Shaoze F et al. (2024) APPLICATION OF COMPUTER VISION TECHNOLOGY IN ATHLETES CLINICAL MEDICAL RECORD IMAGE BIOMETRIC EXTRACTION. Revista Internacional de Medicina y Ciencias de la Actividad Física y el Deporte vol. 24 (96) pp. 520-539.
DOI: https://doi.org/10.15366/rimcafd2024.96.032

ORIGINAL

APPLICATION OF COMPUTER VISION TECHNOLOGY IN ATHLETES CLINICAL MEDICAL RECORD IMAGE BIOMETRIC EXTRACTION

Fu Shaoze*, Gao Gang, Zhao Hongmei

College of Physical Education, Xinjiang Normal University, Urumqi 830054, Xinjiang Uygur Autonomous Region, China
E-mail: a15515706123@163.com

Recibido 20 de octubre de 2023 **Received** October 20, 2023
Aceptado 20 de puede de 2024 **Accepted** May 20, 2024

ABSTRACT

With the development of clinical informatization technology, a large amount of data resources has been accumulated in the medical field, among which Athletes' electronic medical records (EMRs) are one of the important data sources of clinical informatization and contain rich medical knowledge, how to obtain valuable biometric features from these data has become the basis of medical intelligence research. Therefore, this paper takes structured text as the entry point, and firstly, the research status of information extraction is elaborated. Moreover, in the work of e-medical biometrics recognition, medical text is added to the language model BERT for pre-training and fine-tuning, and a multi-head self-attention mechanism is introduced to incorporate a bidirectional LSTM model for feature extraction, and biometric features are extracted by using a stochastic conditional field as a classification constraint. This design bypasses the character-level image segmentation step for text line images, thus avoiding the overall accuracy degradation caused by the backward accumulation of errors in character segmentation. Finally, the experimental results show that the model can effectively accomplish the related biological feature extraction tasks.

KEYWORDS: Biological Features; Electronic Medical Record Image; Automatic Extraction; Computer Vision

1. INTRODUCTION

With the booming development of Internet information technology (Choi et al., 2020; Ratta et al., 2021; Singh et al., 2020), artificial intelligence technology (Benjamens et al., 2020; Meskó & Görög, 2020; Mirchi et al., 2020) with computer vision as the main driving force has been applied more deeply in many fields, and the intelligence in the medical field has become one of the key areas of researchers' attention. As one of the most important and complex data in the medical and healthcare field, the focus of its processing technology research has gradually transitioned from simple medical entity extraction (Bhatia et al., 2019; Ji et al., 2019; Zhang et al., 2019), entity normalization and other data standardization-oriented technologies to automatic extraction of biometric features that are closer to the needs of clinical assisted decision-making, such as complex entity extraction, entity relationship mining and knowledge graph construction. At now, the healthcare industry has transitioned into the stage of clinical informatization, resulting in the accumulation of a substantial volume of medical data. Medical data refers to the comprehensive collection of information produced by healthcare professionals during the diagnosis and treatment of patients (Desai et al., 2020; Nirjhor et al., 2021). This encompasses various types of data, such as basic patient information, athletes' electronic medical records, diagnostic and treatment data, medical imaging data, medical management data, economic data, and medical equipment and instrument data. Among these, athletes' electronic medical records play a crucial role in clinical informatization and serve as a significant resource of medical data (Langlotz et al., 2019; Patel, 2019; Willemink et al., 2020). Due to the traditional paper medical records are not easy to save and retrieve the characteristics of such medical records are not conducive to the subsequent application and research, as the medical structure is gradually scientific, standardized storage and management of complex medical information, which is mainly based on the patient as the main body, in accordance with certain rules of analysis, collation, storage and sharing, more complete and long-lasting preservation of patient's medical information. Based on athletes' electronic medical records, hospitals have opened up administrative information systems such as outpatient, inpatient, pharmacy and financial systems, as well as diagnostic information systems such as examination, ICU guardianship and surgical anesthesia management, to realize information sharing and interoperability. The realization of these systems is closely related to information extraction technology (Han et al., 2019; Saadi & Belhadef, 2020; Yang et al., 2019). The automated extraction of biometric features is a crucial undertaking within the field of information extraction. This work holds great importance for both the advancement of research in information extraction techniques and their practical implementation. The named entity recognition task for athletes' electronic medical records primarily focuses on identifying entities related to examinations, surgeries, diseases, symptoms, drugs, and sites. Entity relationship extraction refers to the process of extracting organized information from text that lacks structure. The primary objective of this task is not only to extract relationships between entities mentioned in the text, but also to ascertain the nature of these interentity relationships. For instance, in the sentence "A patient with diabetes mellitus for 3 years was administered metformin treatment, resulting in satisfactory glycemic control," entity relationship extraction would involve identifying and extracting the relationship between the patient and the treatment administered. The association between the medical condition known as "diabetes" and the pharmaceutical agent "metformin" is characterized by the amelioration of the disease through therapeutic intervention. The investigation of electronic medical record entity recognition and entity connection extraction holds significant value as a crucial means of medical information. This research has the potential to contribute to the advancement of the medical sector to a certain degree (Dai et al., 2019; Gonzalez-Agirre et al., 2019; Zhao et al., 2019). Therefore, this paper takes diabetes EHRs and public EHR datasets as research objects, firstly, obtains real EHRs from hospitals, and carries out a series of analysis, annotation and review to form EHR entity and relationship datasets. On the basis of the research dataset, for the named entity recognition task, migration learning is used to improve the performance of the model, for the entity-relationship extraction task, external knowledge is organized, medical text is added to the language model BERT for pre-training fine-tuning (Ji et al., 2021; Rasmy et al., 2021), and the mechanism of multi-head self-attention is introduced, and a bidirectional LSTM model (Rajeev et al., 2019) is fused for feature extraction, and a stochastic conditional field is used as a classification constraint for biological feature extraction, which can not only improve the ability of the model to automatically obtain medical entities and relationships, but also provide a data base for the construction of medical knowledge base and medical knowledge graph, and obtain valuable biological features from these data to become the basis of medical intelligence research. The main contributions are as follows: (1) An EMRs-BERT model is constructed for automatic extraction of biometric features. In addition to this, in biometric feature relationship extraction from athletes' electronic medical records, a convolution-based Conv-BiLSTM model is proposed for sentence semantic feature extraction by combining the current powerful pre-trained language model BERT to construct the initial semantic representations, and combining it with the character-level attention mechanism to form the semantic features. (2) The proposed algorithm bypasses the character-level image segmentation step for text line images, thus avoiding the overall accuracy degradation due to the backward accumulation of errors in character segmentation, and the experimental results show that the model is able to effectively accomplish the task of relevant biological feature extraction.

