULTS OF DIFFERENT RECURRENT NEURAL NETWORK MODELS

Models	Precision	Recall	F_1
LSTM-CRF	79.41	79.24	79.32
BiLSTM-CRF	80.37	80.45	80.41
BiGRU-CRF	80.65	80.57	80.61
BERT-BiGRU-CRF	82.94	83.83	83.13
BERT-BiGRU-Att-CRF	84.05	84.17	84.11
KGBERT-BiGRU-Att-CRF	87.84	86.10	86.97

TABLE III
RESULTS OF DIFFERENT RECURRENT NEURAL NETWORK MODELS

Models	Precision	Recall	F_1
KGBERT	84.68	84.13	84.41
BERT	84.59	84.02	84.31

(2) Compare the experimental results of BiGRU-CRF model with BiLSTM-CRF model. In the experiments BiGRU-CRF model has little improvement over BiLSTM-CRF model, but the training time of BiLSTM-CRF model is shortened by about 25% on average, which indicates that the training time can be shortened dramatically and the time cost can be saved when the recognition effect is similar.

(3) Comparing the experimental results of the BERT-BiGRU-CRF model with the BiGRU-CRF model, the BERT-BiGRU-CRF model improves the accuracy by 2.29%, the recall by 2.76%, and the F1 value by 2.52% compared to the BiGRU-CRF model. It shows that the pre-trained language model BERT can effectively fuse multiple information such as vocabulary, sentence and location to provide vector encoding with semantic features for downstream tasks.

(4) Comparing the experimental results of the BERT-BiGRU-Att-CRF model with the BERT-BiGRU-CRF model. The BERT-BiGRU-Att-CRF model that incorporates the attention mechanism at the text encoding layer, the model improves the three evaluation parameters by 1.11%, 0.84%, and 0.98%, respectively, compared to the BERT-BiGRU-CRF model [20]. It indicates that although BiGRU can extract contextual information in both directions at the text encoding layer, it is unable to capture it over a long distance, whereas the attention mechanism can break through the limitation of the text length and compute the attention value of any two vectors in the sequence, so as to capture the global information and improve the model's ability of extracting textual features.

(5) Comparing the experimental results of the KGBERT-BiGRU-Att-CRF model with the BERT-BiGRU-Att-CRF model, the KGBERT-BiGRU-Att-CRF model embeds the knowledge graph into the BERT model on the basis of the BERT-BiGRU-Att-CRF model, and obtains 87.84% in the experimental accuracy, 86.10% recall and 86.97% F1 value, comparing with the BERT-BiGRU-Att-CRF model without embedded knowledge graph, the accuracy, rate recall, and F1 value are improved, which are 3.79%, 1.93% and 2.86%, respectively. It is fully illustrated that incorporating professional domain knowledge for the BERT model at the embedding layer stage and improving linguistic representation with multi-information entities in the knowledge graph as external knowledge make the BERT model conform to human language not only at the level of words, phrases, and grammars, but also at the level of professional domain knowledge, so as to improve the recognition effect of named entities in specific domains.

Combined with the analysis of the results of several comparative experiments mentioned above, the KGBERT-BiGRU-Att-CRF model proposed in this paper performs better on the task of named entity recognition in electronic medical records and has better recognition effect compared to the existing mainstream models for named entity recognition.

2) Experiment two: As can be seen from Table 3, the results achieved by the KGBERT model on the generic domain are comparable to those of the BERT model, which on the one hand indicates that no external knowledge embedding is needed for semantic representation in the generic domain because the generic domain does not contain domain-specific

knowledge structures, and the BERT model itself has a strong semantic representation capability, which can fully satisfy the requirements for recognising information such as place names and people's names. On the other hand, the embedded KGBERT model does not lose the original text information after heterogeneous information is embedded, and even if there is no matching entity input in the knowledge graph, the KGBERT model can still rely on extracting the text's own features to complete the task of recognising the named entities in the generic domain.

V. Conclusion

Aiming at the problem that character vectors are not integrated with professional domain knowledge, this paper proposes a KGBERT-BiGRU-Att-CRF model that integrates the attention mechanism and knowledge graph. The fusion of entity knowledge and entity relations in electronic medical records enables the model to obtain better annotation sequences from both text-level and knowledge-level granularities, thus achieving improved recognition of text by the model. The accuracy, recall, and F1 value of the model are tested to be 87.84%, 86.10%, and 86.97%, respectively, which are 3.79%, 1.93%, and 2.86% higher than that of the unembedded knowledge graph, proving the effectiveness of knowledge embedding.

References

[1] D. CM, C. EG, R. SR, D. K, and F. T. et al., "Electronic Health Records in Ambulatory Care–a National Survey of Physicians," *2008 Massachusetts Medical Society*, vol. 359(1), pp. 50–60, 2008.

