Type Information Utilized Event Detection via Multi-Channel GNNs in Electrical Power Systems

QIAN LI and JIANXIN LI, School of Computer Science and Engineering, Beihang University, China
LIHONG WANG, National Computer Network Emergency Response Technical Team/Coordination Center of China, China
CHENG JI, School of Computer Science and Engineering, Beihang University, China
YIMING HEI, School of Cyber Science and Technology, Beihang University, China
JIAWEI SHENG, Institute of Information Engineering, Chinese Academy of Sciences, China
QINGYUN SUN, School of Computer Science and Engineering, Beihang University, China
SHAN XUE, School of Computing, Macquarie University, Australia
PENGTAO XIE, Department of Electrical and Computer Engineering, UC San Diego, United States

Event detection in power systems aims to identify triggers and event types, which helps relevant personnel respond to emergencies promptly and facilitates the optimization of power supply strategies. However, the limited length of short electrical record texts causes severe information sparsity, and numerous domain-specific terminologies of power systems makes it difficult to transfer knowledge from language models pre-trained on general-domain texts. Traditional event detection approaches primarily focus on the general domain and ignore these two problems in the power system domain. To address the above issues, we propose a **Multi-Channel graph neural network utilizing Type information for Event Detection** in power systems, named **MC-TED**, leveraging a semantic channel and a topological channel to enrich information interaction from short texts. Concretely, the semantic channel refines textual representations with semantic similarity, building the semantic information interaction among potential event-related words. The topological channel generates a relation-type-aware graph modeling word dependencies, and a word-type-aware graph integrating part-of-speech tags. To further reduce errors worsened by professional terminologies in type analysis, a type learning mechanism is designed for updating the representations of both the word type and relation type in the topological channel. In this way, the information sparsity and professional term occurrence problems can be alleviated by enabling interaction between topological and semantic information. Furthermore,

The authors of this paper were supported by the NSFC through grant 6187202, 62106059 and the Academic Excellence Foundation of Beihang University for PhD Students.
Authors' addresses: Q. Li, J. Li (corresponding author), C. Ji, and Q. Sun, School of Computer Science and Engineering, Beihang University, No. 37 Xueyuan Road, Haidian, Beijing, Beijing, 100083, China; emails: {liqian, lijx, jicheng, sunqy}@act.buaa.edu.cn; L. Wang, National Computer Network Emergency Response Technical Team/Coordination Center of China, No. 3, Yumin Road, Chaoyang, Beijing, Beijing, 100029, China; email: wlh@isc.org.cn; Y. Hei, School of Cyber Science and Technology, Beihang University, No. 37 Xueyuan Road, Haidian, Beijing, Beijing, 100083, China; email: black@buaa.edu.cn; J. Sheng, Institute of Information Engineering, Chinese Academy of Sciences, 89 Minzhuang Road, Haidian, Beijing, Beijing, 100093, China; email: shengjiawei@iie.ac.cn; S. Xue, School of Computing, Macquarie University; Ryde, St Leonards, New South Walet, Sydney, New South Wales, 2109, Australia; email: emma.xue@data61.csiro.au; P. Xie, Department of Electrical and Computer Engineering, UC San Diego, 9500 Gilman Dr, La Jolla, CA 92093-0001, United States, La Jolla, California, 92093-0001, United States; email: p1xie@eng.ucsd.edu.

1559-1131/2023/05-ART20 $15.00
https://doi.org/10.1145/3577031

to address the lack of labeled data in power systems, we built a Chinese event detection dataset based on electrical Power Event texts, named **PoE**. In experiments, our model achieves compelling results not only on the PoE dataset, but on general-domain event detection datasets including ACE 2005 and MAVEN.

CCS Concepts: • **Computing methodologies → Information extraction;**

Additional Key Words and Phrases: Event detection, power systems, multi-channel, topological channel, semantic channel

ACM Reference format:
Qian Li, Jianxin Li, Lihong Wang, Cheng Ji, Yiming Hei, Jiawei Sheng, Qingyun Sun, Shan Xue, and Pengtao Xie. 2023. Type Information Utilized Event Detection via Multi-Channel GNNs in Electrical Power Systems. *ACM Trans. Web* 17, 3, Article 20 (May 2023), 26 pages.
https://doi.org/10.1145/3577031

1 INTRODUCTION

Electrical power systems [1, 10] provide the whole electricity supply across the country to ensure the basic needs of people's livelihood. The application of automated monitoring and analysis on the device-related event reports in the electrical power systems can improve management efficiency, save labor cost, and maintain power supplement stability [17, 47]. One critical step for this is **Event Detection (ED)** [12, 32, 35] and specifically, the electrical event detection. The electrical event detection task refers to identifying triggers from power records and correctly classifying the device-related events. Through precisely detecting electrical events, we can swiftly obtain essential information, e.g., latent device risks, and help execute subsequent analysis. Thus, we focus on the event detection task in the electrical power systems.

There are two main problems for electrical event detection: short texts and professional terminology. Due to the advantage of short texts being easy to read and spread, device events are often described in short sentences for efficient communication in the electrical power systems. However, the descriptions of short texts are inadequate, sparse, and irregular, which imposes higher requirements on feature learning. Furthermore, electricity texts contain many terminologies, such as "high voltage" and "laying path" (shown in Figure 1), which are rarely mentioned in general-domain texts.

Event detection methods based on graph representation learning learn semantic structural information in text by composing graphs. Graph representation learning aims to map nodes, edges, and even the entire graph to a low-dimensional space. It maintains the structure and properties of the graph itself, which in turn serves the downstream graph data mining task. Homogeneous graphs can only characterize data with simple structure and semantic homogeneity, while textual composition graphs will contain multiple types of nodes and relations with rich semantic information. Heterogeneous graphs can fully portray such responsible nonlinear relationships with stronger representational and inference capabilities. Each type of node in a heterogeneous graph contains a variety of rich semantic information and complex interactions between nodes. Each type of semantic information also reflects only one aspect of the node's characteristics. In order to learn a more comprehensive node representation, the model needs to fuse multiple aspects of information to enhance the node representation.

Existing methods for power system event detection fail to address the above two problems. In recent years, researchers have proposed many methods for event detection and achieved promising results [2, 33, 63]. However, most of these methods focus on dealing with the long-tail distribution [57] of data and few-shot data [4, 63], ignoring the above-mentioned problems in power systems. Existing methods can be categorized into two main paradigms: BERT-based methods [55] and graph-based methods [5, 36, 60]. (1) BERT-based models perform well in general domain event detection due to the large amount of data available for pre-training. However, in the domain of

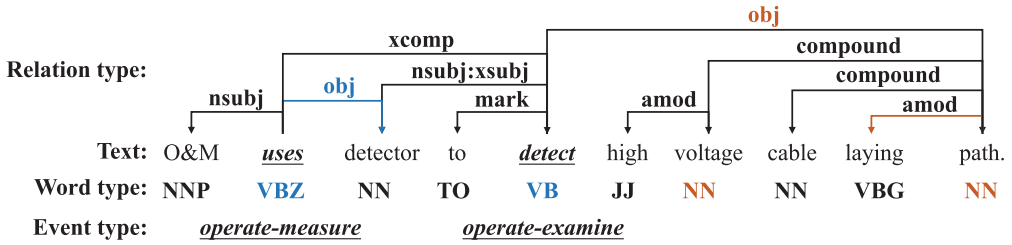


Fig. 1. An example in our dataset. Word type means Part-of-Speech, and relation type means dependencies between words with relation. The triggers are "uses" and "detect".

power systems, some combinations of common words can have different meanings. For example, "laying path" means "put something on a path" in generation domain, but it means "laying a route of cable" in a power system. The distinction between general and electrical domains causes severe semantic confusion. This problem can be potentially resolved by pre-training models on labeled data in the power field, which takes time and experience. (2) Graph-based models integrate various methods to design a graph with texts [5, 36, 38]. As shown in Figure 1, we give an example of words, and relation is obtained by the Stanford CoreNLP tool.[1] It can bring a variety of interactive information among words. Nevertheless, graphs constructed from short electricity texts contain few nodes and relations, which hinders the effectiveness of graph representation learning.

To solve these problems, we propose a **M**ulti-**C**hannel GNN Utilizing **T**ype Information for **E**vent **D**etection (**MC-TED**). It contains a semantic channel and a topological channel to fuse multiple sources of information to learn better node representations. The semantic channel reflects the similarity of words through adding relation with lower cosine distance. The topological channel constructs a relation-type-aware graph and a word-type-aware graph to learn more comprehensive node representations. They consider the word type and relation type to portray nonlinear relationships with stronger representative and inference capabilities fully. Each type of word and relation contains a variety of rich semantic information and complex interaction between nodes. Furthermore, we use the Stanford CoreNLP tool to initialize the word and relation type, which is updated in training to mitigate the errors caused by Stanford CoreNLP. To address the lack of labeled data in the domain of the power systems, we build a **Po**wer **E**vent detection dataset (**PoE**). Experimental results show that our approach can achieve excellent performance on both the PoE dataset and public event detection datasets, including the ACE 2005 [7] and MAVEN [57] datasets. We also demonstrate that our word type and relation type learning strategies are particularly beneficial for the word-type-aware graph and relation-type-aware graph.

This paper makes the following contributions.

- We design a novel **multi-channel event detection model (MC-TED[2])** utilizing word and relation type information for the power systems from semantic and topological channels with type utilized multi-channel GNNs to capture richer information.
- Topological channel designs a relation-type-aware graph to utilize the relation type and a node-type-aware graph considering the node type. The two types can be learnable to mitigate the errors caused by the CoreNLP syntactic analysis tools.
- We construct a Chinese event detection dataset – PoE, in the domain of power systems, derived from power system records. Our model achieves state-of-the-art performance on the PoE as well as on the general-domain datasets including the ACE 2005 and MAVEN.

[1] http://corenlp.run/.
[2] The source code is available at https://anonymous.4open.science/r/MC-TED-3C77.

2 RELATED WORK

The electrical power systems contain numerous event text, but engineers only focus on the core events regularly. It requires models to capture certain types of events, which is a schema-based event detection task [5, 28, 36, 44, 60]. Thus, event detection in power systems is to find triggers from the text and identify the event type corresponding to the triggers.

2.1 Event Detection in Power Systems

Large-scale new energy is integrated into the power grid system, which introduces strong uncertainty to the power system [18]. It brings significant challenges to detect events in the modern power systems. Traditional methods convert it to a constrained optimization problem [51], or non-convex optimization learning algorithms [65], and the like. To detect the abnormal events in the power systems using reduced phasor measurement units data, Liu et al. [34] propose the unequal-interval reduction method with local outlier factor. Zhou et al. [65] propose a hidden structure semi-supervised machine for the event detection task. It incorporates information from labeled data and unlabeled data for the power distribution network. HS^3M [65] combines unlabeled data and partly labeled data in a large margin learning objective, which is a classical method for event detection in the power systems. It incorporates partial knowledge for event detection. SS-LOF [34] is an event detection algorithm in the power systems using reduced data and local outlier factor for detecting the events in power systems. It is employed to determine the region of the event source. There are some graph-based methods for spatio-temporal solar irradiance forecasting [21], and short-term wind speed prediction [22], and so on. To our best knowledge, there are few graph representation learning methods for the event detection task [20, 40].

2.2 Graph-Based Event Detection

In recent years, there are two main branches: BERT based models [15, 29, 55] and graph based models [25, 26, 41, 42, 46]. The BERT-based methods learn a better context representation through self-attention mechanisms, and large-scale pre-train [6]. However, for the text with professional terms and short text, there are no apparent advantages. The graph based methods commit to constructing syntactic dependency graph [5, 39, 60], and building event correlations [9, 59], and so on. There have been previous approaches to incorporate syntactic information into the **Graph Convolutional Network (GCN)** [60], often ignoring dependency tag information to convey rich and valuable language knowledge to ED. Nguyen et al. [39] propose a GCN-based event detection model over syntactic dependency trees and entity mention-guided pooling. It operates a pooling over the graph-based convolution vectors of the current word and the entity mentioned in the sentences. MOGANED [60] uses a dependency tree-based GCN with aggregative attention to explicitly model and aggregate multi-order syntactic representations in sentences. It models multi-order representations via **graph attention network (GAT)**. It utilizes both first-order syntactic graphs and high-order syntactic graphs to explicitly model multi-order representations of candidate triggers. These methods for ED fail to exploit the overall contextual importance of the words, which can be obtained via the dependency tree, to boost the performance. Lai et al. [27] propose a novel gating mechanism to filter noisy information in the hidden vectors of the GCN models for ED based on the information from the candidate triggers. They also introduce a mechanism to achieve the contextual diversity for the gates and the importance of score consistency for the graphs and models in ED. They demonstrate how gating mechanisms, gate diversity, and graph structure can integrate syntactic information and improve the hidden vectors for ED models. GatedGCN [26] is a gating mechanism to filter noisy information in the hidden vectors of the GCN models for ED. Lai et al. [25] present how transferring open-domain knowledge from word sense disambiguation and regulating representation based on pruned dependency graphs can improve

few-shot learning on large-scale datasets. They propose a training signal derived from dependency graphs to regularize the representation learning.

