EEG: Knowledge Base for Event Evolutionary Principles and Patterns

Zhongyang Li, Sendong Zhao, Xiao Ding, and Ting Liu(✉)

Research Center for Social Computing and Information Retrieval,
Harbin Institute of Technology, Harbin, China
{zyli,sdzhao,xding,tliu}@ir.hit.edu.cn

Abstract. The evolution and development of events has its underlying principles, leading to events happened sequentially. Therefore, the discovery of such evolutionary patterns between events are of great value for event prediction, decision-making and scenario design of dialog system. In this paper, we propose **Event Evolutionary Graph** (EEG), which reveals evolutionary patterns and development logics between events. Specifically, we propose to construct EEG by recognizing the sequential relation between events and the direction of each sequential relation. For sequential relation and direction recognition, we explore the effectiveness of 4 categories of features: count-based, ratio-based, context-based and association-based features for correctly identifying sequential relations and corresponding directions. Experimental results show that (1) the framework we proposed is promising for EEG construction and (2) methods we proposed are effective for both sequential relation and direction recognition.

Keywords: Event Evolutionary Graph · Sequential relation between events · Social media · Knowledge base

1 Introduction

The evolution and development of events have its underlying principles, making events happen sequentially. This kind of evolutionary patterns of events is of great value. For example, the sentence "After having lunch, Tom paid the bill and left the restaurant" shows a sequence of events evolution ① "have lunch", ② "pay the bill" and ③ "leave the restaurant". This event series is a very ordinary pattern for the scenario of having lunch in a restaurant. This kind of pattern is very common in our daily life, which usually indicate the basic patterns of events evolution and human behaviors. Hence, these patterns of evolutionary events are of great value and important for many tasks, such as event prediction, decision-making and scenario design of dialog system.

Numerous efforts have been dedicated to extracting temporal and causal relations from texts. As the most commonly used corpus, the TimeBank corpus [15] has been adopted in a lot of temporal relation extraction studies. Mani et al. [9]

X. Cheng et al. (Eds.): SMP 2017, CCIS 774, pp. 40–52, 2017.
https://doi.org/10.1007/978-981-10-6805-8_4

applied the temporal transitivity rule to greatly expand the corpus. Chambers et al. [4] used previously learned event attributes to classify the temporal relationship. For causality relation extraction, Zhao et al. [19] extracted multiple kinds of features to recognize causal relations between two events in the same sentence. Radinsky et al. [16] automatically extracted cause-effect pairs from large quantities of news titles by predefined causal templates, and then they used them to predict news events. However, these studies have some limitations. First, this line of work can only extract relations from single sentences. Second, these studies extract relations based on the semantic of specific context rather than discover the underlying patterns of event evolution from large-scale user generated documents.

In order to discover patterns of evolutionary events, we propose **Event Evolutionary Graph** (EEG) and the framework to construct EEG. Specifically, our definition of EEG involves evolutionary patterns and development logics of events. The EEG is composed of events and relations between them. For the sake of generality, we consider sequential and causal relations in EEG. The value of each relation denotes the transition probability between events. Therefore, the construction of EEG can be simplified as two key problems. The first is to recognize relations between each two events. The second is to distinguish the direction of each relation between events. Both problems can be solved based on the classification framework.

The main contributions of this paper are as follows. First, we propose EEG and give its detailed definitions. Second, we propose a promising construction framework to build EEG from large-scale unstructured web corpus. Third, extensive experiments are conducted to solve the central problems of sequential relation and direction recognition. Experimental results show that the methods we developed are effective for both sequential relation and direction recognition.

2 Related Work

2.1 Statistical Script Learning

The use of scripts in Artificial Intelligence dates back to the 1970s [10]. In this conception, scripts are composed of complex events without probabilistic semantics. In recent years, a growing body of research has investigated learning probabilistic co-occurrence-based models with simpler events. Chambers et al. [3] proposed a co-occurrence-based model of (verb, dependency) pairs, which can be used to infer such pairs from documents. Pichotta et al. [12] described a method of learning a co-occurrence-based model of verbs with multiple arguments.

