

Article

LSLSD: Fusion Long Short-Level Semantic Dependency of Chinese EMRs for Event Extraction

Pengjun Zhai , Chen Wang and Yu Fang *

Department of Computer Science and Technology, Tongji University, Shanghai 200082, China; 1810369@tongji.edu.cn (P.Z.); wangchen0204@163.com (C.W.)

* Correspondence: fangyu@tongji.edu.cn

Abstract: Most existing medical event extraction methods have primarily adopted a simplex model based on either pattern matching or deep learning, which ignores the distribution characteristics of entities and events in the medical corpus. They have not categorized the granularity of event elements, leading to the poor generalization ability of the model. This paper proposes a diagnosis and treatment event extraction method in the Chinese language, fusing long short-level semantic dependency of the corpus, LSLSD, for solving these problems. LSLSD can effectively capture different levels of semantic information within and between event sentences in the electronic medical record (EMR) corpus. Moreover, the event arguments are divided into short word-level and long sentence-level, with the sequence annotation and pattern matching combined to realize multi-granularity argument recognition, as well as to improve the generalization ability of the model. Finally, this paper constructs a diagnosis and treatment event data set of Chinese EMRs by proposing a semi-automatic corpus labeling method, and an enormous number of experiment results show that LSLSD can improve the F1-value of event extraction task by 7.1% compared with the several strong baselines.



Citation: Zhai, P.; Wang, C.; Fang, Y. LSLSD: Fusion Long Short-Level Semantic Dependency of Chinese EMRs for Event Extraction. *Appl. Sci.* **2021**, *11*, 7237. <https://doi.org/10.3390/app11167237>

Academic Editor: Kao-Shing Hwang

Received: 10 April 2021

Accepted: 4 August 2021

Published: 5 August 2021

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Copyright: © 2021 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

Keywords: event extraction; event argument recognition; long short-level semantic dependency; event argument granularity

1. Introduction

Information extraction is a critical task in natural language processing, which aims at extracted structured information from unstructured or semi-structured texts. Medical research, such as disease source [1,2], disease prediction [3], and clinical decision [4,5], needs to extract structured information from underlying medical corpus. This kind of effective extraction and learning for knowledge has become a significant task. Traditionally, the medical knowledge graph (KG) takes entities as nodes, edges as relations to describe medical concepts. However, in addition to the involved medical concepts, the rich and closely related diagnosis and treatment events (DTE) in electronic medical records (EMRs) also contain dynamic knowledge and disease development logic, which cannot be captured by a fine-grained conceptual structure. Therefore, it is necessary to study the extraction method of DTE [6,7].

At present, automatic content extraction (ACE) [8] is the most famous event extraction conference. It divides event extraction (EE) into two subtasks: trigger word detection and event argument recognition. The event trigger word (ETW) is the word that can indicate the event type in the sentence, and the event argument (EA) is the participant that describes the event. For example, the sentence “On 12 August 2015, the patient underwent radical resection of rectal cancer in our hospital due to rectal cancer” is an operation event, “underwent” is an ETW, and the EAs include “12 August 2015”, “rectal cancer”, “radical resection of rectal cancer”.

Early research on medical DTE extraction mostly utilizes pattern matching [9,10]. Recent studies have shown that the sequence annotation methods based on the deep neural network have positive effects on the medical field [11,12]. These methods have

incorporated the context of the entities in the form of distributed word representation and illustrated the importance of domain knowledge in clinical medical event extraction.

However, the basic-feature based EE methods merely adopt pattern matching or sequence annotation, which have limitations in fully mining the important but implied semantic dependency information contained in and between sentences. As shown in Table 1, both example 1 and example 2 are operation events. The ETWs and EAs contained in this type of event sentence have a certain regularity, such as that the ETWs are words of “underwent”. EAs often contain entity words such as disease and operation name, while example 3 of pathological examination usually takes a long sentence-level test result as one of the EAs, and it is mostly located after the operation event sentence.

Table 1. Examples of event sentences from Chinese present medical history.

<p>Example #1</p> <p>ES¹: On 12 August 2015, the patient underwent radical resection of rectal cancer in our hospital due to “rectal cancer”.</p> <p>ET²: operation event ETW³: underwent</p> <p>EA⁴: 12 August 2015, rectal cancer, radical resection of rectal cancer</p>
<p>Example #2</p> <p>ES: On 1 January 2014, the patient underwent “radical resection of gastric stump cancer” under general anesthesia in our hospital due to “gastric stump cancer”.</p> <p>ET: operation event ETW: underwent under general anesthesia</p> <p>EA: 1 January 2014, gastric stump cancer, radical resection of gastric stump cancer</p>
<p>Example #3</p> <p>ES: On 25 June 2015, the postoperative pathology showed: “moderately differentiated adenocarcinoma of sigmoid colon, infiltrating into the upper mucosa of intestinal wall, without nerve bundle invasion, one intermediate lymph node metastasis (1/7), central lymph node (0/5), para intestinal lymph node (0/2)”.</p> <p>ET: pathological test event ETW: pathology</p> <p>EA: 25 June 2015, moderately differentiated adenocarcinoma of sigmoid colon . . . para intestinal lymph node (0/2)</p>

¹ ES, ² ET, ³ ETW, and ⁴ EA stand for event sentence, event type, trigger, and argument.

In this paper, an elegant framework is proposed to recognize the complex ETWs and EAs in the rich DTEs in the history of present illness (HPI) texts of Chinese EMRs. The key idea of our work is to introduce different feature information to train the ETW classifier according to the distribution of entities and events at different corpus levels, and adopt different modules to realize the recognition of EAs with different granularity. Specifically, the contribution of this paper is three-fold:

1. Filling up the gap in the current representation framework and annotated corpus of HPI event in Chinese EMRs based on the combined source from the ACE task definition and Informatics for Integrating Biology and the Bedside (I2B2) [13], through adopting the reverse maximum matching algorithm performing semi-automatically labeling of the PME corpus of Chinese EMRs.
2. The proposed Chinese HPI event extraction method which fusing long short-level semantic dependency (LSLSD) of corpus not only highlights the semantic dependence between the components in sentences, but also emphasizes the semantic dependence between sentences. Meanwhile, it distinguishes the granularity of EAs at the short word-level and long sentence-level can better capture the semantic information of the corpus and enhance the generalization ability of the model.
3. A series of experiments on the annotated dataset show that the proposed model LSLSD has significant performance in trigger word detection and event argument recognition. It can effectively recognize EAs of different granularity at the same time.

2. Related Works

Event extraction [14] is an information extraction task that can be traced back to the 1980s. With the emergence of big data and the development of text mining and natural language processing, event extraction technology has been popularized. The technology of

event extraction of the English context is mature [15], while that of the Chinese context, especially in clinical diagnosis and treatment events, is relatively limited. Zheng et al. [16] first proposed the language issues regarding Chinese event extraction, and divides event extraction into four sub-tasks of event recognition, event classification, argument recognition, and argument role classification. We follow ACE by dividing the event extraction into two sub-tasks of event trigger word detection and event argument recognition.

2.1. Event Trigger Word Detection

The approaches of the ETW detection task are mainly divided into two categories. One is the traditional pattern matching method based on syntactic analysis or the clustering method. Marco et al. [17] completed ETW detection of the open domain through rules of event extraction established by using syntactic analysis, Bui et al. [10] and Huang et al. [18] are completed ETW detection by building extraction rules according to the characteristics of biomedical text and using the method of joint constraint clustering, respectively. The other one is the current popular machine learning classification methods, which classify the ETWs after extracting them from sentences. For example, Xia et al. [19] detect ETW through multi-type features of lexical, dictionary, and syntactic according to the characteristics of biological literature, Yang et al. [20] and Zheng et al. [21] apply the remote supervision method to automatically label the financial texts from the knowledge base of Chinese financial event to realize the detection of financial ETWs. Jindal et al. [22] adopt the lexical features such as words and semantic relations in clinical terms to identify events. Wei et al. [23] construct a method based on sequence annotation and combine the static pre-training word vector represented by character level words with the dynamic context word representation based on the pre-training language model as the model input. However, most studies detect ETW by only using intra sentence information and ignore the information of inter sentences, while some scholars believe that the event distribution information at the document level can improve the accuracy of ETW detection of long-level semantic dependency. Xia et al. [24] construct event reasoning rules by using the document consistency feature to improve the classification results after recognizing and classifying ETW based on the semi-supervised model. Li et al. [25] utilize the combination semantics of ETWs and document consistency in Chinese training set to improve the performance of ETW detection, the former is used to infer unknown ETWs in the test set, and the latter is used to infer the situation that is difficult to deal with based on feature method.

