

Using dual-layer CRFs for event causal relation extraction

Jianfeng Fu^{1,2a)}, Zongtian Liu², Wei Liu², and Qiang Guo²

¹ School of Mathematics & Information, Shanghai Lixin University of Commerce, Shanghai, 201620, China

² School of Computer Engineering & Science, Shanghai University, Shanghai, 200072, China

a) fujianfeng2000@gmail.com

Abstract: Traditional methods for event causal relation extraction covered only part of the explicit causal relation in text. This paper presents a method for event causal relation extraction by using dual-layer Conditional Random Fields (CRFs). The method casts the problem of event causal relation extraction as event sequence labeling and employs dual-layer CRFs model to label the causal relation of event sequence. The first layer of the CRFs model is used to label the semantic role of causal relation of the events, and then the output of the first layer is passed to the second layer for labeling the boundaries of the event causal relation. Experimental results show that our method not only covers each class of explicit event causal relation in the text, but also achieves good performance and the F-Measure of the overall performance arrives at 85.3%.

Keywords: event causal relation, event sequence, dual-layer CRFs, Conditional Random Fields

Classification: Science and engineering for electronics

References

- [1] R. Girju, “Automatic detection of causal relations for question answering,” *Proc. ACL 2003 Workshop on Multilingual Summarization and Question Answering*, pp. 76–83, 2003.
- [2] T. Inui, K. Inui, and Y. Matsumoto, “What kinds and amounts of causal knowledge can be acquired from text by using connective markers as clues?,” *Lecture Notes in Computer Science*, pp. 180–193, 2003.
- [3] C. Pechsiri and A. Kawtrakul, “Mining causality from texts for question answering system,” *IEICE Trans. Inf. & Syst.*, vol. 90, no. 10, pp. 1523–1533, 2007.
- [4] C. Khoo, S. Chan, and Y. Niu, “Extracting causal knowledge from a medical database using graphical patterns,” *Proc. 38th Annual Meeting on Association for Computational Linguistics*, pp. 336–343, 2000.
- [5] E. Blanco, N. Castell, and D. Moldovan, “Causal relation extraction,” *Language Resources and Evaluation*, pp. 310–313, 2008.
- [6] J. Lafferty, A. McCallum, and F. Pereira, “Conditional random fields: Probabilistic models for segmenting and labeling sequence data,” *Machine Learning*, vol. 951, pp. 282–289, 2001.

1 Introduction

The automatic extraction of event causal relation in text has become increasingly important in question answering and received considerable attention in recent years [1, 2, 3]. Consider the following example.

A Cherry Hill woman died from fire on Feb. 1, 2010.

The question answering systems need to extract the causal relation pair of “fire $\rightarrow$ death” from the sentence to answer user query such as “What led to the woman’s death?”.

Causation is a semantic relation between two events: *cause* and *effect*. If both *cause* and *effect* are represented in text, then the causal relation is *explicit*, otherwise it is *implicit*. The explicit causal relation also can be classified as *marked* and *unmarked*. An explicit causal relation is marked if there is a specific linguistic unit that indicates the relation; unmarked otherwise. For example, “The driver died *because of* the accident” is marked; “Robert was killed in a car accident” is not. Besides, the explicit causal relation has style of “*one cause, one effect*” and “*one cause, many effects*”. In this paper, we focus on explicit causal relation. For the sake of convenience, the explicit causal relation is simplified as causal relation in the rest of the paper.

Many previous studies have attempted to extract causal relation from text. The approach of pattern matching was used to extract the marked causal relation which with $\langle \textit{cause indicator effect} \rangle$ pattern in text [1, 4]. Besides, the machine learning technique was also introduced to extract the marked causal relation [5]. However, traditional methods for event causal relation extraction were limited to the marked and “one cause, one effect” which covered only part of the causal relation in text. To overcome this limitation, we present a new method for event causal relation extraction. The method casts the problem of event causal relation extraction as the labeling of event sequence which using dual-layer Conditional Random Fields [6] (CRFs). By using this method, not only the causal relation of marked and “one cause, one effect” but also the unmarked and “one cause, many effects” can be extracted. Experimental results show that our method achieves good performance.

2 Proposed method

2.1 Event sequence labeling

Traditional machine learning method often took the problem of event causal relation extraction as a classification task. Classifiers, for example, *Decision Tree* and *Support Vector Machine*, were trained to identify the causation components of cause and effect in text. Different from this method, our method takes the problem of event causal relation extraction as event sequence labeling. Before the detailed discussing of our method, we would like to give some useful definitions as follows.

- **Event:** An event is something that happens or a specific occurrence involving participants.

- **Event trigger:** An event trigger is the word that most clearly expresses an event occurrence. For example: earthquake, die, injure, and so on.
- **Event sequence:** An event sequence is a chain that consists of event triggers according to their order in text. Consider the following example.

*Yesterday morning, a house **caught fire**, two elderly people were **killed** and the hostess was **injured**. It is estimated that the **fire** caused by a **short circuit**. The police are **investigating** the accident.*

The event sequence of above sentences can be described as: *caught fire, killed, injured, fire, short circuit, investigating*.

The method of event sequence labeling is to assign two labels to each event in the sequence with two steps. In the first step, the label is tagged to indicate semantic role of causal relation of the event (whether the event is a “cause” (C), an “effect” (E) or “others” (N)). In the second step, the label is tagged to indicate boundaries of event causal relation (whether the event is outside causation (O), starts causation (B) or continues causation (I)).