2. Related Works

2.1 Named entity recognition for athletes' electronic medical records

Within the realm of medicine, the initial stage of EHR biometric feature extraction involves the detection of named entities in electronic health records

(EHRs). Subsequently, the extraction of entity connections can be pursued building upon this foundational phase. The primary objective of this system is to identify entities, such as diseases, treatments, symptoms, and places, inside medical record texts. Currently, the process of performing named entity recognition for electronic health records (EHR) primarily involves the utilization of dictionary and rule-based methods, statistical learning-based methods, and computer vision-based methods. Among these approaches, computer vision methods have proven to be more effective in this task. In particular, the introduction of Pre-Trained Models (PTM) has significantly enhanced the performance of EHR named entity recognition to a certain degree. The next section provides a description of the various research methods that are relevant to the topic at hand. Dictionary-based approaches are based on medical terminology dictionaries and utilize dictionary matching methods for named entity recognition. In the medical field, researchers use the Integrated Medical Language System (UMLS) metadictionary terminology to extract clinical medical concepts by string matching and analyze the corpus. However, it is often used as a feature to improve the recognition performance of the model due to reasons such as the large amount of human and material resources required to construct, maintain and update the annotated corpus. The rule-based approach uses matching for named entity recognition based on the constructed rule templates. In the medical field, different categories of medical texts have their own characteristics, and the rules constructed for one category of medical texts cannot be directly extended to other categories of medical texts. Therefore, rule-based methods are often used as feature information to assist the automatic extraction of biometric features related to athletes' electronic medical records. Statistical learning-based methods construct statistical models based on data, and then predict and analyze the data. In the medical named entity recognition task, statistical learning methods can be classified into three categories, supervised learning, semi-supervised learning and unsupervised learning, according to whether the corpus used by the model contains labels or not. Among them, supervised learning is trained using labeled datasets to transform the named entity recognition task into a classification problem. The training of supervised learning models relies on high-quality labeled datasets, and due to the lack of labeled datasets for EHRs, some studies have also used semi-supervised learning methods for named entity recognition on EHRs. Unsupervised learning methods cluster entities with similar format or content in the named entity recognition task mainly through similar contextual information. Computer vision Based Approach. In recent years, researchers and scholars have begun to employ neural networks for EHR named entity recognition research. The researchers used unsupervised learning methods to extract word embeddings from large-scale unlabeled corpus to significantly improve the recognition performance of randomly embedded Deep Neural Network (DNN) for medical entities in Chinese clinical documents, which verified the effectiveness of unsupervised feature learning.

A supervised, multi-task Convolutional Neural Networks (CNN) model is also proposed and applied to a biomedical named entity dataset, which is shown to be effective for the medical named entity recognition task by comparing the single-task and multi-task models. At present, the limited availability of corpora significantly affects the training process of electronic health record (EHR) entity recognition models. Given the challenges associated with acquiring electronic health records (EHRs), a prominent area of research pertains to leveraging existing data to enhance the efficacy of named entity recognition. This entails facilitating the exchange of data within a specific domain or across disparate domains. To enhance the training efficacy of the target model, this study suggests utilizing diverse medical data sources, as opposed to relying solely on a single source of athletes' electronic medical records. Multi-source medical data offers a broader spectrum of medical knowledge. Consequently, this paper introduces a neural network-based model for named entity recognition in athletes' electronic medical records. By incorporating medical knowledge from various multi-source medical datasets, this approach addresses the scarcity of labeled electronic medical record data to some extent. The insufficiency of medical record labeling data negatively impacts the model's learning capability.

2.2 Biometric Feature Relationship Extraction from Athletes' electronic medical records

EHR biometric feature extraction is another key task in EHR information extraction, which not only needs to determine whether there is a relationship between the entities in the EHR text, but also needs to identify which class of relationship it belongs to, at present, the relationship categories are mainly predefined according to the entity types, and the relationship extraction is converted into a multiclassification task, and with the development of computer vision, a joint extraction model emerges, which considers the two tasks as one complete process. According to the training method of the model, the EHR entity relationship extraction methods can be divided into three categories: symbiosis-based methods, statistical machine learning-based methods and computer vision-based methods. Among them, symbiosis-based methods are statistical learning methods based on the assumption that the higher the frequency of two entities appearing at the same time, the stronger the relationship is. Statistical machine learning based approaches mainly employ a labeled corpus to train out a classifier for relation classification. Among the machine learning methods used for the task of relation extraction, SVM-based methods are the most widely used.The researchers used a supervised machine learning approach to extract the relations between genetic diseases in biomedical literature.The approach utilizes an integrated SVM to train a rich set of features containing conceptual, syntactic, and semantic attributes, which are jointly learned with word embeddings, and achieves better results. Computer vision Based Approach. With the increasing application of computer vision, researchers have started to utilize computer vision for relationship extraction studies. They used CNN to automatically learn features to extract relationships from clinical discharge summaries, reducing the dependence on manually constructed features. With the emergence and development of Attention mechanisms, the introduction of Attention mechanisms in relationship extraction models has received more and more attention from research scholars. In this study, a novel approach is introduced that combines a deep residual network (ResNet) with the Attention mechanism to extract medical conceptual relations from Chinese athletes' electronic medical records. This approach aims to mitigate the adverse effects of noise in the corpus on parameter learning. Additionally, it integrates the character Attention mechanism to enhance the recognition capabilities of various entities. The existing body of research on relationship extraction using computer vision predominantly relies on shallow networks, which are limited in their ability to comprehensively capture the intricate semantic characteristics present in Chinese electronic health record (EHR) texts. The advent of computer vision has led to the utilization of pre-training models, which have enhanced the efficacy of relationship extraction to some degree. However, when employing pre-training models that are trained on extensive natural language text to extract entity relationships from intricate Chinese electronic health records (EHRs) of limited scale, and subsequently fine-tuning them for downstream tasks, there may still arise instances where the models struggle to effectively learn the textual information.

