

# Joint Extraction Model of Entities and Events

Can Tian<sup>1, a</sup>, Yawei Zhao<sup>1, b, \*</sup>, Liang Ren<sup>2, c</sup>

<sup>1</sup>Big Data Analysis Technology Lab University of Chinese Academy of Sciences Beijing, China

<sup>2</sup>Big Data Analysis Technology Lab Beijing Knowlegene Data Technology Co., Ltd. Beijing, China

<sup>a</sup>tiancan@hust.edu.cn, <sup>b</sup>\*zhaoyw@ucas.ac.cn, <sup>c</sup>renliang@knowlegene.com

**Abstract.** Joint extraction of entities and events is an important task in information extraction. In order to obtain entities and events in the text simultaneously, in this paper we firstly propose a novel tagging scheme that can transform the joint extraction task to a tagging problem. Then, based on our tagging scheme, we use different end-to-end models to extract entities and events directly and we also propose an improved objective function with different parameters to express the importance of different labels. We conduct experiments on a financial dataset and the results show that our methods are better than other existing models.

**Keywords:** entity recognition, event extraction, the joint model.

## 1. Introduction

Entities and events are the basis for building the knowledge graph. Accurately identifying entities and events in the text can improve the effectiveness of tasks such as risk conduction and graph completion. Entity recognition is the pre-order task of event extraction, but most of existing methods to solve the two tasks are pipeline manner, recognizing the entities first and then extracting their events, each task is an independent model makes the task simple to deal with, and each component can be more flexible, but it ignores the connection between the two tasks, the results of entity recognition may affect the performance of event extraction and lead to erroneous delivery. Thus, consider a joint model which can combine the two tasks to become a possibility.

There are many models about entity recognition research, which is mainly based on the LSTM+CRF framework [1] and new features in different fields or languages are added, such as English character vectors [2], Chinese radicals [3], the new mechanism attention [4] and model fusion [5], as well as the latest BERT model [6] also has improved the effect of entity recognition. In general, event extraction is divided into two parts, the recognition of event trigger and event argument extraction, traditional methods to deal with them in a pipelined manner, e.g., the Dynamic Multi-Pooling Convolutional Neural Networks [7]. There also are some researcher combine the two parts together to extract them, Q Li proposed a joint framework based on structured prediction [8] which extracts trigger and arguments together. And there are some related works about event detection [9] which only need to identify if an event in a sentence. Compared to English event extraction, there are few works about Chinese event extraction for lacking corpus, but small researches [10].

Different from above methods, the method proposed in this paper is based on a novel tagging scheme that can convert the two tasks to a tagging problem so that we can use the end-to-end model to extract entities and events together and we also propose a new objective function with different parameters to express the importance of different labels. We conduct experiments on a financial dataset to validate our method, the experimental results show that our approach is effective to extract entities and events, and better than other existing models. The major contributions of this paper are: (1) A novel tagging scheme combine entity and event is proposed. (2) Latest BERT model is used to settle the joint extraction problem. (3) An improved objective function is proposed to express the importance of different labels.

## 2. Task Description

The entity recognition and event extraction are the subtasks of information extraction which the concepts are defined in ACE, we introduce some ACE terminologies to facilitate the understanding of the tasks:

Entity: an object or a set of objects in one of the semantic categories of interests.

Entity mention: a reference to an entity, it is a noun phrase in most cases.

Event trigger: the main word that most clearly expresses an event occurrence, it can be verbs and nouns.

Event arguments: the mentions that are involved in an event, all of them are entities.

Event mention: a phrase or sentence within which an event is described, including the event trigger and arguments.

## 3. The Proposed Method

In this section, we first introduce the tagging scheme which convert the joint extraction task to a tagging problem. Then we discuss how to use the end-to-end model to deal with the problem and describe the improved objective function that we proposed in this paper.

### 3.1 The Tagging Scheme

For a given text, each token is assigned a label that contributes to extract the results, we use “BIO” annotation to indicate the beginning, internal, and the “other” labels of the entities in this paper. The entity recognition task is mainly for the purpose of better assisting the extraction of event arguments here. Therefore, for the text of an event, in addition to getting the event trigger and event arguments, it is necessary to mark the entities with the same entity type as the argument of the event. At the same time, by adding a pair of event type identifiers at the beginning and the end of text, it is made to mix the corpus of multiple events to training become a possibility, the two identifiers will prompt the model which event the text belongs to, and use the standard of the event to label it. The specific tagging scheme is shown in Fig. 1.