[2] L. Y, W. X, H. L, Z. L, L. H, and X. L. et al., "Chinese Clinical Named Entity Recognition in Electronic Medical Records: Development of a Lattice Long Short-Term Memory Model with Contextualized Character Representations," *JMIR Med Inform. 2020 Sep 4*, vol. 8(9), p. e19848, 2020.

[3] B. Ji, S. Li, J. Yu, R. Liu, J. Tang, Q. Li, and W. Xu, "A BiLSTM-CRF Method to Chinese Electronic Medical Record Named Entity Recognition," *Proceedings of the 2018 International Conference on Algorithms, Computing and Artificial Intelligence*, 2018.

[4] G. Wu, G. Tang, Z. Wang, Z. Zhang, and Z. Wang, "An Attention-Based BiLSTM-CRF Model for Chinese Clinic Named Entity Recognition," *IEEE Access*, vol. 7, pp. 113 942–113 949, 2019.

[5] "A dictionary-guided attention network for biomedical named entity recognition in Chinese electronic medical records," *Expert Systems with Applications*, vol. 231, p. 120709, 2023.

[6] S. Morwal, N. Jahan, and D. Chopra, "Named Entity Recognition using Hidden Markov Model (HMM)," *International Journal on Natural Language Computing*, vol. 1, pp. 15–23, 12 2012.

[7] W. Khan, A. Daud, K. Shahzad, T. Amjad, A. Banjar, and H. Fasihuddin, "Named Entity Recognition Using Conditional Random Fields," *Applied Sciences*, vol. 12, no. 13, 2022. [Online]. Available: https://www.mdpi.com/2076-3417/12/13/6391

[8] T. Gui, R. Ma, L. Zhao, Y.-G. Jiang, and X. Huang, "CNN-Based Chinese NER with Lexicon Rethinking," 08 2019, pp. 4982–4988.

[9] Y. Zhang and J. Yang, "Chinese NER Using Lattice LSTM," *ArXiv*, vol. abs/1805.02023, 2018.

[10] H. Wang, W. Yang, W. Feng, L. Zeng, and Z. Gu, "Threat intelligence named entity recognition techniques based on few-shot learning," *Array*, vol. 23, p. 100364, 2024.

[11] G. Yang and H. Xu, "A Residual BiLSTM Model for Named Entity Recognition," *IEEE Access*, vol. PP, pp. 1–1, 12 2020.

[12] L. Jimmy, K. Nongmeikappam, and S. K. Naskar, "BiLSTM-CRF Manipuri NER with Character-Level Word Representation," *Arabian Journal for Science and Engineering*, vol. 48, pp. 1715–1734, 2022.

[13] J.-j. Guo, S.-p. Wang, C.-h. Yu, and J.-y. Song, "Chinese POS Tagging Method Based on Bi-GRU+CRF Hybrid Model," in *Advances in Intelligent Networking and Collaborative Systems*, F. Xhafa, L. Barolli, and M. Greguš, Eds. Cham: Springer International Publishing, 2019, pp. 453–460.

[14] C. Song, Y. Xiong, W. Huang, and L. Ma, "Joint Self-Attention and Multi-Embeddings for Chinese Named Entity Recognition," in *2020 6th International Conference on Big Data Computing and Communications (BIGCOM)*, 2020, pp. 76–80.

[15] D. Zhao, X. Chen, and Y. Chen, "Named Entity Recognition for Chinese Texts on Marine Coral Reef Ecosystems Based on the BERT-BiGRU-Att-CRF Model," *Applied Sciences*, vol. 14, p. 5743, 07 2024.

[16] C. Sun, Z. Yang, L. Wang, Y. Zhang, H. Lin, and J. Wang, "Biomedical named entity recognition using BERT in the machine reading comprehension framework," *Journal of Biomedical Informatics*, vol. 118, p. 103799, 2021.

[17] J. Devlin, M.-W. Chang, K. Lee, and K. Toutanova, "BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding," in *Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long and Short Papers)*. Minneapolis, Minnesota: Association for Computational Linguistics, Jun. 2019, pp. 4171–4186.

[18] X. Li, Q. Wen, H. Lin, Z. Jiao, and J. Zhang, "Overview of CCKS 2020 Task 3: Named Entity Recognition and Event Extraction in Chinese Electronic Medical Records," *Data Intelligence*, vol. 3, no. 3, pp. 376–388, 09 2021.

[19] P. Lu, C. Shao, S. Deng, J. Zeng, and K. Lin, "IBNNER: A Biaffine Model-Based Chinese Nested Named Entity Recognition Method for Medical Texts," *IAENG International Journal of Computer Science*, vol. 51, no. 11, pp. 1686 – 1699, 2024.

[20] Y. Ma, M. Wen, and H. Liu, "Named Entity Recognition Model of Traditional Chinese Medicine Medical Texts based on Contextual Semantic Enhancement and Adversarial Training," *IAENG International Journal of Computer Science*, vol. 51, no. 8, pp. 1137 – 1143, 2024.