3. Methodology

The process of extracting information from athletes' electronic medical records mainly involves the representation of word vectors and the Transformers framework. In this section, the technical concepts are theoretically elaborated.

3.1 Representation of word vectors

Word embedding is an important concept for transforming text into vectors. The formation of word vectors requires the foundation of word embedding. In NLP, we can only analyze the text by transforming it into vectors. The development of word vectors can be traced back to the original One-hot encoding, then to the text representation based on word embeddings and the subsequent contextualized ELMo text encoding, which are all methods to transform text into word vectors, and the following three methods are introduced.

3.1.1 One-hot encoding representation

The unicode representation is a vector of 1 and 0 numbers used to represent text. After the corpus has been segmented, a corpus of words is obtained, and the number of words in the corpus is the length of the text vector representation. The occurrence of 1 at different positions in the vector means that they correspond uniquely to words in the corpus, thus constituting the index of the query word. For the text vector, the words appearing in the text, the corresponding position of the text vector can be represented by the corresponding word vector, as shown in Figure 1.

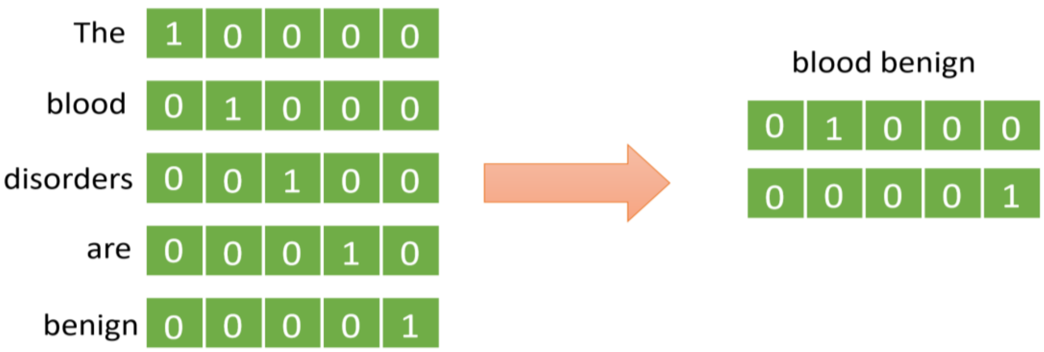

Figure 1: Textual representation of one-hot encoding.

Based on the unique heat encoding, the Term-Frequency-Inverse Document Frequency (TF-IDF) representation has gradually become the main way to represent word vectors. In TF-IDF, Term-Frequency (TF) is the frequency of occurrence of a word. Term-Frequency is the frequency of occurrence of a word in a sentence. If a word occurs very frequently in a sentence, it means that the center word of the sentence is most likely to be this word. The resulting word frequency is expressed by the following equation:

$$TF_{i,j} = \frac{num_{i,j}}{\sum_k num_{k,j}} \qquad (1)$$

where $num_{i,j}$ denotes the number of occurrences of phrase t_i in document j or sentence j, and $\sum_k num_{k,j}$ denotes the sum of the number of occurrences of each individual word across the corpus. Although phrases with a high word frequency reflect importance to some extent, auxiliary words such as "the", "they", "you" also have a high word frequency. These auxiliaries do not refine the important information of a sentence well, so word frequency cannot accurately recognize the importance of a word. Therefore, Inverse Document Frequency (IDF) is derived from it. IDF represents the prevalence of a word. The formula of IDF is as follows:

$$IDF_i = log \frac{D}{1+|1:t_j \in d_j|} \qquad (2)$$

where D denotes the sum of the number of all documents or sentences in the corpus, and $|j: t_{j\in} d_j|$ denotes the number of all documents containing the phrase t_i. If the phrase is included in more documents, the lower the IDF value is, which means that the universality of the word is higher, and thus the discriminative ability of the phrase is weaker. In order to find the keywords, we need in text analysis, TF and IDF are combined to form TF-IDF to filter out the most distinguishable keywords in the specialized field. The TF-IDF equation is as follows:

$$TF - IDF = TF \bullet IDF \quad (3)$$

Using TF-IDF, we obtain an importance score for each word in the corpus, which is used as a vector representation of the word and then combined into a vector representation of the text.

3.1.2 Word vector representation under language model

Although we can use unique thermal encoding or TF-IDF to represent the text as vectors, the semantic information contained in such a representation is meaningless, which only encodes the words and cannot express the correlation between the words, and the unique thermal encoding will have the problem of sparse vectors, which is not conducive to the training of the model. Based on this, the emergence of word embedding can alleviate these problems to a certain extent. Word embedding can map the word vectors of unique heat coding into a relatively low-dimensional vector space according to the rules of semantic information. The mapping in word embeddings is a byproduct of the previously mentioned training language model - weight vectors. In 2013, the Google ai team first proposed the Word2Vec toolkit. Word2Vec consists of two models: the Continuous Bagof-Words (CBOW) model and the Skip-Gram model. In CBOW model, the training task of the model is to predict the probability distribution of the vacant center word by taking the encoding of this paper in the context of the center word as the input, and its model structure is shown in Figure 2.