There have been a number of recent published neural models for script learning. Pichotta et al. [13] showed that an LSTM event sequence model outperformed previous methods for predicting verbs with arguments. Pichotta et al. [14] used a Seq2Seq model directly operating on raw tokens to predict sentences. Mark and Clark [7] described a feed-forward neural network which composed verbs and arguments into low-dimensional vectors.

Script learning is very similar to EEG in concepts. However, script learning usually extracts event chains without considering their temporal orders. Event definition and representation are also different. EEG aims to organize event evolutionary patterns into a commonsense knowledge base, which is the biggest difference between them.

2.2 Temporal Relation Extraction

A lot of annotation efforts have been devoted to construct corpora for building event ordering models. However, most of existing corpora focus on English. As the commonly used corpus in temporal relation extraction, the TimeBank corpus [15] has been adopted in a series of TempEval competitions [17,18], facilitating the development and evaluation of temporal relation extraction systems. But TimeBank corpus only annotates a small subset of easily-identified event mention pairs, which much limit its applications. To overcome this problem, Do et al. [6] produced an event-driven corpus on the ACE 2005 English corpus. Cassidy et al. [1] enriched the TimeBank-Dense corpus on top of TimeBank. In comparison, there are few corpora for Chinese temporal relation extraction.

Due to the corpus limitation, previous studies on temporal relation extraction focus on inferring temporal relations between event mentions in the same sentence or neighboring sentences from English text, dominated by feature-based approaches [4,8,9]. Chambers et al. [2] proposed a sieve-based architecture to joint those different tasks of temporal relation extraction. Mirza and Tonelli [11] used multiple cascaded classifiers to simultaneously solve the temporal and causal relation classification, and achieved the best experimental results.

All these studies solve temporal relation extraction from specific context. However, we solve this problem by frequency-based inference from multiple sentences, which is a commonsense reasoning process.

3 Event Evolutionary Graph

In this section, we give the detailed definition of EEG, which consists of event and relations between them. We consider two types of relation here, i.e., sequential and causal relations.

In EEG, events are represented by **abstract, generalized and semantic complete verb phrases**. Each event must contain a trigger word, which mainly indicates the occurrence of the event, and some other necessary components, such as the subject, object or modifier, to ensure the semantic completeness. **Abstract and generalized** means that we don't focus on the accurate happening location and time, and the exact subject of the event. **Semantic complete** means that human beings can understand the meaning of the event without vague and ambiguity. For example, "have hot pot", "watch movies", "go to the airport", are reasonable verb phrases to represent events. While "go somewhere", "do the things", "eat" are unreasonable or incomplete event representations, as they are too vague to understand.

Fig. 1. Tree structured event evolutionary graph under the scenario of "plan a wedding".

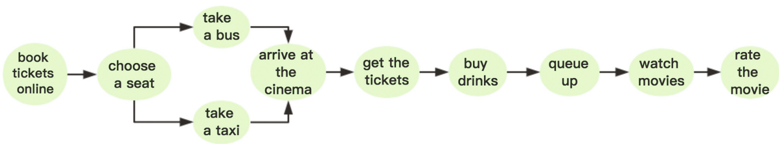

Fig. 2. Chain structured event evolutionary graph under the scenario of "watch movies".

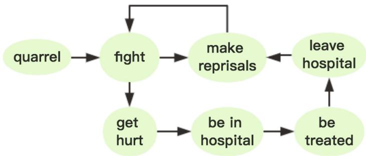

Fig. 3. Cyclic structured event evolutionary graph under the scenario of "fight".

The sequential relation between two events refers to their partial temporal orderings. For example, "After having lunch, Tom paid the bill and left the restaurant." "have lunch", "pay the bill" and "leave the restaurant" compose a sequential relation event chain.

Causality is the relation between one event (the cause) and a second event (the effect), where the second event is understood as a consequence of the first. It is obvious that the causal relation between events must be sequential. For example, "The nuclear leak in Japan leads to serious ocean pollution". "Nuclear leak in Japan" is the cause event, and "ocean pollution" is the effect event. Besides, the cause event happens before the effect event in temporal ordering. Hence, causal relation is a subset of sequential relation between events, and sequential relation is more general than causal relation.

