AI Open 1 (2020) 22–39

Contents lists available at ScienceDirect

AI Open

journal homepage: www.keaipublishing.com/en/journals/ai-open

# Extracting Events and Their Relations from Texts: A Survey on Recent Research Progress and Challenges

Kang Liu[a,b,*], Yubo Chen[a,b], Jian Liu[c], Xinyu Zuo[a,b], Jun Zhao[a,b]

[a] National Laboratory of Pattern Recognition, Institute of Automation, Chinese Academy of Sciences, Beijing, 100190, China
[b] School of Artificial Intelligence, University of Chinese Academy of Sciences, Beijing, 100049, China
[c] School of Computer and Information Technology, Beijing Jiaotong University, No.3 Shangyuancun Haidian District Beijing, 100044, PR China

ARTICLE INFO

*Keywords:*
Event extraction
Event relation extraction
Knowledge graph

ABSTRACT

Event is a common but non-negligible knowledge type. How to identify events from texts, extract their arguments, even analyze the relations between different events are important for many applications. This paper summaries some constructed event-centric knowledge graphs and the recent typical approaches for event and event relation extraction, besides task description, widely used evaluation datasets, and challenges. Specifically, in the event extraction task, we mainly focus on three recent important research problems: 1) how to learn the textual semantic representations for events in sentence-level event extraction; 2) how to extract relations across sentences or in a document level; 3) how to acquire or augment labeled instances for model training. In event relation extraction, we focus on the extraction approaches for three typical event relation types, including coreference, causal and temporal relations, respectively. Finally, we give out our conclusion and potential research issues in the future.

## 1. Introduction

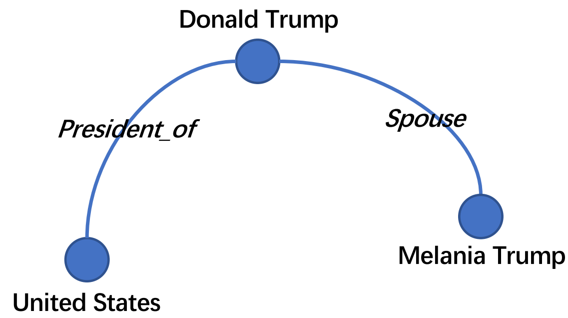
Knowledge Graph (KG), a recent well-known knowledge form, describes and stores the facts in the world with a graph structure. Since knowledge graph could provide the behind and additional semantic information besides the input data, it has shown its power and necessity in many applications, such as textual meaning understanding and logical reasoning. Therefore, starting from the first large-scale industrial knowledge graph, i.e. Google's knowledge graph, it has attracted more and more widespread attention.

Basically, most existing KGs were entity-centric, which described factual knowledge through relational triplets with nodes (entities) and edges (the relations between entities or concepts). As shown in Fig. 1(a), *Donald Trump* and *United States* are two entities, named as a head entity and a tail entity, respectively. *President_of* is their relation. Therefore, such KGs could be regarded as an amount of structured data with linked entities and concepts. However, only storing entities and their relations in KGs is insufficient. There are several other knowledge types in the world, including events, scenarios, frames, rules, etc. Especially for events, it is a common, important, and non-negligible element for analyzing textual meaning in many fields. Therefore, several researchers try to construct event-centric KGs, where they organize knowledge centering on events, like *EventKG* (Gottschalk and Demidova, 2018), *Event Logic Graph* (Ding et al., 1907), *ASER* (Zhang et al., 2020), etc. In an expected event-centric KG, the nodes denote events (including all event arguments) and the edges/links between any two nodes represent their semantical relationships. Fig. 1(b) gives an event-centric KG example. *USA_Attacked_Syrian* and *Trump_Assigned_Executive_Order* are two events with different types, where both of them are denoted as nodes in an event-centric knowledge graph. And there is a *causal relation* between them as an edge in this graph.

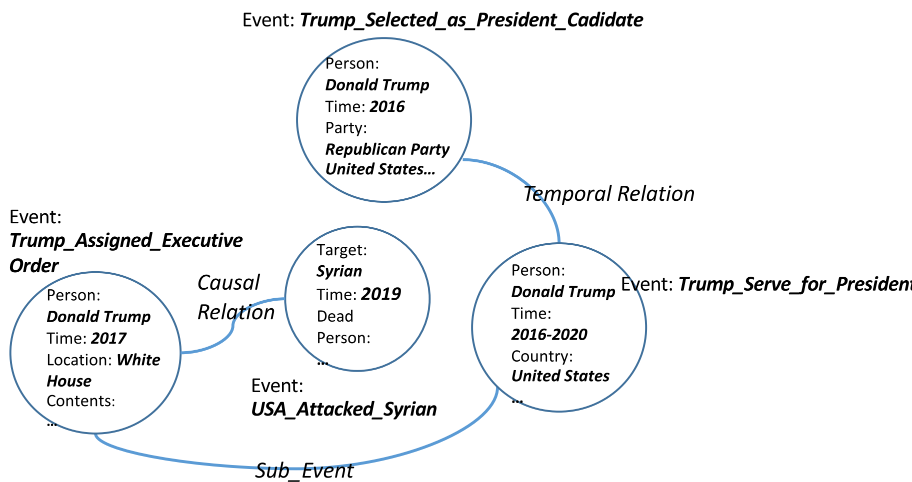
Actually, constructing such event-centric KGs is not an easy task. Plenty of events usually are mentioned in plain texts. Different from the entity textual expression that an entity is expressed by one word or phrase, an event is usually mentioned in the whole sentence or across multiple sentences. Therefore, identifying and extracting events, especially for those structured event information ("Person", "Time", "Location", etc. in *Trump_Assigned_Executive_Order*), needs a more deep understanding on the sentence meanings. Currently, because of the unsatisfactory current extraction methods, the events in most existing event-centric KG are still represented by unstructured phrases/sentences, such as *Event2Mind* (Rashkin et al., 2018) and *ATOMIC* (Sap et al.,

* Corresponding author. National Laboratory of Pattern Recognition, Institute of Automation, Chinese Academy of Sciences, Beijing, 100190, China.
*E-mail addresses:* kliu@nlpr.ia.ac.cn (K. Liu), yubo.chen@nlpr.ia.ac.cn (Y. Chen), jianliu@bjtu.edu.cn (J. Liu), xinyu.zuo@nlpr.ia.ac.cn (X. Zuo), jzhao@nlpr.ia.ac.cn (J. Zhao).

https://doi.org/10.1016/j.aiopen.2021.02.004
Received 23 December 2020; Received in revised form 26 February 2021; Accepted 26 February 2021
Available online 18 March 2021




(a) Entity-centric Knowledge Graph




(b) Event-centric Knowledge Graph

**Fig. 1.** Entity-centric vs. Event-centric Knowledge Graph.

1811a). As a result, much more fine-grained event information is under-investigated. Moreover, how many relation types between events there are is still a controversial problem, and the relationship textual expressions are diverse and implicit. Identifying event relations with different types from texts are challenging and needs deep reasoning.

Therefore, this survey mainly summarizes the recent research progress on event extraction and event relation extraction. The following second section presents the task description of event extraction and event relation identification. Then the third and fourth sections mainly present the main challenges, current evaluation datasets, and corresponding recent research progress for two tasks, respectively. The fifth section introduces the existing event-centric KGs shortly. At last, the conclusion is summarized and the future research issues are discussed.

## 2. Task description of event extraction and event relation extraction

### 2.1. Event definition

Before giving a clear-cut description of the event extraction task, we must first make sure what is an event in the texts. However, there is still so far no common agreement on the event definition, which usually varies in different applications and tasks. In this survey, we refer to the definition in Automatic Content Extraction (ACE) 2005[1] evaluation, i.e., an event is "a specific occurrence involving participants". Actually, an event in ACE 2005 was defined as a structure containing several arguments with different roles. The related terminologies are defined as follows:

- **Event Mention**: An event mention is an entire sentence within which an event is described.
- **Event Trigger**: An event trigger is a word that most clearly expresses an event that happens.
- **Event Argument(s)**: Event arguments are taggable entities that are involved in the event. Most often, there is a specific set of participant roles that can be filled for each type and subtype of event.

For example, the following sentence A number of *demonstrators* **threw** *stones and empty bottles* at *Israeli soldiers* positioned near *a Jewish holy site at the town's entrance*.describes an ATTACK event. The entire sentence could be regarded as an event mention. Moreover, this event is triggered by the word *threw*. So *threw* is an event trigger. Additionally, four event arguments are included: the ATTACK-Attacker (*demonstrators*), the ATTACK-Target (*Israeli soldiers*), the ATTACK-Instrument (*stones and empty bottles*), and the ATTACK-Place (*a Jewish holy site at the town's entrance*). And the structured event information is represented in Table 1.

[1] https://catalog.ldc.upenn.edu/LDC2006T06.

**Table 1**
An example of event structured information in ACE definition.

| Event Trigger | |
|---|---|
| *threw* (*event type*=ATTACK) | |
| Event Arguments | |
| **Roles** | **Values** |
| ATTACK-Attacker | *demonstrators* |
| ATTACK-Target | *Israeli soldiers* |
| ATTACK-Instrument | *stones and empty bottles* |
| ATTACK-Place | *a Jewish holy site at the town's entrance* |

### 2.2. Event extraction task

Accordingly, event extraction is to extract such structured information from texts. In ACE 2005, there are two subtasks for event extraction: 1) **event detection**, which focuses on detecting the events and determining the corresponding event types; and 2) **event argument extraction**, which aims to extract words/phrases/entities playing different roles in a target event. For the aforementioned sentence, the structured extraction results are shown in Table 1.

### 2.3. Event relation extraction task

As aforementioned in the first section, extracting event relations is an important task for event-centric knowledge graph construction, which aims to identify the semantical relationships among different events. For example,

Kimani Gray, a young man who likes football, was $killed_{E1}$ in a police $attack_{E2}$ shortly after a tight match.

An event E1 (trigger by *killed*) is caused by another event E2 (trigger by *attack*). Thus, there is a causal relation between them.

However, the number of event relation types is still an argumentative question. Recently, the most concerning event relations in the research area include three types: *coreferential relation*, *causal relation*, and *temporal relation*. Accordingly, the current event relation task is to judge whether there is a specific relation or not, by giving assigned two events and their contexts. Thus, this task could be naturally regarded as a "0–1" classification problem.

## 3. Recent research progresses in event extraction

In this section, we will introduce the recent typical event extraction methods. Before that, we first present some widely used evaluation datasets and recent investigated research issues.

### 3.1. Datasets for event extraction

A series of evaluations have been proposed to inspire the study of event extraction. We here introduce two event datasets, i.e., ACE 2005 and TAC KBP, that are commonly used in the current event extraction research. Table 2 compares the data statistics of the two datasets.

**Table 2**
Data statistics of ACE 2005 and TAC KBP 2016 (part of LDC2017E02), and TAC KBP 2017 (LDC2017E55). In the TAC KBP datasets, NW and DF are two domains. The newswire (NW) portion was selected from a collection of New York Times and Xinhua articles, while the discussion forum (DF) part was selected from online threads.

| Datasets | | # Doc. | # Sen. | # Trigger | # Argument |
|---|---|---|---|---|---|
| | Train | 529 | 16,473 | 4420 | 7945 |
| ACE 2005 | Dev | 30 | 930 | 505 | 949 |
| | Test | 40 | 714 | 424 | 899 |
| TAC KBP 2016 | NW | 85 | 2601 | 2505 | 5085 |
| | DF | 84 | 2501 | 1650 | 2834 |
| TAC KBP 2017 | NW | 83 | 2541 | 2105 | 5861 |
| | DF | 84 | 2653 | 2270 | 5068 |

*ACE 2005 Event Corpus.* Automatic Content Extraction (ACE) 2005,[2] focuses on developing information extraction (IE) techniques. The ACE 2005 (English) event corpus consists of 599 documents in 6 different domains, including newswire, broadcast, conversation, weblog, usenet, and telephone speech. Moreover, it defines 8 different event types and 33 different event subtypes, which are summarized in Fig. 2 (For each subtype, a finite set of event arguments are defined). The current studies often adopt the data split devised by Ji and Grishman (2008) for evaluation and comparison. The details are shown in Table 2.

*TAC KBP Corpus.* TAC KBP[3] is another evaluation that aims to develop technologies for populating knowledge bases (KBs). In TAC KBP, the event extraction task (from 2015 to 2018) aims to extract event information from unstructured texts, and that the information would be suitable for a structured KB. TAC KBP 2015 defines 9 different event types and 38 event subtypes. And TAC KBP 2016 and 2017 reduce the types to 8 event types and 18 event subtypes, for more efficient dataset creation/evaluation. Fig. 3 shortly summarizes all the event types/subtypes defined in the TAC KBP (2016 and 2017) event ontology.

Notably, TAC KBP event corpus shares many common aspects with ACE 2005 event corpus. A major difference in TAC KBP is that a single event span may be tagged with multiple event types which is referred as a double tagging problem (Mitamura et al., 2017).

### 3.2. Recent focused research problems in event extraction

As the same as other natural language processing tasks, the current event extraction models mainly depended on statistical machine learning, where neural models were widely exploited in recent years. Three explicit and common questions were mostly investigated:

1) *How to learn the semantic representations for events from the given texts?* For event extraction, exploiting effective features are important for statistical learning models. Early approaches designed exquisite features, such as lexical, syntactic and kernel-based features (Ahn, 2006a). Recently, with the development of deep learning, researchers have employed various neural networks, including CNNs (Chen et al., 2015), RNNs (Nguyen et al., 2016) and Transformers Yang et al. (2019) on this task. Moreover, since the event structure is intuitively complex, recent researches work started to employ additional information, like entities, document-level information and syntactic structure, to boost semantic representation learning in the neural models. Consequently, various neural variations and several strategies are employed. We will introduce them in Section 3.3,
2) *How to extract events across sentences or in a document-level?* Most recent approaches focused on this task only on the sentence-level, where the basic assumption is that an event is represented in a sentence. However, in most cases, writers usually use several even non-continued sentences to express single or multiple events. In contrast to sentence-level event extraction, document-level event extraction needs to consider more complex problems: including arguments-scattering, *multi-events expression*, etc. We will address these issues in Section 3.4.
3) *How to train event extraction models based on insufficient labeled instances?* Existing supervised learning based extraction models heavily rely on the training data. However, the human annotation for all event types is very costly. Therefore, researchers proposed several data augmentation approaches for event extraction, including distant supervision-based approaches, acquiring relevant labeled data from different languages, and external knowledge. We will briefly present them in Section 3.5.

[2] https://catalog.ldc.upenn.edu/LDC2006T06.
[3] https://tac.nist.gov/2017/KBP/data.html.