To make the most of semantic information implied in the corpus, different from the previous models, we designed some extended features based on the basic features according to the characteristics of the data, including intra sentence features, inter sentence features, dependency syntax, and other features, to improve the accuracy of ETW detection model from the semantic perspective of entity information and relationship information.

2.2. Event Argument Recognition

Recently, the most popular methods of EA task have been based on sequence annotation; the model of bidirectional long short-term memory (BiLSTM) combined with conditional random field algorithm (CRF) is most frequently employed in sequence annotation task and has achieved excellent results. BiLSTM can capture useful context information from forward and backward of sentences, while CRF has the advantage of using sentence-level and adjacent labeled information when predicting current label. Xu et al. [26] construct an event model of clinical guidelines. The model extracts treatment events of Chinese clinical guidelines by through Word2Vec, LSTM and CRF technologies. Zeng et al. [27] propose a convolution BiLSTM neural network that combines LSTM and convolutional neural network (CNN) for Chinese event extraction, which can capture both sentence-level and local lexical information without any hand-craft features, in addition to the design of ETW location feature and ETW type feature and connect them with the original word vector. However, these models only extract treatment event contains short word-level arguments and do not divide EAs into long short-level granularity, and the

arguments of long sentence-level are hardly extracted by a single sequence annotation model. Wei et al. [24] propose a multi-classification model based on self-attention, which fully utilizes the features of entity and entity type. However, the model only considers the granularity of the corpus when obtaining the representation vector. Jari et al. [28] focus on the edge detection, entity context features and the relation path feature between entities in the biomedical field. The hybrid method proposed by Xuan et al. [29] combines rule-based and feature-based classifier, which is effective in the long sentence-level EAs on biological EA recognition. We follow a classic method of BiLSTM-CRF [30] in the EA recognition task for short word-level arguments; the differences are that we employ the transformer [31] instead of the original encoder and we integrate extended semantic features and design a method of joint sequence annotation and pattern matching to recognize multi granularity EAs. For the Chinese EMR text purpose, we improve the generalization of the EA recognition model by dividing the EAs into long- and short-level granularity, and joint pattern matching and sequence annotation methods, so that the model can not only accurately identify the short word-level EAs in the corpus, but also effectively recognize the long sentence-level EAs widely existing in Chinese EMRs, to improve the performance of event argument recognition model.

3. Materials and Methods

The event extraction model in this paper is applied to the DTE extraction of the HPI texts in Chinese EMRs. In that sense, we complete the semi-automatic annotation of entity, ETW, and EA according to the annotation standard for entity and event form ACE and I2B2. Moreover, an event extraction model fusion long short-level semantic dependence of corpus has been proposed to capture the feature information of the corpus at different levels. Figure 1 illustrates our framework. To simplify the expression, the common nouns in the paper are abbreviated, as shown in Table 2. We will first introduce the semi-automatic annotation of the corpus.

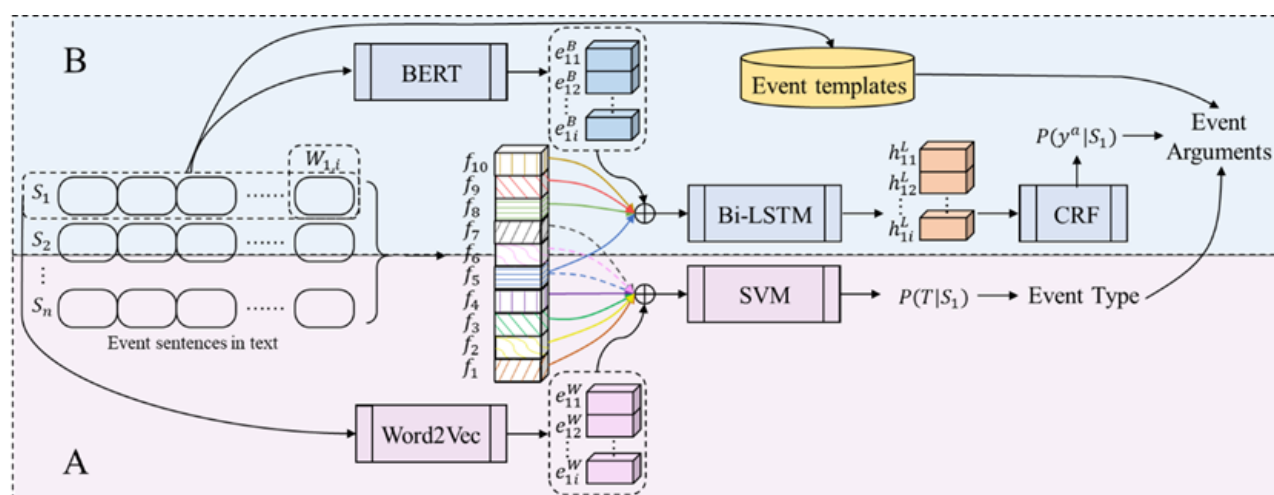


Figure 1. The framework of LSLSD. This model is composed of two modules: trigger word detection (A) and event argument recognition (B) (the pink and blue background color, respectively). $f_1 \sim f_4$ are the basic features commonly used in trigger word detection task; $f_5 \sim f_{10}$ are the extended features proposed in this paper. By fusing these semantic features, the potential association information within and between the event sentences in HPI texts can be fully mined.

Table 2. Abbreviations of nouns are commonly used in the paper.

Nouns	Abbreviations
Diagnosis and Treatment Event	DTE
Event Extraction	EE
Event Trigger Word	ETW
Event Argument	EA
History of Present Illness	HPI

3.1. Corpus Annotation

- **Data preprocessing:** the HPI text of Chinese EMRs records the whole process of patients from illness to hospitalization and treatment, including symptoms, disease development, diagnosis results, detailed process of examination and treatment, etc. Considering the varied writing habits of doctors, the expression of the same disease, operation, or drug name may differ. There are abbreviations and variants which bring difficulties to the follow-up work. Therefore, the data need to be preprocessed. For example, the sentence segmentation for long sentences containing multiple events, deleting sentences unrelated to medical events, unifying medical vocabulary such as disease, operation, drug, standardizing disordered punctuation;
- **Defining event types and element arguments:** I2B2 [13] held in 2012 defines a medical event as status, process, occurrence, and change, which are medically related. According to the definition of event extraction task in ACE and the definition of the medical event in I2B2, considering the content of the data set, the DTE of HPI are divided into six categories: admission, examination, test, treatment, operation, and diagnosis. The test events are subdivided into pathological test and immunohistochemistry, and the treatment events are subdivided into general treatment and chemotherapy. The event types and their event arguments are shown in Table 3;
- **Building a dictionary of candidate ETWs:** the ETW is the core word that indicates the occurrence of a certain event in the sentence, generally verb or gerund. The words with high frequency in each type of events are selected in this paper as the candidate ETWs for each type of event. As shown in Table 3, the dictionary is expanded by using the synonyms of the candidate ETWs in the synonym forest;
- **Defining entity types:** I2B2 and CCKS2017 (http://www.sigkg.cn/ccks2017/?page_id=51 (accessed on 18 January 2020)), respectively, divide medical named entities into three categories (medical problem, examination, treatment) and five categories (treatment, body part, symptom, examination, disease). Symptom entities are divided into body parts and symptom descriptions, and treatment entities into drug and operation in CCKS2018 (http://www.sigkg.cn/ccks2018/?page_id=16 (accessed on 18 January 2020)). Therefore, the entities in the data set are divided into three categories disease, symptom, and treatment. The symptom includes the body part and symptom description, and treatment includes drug, operation, and general treatment;
- **Corpus semi-automatic annotation:** each HPI text contains 30–40 entities; manual annotation is time-consuming and labor-consuming. Therefore, this paper annotates corpus by a semi-automatic method based on a medical dictionary. Firstly, entity dictionary is collected from medical dictionaries (e.g., Dingxiangyuan (<https://portal.dxy.cn/> (accessed on 21 September 2019)), Sogou Medical (<https://pinyin.sogou.com/dict/cate/index/132> (accessed on 1 September 2019)) and 39 Health Net (<http://www.39.net/> (accessed on 18 September 2019))), then a reverse maximum matching algorithm is employed to tag the texts automatically. Secondly, the ETWs, EAs, and ETW types in texts are manually annotated, referring to the definition of the event type. Finally, two doctors are invited to check and correct the above annotation results.