EEG is a Directed Cyclic Graph, whose nodes are events, and edges stand for the sequential and causal relations between events. Essentially, EEG is a commonsense knowledge graph, which describes the event evolutionary patterns. Figures 1, 2 and 3 demonstrate three different event evolutionary subgraphs of three different scenarios. Concretely, Fig. 1 describe some event evolutionary patterns under the scenario of "plan a wedding", which happen repeatedly in people's daily life, and have evolved into some fixed human behavior patterns. For example, "plan a wedding" usually follows by "buy a house", "buy a car" and "plan a travel".

4 Methods for Constructing EEG

In this section, we propose a construction framework to construct EEG from large-scale unstructured text, including data cleaning, natural language processing, event extraction, event pair candidates extraction, sequential relation and direction recognition, causality recognition and transition probability computation. Figure 4 sketches this framework. Details about the main construction steps are described below. Causality is rare in the travel domain corpus we used. Hence, causality recognition is not covered in this paper.

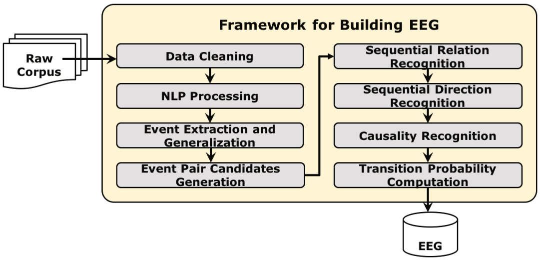

Fig. 4. Out proposed framework for building EEG from large-scale unstructured text.

4.1 Event Extraction

After cleaning the data, a series of natural language processing steps including segmentation, part-of-speech tagging, and dependency parsing are conducted for event extraction. Tools provided by Language Technology Platform [5] are used for this preprocessing.

Event extraction strategy is different from task to task, and it is mainly determined by the underlying task. As defined in Sect. 3, events in EEG are represented by abstract, generalized and semantic complete verb phrases, considering the corpus used in this paper. We extract verb-object phrases from the dependency-parsed tree. Though this is a simple strategy, we do find that a lot of high quality verb phrases are extracted.

We filter the low-frequency verb-object phrases by a proper threshold, to exclude the verb phrases extracted due to segmentation and dependency parsing errors. Some too general events such as "go somewhere" and "do something" are removed by regular expressions with a dictionary. Note that event extraction is not the key problem in this paper.

4.2 Event Pair Candidates Generation

Based on event extraction results, candidate event pairs are generated using two heuristic rules. First, every two events from single sentences are considered as an event pair candidate. Second, every two events from two consecutive sentences are considered as an event pair candidate as well. The event from the first sentence is taken as the first element of the pair, and the event from the second sentence as the second element.

For example, events A and B are extracted form the former sentence, and events C, D, E are extracted from the latter adjacent sentence. Ten event pairs are constructed as (A, B), (A, C), (A, D), (A, E), (B, C), (B, D), (B, E), (C, D), (C, E), (D, E).

Table 1. The features used for sequential relation and direction classification.

Count-based features	**Ratio-based features**
T1: count of (A, B) T2: count of (A, B) where A occurs before B T3: count of (A, B) where B occurs before A T4: count of A T5: count of B T6: count of verb-A T7: count of object-A T8: count of verb-B T9: count of object-B	R1: T2/T1, R2: T1/T4 R3: T1/T5, R4: T1/T6 R5: T1/T7, R6: T1/T8 R7: T1/T9, R8: T6/T4 R9: T7/T4, R10: T8/T5 R11: T9/T5
Context-based features	**Association-based features**
C1: count of all unique contexts C2: average length of all contexts C3: count of one-sentence contexts C4: count of two-sentence contexts C5: C3/C1 C6: contain relation of verb-A and verb-B C7: contain relation of object-A and object-B C8: postag of object-A C9: postag of object-B	A1: PMI of A and B A2: PMI of verb-A and verb-B A3: PMI of verb-A and object-B A4: PMI of object-A and verb-B A5: PMI of object-A and object-B