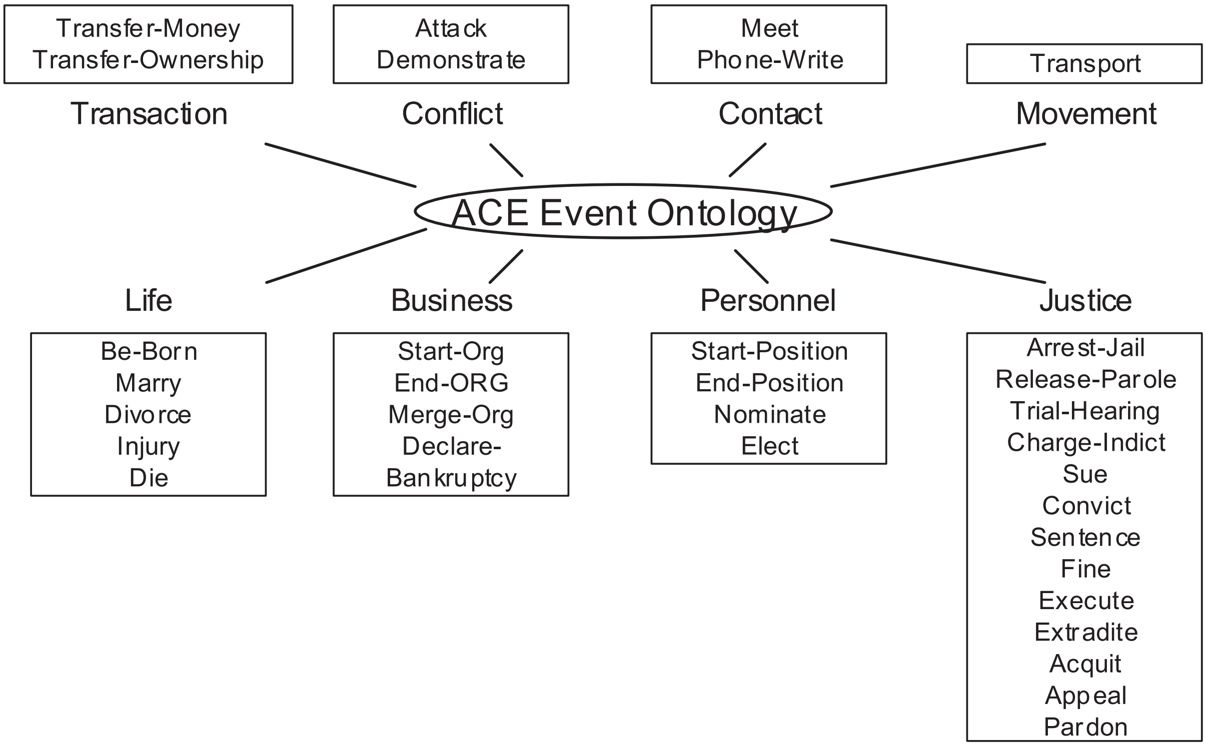



**Fig. 2.** Event types/subtypes defined in the ACE event ontology.

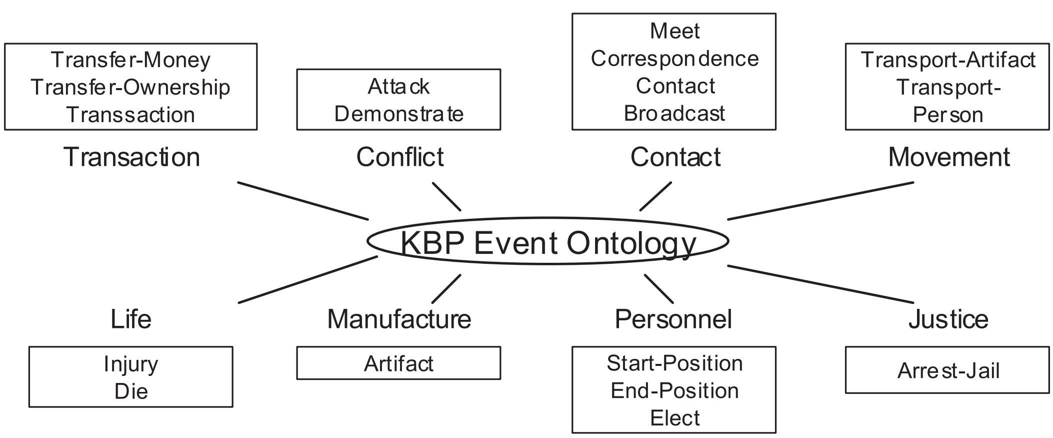



**Fig. 3.** Event types/subtypes defined in the TAC KBP (2016–2017) event ontology.

### *3.3. Neural models for learning semantic representation in events extraction*

Most existing event extraction models are based on the supervised setting which assumes that sufficient labeled examples are available for training. Then the key problem is to learn semantic representations from the texts for events (the first question in subsection 3.2). Besides learning features from raw texts through some typical neural networks (CNNs, RNNs, etc.) (subsection 3.3.1), some additional fine-grained information is employed to improve the representation, such as entity-level features, document-level features, and syntax-level features (subsection 3.3.2, 3.3.3 and 3.3.4). And many recent works have explored using pre-trained language models for feature learning (subsection 3.3.5). Table 3 categories the existing neural models for event extraction, based on different information that they employed.

#### *3.3.1. Neural models exploring raw textual features*

The majority of existing methods adopt CNNs, RNNs to model textual features for event extraction.

*CNNs Based Methods.* Chen et al. (2015) proposed a CNN-based model called Dynamic Multi-pooling Convolutional Neural Network (DMCNN). The main framework contained two tunnels to learn semantical features from raw texts, including lexical-level and sentence-level features. In specific, as shown in Fig. 4, DMCNN introduced a word-representation model to capture semantic regularities for words and adopt a framework based on a CNN to capture sentence-level clues.

Instead of adopting the original CNN structure, Chen et al. (2015) leveraged a dynamic multi-pooling layer to explore the information missed by the CNN. Their model devised the following features: 1) *Context-word features*, the embedding vectors of each word (Pennington et al., 2014). 2) *Position features*, defined as the relative distance of the current word to the predicted trigger or candidate arguments. Each distance value is represented by an embedding vector. 3) *Event-type features*.

**Table 3**
Current different Neural Models using different Features.

| Feature | Neural Models |
|---|---|
| Raw Texts | Convolutional Neural Networks (CNNs) (Chen et al., 2015, 2016; Nguyen et al., 2016)<br>Recurrent Neural Networks (RNNs) (Nguyen and Grishman, 2015)<br>Hybrid Neural Networks (Feng et al., 2016) |
| Raw Texts | Attention Mechanisms (Duan et al., 2017; Zhao et al., 2018; Chen et al., 2018) |
| + Document-level features | Memory Networks (Liu et al., 2018a) |
| Raw Texts | Attention Mechanisms (Liu et al., 1789) |
| + Entity-level features | Adversarial Networks (Hong et al., 2018; Liu et al., 2019a; Hong et al., 2018; Zhang and Ji, 1804)<br>Hybrid Networks (Zhang et al., 2019)<br>Multi-task Learning (Nguyen and Nguyen, 1812; Wadden et al., 2019) |
| Raw Texts | Recurrent Neural Networks (RNNs) (Sha et al., 2018; Orr et al., 2018) |
| + Syntax-level features | Graph Convolutional Networks (GCNs) (Nguyen and Nguyen, 1812; Liu et al., 2018b; Yan et al., 2019; Cui et al., 2020) |
| Pre-Trained Language Models | BERT Architectures (Yang et al., 2019; Liu et al., 2020a, 2020b; Du and Cardie, 2020a)<br>ELMo Architectures (Zhang and Ji, 1804) |

The event type predicted in the trigger classification stage is also considered as an important feature for event argument identification.

Assume $x_i \in R^d$ is the $d$-dimensional vector representation of the $i$-th word in the sentence (the concatenation of context-word feature, position feature, and event-type feature). DMCNN represents a sentence of length $n$ as $x_{1:n} = x_1 \oplus x_2 \oplus \ldots \oplus x_n$, where $\oplus$ refers to vector concatenation computation. Then DMCNN employed a filter $w \in R^{h\times d}$, which was applied to a window of $h$ words, to produce a new feature. For example, a feature $c_i$ was generated from a window of words $x_{i:i+h-1}$ by the following operator:

$$c_i = f(w \cdot x_{i:i+h-1} + b) \tag{1}$$

Then, DMCNN employed multiple filters in a convolution process. Assume there are $m$ filters $w_1, w_2, \ldots w_m$, the convolution operation can be expressed as:

$$c_{ji} = f\left(w_j \cdot x_{i:i+h-1} + b_j\right) \tag{2}$$

where $j$ ranges from 1 to $m$. Next, DMCNN employs “dynamic multi-pooling” to split each feature map into three parts according to the positions of argument and trigger. Here, “dynamic” means the split points are not fixed because the positions of argument and trigger will vary in different sentences. In each part, a max-pooling operation is conducted, which means that there are totally three max-pooling operations in DMCNN. Finally, DMCNN concatenated the automatically learned features to feed them into a final classifier for event trigger prediction and event argument identification.

Nguyen and Grishman (2016) further believed that the original CNNs can only model continuous k-grams other than non-continuous ones that are important for event detection. Therefore, they modified CNNs to model skip-grams to enhance the performance.

*RNNs Based Methods.* Nguyen et al. (2016) proposed a Recurrent Neural Networks (RNNs) to learn richer representations from contexts for predicting event triggers and arguments. The overall approach involved two RNNs in forward and reverse directions. As shown in Fig. 5, The method contained three major steps as follows.

1) *Sentence Embedding.* It first transformed each word$w_i$ into a real-valued vector $x_i$ by concatenating: the word embedding vector of $w_i$; the embedding vector for the entity type of $w_i$; the binary vector corresponding to the dependency relation between words in the parse tree (Li et al., 2013). Such concatenation operation is a simple and effective way to integrate different kinds of features together. The more sophisticated solutions will be discussed in subsection 3.3.2, 3.3.3 and 3.3.4.
2) *RNN Encoding.* It encoded each input vector $x_i$ to a hidden vector $h_i$. Assume the input sequence $X = (x_1, x_2, \cdots, x_n)$, then the model applied RNN over $X$ to generate the hidden vector sequence $(\alpha_1, \alpha_2, \cdots, \alpha_n)$ by $\overrightarrow{RNN}(x_1, x_2, \cdots, x_n) = (\alpha_1, \alpha_2, \cdots, \alpha_n)$. A reverse RNN was used to process $X$ by computing $\overleftarrow{RNN}(x_n, x_{n-1}, \cdots, x_1) = (\alpha_n', \alpha_{n-1}', \cdots, \alpha_1')$. Lastly, the hidden representation for $x_i$ is $h_i = [\alpha_i, \alpha_i']$.
3) *Prediction.* In order to jointly predict triggers and argument roles, the model maintained a binary memory vector for triggers, and binary memory matrices for arguments at each time. These memory components were computed based on the previous matrices to enhance reasoning.

Moreover, Feng et al. (2017) adopt a hybrid neural network combing CNNs with LSTMs. They showed that the architecture was language-independent and could achieve promising results in extraction events in English, Chinese and Spanish. Chen et al. (2016) adopt a similar architecture as that of Nguyen et al. (2016) for event detection and showed improvement over CNNs based architecture. Hong et al. (2018) adopt a different approach based on adversarial training, aiming to exclude spurious features inherent in event detection. They also leveraged RNNs based event detection architecture similar to (Nguyen et al.,

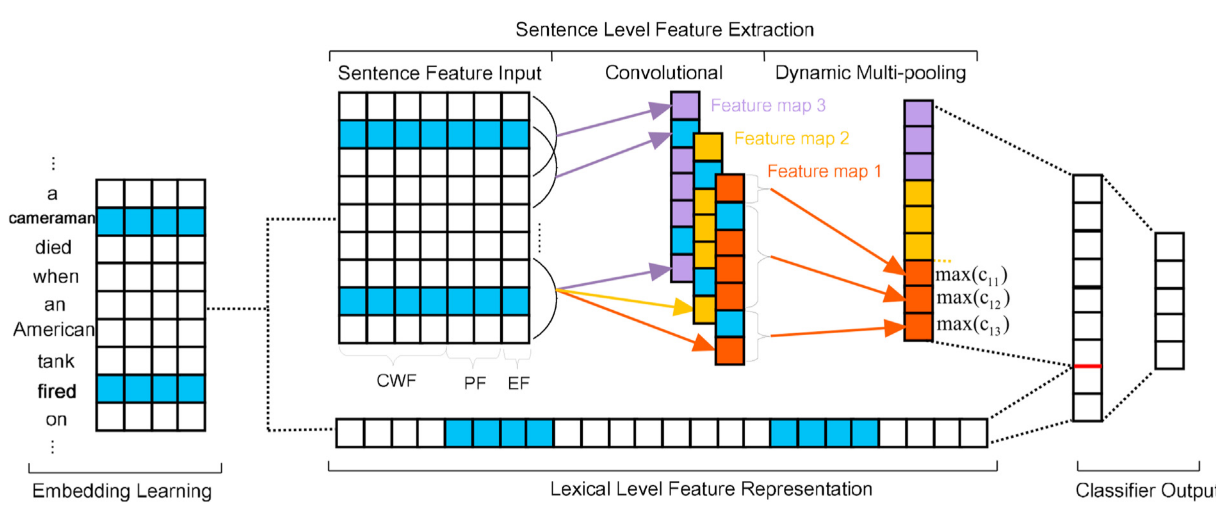



**Fig. 4.** The architecture of the dynamic multi-pooling convolutional neural network (DMCNN) proposed in (Chen et al., 2015).

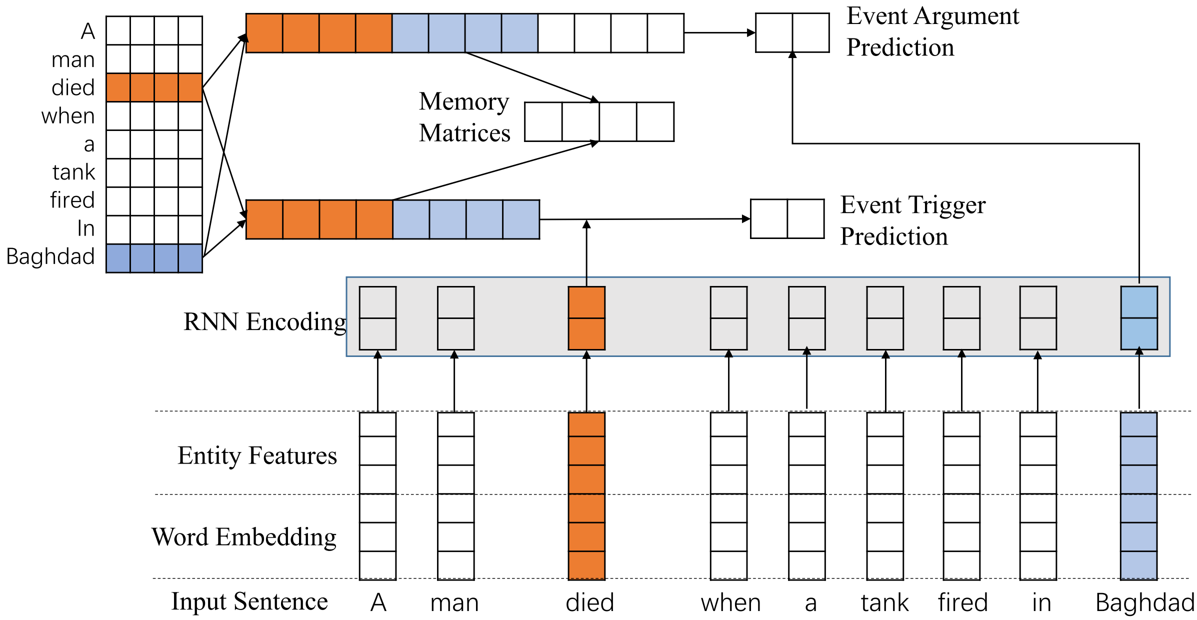



**Fig. 5.** The model architecture presented in (Nguyen et al., 2016).

2016; Chen et al., 2016).

Besides, more advanced neural architectures, the variations of CNNs and RNNs, are proposed to improve the sentence representations. Their models mainly focus on incorporating fine-grained information into the networks for event extraction. We will introduce them in the following three subsections.

#### 3.3.2. *Exploring entity-level information*

In some cases, entity information plays a crucial role in event extraction. For example, *Danny Baker* fired by *BBC*.

If we have known that *Danny Baker* is a person entity and *BBC* is an organization entity, this information could help to identify that "fired" evokes an event of *End-Of-Position* rather than *Attack* type. Traditional approaches usually designed fine-grained rules to exploit entity information to enhance event extraction (Liao and Grishman, 2010; Hong et al., 2011). In neural models, researchers usually exploited three common means including attention mechanisms, adversarial training and multi-task learning to capture entity-level information for event extraction.

Liu et al. (1789) proposed to build a supervised attention mechanism to force the model to focus more on entities than other parts for identifying event triggers. And the results showed that supervised attention leads to improved performance.

Liu et al. (2019a), Zhang and Ji (1804) investigated adversarial training to exploit entity-level information. The difference is that Liu et al. (2019a) adopt knowledge distillation for feature learning, using entity-information as a supervision signal to enhance learning, while Zhang and Ji (1804) investigated a reverse imitation learning, in deep reinforcement learning, to jointly identify entities and events.

Zhang et al. (2019), Nguyen and Nguyen (1812), Wadden et al. (2019) investigated multi-task learning to exploit entity information for event extraction. Zhang et al. (2019) adopt entity enhanced language modeling with extraction. They showed that the two tasks are related to each other, and jointly learning them can bring improvements. Nguyen and Nguyen (1812) used one network to jointly learning entity recognition and event extraction, and showed improvement over the single-task learning approach. Wadden et al. (2019) proposed a joint model including entity recognition, relation extraction and event extraction. Their experimental results showed these three tasks could improve each other.