Table 3. Examples of representation framework and trigger works for HPI events.

Event Types	Sub-Types	EAs	Example of ETWs
Admission	Admission	Time, Synonym, Operation name, Disease	admission, see a doctor, hospitalization, receive and cure
Test	Pathological test	Time, Disease, Result of pathological test	pathology, biopsy, slicing, pathological examination
	Immunohistochemistry	Time, Result of immunohistochemistry	immunological staining, staining test
Examination	Examination	Time, Disease, Result of examination	gastroscopy, blood routine, CT, colonoscopy
Treatment	General treatment	Time, treatment method	adjuvant therapy, healing, nursing
	Chemotherapy	Time, Drug	chemotherapy, targeted therapy, radiation
Operation	Operation	Time, Synonym, Operation name, Disease	underwent, surgery, general anesthesia
Diagnosis	Diagnosis	Time, Synonym, Operation name, Disease	diagnosis, judgment, determine, results

3.2. Event Trigger Word Detection Module

Event trigger word detection is a vital step in the event extraction task. Its result directly affects the accuracy of event extraction. In this paper, ETW detection is regarded as a multi-classification task. By analyzing the long short-level semantic dependence intra-event sentence and inter sentences in HPI texts, extended features are proposed to support the training of classifiers, and to obtain more accurate ETW detection results.

Firstly, the words existing in the sentence contained in the candidate ETW dictionary are condemned as the ETW of the sentence. If the words are not found, the ETW will be replaced by words with higher similarity in the sentence and the candidate ETW dictionary. Then, several types of features commonly utilized in ETW detection are selected as the basic features, as shown in Table 4 (1–4). For details, please refer to paper [32].

Table 4. Semantic features faced to event trigger word detection.

Feature Types	Number	Symbols	Names	Describes
Basic feature	1	f_1	Trigger word	Word vector by LTP ¹ word segmentation tool
	2	f_2	Part-of-speech	Part-of-speech tagged by LTP
	3	f_3	context	Word and part-of-speech in window of length 3
	4	f_4	dictionary	Event types corresponding to candidate ETW dictionary
Extended feature	Intra sentence	5	Entity types	Distribution of entities in different types of event sentences
		6	# ² different entity types	Number of entities in different types of event sentences
	Inter sentences	7	Document consistency	Distribution of entity types before and after event sentences

¹ LTP is a language technology platform constructed by Harbin Institute of Technology (<http://ltp.ai/>) (accessed on 6 March 2020)); ²# is the number.

Based on the basic features suitable for general texts, extended features are designed for the entity distribution information in medical event sentences by analyzing the linguistic characteristic of Chinese HPI event sentences, as shown in Table 4 (5,6). Where feature f_5 and f_6 , respectively, refer to the distribution and number of entities in different types of event sentences. These two extended features of short-level semantic dependency are proposed because each type of event sentence does not cover all six types of entities, and the number of entities is not the same; these distribution regulars play an important role in the classification of ETWs.

In particular, there is a correlation between the events because HPI text describes the occurrence and evolution of the patient's disease. For example, the admission event does not occur after the operation event, and the pathological examination event does not occur before the operation event. Therefore, inspired by [24,25], the extended feature of long-level semantic dependence refers to the consistency between event sentences in the HPI text is proposed named document consistency feature, as shown in Table 4 (7). Taking the operation event as an example, Figure 2 statistics the distribution of different event types before and after the operation event in the HPI texts. It can be seen that the probability of admission, diagnosis, and examination event occurring is higher before the operation event, but the probability of pathological test, chemotherapy, and general treatment event occurring is higher after the operation event.

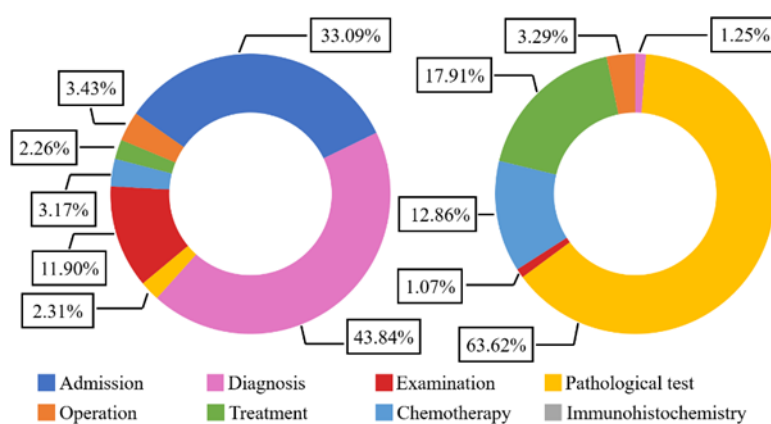


Figure 2. Distribution of event types before and after operation events.

Based on the above semantic features, for a given HPI text R , $R = \{S_1, S_2, \dots, S_n\}$, where $S_n = \{W_{n1}, W_{n2}, \dots, W_{ni}\}$ is the n th event sentence and W_{ni} is the word in the sentence, where i express the sentence length (total number of words). The goal of ETW detection task is to obtain the classification $P(T|S_n)$ of ETW in the event sentence to determine the event type. Firstly, R is transformed into the corresponding word vector $[[e_{n1}^W], [e_{n2}^W], \dots, [e_{ni}^W]]$ through the pre-trained word embedding Word2Vec (<https://github.com/tensorflow/tensorflow/tree/r1.1/tensorflow/examples/tutorials/word2vec> (accessed on 22 April 2020)). Secondly, SVM is adopted as a classifier to classify ETWs. The one versus rest (OVR) method is adopted to realize the multi-classification task of ETW, K SVMs are trained for K ETWs categories, and the j th SVM is used to judge whether the ETWs belong to category j .

Specifically, SVM classifies the given x into category j by searching the maximum value of $W_j^T x + b_j$. To mine the long short-level semantic dependence information in the corpus, we will concatenate the feature vector f_m ($m = 1, 2, \dots, 7$) of input HPI text R and the word vector $[[e_{n1}^W], [e_{n2}^W], \dots, [e_{ni}^W]]$ of R to obtain the semantic word vector S_e :

$$S_e = [[x_1], [x_2], \dots, [x_{n*i}]] = [[e_{n1}^W; f_m], [e_{n2}^W; f_m], \dots, [e_{ni}^W; f_m]] \quad (1)$$

where ‘[.]’ is the concatenation operation, $S_e \in \mathbb{R}^{d_{n*i}}$, $x_{n*i} \in \mathbb{R}^{d_{\hat{x}}}$ ($d_{\hat{x}} = d_{n*i} + d_m$). Therefore, the dual form of the SVM optimization problem can be expressed as:

$$\max : \mathcal{L}_T = \sum_{q=1}^{n*i} \alpha_q - \frac{1}{2} \sum_{q,j=1}^{n*i} \alpha_q \alpha_j y_q y_j K(x_q, x_j) \quad (2)$$

$$s.t. \sum_{q=1}^{n*i} \alpha_q y_q = 0, 0 \leq \alpha_q \leq c \quad (3)$$

where $\alpha_q, \alpha_j \in C$, y_q, y_j are ETW types of x_q and x_j , respectively, $K(x_q, x_j)$ is the kernel function, and c is the penalty coefficient.