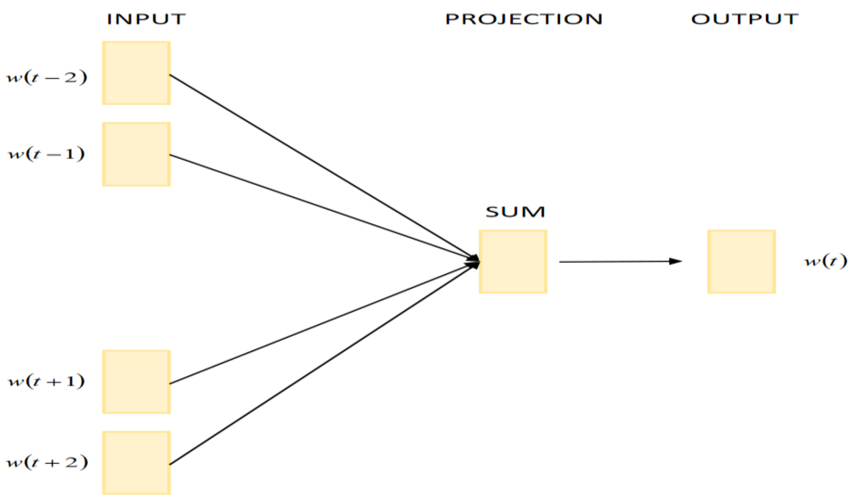

Figure 2: Schematic representation of the CBOW model structure.

After the neural network is trained, the weight matrix is obtained, and this weight matrix is the word vector we use. The word embedding vector is obtained by multiplying the word vector with the weight matrix based on the unique heat encoding of the word vector, and the weight matrix becomes the spatial vector of the word embedding. Similarly, in the Skip-Gram model, the input is just the center word, and the output of the model is the distribution of the contextual text of the input center word.

3.1.3 Word vector representation under neural networks

Although the representation of word embedding solves the high-dimensional problem, its semantic connection still performs poorly in the model, and also generates the ambiguity problem, the meanings of words in different contexts are biased, so the contextual information is important for the expression of word vectors. To solve the ambiguity problem, Allen Institute proposed ELMo text vector representation model in 2018, which is based on the two-layer BiLSTM model. The BiLSTM model recognizes the text as time series data, extracts the special diagnosis of text information from two directions, i.e., fuses the contextual text information, and then combines the word embedding vectors with the word vector weighted sum to get the word vectors of ELMo, and the network structure of ELMo is shown in Figure 3.

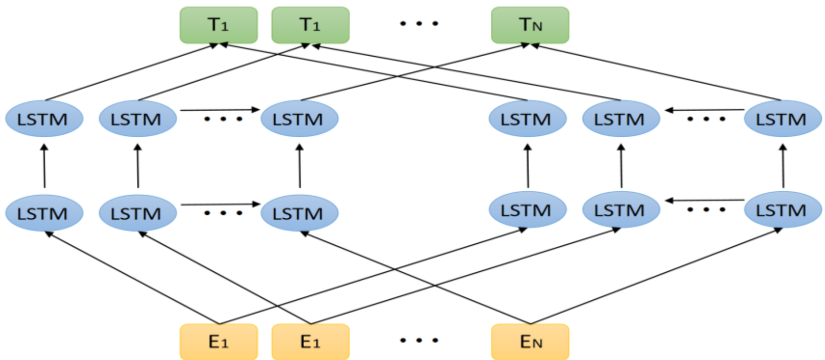

Figure 3: Schematic representation of the structure of the ELMo language model.

The ELMo model is a pre-trained model, and the ELMo word vectors for the same word are different in different downstream tasks. In many of the subsequent tasks, the ELMo model plays a role in the text feature extraction process. The disadvantages of the ELMo model are obvious. The basic framework of ELMo is BiLSTM, which still has the problem of forgetfulness in ultra-long distance feature extraction, and the problem of information redundancy still exists.

3.2 Transformer and attention

3.2.1 Self-attention mechanism

The Transformer framework is a modeling framework proposed by

Google AI team based on the pre-trained language model, and the main modules of the framework are Encoder, Decoder and Attention mechanism. In the Encoder-Decoder framework, Self-attention is the basis of the framework, and is also a very important tool for feature extraction in natural language processing, computer vision and speech recognition. Self-attention represents the degree of attention to the data itself, the higher the degree of attention, the higher the relevance of the information, and the attention mechanism allows the model to focus on the key information to improve the computational efficiency of the model. In this paper, the task of named entity recognition belongs to the task of sequence generation, i.e., the Sqe2Seq task, which needs to focus on the amount of information over a long distance, so the attention mechanism is also the core point of this paper. For the input text vector $X = (x_1, x_2, \dots x_n)$, the formula for calculating the single-headed self-attention score is as follows:

$$Attention(Q, K, V) = Soft\,max\left(\frac{QK^T}{\sqrt{d_k}}\right)V \quad (4)$$

$$Q = X \bullet W^q \quad (5)$$

$$K = X \bullet W^k \quad (6)$$

$$V = X \bullet W^v \quad (7)$$

where $Q = (q_1, q_2, \cdots q_n)$, $Q = (k_1, k_2, \cdots k_n)$, $V = (v_1, v_2, \cdots v_n)$ are all obtained by dot-multiplication based on the input vectors and the corresponding weight matrices. And calculating the similarity score within the text has:

$$score(x_i, x_j) = a_{i,j} = q_i k_j^T \quad (8)$$

The similarity score refers to the similarity of the text components, and the similarity involves the q-vector and k-vector which are the relevant information extracted from the text vectors. The relevance score is obtained and normalized to get the final attention score. The V-vector is a vector of valuable features extracted from the text vectors, and the attention score is multiplied by the V-vector to get the final output of the attention model. The output of the attention can be understood as the hidden state of the cells appearing in the recurrent neural network, which contains the long-distance relevant information. It is worth noting that the three weights in the attention mechanism W^q, W^k, W^v are the parameters that the model needs to learn, and for the case of multiple attention, the updating of weights is also an area that the model needs to take into account. In the framework of Transformer, the mechanism of multi-attention is involved, which is based on the self-attention and adds various W^q, W^k, W^v weights. Thus, different weights extract different information, i.e., the information noticed is not the same, and the final attention score needs to be normalized.

