# Survey on Event Extraction Technology in Information Extraction Research Area

Liying Zhan,Xuping Jiang
Institute of Information and Communication of National University of Defense Technology, Wuhan Hubei , China
1241761652@qq.com

**Abstract—Event extraction is one of the most challenging tasks in the field of information extraction. The research of event extraction technology has important theoretical significance and wide application value. This paper makes a survey of event extraction technology, describes the tasks and related concepts of event extraction, analyzes, compares and generalizes the relevant descriptions in different fields. Then analyzes, compares and summarizes the three main methods of event extraction. These methods have their own advantages and disadvantages. The methods based on rules and templates are more mature, the method based on statistical machine learning is dominant, and the method based on deep learning is the future development trend. At the same time, this paper also reviews the research status and key technologies of event extraction, and finally summarizes the current challenges and future research trends of event extraction technology.**

**Keywords—event extraction; information extraction; pattern matching; machine learning; deep learning**

## I. INTRODUCTION

With the rapid development of information technology, such as computers and networks, and the advent of cloud computing and big data eras, information data is exploding. Obtaining valuable information from massive information data has become a focus of attention, and information extraction technology has come into being. Information extraction belongs to the field of natural language processing, and with the development of natural language processing, it has become a hot topic of current research. As the name implies, information extraction refers to extracting information that people need and are interested in from a large amount of texts and documents, and storing it structurally [1]. Event extraction is an important branch of information extraction, one of the most challenging tasks and a problem in artificial intelligence. The main research is to extract the event information of interest from various texts and store it in a structured way, for use in other information extraction business or direct practical application.

The event extraction technology combines multidisciplinary development results and practical application requirements, and has important theoretical research significance and practical value. Event extraction is an important part of the field of natural language processing, involving information processing, artificial intelligence, pattern matching and data processing. The development of event extraction technology can promote the integration and development of related disciplines and promote the deep development of natural language processing technology. In practical applications, event extraction has been widely used in the fields of automatic question and answer [2], information retrieval [2], human-machine interface, trend analysis and so on.

## II. DESCRIPTION OF EVENT EXTRACTION ISSUES

### A. *Event extraction task description*

Event extraction task description, different fields have different definitions and different understandings, mainly reflected in the relevant descriptions of Yale University and Message Understanding Conference (MUC), Automatic Content Extraction Conference (ACE) and TAC KBP 2015 Evaluation Conference. The comparative analysis is shown in Table I.

978-1-5386-6243-4/19/$31.00 ©2019 IEEE

TABLE I RELATED DESCRIPTION OF EVENT EXTRACTION

| | Event extraction task description work | Application area | Achievement |
|---|---|---|---|
| Yale University | Yale University pioneered the study of event type identification and conducted research on event understanding. | News topic hotspots classified information | Designed information extraction system |
| MUC | Various specifications, task objectives and a complete evaluation system are defined. It is proposed that the entire extraction target is to form an event scene description in the form of a template. | News report | Laid a solid foundation for event extraction development |
| ACE | The conference describes the event extraction as a task from identifying events in the text and extracting elements of the event. | Unconstrained by specific areas and scenarios | Great progress has been made in the accuracy and versatility of event extraction |
| TAC KBP | The conference describes event extraction as the process of automatically extracting specific types of event attributes in the text. | Chinese English Spanish | Specifically for the field of NLP, and provides a large number of corpora. |

From the table, we can see that the description of event extraction task is a developing process, the applicable fields are expanding, the applied languages are increasing, and the evaluation system is improving. In summary, The event extraction tasks mainly include: event type identification, event element identification, and event element role assignment. The general flow of event extraction is shown in Fig.1.

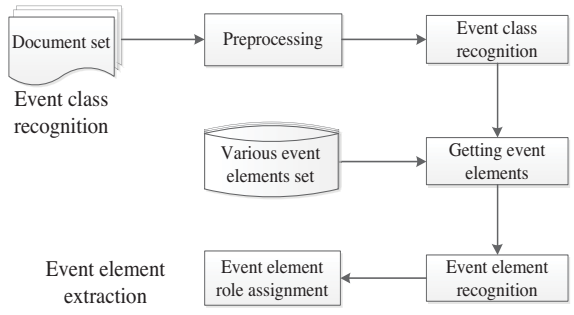


Fig.1 Event extraction flow chart

### *B Event Extraction Related Concept Description*

The related concepts of event extraction have different descriptions in different fields. The more important ones are the ACE evaluation conference and the TAC KBP evaluation conference, all of which describe the related concepts of event extraction. The event and its components are described separately below.