There have been previous approaches to incorporate syntactic information into the GCN, often ignoring dependency tag information to convey rich and valuable language knowledge to ED. Cui et al. [5] propose an edge-enhanced graph convolutional network, which uses both syntactic structures and dependent label information to perform ED. For the event correlations-based graph for event detection task, Xie et al. [59] formulate event detection task as a graph parsing problem, to overcome the inherent issues with existing trigger classification-based models. It can explicitly model the multiple event correlations and naturally utilize the rich information conveyed by event type and sub-type. Dutta et al. [9] propose a framework for incorporating both dependencies and their labels using the graph transformer networks [61]. However, these methods cannot be completely applied to electrical power systems due to the characteristics of short text and professional terms.

2.3 Topological Graph Learning

Graph neural network [2, 48, 64] can fully describe the complex nonlinear structure on the graph and has a more vigorous representation and reasoning ability [58, 64]. Through topological graph learning and utilizing the heterogeneity, the high-level information can be mined to overcome the short text [13, 43]. Graph representation learning and graph neural networks have gradually become the hot field of graph data mining. Traditional graph representation algorithms are mainly shallow models, such as Metapath2vec [8] and HERec [50] based on the random walk and Skip-Gram model. They cannot effectively capture complex nonlinear structures on the graph. Graph neural network aims to map nodes, edges, or even the whole graph to low-dimensional vector space and keep the structure and properties of the graph itself. It can fully describe the complex nonlinear structure on the graph and has a more vigorous representation and reasoning ability. The spatial embedding fusion method based on graph convolution neural network mainly realizes the aggregation and updating of node information in a multi-layer network. In recent years, multi-graph [16, 19, 31, 37] and heterogeneous graph [11, 13, 14, 52, 56, 62] are effective methods to mining high-order and hidden information. Khan et al. [19] propose a multi-graph convolutional network model to solve the embedding problem of the existing multivariate relational network model. Ma et al. [37] propose a graph convolution model of multi-dimensional networks to capture more abundant node-level information. Considering that most of the existing methods are over-parameterized and limited to learning node representation, Vashishth et al. [53] propose a **Composition-based multi-relational Graph Convolutional Network (CompGCN)** in which nodes and relations are embedded into the relational graph. Huang et al. [16] propose the framework of a multi-graph convolutional network by developing new convolution operators on multi-graph. To solve the problem of short text classification, Hu et al. [13] convert text into a heterogeneous graph and design a dual-level heterogeneous graph attention network for learning text representation.

Heterogeneous graph neural networks can fully mine the complex structure and rich semantics to learn node representation and improve the performance of subsequent tasks. PTE [52] decomposes the text corpus into multiple heterogeneous text graphs and realizes the representation learning of text graphs through the joint decomposition of multiple graphs. HAN [56] is a heterogeneous graph neural network integrating hierarchical attention mechanism, which learns node representation by weighted fusion from node level and semantic level, respectively. MEIRec [11] is a classical heterogeneous graph neural network, which can learn node representation by integrating rich semantic information provided by multi-element paths. HGT [14] learns node representation based on heterogeneous graph neural network by aggregating heterogeneous relation triples.

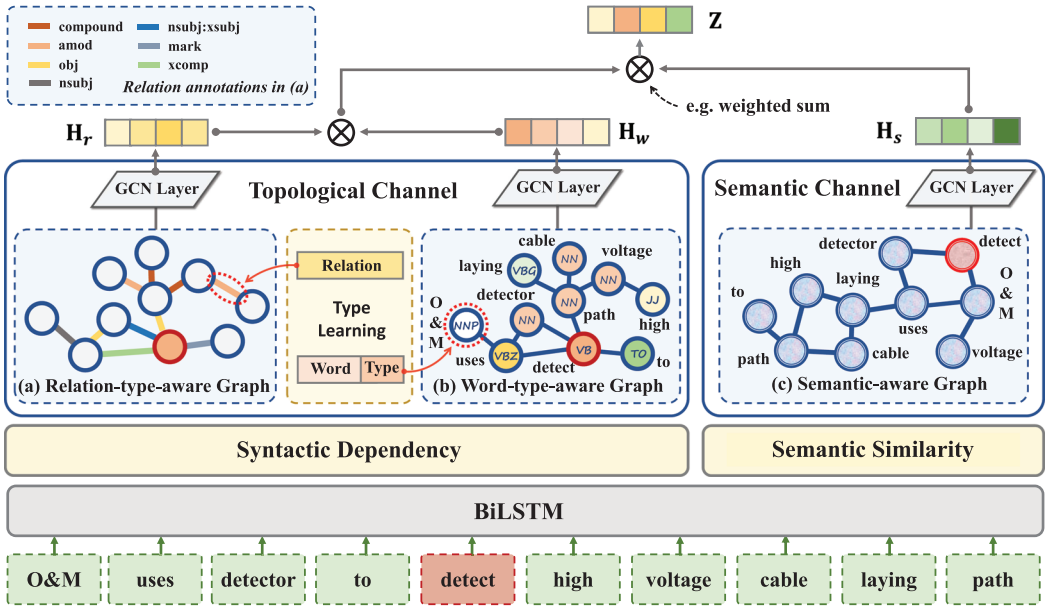


Fig. 2. Our multi-channel event detection framework, MC-TED. It contains a topological channel and semantic channel. The topological channel includes two heterogeneous graphs utilizing word dependencies and part-of-speech tags. A relation-type-ware graph updates relation type, and a node-type-aware graph updates node type. The semantic channel refines the textual representations with semantic similarity.

3 FRAMEWORK

We build a **multi-channel event detection model for the power systems (MC-TED)**, as shown in Figure 2. The inputs of our MC-TED model are the words obtained by the CoreNLP syntactic dependency analysis tool. All words are fed into a **Bidirectional Long Short-Term Memory (BiLSTM)** [49] to get their initial representations. Then, the MC-TED model learns from the topological and semantic channels through the type utilized multi-channel GNN:

- In the topological channel, we construct a syntactic graph based on the syntactic dependency analysis results generated by the Stanford CoreNLP tool. To consider the heterogeneity of relations and words, we further design a relation-type-aware graph and a word-type-aware graph, considering the relation types between words and the word types, respectively. In the two graphs, the representations of relation types and node types are both learnable.
- In the semantic channel, we build a semantic-aware graph according to the semantic distance of the current word representations. The structure of the semantic-aware graph dynamically changes during the training process of the MC-TED. It reflects the similarity of words through adding relation with lower cosine distance.

Through the topological and semantic channels, the MC-TED integrates multiple aspects of multiple sources of information (e.g., node and context) to learn better word representations through the type utilized multi-channel GNN. The type utilized multi-channel GNN considers the word type and relation type to portray nonlinear relationships with stronger representative and inference capabilities fully. At the final step of the MC-TED, the weighted average of the two channels is normalized to obtain the final representation of each word. It then detects triggers and event types in each sentence through a weighted mechanism to obtain the optimal weights.

3.1 Word Feature Representation

We use the CoreNLP syntactic analysis tool to divide a sentence T into words $[w_1, w_2, \ldots, w_N]$, where N is the number of words in T. The word representation can be contained by the GloVe 100-dimensional word vectors [30, 45]. For our Chinese power event detection dataset PoE, a sentence needs to be firstly segmented through word segmentation tools. The word with multiple characters is initialized randomly to keep dimensions consistent. Then the word embeddings are fed into a BiLSTM to obtain the contextual representations $\mathbf{H}^0$.

3.2 Type Utilized Multi-Channel GNNs

Graph neural network [58] can mine the complex structure and rich semantics of the graph to learn node representation and improve the performance of subsequent tasks. In event detection, by modeling multiple syntactic dependency associations between words, GNN-based models can aggregate diverse interaction information to understand the multi-semantic expression of words.

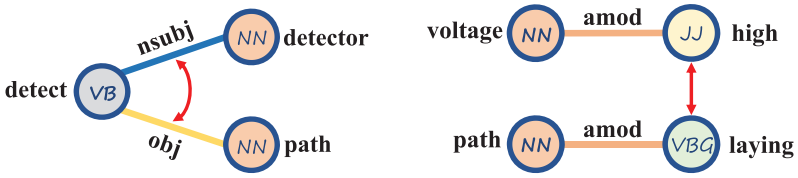
To harness the great power of GNN, we design a multi-channel graph neural network for event detection through semantic and topological channels. As illustrated in Figure 2, it contains: (1) **topological channel**, where a relation-type-aware graph (Figure 2(a)) and a word-type-aware graph (Figure 2(b)) are generated to achieve a more detailed and comprehensive description considering the semantics of word and relation types, and (2) **semantic channel**, where a semantic-aware graph is constructed based on the semantic similarity to refine the textual representations for the event detection. With the two channels, MC-TED integrates multiple aspects of information to learn a more comprehensive word representation.

3.3 Topological Channel

The text contains rich syntactic (topological) information, and the representations of words change in different contexts. To improve comprehensively word representation, we use the Stanford CoreNLP syntactic analysis tool to construct a syntactic graph where the nodes denote the words in a sentence and edges denote the dependencies between words, with the part-of-speech tags (e.g., noun and verb) as node types and the dependency classes (e.g., nominal subject and direct object) as relation types. Formally, the syntactic graph is a form of heterogeneous graph $\mathcal{G} = (\mathcal{V}, \mathcal{E})$ with a word type mapping function $\psi_w : \mathcal{V} \rightarrow \mathcal{A}$ and a relation type mapping function $\psi_r : \mathcal{E} \rightarrow \mathcal{R}$, where $\mathcal{V}$ is the set of words and $\mathcal{E}$ is the set of relations, $\mathcal{A}$ is the set of word types and $\mathcal{R}$ is the set of relation types, and $|\mathcal{A}| + |\mathcal{R}| > 2$.

Insights on different type information. Different from the common heterogeneous graphs, the syntactic graph contains two kinds of type information that cannot be substituted or deduced from each other, the word type and the relation type. It is noticed that there are two phenomenons in the syntactic graph: (1) some relations between two words with the same type belong to different types (i.e., "$A_1 \xrightarrow{R_1} A_2$ and $A_1 \xrightarrow{R_2} A_2$"), and (2) relations with the same type may connect two words whose types are different (i.e., "$A_1 \xrightarrow{R} A_2$ and $A_1 \xrightarrow{R} A_3$"). Taking the text in Figure 1 as an example, as shown in Figure 3, the relation types between the words with type "VB" and type "NN" are from {"nsubj", "obj"}, and the relation with type "amod" connects two words with types {"NN", "JJ"} and {"NN", "VBG"}. Therefore, only obtaining relation type information or word type information is inadequate, as the two types cannot be substituted or deduced.

To address the above issue, we divide the syntactic graph into a relation-type-aware graph and a word-type-aware graph to utilize different type information for the type utilized multi-channel GNN. Specifically, (1) the **relation-type-aware graph** has only one word type (i.e., $|\mathcal{A}| = 1$) and multiple relation types (i.e., $|\mathcal{R}| > 1$) to learn the topological information under the relation type



(a) Different relation types between the same node type pair.

(b) Different word types connected by the same relation type.

Fig. 3. The heterogeneity of relations and words. For one sentence, there are both multi-relationship types for the same node type as the left, and multi-node types for the same relation type as the right.

constraint, and (2) the **word-type-aware graph** has multiple word types (i.e., $|\mathcal{A}| > 1$) and only one relation type (i.e., $|\mathcal{R}| = 1$) for word type learning.

Furthermore, the type information in syntactic is obtained by the CoreNLP syntactic analysis tool, which leads to type fault in some cases, especially for the electrical power system with numerous professional terminology (see an example in Section 5.7). We design a type learning mechanism in both the relation-type-aware graph and word-type-aware graph for the type utilized multi-channel GNNs to lessen this error caused by short text and professional terms.