4.3 Sequential Relation and Direction Recognition

Given an event pair candidate (A, B), sequential relation recognition is to judge whether they have a sequential relation or not. For the ones having a sequential relation, direction recognition should be conducted to distinguish the direction. For example, we need to recognize there is a directed edge from "buy tickets" to "watches movies". We regard the sequential relation and direction recognition as two separate binary classification tasks. Previous temporal relation classification studies judge the relation and direction from single sentences. Alternatively, we solve this problem by frequency-based inference from multiple sentences. Specifically, we achieve this by forcing event A and B to co-occur more than k times. It is reasonable because event A and B are strongly associated with a high-frequency co-occurrence.

Multiple kinds of features are extracted for these two supervised classification tasks. All the features used are listed in Table 1. Details about the intuition why we choose these features are described below.

Count-Based Features: For a strongly associated event pair (A, B), there are many different statistics for them. They are directly counted from contexts that they co-occur or the whole corpus. We believe these statistics are effective to measure how strong and in which way event A and B are associated with each other. Therefore, they can be useful features for sequential relation and direction classification. These statistics include co-occur counts (T1 to T3), counts for events in the whole corpus (T4 and T5), and counts for verb and object in the events (T6 to T9).

Ratio-Based Features: The count-based features are numbers that directly counted from the context and whole corpus. Some meaningful combinations between these count numbers may provide extra information that is useful for sequential relation and direction classification. For example, if T2/T1 is close to 1, we can believe that A almost always occurs before B in the context they co-occur. It is a significant signal that the sequential direction is from A to B. These combination features are listed in Table 1 as ratio-based features (R1 to R11).

Context-Based Features: We believe that contexts exist where sequential candidates are more likely to appear. We developed features that capture the characteristics of likely contexts for sequential relations. In a nutshell, they include the count of unique context that A and B co-occur (C1), average length of all contexts (C2), the count of contexts that A and B are in the same sentence or not (C3 and C4), the ratio of one-sentence context (C5), the contain relation between verbs and objects (C6 and C7), and postag of objects (C8 and C9).

Association-Based Features: These features measure the association strength between event A and event B, including PMI scores from two different aspects. First, PMI score of (A, B) is computed as the macro measure of how strong event A and event B are associated with each other (A1). Second, we further consider the fine-grained PMI scores of (verb-A, verb-B), (verb-A, object-B), (object-A, verb-B) and (object-A, object-B), which measure the partial association strength of event components (A2 to A5).

4.4 Transition Probability Computation

Given an event pair (A, B), we use the following equation to approximate the transition probability from event A to event B:

$$P(B|A) = \frac{count(A, B)}{count(A)}, \tag{1}$$

where $count(A, B)$ is the co-occurrence count of event pair (A, B), and $count(A)$ is the occurrence count of event A in the whole corpus.

5 Experiments

In this section, we conducted two kinds of experiments. The first is to judge whether two events has sequential relation. And the second is to judge the partial ordering between two sequential events. These two steps are crucial and central for constructing clean and accurate EEG.

5.1 Dataset Description

We crawled 320,702 question-answer pairs from travel topic on Zhihu[1] as our experimental dataset. Travel is a relatively high level topic, which covers a wide range of things about traveling. Therefore, a lot of commonsense event evolutionary knowledge can be discovered from this data source.

Table 2. The detailed data statistics.

	Total	Positive	Negative
Sequential relation	2,173	1,563	610
Sequential direction	1,563	1,349	214

We annotate 2,173 event pairs with high co-occurrence frequency (>= 5) as our experiment corpus. Each event pair (A, B) is ordered that A occurs before B with a higher frequency than B occurs before A. In the annotation process, the annotators are provided with the event pairs and their corresponding contexts. They need to judge whether there is a sequential relation between two events from a commonsense perspective. If true, they also need to give the sequential direction. For example, "watch movies" and "listen to music" are tagged as no sequential relation (negative), while "go to the railway station" and "by tickets" are tagged as having a sequential relation (positive), and the sequential direction is from the former to the latter (positive). The detailed corpus statistics are listed in Table 2. The positive and negative examples are very imbalanced. So we over sample the negative examples in training set to ensure the number of positive and negative training examples are equal.