1) Event: The concept of event has no unified definition yet. It has different understandings in different fields. The term originated from cognitive psychology, and cognitive scientists regard events as the basic unit of human understanding and memory of the real world. The ACE2005 evaluation conference defines an event as an objective fact that a particular person or thing interacts at a specific time and place[3].In the field of natural language processing, events represent one or more actions or changes in state that occur at a certain time or region. Regardless of how it is defined, events contain basic elements—time, place, person, and actions triggered by verbs, nouns, or phrases.

2) Trigger: Also known as the event indicator, the ACE2005 evaluation conference described it as a word that can trigger an event, usually a verb or a noun and phrase that represents an action; TAC KBP Assessment Conference described it as the core word for the trigger event. In short, the trigger word can determine the type of event, and reflect the most important characteristics of the event, generally verbs or nouns and phrases with verb nature, which play a key role in the recognition of event types.

3) Event elements: Also known as event argument, the ACE Assessment Conference describes it as entity and entity attribute information related to the occurrence of an event, including time, place, person, etc. The TAC KBP Assessment Conference defines it the participants of events, mainly consisting of information such as entities and time. Event elements generally refer to the various elements of an event, which can reflect the subject information of the event.

4) Event categories: Different areas have different definitions. ACE2005 corpus, defines 8 event categories and 33 seed categories. The eight categories are life, movement, business, contact, justice, personnel, transaction, and conflict. The TAC KBP Evaluation Conference defines 9 categories and 38 seed categories, one major category (manufacture) and 5 seed categories more than ACE. The five seed categories are 1 seed category of the manufacture, 2 seed categories of the interaction contact, 1seed category of the movement, and 1

seed category of transaction.

## III. MAIN EVENT EXTRACTION METHODS

In recent years, under the impetus of the ACE evaluation conference, the research on event extraction has developed rapidly, and some theoretical results have been achieved, and some practical systems have been developed. Compared with English, especially in English, the Chinese event extraction research started earlier and the theory is more mature, but the Chinese event extraction research has also achieved certain results. The domestic research on Chinese event extraction technology started late, but it has also achieved certain achievements. The initial research was mainly methods based on rules and templates, and later developed into methods based on statistical machine learning. The current research mainly tends to be methods based on deep learning.

### *A. Event extraction methods based on rules and templates*

The early method of event extraction was mainly based on rule-based methods, and later developed into a method based on pattern matching. These methods are essentially the same, that is, they need to build rules or templates. The event extraction method based on pattern matching refers to a method of matching the event sentence to be extracted with the corresponding template, it basic principle is shown in Fig. 2.

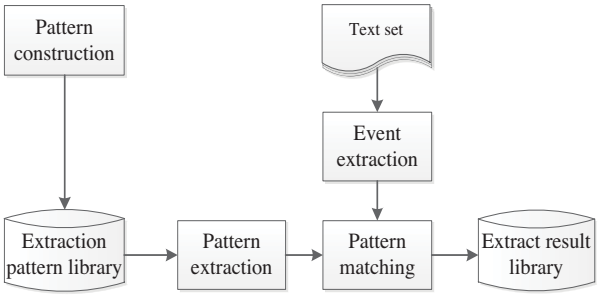


Fig. 2 Schematic diagram of event extraction based on pattern matching

Meiying Jia of Beijing University of Science and Technology used pattern matching method to study the extraction of military exercise information [4]. It uses hierarchical automatic classification method, seed mode -based bootstrapping method and corpus -based labeling i different stages of extraction. Its pattern matching method focuses on the pattern acquisition and matching algorithm, as shown in Fig. 3.