3.3. Event Argument Extraction

Inspired by the research of [24], the EAs are divided into long sentence-level and short word-level according to the long short-level of the corpus, and the EAs are recognized by the method of joint sequence annotation and pattern matching. Here, the examination, pathological test, immunohistochemistry event result includes 20~50 words, defined as long sentence-level EA, and other EAs are short word-level EA.

In addition to the intra sentence and inter sentence semantic features described in Section 3.2, there are other semantic features in the event sentence of HPI, such as the location feature and type feature of ETWs and the dependency syntactic feature of event sentences shown in Table 5. The location feature refers to the distance between EAs and ETW in event sentence can provide syntactic information about the event for EA recognition because the distribution of EAs usually surrounds ETW. The ETW position is encoded by the transformer, and the even and odd position is encoded by $PE_{(pos,2i)} = \sin(pos/10,000^{2i/d})$ and $PE_{(pos,2i+1)} = \cos(pos/10,000^{2i/d})$, respectively, where pos is the ETW position in the sequence, d and i are vector dimension, and i th item in the sequence, respectively.

Table 5. Semantic features faced to argument recognition.

Feature Types	Number	Symbols	Names	Describes
Extended features	8	f_8	Location of ETW	Distance between EAs and ETW
	9	f_9	Type of ETW	Types of ETW obtained by module A
	10	f_{10}	Dependency syntactic	Dependency syntactic in event sentences

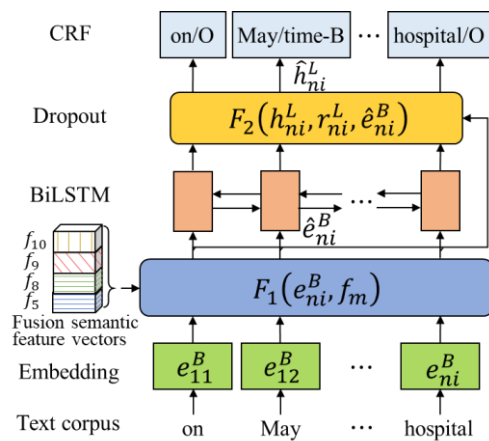
Meanwhile, different event types often correspond to different kinds of EAs, which are closely related, while ETW type implies the type of event sentence. Moreover, under the writing standard of HPI text, the same type of event sentences usually follows a similar syntactic structure, such as the treatment event sentence “On May 14, 2015, the patient was given acid making, anti-inflammatory, and antiemetic nursing.” and “The patient was given treatment for symptomatic after the operation, such as protecting liver and stomach and improving immunity.”. The former ETW is “nursing”, EAs are “acid making”, “anti-inflammatory” and “antiemetic”, the latter ETW is “treatment for symptomatic”, EAs are “protecting liver” and “protecting stomach” and “improving immunity”. Although the ETWs of the two event sentences are not the same, the EAs in each sentence are juxtaposed, and the EA and ETW are modifier-head constructions. Thus, the syntactic components such as “subject predicate object” and “attribute adverbial complex” in the event sentence can be identified by dependency parsing. The syntactic relations of dependency parsing, and their corresponding label are shown in Table 6; the specific types of relations can refer to the language technology platform LTP.

Table 6. Relations under dependency syntactic.

Relation Types	Labels	Relation Types	Labels
Subject–predicate relation	SBV	Verb–complement construction	CMP
Verb–object relation	VOB	Juxtaposition	COO
Indirect object relation	IOB	Preposition–object relation	POB
Preposed–object	FOB	Left additional relation	LAD
Concurrent remark	DBL	Right additional relation	RAD
Modifier–head construction	ATT	Independent structure	IS
Mesomorphic construction	ADV	Punctuation	W

(1) For EAs of short word-level, recognized by the sequence annotation method fusion semantic feature vectors of corpus, the model framework is given in Figure 3. Firstly, the vector representation $[[e_{n1}^B], [e_{n2}^B], \dots, [e_{ni}^B]]$ of all words in each event sentence from HPI text R is obtained by BERT [33]. Secondly, the word vector is fused with semantic feature vectors f_m ($m = 5, 8, 9, 10$); the fusion process is as follows:

$$F_1(e_{ni}^B, f_m) = e_{ni}^B = [[e_{n1}^B; f_m], [e_{n2}^B; f_m], \dots, [e_{ni}^B; f_m]] \quad (4)$$

**Figure 3.** Recognition model for short word-level by fusing semantic feature vectors.

Then, encoding the fused vectors by a bi-directional LSTM network with a hidden layer size of $d_{\tilde{n}}$ can learn how to judge the key fusion vectors to obtain the semantic information of short word-level, where $d_{\tilde{n}} = 1/2d_{n*i}$. In the t th time step, the hidden layer state of BiLSTM output is h_{ni}^t with the input \hat{e}_{ni}^t ; this process is as following:

$$[h_{n1}^L], [h_{n2}^L], \dots, [h_{ni}^L] = BiLSTM\left([e_{n1}^L], [\hat{e}_{n2}^L], \dots, [\hat{e}_{ni}^L]\right) \quad (5)$$

The dropout layer is adopted to make some bi-directional long short-term memory units randomly inactivated for avoiding over-fitting of training results, and the semantic information of short word-level is enhanced by concatenating the vectors fused semantic features.

$$F_2(h_{ni}^L, r_{ni}^L, \hat{e}_{ni}^L) = f(W^L r^L * (h_{ni}^L \parallel \hat{e}_{ni}^L) + b^L) \quad (6)$$

where $W^L \in \mathbb{R}^{d_{n*i} \times d_{\hat{x}}}$ is the weight matrix, b^L is bias, and $f(*)$ is activation function, $r^L = [r_{n1}^L, \dots, r_{ni}^L]$, $r_{ni}^L \sim \text{Bernoulli}(p)$. Here, Bernoulli is utilized to randomly generate a vector of 0 or 1.

Finally, the output \hat{h}_{ni}^L of dropout is entered into CRF and the corresponding label sequence is $y^a = (y_{n1}^a, \dots, y_{ni}^a)$. Then, all parameters of CRF can be estimated by maximizing $P(y^a | \hat{h}^L)$ for the given HPI text R :

$$P(y^a | \hat{h}^L) = \frac{1}{Z(\hat{h}_{ni}^L)} \exp \left(\sum_{g=1}^{ni} \sum_{v=1}^{ni} \left(\lambda_g E_{y_g^a, g} + \mu_v \tilde{T}_{y_{v-1}^a, y_v^a} \right) \right) \quad (7)$$

$Z(\hat{h}_{ni}^L)$ is the normalization factor, $E_{y_g^a, g}$ is the probability of label y_g^a corresponds to \hat{h}_g^L , $\tilde{T}_{y_{v-1}^a, y_v^a}$ is the probability of label y_{v-1}^a corresponds to \hat{h}_{v-1}^L under the premise of label y_v^a corresponds to \hat{h}_v^L , λ_g and μ_v are hyperparameters. Therefore, the CRF is trained by calculating the maximum log-likelihood function:

$$\max : \mathcal{L}_A = \sum_{(S, y^a) \in R} \log P(y^a | \hat{h}^L) \quad (8)$$

To sum up, we train the event extraction model LSLSD in Chinese DTE by minimizing the following loss function:

$$\mathcal{L}(\Theta) = -(\lambda_1 \mathcal{L}_T + \lambda_2 \mathcal{L}_A) \quad (9)$$