[1] https://www.zhihu.com/.

5.2 Compared Methods and Evaluation Metrics

For sequential relation recognition, PMI score of an event pair is used as the baseline method. For sequential direction recognition, if event A occurs before B with a higher frequency than B occurs before A, we think the sequential direction is from event A to event B. This is called the **Preceding Assumption**, which is used as the baseline method for sequential direction recognition.

For two experiments, four classifiers are used for these classification tasks, which are naive bayes classifier (NB), logistic regression (LR), multiple layer perceptron (MLP) and support vector machines (SVM). We explored different feature combinations to find the best feature set for both classification tasks. All experiments are conducted using five-fold cross validation. The final experiment result is the average performance of ten times of implementations.

Two kinds of evaluation metrics are used to evaluate the performance of our proposed methods. They are accuracy, and the precision, recall and F1 value.

5.3 Results and Analysis

Table 3 shows the experimental results for sequential relation classification, and we find that the pure PMI baseline achieves very good performance. Indeed, due to the imbalance of positive and negative test examples, PMI baseline chooses a threshold to classify all test examples as positive, and get a recall of 1. Four different classifiers with all features in Table 1 achieve poor results, and only the NB achieves higher performance than the baseline method. We explored all combinations of four kinds of features to find the best feature set for different classifiers. Still, the NB classifier achieves the best performance with a 0.776 accuracy and a 0.857 F1 score.

Table 3. Sequential relation classification results. Baseline result is given at the top row. Results of four classifiers with all features in Table 1 are in the middle. Results of four classifiers with the best feature combinations are given at the bottom.

Features	Methods	Accuracy	Precision	Recall	F1
	Baseline	0.719	0.719	1.000	0.837
All features	NB	**0.763**	0.784	**0.924**	**0.848**
	LR	0.690	0.795	0.765	0.779
	MLP	0.683	0.841	0.692	0.756
	SVM	0.523	**0.849**	0.409	0.551
Ratio+Association	NB	**0.776**	0.789	**0.939**	**0.857**
Ratio	LR	0.770	0.800	0.907	0.850
Association	MLP	0.747	**0.808**	0.852	0.829
Ratio	SVM	0.765	0.789	0.919	0.849

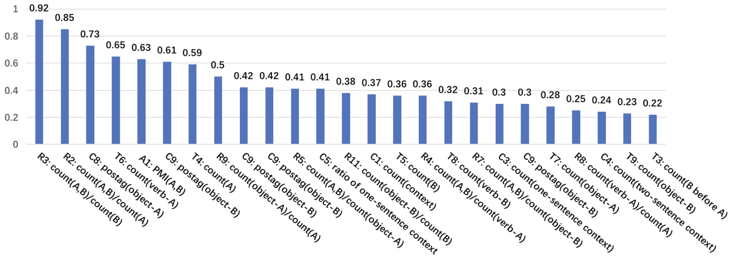

Fig. 5. The 25 most important features for relation classification and their relative importance scores.

Besides, we compute the relative importance scores for all features in Table 1, and the top 25 most important features for relation classification and their relative importance scores are illustrated in Fig. 5. These importance scores are computed by six different methods, including chi-square test, ANOVA F-value, maximal information coefficient, random forest, recursive feature elimination and stability selection. Their individual scores are normalized into the range of 0 and 1, and the average of six scores is computed as the last importance score. We find that ratio-based features are the most important features for sequential relation classification. Experimental results in Table 3 verify this conclusion, with three classifiers achieve the best performance using ratio-based features. And the association-based features are the second most important features.

Table 4. Sequential direction classification results. Baseline result is given at the top row. Results of four classifiers with all features in Table 1 are in the middle. Results of four classifiers with the best feature combinations are given at the bottom.