#### 3.3.3. *Exploring syntax-level information*

More advanced neural approaches have exploited syntax-level information for event extraction and have demonstrated encouraging results. An important problem is how to encode syntactic structure into the event-related distributional representations. According to recent methods, RNNs and Graph Convolutional Networks (GCNs) are two commonly used architectures.

Liu et al. (2018b) proposed a Jointly Multiple Events Extraction (JMEE) framework, which could jointly extract multiple event triggers and arguments. The methods introduced syntactic shortcut arcs to enhance information flow, and an attention-based graph convolution network was devised to enable information integration. Liu et al. (2019b) further extended the method to model cross-lingual event detection, and they found that syntax information had the property of multi-lingual invariant, which could be used to effective multilingual transfer learning.

Orr et al. (2018), Sha et al. (2018) devised a similar approach as that of (Liu et al., 2018b), but they modeled the syntax information via BiLSTM networks instead of graph convolution networks. Compared with graph network-based approaches, BiLSTM networks have a relatively fast inference process.

#### 3.3.4. *Exploring document-level information*

Sentence-level features may not always capture enough information for event extraction. For example, it is difficult to identify that "leave" in "I knew it was time to leave" triggers an *End-of-Position* event when not given long contexts. A lot of existing works for event extraction also incorporated document-level information that are based on human-designed features (Ji and Grishman, 2008).

To incorporate document level information, Duan et al. (2017) combined document embeddings with words embeddings. Therefore, the representation of a word is:

$$\boldsymbol{h}_i = \boldsymbol{d} \oplus \boldsymbol{x}_i \tag{3}$$

where $\boldsymbol{d}$ is the global document vector learned by document modeling methods (Le and Mikolov, 1405), and it is shared by each word; $\boldsymbol{x}_i$ is the word embedding of $w_i$. Then Duan et al. (2017) leveraged a model based on RNNs based architecture similar to (Nguyen et al., 2016) compute feature vectors and conducted the final event type prediction.

To obtain better integration performance, hierarchical and multi-hop architectures are commonly adopted strategies. In detail, Chen et al. (2018), Zhao et al. (2018) adopt hierarchical structures to capture document-level information for event extraction. The difference is that Zhao et al. (2018) also used supervised attention to force the model to focus on entity information, while Chen et al. (2018) used unsupervised attention to learn features. Liu et al. (2018a) used a memory network to encode the document-level formation, where the process of event detection is achieved vie multi-step inference. The results show that the multi-hop mechanism of the memory network can enhance learning.

Moreover, Liu et al. (2016a) devised a probabilistic soft logic (PSL) model, which considered event-event association and topic-event association into the reasoning process. The model can jointly consider local predictions from multiple classifiers to generate final EE results.

Nevertheless, we note the improvement of integrating document-level information is modest, and it still remains an open challenge how to effectively incorporate document-level information for event extraction.

#### 3.3.5. *Exploring pre-trained language models*

Recently, more and more studies have investigated using pre-trained language models (e.g., BERT (Devlin et al., 2019) and EMLo (Peters et al., 2018)) for event extraction. Owing to that pre-trained language models can learn universal language representations using a large amount of unlabeled data, using them for feature learning often leads to considerable improvements over methods using traditional neural networks for feature learning.

Yang et al. (2019) directly applied BERT representations for the event extraction task, and their model has achieved state-of-the-art performance without designing task-specific architectures or using external resources. Wang et al. (2019a) combined BERT representation and adversarial learning to explore weakly-supervised data for event trigger extraction. Their method can cooperate to obtain more diverse and accurate training data to enhance learning. Wang et al. (2019b) designed a role-oriented modular network for specifically event argument extraction task, and used both CNNs and BERT to learn features. Liu et al. (2020a) used BERT architectures to explore context-specific information for learning, which is shown to greatly enhance the generalization and robustness of the event extraction model. Recently, Du and Cardie (2020a), Liu et al. (2020b) formulated event extraction as a machine reading comprehension (MRC) task, where the extraction of event information is seen as a question answering process.

#### 3.3.6. *Comparison of current neural methods for sentence-level event extraction*

Table 4 compares different event extraction methods introduced above on the ACE 2005 dataset. We use a traditional feature-based method designed by Li et al. (2013) as a baseline model. From the results, we could observe that all of the deep learning-based methods yield better performance than traditional feature-based methods, owing to their ability in automatic feature representation.

We further observe that the methods exploring document-level (e.g. (Zhao et al., 2018),)) and entity-level (Hong et al., 2018) features generally outperform methods exploring raw texts only. Moreover, adopting syntax-level information can further boost performance. It implies that syntax information is closely related to the event extraction task, and fine-grained information including entities and document contexts can generally benefit the event extraction task. Lastly, we note the models achieving the best performance, i.e., the systems designed by Yang et al. (Yang et al., 2019), Liu et al. (Liu et al., 2020b), adopt only textual features which are however learned by pre-trained language models. This suggests that learning universal language representations from unlabeled sentences can usually benefit feature representation in event extraction. Nevertheless, to date, there is less work discussing combing pre-trained language models with entity-level or document-level information.

**Table 4**
Comparison of different methods for event extraction on ACE 2005 dataset. [†] indicates adopting textual feature; [‡] indicates adopting entity-level feature; * indicates adopting document-level feature; [††] indicates adopting syntax-level feature; [§] indicates methods adopting pre-trained language models; - indicates that the performance is not given.

| Model | Trigger **Identification** | | | Trigger **Classification** | | | Argument **Identification** | | | Argument **Classification** | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | P | R | F | P | R | F | P | R | F | P | R | F |
| MaxEnt (Li et al., 2013) | 76.2 | 60.5 | 67.4 | 74.5 | 59.1 | 65.9 | 74.1 | 37.4 | 49.7 | 65.4 | 33.1 | 43.9 |
| Li et al. (Li et al., 2013) | 76.9 | 65.0 | 70.4 | 73.7 | 62.3 | 67.5 | 69.8 | 47.9 | 56.8 | 64.7 | 44.4 | 52.7 |
| Chen et al. (Chen et al., 2015)[†] | 80.4 | 67.7 | 73.5 | 75.6 | 63.6 | 69.1 | 68.8 | 51.9 | 59.1 | 62.2 | 46.9 | 53.5 |
| Nguyen et al. (Nguyen and Grishman, 2015)[†] | 68.5 | 75.7 | 71.9 | 66.0 | 73.0 | 69.3 | 61.4 | 64.2 | 62.8 | 54.2 | 56.7 | 55.4 |
| Nguyen et al. (Nguyen et al., 2016)[†] | – | – | – | – | – | 71.3 | – | – | – | – | – | – |
| Chen et al. (Chen et al., 2016)[†] | – | – | 72.2 | – | – | 68.9 | – | – | 60.0 | – | – | 54.1 |
| Feng et al. (Feng et al., 2017)[†] | 80.8 | 71.5 | 75.9 | 84.6 | 64.9 | 73.4 | – | – | – | – | – | – |
| Liu et al. (Liu et al., 2017)* | – | – | 72.3 | – | – | 69.6 | – | – | – | – | – | – |
| Liu et al. (Liu et al., 1789)[‡] | – | – | – | 81.4 | 66.9 | 73.4 | – | – | – | – | – | – |
| Duan et al. (Duan et al., 2017)* | – | – | – | 77.2 | 64.9 | 70.5 | – | – | – | – | – | – |
| Sha et al. (Sha et al., 2018)[††] | – | – | – | 74.1 | 69.8 | 71.9 | 71.3 | 64.5 | 67.7 | 66.2 | 52.8 | 58.7 |
| Zhao et al. (Zhao et al., 2018)* | – | – | – | 72.3 | 75.8 | 74.0 | – | – | – | – | – | – |
| Chen et al. (Chen et al., 2018)* | – | – | – | 77.9 | 69.1 | 73.3 | – | – | – | – | – | – |
| Liu et al. (Liu et al., 2018b)[††] | 80.2 | 72.1 | 75.9 | 76.3 | 71.3 | 73.7 | 71.4 | 65.6 | 68.4 | 66.8 | 54.9 | 60.3 |
| Hong et al. (Hong et al., 2018)[†] | 75.3 | 78.8 | 77.0 | 71.3 | 74.7 | 73.0 | – | – | – | – | – | – |
| Nguyen et al. (Nguyen and Nguyen, 1812)[‡] | 70.5 | 74.5 | 72.5 | 68.0 | 71.8 | 69.8 | 59.9 | 59.8 | 59.9 | 52.1 | 52.1 | 52.1 |
| Zhang and Ji (Zhang and Ji, 1804)[‡] | 76.4 | 68.2 | 72.1 | 74.2 | 65.3 | 69.5 | 66.2 | 51.4 | 57.8 | 65.6 | 48.7 | 55.9 |
| Zhang et al. (Zhang et al., 2019)[‡] | – | – | 72.9 | – | – | 71.6 | – | – | – | – | – | – |
| Liu et al. (Liu et al., 2019a)[‡] | – | – | – | 76.8 | 72.9 | 74.8 | – | – | – | – | – | – |
| Yan et al. (Yan et al., 2019)[‡] | – | – | – | 79.5 | 72.3 | 75.7 | – | – | – | – | – | – |
| Cui et al. (Cui et al., 2020)[††] | – | – | – | 76.7 | 78.6 | 77.6 | – | – | – | – | – | – |
| Wadden et al. (Wadden et al., 2019)[††] | – | – | 76.5 | – | – | 73.6 | – | – | 55.4 | – | – | 52.5 |
| Wang et al. (Wang et al., 2019b)[§] | – | – | – | – | – | – | – | – | – | 62.2 | 56.6 | 59.3 |
| Yang et al. (Yang et al., 2019)[§] | 84.8 | 83.7 | 84.2 | 81.0 | 80.4 | 80.7 | 71.4 | 60.1 | 65.3 | 62.3 | 54.2 | 58.0 |
| Liu et al. (Liu et al., 2020a)[§] | – | – | – | 75.2 | 74.4 | 74.8 | – | – | – | – | – | – |
| Du et al. (Du and Cardie, 2020)[§] | 74.3 | 77.4 | 75.8 | 71.1 | 73.7 | 72.4 | 58.9 | 52.1 | 55.3 | 56.8 | 50.2 | 53.3 |
| Liu et al. (Liu et al., 2020b)[§] | – | – | – | 75.6 | 74.2 | 74.9 | – | – | – | 63.0 | 64.2 | 63.6 |

### 3.4. Methods for document-level event extraction

The aforementioned approaches extract events in a single sentence. However, writers usually use non-continued sentences to express an event or several events. Thus, the system needs to extract events in a wider scope, i.e. document-level event extraction (the second question in subsection 3.2). As shown in Fig. 6, an *Equity Freeze* event (*Event*1) with five arguments are described in two sentences $S_4$ and $S_5$. Moreover, *Equity Freeze* event (*Event*2) are described in $S_4$ and $S_6$ which have two shared arguments with *Event*1. Thus, document-level event extraction aims to extract and split such two events from the given document.

In contrast to sentence-level event extraction, document-level event extraction has more challenges as follows. 1) *Arguments-scattering*: arguments of one event may scatter across multiple sentences in one document. Thus in contrast to extracting events in one sentence, document-level event extraction needs to capture long-distance (even across sentences) dependencies among event information. 2) *Multi-events*: there are multiple events mixed in a document, that require document-level event extraction systems to discriminate them and split the corresponding arguments for different events. The intuitive illustration of these challenges are shown in Fig. 6, the two events share the same entities *Mr. LianWeife* and *Shenzhen Intermediate People's Court* as event role *Equity Holder* and *Legal Institution*, respectively, and these arguments are distributed in different sentences ($S_4 - S_6$) separately.

However, there are few works in the literature that have gone beyond individual sentences to make the extraction. These works could be roughly divided into two categories as follows. (1) *The methods of document-level event role filler extraction*. These works focused on identifying event-specific role fillers, i.e., arguments, in an article and were mainly evaluated on the MUC-4 dataset. (2) *The methods of document-level pre-defined event extraction*. These works focused on identity events of pre-defined types with their event-specific arguments from the given whole document in a specific domain, such as the financial domain. We will introduce them specifically in the following subsections.

#### 3.4.1. Methods for document-level event role filler extraction

The methods of document-level event role filler extraction mainly followed the task setting in MUC evaluation (Sundheim, 1992). The full event extraction task of MUC involved three parts: (1) *role filler extraction*, which aimed to extract arguments of specific event types from the given whole document. In this subtask, the systems needed to find all argument mentions in a document; (2) *role filler mention coreference resolution*, which aimed to find all mentions for one argument in a specific event; (3) *event tracking*, which aimed to group the argument mentions into a specific event.

Though the task definition in MUC covered the full process of the document-level event extraction, it assumed that there was only one event in a given document. Therefore, existing methods usually regarded this task as a task of the slot filling or document-level relation extraction. The corresponding problem is how to discriminate the extracted role filler mentions in a document (*arguments-scattering*).

To solve the *arguments-scattering* problem, existing approaches usually adopt a clue-selection strategy that selects event-related clues in the document and filter the unrelated argument mentions. Patwardhan and Riloff (2009) proposed GLACIER to model all event-related clues in a probabilistic framework. Huang and Riloff (2011) proposed TIER to extract role filler. In TIER, they first used a classifier to determine the genre of the whole document, then they identified event-related sentences based on the genre information and extracted the role fillers in them. To filter more noises from multiple event instances, Huang and Riloff (2012) proposed a model with a bottom-up framework. First, they identified role filler candidates. Then, they used a cohesion classifier to remove the candidates in the spurious sentences. Du and Cardie (2005) proposed an end-to-end neural sequence model. In their model, they used a multi-granularity reader to dynamically incorporate paragraph-level and sentence-level contextualized representations. Evaluations on the MUC-4 dataset proved that their model could achieve substantial improvement over prior works.

#### 3.4.2. Methods of document-level pre-defined event extraction

The methods of document-level pre-defined event extraction so far mainly focused on identifying events with pre-defined types and extracting the corresponding arguments from a given document. Existing methods mainly conducted this task in the financial domain. Unlike the proposed works on the MUC-4 dataset, these works usually assumed that there are multiple events with different types in a document. Moreover, an event is possibly expressed several times in multiple and even non-continued sentences.

Yang et al. (2018) proposed a document-level Chinese financial event extraction framework, shortly named DCFEE. DCFEE first used a weak-supervised sentence-level event extractor to label event arguments and triggers in each sentence. Note that, DCFEE employs distant supervision to generate sentence-level training data (event trigger and event arguments) for different event types. Then, a key event detection model was proposed to discover the major sentence that presents most of the arguments for an individual event in the document. And an arguments-completion strategy was used to automatically pad other missing arguments from the surrounding sentences for a target

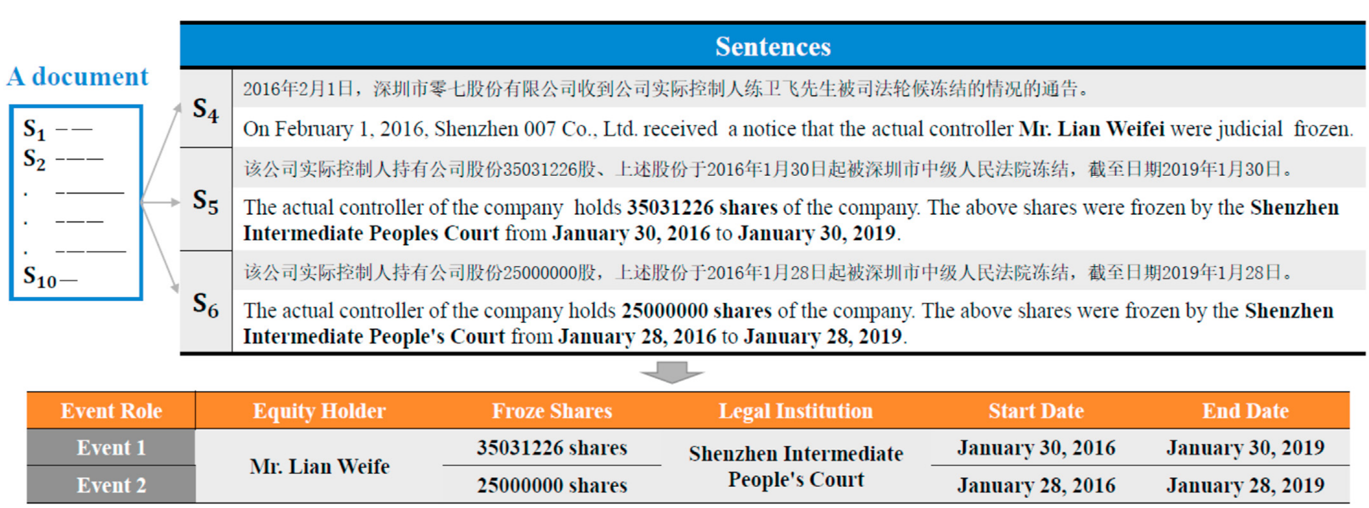



**Fig. 6.** A sample of document-level event extraction with two *Equity Freeze* events whose arguments scatter across multiple sentences. In this document, only three sentences ($S_4$ - $S_6$) are shown, Event 1 and Event 2 are targets to be extracted and words in bold-faced are event arguments with specific roles.

pre-defined event. Their basic assumption is that all arguments are not far away and are mentioned surrounding the major sentence. In this way, the *arguments-scattering* problem could be naturally handled. At the same time, *multi-events* could be discriminated through discovering the major sentence.