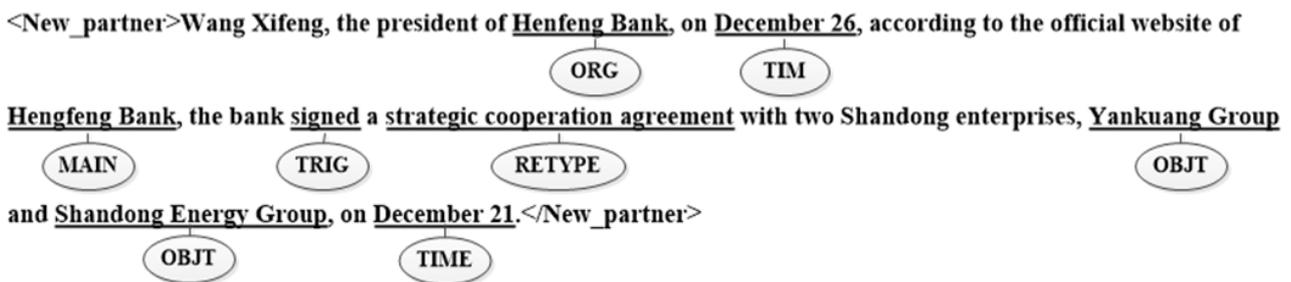


Fig. 1 An example of tagging scheme

The “New\_partner” indicates the event type, and “ORG” and “TIM” indicate irrelevant organizations and time that are not related to the event. “TRIG” expresses the event trigger, “MAIN”, “TIME”, “OBJT”, “RETYPE” represent the argument of the event (host organization, time of occurrence, object organization and the type of cooperation). It can be found that in the unrelated entries marked "O", "Wang Xifeng" and "Shandong" are entities of person and place, it's unnecessary to label them because the entity type of the event argument does not involve these entities.

### 3.2 End-to-end Model

The End-to-end model is widely used to solve the sequence tagging problem. The LSTM+CRF framework is the most typical model, and it has achieved better results in different sequence tagging tasks. On this basis, by adding word2vec, attention mechanism, advanced transformer, and the latest BERT model have improved the effect of different tasks, the following lists several commonly used models.

•word2vec+LSTM+CRF: Since word2vec was proposed and replaced with the one-hot word vector, most of the better results obtained on the sequence tagging tasks are based on or related to the model, see Fig. 2.

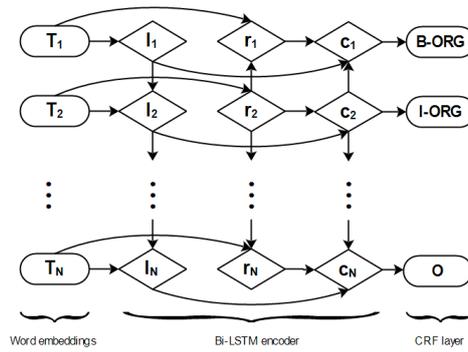


Fig. 2 word2vec+LSTM+CRF model

•BERT+ SoftMax: Google proposed the BERT model recently, which is considered to be a milestone in the NLP field. Similar to word2vec, the model uses a large number of corpus training models to represent language and provide the representation as a feature to downstream tasks. Different from the existing represent models, BERT has different degrees of improvement in the model framework, pre-training objective function and training method, and achieves better results on many NLP tasks. The output is directly connected to the BERT model followed by SoftMax on tagging problem, which also enhances the effect of the model, as shown in Fig. 3.

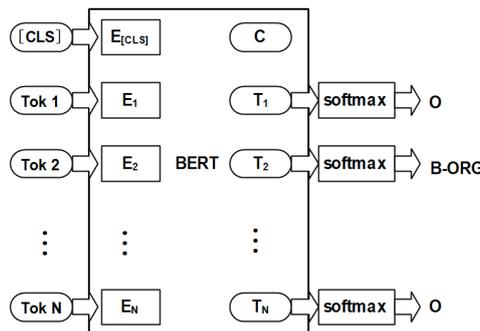


Fig. 3 BERT+ SoftMax model

•BERT+LSTM+CRF: When multiple events are trained together, the arguments of each event are different, the number of entities of the joint extraction task will increase. The result will be greatly influenced if continue to use SoftMax as the output function, it is difficult to capture the connection of labels, such as the beginning of the entity "B" cannot be followed by a "B" and "I" cannot appear independently. So, continue to use the framework of LSTM + CRF, while BERT is just a representation of language to replace the word vector generated by word2vec, as shown in Fig. 4.