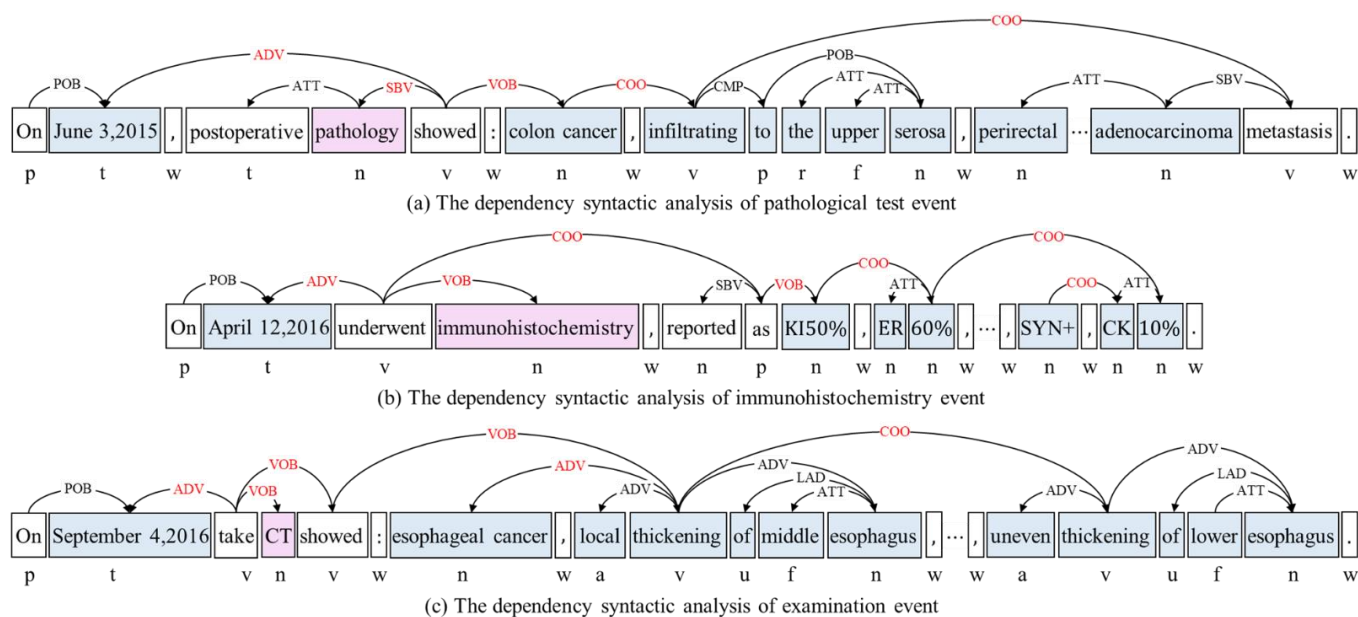
where λ_1 and λ_2 are weights, Θ is the sum of the parameters of the model LSLSD.

(2) For long sentence-level EAs, being recognized by the pattern matching rules that are given in Table 7 is not only resulted from the low detection accuracy of machine learning methods as the EAs contain a generous number of words, but also because the sentence structure of examination and test event corresponding to long sentence-level EA is relatively regular and has distinct syntactic characteristics. For example, pathological test (PT) event “On 3 June 2015, postoperative pathology showed: colon cancer, infiltrating to the upper serosa, perirectal . . . adenocarcinoma metastasis”, where the ETW is “pathology” and the EAs are time argument “3 June 2015”, disease argument “colon cancer” and test result argument “infiltrating to the upper serosa . . . adenocarcinoma metastasis”. From the dependency syntactic analysis of the event sentence, Figure 4a, time argument and “showed” have adverbial structure, while “showed” has the subject–predicate structure with ETW “pathology”; the disease argument has the verb–object structure with “showed”; the long sentence-level test result argument contains all words in the recursive juxtaposition structure with disease argument. The logical expression of the rule is shown in Table 7-E1. The dependency syntactic analysis of immunohistochemistry (IHC) event “On 12 April 2016, underwent immunohistochemistry, reported as KI50%, ER60%, . . . , SYN+, CK10%” is shown in Figure 4b, where the ETW is “immunohistochemistry” and the EAs are time argument “12 April 2016” and test result argument “KI50%, ER60%, . . . , SYN+, CK10%”. The logical expression of the rule for such sentence pattern is shown in Table 7-IHC1. “CT”, “4 September 2016”, “esophageal cancer” and “showing local thickening of mid-esophageal, uneven thickening of the lower esophagus” in CT examination event “On 4 September 2016, take CT showed: esophageal cancer, local thickening of middle esophageal, uneven thickening of lower esophagus” are the ETW, time argument, disease argument, and examination result argument, respectively, and its dependency syntactic analysis and logical expression of rules are shown in Figure 4c and Table 7-E1, respectively.

Table 7. Relations under dependency syntactic.

Event Types	Logical Expression
E1	$\tau \xrightarrow{ADV} VOB \xrightarrow{VOB} \dots VOB \xrightarrow{ADV} D \xrightarrow{w} RS^1, T \xrightarrow{VOB} \dots VOB \xrightarrow{COO} RE^2$
E2	$T \xrightarrow{w} RS \xrightarrow{SBV} COO \dots COO \xrightarrow{RE} VOB \xrightarrow{D}$
E3	$\tau \xrightarrow{ADV} \dots ADV \xrightarrow{ADV} VOB \xrightarrow{D} \xrightarrow{w} RS \xrightarrow{ADV} ATT \dots ATT \xrightarrow{ADV} RE$
E4	$\tau \xrightarrow{ADV} VOB \xrightarrow{T} \xrightarrow{w} RS \xrightarrow{w} COO \dots COO \xrightarrow{RE}$
PT1	$\tau \xrightarrow{ADV} SBV \xrightarrow{T} \xrightarrow{SBV} VOB \xrightarrow{COO} RS \xrightarrow{COO} \dots COO \xrightarrow{RE}$
PT2	$T \xrightarrow{SBV} VOB \xrightarrow{D} \xrightarrow{w} RS \xrightarrow{COO} SBV \xrightarrow{D} \xrightarrow{SBV} VOB \xrightarrow{RE}$
PT3	$\tau \xrightarrow{ATT} \dots ATT \xrightarrow{T} \xrightarrow{ATT} SBV \xrightarrow{w} RS \xrightarrow{COO} \dots COO \xrightarrow{RE}$
PT4	$T \xrightarrow{ATT} \xrightarrow{w} RS \xrightarrow{ATT} \dots ATT \xrightarrow{COO} \dots COO \xrightarrow{RE} \xrightarrow{CMP} POB \xrightarrow{D}$
PT5	$T \xrightarrow{ATT} CMP \xrightarrow{D} \xrightarrow{w} RS \xrightarrow{ATT} COO \dots COO \xrightarrow{RE}$
IHC1	$\tau \xrightarrow{ADV} VOB \xrightarrow{T} \xrightarrow{COO} VOB \xrightarrow{COO} RS \xrightarrow{COO} \dots COO \xrightarrow{RE}$
IHC2	$T \xrightarrow{SBV} VOB \xrightarrow{RS} \xrightarrow{COO} \xrightarrow{w} \dots COO \xrightarrow{w} RE$
IHC3	$T \xrightarrow{SBV} ADV \xrightarrow{\tau}, T \xrightarrow{SBV} VOB \xrightarrow{RS} \xrightarrow{COO} \dots COO \xrightarrow{RE}$

¹ RS, ² RE are the beginning and end of the result argument, respectively.

**Figure 4.** The dependency syntactic analysis of event sentences.

4. Experiments

4.1. Dataset

Annotating the data from 1000 HPI texts of Chinese EMRs in a top-three hospital in Shanghai by the semi-automatic annotation method proposed in Section 3.1, and the data set of Chinese medical event (CME_1000) is obtained, which contains 7035 event sentences. The number of ETWs and EAs in different event sentences is shown in Table 8; “#” represents the number. Here, 70%, 20%, and 10% of the CME_1000 are randomly adopted as the training set, test set, and validation set, respectively, and there are no duplicate items in these three sets.

Table 8. Statistics of CME_1000 for event extraction in Chinese HPI text.

Dataset	# HPI Text	# Entity	# Event Sentence	# ETW	# EA
Training	700	16,071	4822	4822	14,296
Validation	100	2194	728	728	1681
Test	200	3528	1485	1485	3607
Total	1000	21,793	7035	7035	19,584

4.2. Details

There are several comparative experiments designed for the subtask of event trigger word detection and event argument recognition of event extraction; the experiment results are evaluated by the index value of precision (P), recall (R), and F1-value.