For gathering arguments for different events, Zheng et al. (1904) proposed an end-to-end neural model named Doc2EDAG to capture the global information of the same entity and predicted multiple events in the document. The key behind the idea is to transform the structured event table into an entity-based directed acyclic graph (EDAG). In this way, the hard table-filling task could be transformed into several sequential path-expanding and more tractable sub-tasks. Then, they encoded candidate arguments with global document-level contexts and used memory for path expanding.

For *multi-events* problem, Chen et al. (2020) believe that the target event arguments tend to be mentioned in several groups of adjacent sentences. They define those adjacent relevant sentences as different event regions. They build a graph to directly model the multiple regions in a document, where the nodes are candidate event arguments and the edges between two nodes reflect their positional relation (in adjacent sentences or within the same sentence) or the coreference relation. Then a graph attention network (EE-GAT) was used to identify event regions in a document and aggregate event information. In this way, redundant or irrelevant extractions in the sentence-level could be effectively avoided.

#### 3.4.3. Comparison of document-level event extraction methods

Table 5 and Table 6 compares different types of methods introduced above on MUC-4 and CFEDD (Yang et al., 2018) datasets. From the results, we have the following observations. (1) The methods that capture interaction and filter noises from multiple events (TIER, CohesionExt, EE-GAT, EDAG) outperform the methods only considering one event (GLACIER, DCFEE), which proves that considering the *multi-event* problem in one document is important for document-level event extraction. (2) Document-level information, such as document structure (TIER, DCFEE), entity coreference (EE-GAT, EDAG), entity co-occurrence (CohesionExt), are important for solving the *arguments-scattering* problem. (3) The deep learning methods (MGReader, EE-GAT, DCFEE, EDAG) gain higher performance than the feature-based methods (GLACIER, TIER, CohesionExt), which proves the effectiveness of the deep learning methods on capturing semantics and structures of the given document.

### 3.5. Data augmentation for event extraction

Most current event extraction models were often based on a supervised learning setting. However, hand-labeled training data is expensive to produce, in low coverage of event types and limited in size, which makes the supervised methods hard to extract large-scale events. There are some works that try to solve the data labeling problem (the third question in subsection 3.2).

Totally, these works can be roughly divided into three types: (1) *Distant Supervised Data Generation*, which aims to automatically generate large scale labeled data by distant supervision; (2) *Cross-lingual Data Augmentation*, which aims to use cross-lingual information to augment data; (3) *Data Augmentation with External Knowledge*, which aims use external knowledge to alleviate the data sparsity problem. We will introduce works in these three areas specified in the following subsections.

**Table 5**
Experimental Results of the Current Approaches on the MUC-4. Here, GLACIER, TIER, CohesionExt, MGReader and EE-GAT are all introduced in subsection 3.4.1. P, R and F mean Precision, Recall and F1-value, respectively.

| | P | R | F |
|---|---|---|---|
| GLACIER (Patwardhan and Riloff, 2009) | 47.8 | 57.2 | 52.8 |
| TIER (Huang and Riloff, 2011) | 50.8 | 61.4 | 55.6 |
| CohesionExt (Huang and Riloff, 2011) | 57.8 | 59.4 | 58.6 |
| MGReader (Du and Cardie, 2005) | 56.4 | 62.8 | 59.4 |
| EE-GAT (Chen et al., 2020) | 63.0 | 66.0 | 65.0 |

#### 3.5.1. Distant supervised data generation

Works on distant supervised data generation aim to generate large-scale labeled data by distant supervision automatically. Chen et al. (2017) firstly proposed to automatically label data for large-scale event extraction via world knowledge and linguistic knowledge. Compared with relation extraction, the main problem of distant supervised data generation for event extraction is that an event may be expressed by several sentences and it is hard to find all arguments of one event instances in one sentence. Thus, they first proposed an *tf-idf* based approach to figuring out key arguments of an event in Freebase. Then the events in the sentence and corresponding trigger words could be extracted if the sentence includes those key arguments. They further employed FrameNet to filter noisy triggers and expand more useful trigger words. Finally, they used a soft distant supervision to label events in sentences, which assumed that any sentence containing all key arguments in Freebase and a corresponding trigger word is likely to express that event in some ways. Based on such simple rules, many labeled training examples are generated.

Zeng et al. (1712) also developed a training data generation approach for event extraction. Their approach can scale up event extraction training instances from thousands to hundreds of thousands with a much lower cost than a manual approach. Similar to (Chen et al., 2017), they first extracted the key arguments for an event from existing structured knowledge bases. Then, they used the key arguments to automatically infer the occurrence of an event without explicit trigger identification.

Moreover, to avoid adapting sophisticated pre-defined rules and heavy toolkits in the data generation process, Wang et al. (2019a) proposed a simple trigger-based latent instance discovery strategy. Their assumption is that if a given word serves as the trigger in a known event instance, all instances mentioning this word may also express that event. Therefore, they build a large event-related candidate set with good coverage. And an adversarial training mechanism is employed to iteratively identify informative instances from the candidate set and filter out those noisy ones.

As mentioned in subsection 3.4 (Yang et al., 2018; Zheng et al., 1904), also used the strategies of distant supervision to generate large-scale labeled data for event extraction in the financial domain. Yang et al. (2018) used a sentence-level data generated model to label the event trigger and event arguments in the sentences. And a document-level data generated model is proposed to label the event mentions in a whole document. Zheng et al. (1904) used a similar strategy, where they did not label the trigger words for each event.

#### 3.5.2. Cross-lingual data augmentation

In many cases, exploiting large-scale labeled data from different languages could improve the event extraction performance on the target language. Thus, cross-lingual event extraction aims to employ this benefit by transferring knowledge across different languages. These work can be roughly divided into two types: *Translation-based methods* and *Corpora-based methods*.

*Translation-based methods:* These methods used the translation model to generate more training data from other languages. Wei et al. (2017) proposed radical features from automatic translation for event extraction. They derived meaningful subword features from automatic translations into the target language. Experimental results showed that their method was particularly useful when using languages with writing systems that facilitate easy decomposition into subword features. Liu et al. (2018c) proposed a framework to fulfill the cross-lingual event extraction by using a gated neural network and machine translation engine. In specific, they exploited the consistent information in multilingual data via context attention mechanism to alleviate the data scarcity problem. They also proposed gated cross-lingual attention to exploiting the

**Table 6**
Experimental Results of the Current Approaches on the CFEDD. Here, DCFEE and EDAG are all introduced in subsection 3.4.2. EF, EP, ER, EO are different event types in the financial domain (EF: Equity Freeze; EP: Equity Pledge; ER: Equity Repurchase; EO: Equity Overweight). P, R and F mean Precision, Recall and F1-value, respectively.

| Model | EF | | | ER | | | EU | | | EP | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | P | R | F | P | R | F | P | R | F | P | R | F |
| DCFEE (Yang et al., 2018) | 66.0 | 41.6 | 51.1 | 84.5 | 81.8 | 83.1 | 62.7 | 35.4 | 45.3 | 64.3 | 63.6 | 63.9 |
| EDAG (Zheng et al., 1904) | 77.1 | 64.5 | 70.2 | 91.3 | 83.6 | 87.3 | 80.2 | 65.0 | 71.8 | 80.0 | 74.8 | 77.3 |

complement information from multilingual data with different confidence.

*Corpora-based methods:* These methods used cross-lingual parallel corpora instead of the translations as the supervised information. Ji (2009) proposed to use cross-lingual predicates to improve the performance of the event extraction. Hsi et al. (2016) proposed an event extraction approach that trained on a combination of both language-dependent and language-independent features. Liu et al. (2019c) proposed a cross-lingual event extraction method, demonstrating a minimal dependency on parallel resources. Specifically, to construct a lexical mapping between different languages, they devised a context-dependent translation method. In details, they proposed a shared syntactic order event detector to deal with the word order difference problem in multilingual co-training.

#### 3.5.3. Data augmentation with external knowledge

To alleviate the data sparsity problem, some works try to introduce external knowledge bases with different structures, such as FrameNet [4] and Abstract Meaning Representation (AMR) .[5]

Liu et al. (2016b) proposed to use FrameNet to alleviate the data sparsity problem in event extraction. In FrameNet, verbs are annotated into different frames which they could trigger. These frames are naturally related to some events in ACE. Then they proposed a global inference approach to detect frame examples in FrameNet which could indicate specific events. Further, they combined the detected examples from FrameNet with labeled instances in ACE as the training set.

Huang et al. (2016) proposed to use AMR to alleviate the data sparsity problem in event extraction. They incorporated AMR symbolic and distributional semantics to detect and represent event structures. Then a joint typing framework was proposed to simultaneously identify event types, argument roles and discover an event schema.

## 4. Recent research progresses in event relation extraction

This section will introduce the recent typical methods for event relation extraction which is less investigated compared with event extraction. So far, existing methods usually focused on three relation types, including event coreference, causal and temporal relations. For each relation type, the following subsections will introduce the existing extraction approaches, besides task description and used evaluation datasets.

### 4.1. Event coreference resolution

Event coreference resolution (ECR) is to determine whether the identified events refer to the same real-world event, where those events may occur within a sentence/document and across multiple sentence/documents. As shown in Fig. 7, the *violence* (E1) and *disperse* (E2) refer to the same incident that violently dispersed the march.

#### 4.1.1. Datasets

The current widely used evaluation datasets are as follows. (1) **MUC**: MUC produces the earliest corpora for supporting the event coreference task (Sixth Message Understandi, 1995; Seventh Message Understan, 1998), where event coreference is performed as part of the scenario template filling task. (2) **ACE 2005**: ACE 2005 is the most widely used version of the ACE corpora for within-document event coreference evaluations, includes both English and Chinese documents. (3) **OntoNotes**: OntoNotes is a large-scale corpus which provides both within and cross-document entity and event coreference annotations (Pradhan et al., 2007). (4) **EventCorefBank**: EventCorefBank (ECB) corpus (Lee et al., 2012) and its extended version ECB + corpus (Cybulska and Vossen, 2014) annotate within and cross-document event coreferential relations. (5) **TAC KBP**: TAC Knowledge Base Population (KBP) includes within-document multilingual event coreference-annotated corpora started from 2015 to 2017, which is annotated with the RichERE annotation style (Song et al., 2015). The statistics of commonly used datasets for event coreference resolution are listed in Table 7.

#### 4.1.2. Methods for event coreference resolution

Existing methods usually regarded ECR as a classification or ranking problem. Thus, classical machine learning models are widely employed, such as the decision trees classifier (Cybulska and Vossen, 2015), the ensemble one-nearest-neighbor classifier (Lu and Ng, 2016), the maximum entropy pairwise classifier (Ahn, 2006b), the information propagation model (Liu et al., 2014), the hierarchical distance-dependent bayesian model (Yang et al., 2015), the latent antecedent method (Liu et al., 2016), the multi-loss neural model (Zuo et al., Zhao) and ranking model (Lu and Ng, 2017a). There models mainly focused on understanding the contexts around two events. To fulfill this aim, existing methods have several intuitive perspectives, such as event-related features, syntactic features, event topic information, linguistic features, etc. (Bejan and Harabagiu, 2010). However, identifying event coreferential relation not only relies on the contextual features. Existing models have considered more complex clues, including document-level or topical structures (Choubey and Huang, 2018) and event argument information (Chen et al., 2009; Chen and Ji, 2009; Choubey and Huang, 2017; Huang et al., 2019), even other related tasks (Chen and Ng, 2016; Araki and Mitamura, 2015; Barhom et al., 2019; Lu et al., 2016; Lu and Ng, 2017b) etc.

*Considering document-level or topical structure for ECR.* Choubey and Huang (2018) proposed a holistic approach for this task by considering their correlations with document topic structures. The key observation is that events make the backbone of a document and coreferent mentions of the same event play a key role in achieving a coherent content structure. This method modeled several aspects of correlations between event coreference chains and document level topic structures, including *correlations between main event chains and topic transition sentences*, *correlations across semantically associated event chains*, etc. Then an integer linear programming (ILP) joint inference framework was employed to combine all aspects of correlations.

*Considering event argument information for ECR.* The straightforward assumption is that if two events have incompatible arguments in any of the argument roles, they cannot be coreferent. Therefore, some methods utilized the event argument information (e.g., argument similarities, argument compatibility, and others) of the arguments participating to understand the coreference among different events (Chen et al., 2009; Chen and Ji, 2009; Choubey and Huang, 2017). In particular, Huang et al. (2019) employed an interactive inference network to iteratively learn argument compatibility and event coreference resolution.

[4] https://framenet.icsi.berkeley.edu/fndrupal/.
[5] https://github.com/amrisi/amr-guidelines/blob/master/amr.md.

| | |
|---|---|
| **Event Corefernece.** | The latest *violence*$_{E1}$ against thousands of Morsi supporters when Egyptian military forcefully *disperse*$_{E2}$ the demonstrators staging *protests*$_{E3}$ in square with tanks. |
| **Event Causal Relation.** | Kimani Gray, a young man who likes football, was *killed*$_{E4}$ in a police *attack*$_{E5}$ shortly after a tight match. |
| **Event Tempral Relation.** | President Obama *paid*$_{E6}$ tribute *Sunday*$_{T1}$ to 29 workers *killed*$_{E7}$ in an *explosion*$_{E8}$ at a West Virginia coal mine earlier this *month*$_{T2}$. |

**Fig. 7.** Examples of event coreferential, causal and temporal relation. *E* denotes *event* and *T* denotes *time expression* in text.

**Table 7**
Statistics of the commonly used datasets for event coreference resolution. **#Doc** denotes the number of documents, **#Event** denotes the number of events, **Typed** denotes whether the event type is annotated, **#Chain** denotes the number of event coreference chains, and **Language** denotes the language of the annotated document.

| Dataset | #Doc | #Event | Typed | #Chain | Language |
|---|---|---|---|---|---|
| MUC6/7 (Sixth Message Understandi, 1995; Seventh Message Understan, 1998) | 60/ 50 | – | Yes | – | EN |
| ACE2005 (E (Automatic Content Ex, 2005) | 639 | 5557 | Yes | 4268 | EN,CN |
| OntoNotes (Pradhan et al., 2007) | 600 | – | No | – | EN |
| ECB/ECB+ (Cybulska and Vossen, 2014) | 982 | 7268 | No | 4953 | EN |
| TAC KBP 2015 (Mitamura et al., 2015) | 360 | 12976 | Yes | 7460 | EN |
| TAC KBP 2016 (Mitamura et al., 2016) | 505 | 9042 | Yes | 6799 | EN,CN,ES |
| TAC KBP 2017 (Mitamura et al., 2017) | 500 | 8951 | Yes | 8022 | EN,CN,ES |

*Considering other related tasks for ECR.* There are many works that consider resolve ECR with other related tasks jointly, like event extraction, entity recognition, etc. They believe that the relatedness between different tasks could be employed to enhance the performance. In specific, Araki and Mitamura (2015) jointly identified event trigger and resolve event coreference with a structured perceptron training algorithm. Chen and Ng (2016) performed joint inference via integer linear programming (ILP) over the outputs of the models trained for entity extraction, entity coreference, event extraction, and event coreference. Additionally, Lu et al. (2016) performed joint inference using Markov Logic Networks (MLNs) over trigger identification, argument extraction, entity coreference, and event coreference. Barhom et al. (2019) proposed a joint neural architecture for the cross-document entity and event coreference resolution. Lu and Ng (2017b) proposed a joint structured CRFs for event coreference resolution, trigger detection, and event anaphoric determination. Lu and Ng (2017b) treated each of these extracted words and phrases as a candidate event mention. And a structured conditional random field is employed to make joint predictions of the aforementioned three tasks for each candidate event mention.