Features	Methods	Accuracy	Precision	Recall	F1
	Baseline	0.861	0.866	0.993	0.925
All Features	NB	0.803	0.891	0.880	0.885
	LR	0.642	0.894	0.663	0.761
	MLP	0.787	**0.903**	0.844	0.872
	SVM	**0.864**	0.866	**0.997**	**0.927**
Association	NB	0.862	0.863	**0.999**	0.926
Ratio+Association	LR	0.713	0.861	0.796	0.826
All Features	MLP	0.787	**0.903**	0.844	0.872
Association+Context	SVM	**0.870**	0.877	0.988	**0.929**

Table 4 shows the experimental results for sequential direction classification, from which we find that the **Preceding Assumption** is a very strong baseline for direction classification, and achieves a accuracy of 0.861 and F1 of 0.925.

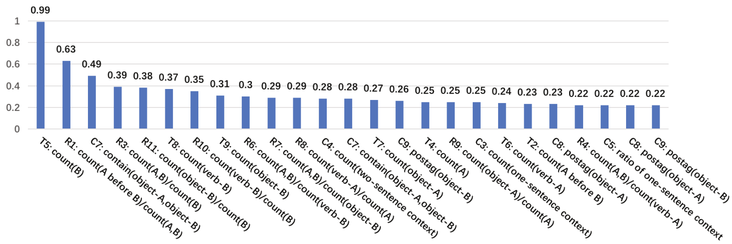

Fig. 6. The 25 most important features for direction classification and their relative effectiveness scores

Four classifiers with all features in Table 1 achieve poor results, and only the SVM achieves higher performance than the baseline method. We explored all combinations of four kinds of features, to find the best feature set for different classifiers. Still, the SVM classifier achieves the best performance with a 0.870 accuracy and a 0.929 F1 score, using the association and context based features.

We also compute the relative feature importance scores for all features in Table 1, and the top 25 most important features for direction classification and their relative importance scores are showed in Fig. 6. We find that the most important two features are T5: count of event B, and R1: count(A before B)/count(A, B). But three of four classifiers achieve their best performance without these two features, which is very interesting. We further experiment with adding these two features, finding that it doesn't help. This is mainly because the feature importance scores examine each feature individually to determine the strength of the relationship between the feature and the response variable. Therefore, they may degrade the performance when combined together due to their opposite correlation with each other.

Based on the experimental results listed above, some useful conclusions can be reached as follows:

- The more features the better performance is not true, and different classifiers capture different kinds of features.
- The two simple baseline methods used in our experiments achieve very good experimental results. However, our proposed feature-based supervised methods achieve the best performance.
- Though certain features are important individually, the performance can be degraded when they are combined, due to their opposite correlation.

6 Case Study

Based on the construction framework proposed, we construct a Chinese travel domain EEG[2] from large-scale unstructured web corpus. A subgraph in this EEG

[2] http://202.118.250.16:60810.

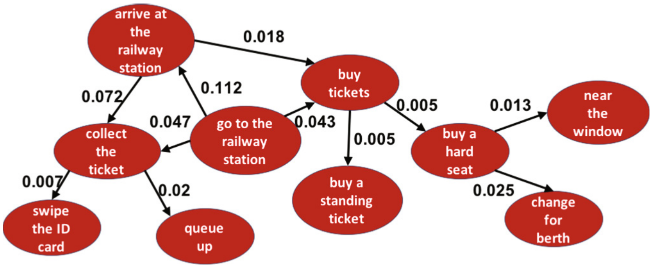

Fig. 7. Subgraph in our constructed travel domain EEG under the scenario of "buy train tickets".

under the scenario of "buy train tickets" is illustrated in Fig. 7. We can find that it covers a lot of important events, such as "go to the railway station", "buy tickets" and "buy a hard seat". They are organized into a directed sequential relation graph, whose edges are labeled with transition probability.

7 Conclusion and Future Work

In this paper, we propose Event Evolutionary Graph (EEG), which reveals evolutionary patterns and development logics between events. We also propose a framework to construct EEG from large-scale unstructured corpus. Extensive experiments are conducted to solve the central problems of sequential relation and direction recognition. Experimental results show that the approaches we proposed are effective for both sequential relation and direction recognition.