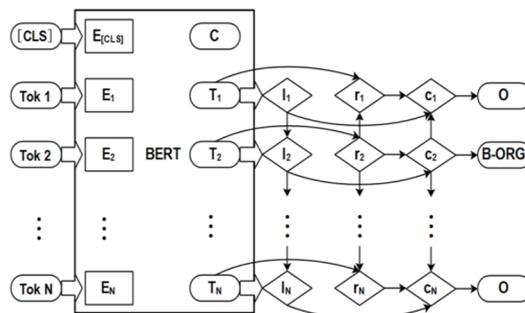


Fig. 4 BERT+LSTM+CRF model

### 3.3 Improved Objective Function

It can be seen from the tagging scheme that the irrelevant organization (ORG), the host organization (MAIN) and the object organization (OBJT) are the same entity type, but we treat them as different tags (time, amount, et al.). To express the correlation between them, two sets of tag sequences are generated during data processing, the first sequence  $y_1$  is the entity type tag, the second  $y_2$  is the event tag, and the losses are calculated in the model by the two tags, then optimize the two losses, and for each of the label  $i$ , its importance is different, for example, in some events, we mainly focus on participants, regardless of the time of occurrence or the amount involved, etc., so it is possible to introduce different weights into each of the tags to calculate the loss. Combining the above two strategies, the objective function is defined as follows.

$$L = \max \sum_{j=1}^{|D|} \sum_{t=1}^{L_j} (\beta_1 (\log(p_{1t}^{(j)} = y_{1t}^{(j)} | x_j, \Theta)) + \beta_2 \alpha_i \cdot \log(p_{2t}^{(j)} = y_{2t}^{(j)} | x_j, \Theta)) \quad (1)$$

The  $|D|$  is the size of training set,  $L_j$  is the length of the sentence  $x_j$ ,  $p_{1t}^{(j)}$ ,  $y_{1t}^{(j)}$  refers to the prediction label and the event label of token  $t$  in sentence  $x_j$  respectively, and  $\beta_1$ ,  $\beta_2$  represent the weight of the losses which are calculated by  $y_1$  and  $y_2$ ,  $\alpha_i$  indicating the weight of each tag, the value can be determined according to the importance of different event arguments.

## 4. Experiments

In this section, we first introduce the experimental setting, including the dataset that we chose, models for comparing and the typical hyperparameters of the models. Then we report the results of different methods and analyze the results of a model.

### 4.1 Experimental Setting

**Dataset:** Most of the papers to extract events are based on ACE2005 corpus. There is little research on Chinese while ACE2005 only has 633 Chinese documents with less than 2000 events, the training data are seriously lacking, and the event type is sparse. We use the financial datasets to evaluate the performance of our methods, including 27 types of events, with a total of 32,480 annotation sentences.

- **Evaluation:** We chose standard Precision, Recall and F1 score to evaluate the experimental results.
- **Baselines:** For comparing with the results of the end-to-end model, we reproduced two benchmark methods structured perceptron and DMCNN based on the dataset.
- **Hyperparameters:** There are five models mentioned, we choose the beam search size equal to 4 in structured perceptron, the dimension of word embeddings, positions, events are 100, 5, 5 and the filter size, feature map size are 3, 100 respectively in DMCNN, the dimension of word embeddings, LSTM are 100, 100 in word2vec+LSTM+CRF, we choose the max sequence length equal to 128 in BERT+ SoftMax, and  $\beta_1 = 1$ ,  $\beta_2 = 1$ ,  $\alpha = (5, 2, 2, 1)$  where the parameter  $\alpha$  corresponding to the weight of event trigger and host organization, other event arguments, other entities, "O" respectively in BERT+LSTM+CRF+L.

### 4.2 Experimental Results

We report the results of different methods as shown in Table 1, it can be seen that the end-to-end models outperform the contrast models in F1 score, this means that our tagging scheme is helpful to the event extraction and it is always useful no matter which end-to-end model you choose. We can conclude that the improved objective function is good for extracting the event arguments from the last two experiments even though the effect of the entities will be reduced. For a single model, the

results are represented in Table 2, it can be seen that the value of  $\alpha$  and  $\beta$  have an impact on different tags.