3.2.2 Encoder in Transformer

The Transformer model is commonly used in natural language processing for Sqe2Sqe tasks, such as machine translation, named entity recognition, and human-robot questioning. It is essentially an Encoder-Decoder structure, as shown in Figure 4.

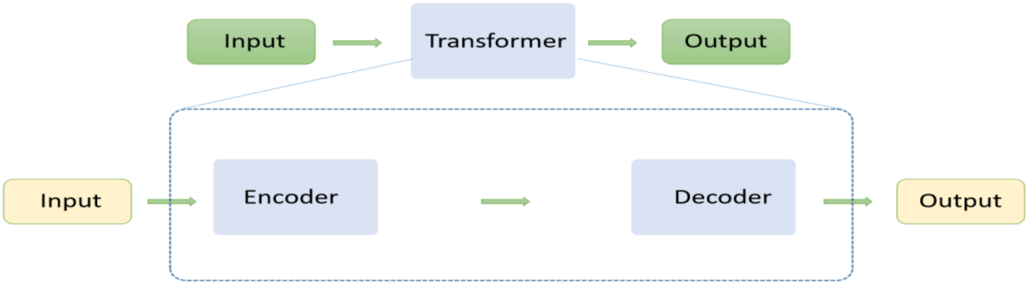

Figure 4: Application of the Transformer Framework to Machine Translation

In the Transformer framework, there are 6 encoders and 6 decoders, and the feature information obtained by the encoders will be sent to the decoders for decoding to accomplish the subsequent tasks. In the Transformer framework, there are four core components: Encoder, Decoder, Position Encoding and Model Output. The encoder encodes the input vectors, and the decoder combines the output vectors of the encoder for subsequent feature processing. The internal structure of encoder and decoder is similar, and its structure is shown in Figure 5.

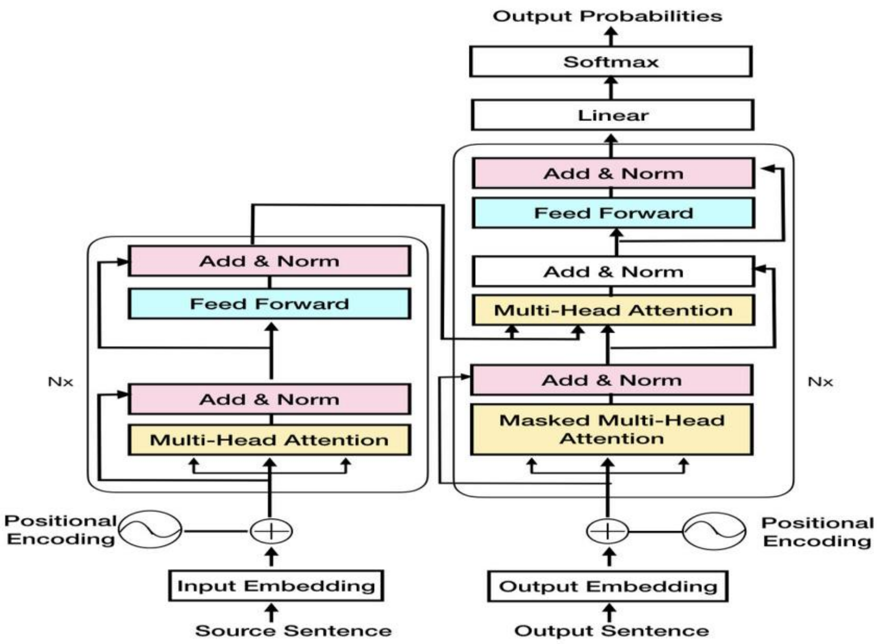

Figure 5: Model structure of encoder and decoder pairs in Transformer

The input of the encoder is the vector obtained by adding the text encoding embedding and positional encoding. Positional encoding mainly identifies the order of each word in the text, and the formula for positional encoding is as follows:

$$PE_{(pos,2i)} = sin\left(\frac{pos}{10000^{\frac{2i}{d}}}\right) \quad (9)$$

$$PE_{(pos,2i+1)} = \cos\left(\frac{pos}{10000^{\frac{2i}{d}}}\right) \quad (10)$$

Where pos denotes the sequential position of the character in the sentence, d denotes the dimension of the text encoding vector. $2i$ denotes the coded position of the even numbered dimensions, the vector X after summing the text and position coding is fed into the Multi-Head Attention module to extract the corresponding attention scores, and then the obtained attention scores are linearly spliced and normalized to obtain the final attention score Z, the matrix Z has the same dimensions as the input X. After that, the X and the attention score matrix are connected to a matrix of residuals, and the mitigated network is normalized to a matrix of residuals. The matrix Z has the same dimension as the input X. The X and the attention score matrices are then residually concatenated and normalized to alleviate the network degradation problem. The normalized vectors are then fed into two fully connected layers, where the activation function of the Feed Forward layer is the ReLU function and the second layer has no activation function. The corresponding equations are as follows:

$$output = max(0, XW_1 + b_1)\, W_2 + b_2 \quad (11)$$

where X is the output of the multi-attention layer. The Add&Norm layer mainly consists of a summing and normalization module, which is calculated as follows:

$$Layer Norm = \big(X + MultiHeadAttention(X)\big) \quad (12)$$

$$Layer Norm = \big(X + FeedForward(X)\big) \quad (13)$$

In summary, the encoder consists of the Multi-attention, Feed Forward and Add Norm layers. In the Transformer framework, the text is encoded by 6 encoders, each encoder takes its input from the output vector of the previous encoder, and the first encoder receives the embedding vector of the text.