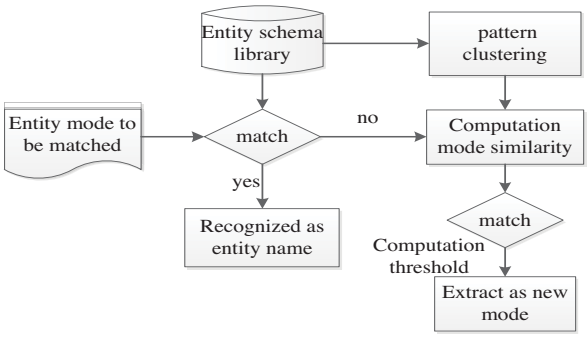


Fig.3 Schematic diagram of the pattern acquisition and matching algorithm

The above research shows that the core of the pattern matching method is the construction of the event extraction mode. Jifa Jiang studied the automatic acquisition of patterns [5], and proposed an event extraction model learning method, which based on the domain-independent concept knowledge base. The system he built can automatically learn the IE mode from the original corpus as long as the IE task is defined, without providing seed mode and preprocessing of the corpus, which greatly improves the efficiency. Another scholar Ming Luo constructed a hierarchical lexical-semantic rule model based on finite state machine driven for the automatic extraction of various financial event information[6], with high accuracy. Liao et al.[7] used the predicate-argument pattern when constructing event extraction templates, and extended the original template by similarity.

The method based on pattern matching is better applied in a specific field, but this method has poor portability and flexibility. It needs to rebuild the model when it is cross-domain. The construction of the model takes a lot of time and manpower. Using machine learning and other methods can speed up the acquisition of patterns, but it will bring conflicts between different modes.

### *B. Event extraction methods based on statistical machine learning*

To extract events by machine learning is essentially to treat event extraction as a classification problem. The main task is to select appropriate features and construct appropriate classifiers. Compared with the pattern matching method, the machine learning method can be applied in different fields, and has high portability and flexibility, and has been widely used.

The classifier is generally constructed on the basis of statistical models. The main statistical models in event

extraction mainly include maximum entropy model, hidden Markov model, conditional random field model and support vector machine model.

For example, in 2002, Chieu et al. applied the maximum entropy model for the first time in the recognition of event elements, and extracted lecture announcements and personnel management events. Another scholar, H. Llorens, introduced the conditional random field model (CRF) in semantic role annotation and applied it to TimeML event extraction to improve the performance of the system. Domestic Chinese Jiangde Yu et al. [8] proposed a Chinese text event extraction method based on hidden Markov model (HMM). This method constructs an independent hidden Markov model when extracting each type of event element.

In order to improve the effect of event extraction, a variety of machine learning algorithms are sometimes used in combination. In 2006, David Ahn [2] integrated the MegaM and Timbl machine learning methods to identify event categories and event elements. Event type recognition has the problem of backward dependence on event elements recognition. In 2012, Bolei Hu et al. [9] solved this problem very well. They regarded event extraction as sequence labeling and constructed an improved conditional random domain joint labeling model. The main idea is to simultaneously tag event types and event elements in the graph model. The improved CRF model is shown in Fig. 4.

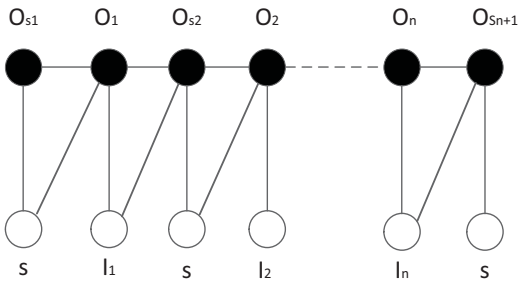


Fig. 4 Improved CRF model

Many machine learning methods are based on trigger words for event recognition. The method based on trigger words introduces a large number of counterexamples in training, resulting in imbalances between positive and negative examples. I In order to solve this problem, Yanyan Zhao [10] of Harbin Institute of Technology identified the event categories by combining trigger word expansion and binary classification. In addition, Honglei Xu proposed an event type recognition method based on event instances [11]. This method overcomes the problem of positive and negative case imbalance and data sparseness by using sentences instead of words as identification examples.

The current dominant role in event extraction research is method based on machine learning, but this method requires large-scale labeled training corpus. If the training corpus is not enough or the category is single, it will seriously affect the extraction effect of the event, and the corpus construction becomes an important task. However, the construction of the corpus takes a lot of manpower and time. In order to alleviate this problem, the scholars further explored the method of deep learning.