In the event trigger word detection module, the software package LibSVM [34] is adopted to train and test the SVM; select RBF kernel for $K(x_q, x_j)$ in the formula (2); the constant value coefo of kernel function is 0; the penalty coefficient c is 1.

In the event argument recognition module, the neural network framework Tensorflow 2.0 (<https://www.tensorflow.org/install> (accessed on 13 October 2019)) is employed for model training and the pre-training language model adopts BERT_base. Adam optimizer is used in the BiLSTM layer, and the batch size, epochs, learning rate, and dropout rate are set to 16, 20, 2×10^{-5} , and 0.5, respectively.

4.3. Results

In the aspect of ETW detection, aiming at proving the positive effect of the proposed extended features fusion of long short-level semantic dependency on the detection results, the model performance of adding extended features of short-level semantic dependency is compared based on the baseline model without fusion of extended features (only feature $f_1 \sim f_4$ are included). The performance of the ETW detection module under different feature combinations on CME_1000 is shown in Table 9. From Table 9, the experiment results of the baseline are not ideal because the entity information in the medical text is not fully utilized. After adding the entity type feature f_5 (T) of intra sentence, the classification results of all kinds of events are improved because different entity types only appear in specific event types. For example, the P and F1 value of the operation event increased by 5.3% and 3.61%. This is because the ETW of operation event often appears in chemotherapy, test, and examination event, and the entities of operation type only appear in admission, diagnosis, and operation event, ensuring the entity type feature T of intra sentence can improve the ETW detection results of operation event. Moreover, after adding the number feature f_6 (N) of different entities of inter sentences, the average F1 value for all events increased from 81.07% to 82.45%. For example, the F1 value of the examination event is increased by 3.44%. This is because the feature N can effectively distinguish test and examination events with similar entity type distribution.

In the aspect of EA extraction, our model LSLSD is compared with word-based C-BiLSTM and char-based C-BiLSTM proposed by Zeng et al. [27] to prove the better performance of LSLSD in EA extraction. As shown in Table 10, the short word-level arguments and long sentence arguments of the corpus are better captured by LSLSD, which joins sequence annotation and pattern matching. In the EA recognition task, the F1 value of LSLSD is 6.5% and 5.1% higher than word-based C-BiLSTM and char-based C-BiLSTM respectively. The F1 value of LSLSD is 11.2% and 8.2% higher than word-based C-BiLSTM and char-based C-BiLSTM, respectively, in the EA classification task.

Table 9. Results of trigger detection under different feature combinations.

Methods	Event Types	Chemotherapy			Examination		
		P	R	F1	P	R	F1
Baseline		94.71	82.06	87.93	87.14	60.39	71.34
Baseline + T		95.29	85.11	89.91	87.45	63.23	73.39
Baseline + T + N		95.31	86.06	90.44	89.47	67.32	76.83
Methods	Event Types	Operation			General Treatment		
		P	R	F1	P	R	F1
Baseline		68.96	86.25	76.64	86.98	76.95	81.66
Baseline + T		74.26	87.28	80.25	87.22	81.48	84.25
Baseline + T + N		75.89	87.62	81.33	87.24	81.57	84.31
Methods	Event Types	Admission			Immunohistochemistry		
		P	R	F1	P	R	F1
Baseline		78.34	73.44	75.81	69.91	76.51	73.06
Baseline + T		81.22	75.47	78.24	70.27	81.29	75.38
Baseline + T + N		82.16	76.50	79.23	75.46	82.60	78.87
Methods	Event Types	Pathological Test			Diagnosis		
		P	R	F1	P	R	F1
Baseline		87.84	77.51	82.35	86.91	81.77	84.26
Baseline + T		88.13	77.28	82.35	89.09	80.90	84.80
Baseline + T + N		91.06	76.81	83.33	91.26	80.05	85.29

Table 10. Results of event argument extraction under different methods.

Methods	Event Types	EA Recognition			EA Classification		
		P	R	F1	P	R	F1
Word-based C-BiLSTM		59.8	52.7	59.8	59.8	52.7	59.8
Char-based C-BiLSTM		61.2	58.5	61.2	61.2	58.5	61.2
LSLSD (our)		66.3	67.5	66.3	67.5	66.3	67.5

4.4. Discussion

In terms of ETW detection, the low accuracy of ETW detection is directly reduced by the ETWs that infrequently appear in the candidate ETW dictionary that exists in HPI texts. The accuracy of ETW module is improved by setting a two-stage experiment; the first stage determines whether each candidate ETW, verb, and gerund are ETW through direct binary classification by using features of word, part of speech, and the number of different entities type. The second stage determines the type by multi-classification for the ETWs selected in the first stage. As shown in Table 11, the F1 value of the two-stage ETW detection has been significantly improved under the same feature combination. For example, compared with the result in Table 10, the F1 value of the chemotherapy event increased from 89.91% to 91.48% through the two-stage method under the feature combination of “Baseline + T”.

Table 11. Results of trigger detection by two-stage method.

Event Types		Chemotherapy			Examination		
		P	R	F1	P	R	F1
Two-stage	Baseline + T	95.06	88.16	91.48	88.61	69.31	77.78
	Baseline + T + N	95.12	88.32	91.59	89.87	70.29	78.88
Event Types		Operation			General treatment		
		P	R	F1	P	R	F1
Two-stage	Baseline + T	77.87	85.91	81.69	87.87	83.53	85.65
	Baseline + T + N	80.75	87.97	84.21	87.86	84.31	85.91
Event Types		Admission			Immunohistochemistry		
		P	R	F1	P	R	F1
Two-stage	Baseline + T	84.33	75.66	79.76	76.05	83.65	79.67
	Baseline + T + N	85.20	76.62	80.68	76.12	81.87	78.89
Event Types		Pathological Test			Diagnosis		
		P	R	F1	P	R	F1
Two-stage	Baseline + T	90.71	75.67	82.51	91.64	80.22	85.55
	Baseline + T + N	91.45	76.91	83.55	91.56	79.06	84.85

In addition, the event sentence with the accuracy of TWE classification is less than 40% by fusing extended features of short-level semantic dependence named untrusted event sentence; otherwise, it is a trusted event sentence. The untrusted event sentences are classified a second time by adding long-level semantic dependence features f_7 and other extended features of enter sentences to improve the accuracy of ETW detection. From the ETW detection results under different features combinations in Table 12, the result of the “P1 + R1 + B + N + PN + RN” combination is the best. Compared with the two-stage ETW detection method, the R and F1 value of experiment with document consistency features are improved from 79.03% to 80.61% and from 82.23% to 83.57%, respectively, which indicate that document consistency features can supplement the information difficult to extract on the long sentence-level and improve the performance of the model. The trend of F1 value change of different events under different feature combinations are shown in Figure 5, in which the entity type feature T of intra sentence significantly improves the ETW detection effect on chemotherapy, treatment, operation, and examination events, and the number feature N of different entities of inter sentences significantly improves the ETW detection effect on the pathological test, immunohistochemistry, operation, and examination events. The two-stage experiment had a great influence on treatment, operation, examination, and chemotherapy events, while the document consistency features S improved the results of ETW detection of all event types.

Table 12. Results of trigger detection based on long-level semantic dependency feature.

Methods	Event Types	ALL-TOTAL		
		P	R	F1
B ¹ + P1 ² + R1 ³		86.31	79.42	82.72
B + P1 + R1 + T + PT ⁴ + RT ⁵		86.54	80.13	83.21
B + P1 + R1 + T + N + PTN ⁶ + RTN ⁷		85.84	79.36	82.47
B + P1 + R1 + N + PN ⁸ + RN ⁹		86.97	80.61	83.57

¹ B represents the basic features. ² P1, ³ R1, ⁴ PT, ⁵ RT, ⁶ PTN, ⁷ RTN, ⁸ PN and ⁹ RN are the document consistency feature, where P1 and R1 represent the event type of the previous and latter sentence respectively; PT and RT, respectively, represent the feature (T) of entity type in the previous and latter sentence; PN and RN are the number feature (N) of different entity types in the previous and latter sentence; and PTN and RTN represent the extended features of intra sentence in the previous and latter sentence, respectively.