### 4.2. Event causal relation extraction

An important part of text understanding arises from understanding if and how two events are related to semantically (Mirza, 1604). The causal relation is one important relation type among events. For example, as shown in Fig. 7, the event *killed* (E4) is caused by the event *attack* (E5). Actually, causality is not a linguistic notion. Although language can be used to express causality, it exists as a psychological tool for understanding the world independently of language (Everaert et al., 2012).

#### 4.2.1. Datasets

The current widely used evaluation datasets for event causal relation identification are as follows. (1) **Causal-TimeBank**: Mirza and Tonelli (2014) annotated Causal-TimeBank of event-causal relations based on the TempEval-3 corpus. (2) **EventStoryLine**: Caselli and Vossen (2017) annotated the EventStoryLine for event causal relation identification based on the 320 short stories released by Mostafazadeh et al. (2016). (3) **EventCausality**: Do et al. (2011a) adopted a weakly-supervised method to retrieve additional examples for training models and it is an extremely tiny dataset. (4) **BECauSE 2.0**: Dunietz et al. (2017) presented BECauSE 2.0, a new version of the BECauSE corpus (Dunietz et al., 2015) of causal relation and other seven relations between two spans. (5) **SemEval-2007**: SemEval-2007 Task 4, classification of semantic relations between nominals (Girju et al., 2007) gives access to a corpus containing nominal causal relations among others, as causality is one of the considered semantic relations. (6) **Wall Street Journal**: Bethard et al. (2008) collected 1000 conjoined event pairs connected by from the Wall Street Journal corpus. The event pairs are annotated manually with both temporal and causal relations. (7) **FinReason**: Chen et al. (2021) proposed a financial-domain Chinese corpus regarding extracting the causes of major events in the announcements of listed companies. Each document in this corpus contains one or more structural events, and each event has none, one or more causes in the document. In total, there are 3 event types, including Pledge of shares, Overweighting/Underweighting of shares, and Lawsuit. Existing canonical event extraction and machine reading comprehension methods were performed on this task. The results show a 7 percentage point F1 score gap between the best model and human performance. The statistics of the aforementioned datasets are listed in Table 8.

#### 4.2.2. Methods for event causal relation extraction

Most existing methods usually regarded event causal relation extraction (ECE) as a classification task. That is, given two events and their contexts, to identify whether there is a causal relation between them, even including the corresponding causal relation type. Similar to other event relations identification, classical classifier and modern neural models are employed. Generally, the existing model dealt with it in a supervised setting and the main focused problem is how to extract clues or learn the semantical representations for indicating event causal relations in the contexts. Basically, existing models could be classified into two groups, including *considering internal contextual information* and *considering external casual related knowledge*. Moreover, another important

**Table 8**
Statistics of commonly used datasets for event causal relation extraction. **#Doc** denotes the number of documents, **#Event** denotes the number of events, **#Causal-Link** denotes the number of causal link between events, and **#All-Link** denotes the number of the causal and non-causal link between events.

| Dataset | #Doc | #Event | #Causal-Link | #All-Link |
|---|---|---|---|---|
| Causal-TimeBank (Mirza and Tonelli, 2014) | 184 | 6813 | 318 | 7608 |
| EventStoryLine (Mostafazadeh et al., 2016) | 258 | 5334 | 1770 | 7805 |
| EventCausality (Do et al., 2011a) | 25 | 1134 | 414 | 887 |
| BECauSE 2.0 (Dunietz et al., 2017) | 119 | – | 1803 | 2386 |
| FinReason (Chen et al., 2021) | 8794 | – | 12,861 | 11,006 |

problem is how to acquire labeled examples for model training, while there is little work on it.

*Considering internal contextual information for ECE.* To extract effective clues for indicating event causal relations in the contexts, various textual features are exploited, including syntactic features, lexical features, explicit causal patterns (Hashimoto et al., 2014; Riaz and Girju, 2010, 2014a; Do et al., 2011b; Hidey and McKeown, 2016), statistical causal association (Beamer and Girju, 2009; Hu et al., 1708; Hu and Walker, 1708; Mirza et al., 2014; Mirza and Tonelli, 2016) and etc.

In particular, Riaz and Girju (2013) proposed a set of novel metrics (i.e., Explicit Causal Association (ECA), Implicit Causal Association (ICA), and Boosted Causal Association (BCA)) to identify the likelihood of verb pairs to encode causality. Riaz and Girju (2014b) proposed a set of linguistic features to identify causal relations. For example, some semantic classes of nouns could encode causal or non-causal relations, and a verb-noun pair may not encode causality when a verb and a noun represent the same events. Zuo et al. (2020a) designed a pyramid salient-aware network (PSAN) to understand causal explanatory semantics of context. Gao et al. (2019) modeled rich aspects of document-level causal structures, including *main event in a document, the first sentence as the foreground events*, *syntactic relations between event pairs*, *discourse relations between two text units*, *event coreference relations*, for achieving comprehensive causal relation identification in news articles.

*Considering external causal related knowledge for ECE.* To enhance representation for causal relations, especially when the target texts are short or noisy, some models introduce external causal-related knowledge or data to understand causal semantics between events. Kadowaki et al. (2019) identified causality between events with BERT (Devlin et al., 2019) which is pre-trained with causality candidate documents as background knowledge. Moreover, some methods try to establish external resources of event causal-related commonsense (Rashkin et al., 2018; Sap et al., 1811b; Mostafazadeh et al., 2020), which can be used as event-related knowledge resources for ECE. In particular, Liu et al. (2020c) proposed a knowledge enhanced mention masking generalized model to learn event-agnostic, context-specific patterns for event causality identification. Given two events $e_1$ and $e_2$, the overall approach consists of three major components. (1) *Knowledge aware reasoner* retrieves background knowledge of events in ConcepNet (Speer et al., 2017), and then integrates the knowledge with the processed texts via BERT based encoding. (2) *Event masking reasoner* masks event mentions for reasoning, aiming to learn event-agnostic, context-specific patterns for reasoning. (3) *The attentive sentinel* employs an attention mechanism to balance the above two components for the final prediction.

*Data augmentation for ECE.* The lack of training examples is important in ECE. Riaz and Girju (2014b) used the annotations of FrameNet to generate a training corpus of verb-noun instances encoding cause and non-cause relations. Zuo et al. (2020b) employed distant supervision from the knowledge base for data augmentation. They argued that a sentence that contains an event pair with a high probability of causality and expresses its causal semantic can be simply labeled as a training example. Obviously, such assumption may be weak and introduce many noises.

### 4.3. Event temporal relation extraction

Event temporal relation extraction (ETE) aims to understand the temporal order among events and time expressions in texts. As shown in Fig. 7, the event *paid* (E6) happened on *Sunday* (T1), the event *killed* (E7) is included in the event *explosion* (E8) and happened on *month* (T2), and the event *paid* (E6) happened before the event *killed* and event *explosion*. Temporal relations, or temporal links, are annotations that bring together pieces of markable temporal information in texts, and make a formal representation of temporally ordered events possible (Mirza, 1604).

#### 4.3.1. Datasets

The current widely used evaluation datasets are as follows. (1) **MUC**: The named entity subtasks of MUC-6 and MUC-7 (Sixth Message Understandi, 1995; Seventh Message Understan, 1998) require the identification of absolute (MUC6) and relative (MUC7) time expressions. (2) **TimeBank 1.1**: The annotations of TimeBank 1.1 (Pustejovsky et al., Lazo) follow the 1.1 version of TimeML specifications (Pustejovsky et al., 2003). (3) **TimeBank 1.2**: Timebank 1.2[6] follows the newer TimeML specifications version 1.2.1. The annotation process for Timebank 1.2 is similar to Timebank 1.1, except that all annotations are performed by expert annotators. (4) **AQUAINT**: AQUAINT[6] contains 73 news report documents and is very similar in content to and uses the same specifications as TimeBank 1.2. (5) **TimeBank-Dense**: TimeBank-Dense (Chambers et al., 2014) forces annotators to examine all pairs of events within the same or neighboring sentences to mitigate the sparsity issue. (6) **TempEval**: a) TempEval Corpus[6] (Verhagen et al., 2007a) is created for the temporal relation extraction task at SemEval-2007 based on TimeBank 1.2. b) TempEval-2 Corpus[6] (Verhagen et al., 2010), a multilingual corpus, is created for the Tempeval-2 task at the Semeval-2010 competition. It includes annotations in Chinese, English, French, Italian, Korean, and Spanish. c) TempEval-3 corpus (UzZaman et al., 2013) includes the AQUAINT and a large automatically system annotated â€œsilverâ€ temporal corpus. (7) **MATRES**: MATRES (Ning et al., 2018) enhances the data quality by using a multi-axis annotation scheme and adopting a start point of events to improve inter-annotator agreements. (8) **Dependency Structured Temporal** (Zhang and Xue, 2018): A novel corpus where events and time expressions in a document form a dependency tree (Zhang and Xue, 1808). Specifically, each dependency relation corresponds to an instance of temporal anaphora where the antecedent is the parent and the anaphora is the child. The statistics of commonly used datasets for event temporal relation extraction are listed in Table 9.

**Table 9**
Statistics of commonly used datasets for event temporal relation extraction. **#Doc** denotes the number of documents, **#Event** denotes the number of events, **#Timex** denotes the number of time expressions, and **#Link** denotes the number of the temporal and non-temporal link between events.

| Dataset | #Doc | #Event | #Timex | #Link |
|---|---|---|---|---|
| TimeBank 1.1 (Pustejovsky et al., Lazo) | 300 | 7571 | 1423 | 8242 |
| TimeBank 1.2[6] | 183 | 7935 | 1414 | 9615 |
| AQUAINT[6] | 73 | 4432 | 605 | 6111 |
| TimeBank-Dense (Chambers et al., 2014) | 36 | 1729 | 289 | 12715 |
| TempEval 2007[6] (Verhagen et al., 2007a) | – | 6832 | 1249 | 5790 |
| TempEval 2010[6] (Verhagen et al., 2010) | – | 5688 | 2117 | 4907 |
| TempEval, 2013 (UzZaman et al., 2013) | – | 11145 | 2078 | 11098 |

#### 4.3.2. Methods for event temporal relation extraction

In recent years, the mainstream researches of temporal relation extraction are mainly based on TimeML format (Pustejovsky et al., 2003) that is the most widely used markup language for events, time expressions, and temporal relations. Similar to relation extraction, most of the existing methods modeled ETE as a classification task. In general, existing approaches mainly contained three types, including rule-based, machine learning-based and neural models.

*Temporal rules-based methods for ETE.* To understand the temporal order among events, early ETE models usually relied on temporal rules, such as the rule-based deep syntactic analyzers (Hagège and Tannier, 2007; Strötgen and Gertz, 2010), the rule-based system using knowledge databases (Llorens et al., 2010), the rule-based time expression tagger based on regular expression patterns over tokens (Chang and Manning, 2013), the finite-state rule cascading recognizer and classifier (Zavarella and Tanev, 2013), and the in-house rule-based systems for clinical temporal modeling (Tissot et al., 2015). In particular, Chambers et al. (2014) presented CAEVO, a cascading event ordering architecture, which

[6] http://www.timeml.org/timebank/timebank.html.

included several rule-based classifiers based on linguistic theory. CAEVO naturally integrates these machine-learned classifiers using the sieve architecture.

*Machine learning-based approaches for ETE.* In addition to the temporal rules, some statistical temporal contextual features are used to construct machine learning-based models for ETE. For example, the maximum entropy classifier trained on the data expanded with temporal reasoning (Mani et al., 2006), the pair-wise classifier which avoided the pitfalls of evaluating a graph of inter-related labels by defining three sub-tasks (Verhagen et al., 2007b), the two-stage classifier (Chambers et al., 2007), the automatic events and time expressions identifier (Verhagen and Pustejovsky, 2008), the ClearTK, a pipeline machine-learning model with a simple morpho-syntactic annotation (Bethard, 2013), the UTTime, the logistic regression classifiers with a deep syntactic parser (Laokulrat et al., 2013), the NavyTime, a split classifier approach that breaks the ordering tasks into smaller decision points (Chambers, 2013), the Markov Logic model that jointly identifies relations of three relation types (Yoshikawa et al., 2009), the structured learning approach (Ning et al., 2017), and the scalable structured learning model (Moens and Leeuwenberg, 2017).

*Neural models for ETE.* Recently, a lot of neural models have been proposed to capture temporal relations, such as the basic CNNs and LSTMs approaches (Tourille et al., 2017a; Dligach et al., 2017), the dependency paths based BiLSTM models (Cheng and Miyao, 2017; Meng et al., 2017), the context-aware neural-based model (Meng and Rumshisky, 2018), and the end-to-end neural models on the clinical domain (Chikka, 2016; Li and Huang, 2016). Additionally, some ensemble models which have been proposed, such as the recurrent neural model combined with SVMs and rules on clinical domain (Tourille et al., 2017b; Long et al., 2017). Moreover, other more refined neural models have also been gradually proposed, like the relative time-line constructor (Leeuwenberg and Moens, 2018), the joint event and temporal relation extractor (Han et al., 2019), the improved neural models for temporal relation extraction (Ning et al., 2019), the structured neural network empowered by domain knowledge (Han et al., 2020), and the contextualized neural language models for temporal dependency parsing (Ross et al., 2020). And Han et al. (2019) proposed a neural structured prediction model with joint representation learning to make predictions on events and relations simultaneously. Specifically, a RNN-based multi-task scoring module was exploited which included an event scorer and a relation scorer for both event and relation prediction.

### 4.4. Comparison of methods for different event relation extraction

Since there is no consistent task formulation for different event relation extraction, it is hard to make a summary for the existing approaches. Moreover, we try to take analysis and comparisons of the aforementioned methods for three event relation extraction tasks. As shown in Table 10, we could obtain the following briefly conclusions:

*Sentence-level vs. Document-level Extraction.* Coreferential events are mainly distributed in the document, including within-document and cross-document, and rarely distributed in the same sentence. Therefore, most existing approaches handled ECR at the document level. On the contrary, causal and temporal related events are distributed both within one sentence and cross different sentences of one document. Thus, both sentence-level and document-level proposed were proposed.

*Pipline vs. Joint.* To extract event relations, most methods are pipeline models which extract events first and then identify the relations between the events. To address the error propagation problem of the pipeline model, the methods of jointly extracting events and identifying event relations are proposed in the event coreference resolution and event temporal relation extraction.

*Local vs. Global.* The relations of events is distributed among different sentences in the document, and different events affect each other. Therefore, based on the traditional local models, more global models are gradually proposed to understand the event relations.

*Data Augmentation.* Most event relation extraction approaches were based on the supervised setting, where sufficient labeled instances are needed. To produce sufficient training data, data augmentation methods are proposed, especially for the training of event causal extraction. However, because event coreference and temporal relation are more biased towards the document level, data augmentation is seldom addressed for these relation types.

*Employing Event Argument.* The understanding of event coreference is strongly dependent on event arguments. Relatively, event argument information has little influence on the identification of event causal relation and event temporal relation.

*Employing External Resources.* To date, external resources for event causal relation extraction and event temporal relation extraction are introduced. However, event coreference resolution lacks external resources.

**Table 10**
Comparisons of different methods on three event relation extraction tasks.

| Methods on Different Tasks | Coreference | Causal Relation | Temporal Relation |
|---|---|---|---|
| Sentence/Document Level | ✓/× | ✓/✓ | ✓/✓ |
| Pipeline/Joint | ✓/✓ | ✓/× | ✓/✓ |
| Local/Global | ✓/✓ | ✓/✓ | ✓/✓ |
| Data Augmentation | × | ✓ | × |
| Employing Event Argument | ✓ | × | × |
| Employing External Resources | × | ✓ | ✓ |

## 5. Recent research in event knowledge graph

Most aforementioned event/relation extraction approaches are under investigated, and their performance are still not satisfied for event-centric knowledge graph construction. Nevertheless, several event knowledge graphs have been constructed for different aims. We introduce them briefly as follows.