It is easy to extract event causal relation from a labeled event sequence. Firstly, the causation extent can be identified by the boundary labels. Secondly, the semantic role of event in each causal relation can be identified by the labels of C, E, and O.

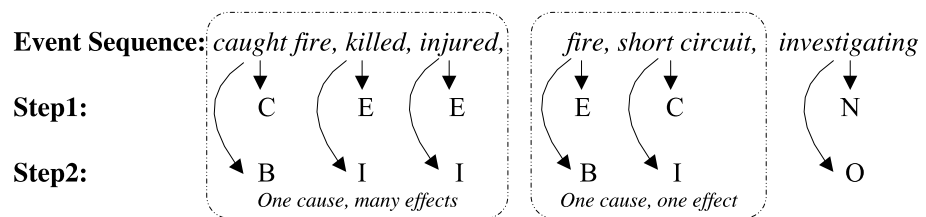


Fig. 1. The labeling process of the event sequence

Fig. 1 shows the labeling process of the event sequence. There are two causal relations contained in the labeled event sequence. One is “caught fire → killed, caught fire → injured”, the other is “short circuit → fire”. The first one is an unmarked, “one cause, many effects” causal relation, and the second one is a marked, “one cause, one effect” causal relation. From Fig. 1, it can be seen that not only “one cause, one effect” but also “one cause, many effects” can be handled with our method. Besides, the method is not limited to the marked causal relation; the unmarked is also appropriated.

2.2 Dual-layer CRFs

Conditional Random Fields [6] are undirected graphical models that are conditionally trained. A common special-case graph structure is a linear chain, which corresponds to a finite state machine, and is suitable for sequence labeling. Given an input sequence $X = x_1 \dots x_N$, a linear chain CRF with parameters $\Lambda = \{\lambda, \dots\}$ defines a conditional probability for a label sequence

$Y = y_1 \dots y_N$ as:

$$P_{\Lambda}(Y|X) = \frac{1}{Z_X} \exp \left(\sum_{n=1}^N \sum_k \lambda_k f_k(y_{n-1}, y_n, X, n) \right) \quad (1)$$

Where Z_X is the per-input normalization that makes the probability of all label sequences sum to one; $f_k(y_{n-1}, y_n, X, n)$ is a feature function which is often binary-valued, and λ_k is a learned weight associated with feature f_k .

The most probable label sequence for an input X ,

$$Y^* = \arg \max_y P_{\Lambda}(Y|X) \quad (2)$$

can be efficiently calculated by dynamic programming using the Viterbi algorithm.

Given training set $\{(X_i, Y_i): i = 1, \dots, M\}$, the parameters can be estimated by maximizing the conditional log-likelihood.

$$\begin{aligned} L_{\Lambda} &= \sum_i \log P_{\Lambda}(Y_i|X_i) \\ &= \sum_i \left(\sum_{n=1}^N \sum_k \lambda_k f_k(y_{n-1}, y_n, X, n) - \log Z_{xi} \right) \end{aligned} \quad (3)$$

The quasi-Newton gradient-climber, BFGS, (which has been found that it converges much faster than traditional maximum entropy learning algorithms, such as GIS and IIS) is used to train CRFs in our experiments.

As mentioned above, the task for labeling the causal relation of event sequence contains two steps. Correspondingly, we employ a dual-layer CRFs model to the task. In the first layer, features, such as “*trigger*”, “*part of speech of trigger*” and “*event type*”, are used to train the model and predict the semantic role of causal relation of event sequence. With the features that are used in the first layer, the output of the first layer is passed to the second layer to train the model and predict the boundaries of the event causal relation.

3 Experiments

For evaluating our method, we built a corpus by using 200 emergency news articles which had been collected from internet. Each article in the collection was annotated and checked by two group annotators. The detailed information of the corpus is shown in Table I. 3/4 of the corpus is used as training set, the rest as test set. The experimental results are measured by recall, precision and F-Measure and summarized in Table II.

Table I. Information of the corpus

Relation type	Amount	Rate (%)
One cause, one effect	166	48.8
One cause, many effects	174	51.2
Marked	133	39.1
Unmarked	207	60.9
Overall	340	100

From Table I, it can be seen that the “one cause, one effect” and marked is a subset of event causal relation in text. There are a large number of “one cause, many effects” (51.2%) and unmarked (60.9%) event causal relation. Therefore, traditional method covers only part of the causal relation in text.

Table II. Experimental results for event causal relation extraction

Relation type	precision	recall	F-Measure
One cause, one effect	95.3%	85.7%	90.2%
One cause, many effects	83.9%	78.3%	81.0%
Marked	90.6%	81.4%	85.8%
Unmarked	88.3%	81.9%	85.0%
Overall	89.2%	81.7%	85.3%

As is shown in Table II, our method achieves 90.2% F-Measure on “one cause, one effect”, which is better than “one cause, many effects” (81.0% F-Measure). In addition, our method obtains good performance on unmarked as well as marked event causal relation (85.8% versus 85.0% F-Measure). Our method covers each class of event causal relation in the text, and the F-Measure of overall performance arrives at 85.3%.

4 Conclusion

This paper presented a new method for event causal relation extraction. The method casts the problem of event causal relation extraction as event sequence labeling by using dual-layer CRFs model. Experimental results show that our method not only covers each class of event causal relation in text, but also achieves good performance and the F-Measure of the overall performance arrives at 85.3%.

Acknowledgments

This work is supported by the projects of National Science Foundation of China (NSFC No.60575035, No.60975033), Shanghai Leading Academic Discipline Project (J50103) and Postgraduate Innovation Fund of Shanghai University (SHUCX091041).