3.2.3 Decoder in Transformer

In Transformer, the features obtained after encoding the text by the encoder need to be decoded to accomplish the downstream tasks. The structure of the decoder is similar to that of the encoder, which also contains a Multi-Head self-attention layer, a feed-forward layer, and a Multi-Head attention layer with a mask, as shown in Figure 6.

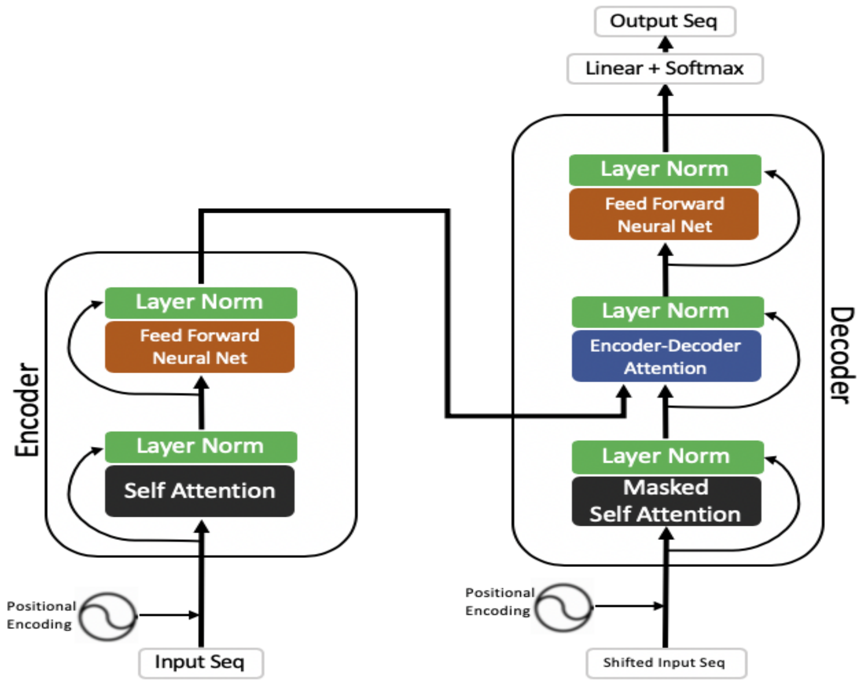

Figure 6: Structure of the decoder in the Transformer framework.

In the second multi-attention computation, the $Q\backslash K\backslash V$ vectors of the input vector Z need to be computed, but the difference here is that the output matrix of the masked multi-attention is used for the computation of the Q matrix, while the final output matrix of the encoder is used for the computation of the K and V matrices. The purpose of this is to share the information encoded in the encoder in order to accomplish the sequence generation task better. After decoding, the decoder produces the final output Z which is fed into the Softmax layer to predict the output word. Because of the mask, Z_0 contains only the information of the first word itself, so it only predicts the generated text at the first position. The information structure of the final output Z is shown in Figure 7.

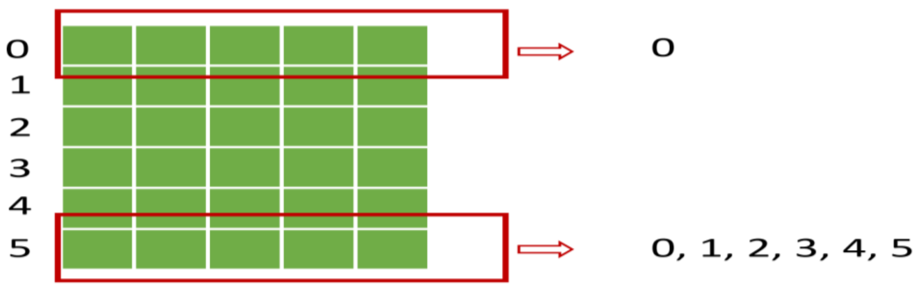

Figure 7: Mask Matrix Workflow.

In conclusion, the encoder and decoder in the Transformer framework can fulfill the task of sequence generation well because of its attention mechanism. In the long text and long sequence tasks, feature extraction is more effective. Based on the framework of Transformer, Google team formally proposed the language model of BERT.

3.3 BERT

The concept of pre-training language model is based on pre-training, the idea of pre-training is based on the model parameters obtained from the previous corpus training, and the trained parameters are used as the base parameters of the autonomous model for model training and optimization. In the field of natural language processing, we input the text parameters of the pre-trained language model as text vector representations into the model, and then fine-tune it with the next task to complete the modeling task. In this paper, we use the BERT language model. The BERT language model utilizes both gap-filling and sentence-pair prediction methods to train the language model to capture both word-level and sentence-level features.

In the fill-in-the-blank training approach, the core idea is similar to predicting masked words in crossword puzzles. In the training model, we feed complete text into the BERT model, the model will mask some words in the sentence with a 15% probability, and all the model has to do is to recover the masked words. In the sentence pair prediction training mode, we input different pairs of sentences to determine whether the two pairs are adjacent or not. The internal structure of BERT is the encoder structure in Transformer, the input text is passed through BERT to get the vector representation of the text, and the text vector of BERT is obtained to complete the subsequent tasks. The BERT model is actually a stack of Encoder encoders, the same as the Encoder encoder in the Transformer, including the Self-Attention and Feed Forward layers to get the final encoding, which can be used as word vectors. The encoding obtained in BERT can be regarded as a certain degree of word embedding, and BERT as a powerful language model, many scholars' subsequent research tasks are based on BERT.

4 Experiment and Results

4.1 Datasets

(1) DEMRC data set: This dataset is constructed from Chapter 3 of this paper, in order to compare the size of the dataset as a whole, the number of corpus and the number of triples in the training, validation and test sets are counted here, which contains a total of 40 sub-relationship categories, among which there are 16 subclasses for entity relationships and 25 subclasses for modifier relationships, and the specific statistics are shown in Table 1.