*EventKG.*Gottschalk and Demidova (2018) presented an EventKG to take an important step to facilitate a global view on events and temporal relations currently spread across entity-centric knowledge graphs and manually curated semi-structured sources. EventKG currently includes data sources in five languages - English, German, French, Russian, and Portuguese.

*ELG.*Ding et al. (1907) presented an Event Logic Graph (ELG), a directed cyclic graph, whose nodes are events, and edges stand for the sequential, causal, conditional, or hypernym-hyponym relations between events. Essentially, ELG is an event logic knowledge base, which reveals evolutionary patterns and development logics of real-world events.

*ASER.*Zhang et al. (2020) proposed an ASER for discovering useful real-world knowledge about Activities (or process, e.g., "I sleep"), States (e.g., "I am hungry"), Events (e.g., "I make a call"), and their relations (e.g., "I am hungry" may result in "I have lunch"). ASER defines a brand new KG where the primitive units of semantics are eventualities. Each eventuality instance is a hyperedge connecting several vertices (words). A relation between two eventualities in ASER represents one of the 14 relation types defined in PDTB or a co-occurrence relation. ASER designed several high-quality patterns based on dependency parsing results and extract all eventualities over large-scale corpora.

Events are important semantical units in commonsense knowledge, like states, activities. Thus, several works try to build commonsense resources surrounding events, such as *Event2Mind* (Rashkin et al., 2018), *GLUCOSE* (Mostafazadeh et al., 2020), *ATOMIC* (Sap et al., 1811a), *(COMET-)ATOMIC*$20^{20}$ (Hwang et al., 2020). In these resources, an event usually is represented as phrases or sentences instead of the definition in ACE.

*Event2Mind.*Rashkin et al. (2018) introduced a new task, corpus, and model supporting commonsense inference on events. It specifically focused on modeling stereotypical intents and reactions of people, described in short free-form text (event phrases). The goal of Event2Mind is to probe whether it is feasible to build computational models that can perform limited, but well-scoped commonsense inference on event phrases.

*GLUCOSE*.Mostafazadeh et al. (2020) introduced the GeneraLized and COntextualized Story Explanations (GLUCOSE) dataset. Given a short story and a sentence X in the story, GLUCOSE captures ten dimensions of causal explanation related to X. These dimensions are designed to focus on causal reasoning around events and states, eliciting event causal chains, character motivations, emotions, naive psychology, and change of attributes such as location and possessions to core story entities.

*ATOMIC*.Sap et al. (1811a) introduced the ATOMIC, an atlas of machine commonsense, as a step toward addressing the rich spectrum of inferential knowledge that is crucial for automated commonsense reasoning. It mainly focuses on inferential *if-then* knowledge. They employed crowdsourcing and proposed a new taxonomy of if-then reasoning types. One way to categorize the types is based on the content being predicted: (1) *If-Event-Then-Mental-State*, (2) *If-Event-Then-Event*, and (3) *If-Event-Then-Persona*. Another way to categorize is based on their causal relations: (1) "causes", (2) "effects", and (3) "stative". Using this taxonomy, ATOMIC gathers over 877K instances of inferential knowledge.

*(COMET-)ATOMIC*$20^{20}$.Hwang et al. (2020) presented ATOMIC$20^{20}$, a commonsense knowledge graph with 1.33M everyday inferential knowledge tuples about entities and events. ATOMIC$20^{20}$ represents a large-scale commonsense repository of textual descriptions that encode both the social and the physical aspects of common human everyday experiences, collected to be complementary to commonsense knowledge encoded in current language models. ATOMIC$20^{20}$ introduces 23 commonsense relations types. They can be broadly classified into three categorical types: 9 commonsense relations of social-interaction, 7 physical-entity commonsense relations, and 7 event-centered commonsense relations concerning situations surrounding a given event of interest.

Table 11 presents the statistics of these event knowledge graphs. Most of them have large scales and have been proved to be effective for reasoning-driven applications.

**Table 11**
The statistics of different event knowledge graphs.

| Event Knowledge Graphs | Statistics |
|---|---|
| EventKG (Gottschalk and Demidova, 2018) | 88,473,111 Triples, 609,247 Nodes (Events) |
| ELG (Ding et al., 1907) | 2173 Relations (Logic Relations) |
| ASER (Zhang et al., 2020) | 64,351,959 Triples, 194,000,677 Nodes (Eventuality) |
| Event2Mind (Rashkin et al., 2018) | 24,716 Unique Event Phrases |
| GLUCOSE (Mostafazadeh et al., 2020) | 670K Annotations, 335K Rules (Causal Commonsense) |
| ATOMIC (Sap et al., 1811a) | 877,108 Triples, 309,515 Nodes (Event Phrases) |
| (COMET-)ATOMIC$20^{20}$ (Hwang et al., 2020) | 1.33M Triples |

## 6. Summary and challenges

This paper introduces a survey on the task of event and event relation extraction. In event extraction, we focus on recent three research topics and corresponding methods, including recent neural models for the sentence level event extraction, methods for across sentences or document-level event extraction and data augmentation approaches. In event relation extraction, we mainly focus on extraction methods for three relation types, such as event coreference, causal relation, and temporal relation. We also give the widely used evaluation datasets and current performances on each dataset. Nevertheless, there are several challenges or problems to be focused on and investigated further.

*How to represent event in the texts?* Most current event extraction approaches are based on the event definition in ACE. In texts, an event is defined to be triggered by a trigger word. As a result, the trigger word is usually regarded as the attorney for the corresponding event in texts. The event arguments extraction is regarded as judging the relations between some words/phrases and the trigger word. And current event relation extraction is also regarded as judging the relations among different given trigger words. However, only representing an event as a single trigger word or phrase is explicitly unreasonable. In many cases, an event is usually triggered/described by multiple words or several sentences. For example, "Kang Inc. declared its pledge of 40,000,000 shares to CRCB bank since Jan 1st, 2016. … Kang's pledge aimed at providing a guarantee for self-financing …". The cause of a PLEDGE event (Kang Inc.'s pledgetion) is another event that is described in a textual span (providing a guarantee for self-financing) other than a single trigger word. Thus, how to represent events in texts rather than a trigger word is still an unsolved problem.

*Document-level event extraction needs deep analysis.* Although we introduce some works for extracting events at the document-level, the researches in such a field is still scarce. Compared with sentence-level event extraction, extracting events across sentences or in the whole document needs to take a large view of texts. *Arguments-scattering* and *multi-events* problems are still serious. Although some graph-based approaches (Zheng et al., 1904; Chen et al., 2020) were proposed to capture the structure of the document and relations between sentences, the solutions are still simple and opaque. It needs deep analysis in the future.

*How to perform multi-modal event extraction?* Only extracting events from texts is limited. In the real world, an event may be described by different modalities, including texts, speech, images, and videos. Jointly using multi-modal information could be helpful for the event extraction system to disambiguate and complement information mutually. However, there is little work focusing on this problem, where how to represent multi-modal information in a unified semantic space and compute their alignments are challenging problems.

*How to define events and perform extraction in open domains?* Currently, the event definition mostly follows that in ACE, where an event is defined as a structure. In Freebase, an event is defined as a *compound value type* (CVT), which is also based on a structured format. Based on such definition, existing approaches perform extraction under the pre-defined event frames. However, not all events could be formulated in such structured formats. Moreover, enumerating all frames for each event type is exhaustive and impractical. Therefore, in some event knowledge bases (like ATOMIC), an event is represented as a sentence instead of a structured format. We believe the reason is that we could not know and predefine event types under many scenarios, such as commonsense reasoning. Moreover, besides coreference, causal and temporal relations mentioned in this paper, are there other event relation types that are still under-investigated. Formulating event relation extraction as a classification problem by given two trigger words or sentences is limited. Therefore, how to define events/relations, especially for those unseen ones, is an important problem for event extraction. In this way, open event extraction like OpenIE may be needed.

In general, the current methods, whatever extracting events or event relations, are still not satisfied for event-centric knowledge graph construction and other downstream tasks. There are still many questions that need to be resolved further. Nevertheless, extracting events and event relations has taken a big step forward, which makes the knowledge extraction not always limited to the entity level. We expect that it could attract more and more attentions in the future.

## Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

## Acknowledgement

This work is supported by the Natural Key R&D Program of China (No.2018YFB1005100), the National Natural Science Foundation of China (No. 61922085, No.U1936207, No.61806201) and the Strategic Priority Research Program of Chinese Academy of Sciences (Grant No. XDA27000000). This research work was also supported by Beijing Academy of Artificial Intelligence (BAAI).

## References

Ahn, D., 2006a. The stages of event extraction. In: Proceedings of the Workshop on Annotating and Reasoning about Time and Events. Association for Computational Linguistics. URL. http://www.aclweb.org/anthology/W06-0901.

Ahn, D., 2006b. The Stages of Event Extraction.

Araki, J., Mitamura, T., 2015. Joint Event Trigger Identification and Event Coreference Resolution with Structured Perceptron. EMNLP.

Barhom, S., Shwartz, V., Eirew, A., Bugert, M., Reimers, N., Dagan, I., 2019. Revisiting Joint Modeling of Cross-Document Entity and Event Coreference Resolution. ACL.

Beamer, B., Girju, R., 2009. Using a bigram event model to predict causal potential. In: International Conference on Intelligent Text Processing and Computational Linguistics. Springer, pp. 430–441.

Bejan, C., Harabagiu, S.M., 2010. Unsupervised Event Coreference Resolution with Rich Linguistic Features. ACL.

Bethard, S., Corvey, W., Klingenstein, S., Martin, J.H., 2008. Building a corpus of temporal-causal structure. In: Proceedings of the Sixth International Conference on Language Resources and Evaluation (LREC'08), European Language Resources Association (ELRA), Marrakech, Morocco. URL. http://www.lrec-conf.org/proceedings/lrec2008/pdf/229_paper.pdf.

Bethard, S., 2013. ClearTK-TimeML: A minimalist approach to TempEval. In: Second Joint Conference on Lexical and Computational Semantics (*SEM), vols. 10–14. Association for Computational Linguistics, Atlanta, Georgia, USA, 2013, vol. 2: Proceedings of the Seventh International Workshop on Semantic Evaluation (SemEval 2013). https://www.aclweb.org/anthology/S13-2002.

Caselli, T., Vossen, P., 2017. The event storyline corpus: a new benchmark for causal and temporal relation extraction. In: Proceedings of the Events and Stories in the News Workshop, vols. 77–86.

Chambers, N., 2013. NavyTime: event and time ordering from raw text. In: Second Joint Conference on Lexical and Computational Semantics (*SEM), Volume 2: Proceedings of the Seventh International Workshop on Semantic Evaluation (SemEval 2013). Association for Computational Linguistics, Atlanta, Georgia, USA, pp. 73–77. https://www.aclweb.org/anthology/S13-2012. URL.

Chambers, N., Wang, S., Jurafsky, D., 2007. Classifying temporal relations between events. In: Proceedings of the 45th Annual Meeting of the ACL on Interactive Poster and Demonstration Sessions, ACL '07, vols. 173–176. Association for Computational Linguistics, USA.

Chambers, N., Cassidy, T., McDowell, B., Bethard, S., 2014. Dense event ordering with a multi-pass architecture. Transactions of the Association for Computational Linguistics 2, 273–284. https://doi.org/10.1162/tacl̀00182. URL. https://www.aclweb.org/anthology/Q14-1022.

Chang, A.X., Manning, C.D., 2013. SUTime: Evaluation in TempEval-3. SemEval@NAACL-HLT.

Chen, Z., Ji, H., 2009. Graph-based event coreference resolution. In: Graph-based Methods for Natural Language Processing.

Chen, C., Ng, V., 2016. Joint Inference over a Lightly Supervised Information Extraction Pipeline: towards Event Coreference Resolution for Resource-Scarce Languages. AAAI.

Chen, Z., Ji, H., Haralick, R., 2009. A Pairwise Event Coreference Model, Feature Impact and Evaluation for Event Coreference Resolution.

Chen, Y., Xu, L., Liu, K., Zeng, D., Zhao, J., 2015. Event extraction via dynamic multi-pooling convolutional neural networks. In: ACL, the Association for Computer Linguistics. ISBN 978-1-941643-72-3, URL. http://dblp.uni-trier.de/db/conf/acl/acl2015-1.html.

Chen, Y., Liu, S., He, S., Liu, K., Zhao, J., 2016. Event Extraction via Bidirectional Long Short-Term Memory Tensor Neural Networks. CCL.

Chen, Y., Liu, S., Zhang, X., Liu, K., Zhao, J., 2017. Automatically labeled data generation for large scale event extraction. In: Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics, vol. 1, pp. 409–419. Long Papers.

Chen, Y., Yang, H., Liu, K., Zhao, J., Jia, Y., 2018. Collective event detection via a hierarchical and bias tagging networks with gated multi-level attention mechanisms. In: Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing. Association for Computational Linguistics, Brussels, Belgium, pp. 1267–1276. https://doi.org/10.18653/v1/D18-1158. URL. https://www.aclweb.org/anthology/D18-1158.

Chen, P., Yang, H., Liu, K., Huang, R., Chen, Y., Wang, T., Zhao, J., 2020. Reconstructing event regions for event extraction via graph attention networks. In: Proceedings of the 1st Conference of the Asia-Pacific Chapter of the Association for Computational Linguistics and the 10th International Joint Conference on Natural Language Processing. Association for Computational Linguistics, Suzhou, China, pp. 811–820. URL. https://www.aclweb.org/anthology/2020.aacl-main.81.

Chen, P., Liu, K., Chen, Y., Wang, T., Zhao, J., 2021. Probing into the root: a dataset for reason extraction of structural events from financial documents. In: Proceedings of the 16th Conference of the European Chapter of the Association for Computational Linguistics.

Cheng, F., Miyao, Y., 2017. Classifying Temporal Relations by Bidirectional LSTM over Dependency Paths. ACL.

Chikka, V.R., 2016. CDE-IIITH at SemEval-2016 Task 12: Extraction of Temporal Information from Clinical Documents Using Machine Learning Techniques. SemEval@NAACL-HLT.

Choubey, P.K., Huang, R., 2017. Event Coreference Resolution by Iteratively Unfolding Inter-dependencies Among Events. EMNLP.

Choubey, P.K., Huang, R., 2018. Improving Event Coreference Resolution by Modeling Correlations between Event Coreference Chains and Document Topic Structures. ACL.

Cui, S., Yu, B., Liu, T., Zhang, Z., Wang, X., Shi, J., 2020. Edge-Enhanced Graph Convolution Networks for Event Detection with Syntactic Relation.

Cybulska, A., Vossen, P., 2014. Using a sledgehammer to crack a nut? Lexical diversity and event coreference resolution. In: Proceedings of the Ninth International Conference on Language Resources and Evaluation (LREC'14), European Language Resources Association (ELRA), Reykjavik, Iceland, pp. 4545–4552. URL. http://www.lrec-conf.org/proceedings/lrec2014/pdf/840_Paper.pdf.

Cybulska, A., Vossen, P.T.J.M., 2015. Translating Granularity of Event Slots into Features for Event Coreference Resolution. EVENTS@HLP-NAACL.

Devlin, J., Chang, M.-W., Lee, K., Toutanova, K., 2019. BERT: pre-training of deep bidirectional Transformers for language understanding. In: Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, vol. 1. Association for Computational Linguistics, Minneapolis, Minnesota, pp. 4171–4186. https://doi.org/10.18653/v1/N19-1423. Long and Short Papers. https://www.aclweb.org/anthology/N19-1423. URL.

X. Ding, Z. Li, T. Liu, K. Liao, ELG: an Event Logic Graph, ArXiv abs/1907.08015.

Dligach, D., Miller, T., Lin, C., Bethard, S., Savova, G., 2017. Neural temporal relation extraction. In: Proceedings of the 15th Conference of the European Chapter of the Association for Computational Linguistics: Volume 2, Short Papers. Association for Computational Linguistics, Valencia, Spain, pp. 746–751. URL. https://www.aclweb.org/anthology/E17-2118.

Do, Q., Chan, Y.S., Roth, D., 2011a. Minimally supervised event causality identification. In: Proceedings of the 2011 Conference on Empirical Methods in Natural Language Processing. Association for Computational Linguistics, Edinburgh, Scotland, UK, pp. 294–303. URL. https://www.aclweb.org/anthology/D11-1027.