To the best of our knowledge, EEG is first proposed by this paper. It is a knowledge base about event evolutionary patterns. Our final goal is to automatically mine this kind of knowledge from open domain large-scale unstructured documents. In future work, we will explore more robust event extraction technique and integrate causality recognition into our construction process. Applying EEG to real world applications is also an interesting research direction.

Acknowledgments. This work was supported by the National Key Basic Research Program of China via grant 2014CB340503 and the National Natural Science Foundation of China (NSFC) via grants 61472107 and 61632011. The authors would like to thank the anonymous reviewers for their insightful comments and suggestions.

References

1. Cassidy, T., McDowell, B., Chambers, N., Bethard, S.: An annotation framework for dense event ordering. Technical report, Carnegie-Mellon University, Pittsburgh, PA (2014)
2. Chambers, N., Cassidy, T., McDowell, B., Bethard, S.: Dense event ordering with a multi-pass architecture. TACL **2**, 273–284 (2014)

3. Chambers, N., Jurafsky, D.: Unsupervised learning of narrative event chains. In: ACL, vol. 94305, pp. 789–797 (2008)
4. Chambers, N., Wang, S., Jurafsky, D.: Classifying temporal relations between events. In: ACL, pp. 173–176. ACL (2007)
5. Che, W., Li, Z., Liu, T.: LTP: a Chinese language technology platform. In: ICCL, pp. 13–16. Association for Computational Linguistics (2010)
6. Do, Q.X., Lu, W., Roth, D.: Joint inference for event timeline construction. In: EMNLP, pp. 677–687. ACL (2012)
7. Granroth-Wilding, M., Clark, S.: What happens next? Event prediction using a compositional neural network model. In: AAAI (2016)
8. Laokulrat, N., Miwa, M., Tsuruoka, Y., Chikayama, T.: UTTime: temporal relation classification using deep syntactic features. In: SemEval-2013, pp. 88–92 (2013)
9. Mani, I., Verhagen, M., Wellner, B., Lee, C.M., Pustejovsky, J.: Machine learning of temporal relations. In: ICCL and ACL, pp. 753–760. ACL (2006)
10. Minksy, M.: A framework for representing knowledge. Psychol. Comput. Vis. **73**, 211–277 (1975)
11. Mirza, P., Tonelli, S.: CATENA: CAusal and TEmporal relation extraction from NAtural language texts. In: ICCL, pp. 64–75 (2016)
12. Pichotta, K., Mooney, R.J.: Statistical script learning with multi-argument events. In: EACL, vol. 14, pp. 220–229 (2014)
13. Pichotta, K., Mooney, R.J.: Statistical script learning with recurrent neural networks. In: EMNLP, p. 11 (2016)
14. Pichotta, K., Mooney, R.J.: Using sentence-level LSTM language models for script inference. In: ACL (2016)
15. Pustejovsky, J., Hanks, P., Sauri, R., See, A., Gaizauskas, R., Setzer, A., Radev, D., Sundheim, B., Day, D., Ferro, L., et al.: The timebank corpus. In: Corpus Linguistics, Lancaster, UK, vol. 2003, p. 40 (2003)
16. Radinsky, K., Davidovich, S., Markovitch, S.: Learning causality for news events prediction. In: WWW, pp. 909–918. ACM (2012)
17. Verhagen, M., Gaizauskas, R., Schilder, F., Hepple, M., Katz, G., Pustejovsky, J.: SemEval-2007 task 15: TempEval temporal relation identification. In: SemEval-2007, pp. 75–80. ACL (2007)
18. Verhagen, M., Sauri, R., Caselli, T., Pustejovsky, J.: SemEval-2010 task 13: TempEval-2. In: SemEval-2010, pp. 57–62. ACL (2010)
19. Zhao, S., Liu, T., Zhao, S., Chen, Y., Nie, J.Y.: Event causality extraction based on connectives analysis. Neurocomputing **173**, 1943–1950 (2016)