Relation-type-aware graph. To consider the heterogeneity of relations, we design a relation-type-aware graph $\mathcal{G}_r = (\mathcal{V}_r, \mathcal{E}_r)$ with the mapping functions $\psi_{r,v} : \mathcal{V}_r \rightarrow \mathcal{A}_r$ and $\psi_{r,r} : \mathcal{E}_r \rightarrow \mathcal{R}_r$, where $|\mathcal{A}_r| = 1$ and $|\mathcal{R}_r| > 1$. The relation-type-aware graph does not introduce node type information that means all nodes are the same type. However, the relation types of this graph are different, which are obtained by the CoreNLP syntactic analysis tool. In this way, the relation-type-aware graph only focus on the relation types. In order to utilize the information of relation types, we propose to learn the corresponding relation type vectors as the edge representations:

$$\mathbf{E}_{ij} \leftarrow \mathbf{R}_{\psi_{r,r}(e_{ij})}, i, j \in [1, N], \tag{1}$$

where $\mathbf{E} \in \mathbb{R}^{N \times N \times d_r}$ is the representation of the edge e, and $\mathbf{R} \in \mathbb{R}^{|\mathcal{R}_r| \times d_r}$ is the representation of the edge type which is a learnable tensor in the type learning mechanism (d_r is the dimension of the relation type representation between two nodes).

In the relation-type-aware graph, the representations of edges and nodes in the l-th layer are

$$\mathbf{E}^l, \mathbf{H}_r^l = \text{GNN}(\mathbf{E}^{l-1}, \mathbf{H}_r^{l-1}), \tag{2}$$

where $\mathbf{H}_r^{l-1}$ is the word (node) representations in layer $l-1$ and $\mathbf{H}_r^0$ is the output of BiLSTM in word feature representation. In this paper, we adapt a two-layer GCN and the edge representation is used as the weight of the edge during the message-passing process.

In order to make the model better learn the representation of nodes under different relationships, our relation-type-aware graph updates the edge representation in each layer, making the model focus on learning edge information. The representation of edge in the layer l is updated as

$$\mathbf{E}_{i,j}^l = \mathbf{W}_r \left[\mathbf{E}_{i,j}^{l-1} || \mathbf{h}_i^l || \mathbf{h}_j^l\right], i, j \in [1, N], \tag{3}$$

where $||$ means the concatenation operator, $\mathbf{h}_i^l$ and $\mathbf{h}_j^l$ denote the representations of the two nodes i and node j connected by the edge $e_{i,j}$ in the l_{th} layer, respectively, $\mathbf{W}_r \in \mathbb{R}^{(2\times d+d_r)\times d_r}$ is a learnable transformation matrix (d is the dimension of word representation).

Word-type-aware graph. We next design a word-type-aware graph $\mathcal{G}_w = (\mathcal{V}_w, \mathcal{E}_w)$ with the mapping functions $\psi_{w,v} : \mathcal{V}_w \rightarrow \mathcal{A}_w$ and $\psi_{w,r} : \mathcal{E}_w \rightarrow \mathcal{R}_w$, where $|\mathcal{A}_w| > 1$ and $|\mathcal{R}_w| = 1$. The word-type-aware graph does not introduce the relation type and considers the word types obtained by the CoreNLP syntactic tool. The node embedding of this graph is the concatenation

of word embedding and word type embedding. Thus, for word type learning, we propose to learn the corresponding word type vectors with the word representation:

$$\mathbf{H}_w^l, \mathbf{A}^l = \text{GNN}([\mathbf{H}_w^{l-1} || \mathbf{A}^{l-1}]), \tag{4}$$

where $\mathbf{A} \in \mathbb{R}^{N \times d_w}$ is a learnable vector for the word type representation in the type learning mechanism (d_w is the dimension of the word type representation).

The nodes and edges in the node-type-aware graph are the same as the relation-type-aware graph, but the difference is that the types of nodes in this graph are considered instead of the types of relations between nodes. Our node-type-aware graph updates the representation for each node by aggregating the information from its neighbors through the adjacency matrix. It helps for learning more context information for each node. In order to make the model better learn the representation of nodes under different node types, our node-type-aware graph channel only updates the node type representation, making the model focus on learning node information.

3.4 Semantic Channel

Topological channel can learn word representations based on syntactic dependencies from two dimensions: node type and relation type. In addition, we design a semantic channel to learn connections based on the semantic distance between words. It metrics the similarity of semantics among words. The semantic channel can establish relationships between words that are far apart and grammatically independent but have similar meanings.

Semantic-aware graph. Different from the topological channel, we establish links between the words based on the semantic similarity. We then generate edge weights based on word semantic embeddings over the corpus. The similarity score can be calculated through cosine distance. If the similarity score exceeds a predefined threshold ρ_{sem}, it means that the two words have a semantic relationship in the current sentence. The edge weight of each pair of words can be obtained by

$$\alpha_{i,j} = \text{sim}(\mathbf{h}_i, \mathbf{h}_j) = \begin{cases} \frac{\mathbf{h}_i \cdot \mathbf{h}_j}{||\mathbf{h}_i|| \cdot ||\mathbf{h}_j||}, & \cos(\theta) \geq \rho_{\text{sem}} \\ 0, & \cos(\theta) < \rho_{\text{sem}} \end{cases}, \tag{5}$$

where $\alpha_{i,j}$ denotes the edge weight between words i and j.

Note that the structure of the semantic-aware graph is dynamically changing in the training process, because the edge weight depends on the current representations of words in each step of the training process. Finally, given the current structure of the semantic-aware graph, a GNN is used for learning the representations of words $\mathbf{H}_s$.

3.5 Event Detection

Event detection is a fundamental task in the field of natural language processing, aiming at identifying trigger words from the text and classifying the text into corresponding event types. Although current methods and models for event detection have achieved increasingly good results, they are mostly trained based on generic domain datasets and often encounter the problem of insufficient annotated data when faced with event detection tasks in a specific domain. In electrical power systems, device events are often described in short sentences for efficient communication and convenience for reading and spreading. Thus, it requires the model to impose higher requirements on feature learning and fuse multiple aspects of information to enhance the node representation.

We pass the representation of the topological channel and the semantic channel through a weighted mechanism to obtain the optimal weights of the three different graphs. The word representations in text are formulated as follows:

$$\mathbf{Z} = \lambda_1 \mathbf{H}_r + \lambda_2 \mathbf{H}_w + \lambda_3 \mathbf{H}_s, \tag{6}$$

Table 1. Statistics of Datasets used in Our Work

Dataset	Field	Language	#Events	#Event Types	#Sentences	#Tokens	#Average Sentence Length
MAVEN [57]	General	English	111,611	168	49,873	1,276,000	25.5
ACE 2005 [7]	General	English	4,090	33	15,789	303,000	19.1
PoE (Ours)	**Power**	**Chinese**	**5,124**	**26**	**4,767**	**53,071**	**11.1**

Different from other datasets, our dataset is a Chinese dataset in electrical power systems.

Table 2. Event Type and Sub-type on PoE

Number	Event Type	Event Sub-type
1	measurement	measurement
2	statistics	total, ratio
3	requirement	stipulate, expect
4	operate	connect, disconnect, modify, check
5	happen	start, complete, cancel, analyze, implement, suffer
6	defect	damage, carve, break, burn, placement, dislocation, deformation, corrosion mismatch, shed, variant

It contains six event types and twenty-six sub-types. The defect event type consists of eleven sub-types and the measurement event type only has one sub-type.

where $\lambda_1, \lambda_2, \lambda_3$ are the parameters. We feed the representation of each word into a fully-connected network. It is followed by a softmax function to compute distribution $p(t \mid \mathbf{Z})$ for event type t:

$$p(t \mid \mathbf{Z}) = \text{softmax}(\mathbf{W}_t \mathbf{Z} + \mathbf{b}_t), \tag{7}$$

where $\mathbf{W}_t$ maps the word representation $\mathbf{Z}$ to the feature score for each event type and $\mathbf{b}_t$ is a bias term. After softmax, the event label with the largest probability is chosen as the event classification result. The loss function is formulated as follows:

$$J(\theta) = -\sum_{i=1}^{N_T} \sum_{j=1}^{N_i} \log p(P_j^t \mid T_i, \theta), \tag{8}$$

where N_T is the number of sentences, N_i is the number of words in the i-th sentence, P_j^t is the event type t in the $j-th$ word, and T_i is the i-th sentence.

4 DATASETS AND SETTINGS

4.1 Datasets

We construct a Chinese event detection dataset in electrical power systems, called PoE. The text is device-related event reports in electrical power systems, written by a professional. Due to the advantage of short texts being easy to read and spread, device events are often described in short sentences for efficient communication in electrical power systems. Furthermore, the event release in power system is usually short and contains numerous professional terms, which is convenient for relevant personnel to respond quickly. Thus, the raw data are the short sentence-level data. It describes daily operation of electrical equipment containing 763 event reports. There are 4,767 sentences, 5,124 events, and 53,071 tokens, containing 26 sub-categories of events, as shown in Table 1. To verify the effect of MC-TED on open datasets, we also conduct experiments on ACE 2005 and MAVEN datasets. The average sentence length are 11.1 tokens in PoE, which is 14.4 and 8.0 less than MAVEN and ACE 2005. Furthermore, compared with ACE 2005 and MAVEN datasets, PoE has a few None event type sentences, which is the most event type in other datasets.

Table 2 shows the event type and sub-type detail of PoE. It contains six categories of event type and 26 categories of event sub-type. The PoE mainly describes the equipment statistics, defects,

Table 3. Statistical Information on PoE

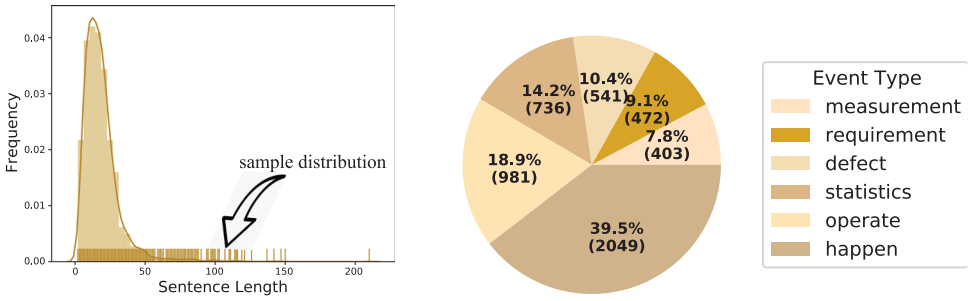
Dataset	#Sentence Number	#Event Sentence Number	#Event Trigger	#No Event Trigger	#Candidate Trigger
Train	3,813	3,521	4,095	1,728	5,823
Valid	477	452	515	217	732
Test	477	451	514	215	729

It is divided into three parts: train, valid, and test sets. The partition rule is to make the ratio of sentence number, event sentence number, and candidate trigger close to 8:1:1.

Table 4. Statistical Information of Co-occurrence Events on PoE Training Data

Dataset	#One Type Number	#Two Type Number	#Three Type Number	#Four Type Number	#Co-occurrence Proportion
Event Type	2,675	423	282	141	0.24
Event Sub-type	2,289	634	352	246	0.35

#One Type Number means the number of sentences having only one event type or one sub-type. #Co-occurrence Proportion means the proportion of co-occurrence sentences.



(a) Sentence length distribution.

(b) Event type.

Fig. 4. Frequency and number of event type. (a) The samples are concentrated between 0 and 50. When the sentence length exceeds 100, the samples become very sparse. (b) The sentence distribution of event type. The largest number of event types is the happen and smallest is the measurement event type.

requirements, and maintenance events in the electric power systems. The data set focuses on nine major equipment defects, including damage, burn, corrosion, long-term placement, dislocation, deformation, fall off, and open circuit. The datasets are divided into training, validation, and test sets at an 8:1:1 ratio according to the sentence number, as shown in Table 3. The proportion of event number, event trigger, and no event trigger in the three datasets are basically the same. The training set contains all event types to ensure that there are no invisible event types in the test set.

Table 4 shows co-occurrence events information of the event type and sub-type detail of PoE on training data. The co-occurrence sentences means a sentence belong to different event types or event sub-types. Our data set contains up to four co-occurrence event types. For the event type and event sub-type, most sentences on the PoE contain only one event type and sub-type, which is the same as the existing event detection data set. However, the PoE contains more multi-event sentences. It has 12% sentences containing two different event types and 18% sentences containing two different event sub-types. Thus, our data set is more suitable for studying the event type detection problem caused by event co-occurrence.