Do, Q.X., Chan, Y.S., Roth, D., 2011b. Minimally supervised event causality identification. EMNLP, ACL 294–303.

X. Du, C. Cardie, Document-Level Event Role Filler Extraction Using Multi-Granularity Contextualized Encoding, arXiv preprint arXiv:2005.06579 .

Du, X., Cardie, C., 2020. Event extraction by answering (almost) natural questions. In: Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing (EMNLP). Association for Computational Linguistics, Online, pp. 671–683. https://doi.org/10.18653/v1/2020.emnlp-main.49. URL. https://www.aclweb.org/anthology/2020.emnlp-main.49.

Duan, S., He, R., Zhao, W., 2017. Exploiting document level information to improve event detection via recurrent neural networks. In: Proceedings of the Eighth International Joint Conference on Natural Language Processing (Volume 1: Long Papers). Asian Federation of Natural Language Processing, Taipei, Taiwan, pp. 352–361. URL. https://www.aclweb.org/anthology/I17-1036.

Dunietz, J., Levin, L., Carbonell, J., 2015. Annotating causal language using corpus lexicography of constructions. In: Proceedings of the 9th Linguistic Annotation Workshop. ACL, Denver, Colorado, USA, pp. 188–196.

Dunietz, J., Levin, L., Carbonell, J., 2017. The BECauSE corpus 2.0: annotating causality and overlapping relations. In: Proceedings of the 11th Linguistic Annotation Workshop, vols. 95–104. ACL, Valencia, Spain.

ACE, 2005. Automatic content extraction) English annotation guidelines for events, technical report, linguistic data consortium. URL. https://www.ldc.upenn.edu/sites/www.ldc.upenn.edu/files/english-relations-guidelines-v6.2.pdf.

Everaert, M., Mar, M., Siloni, T., 2012. The Theta System: Argument Structure at the Interface 9780199602513, pp. 1–432. https://doi.org/10.1093/acprof:oso/9780199602513.001.0001.

Feng, X., Huang, L., Tang, D., Ji, H., Qin, B., Liu, T., 2016. A language-independent neural network for event detection. In: Short Papers). Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics, vol. 2. Association for Computational Linguistics, Berlin, Germany, pp. 66–71. https://doi.org/10.18653/v1/P16-2011. URL. https://www.aclweb.org/anthology/P16-2011.

Feng, X., Guo, J., Qin, B., Liu, T., Liu, Y., 2017. Effective deep memory networks for distant supervised relation extraction. IJCAI 17. https://doi.org/10.24963/ijcai.2017/559.

Gao, L., Choubey, P.K., Huang, R., Huang, R., Huang, R., 2019. Modeling document-level causal structures for event causal relation identification. In: NAACL, ACL, Minneapolis, Minnesota, pp. 1808–1817.

Girju, R., Nakov, P., Nastase, V., Szpakowicz, S., Turney, P., Yuret, D., 2007. SemEval-2007 task 04: classification of semantic relations between nominals. In: Proceedings of the Fourth International Workshop on Semantic Evaluations (SemEval-2007), vols. 13–18. Association for Computational Linguistics, Prague, Czech Republic. URL. https://www.aclweb.org/anthology/S07-1003.

Gottschalk, S., Demidova, E., 2018. EventKG: A Multilingual Event-Centric Temporal Knowledge Graph. ESWC.

Hagège, C., Tannier, X., 2007. XRCE-T: XIP Temporal Module for TempEval Campaign. SemEval@ACL.
Han, R., Ning, Q., Peng, N., 2019. Joint event and temporal relation extraction with shared representations and structured prediction. In: Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing (EMNLP-IJCNLP). Association for Computational Linguistics, Hong Kong, China, pp. 434–444. https://doi.org/10.18653/v1/D19-1041. URL. https://www.aclweb.org/anthology/D19-1041.
Han, R., Zhou, Y., Peng, N., 2020. Domain Knowledge Empowered Structured Neural Net for End-To-End Event Temporal Relation Extraction.
Hashimoto, C., Torisawa, K., Kloetzer, J., Sano, M., Varga, I., Oh, J.-H., Kidawara, Y., 2014. Toward future scenario generation: extracting event causality exploiting semantic relation, context, and association features. ACLPPinforma 987–997.
Hidey, C., McKeown, K., 2016. Identifying causal relations using parallel wikipedia articles. In: ACL, ACL, Berlin, Germany, pp. 1424–1433.
Hong, Y., Zhang, J., Ma, B., Yao, J., Zhou, G., Zhu, Q., 2011. Using cross-entity inference to improve event extraction. In: ACL, Association for Computational Linguistics. URL. http://www.aclweb.org/anthology/P11-1113.
Hong, Y., Zhou, W., Zhang, J., Zhou, G., Zhu, Q., 2018. Self-regulation: employing a generative adversarial network to improve event detection. In: Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers). Association for Computational Linguistics, Melbourne, Australia, pp. 515–526. https://doi.org/10.18653/v1/P18-1048. URL. https://www.aclweb.org/anthology/P18-1048.
Hsi, A., Yang, Y., Carbonell, J.G., Xu, R., 2016. Leveraging multilingual training for limited resource event extraction. In: Proceedings of COLING 2016, the 26th International Conference on Computational Linguistics: Technical Papers, pp. 1201–1210.
Z. Hu, M. A. Walker, Inferring Narrative Causality between Event Pairs in Films, arXiv preprint arXiv:1708.09496 .
Z. Hu, E. Rahimtoroghi, M. A. Walker, Inference of Fine-Grained Event Causality from Blogs and Films, arXiv preprint arXiv:1708.09453 .
Huang, R., Riloff, E., 2011. Peeling back the layers: detecting event role fillers in secondary contexts. In: Proceedings of the 49th Annual Meeting of the Association for Computational Linguistics: Human Language Technologies, vols. 1137–1147.
Huang, R., Riloff, E., 2012. Modeling textual cohesion for event extraction. AAAI 1. Citeseer, 1.
Huang, L., Cassidy, T., Feng, X., Ji, H., Voss, C., Han, J., Sil, A., 2016. Liberal event extraction and event schema induction. In: Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics, vol. 1, pp. 258–268. Long Papers.
Huang, Y., Lu, J., Kurohashi, S., Ng, V., 2019. Improving Event Coreference Resolution by Learning Argument Compatibility from Unlabeled Data. NAACL-HLT.
J. D. Hwang, C. Bhagavatula, R. L. Bras, J. Da, K. Sakaguchi, A. Bosselut, Y. Choi, COMET-ATOMIC 2020: on Symbolic and Neural Commonsense Knowledge Graphs, ArXiv abs/2010.05953.
Ji, H., 2009. Cross-lingual predicate cluster acquisition to improve bilingual event extraction by inductive learning. In: Proceedings of the Workshop on Unsupervised and Minimally Supervised Learning of Lexical Semantics, vols. 27–35.
Ji, H., Grishman, R., 2008. Refining event extraction through cross-document inference. In: ACL, Association for Computational Linguistics. URL. http://www.aclweb.org/anthology/P/P08/P08-1030.
Kadowaki, K., Iida, R., Torisawa, K., Oh, J.-H., Kloetzer, J., 2019. Event causality recognition exploiting multiple annotators' judgments and background knowledge. In: Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing (EMNLP-IJCNLP). Association for Computational Linguistics, Hong Kong, China, pp. 5816–5822. https://doi.org/10.18653/v1/D19-1590. URL. https://www.aclweb.org/anthology/D19-1590.
Laokulrat, N., Miwa, M., Tsuruoka, Y., Chikayama, T., 2013. UTTime: temporal relation classification using deep syntactic features. In: Second Joint Conference on Lexical and Computational Semantics (*SEM), Volume 2: Proceedings of the Seventh International Workshop on Semantic Evaluation (SemEval 2013), vols. 88–92. Association for Computational Linguistics, Atlanta, Georgia, USA. https://www.aclweb.org/anthology/S13-2015. URL.
Le, Q.V., Mikolov, T.. Distributed Representations of Sentences and Documents, CoRR abs/1405.4053. URL. http://arxiv.org/abs/1405.4053.
Lee, H., Recasens, M., Chang, A., Surdeanu, M., Jurafsky, D., 2012. Joint entity and event coreference resolution across documents. In: Proceedings of the 2012 Joint Conference on Empirical Methods in Natural Language Processing and Computational Natural Language Learning. Association for Computational Linguistics. Jeju Island, Korea, 489–500, URL. https://www.aclweb.org/anthology/D12-1045.
Leeuwenberg, A., Moens, M.-F., 2018. Temporal information extraction by predicting relative time-lines. In: Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing. Association for Computational Linguistics, Brussels, Belgium, pp. 1237–1246. https://doi.org/10.18653/v1/D18-1155. URL. https://www.aclweb.org/anthology/D18-1155.
Li, P., Huang, H., 2016. UTA DLNLP at SemEval-2016 Task 12: Deep Learning Based Natural Language Processing System for Clinical Information Identification from Clinical Notes and Pathology Reports. SemEval@NAACL-HLT.
Li, Q., Ji, H., Huang, L., 2013. Joint event extraction via structured prediction with global features. In: Proceedings of the 51st Annual Meeting of the Association for Computational Linguistics, vol. 1. Association for Computational Linguistics, Sofia, Bulgaria, pp. 73–82. Long Papers). https://www.aclweb.org/anthology/P13-1008.
Liao, S., Grishman, R., 2010. Using document level cross-event inference to improve event extraction. In: ACL, ACL '10, Association for Computational Linguistics. URL. http://dl.acm.org/citation.cfm?id=1858681.1858762.
Liu, S., Chen, Y., Liu, K., Zhao, J., 2017a. Exploiting argument information to improve event detection via supervised attention mechanisms, 1789–1798. In: Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers). Association for Computational Linguistics, Vancouver, Canada. https://doi.org/10.18653/v1/P17-1164. URL. https://www.aclweb.org/anthology/P17-1164.
Liu, Z., Araki, J., Hovy, E., Mitamura, T., 2014. Supervised Within-Document Event Coreference Using Information Propagation. LREC.
Z. Liu, J. Araki, T. Mitamura, E. Hovy, CMU-LTI at KBP 2016 Event Nugget Track, Theory and Applications of Categories .
Liu, S., Liu, K., He, S., Zhao, J., 2016a. A probabilistic soft logic based approach to exploiting latent and global information in event classification. In: AAAI. AAAI Press.
Liu, S., Chen, Y., He, S., Liu, K., Zhao, J., 2016b. Leveraging framenet to improve automatic event detection. In: Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics, vol. 1, pp. 2134–2143. Long Papers.
Liu, S., Chen, Y., Liu, K., Zhao, J., Luo, Z., Luo, W., 2017. Improving Event Detection via Information Sharing Among Related Event Types.
Liu, S., Cheng, R., Yu, X., Cheng, X., 2018a. Exploiting contextual information via dynamic memory network for event detection. In: Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing. Association for Computational Linguistics, Brussels, Belgium, pp. 1030–1035. https://doi.org/10.18653/v1/D18-1127. URL. https://www.aclweb.org/anthology/D18-1127.
Liu, X., Luo, Z., Huang, H., 2018b. Jointly multiple events extraction via attention-based graph information aggregation. In: Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing. Association for Computational Linguistics, Brussels, Belgium, pp. 1247–1256. https://doi.org/10.18653/v1/D18-1156. URL. https://www.aclweb.org/anthology/D18-1156.
Liu, J., Chen, Y., Liu, K., Zhao, J., 2018c. Event detection via gated multilingual attention mechanism. In: Thirty-Second AAAI Conference on Artificial Intelligence.
Liu, J., Chen, Y., Liu, K., 2019a. Exploiting the ground-truth: an adversarial imitation based knowledge distillation approach for event detection. Proceedings of the AAAI Conference on Artificial Intelligence 33, 6754–6761. https://doi.org/10.1609/aaai.v33i01.33016754.
Liu, J., Chen, Y., Liu, K., Zhao, J., 2019b. Neural cross-lingual event detection with minimal parallel resources. In: Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing (EMNLP-IJCNLP). Association for Computational Linguistics, Hong Kong, China, pp. 738–748. https://doi.org/10.18653/v1/D19-1068. URL. https://www.aclweb.org/anthology/D19-1068.
Liu, J., Chen, Y., Liu, K., Zhao, J., 2019c. Neural cross-lingual event detection with minimal parallel resources. In: Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing, vols. 738–748. EMNLP-IJCNLP).
Liu, J., Chen, Y., Liu, K., Jia, Y., Sheng, Z., 2020a. How does context matter? On the robustness of event detection with context-selective mask generalization. In: Findings of the Association for Computational Linguistics: EMNLP 2020. Association for Computational Linguistics, Online, pp. 2523–2532. https://doi.org/10.18653/v1/2020.findings-emnlp.229. URL. https://www.aclweb.org/anthology/2020.findings-emnlp.229.
Liu, J., Chen, Y., Liu, K., Bi, W., Liu, X., 2020b. Event extraction as machine reading comprehension. In: Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing (EMNLP). Association for Computational Linguistics, Online, pp. 1641–1651. https://doi.org/10.18653/v1/2020.emnlp-main.128. URL. https://www.aclweb.org/anthology/2020.emnlp-main.128.
Liu, J., Chen, Y., Zhao, J., 2020c. Knowledge enhanced event causality identification with mention masking generalizations. In: Bessiere, C. (Ed.), IJCAI-20, International Joint Conferences on Artificial Intelligence Organization, 3608–3614, main track.
Llorens, H., Boró, E.S., Navarro-Colorado, B., 2010. TIPSem (English and Spanish): Evaluating CRFs and Semantic Roles in TempEval-2. SemEval@ACL.
Long, Y., Li, Z., Wang, X., Li, C., 2017. XJNLP at SemEval-2017 Task 12: Clinical Temporal Information Ex-Traction with a Hybrid Model. SemEval@ACL.
Lu, J., Ng, V., 2016. UTD's event nugget detection and coreference system at KBP 2016. In: Proceedings of the 2016 Text Analysis Conference. TAC 2016, Gaithersburg, Maryland, USA. November 14-15, 2016, NIST, URL. https://tac.nist.gov/publications/2016/participant.papers/TAC2016.UTD.proceedings.pdf.
Lu, J., Ng, V., 2017a. Learning Antecedent Structures for Event Coreference Resolution, 2017 16th IEEE International Conference on Machine Learning and Applications. ICMLA), pp. 113–118.
Lu, J., Ng, V., 2017b. Joint Learning for Event Coreference Resolution. ACL.
Lu, J., Venugopal, D., Gogate, V., Ng, V., 2016. Joint Inference for Event Coreference Resolution. COLING.
Mani, I., Verhagen, M., Wellner, B., Lee, C.M., Pustejovsky, J., 2006. Machine learning of temporal relations. In: Proceedings of the 21st International Conference on Computational Linguistics and the 44th Annual Meeting of the Association for Computational Linguistics, ACL-44. Association for Computational Linguistics, USA, pp. 753–760. https://doi.org/10.3115/1220175.1220270. URL.
Meng, Y., Rumshisky, A., 2018. Context-Aware Neural Model for Temporal Information Extraction. ACL.
Meng, Y., Rumshisky, A., Romanov, A., 2017. Temporal Information Extraction for Question Answering Using Syntactic Dependencies in an LSTM-Based Architecture. EMNLP.
Mirza, P.. Extracting Temporal and Causal Relations between Events, CoRR abs/1604.08120. URL. http://arxiv.org/abs/1604.08120.