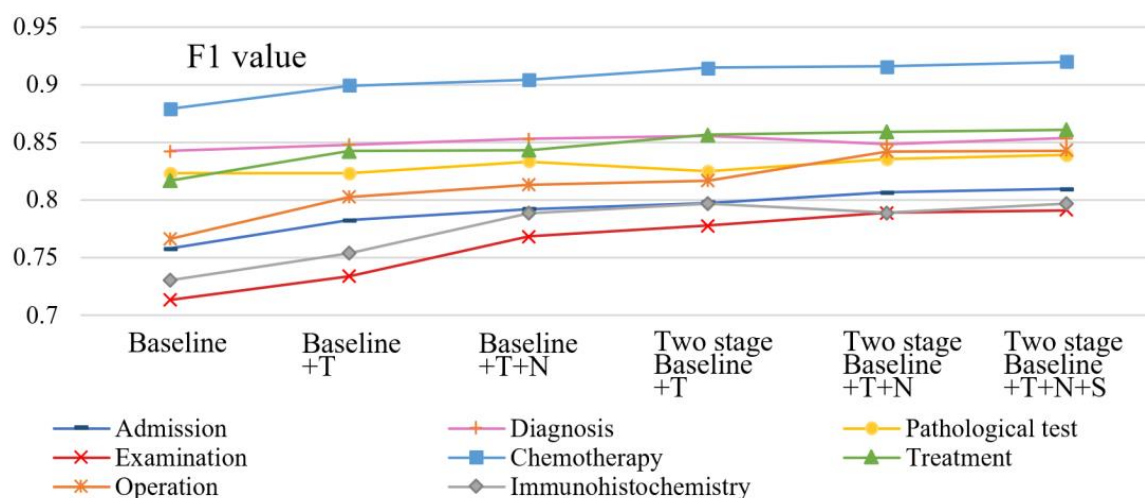


Figure 5. F1-value variation tendency under different feature combinations.

In terms of EA extract, the influence of different word vector pre-training models of Word2Vec, ELMO, and BERT on the recognition of ETW and EA are compared in this paper; as shown in Table 13, the effect of BERT is better than that of Word2Vec and ELMO, which is because Word2Vec does not consider the influence of context information on the word vector when constructing the text vector so that the word vector generated by the same word is invariant. ELMO employs stacked BiLSTM to make the generated word vectors contain more context information of the vocabulary, so the ELMO is better than Word2Vec. The model effect of using BERT is better than ELMO and Word2Vec, due to the generated text sequence of BERT having deeper context information and lexical-semantic by feature extractor of transformer. The experiment results of removing the CRF model are obtained to highlight the positive effect of CRF on improving the accuracy of label prediction by adding constraint rules, and the results show that the effect of using BERT-BiLSTM along is far less than that of adding the CRF model.

Table 13. Results of event argument extraction under different methods.

Methods	Tasks	EA Recognition			EA Classification		
		P	R	F1	P	R	F1
Word2Vec + BiLSTM + CRF		45.2	51.3	48.1	41.3	48.6	44.7
ELMO + BiLSTM + CRF		52.7	57.4	54.9	51.6	57.2	54.3
BERT + BiLSTM + CRF		53.5	57.8	55.6	52.6	57.1	54.8
BERT + BiLSTM		51.0	54.7	52.8	48.9	53.1	50.9

In addition, the ablation experiment is designed to better study the influence of the four extra input vectors of BiLSTM on the EA recognition task. This experiment utilizes the sequence annotation model of BERT-BiLSTM-CRF for all event types, the experiment results are shown in Table 14, where D represents the distance vector between the word and the ETW, and C and Y represent the type vector of the ETW and the dependency syntactic information vector, respectively.

Table 14. Results of event argument extraction under different feature combinations.

Methods	Tasks	EA Recognition			EA Classification		
		P	R	F1	P	R	F1
D		54.2	58.3	56.2	53.1	57.6	55.3
DC		57.8	63.4	60.5	56.2	59.4	57.8
DY		54.5	57.9	56.1	52.7	57.1	54.8
DCY		55.1	58.0	56.5	53.6	58.2	55.8
DCT		60.5	62.6	61.5	58.5	60.1	59.3
DCTY		58.7	59.8	59.2	55.6	57.3	56.4

From Table 14, the addition of four types of vectors improves the performance of the model in varying degrees, which proves the effectiveness of the proposed feature vectors. It is noted that the experimental effect is higher than that of BERT-BiLSTM-CRF when the dependency syntactic information vector Y is added into the combination of the distance vector D, type vector C of ETW and entity type vector T of intra sentence, but it is slightly lower than before. The same situation occurs again for the vector combination of DC and DCY, which illustrates that the addition of dependency syntactic information vector will harm the combination of other features that may be caused by the characteristics of syntactic structure is not obvious due to the different writing styles of different doctors, although the medical text has a certain degree of standardization. Compared with the BERT-BiLSTM-CRF, the F1 values of the best combination DCT among all vector combinations on EA recognition and EA classification tasks are improved by 5.9% and 4.5%, respectively.

5. Conclusions

A novel model for event extraction on Chinese history of present illness texts by fusing its long short-level semantic dependency is proposed in this paper, and the experiment results demonstrated that the proposed extended features with implicit information of intra-sentence or inter-sentences help improve the performance of LSLSD. Meanwhile, LSLSD has excellent generalization performance, due to the fact that the event arguments recognition module divides event arguments into the long sentence-level and short word-level, and it recognizes event arguments by the method of joint sequence annotation and pattern matching.

We also noticed in the experiment results analysis that, overall, the dependence syntactic feature proposed in this paper has negative effects on some event arguments recognition. The reason for this effect is that the HPI text data set is small; on the other hand, the writing styles of different clinicians are different. Further improvement might be made by expanding the scale of Chinese HPI texts, to explore more profound semantic features. We will study advanced deep learning algorithms to improve the accuracy of event argument recognition.

Author Contributions: Methodology, P.Z., C.W.; software, P.Z., C.W.; formal analysis, P.Z., Y.F.; investigation, P.Z., C.W.; data curation, P.Z., C.W.; writing—original draft preparation, P.Z., C.W.; writing—review and editing, P.Z., Y.F.; visualization, P.Z., C.W.; supervision, Y.F.; project administration, Y.F.; funding acquisition, Y.F. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by the National Key Research and Development Program of China (No. 2019YFB2101600).

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: The electronic medical record data obtained from the hospital has not been made available due to privacy concerns.

Acknowledgments: We would like to thank two doctors from a First-class Hospital at Grade III in Shanghai, who manually checked and corrected the data set that is labeled by our semi-automatic annotation method for the experiments in this paper.

Conflicts of Interest: The authors declare no conflict of interest.