Furthermore, we analysis the sentence length and event type distributions in Figures 4 and 5. In Figure 4(a), the maximum and shortest sentence length are 224 and 6. The most sentences are in [10,25], accounting for 72% of the total data set. The average sentence length is 18.88. It can be seen that the text length is short in PoE, increasing the difficulty of event detection. In Figure 4(b), the event "operate" is the largest number of events in the six categories, with a total of 2,049 sentences,

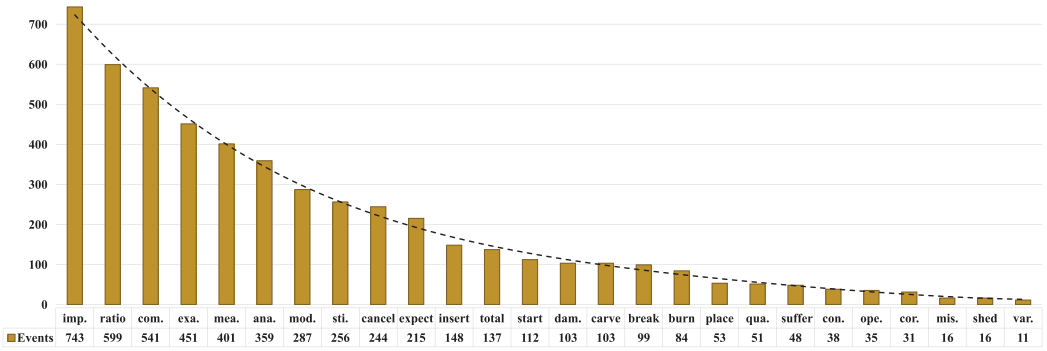


Fig. 5. Sub-type event numbers. The distribution of sub-type events is uneven. The implement sub-type that belongs to the happen event type contains more than 700 sentences and the variant sub-type coming from the defect event type has fewer than 10 sentences.

accounting for 39.98%. The events with the least number are "measure" events (403 events in total, accounting for 7.86%). In Figure 5, the type with the largest number of events is "implementation" in 26 sub-categories, with a total of 743 events. The least number of events is the "deformation" event, 11 events in total. The difference between the type with the largest number of events and the type with the least number of events is very large, indicating that the long tail phenomenon is obvious in the sub-categories, the same as ACE 2005 [7] and MAVEN [57].

Compared with the common ED datasets, PoE contains the following four differences:

- **Power Field.** Different from ACE 2005 and MAVEN, PoE is an event detection dataset from the power systems instead of the general field. The events in PoE, including electrical equipment maintenance measures and recommendations, are closely related to the business in the electric power field.
- **Short Text.** The texts in PoE are collected from the electrical records, which are usually the short sentences with non-compliant grammar. The little information brings difficulty to semantic understanding. (In PoE, the average text length is 18.8.)
- **Professional Terms.** PoE is a Chinese sentence-level event detection dataset, containing numerous professional terms in the power systems, such as "laying path", "direct buried", and "penetration seal". It brings limitations to the existing methods in the general field, since the semantics of some terms change when they are used in the power systems.
- **Colloquial expressions.** The existing data sets are crawling from news, whose expression is more formal, basically in accordance with the grammar rules. The PoE comes from equipment maintenance records of electrical worker. Therefore, it contains numerous colloquial expressions, and missing sentence elements.

4.2 Implementation Details

All models are implemented based on PyTorch, an open-source deep learning framework. The model parameters are initialized as Gaussian distribution and updated and optimized by Adam algorithm [23]. For re-implementation, Table 5 reports our hyper-parameter settings in MC-TED on ACE 2005, MAVEN, and PoE. The common parameters are set as follows: the learning rate is 0.15, the regularization coefficient is 0.001, the model layer is 2, the word representation dimension is 100, the word type representation dimension is 25, and the relation type representation dimension is 25. The batch size is 10 on ACE 2005 and PoE, and 32 on MAVEN. For our MC-TED, we adopt a two-layer GCN [56] as the graph neural network in our framework. The random dropout ratio is

Table 5. Hyper-parameter Settings in MC-TED

Hyper-parameters	ACE 2005	MAVEN	PoE
GCN initial learning rate	0.15	0.15	0.15
Train epoch	100	100	100
Model dimensions	150	150	150
Batch size	10	32	10
Model layer	2	2	2
Dropout rate	0.55	0.50	0.60
Threshold ρ_{sem}	0.10	0.12	0.15
λ_1	2	2	2
λ_2	1	1	1
λ_3	2	2	2

Due to the differences among the three datasets, the optimal hyper-parameters are different.

set to 0.55, 0.50, and 0.60 on ACE 2005, MAVEN, and PoE. The similarity score thresholds ρ_{sem} are 0.10, 0.12, and 0.15 on ACE 2005, MAVEN, and PoE. The early stop strategy is selected to prevent the model from overfitting. Specifically, if the model does not decrease the loss function of the validation set for 100 consecutive turns, the operation is stopped.

To ensure the fairness, all baseline models use the same dataset partitioning as stated in Section 4.1. We use the same value for the common hyper-parameters, including the optimizer, learning rate, batch size, and epoch. All graph-based, LSTM-based, and CNN-based models use the same node representation dimension, which is set to 150 dimensions to keep dimensions consistent with the GNN-based models. For CNN-based methods, the filter kernel size is set to 3×3 with two layers. For LSTM-based methods, the model layer is two. For BERT-based approaches, we use BERT [6] as sequence encoder, which has 12 layers, 768-dimensional hidden embeddings, and 12 attention heads. All the experiments are conducted on one NVIDIA V100 32GB GPU. Note that all the hyper-parameter settings are tuned on the validation set by the grid search with five trials.

4.3 Baselines

Many event detection models have been designed for these years, which can be divided into CNN-based, LSTM-based, GNN-based, BERT-based, and power systems baselines according to model structure. We compare our method with the following four types of ED baselines:

CNN-based baselines. (1) CNN is a two-layer CNN model. It uses sequence tagging to implement event detection. (2) DMCNN [3] leverages a CNN to automatically learn sequence representations with a dynamic multi-pooling mechanism.

LSTM-based baselines. (3) BiLSTM (2-layer) is a two layer BiLSTM. (4) MLBiNet (2-layer) [36] uses a multi-layer bidirectional network to model document-level dependencies. We use CoreNLP to obtain the entity for embedding.

GNN-based baselines. (5) GCN (2-layer) [24] is a semi-supervised graph convolutional network without considering node type and relation type. (6) GAT (2-layer) [54] uses attention mechanism for learning neighbor weights. (7) MOGANED [60] uses a multi-order graph attention network to effectively model the multi-order syntactic relations in dependency trees. (8) GatedGCN [26] designs a graph based on dependency tree and distance to the candidate trigger of each word.

Table 6. Event Detection Results on our Electrical Power Dataset PoE and PoE-Multi

Datasets	PoE (Our Dataset)			PoE-Multi (only multiple event sentences)		
Metric	F1 (%)	Precision (%)	Recall (%)	F1 (%)	Precision (%)	Recall (%)
CNN	61.03 ± 1.24	59.17 ± 1.71	63.02 ± 1.10	55.23 ± 0.21	54.19 ± 1.01	58.63 ± 1.34
DMCNN [3]	62.34 ± 1.53	59.83 ± 1.57	64.27 ± 1.93	55.83 ± 1.21	57.39 ± 0.94	58.30 ± 0.74
BiLSTM	62.50 ± 1.66	56.98 ± 1.72	65.89 ± 1.56	54.83 ± 1.02	53.39 ± 0.26	57.73 ± 1.07
MLBiNet (2-layer) [36]	62.50 ± 1.66	56.98 ± 1.72	65.89 ± 1.56	60.78 ± 1.12	54.89 ± 1.20	61.83 ± 1.04
GCN [24]	60.36 ± 1.13	61.14 ± 1.09	59.60 ± 1.54	57.48 ± 1.30	57.89 ± 1.98	54.38 ± 1.48
GAT [54]	58.88 ± 1.47	57.43 ± 1.93	60.42 ± 1.27	54.89 ± 2.73	56.84 ± 2.83	57.38 ± 1.03
MOGANED [60]	61.95 ± 1.78	58.94 ± 1.73	64.84 ± 1.34	58.87 ± 0.84	56.48 ± 0.92	59.89 ± 1.26
GatedGCN [26]	61.83 ± 1.92	60.73 ± 1.34	65.89 ± 1.92	59.34 ± 1.43	58.72 ± 2.04	60.73 ± 1.44
EE-GCN [5]	62.46 ± 1.17	57.41 ± 1.42	67.80 ± 1.21	60.02 ± 2.32	56.37 ± 1.93	60.78 ± 2.34
BERT+CRF	63.48 ± 1.05	61.13 ± 1.52	67.48 ± 1.81	60.37 ± 0.83	61.08 ± 0.89	63.92 ± 0.38
BERT+BiLSTM	63.23 ± 1.87	61.83 ± 1.30	66.83 ± 1.93	61.73 ± 2.84	60.47 ± 1.93	63.78 ± 0.47
BERT+CNN	63.21 ± 1.92	61.83 ± 0.37	66.98 ± 1.32	61.87 ± 1.37	60.29 ± 1.93	63.37 ± 2.29
DMBERT [55]	63.56 ± 1.72	62.21 ± 1.78	67.70 ± 1.33	61.28 ± 0.73	59.48 ± 1.34	64.28 ± 1.27
HS^3M [65]	60.23 ± 1.21	57.95 ± 1.25	64.93 ± 1.75	56.32 ± 1.02	56.76 ± 1.52	56.69 ± 1.59
SS-LOF [34]	62.53 ± 1.17	59.65 ± 1.83	62.90 ± 1.76	58.25 ± 0.70	58.19 ± 1.54	55.29 ± 1.65
MC-TED (Ours)	**65.58 ± 1.24**	**63.71 ± 1.29**	**69.93 ± 1.61**	**62.82 ± 1.23**	**61.03 ± 1.72**	**65.38 ± 1.29**

The PoE-Multi means the sentences set that contain multiple events in a sentence. Our approach performs best, even when only considering multiple event sentences, which shows MC-TED is the best choice in PoE.

It considers the important scores to build the graph. (9) EE-GCN [5] is an edge-enhanced graph convolutional network using the syntactic structure, which uses the syntactic structure and the typed dependent label information.

BERT-based baselines. (10) BERT+CRF is a vanilla BERT model. It uses a multi-classification approach to implement event detection. (12) BERT+BiLSTM uses a BERT based model and a two layer BiLSTM after the BERT. (13) BERT+CNN adapts a BERT model and a two layer CNN model after the BERT. (14) DMBERT [55] takes advantage of BERT and adopts the dynamic multi-pooling mechanism to aggregate features.

Power systems baselines. (15) HS^3M [65] combines unlabeled data and partly labeled data in a large margin learning objective, which is a classical method for event detection in the power systems. It incorporates partial knowledge for event detection. (16) SS-LOF [34] is an event detection algorithm in the power systems using reduced data and local outlier factor for detecting the events in power systems. It is employed to determine the region of the event source.

5 EXPERIMENTS AND RESULTS

5.1 Experimental Results

In order to verify the effectiveness of MC-TED, the latest CNN-based, LSTM-based, graph-based, and BERT-based event detection models are selected as the baseline models. We perform event detection on PoE, MAVEN, and ACE 2005 datasets and report the average results with the standard deviation across the 10-folds cross-validation in Tables 6 and 7. Part of the results are from the original papers [3, 5, 36, 55, 57, 60].

MC-TED consistently outperforms all comparison algorithms on the **electrical power datasets (PoE)**, as well as only having multiple event sentences on the PoE in Table 6. Compared with CNN-based models (CNN and DMCNN), MC-TED achieves 4.55% and 3.24% improvements in terms of the F1-score on PoE, and promotes at least 6% points on only multiple event sentences, respectively. It should be contributed to semantic dependencies modeling and syntactic topology preserving ability of the proposed model. Compared with BERT+CRF, BERT+BiLSTM, BERT+CNN, and DMBERT, MC-TED improves the F1-score by at least 2% on PoE and 1% on only multiple

Table 7. Event Detection Results on General Datasets MAVEN and ACE 2005

Datasets	**MAVEN**			**ACE 2005**		
Metric	F1 (%)	Precision (%)	Recall (%)	F1 (%)	Precision (%)	Recall (%)
DMCNN [3]	60.60 ± 0.20	66.30 ± 0.89	55.90 ± 0.50	68.00 ± 1.95	73.70 ± 2.42	63.30 ± 3.30
MLBiNet (2-layer) [36]	63.69 ± 1.29	62.52 ± 1.61	66.91 ± 2.75	74.60 ± 1.82	72.62 ± 1.35	76.74 ± 1.83
MOGANED [60]	63.80 ± 0.18	63.40 ± 0.88	64.10 ± 0.90	72.10 ± 0.39	70.40 ± 1.38	73.90 ± 2.24
GatedGCN [26]	65.82 ± 1.37	63.72 ± 1.73	67.78 ± 1.83	73.40 ± 1.89	76.70 ± 1.48	70.50 ± 1.72
EE-GCN [5]	66.92 ± 1.53	64.29 ± 1.67	69.82 ± 1.38	74.84 ± 1.87	**75.84 ± 1.92**	72.93 ± 1.62
BERT+CRF	67.80 ± 0.15	65.00 ± 0.84	70.90 ± 0.94	74.10 ± 1.56	71.30 ± 1.77	77.10 ± 1.99
DMBERT [55]	67.10 ± 0.41	62.70 ± 1.01	**72.30 ± 1.03**	74.30 ± 0.81	70.20 ± 1.71	**78.90 ± 1.64**
MC-TED (Ours)	**68.47 ± 1.04**	**66.55 ± 1.43**	70.09 ± 1.78	**75.26 ± 1.53**	74.62 ± 1.73	73.38 ± 1.82

Our model achieves the best performance on F1, which demonstrates the universality of our method.

event sentences through modeling multi-channel graphs catching more semantic information in a complex context. MC-TED models the heterogeneity of the graph compared with other graph-based models. It fully considers the semantic information of multiple node types and relations and performs weighted fusion, achieving better results.