*C. Event extraction methods based on deep learning*

Deep learning is a new direction in the field of machine learning research. Compared with shallow neural networks, deep neural network (DNN) has better feature learning ability, unsupervised layer-by-layer pre-training of its abstract mathematics. Features that more effectively characterize the essential characteristics of raw data . Yajun Zhang et al.[12] constructed an event recognition model based on deep learning, and used BP neural network to identify events, in which the deep semantic information of words was extracted through deep belief network. At the same time, the literature also proposed a hybrid supervision deep belief network, which integrated both supervised and unsupervised learning methods, which can improve the recognition effect and control the training time.

The traditional feature-based event extraction method requires a large number of feature design work manually, and requires complex natural language processing tools, which consume a lot of manpower and time and generate data sparse problems. In this regard, Kai Wang [13] proposed an event extraction method based on recurrent neural network (RNN), which can automatically learn the features in the sentence, without a lot of artificial feature design work, and overcome the complex feature engineering.

Recurrent neural network (RNN) is widely used in the field of natural language processing. It is mainly used to solve sequence problems, and it has good effects for event extraction. This is because the structure of the recurrent neural network

model consists of three layers, namely the input layer x, the hidden layer h, and the output layer y, where in the hidden layer h represents the internal state of the recurrent neural network, as shown in Fig. 5.

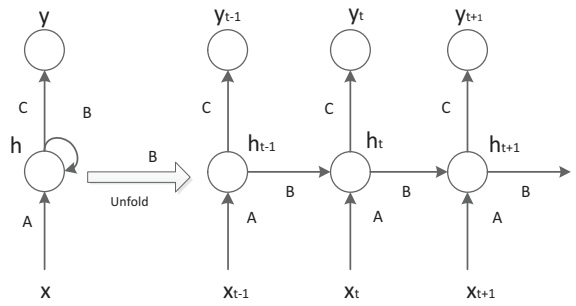


Fig. 5 recurrent neural network structure model

At time t, the input h(t) of the hidden layer consists of the input x(t) at the current time and the output h(t-1) of the hidden layer at the previous time, and h(t-1) contains the input information of the previous moment and the information in the previous hidden layer. In this way, by adding the hidden layer input at the previous moment, the history information of the sequence is added, so that the information of the longer distance can be utilized.

In addition, in order to avoid complex feature engineering, the relevant scholars constructed a neural network model of joint learning, and proposed a neural network event extraction method based on the joint model. For example, Nguyen et al. propose a joint learning based on RNN model for event type recognition and event element recognition. Zhengkuan Zhang[14] of Beijing University of Posts and Telecommunications designed a new event extraction framework, combined with the window-winding convolutional neural network and the recurrent neural network to form a connected learning method, which can simultaneously extract event trigger words and event elements, not only avoiding the complex feature engineering also solves the problem of error propagation.

The deep learning method overcomes the limitation of shallow machine learning, can learn more abstract mathematical features, and make the data have better feature expression, so as to achieve efficient extraction of text events. Compared with shallow machine learning, the deep learning framework can effectively capture data features exponentially, which has been applied in the field of event extraction.

The above three major categories of major event extraction methods are reviewed, each with advantages and disadvantages, and comparative analysis is shown in Table II.

TABLE II COMPARISON AND ANALYSIS OF THE MAIN METHODS OF EVENT EXTRACTION

| Method | Main idea | Advantage | Disadvantage |
|---|---|---|---|
| Methods based on rules and templates | Construction of rules and patterns | The research started early, the method is mature, the application in the specific field is better, and the accuracy is high. | Poor portability, Poor flexibility, time-consuming and labor-intensive pattern construction, conflicts between different modes |
| Methods based on statistical machine learning | Choosing The right features and constructing the right classifier | Strong applicability, high portability and flexibility, relevant technologies are relatively mature, occupying a dominant position | Large-scale labeled training corpus Is needed. The construction of corpus is time-consuming and laborious, and there are positive and negative cases of imbalance and data sparseness. |
| Methods based on deep learning | Selection of neural network models and learning of data characteristics | Can learn More abstract mathematical features, better express data features, advanced technology, high efficiency | Related technology research is in its infancy, with less application development |

## IV. CHALLENGES IN EVENT EXTRACTION

With the in-depth development of event extraction research, event extraction has made great progress both in theory and in application. However, the development of artificial intelligence and big data technology has put forward higher requirements for the accuracy of event extraction. The research and development still faces many challenges, mainly in the following aspects:

1) Research on related technologies such as entity, relationship identification, and syntax analysis is not mature enough, leading to cascading errors. Event extraction has developed on the basis of entity and relationship identification. It depends to some extent on the effects of entities, relationship recognition and text preprocessing, but these underlying technologies are still not mature enough.