Table 1. Experimental Results of Different Models

Methods	Entity			Trigger			Argument		
	P	R	F	P	R	F	P	R	F
Structured perceptron	-	-	-	75.30	80.10	77.60	54.30	61.54	57.69
DMCNN	-	-	-	85.53	87.39	86.43	63.55	64.19	63.87
Word2vec+LSTM+CRF	82.59	91.07	86.62	82.31	85.21	83.74	66.46	69.51	67.95
BERT+softmax	87.27	91.43	89.30	84.02	89.87	86.85	70.74	77.09	72.96
BERT+LSTM+CRF	<b>89.46</b>	<b>92.88</b>	<b>91.14</b>	<b>86.93</b>	91.24	89.03	<b>73.59</b>	77.58	75.53
BERT+LSTM+CRF+L	88.71	91.67	90.16	85.99	<b>93.51</b>	<b>89.59</b>	72.37	<b>82.12</b>	<b>76.94</b>

Table 2. Experimental Results of a Single Model

Entity	BERT+LSTM+CRF+L			Trigger and Event argument	BERT+LSTM+CRF+L		
	P	R	F		P	R	F
ORG	86.54	89.92	88.20	TRIG	85.99	93.51	89.59
TIM	93.24	94.81	94.02	MAIN	81.39	86.46	83.85
MON	90.32	93.03	91.65	OBJT	74.73	79.53	77.05
PERC	75.51	88.56	81.52	TIME	69.05	75.75	72.24
PER	85.71	90.32	89.96	MONEY	73.29	76.01	74.63
OTH	68.35	73.40	70.78	RETYPE	87.81	93.51	90.57
...	...	...	...	QUTYPE	53.40	53.12	53.26

### 4.3 Error Analysis

The experimental results of each model the precision rate is lower than the recall rate, the main reason is the terrible quality of the experimental dataset, especially the entities that are not event arguments are not marked. We will not extract the event arguments if the event trigger is not found for the DMCNN as a pipelined method, so the results is unexpected for the arguments even though the extraction of event trigger is better, this is why we choose the joint model to extract entities, event trigger, and arguments.

For the results of a single model, the extraction of event triggers and host organizations are better because we choose bigger weight parameters for them. The effect of time and amount about the events is lower while the results of related entities are better, for there are more than one related entity in the sentence which are made the model confused. The results of other special entities are staggered for the dataset lack of diversity.

### 4.4 Analysis of Parameters

In order to evaluate the impact of  $\alpha$  and  $\beta$ , we change the parameter  $\beta$  from 0 to 2.5 and choose several sets of  $\alpha$ , the experimental results (including trigger and event arguments) as shown in Fig. 5 and Fig. 6.

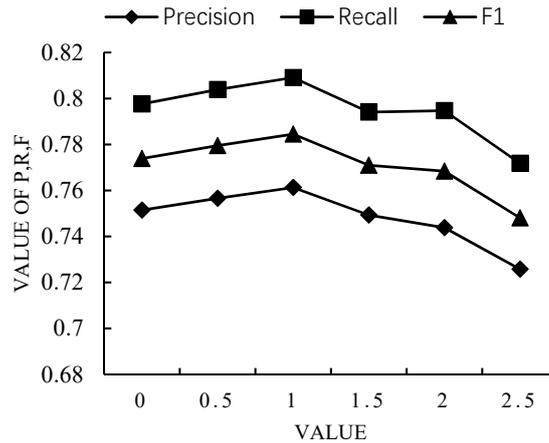


Fig. 5 The results on different parameter  $\beta$

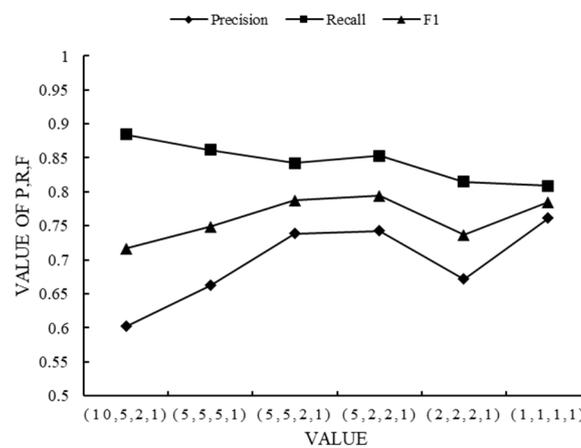


Fig. 6 The results on different parameter  $\alpha$

It can be seen from Fig. 5 that the result of  $\beta=1$  achieves a 1% improvement in F1 score over the  $\beta=0$ , and it is inversely proportional to  $\beta$  when  $\beta$  become bigger. Besides, in order to express the importance of event trigger and host organization in an event, we assign a bigger weight to them, we can conclude from Fig. 6 that if  $\alpha$  is too large, it will improve the rate of recall but affect the accuracy of prediction. When  $\alpha = (5, 2, 2, 1)$ , it can achieve a 1% improvement in F1 score.