Table1: Demrc Dataset

TYPE	TRAINING SET	VALIDATION SET	TEST SET
SUB-RELATIONSHIP CATEGORIES	40	40	40
CORPUS	5104	638	638
TERNARY GROUPS	24937	3057	2670

(2) DiaKG data set: This dataset is a Ruijin diabetes research literature dataset centered on the disease and drug name entities, which are given pre-defined relationship types for the entities appearing in the sourced medical resources. It mainly expands the relationship between the two entities, disease and drug name, and contains 15 types of relationships, such as examination method-disease, clinical manifestation-disease, surgery-disease, examination index-disease, frequency of medication-drug name, dosage of medication-drug name, and adverse reaction-drug name. Since the officially provided data only contains training set and the test set is not provided, this paper divides the training set into training set, counts and processes 43 documents containing 3, 498 sentences, among which the documents with relations are 1, 561 sentences, and divides the dataset into training set, test set and validation set according to 8:1:1. The specific statistical results are shown in Table 2.

Table 2: DiaKG DATASET.

TYPE	TRAINING SET	VALIDATION SET	TEST SET
SUB-RELATIONSHIP CATEGORIES	15	15	15
CORPUS	1249	156	156
TERNARY GROUPS	6861	871	909

4.2 Experimental setup

The experimental setup consists of two parts, the experimental environment setup used during the experiment and the hyperparameter setup of the model used, as shown in Tables 3 and 4.

Table 3: Experimental environment setup.

TYPE	PARAMETERS
OS	Ubuntu 16.04
GPU	Nvidia GeForce 2080Ti
RAM	32G
PYTHON	3.7.7
TRANSFORMERS	3.0.2
PYTORCH	1.4.0
CUDA	9.2

Table 4: Hyperparameter settings.

HYPERPARAMETER	VALUE
MAX_LEN	310
BATCH_SIZE	16
DROPOUT	0.5
OPTIMIZER	Adam
LEARNING RATE	3e-5
EPOCH_NUM	30

4.3 Experimental results and analysis

In order to evaluate the effectiveness of the joint entity-relationship extraction model fusing structural information proposed in this chapter on small-scale complex biometric feature extraction tasks, this paper utilizes the DEMRC and DiaKG datasets to train and evaluate the proposed ConvBiLSTM model. The comparative experimental models mainly include BERT-BiLSTM-CRF, BERT-CNN and Word2Vec-LSTM models for comparison experiments.

After pre-processing the labeled dataset and analyzing it, this paper proposes a machine learning based ConvBiLSTM model for information extraction experiments on athletes' electronic medical records. In this paper, the pre-training model is first fine-tuned by adding the electronic medical record corpus to the original pre-training corpus to update the BERT weights of the original pre-training model, so that the pre-training language model can be more effective in encoding text in the medical domain.

Based on this, a word-level attention mechanism is added to increase the weights of keywords involved in the classification task. For sentence-level feature extraction, the gating unit in the BiLSTM model is improved by replacing the weight multiplication with a convolutional operation, so as to better filter and extract semantic information in the sentence. In the feature fusion stage, we introduce Maxpooling to maximize the pooling layer feature vectors for fusion, so as to better complete the subsequent relationship classification task. In the relationship classification layer, since the task is a multi-classification model, we use Softmax and Dropout strategies for relationship classification.

In the comparison experiments of our proposed Conv-BiLSTM model, we choose the current Bert-CNN model and Bert-BiLSTM model for relational classification. In the comparison experiments, the same dataset and the same batch size are used for model training and evaluation, and the highest values are recorded. The results of the comparison experiments are shown in Tables 5, 6 and 7.

Table 5: Comparison of experimental results.

MODEL	PRECISION(%)	RECALL(%)	F1-SCORE(%)
WORD2VEC-LSTM	71.42	72.33	71.49
BILSTM-CRF	74.12	75.32	74.71
BERT-CNN	76.52	76.51	77.04
BERT-BILSTM	77.99	77.58	78.20
CONV-BILSTM	**79.19**	**81.22**	**80.19**

Table 6: Experimental results of biometric feature extraction on DiaKG dataset.

MODEL	PRECISION(%)	RECALL(%)	F1-SCORE(%)
WORD2VEC-LSTM	59.59	24.21	34.59
BILSTM-CRF	60.53	29.91	39.83
BERT-CNN	56.74	31.21	40.27
BERT-BILSTM	51.01	39.53	40.76
CONV-BILSTM	61.20	40.31	45.22

Table 7: Experimental results of biometric feature extraction on DEMRC dataset.

MODEL	PRECISION(%)	RECALL(%)	F1-SCORE(%)
WORD2VEC-LSTM	44.90	60.67	62.41
BILSTM-CRF	64.26	61.44	62.98
BERT-CNN	64.16	62.44	63.29
BERT-BILSTM	63.57	62.52	63.04
CONV-BILSTM	64.60	63.01	63.53

From the experimental data, it can be seen that the results of the classification method combined with the fine-tuned pre-trained language model are higher than the other methods in the comparison experiments, in which the F1 score reaches 80.19%. For the text of athletes' electronic medical records, the length of the text is relatively long compared to the text of microblogging comments and product reviews, which requires the model to be able to extract the semantic information of the text over a long distance.

In addition, for the medical domain, the textual features of the medical domain also need to be extracted by the model, therefore, the pre-training model BERT is fine-tuned in this paper, and the results outperform the other models, with an increase of 2.49% in the F1 score of the Bert-BiLSTM model compared to the one without fine-tuning. Overall, the ConvBiLSTM model proposed in this paper has better performance in both precision and recall, which indicates that the model proposed in this paper is capable of further feature fusion and deeper automatic extraction of clinically relevant biometric features.

5. Conclusion

In this paper, we take structured text as the entry point, firstly, we explain the current research status of information extraction. And in the work of e-medical biometrics, medical text is added to the language model BERT for pre-training and fine-tuning, and introduces the mechanism of multiple self-attention, fuses the bidirectional LSTM model for feature extraction, and performs the biometric feature extraction by using the stochastic conditional field as the constraints for classification.