References

- Guo, S.; Jin, Z.; Gotz, D.; Du, F.; Zha, H.; Cao, N. Visual progression analysis of event sequence data. *IEEE Trans. Vis. Comput. Graph.* **2018**, *25*, 417–426. [\[CrossRef\]](#) [\[PubMed\]](#)
- Kwon, B.C.; Anand, V.; Severson, K.A.; Ghosh, S.; Sun, Z.; Forhnert, B.I.; Lundgren, M.; Ng, K. DPVis: Visual exploration of disease progression pathways. *arXiv* **2019**, arXiv:1904.11652.
- Jin, Z.; Cui, S.; Guo, S.; Gotz, D.; Sun, J.; Cao, N. Carepre: An intelligent clinical decision assistance system. *ACM Trans. Comput. Healthc.* **2020**, *1*, 1–20. [\[CrossRef\]](#)
- Yazhini, K.; Loganathan, D. A state of art approaches on deep learning models in healthcare: An application perspective. In Proceedings of the 2019 3rd International Conference on Trends in Electronics and Informatics (ICOEI), IEEE, Tirunelveli, India, 23–25 April 2019; pp. 195–200.
- Harerimana, G.; Kim, J.W.; Yoo, H.; Jang, B. Deep learning for electronic health records analytics. *IEEE Access* **2019**, *7*, 101245–101259. [\[CrossRef\]](#)
- Yadav, S.; Ramteke, P.; Ekbal, A.; Saha, S.; Bhattacharyya, P. Exploring disorder-aware attention for clinical event extraction. *ACM Trans. Multimed. Comput. Commun. Appl.* **2020**, *16*, 1–21. [\[CrossRef\]](#)
- Fócil-Arias, C.; Sidorov, G.; Gelbukh, A. Medical events extraction to analyze clinical records with conditional random fields. *J. Intell. Fuzzy Syst.* **2019**, *36*, 4633–4643. [\[CrossRef\]](#)
- Doddington, G.R.; Mitchell, A.; Przybocki, M.A.; Ramshaw, L.A.; Strassel, S.M.; Weischedel, R.M. The automatic content extraction (ace) program-tasks, data, and evaluation. In Proceedings of the 4th International Conference on Language Resources and Evaluation, Lisbon, Portugal, 26–28 May 2004; Volume 2, pp. 837–840.
- Kovačević, A.; Dehghan, A.; Filannino, M.; Keane, J.A.; Nenadic, G. Combining rules and machine learning for extraction of temporal expressions and events from clinical narratives. *J. Am. Med. Inform. Assoc.* **2013**, *20*, 859–866. [\[CrossRef\]](#) [\[PubMed\]](#)
- Bui, Q.C.; Campos, D.; van Mulligen, E.; Kors, J. A fast rule-based approach for biomedical event extraction. In Proceedings of the BioNLP Shared Task 2013 Workshop, Sofia, Bulgaria, 9 August 2013; pp. 104–108.
- Viani, N.; Miller, T.A.; Napolitano, C.; Priori, S.G.; Savova, G.K.; Bellazzi, R.; Sacchi, L. Supervised methods to extract clinical events from cardiology reports in Italian. *J. Biomed. Inform.* **2019**, *95*, 103219. [\[CrossRef\]](#)
- Zhou, X.; Xiong, H.; Zeng, S.; Fu, X.; Wu, J. An approach for medical event detection in Chinese clinical notes of electronic health records. *BMC Med. Inform. Decis. Mak.* **2019**, *19*, 31–37. [\[CrossRef\]](#)
- Uzuner, Ö.; Luo, Y.; Szolovits, P. Evaluating the state-of-the-art in automatic deidentification. *J. Am. Med. Inform. Assoc.* **2007**, *14*, 550–563. [\[CrossRef\]](#) [\[PubMed\]](#)
- Hogenboom, F.; Frasinca, F.; Kaymak, U.; De Jong, F.; Caron, E. A survey of event extraction methods from text for decision support systems. *Decis. Support. Syst.* **2016**, *85*, 12–22. [\[CrossRef\]](#)
- Devlin, J.; Zbib, R.; Huang, Z.; Lamar, T.; Schwartz, R.; Makhoul, J. Fast and robust neural network joint models for statistical machine translation. In Proceedings of the 52nd Annual Meeting of the Association for Computational Linguistics, Baltimore, MD, USA, 23–25 June 2014; Volume 1, pp. 1370–1380.
- Chen, Z.; Ji, H. Language specific issue and feature exploration in Chinese event extraction. In Proceedings of the Human Language Technologies: The 2009 Annual Conference of the North American Chapter of the Association for Computational Linguistics, Boulder, CO, USA, 31 May–5 June 2009; pp. 209–212.
- Valenzuela-Escárcega, M.A.; Hahn-Powell, G.; Surdeanu, M.; Hicks, T. A Domain-independent rule-based framework for event extraction. In Proceedings of the ACL-IJCNLP 2015 System Demonstrations, Beijing, China, 26–31 July 2015; pp. 127–132.
- Huang, L.; Cassidy, T.; Feng, X.; Ji, H.; Voss, C.; Han, J.; Sil, A. Liberal event extraction and event schema induction. In Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics, Berlin, Germany, 7–12 August 2016; Volume 1, pp. 258–268.
- Xia, J.; Fang, A.C.; Zhang, X. A novel feature selection strategy for enhanced biomedical event extraction using the Turku system. *BioMed Res. Int.* **2014**, *2014*, 205239. [\[CrossRef\]](#) [\[PubMed\]](#)
- Yang, H.; Chen, Y.; Liu, K.; Xiao, Y.; Zhao, J. DCFEE: A document-level Chinese financial event extraction system based on automatically labeled training data. In Proceedings of the ACL 2018, System Demonstrations, Melbourne, Australia, 15–20 July 2018; pp. 50–55.
- Zheng, S.; Cao, W.; Xu, W.; Bian, J. Doc2EDAG: An end-to-end document-level framework for chinese financial event extraction. *arXiv* **2019**, arXiv:1904.07535.
- Jindal, P.; Roth, D. Extraction of events and temporal expressions from clinical narratives. *J. Biomed. Inform.* **2013**, *46*, S13–S19. [\[CrossRef\]](#)
- Wei, Y.; Liu, M.; Hu, H. Biomedical event extraction based on deep contextual word representation and self-attention. *Comput. Eng. Sci.* **2020**, *42*, 1670–1679. (In Chinese)

24. Xu, X.; Li, P.F.; Zheng, X.; Zhu, Q. Event inference for semi-supervised Chinese event extraction. *J. Shandong Univ. Nat. Sci.* **2014**, *12*, 3.
25. Li, P.; Zhou, G.; Zhu, Q.; Hou, L. Employing compositional semantics and discourse consistency in Chinese event extraction. In Proceedings of the 2012 Joint Conference on Empirical Methods in Natural Language Processing and Computational Natural Language Learning, Jeju, Korea, 12–14 July 2012; pp. 1006–1016.
26. Yu, H.; Xu, C.; Liu, Y.; Fu, Y.; Gao, D. BiLSTM and CRF-based extraction of therapeutic events from Chinese clinical guidelines. *Chin. J. Med. Libr. Inf. Sci.* **2020**, *29*, 9–14. (In Chinese)
27. Zeng, Y.; Yang, H.; Feng, Y.; Wang, Z.; Zhao, D. A convolution BiLSTM neural network model for Chinese event extraction. In *Natural Language Understanding and Intelligent Applications*; Springer: Kunming, China, 2–6 December 2016; pp. 275–287.
28. Björne, J.; Salakoski, T. Generalizing biomedical event extraction. In Proceedings of the BioNLP Shared Task 2011 Workshop, Portland, OR, USA, 24 June 2011; pp. 183–191.
29. Pham, X.Q.; Le Minh, Q.; Ho, B.Q. A hybrid approach for biomedical event extraction. In Proceedings of the BioNLP Shared Task 2013 Workshop, Sofia, Bulgaria, 9 August 2013; pp. 121–124.
30. Huang, Z.; Xu, W.; Yu, K. Bidirectional lstm-crf models for sequence tagging. *arXiv* **2015**, arXiv:1508.01991.
31. Vaswani, A.; Shazeer, N.; Parmar, N.; Uszkoreit, J.; Jones, L.; Gomez, A.N.; Kaiser, L.; Polosukhin, I. Attention is all you need. *arXiv* **2017**, arXiv:1706.03762.
32. Wang, C.; Zhai, P.; Fang, Y. Chinese medical event detection based on feature extension and document consistency. In Proceedings of the 5th International Conference on Automation, Control and Robotics Engineering (CACRE), IEEE, Dalian, China, 18–20 September 2020; pp. 753–758.
33. Devlin, J.; Chang, M.W.; Lee, K.; Toutanova, K. Bert: Pre-training of deep bidirectional transformers for language understanding. *arXiv* **2018**, arXiv:1810.04805.
34. Chang, C.C.; Lin, C.J. LIBSVM: A library for support vector machines. *ACM Trans. Intell. Syst. Technol.* **2011**, *2*, 1–27. [[CrossRef](#)]