Mirza, P., Tonelli, S., 2014. An Analysis of Causality between Events and its Relation to Temporal Information. COLING.
Mirza, P., Tonelli, S., 2016. Catena: causal and temporal relation extraction from natural language texts. COLING 64–75.
Mirza, P., Sprugnoli, R., Tonelli, S., Speranza, M., 2014. Annotating causality in the TempEval-3 corpus. In: EACL 2014 Workshop on Computational Approaches to Causality in Language (CAtoCL), ACL, vols. 10–19.
T. Mitamura, Z. Liu, E. Hovy, Overview of TAC KBP 2015 Event Nugget Track, Theory and Applications of Categories .
Mitamura, T., Liu, Z., Hovy, E.H., 2016. Overview of TAC-KBP 2016 event nugget track. In: Proceedings of the 2016 Text Analysis Conference, TAC 2016, Gaithersburg, Maryland, USA, November 14-15, 2016, NIST. URL. https://tac.nist.gov/publications/2016/additional.papers/TAC2016.KBP_Event_Nugget_overview.proceedings.pdf.
T. Mitamura, Z. Liu, E. Hovy, Events Detection, Coreference and Sequencing: What's Next? Overview of the TAC KBP 2017 Event Track, Theory and Applications of Categories .
Moens, M.-F., Leeuwenberg, A., 2017. Structured Learning for Temporal Relation Extraction from Clinical Records. EACL.
Mostafazadeh, N., Grealish, A., Chambers, N., Allen, J., Vanderwende, L., 2016. CaTeRS: causal and temporal relation scheme for semantic annotation of event structures. In: Proceedings of the Fourth Workshop on Events, vols. 51–61.
Mostafazadeh, N., Kalyanpur, A., Moon, L., Buchanan, D., Berkowitz, L., Biran, O., Chu-Carroll, J., 2020. GLUCOSE. GeneraLized and COntextualized Story Explanations.
Nguyen, T.H., Grishman, R., 2015. Event detection and domain adaptation with convolutional neural networks. In: ACL, Association for Computational Linguistics. URL. http://www.aclweb.org/anthology/P15-2060.
Nguyen, T.H., Grishman, R., 2016. Modeling skip-grams for event detection with convolutional neural networks. In: Proceedings of the 2016 Conference on Empirical Methods in Natural Language Processing. Association for Computational Linguistics, Austin, Texas, pp. 886–891. https://doi.org/10.18653/v1/D16-1085. URL. https://www.aclweb.org/anthology/D16-1085.
Nguyen, T.M., Nguyen, T.H.. One for all: neural joint modeling of entities and events, CoRR abs/1812.00195. URL. http://arxiv.org/abs/1812.00195.
Nguyen, T.H., Cho, K., Grishman, R., 2016. Joint event extraction via recurrent neural networks. In: NAACL, Association for Computational Linguistics. URL. http://www.aclweb.org/anthology/N16-1034.
Ning, Q., Feng, Z., Roth, D., 2017. A Structured Learning Approach to Temporal Relation Extraction. EMNLP.
Ning, Q., Wu, H., Roth, D., 2018. A multi-Axis Annotation scheme for event temporal relations. In: Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers). Association for Computational Linguistics, Melbourne, Australia, pp. 1318–1328. https://doi.org/10.18653/v1/P18-1122. URL. https://www.aclweb.org/anthology/P18-1122.
Ning, Q., Subramanian, S., Roth, D., 2019. An improved neural baseline for temporal relation extraction. In: Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing (EMNLP-IJCNLP). Association for Computational Linguistics, Hong Kong, China, pp. 6203–6209. https://doi.org/10.18653/v1/D19-1642. URL. https://www.aclweb.org/anthology/D19-1642.
Orr, W., Tadepalli, P., Fern, X., 2018. Event detection with neural networks: a rigorous empirical evaluation. In: Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing. Association for Computational Linguistics, Brussels, Belgium, pp. 999–1004. https://doi.org/10.18653/v1/D18-1122. URL. https://www.aclweb.org/anthology/D18-1122.
Patwardhan, S., Riloff, E., 2009. A unified model of phrasal and sentential evidence for information extraction. In: Proceedings of the 2009 Conference on Empirical Methods in Natural Language Processing, vols. 151–160.
Pennington, J., Socher, R., Manning, C.D., 2014. GloVe: global vectors for word representation. Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing 1532–1543. https://doi.org/10.3115/v1/D14-1162. URL. http://www.aclweb.org/anthology/D14-1162, 2014.
Peters, M., Neumann, M., Iyyer, M., Gardner, M., Clark, C., Lee, K., Zettlemoyer, L., 2018. Deep contextualized word representations. In: Proceedings of the 2018 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, vol. 1. Association for Computational Linguistics, New Orleans, Louisiana, pp. 2227–2237. https://doi.org/10.18653/v1/N18-1202 (Long Papers). https://www.aclweb.org/anthology/N18-1202. URL.
Pradhan, S., Ramshaw, L., Weischedel, R., MacBride, J., Micciulla, L., 2007. Unrestricted coreference: identifying entities and events in OntoNotes. In: International Conference on Semantic Computing. ICSC, pp. 446–453, 2007.
Pustejovsky, J., Castaño, J.M., Ingria, R., Saurí, R., Gaizauskas, R., Setzer, A., Katz, G., Radev, D.R., 2003. TimeML: robust specification of event and temporal expressions in text. In: New Directions in Question Answering.
J. Pustejovsky, P. Hanks, R. Saurí, A. See, R. Gaizauskas, A. Setzer, D. Radev, B. Sundheim, D. Day, L. Ferro, M. Lazo, The TimeBank corpus, Proceedings of Corpus Linguistics .
Rashkin, H., Sap, M., Allaway, E., Smith, N.A., Choi, Y., 2018. Event2Mind: commonsense inference on events, intents, and reactions. In: ACL, ACL, Melbourne, Australia, vols. 463–473.
Riaz, M., Girju, R., 2010. Another look at causality: discovering scenario-specific contingency relationships with no supervision. In: 2010 IEEE Fourth International Conference on Semantic Computing, IEEE, vols. 361–368.
Riaz, M., Girju, R., 2013. Toward a better understanding of causality between verbal events: extraction and analysis of the causal power of verb-verb associations. Proceedings of the SIGDIAL 2013 Conference 21–30.
Riaz, M., Girju, R., 2014a. In-depth exploitation of noun and verb semantics to identify causation in verb-noun pairs. In: Proceedings of the 15th Annual Meeting of the Special Interest Group on Discourse and Dialogue (SIGDIAL), vols. 161–170.
Riaz, M., Girju, R., 2014b. Recognizing causality in verb-noun pairs via noun and verb semantics. In: Proceedings of the EACL 2014 Workshop on Computational Approaches to Causality in Language (CAtoCL), vols. 48–57.
Ross, H., Cai, J., Min, B., 2020. Exploring Contextualized Neural Language Models for Temporal Dependency Parsing.
M. Sap, R. L. Bras, E. Allaway, C. Bhagavatula, N. Lourie, H. Rashkin, B. Roof, N. A. Smith, Y. Choi, ATOMIC: an Atlas of Machine Commonsense for If-Then Reasoning, ArXiv abs/1811.00146.
M. Sap, R. LeBras, E. Allaway, C. Bhagavatula, N. Lourie, H. Rashkin, B. Roof, N. A. Smith, Y. Choi, ATOMIC: an Atlas of Machine Commonsense for If-Then Reasoning, CoRR abs/1811.00146.
Seventh message understanding conference (MUC-7). In: Proceedings of a Conference Held in Fairfax, Virginia, April 29 - May 1, 1998, 1998. URL. https://www.aclweb.org/anthology/M98-1000.
Sha, L., Qian, F., Chang, B., Sui, Z., 2018. Jointly Extracting Event Triggers and Arguments by Dependency-Bridge RNN and Tensor-Based Argument Interaction. AAAI.
Sixth message understanding conference (MUC-6). In: Proceedings of a Conference Held in Columbia, Maryland, November 6-8, 1995, 1995. URL. https://www.aclweb.org/anthology/M95-1000.
Song, Z., Bies, A., Strassel, S., Riese, T., Mott, J., Ellis, J., Wright, J., Kulick, S., Ryant, N., Ma, X., 2015. From light to rich ERE: annotation of entities, relations, and events. In: Proceedings of the the 3rd Workshop on EVENTS: Definition, Detection, Coreference, and Representation. Association for Computational Linguistics, Denver, Colorado, pp. 89–98. https://doi.org/10.3115/v1/W15-0812. URL. https://www.aclweb.org/anthology/W15-0812.
Speer, R., Chin, J., Havasi, C., 2017. ConceptNet 5.5: an open multilingual graph of general knowledge. In: Proceedings of the Thirty-First AAAI Conference on Artificial Intelligence, AAAI'17. AAAI Press, 4444–4451.
Strötgen, J., Gertz, M., 2010. HeidelTime: High Quality Rule-Based Extraction and Normalization of Temporal Expressions. SemEval@ACL.
Sundheim, B.M., 1992. Overview of the Fourth Message Understanding Evaluation and Conference, Tech. Rep. NAVAL COMMAND CONTROL AND OCEAN SURVEILLANCE CENTER RDT AND E DIV, SAN DIEGO CA.
Tissot, H., Gorrell, G., Roberts, A., Derczynski, L., Fabro, M.D.D., 2015. UFPRSheffield: Contrasting Rule-Based and Support Vector Machine Approaches to Time Expression Identification in Clinical TempEval. SemEval@NAACL-HLT.
Tourille, J., Ferret, O., Névéol, A., Tannier, X., 2017a. Neural Architecture for Temporal Relation Extraction: A Bi-LSTM Approach for Detecting Narrative Containers. ACL.
Tourille, J., Ferret, O., Tannier, X., Névéol, A., 2017b. LIMSI-COT at SemEval-2017 task 12: neural architecture for temporal information extraction from clinical narratives. In: Proceedings of the 11th International Workshop on Semantic Evaluation (SemEval-2017). Association for Computational Linguistics, Vancouver, Canada, pp. 597–602. https://doi.org/10.18653/v1/S17-2098. URL. https://www.aclweb.org/anthology/S17-2098.
UzZaman, N., Llorens, H., Derczynski, L., Allen, J., Verhagen, M., Pustejovsky, J., 2013. SemEval-2013 task 1: TempEval-3: evaluating time expressions, events, and temporal relations. In: Second Joint Conference on Lexical and Computational Semantics (*SEM), Volume 2: Proceedings of the Seventh International Workshop on Semantic Evaluation (SemEval 2013), vols. 1–9. Association for Computational Linguistics, Atlanta, Georgia, USA. https://www.aclweb.org/anthology/S13-2001. URL.
Verhagen, M., Pustejovsky, J., 2008. Temporal Processing with the TARSQI Toolkit, COLING '08. Association for Computational Linguistics, USA, pp. 189–192.
Verhagen, M., Gaizauskas, R., Schilder, F., Hepple, M., Katz, G., Pustejovsky, J., 2007a. SemEval-2007 task 15: TempEval temporal relation identification. In: Proceedings of the Fourth International Workshop on Semantic Evaluations (SemEval-2007). Association for Computational Linguistics, Prague, Czech Republic, pp. 75–80. URL. https://www.aclweb.org/anthology/S07-1014.
Verhagen, M., Gaizauskas, R., Schilder, F., Hepple, M., Katz, G., Pustejovsky, J., 2007b. SemEval-2007 Task 15: TempEval Temporal Relation Identification, SemEval '07, vols. 75–80. Association for Computational Linguistics, USA.
Verhagen, M., Saurí, R., Caselli, T., Pustejovsky, J., 2010. SemEval-2010 task 13: TempEval-2. In: Proceedings of the 5th International Workshop on Semantic Evaluation. Association for Computational Linguistics, Uppsala, Sweden, pp. 57–62. URL. https://www.aclweb.org/anthology/S10-1010.
Wadden, D., Wennberg, U., Luan, Y., Hajishirzi, H., 2019. Entity, relation, and event extraction with contextualized span representations. In: Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing (EMNLP-IJCNLP). Association for Computational Linguistics, Hong Kong, China, pp. 5784–5789. https://doi.org/10.18653/v1/D19-1585. URL. https://www.aclweb.org/anthology/D19-1585.
Wang, X., Han, X., Liu, Z., Sun, M., Li, P., 2019a. Adversarial training for weakly supervised event detection. In: Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, vol. 1, pp. 998–1008 (Long and Short Papers).
Wang, X., Wang, Z., Han, X., Liu, Z., Li, J., Li, P., Sun, M., Zhou, J., Ren, X., 2019b. HMEAE: hierarchical modular event argument extraction. In: Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing (EMNLP-IJCNLP). Association for Computational Linguistics, Hong Kong, China, pp. 5777–5783. https://doi.org/10.18653/v1/D19-1584. URL. https://www.aclweb.org/anthology/D19-1584.

Wei, S., Korostil, I., Nothman, J., Hachey, B., 2017. English event detection with translated language features. In: Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics, vol. 2, pp. 293–298. Short Papers.

Yan, H., Jin, X., Meng, X., Guo, J., Cheng, X., 2019. Event detection with multi-order graph convolution and aggregated attention. In: Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing (EMNLP-IJCNLP). Association for Computational Linguistics, Hong Kong, China, pp. 5766–5770. https://doi.org/10.18653/v1/D19-1582. URL. https://www.aclweb.org/anthology/D19-1582.

Yang, B., Cardie, C., Frazier, P., 2015. A hierarchical distance-dependent bayesian model for event coreference resolution. Transactions of the Association for Computational Linguistics 3, 517–528.

Yang, H., Chen, Y., Liu, K., Xiao, Y., Zhao, J., 2018. DCFEE: a document-level Chinese financial event extraction system based on automatically labeled training data. In: Proceedings of ACL 2018, System Demonstrations, vols. 50–55.

Yang, S., Feng, D., Qiao, L., Kan, Z., Li, D., 2019. Exploring pre-trained language models for event extraction and generation. In: Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics. Association for Computational Linguistics, Florence, Italy, pp. 5284–5294. https://doi.org/10.18653/v1/P19-1522. URL. https://www.aclweb.org/anthology/P19-1522.

Yoshikawa, K., Riedel, S., Asahara, M., Matsumoto, Y., 2009. Jointly Identifying Temporal Relations with Markov Logic. ACL/IJCNLP.

Zavarella, V., Tanev, H., 2013. FSS-TimEx for TempEval-3: Extracting Temporal Information from Text. SemEval@NAACL-HLT.

Y. Zeng, Y. Feng, R. Ma, Z. Wang, R. Yan, C. Shi, D. Zhao, Scale up Event Extraction Learning via Automatic Training Data Generation, arXiv preprint arXiv:1712.03665 .

Zhang, T., Ji, H.. Event extraction with generative adversarial imitation learning, CoRR abs/1804.07881. URL. http://arxiv.org/abs/1804.07881.

Y. Zhang, N. Xue, Structured Interpretation of Temporal Relations, ArXiv abs/1808.07599.

Zhang, Y., Xue, N., 2018. Structured interpretation of temporal relations. In: Proceedings of the Eleventh International Conference on Language Resources and Evaluation (LREC 2018), European Language Resources Association (ELRA), Miyazaki, Japan. URL. https://www.aclweb.org/anthology/L18-1490.

Zhang, Y., Xu, G., Wang, Y., Liang, X., Wang, L., Huang, T., 2019. Empower event detection with bi-directional neural language model. Knowl. Base Syst. 167, 87–97. https://doi.org/10.1016/j.knosys.2019.01.008. ISSN 0950-7051. http://www.sciencedirect.com/science/article/pii/S0950705119300097.

Zhang, H., Liu, X., Pan, H., Song, Y., ki Leung, C.W., 2020. ASER: a large-scale eventuality knowledge graph. In: Proceedings of The Web Conference 2020 201–211. https://doi.org/10.1145/3366423.3380107.

Zhao, Y., Jin, X., Wang, Y., Cheng, X., 2018. Document embedding enhanced event detection with hierarchical and supervised attention. In: Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics, vol. 2. Association for Computational Linguistics, Melbourne, Australia, pp. 414–419. https://doi.org/10.18653/v1/P18-2066. Short Papers. https://www.aclweb.org/anthology/P18-2066. URL.

S. Zheng, W. Cao, W. Xu, J. Bian, Doc2EDAG: an End-To-End Document-Level Framework for Chinese Financial Event Extraction, arXiv preprint arXiv:1904.07535 .

Zuo, X., Chen, Y., Liu, K., Zhao, J., 2020a. Towards Causal Explanation Detection with Pyramid Salient-Aware Network.

Zuo, X., Chen, Y., Liu, K., Zhao, J., 2020b. KnowDis: knowledge enhanced data augmentation for event causality detection via distant supervision. In: Proceedings of the 28th International Conference on Computational Linguistics, International Committee on Computational Linguistics, Barcelona, Spain (Online), 1544–1550, URL. https://www.aclweb.org/anthology/2020.coling-main.135.

X. Zuo, Y. Chen, K. Liu, J. Zhao, Event co-reference resolution via a multi-loss neural network without using argument information, Sci. China Inf. Sci. 62.