This design bypasses the character-level image segmentation step for text line images, thus avoiding the overall accuracy degradation caused by the backward accumulation of errors in character segmentation. Finally, the experimental results show that the model can effectively accomplish the related athletes biological feature extraction tasks.

REFERENCES

Benjamens, S., Dhunnoo, P., & Meskó, B. (2020). The state of artificial intelligence-based FDA-approved medical devices and algorithms: an online database. *NPJ digital medicine*, *3*(1), 118.

Bhatia, P., Celikkaya, B., Khalilia, M., & Senthivel, S. (2019). Comprehend medical: a named entity recognition and relationship extraction web service. 2019 18th IEEE International Conference On Machine Learning And Applications (ICMLA),

Choi, N. G., DiNitto, D. M., Lee, O. E., & Choi, B. Y. (2020). Internet and health information technology use and psychological distress among older adults with self-reported vision impairment: case-control study. *Journal of Medical Internet Research*, *22*(6), e17294.

Dai, Z., Wang, X., Ni, P., Li, Y., Li, G., & Bai, X. (2019). Named entity recognition using BERT BiLSTM CRF for Chinese electronic health records. 2019 12th international congress on image and signal processing, biomedical engineering and informatics (cisp-bmei),

Desai, R. J., Wang, S. V., Vaduganathan, M., Evers, T., & Schneeweiss, S. (2020). Comparison of machine learning methods with traditional models for use of administrative claims with electronic medical records to predict heart failure outcomes. *JAMA network open*, *3*(1), e1918962-e1918962.

Gonzalez-Agirre, A., Marimon, M., Intxaurrondo, A., Rabal, O., Villegas, M., & Krallinger, M. (2019). Pharmaconer: Pharmacological substances, compounds and proteins named entity recognition track. Proceedings of The 5th Workshop on BioNLP Open Shared Tasks,

Han, J., Chen, K., Fang, L., Zhang, S., Wang, F., Ma, H., Zhao, L., & Liu, S. (2019). Improving the efficacy of the data entry process for clinical research with a natural language processing–driven medical information extraction system: Quantitative field research. *JMIR medical informatics*, *7*(3), e13331.

Ji, B., Liu, R., Li, S., Yu, J., Wu, Q., Tan, Y., & Wu, J. (2019). A hybrid approach for named entity recognition in Chinese electronic medical record. *BMC medical informatics and decision making*, *19*, 149-158.

Ji, S., Hölttä, M., & Marttinen, P. (2021). Does the magic of BERT apply to medical code assignment? A quantitative study. *Computers in biology and medicine*, *139*, 104998.

Langlotz, C. P., Allen, B., Erickson, B. J., Kalpathy-Cramer, J., Bigelow, K., Cook, T. S., Flanders, A. E., Lungren, M. P., Mendelson, D. S., & Rudie, J. D. (2019). A roadmap for foundational research on artificial intelligence in medical imaging: from the 2018 NIH/RSNA/ACR/The Academy Workshop. *Radiology*, *291*(3), 781-791.

Meskó, B., & Görög, M. (2020). A short guide for medical professionals in the era of artificial intelligence. *NPJ digital medicine*, *3*(1), 126.

Mirchi, N., Bissonnette, V., Yilmaz, R., Ledwos, N., Winkler-Schwartz, A., & Del Maestro, R. F. (2020). The Virtual Operative Assistant: An explainable artificial intelligence tool for simulation-based training in surgery and medicine. *Plos one*, *15*(2), e0229596.

Nirjhor, M. K. I., Yousuf, M. A., & Mhaboob, M. S. (2021). Electronic medical record data sharing through authentication and integrity management. 2021 2nd International Conference on Robotics, Electrical and Signal Processing Techniques (ICREST),

Patel, V. (2019). A framework for secure and decentralized sharing of medical imaging data via blockchain consensus. *Health informatics journal*, *25*(4), 1398-1411.

Rajeev, R., Samath, J. A., & Karthikeyan, N. (2019). An intelligent recurrent neural network with long short-term memory (LSTM) BASED batch normalization for medical image denoising. *Journal of medical systems*, *43*(8), 234.

Rasmy, L., Xiang, Y., Xie, Z., Tao, C., & Zhi, D. (2021). Med-BERT: pretrained contextualized embeddings on large-scale structured electronic health records for disease prediction. *NPJ digital medicine*, *4*(1), 86.

Ratta, P., Kaur, A., Sharma, S., Shabaz, M., & Dhiman, G. (2021). Application of blockchain and internet of things in healthcare and medical sector: applications, challenges, and future perspectives. *Journal of Food Quality*, *2021*(1), 7608296.

Saadi, A., & Belhadef, H. (2020). Deep neural networks for Arabic information extraction. *Smart and Sustainable Built Environment*, *9*(4), 467-482.

Singh, R. P., Javaid, M., Haleem, A., Vaishya, R., & Ali, S. (2020). Internet of Medical Things (IoMT) for orthopaedic in COVID-19 pandemic: Roles, challenges, and applications. *Journal of clinical orthopaedics and trauma*, *11*(4), 713-717.

Willemink, M. J., Koszek, W. A., Hardell, C., Wu, J., Fleischmann, D., Harvey, H., Folio, L. R., Summers, R. M., Rubin, D. L., & Lungren, M. P. (2020). Preparing medical imaging data for machine learning. *Radiology*, *295*(1), 4-15.

Yang, Y., Agarwal, O., Tar, C., Wallace, B. C., & Nenkova, A. (2019). Predicting annotation difficulty to improve task routing and model performance for biomedical information extraction. *arXiv preprint arXiv:1905.07791*.

Zhang, Z., Zhou, T., Zhang, Y., & Pang, Y. (2019). Attention-based deep residual learning network for entity relation extraction in Chinese EMRs. *BMC medical informatics and decision making*, *19*, 171-177.

Zhao, S., Liu, T., Zhao, S., & Wang, F. (2019). A neural multi-task learning framework to jointly model medical named entity recognition and normalization. Proceedings of the AAAI Conference on Artificial Intelligence,