Table 7 shows the overall results of each approach per evaluation dataset using uniform parameter settings in Section 4.2. MC-TED achieves a significant improvement in most cases on MAVEN and ACE 2005. Compared with DMCNN and MLBiNet, our model consistently outperforms in both ACE 2005 and MAVEN on F1, precision, and recall. Thus, our method performed significantly better than the CNN-based and LSTM-based methods, which learns word representation from topological and semantic perspectives. Compared with BERT+CRF and DMBERT, MC-TED has the best performance and uses significantly fewer word representation dimensions through building graph for each sentence. Compared with MOGANED [60], GatedGCN [26], and EE-GCN [5], three recent graph neural network models, our model respectively achieves 4.67%, 2.65%, and 1.55% improvements on the F1-score on MAVEN, and improves at least 0.42% points on ACE 2005. It shows that our method is significantly superior to the graph-based methods due to utilizing node and edge types information.

Specifically, compared with EE-GCN, it achieves 3.12%, 1.55%, and 0.42% improvements of F1-score on PoE, MAVEN, and ACE 2005. It demonstrates that MC-TED fully accounts for the rich semantics reflected by multiple node types and relations. Therefore, the node representation is more comprehensive and differentiated. GCN, GAT, EE-GCN, MOGANED, MLBiNet, GatedGCN, and MC-TED consider the importance of neighbors in learning node representations, and the performance of our model is more competitive due to its ability to capture multiple semantic information on multi-channel heterogeneous graphs.

It should be noted that the improvement of MC-TED is more significant on the PoE dataset. It improves at least 1.5 points on the three evaluation metrics. The major reason is that PoE is a Chinese power field dataset, which increases the difficulty of syntactic analysis and type recognition. Therefore, even though MC-TED can fuse multiple node types and relations types, its improvement is not obvious because ACE 2005 does not provide validity information. Therefore, for Precision and Recall evaluation metrics, our model is not optimal. It has the best performance in F1 evaluation metric. Taken together, MC-TED achieved peak performance across all datasets, demonstrating the need to model relation-type-aware graph, node-type-aware graph as well as semantic-aware graph.

5.2 Ablation Study on Three Graphs

For further evaluation on the components of MC-TED, we conduct the ablation experiments on PoE and report the results in Table 8.

Table 8. Ablation Study of Topological Channel and Semantic Channel

Method	F1 (%)	Precision (%)	Recall (%)
Relation-type-aware graph (G1)	62.94 ± 1.25	59.00 ± 1.91	67.45 ± 1.65
Word-type-aware graph (G2)	61.95 ± 1.93	59.91 ± 1.26	64.14 ± 1.38
Semantic-aware graph (G3)	62.90 ± 1.33	57.42 ± 1.33	69.53 ± 1.25
Homogeneous graph (G0)+G3	62.64 ± 1.15	59.48 ± 1.64	66.15 ± 1.48
G1+G2	63.94 ± 1.62	**64.77 ± 1.52**	63.13 ± 1.38
G1+G3	64.07 ± 1.12	62.27 ± 1.79	**70.25 ± 1.61**
G2+G3	63.48 ± 1.42	60.18 ± 1.92	67.17 ± 1.43
G1+G2+G3	**65.58 ± 1.24**	63.71 ± 1.29	69.93 ± 1.61

The bold values are the best result and the underlined values are the second best result.

Table 9. Ablation Study of Topological Channel

Method	F1 (%)	Precision (%)	Recall (%)
MC-TED (only Topological Channel)	**63.94 ± 1.62**	64.77 ± 1.52	63.13 ± 1.38
w/o relation type module	62.82 ± 1.39	62.78 ± 1.23	62.20 ± 1.89
w/o node type module	63.28 ± 1.90	64.29 ± 2.14	63.01 ± 1.29
w/o relation and node types module	61.83 ± 1.27	62.51 ± 1.43	64.10 ± 1.03
w/o relation type learning module	63.14 ± 1.19	**65.94 ± 2.37**	63.26 ± 1.43
w/o node type learning module	62.39 ± 1.04	63.30 ± 1.58	63.18 ± 1.13
w/o relation and node types learning module	62.02 ± 2.05	62.21 ± 1.61	**64.47 ± 1.75**

The bold values are the best result and the underlined values are the second best result.

Topological channel. Both the relation-type-aware graph and the word-type-aware graph are vitally essential for the type utilized multi-channel GNN. As shown in Table 8, on the one hand, compared with the full MC-TED, removing the relation-type-aware graph (G1 in Table 8) results in low performance (e.g., a 2.10% decrease in terms of F1 score). On the other hand, considering inside the topological channel, compared with only applying the word-type-aware (G2 in Table 8), G1 plus G2 makes an improvement of 1.99%, implying the relation-type-aware graph contributes to the topological information capture. The conclusion is consistent for the word-type-aware graph. The main reason is that it achieves a more detailed and comprehensive description considering the semantics of word and relation types. Furthermore, we replace the two graphs with a single GCN and use a homogeneous graph (i.e., without any word type information and relation type information, G0 in Table 8) as its input. The results show a 2.94% decrease after the replacement and a 2.68% decrease when only applying the semantic channel, which reflect the importance of the topological channel with type utilizing multi-channel GNN. It may be because the topological channel designs two different heterogeneous graphs utilizing word dependencies and part-of-speech tags. The relation-type-aware graph updates the edge representation in each model layer, making the model focus on learning edge information. All the observations indicate that our model through utilizing the relation type improves the event detection performance.

Semantic channel. The semantic-aware graph helps capture richer semantic information beyond the original syntactic topology. From Table 8, it is noticed that the performance drops by 1.64% in terms of F1 score when not including the semantic-aware graph (G3 in Table 8). It demonstrates that the semantic channel can establish relationships between words that are far apart and grammatically independent but have similar meanings. The structure of the semantic-aware graph dynamically changes during the training process. Furthermore, after adding the semantic-aware graph, the performance of only applying the relation-type-aware graph or word-type-aware graph increases by 1.13% and 1.53%, respectively. Therefore, the semantic channel improves the

superiority of MC-TED, independent of the topological channel. It also demonstrates that the semantic channel refines textual representations with semantic similarity, building the semantic information interaction among potential event-related words.

All the observations demonstrate the effectiveness of two channel in our model. With the two channels, MC-TED integrates multiple aspects of information to learn a more comprehensive word representation through fusing multiple sources of information. The semantic channel reflects the similarity of words through adding relation with lower cosine distance. The topological channel constructs a relation-type-aware graph and a word-type-aware graph to learn more comprehensive node representations. They consider the word type and relation type to portray nonlinear relationships with stronger representative and inference capabilities fully.

5.3 Ablation Study on Topological Channel

We evaluate variants on topological channel of our approach, as shown in Table 9. We show the experimental results and observe that:

Type module. We exclude the relation type module of relation-type-aware graph for the type utilized multi-channel GNN. As a consequence, excluding the module decreases the F1-scores by 1.12%, as well as precision by 1.99%, and recall by 0.97%. Thus, the F1-scores decrease in all metrics, thereby showing that the relation type module could significantly improve the performance. The relation type module positively affected the learning of word relationship, and relation type knowledge is helpful for event detection with the type utilized multi-channel GNN. We further exclude the node type module, which improves event detection by utilizing node type knowledge. The decreases for the F1, precision, and recall are 0.66%, 0.48%, and 0.12%, respectively, with significant effects on performance. When both the relation type and node type are excluded, the performance is decreased dramatically. These findings indicate that relation type module in G1 and word type module in G2 can effectively improve event detection.

Type learning module. The topological channel has relation type learning and node type learning module for learning more appropriate word representations according to context. It is designed for weaken word and relation type errors in the electric power texts caused by the terminology vocabulary. When the relation type learning module is excluded, the F1-score is decreased by 0.80%. Compared with removing the whole relation type module, the performance is increased which demonstrates the validity of adding relation type. We further exclude both relation and node type learning modules, the F1-score decreases 1.92% points. These results suggest that all of the modules are useful, and that the relation type module is the most important for event detection because excluding them dramatically degraded the performance.

5.4 Discussion on Generalized Variants

To evaluate the effectiveness and generality of components, we experiment various variants on our dataset, as shown in Table 10. Specifically, for word initial representation, we employ six models: Random, Random+BiLSTM, Glove, Glove+BiLSTM, BERT and BERT+BiLSTM. It consists of two branches, that is whether to use BiLSTM. For the encoder, we apply GAT, GCN, and BERT to learn word embedding. For the aggregation method, we adopt concatenation and average methods to merge the representation of three graphs. We analyze the experimental results and observe that:

Initial representation. In the most circumstance, using BiLSTM can boost word representation through learning context information. Furthermore, based on different static word representation methods, the random method acquires the best performance in most instances, which utilizes significantly fewer dimensions of words than BERT.

Encoder. GCN encoder is better than GAT encoder in our scheme. This is mainly because GCN structure alleviates the effect of word representation smoothing, which is more suitable for our

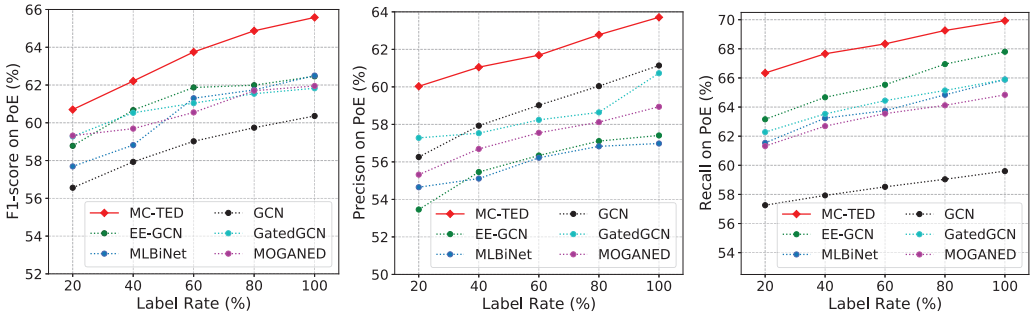
Table 10. Evaluation for Variants on PoE

Initial Representation	Encoder	Aggregation	F1 (%)	Precision (%)	Recall (%)
Random	**GAT**	Concatenation	62.38 ± 1.66	61.28 ± 1.98	65.21 ± 1.46
		Weighted	60.02 ± 1.62	58.24 ± 1.20	62.32 ± 1.74
	GCN	Concatenation	61.19 ± 1.08	60.23 ± 1.60	62.79 ± 1.04
		Weighted	61.76 ± 1.01	60.23 ± 1.01	59.37 ± 1.63
	BERT	Concatenation	62.53 ± 1.12	60.93 ± 1.29	66.93 ± 1.32
		Weighted	63.83 ± 1.27	61.92 ± 1.72	67.28 ± 1.28
Random+BiLSTM	**GAT**	Concatenation	60.87 ± 1.76	61.39 ± 1.03	64.29 ± 1.84
		Weighted	60.06 ± 1.44	60.90 ± 1.37	63.05 ± 1.17
	GCN	Concatenation	59.64 ± 1.81	60.45 ± 1.10	62.82 ± 1.39
		Weighted (Ours)	**65.58 ± 1.24**	**63.71 ± 1.29**	**69.93 ± 1.61**
	BERT	Concatenation	63.82 ± 1.01	60.92 ± 1.73	66.38 ± 1.83
		Weighted	63.23 ± 1.87	61.83 ± 1.30	66.83 ± 1.93
Glove	**GAT**	Concatenation	60.76 ± 1.19	59.74 ± 1.29	62.22 ± 1.98
		Weighted	60.20 ± 1.92	58.20 ± 1.76	62.72 ± 1.93
	GCN	Concatenation	59.32 ± 1.90	62.36 ± 1.81	61.01 ± 1.89
		Weighted	62.14 ± 1.25	57.18 ± 1.53	56.67 ± 1.65
	BERT	Concatenation	63.34 ± 1.74	62.03 ± 1.24	67.38 ± 1.74
		Weighted	62.93 ± 1.83	60.35 ± 1.39	66.87 ± 1.39
Glove+BiLSTM	**GAT**	Concatenation	63.85 ± 1.52	60.13 ± 1.39	61.71 ± 1.21
		Weighted	60.91 ± 1.63	53.52 ± 1.94	61.62 ± 1.75
	GCN	Concatenation	59.16 ± 1.51	60.73 ± 1.09	59.58 ± 1.02
		Weighted	62.55 ± 1.14	<u>62.64 ± 1.12</u>	61.01 ± 1.60
	BERT	Concatenation	63.04 ± 1.37	60.65 ± 1.63	65.26 ± 1.35
		Weighted	63.84 ± 1.84	60.74 ± 1.63	67.83 ± 1.34
BERT	**GAT**	Concatenation	61.68 ± 1.25	60.28 ± 1.29	65.12 ± 1.78
		Weighted	60.90 ± 1.84	58.49 ± 1.02	59.94 ± 1.82
	GCN	Concatenation	62.96 ± 1.53	61.37 ± 1.35	66.40 ± 1.06
		Weighted	61.61 ± 1.10	60.46 ± 1.30	65.50 ± 1.43
	BERT	Concatenation	62.84 ± 1.23	61.05 ± 1.35	66.92 ± 1.83
		Weighted	63.48 ± 1.05	61.13 ± 1.52	<u>67.48 ± 1.81</u>
BERT+BiLSTM	**GAT**	Concatenation	63.64 ± 1.83	60.38 ± 1.23	61.70 ± 1.87
		Weighted	61.74 ± 1.64	59.53 ± 1.92	62.76 ± 1.85
	GCN	Concatenation	59.60 ± 1.34	60.63 ± 1.29	63.01 ± 1.54
		Weighted	<u>64.16 ± 1.81</u>	60.25 ± 1.62	64.17 ± 1.01
	BERT	Concatenation	61.63 ± 1.53	60.74 ± 1.26	66.26 ± 1.25
		Weighted	63.06 ± 1.03	61.24 ± 1.53	66.42 ± 1.32

The bold values are the best result and the underlined values are the second best result.