## 5. Conclusion

Entity recognition is the pre-order task of event extraction, the results of entity recognition may affect the performance of event extraction and lead to erroneous delivery, the same to event trigger and argument. In this paper, a novel tagging scheme is proposed to covert the joint extraction task to a tagging problem, then we use the end-to-end model to extract entities and events together and we also propose a new objective function with different parameters to express the importance of different labels. The experimental results show that the results of the end-to-end model is better than other existing pipelined models, the new parameters are also improved the results. However, sometimes this model cannot extract the event arguments completely or cannot recognize multiple identical arguments, and there are some problems to handle multiple different events in a sentence, which can be improved in the future works.

## Acknowledgments

We want to thank Beijing Knowlegene Data Technology Co., Ltd. for dataset details. This work is also supported by the 13th Five-Year Plan for Information Science of the Chinese Academy of Sciences (No. XXH13504-05) and the National Natural Science Foundation of China (No. 61872331).

## References

- [1]. Huang Z, Xu W, Yu K. Bidirectional LSTM-CRF Models for Sequence Tagging [J]. *Computer Science*, 2015.
- [2]. Lample G, Ballesteros M, Subramanian S, Kawakami K, Dyer C. Neural Architectures for Named Entity Recognition [J]. 2016:260-270.
- [3]. Dong C, Zhang J, Zong C, Hattori M, Di H. Character-Based LSTM-CRF with Radical-Level Features for Chinese Named Entity Recognition [C]// *International Conference on Computer Processing of Oriental Languages*. Springer International Publishing, 2016:239-250.
- [4]. Vaswani A, Shazeer N, Parmar N, Uszkoreit J, Jones L, Aidan N. Attention Is All You Need [J]. 2017.
- [5]. Peters ME, Ammar W, Bhagavatula C, Power R. Semi-supervised sequence tagging with bidirectional language models [J]. 2017.
- [6]. Devlin J, Chang M W, Lee K, Toutanova K. BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding [J]. 2018.
- [7]. Chen Y, Xu L, Liu K. Event Extraction via Dynamic Multi-Pooling Convolutional Neural Networks [C]// *the Meeting of the Association for Computational Linguistics*. 2015.
- [8]. Q Li, H Ji, L Huang. Joint Event Extraction via Structured Prediction with Global Features [C]// *the Meeting of the Association for Computational Linguistics*. 2013.
- [9]. Liu S, Chen Y, Liu K. Exploiting Argument Information to Improve Event Detection via Supervised Attention Mechanisms [C]// *Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*. 2017.
- [10]. Li P, Zhou G. Joint Argument Inference in Chinese Event Extraction with Argument Consistency and Event Relevance. [J]. *IEEE/ACM Transactions on Audio Speech & Language Processing*, 2016, 24(4):612-622.
- [11]. Zheng S, Wang F, Bao H, Hao Y, Zhou P, Xu B. Joint Extraction of Entities and Relations Based on a Novel Tagging Scheme [J]. 2017.
- [12]. Xia X, Peifeng L, Xin Z, Qiaoming Z. Event inference for semi-supervised Chinese event extraction [J]. *Journal of Shandong University*, 2014.
- [13]. Ma X, Hovy E. End-to-end Sequence Labeling via Bi-directional LSTM-CNNs-CRF [J]. 2016.
- [14]. Chiu JPC, Nichols E. Named Entity Recognition with Bidirectional LSTM-CNNs [J]. *Computer Science*, 2015.
- [15]. Yang B, Mitchell T. Joint Extraction of Events and Entities within a Document Context [J]. 2016.
- [16]. Nguyen T H, Cho K, Grishman R. Joint Event Extraction via Recurrent Neural Networks [C]// *Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*. 2016.

- [17]. Zhou D, Zhang X, He Y. Event extraction from Twitter using non-parametric Bayesian mixture model with word embeddings [C]// Conference of the European Chapter of the Association for Computational Linguistics. 2017.
- [18]. Rao S, Marcu D, Knight K, Daume H. Biomedical Event Extraction using Abstract Meaning Representation [C]// Bi-oNLP 2017. 2017.
- [19]. Mikolov T, Chen K, Corrado G, Dean J. Efficient Estimation of Word Representations in Vector Space [J]. *Computer Science*, 2013.
- [20]. Mikolov T, Sutskever I, Chen K, Corrado G, Dean J. Distributed Representations of Words and Phrases and their Compositionality [J]. *Advances in Neural Information Processing Systems*, 2013, 26:3111-3119.
- [21]. Peters M E, Neumann M, Iyyer M, Gardner M, Clark C, Lee K. Deep contextualized word representations [J]. 2018.