2) The field scalability and portability of the event extraction system are not ideal. For example, the relevant research on Chinese event extraction mainly focuses on biomedicine, microblog, news, emergencies, etc. There are few studies in other fields and open domains. There are fewer studies on domain and cross-language event extraction

technologies.

3) Lack of large-scale mature corpus and labeling corpus, the corpus needs further improvement. The manual labeling of corpus is time-consuming and labor-intensive, and the lack of corpus restricts the development of event extraction technology research. Therefore, the automatic construction technology method of large-scale corpus needs further research.

## V. TRENDS IN EVENT EXTRACTION TECHNOLOGY

With the deepening of research and the extensive application of advanced technologies such as artificial intelligence and big data, it is foreseeable that event extraction technology will develop rapidly in future research and will show the following development trends:

1) With the continuous development of related technologies such as entity, relationship recognition, and syntax analysis, the accuracy and recall rate of event extraction will be further improved, and new technical methods such as deep learning will be widely used.

2) With the development of cross-text semantic understanding and multi-language text processing technology, cross-text and cross-language event extraction research will be more extensive, and related application systems will be continuously developed.

3) Future event extraction research will focus on applications, and the field will continue to expand, no longer limited to a specific field, but more oriented to the open field, and the portability of the system will be further improved.

4) The related corpus automatic construction technology will make a breakthrough, no longer need a lot of artificial energy, and the enrichment of corpus will greatly promote the development of event extraction technology.

## REFERENCES


[1] Grishman R. Information extraction [J].The Handbook of Computational Linguistics and Natural Language Processing, 2003: 515-530.

[2] Ahn D. The stages of event extraction [J] . Arte'06 Proceedings of the Workshop on Annotating & Reasoning About Time & Events, 2006:1-8.

[3] ACE ( Automatic Content Extraction )Chinese Annotation Guidelines for Events. National Institute of Standards and Technology [R] . 2005.

[4] Meiying Jia, Bingru Wang, Dequan Zheng. Information Extraction of Military Exercise Based on Pattern Matching [J].Intelligence Analysis and Research, 2009, 183(09):70-75.

[5] Jifa Jiang. Research on the information extraction pattern of free text[D].Chinese Academy of Sciences, 2004.

[6] Ming Luo, Hailiang Huang. A method of extracting financial event information based on lexical-semantic model [J].Computer Applications, 2018, 38 (1): 84-90.

[7] Shasha Liao, Grishman R. Filtered Ranking for Bootstrapping in Event Extraction[C]. Proceedings of the 23rd International Conference on Computational Linguistics, Beijing, China, 2010, 680-688.

[8] Jiangde Yu, Xinfeng Xiao, Xiaozhong Fan. Chinese text event information extraction based on hidden Markov model [J]. Microelectronics and Computer, 2007, 24 (10): 92-94+98.

[9] Bolei Hu, Ruifang He, Hong Sun, Wenjun Wang. Chinese event type recognition based on conditional random field [J].Pattern recognition and artificial intelligence, 2012, 25 (3): 445-449.

[10] Yanyan Zhao. Related technology research on Chinese event extraction [D].Harbin: Harbin Institute of Technology, 2007.

[11] Honglei Xu et al. Research on Chinese Event Extraction Technology for Automatic Recognition of Event Categories [J]. Mind and Computing, 2010, 4 (1): 3444.

[12] Yajun Zhang, Zongtian Liu, Wen Zhou. Event Recognition Based on Deep Belief Network [J]. Journal of Electronics, 2017 (5): 1416-1423.

[13] Kai Wang. Research on English event extraction based on deep learning [D]. Soochow University , 2017.

[14] Zhengkuan Zhang. Research on event extraction based on structured learning [D].Beijing University of Posts and Telecommunications, 2017.