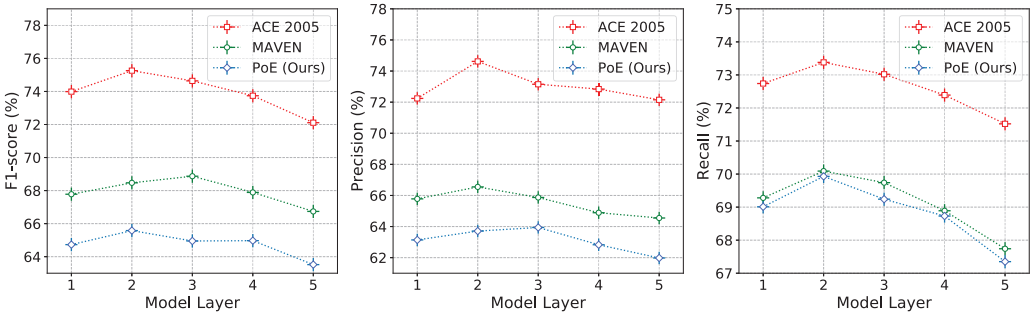
dataset with short sentences. For BERT-based encoder, performance will be reduced compared to Random+BiLSTM+GCN method. The main reason is that event detection in short texts is more dependent on syntactic dependency correlation. In our baseline methods, the BERT-based method has better performance than all LSTM-based models and GNN-based models. However, in the BERT-based models, we try to combine the graph method and the model has a bad effect. It may be because the whole model is strongly dependent on BERT's representation, so it is difficult to further improve. Furthermore, the training time of BERT-based model was significantly longer than that of graph-based model. In order to meet the practical application requirements, we use graph-based model as the encoder.

Aggregation method. Our weighted aggregation method is to set appropriate weights among relation-type-aware graph, word-type-aware graph, and semantic-aware graph for obtaining the



(a) Impact of label rate on F1. (b) Impact of label rate on Precision. (c) Impact of label rate on Recall.

Fig. 6. Parameter study on label rate. The label rate refers to the proportion of labeled data. MC-TED preforms best at all label ratios, which demonstrates our method also works with small samples.



(a) Impact of model layer on F1. (b) Impact of model layer on Precision. (c) Impact of model layer on Recall.

Fig. 7. Parameter study on model layer. We tested the performance of GCN layer between 1 and 5. When the number of model layers is 2, the performance reaches the peak.

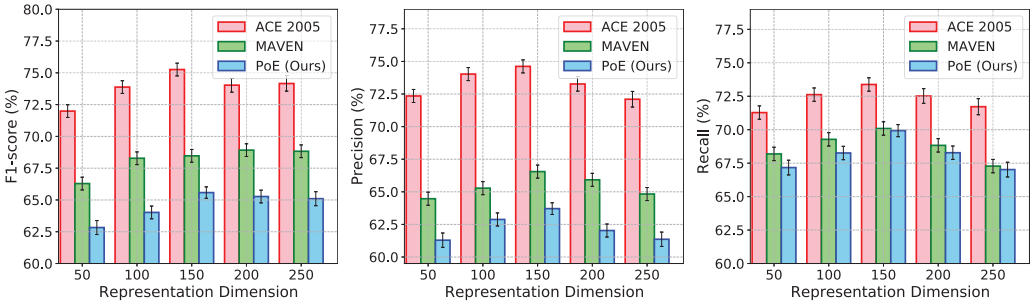
final representation of each word. Our concatenation aggregation method merges the representation of three graphs, which makes the word dimension three times larger. In our scheme, the weighted method usually obtains better results than the concatenation method.

5.5 Parameter Study

In order to further evaluate the sensitivity of MC-TED to parameters, we conduct parameter experiments on PoE. Figures 6, 7, 8, and 9 show the effects of label rate, the GCN layer, the dimension of node representation, and the cosine distance of node representation.

Label rate. As shown in Figure 6, we compare our model with EE-GCN and MLBiNet using two layers. MC-TED is significantly superior to the advanced graph-based methods on most label rates. It shows that our model utilizes the node type and relation type is efficient regardless of the number of labeled samples. MC-TED performs best when we use the full dataset.

Model layer. The number of layers of the heterogeneous graph neural network is relatively shallow. As the number of layers deepens, the effect of heterogeneous graph neural network will decrease significantly (semantic confusion phenomenon). Figure 7 shows the performance of MC-TED with 1-5 layers GCN. It can be seen that as the number of model layers deepens, the performance of our model will decrease significantly (semantic confusion phenomenon). The 2-layer, MC-TED can learn the node representation with distinguishing degree by capturing the



(a) Impact of representation dimension on F1.

(b) Impact of representation dimension on Precision.

(c) Impact of representation dimension on Recall.

Fig. 8. Parameter study on representation dimension. PoE not only have the problems of long tail and multi-event difficulties existing in ACE and MAVEN, but also have the characteristics of short text, many terms, and colloquial language, which increase the difficulty of detection. Therefore, the performance of PoE is lower than that of the other two data sets. In addition, our model performs best on representation dimension of 150.

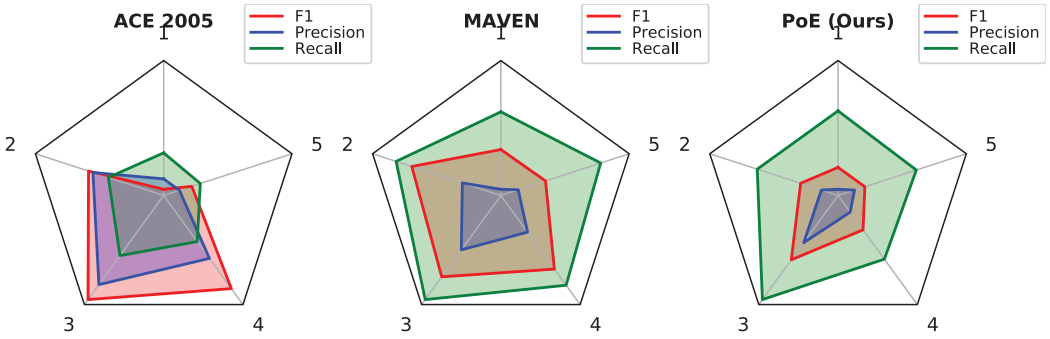


Fig. 9. Parameter study on threshold ρ_{sem}. In all datasets, our model performs best on the threshold of 0.3. F1 values vary significantly in our data set and ACE 2005, which demonstrates the validity of semantic channel.

heterogeneous structure and retain rich semantics on the graph, while the node representation learned by the 5-layer of our model has become indistinguishable. The number of layers of MC-TED is relatively shallow (usually, 1-2 layers are better). The MC-TED model performs best with 2 GCN layers.

Model dimensions. The representation dimension of nodes will directly affect the performance of MC-TED. As shown in Figure 8, with the increase of node representation dimension, the performance of our model increased slowly at first, then remained unchanged, and finally decreased slowly. This is because MC-TED needs enough dimensions to encode semantic and structural information, but too large dimensions may introduce some redundancy and lead to the over-fitting phenomenon. MC-TED performs best when the node representation dimension is 150. Therefore, we finally set the node representation dimension as 150.

Cosine distance. In the semantic-aware channel, we construct the edge between nodes according to cosine distance. We let nodes whose distance is less than the threshold have no edges, and MC-TED updates edge in training. As shown in Figure 9, with the increase of the threshold, the performance of the MC-TED model increased slowly at first, and then decreased slowly. It achieves optimal performance when the threshold value is 0.3.

Table 11. Analysis of Type Learning Mechanism

Method	F1 (%)	Precision (%)	Recall (%)
MC-TED	**65.58 ± 1.24**	63.71 ± 1.29	**69.93 ± 1.61**
w/o relation type learning module	63.75 ± 1.47	**64.92 ± 1.92**	62.63 ± 1.03
w/o word type learning module	63.29 ± 1.58	60.65 ± 1.37	66.16 ± 1.65

The bold values are the best result.

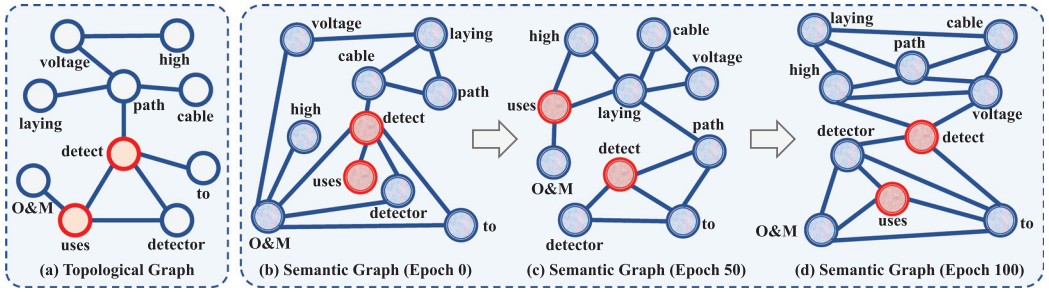


Fig. 10. Analysis of multiple channels: topological graph and learning process of the semantic graph. In the semantic channel, the connection of the graph will change with the increase of model training epochs. The words in one event will be connected, and the connection of words in different events will be disconnected.

5.6 Analysis of Multiple Channels

We further provide more detailed analysis on topological channel and semantic channel. Specifically, we (1) investigate the type learning mechanism in topological channel for the type utilized multi-channel GNN by ablation study, (2) visualize the structure change of the semantic-aware graph to study the effect of network reconstruction, and (3) compare the two graphs to seek the relation between the two channels. We have the following three observations.

Type learning in topological channel is essential. As an important part of reducing the error of wrong analysis of both relation types and word types, the type learning mechanism is proven to be necessary in Table 11. Without the relation type learning or word type learning, the results of event detection have a decline by 1.83% and 2.29% in terms of F1 score. Thus, the type learning mechanism can effectively reduce the cumulative error caused by the analysis tool, which is exacerbated by the professional domain-specific terms in the power systems.

Dynamic structure learning in semantic channel is obviously effective. The semantic channel reconstructs the semantic-aware graph and learns a new structure based on the current representations in each training step. We visualize the structure change of semantic-aware graph during a training process in Figure 10(b), (c), and (d), taking 50 epochs as an interval. The effectiveness of structure learning is obvious that the semantic-relevance words in an event are becoming a cluster during the training process and the corresponding trigger is eventually at the cluster center. The tight connections between the trigger and core arguments makes the trigger identification more accurate, and cluster characteristics help to classify the event type.

Semantic channel and topological channel complement each other. Figure 10(a) and (b) show examples of the topological graph and semantic graph. It is clearly noticed that the semantic channel learns a graph which supplements the non-semantic shortcomings of the topological one in some cases. For instance, the words "voltage" and "cable" are connected with semantic relevance, while ignored by the topological graph. The above conclusion is also the same for the topological channel with the type utilized multi-channel GNN.

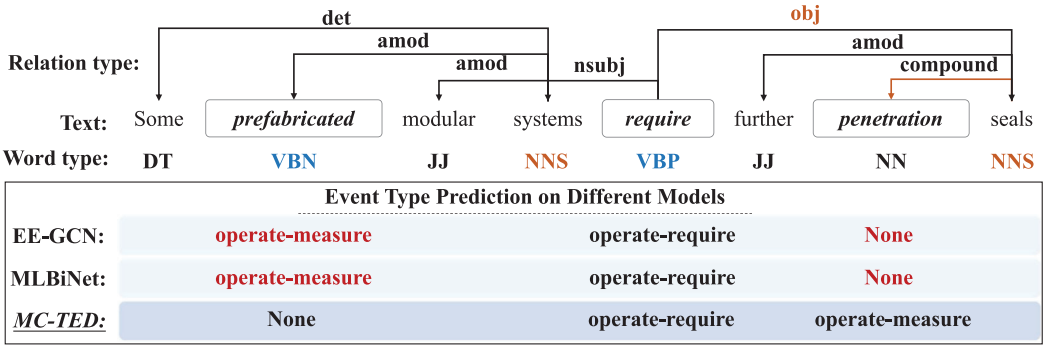


Fig. 11. An example of event type prediction. The red color represents wrong prediction. Our method identifies the candidate trigger penetration is a real trigger and the event type is operate-measure, and judges the candidate trigger prefabricated is not a real trigger. It demonstrates that our model can accurately predict some easily confused events.

5.7 Case Study

As shown in Figure 11, we give the event detection results of EE-GCN, MLBiNet, and MC-TED. Through the CoreNLP tools, an event "penetration seals" is a "compound" and misrepresented as an element of the clause. Therefore, it is difficult to detect the trigger word "penetration" based on the syntactic dependency graph. MC-TED utilizes the type information and makes the node type and relation type learnable. MC-TED can accurately classify the events corresponding to the trigger "penetration" by introducing node and relation type. It shows that MC-TED can optimize the representation of nodes and edges with a relation-type-aware graph and node-type-aware graph. Furthermore, the topological graph is constructed based on a CoreNLP syntactic analysis tool, thus there are some node type and relation type that are easily to be classified to be the trigger words, such as the word "prefabricated" where the word type are classified as "VBN". It demonstrates that MC-TED introduces nodes type and edges type learnable mechanism, weakening thes misleading effect of node type.

6 CONCLUSION

In this paper, we propose MC-TED, a multi-channel GNN which utilizes type information for event detection in power systems. To solve the problems incurred by short texts and professional terminologies in power systems, MC-TED learns representations from topological and semantic channels to capture richer information. The topological channel generates a relation-type-aware graph and a word-type-aware graph, learning relation-type- and word-type-constrained features. The relation type and word type can both be learned in these two graphs to reduce errors generated by syntactic analysis tools. Furthermore, we build a Chinese event detection dataset in power systems, called PoE, which is derived from electrical power records. Extensive experimental results demonstrate that MC-TED achieves state-of-the-art performance on ACE, MAVEN, and PoE datasets.

ACKNOWLEDGMENTS

We thank the anonymous reviewers for their insightful comments and suggestions.

REFERENCES

[1] J. Arrillage, Charles P. Arnold, and B. J. Harker. 1983. Computer modelling of electrical power systems. (1983).

[2] Yuwei Cao, Hao Peng, Jia Wu, Yingtong Dou, Jianxin Li, and Philip S. Yu. 2021. Knowledge-preserving incremental social event detection via heterogeneous GNNs. In *WWW'21: The Web Conference 2021, Virtual Event/Ljubljana, Slovenia, April 19–23, 2021*. 3383–3395. https://doi.org/10.1145/3442381.3449834

[3] Yubo Chen, Liheng Xu, Kang Liu, Daojian Zeng, and Jun Zhao. 2015. Event extraction via dynamic multi-pooling convolutional neural networks. In *Proceedings of the 53rd Annual Meeting of the Association for Computational Linguistics and the 7th International Joint Conference on Natural Language Processing (Volume 1: Long Papers)*. 167–176.

[4] Xin Cong, Shiyao Cui, Bowen Yu, Tingwen Liu, Yubin Wang, and Bin Wang. 2021. Few-shot event detection with prototypical amortized conditional random field. In *Findings of the Association for Computational Linguistics: ACL/IJCNLP 2021, Online Event, August 1–6, 2021 (Findings of ACL, Vol. ACL/IJCNLP 2021)*. 28–40. https://doi.org/10.18653/v1/2021.findings-acl.3

[5] Shiyao Cui, Bowen Yu, Tingwen Liu, Zhenyu Zhang, Xuebin Wang, and Jinqiao Shi. 2020. Edge-enhanced graph convolution networks for event detection with syntactic relation. In *Findings of the Association for Computational Linguistics: EMNLP 2020, Online Event, 16–20 November 2020 (Findings of ACL, Vol. EMNLP 2020)*. 2329–2339. https://doi.org/10.18653/v1/2020.findings-emnlp.211

[6] Jacob Devlin, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova. 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, NAACL-HLT 2019, Minneapolis, MN, USA, June 2–7, 2019, Volume 1 (Long and Short Papers)*. 4171–4186. https://doi.org/10.18653/v1/n19-1423

[7] George R. Doddington, Alexis Mitchell, Mark A. Przybocki, Lance A. Ramshaw, Stephanie M. Strassel, and Ralph M. Weischedel. 2004. The automatic content extraction (ACE) program - tasks, data, and evaluation. In *Proceedings of the Fourth International Conference on Language Resources and Evaluation, LREC 2004, May 26–28, 2004, Lisbon, Portugal*. http://www.lrec-conf.org/proceedings/lrec2004/summaries/5.htm.

[8] Yuxiao Dong, Nitesh V. Chawla, and Ananthram Swami. 2017. metapath2vec: Scalable representation learning for heterogeneous networks. In *Proceedings of the 23rd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, Halifax, NS, Canada, August 13–17, 2017*. 135–144. https://doi.org/10.1145/3097983.3098036

[9] Sanghamitra Dutta, Liang Ma, Tanay Kumar Saha, Di Lu, Joel R. Tetreault, and Alex Jaimes. 2021. GTN-ED: Event detection using graph transformer networks. *CoRR* abs/2104.15104 (2021). arXiv:2104.15104 https://arxiv.org/abs/2104.15104

[10] Mohamed E. El-Hawary. 1995. *Electrical Power Systems: Design and Analysis*. Vol. 2.

[11] Shaohua Fan, Junxiong Zhu, Xiaotian Han, Chuan Shi, Linmei Hu, Biyu Ma, and Yongliang Li. 2019. Metapath-guided heterogeneous graph neural network for intent recommendation. In *Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining, KDD 2019, Anchorage, AK, USA, August 4–8, 2019*. 2478–2486. https://doi.org/10.1145/3292500.3330673

[12] Xiaocheng Feng, Lifu Huang, Duyu Tang, Heng Ji, Bing Qin, and Ting Liu. 2016. A language-independent neural network for event detection. In *Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics, ACL 2016, August 7–12, 2016, Berlin, Germany, Volume 2: Short Papers*. https://doi.org/10.18653/v1/p16-2011

[13] Linmei Hu, Tianchi Yang, Chuan Shi, Houye Ji, and Xiaoli Li. 2019. Heterogeneous graph attention networks for semi-supervised short text classification. 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 2019, Hong Kong, China, November 3–7, 2019*. 4820–4829. https://doi.org/10.18653/v1/D19-1488

[14] Ziniu Hu, Yuxiao Dong, Kuansan Wang, and Yizhou Sun. 2020. Heterogeneous graph transformer. In *WWW'20: The Web Conference 2020, Taipei, Taiwan, April 20–24, 2020*. 2704–2710. https://doi.org/10.1145/3366423.3380027

[15] Lifu Huang, Heng Ji, Kyunghyun Cho, Ido Dagan, Sebastian Riedel, and Clare R. Voss. 2018. Zero-shot transfer learning for event extraction. In *Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics, ACL 2018, Melbourne, Australia, July 15–20, 2018, Volume 1: Long Papers*. 2160–2170. https://doi.org/10.18653/v1/P18-1201

[16] Zhichao Huang, Xutao Li, Yunming Ye, and Michael K. Ng. 2020. MR-GCN: Multi-relational graph convolutional networks based on generalized tensor product. In *Proceedings of the Twenty-Ninth International Joint Conference on Artificial Intelligence, IJCAI 2020*. 1258–1264. https://doi.org/10.24963/ijcai.2020/175

[17] Shahbaz Hussain, Rashid Alammari, Atif Iqbal, and Abdullatif Shikfa. 2020. Application of artificial intelligence in electrical power systems. In *IEEE International Conference on Informatics, IoT, and Enabling Technologies, ICIoT 2020, Doha, Qatar, February 2–5, 2020*. 13–17. https://doi.org/10.1109/ICIoT48696.2020.9089447

[18] Yazhou Jiang. 2021. Data-driven probabilistic fault location of electric power distribution systems incorporating data uncertainties. *IEEE Trans. Smart Grid* 12, 5 (2021), 4522–4534. https://doi.org/10.1109/TSG.2021.3070550

[19] Muhammad Raza Khan and Joshua E. Blumenstock. 2019. Multi-GCN: Graph convolutional networks for multi-view networks, with applications to global poverty. In *The Thirty-Third AAAI Conference on Artificial Intelligence, AAAI 2019, The Thirty-First Innovative Applications of Artificial Intelligence Conference, IAAI 2019, The Ninth AAAI Symposium on Educational Advances in Artificial Intelligence, EAAI 2019, Honolulu, Hawaii, USA, January 27–February 1, 2019*. 606–613. https://doi.org/10.1609/aaai.v33i01.3301606

[20] Mahdi Khodayar, Guangyi Liu, Jianhui Wang, and Mohammad E. Khodayar. 2020. Deep learning in power systems research: A review. *CSEE Journal of Power and Energy Systems* (2020).

[21] Mahdi Khodayar, Saeed Mohammadi, Mohammad E. Khodayar, Jianhui Wang, and Guangyi Liu. 2019. Convolutional graph autoencoder: A generative deep neural network for probabilistic spatio-temporal solar irradiance forecasting. *IEEE Transactions on Sustainable Energy* 11, 2 (2019), 571–583.

[22] Mahdi Khodayar and Jianhui Wang. 2018. Spatio-temporal graph deep neural network for short-term wind speed forecasting. *IEEE Transactions on Sustainable Energy* 10, 2 (2018), 670–681.

[23] Diederik P. Kingma and Jimmy Ba. 2015. Adam: A method for stochastic optimization. In *3rd International Conference on Learning Representations, ICLR 2015, San Diego, CA, USA, May 7–9, 2015, Conference Track Proceedings*. http://arxiv.org/abs/1412.6980

[24] Thomas N. Kipf and Max Welling. 2017. Semi-supervised classification with graph convolutional networks. In *5th International Conference on Learning Representations, ICLR 2017, Toulon, France, April 24–26, 2017, Conference Track Proceedings*. https://openreview.net/forum?id=SJU4ayYgl.

[25] Viet Dac Lai, Minh Van Nguyen, Thien Huu Nguyen, and Franck Dernoncourt. 2021. Graph learning regularization and transfer learning for few-shot event detection. In *SIGIR'21: The 44th International ACM SIGIR Conference on Research and Development in Information Retrieval, Virtual Event, Canada, July 11–15, 2021*. 2172–2176. https://doi.org/10.1145/3404835.3463054

[26] Viet Dac Lai, Tuan Ngo Nguyen, and Thien Huu Nguyen. 2020. Event detection: Gate diversity and syntactic importance scores for graph convolution neural networks. In *Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing, EMNLP 2020, Online, November 16–20, 2020*. 5405–5411. https://doi.org/10.18653/v1/2020.emnlp-main.435

[27] Viet Dac Lai, Tuan Ngo Nguyen, and Thien Huu Nguyen. 2020. Event detection: Gate diversity and syntactic importance scores for graph convolution neural networks. *CoRR* abs/2010.14123 (2020). arXiv:2010.14123 https://arxiv.org/abs/2010.14123

[28] Qian Li, Jianxin Li, Jiawei Sheng, Shiyao Cui, Jia Wu, Yiming Hei, Hao Peng, Shu Guo, Lihong Wang, Amin Beheshti, and Philip S. Yu. 2022. A survey on deep learning event extraction: Approaches and applications. *IEEE Transactions on Neural Networks and Learning Systems* (2022), 1–21. https://doi.org/10.1109/TNNLS.2022.3213168

[29] Qian Li, Hao Peng, Jianxin Li, Jia Wu, Yuanxing Ning, Lihong Wang, Philip S. Yu, and Zheng Wang. 2022. Reinforcement learning-based dialogue guided event extraction to exploit argument relations. *IEEE ACM Trans. Audio Speech Lang. Process.* 30 (2022), 520–533. https://doi.org/10.1109/TASLP.2021.3138670

[30] Qian Li, Hao Peng, Jianxin Li, Congying Xia, Renyu Yang, Lichao Sun, Philip S. Yu, and Lifang He. 2022. A survey on text classification: From traditional to deep learning. *ACM Trans. Intell. Syst. Technol.* 13, 2, Article 31 (Apr. 2022), 41 pages. https://doi.org/10.1145/3495162

[31] Xueting Liao, Danyang Zheng, and Xiaojun Cao. 2021. Coronavirus pandemic analysis through tripartite graph clustering in online social networks. *Big Data Min. Anal.* 4, 4 (2021), 242–251. https://doi.org/10.26599/BDMA.2021.9020010

[32] Hongyu Lin, Yaojie Lu, Xianpei Han, and Le Sun. 2018. Nugget proposal networks for Chinese event detection. In *Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics, ACL 2018, Melbourne, Australia, July 15–20, 2018, Volume 1: Long Papers*. 1565–1574. https://doi.org/10.18653/v1/P18-1145

[33] Jian Liu, Yubo Chen, Kang Liu, and Jun Zhao. 2018. Event detection via gated multilingual attention mechanism. In *Proceedings of the Thirty-Second AAAI Conference on Artificial Intelligence, (AAAI-18), the 30th Innovative Applications of Artificial Intelligence (IAAI-18), and the 8th AAAI Symposium on Educational Advances in Artificial Intelligence (EAAI-18), New Orleans, Louisiana, USA, February 2–7, 2018*. 4865–4872. https://www.aaai.org/ocs/index.php/AAAI/AAAI18/paper/view/16371.

[34] Shengyuan Liu, Yuxuan Zhao, Zhenzhi Lin, Yilu Liu, Yi Ding, Li Yang, and Shimin Yi. 2020. Data-driven event detection of power systems based on unequal-interval reduction of PMU data and local outlier factor. *IEEE Trans. Smart Grid* 11, 2 (2020), 1630–1643. https://doi.org/10.1109/TSG.2019.2941565

[35] Yaopeng Liu, Hao Peng, Jianxin Li, Yangqiu Song, and Xiong Li. 2020. Event detection and evolution in multi-lingual social streams. *Frontiers Comput. Sci.* 14, 5 (2020), 145612. https://doi.org/10.1007/s11704-019-8201-6

[36] Dongfang Lou, Zhilin Liao, Shumin Deng, Ningyu Zhang, and Huajun Chen. 2021. MLBiNet: A cross-sentence collective event detection network. In *Proceedings of the 59th Annual Meeting of the Association for Computational Linguistics and the 11th International Joint Conference on Natural Language Processing, ACL/IJCNLP 2021, (Volume 1: Long Papers), Virtual Event, August 1–6, 2021*. 4829–4839. https://doi.org/10.18653/v1/2021.acl-long.373

[37] Yao Ma, Suhang Wang, Charu C. Aggarwal, Dawei Yin, and Jiliang Tang. 2019. Multi-dimensional graph convolutional networks. In *Proceedings of the 2019 SIAM International Conference on Data Mining, SDM 2019, Calgary, Alberta, Canada, May 2–4, 2019*. 657–665. https://doi.org/10.1137/1.9781611975673.74

[38] Qianren Mao, Xi Li, Hao Peng, Jianxin Li, Dongxiao He, Shu Guo, Min He, and Lihong Wang. 2021. Event prediction based on evolutionary event ontology knowledge. *Future Gener. Comput. Syst.* 115 (2021), 76–89. https://doi.org/10.1016/j.future.2020.07.041

[39] Thien Huu Nguyen and Ralph Grishman. 2018. Graph convolutional networks with argument-aware pooling for event detection. In *Proceedings of the Thirty-Second AAAI Conference on Artificial Intelligence, (AAAI-18), the 30th Innovative Applications of Artificial Intelligence (IAAI-18), and the 8th AAAI Symposium on Educational Advances in Artificial Intelligence (EAAI-18), New Orleans, Louisiana, USA, February 2–7, 2018.* 5900–5907. https://www.aaai.org/ocs/index.php/AAAI/AAAI18/paper/view/16329.

[40] Asiye K. Ozcanli, Fatma Yaprakdal, and Mustafa Baysal. 2020. Deep learning methods and applications for electrical power systems: A comprehensive review. *International Journal of Energy Research* 44, 9 (2020), 7136–7157.

[41] Hao Peng, Jianxin Li, Qiran Gong, Yangqiu Song, Yuanxing Ning, Kunfeng Lai, and Philip S. Yu. 2019. Fine-grained event categorization with heterogeneous graph convolutional networks. In *Proceedings of the Twenty-Eighth International Joint Conference on Artificial Intelligence, IJCAI 2019, Macao, China, August 10–16, 2019.* 3238–3245. https://doi.org/10.24963/ijcai.2019/449

[42] Hao Peng, Jianxin Li, Yangqiu Song, Renyu Yang, Rajiv Ranjan, Philip S. Yu, and Lifang He. 2021. Streaming social event detection and evolution discovery in heterogeneous information networks. *ACM Trans. Knowl. Discov. Data* 15, 5 (2021), 89:1–89:33. https://doi.org/10.1145/3447585

[43] Hao Peng, Jianxin Li, Senzhang Wang, Lihong Wang, Qiran Gong, Renyu Yang, Bo Li, Philip S. Yu, and Lifang He. 2021. Hierarchical taxonomy-aware and attentional graph capsule RCNNs for large-scale multi-label text classification. *IEEE Trans. Knowl. Data Eng.* 33, 6 (2021), 2505–2519. https://doi.org/10.1109/TKDE.2019.2959991

[44] Hao Peng, Ruitong Zhang, Shaoning Li, Yuwei Cao, Shirui Pan, and Philip Yu. 2022. Reinforced, incremental and cross-lingual event detection from social messages. *IEEE Transactions on Pattern Analysis and Machine Intelligence* (2022).

[45] Jeffrey Pennington, Richard Socher, and Christopher D. Manning. 2014. GloVe: Global vectors for word representation. In *Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing, EMNLP 2014, October 25–29, 2014, Doha, Qatar.* 1532–1543. https://doi.org/10.3115/v1/d14-1162

[46] Jiaqian Ren, Lei Jiang, Hao Peng, Zhiwei Liu, Jia Wu, and Philip S. Yu. 2022. Evidential temporal-aware graph-based social event detection via Dempster-Shafer theory. *IEEE ICWS* (2022).

[47] Sophia Boing Righetto, Marcos Aurelio Izumida Martins, Edgar Gerevini Carvalho, Leandro Takeshi Hattori, and Silvia de Francisci. 2021. Predictive maintenance 4.0 applied in electrical power systems. In *IEEE Power & Energy Society Innovative Smart Grid Technologies Conference, ISGT 2021, Washington, DC, USA, February 16–18, 2021.* 1–5. https://doi.org/10.1109/ISGT49243.2021.9372230

[48] Franco Scarselli, Marco Gori, Ah Chung Tsoi, Markus Hagenbuchner, and Gabriele Monfardini. 2009. The graph neural network model. *IEEE Trans. Neural Networks* 20, 1 (2009), 61–80. https://doi.org/10.1109/TNN.2008.2005605

[49] Mike Schuster and Kuldip K. Paliwal. 1997. Bidirectional recurrent neural networks. *IEEE Trans. Signal Process.* 45, 11 (1997), 2673–2681. https://doi.org/10.1109/78.650093

[50] Chuan Shi, Binbin Hu, Wayne Xin Zhao, and Philip S. Yu. 2019. Heterogeneous information network embedding for recommendation. *IEEE Trans. Knowl. Data Eng.* 31, 2 (2019), 357–370. https://doi.org/10.1109/TKDE.2018.2833443

[51] Michael J. Smith and Kevin Wedeward. 2009. Event detection and location in electric power systems using constrained optimization. In *2009 IEEE Power & Energy Society General Meeting*. IEEE, 1–6.

[52] Jian Tang, Meng Qu, and Qiaozhu Mei. 2015. PTE: Predictive text embedding through large-scale heterogeneous text networks. In *Proceedings of the 21st ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, Sydney, NSW, Australia, August 10–13, 2015.* 1165–1174. https://doi.org/10.1145/2783258.2783307

[53] Shikhar Vashishth, Soumya Sanyal, Vikram Nitin, and Partha P. Talukdar. 2020. Composition-based multi-relational graph convolutional networks. In *8th International Conference on Learning Representations, ICLR 2020, Addis Ababa, Ethiopia, April 26–30, 2020.* https://openreview.net/forum?id=BylA_C4tPr.

[54] Petar Velickovic, Guillem Cucurull, Arantxa Casanova, Adriana Romero, Pietro Liò, and Yoshua Bengio. 2018. Graph attention networks. In *6th International Conference on Learning Representations, ICLR 2018, Vancouver, BC, Canada, April 30–May 3, 2018, Conference Track Proceedings.* https://openreview.net/forum?id=rJXMpikCZ.

[55] Xiaozhi Wang, Xu Han, Zhiyuan Liu, Maosong Sun, and Peng Li. 2019. 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, NAACL-HLT 2019, Minneapolis, MN, USA, June 2–7, 2019, Volume 1 (Long and Short Papers).* 998–1008. https://doi.org/10.18653/v1/n19-1105

[56] Xiao Wang, Houye Ji, Chuan Shi, Bai Wang, Yanfang Ye, Peng Cui, and Philip S. Yu. 2019. Heterogeneous graph attention network. In *The World Wide Web Conference, WWW 2019, San Francisco, CA, USA, May 13–17, 2019.* 2022–2032. https://doi.org/10.1145/3308558.3313562

[57] Xiaozhi Wang, Ziqi Wang, Xu Han, Wangyi Jiang, Rong Han, Zhiyuan Liu, Juanzi Li, Peng Li, Yankai Lin, and Jie Zhou. 2020. MAVEN: A massive general domain event detection dataset. In *Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing, EMNLP 2020, Online, November 16–20, 2020.* 1652–1671. https://doi.org/10.18653/v1/2020.emnlp-main.129

[58] Zonghan Wu, Shirui Pan, Fengwen Chen, Guodong Long, Chengqi Zhang, and Philip S. Yu. 2021. A comprehensive survey on graph neural networks. *IEEE Trans. Neural Networks Learn. Syst.* 32, 1 (2021), 4–24. https://doi.org/10.1109/TNNLS.2020.2978386

[59] Jianye Xie, Haotong Sun, Junsheng Zhou, Weiguang Qu, and Xinyu Dai. 2021. Event detection as graph parsing. In *Findings of the Association for Computational Linguistics: ACL/IJCNLP 2021, Online Event, August 1–6, 2021 (Findings of ACL, Vol. ACL/IJCNLP 2021)*. 1630–1640. https://doi.org/10.18653/v1/2021.findings-acl.142

[60] Haoran Yan, Xiaolong Jin, Xiangbin Meng, Jiafeng Guo, and Xueqi Cheng. 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 2019, Hong Kong, China, November 3–7, 2019*. 5765–5769. https://doi.org/10.18653/v1/D19-1582

[61] Seongjun Yun, Minbyul Jeong, Raehyun Kim, Jaewoo Kang, and Hyunwoo J. Kim. 2019. Graph transformer networks. (2019), 11960–11970. https://proceedings.neurips.cc/paper/2019/hash/9d63484abb477c97640154d40595a3bb-Abstract.html.

[62] Jintao Zhang and Quan Xu. 2021. Attention-aware heterogeneous graph neural network. *Big Data Min. Anal.* 4, 4 (2021), 233–241. https://doi.org/10.26599/BDMA.2021.9020008

[63] Jianming Zheng, Fei Cai, Wanyu Chen, Wengqiang Lei, and Honghui Chen. 2021. Taxonomy-aware learning for few-shot event detection. In *WWW'21: The Web Conference 2021, Virtual Event/Ljubljana, Slovenia, April 19–23, 2021*. 3546–3557. https://doi.org/10.1145/3442381.3449949

[64] Jie Zhou, Ganqu Cui, Shengding Hu, Zhengyan Zhang, Cheng Yang, Zhiyuan Liu, Lifeng Wang, Changcheng Li, and Maosong Sun. 2020. Graph neural networks: A review of methods and applications. *AI Open* 1 (2020), 57–81. https://doi.org/10.1016/j.aiopen.2021.01.001

[65] Yuxun Zhou, Reza Arghandeh, and Costas J. Spanos. 2018. Partial knowledge data-driven event detection for power distribution networks. *IEEE Trans. Smart Grid* 9, 5 (2018), 5152–5162. https://doi.org/10.1109/TSG.2017.2681962

Received 18 January 2022; revised 10 July 2022; accepted 20 October 2022