KatGCN: Knowledge-Aware Attention based Temporal Graph Convolutional Network for Multi-Event Prediction

Xin Song, Haiyang Wang, Kang Zeng, Yujia Liu, and Bin Zhou*
College of Computer, National University of Defense Technology
Changsha,China
Email: {songxin, wanghaiyang19, zengkang, liuyujia, binzhou}@nudt.edu.cn

***Abstract*—Social events are due to gradually changing relations between entities including citizens, organizations, and national governments. Predicting multiple co-occurring events of different types in the future can help analysts understand social dynamics better and make quick and accurate decisions in advance. However, due to the overlook of the knowledge (e.g., event actors and different relations between them), existing methods are insufficient to model the structural and temporal dependence of events with different types simultaneously to better realize the prediction of future multiple co-occurring events. In the paper, we propose a novel Knowledge-aware attention based temporal Graph Convolutional Network (KatGCN) for predicting multiple co-occurring events of different types. We model social events as temporal event graph and extract static features (e.g., event background, topic keywords) from event content to enhance semantic of event graph. We design knowledge-aware attention based graph aggregation method to capture the structure dependence of co-occurring events with different types. We apply temporal encoding to capture the temporal dependence between temporally adjacent events. Empirical results on five-country datasets show that KatGCN outperforms state-of-the-art methods. Further studies verify the effectiveness and interpretability of our model.**

***Index Terms*—multi-event prediction, knowledge-aware attention, temporal Graph Convolutional Network**

I. Introduction

Social events such as protests, cooperation, and fights occur frequently and have a significant impact on society. It is highly desirable to predict multiple co-occurring events of different types, aka, multi-event, in advance to reduce the potential social upheaval and damage caused. Prior work [1] [2] mainly focused on predicting the scale of events or whether a given-type event will occur in the future. They have achieved good performance in the prediction of given-type events. However, as for multi-event prediction, they ignore the potential dependence of multiple co-occurring events given that different models are trained for different types of events.

Currently, social events are often extracted from news articles and structured as temporal knowledge graph with additional textual features [3], also called temporal event graph. As shown in Fig. 1, temporal event graph is a sequence of event graph in ascending time order. Each event graph with a timestamp is composed of multiple co-occurring events of different types, including event actors (as nodes) and event types (as edges). For instances, an event that occurred on $t-1$, was *Iran reject British*. Identifying knowledge of temporal event graph, such as event actors and their relations, can provide historical clues for predicting multiple co-occurring events of different types at the future timestamp t. In addition, event content also contains some important features (e.g., background information), such as *Gulf, oil tanker* etc. Incorporating such information can enhance the semantic expression of temporal event graph for better prediction.

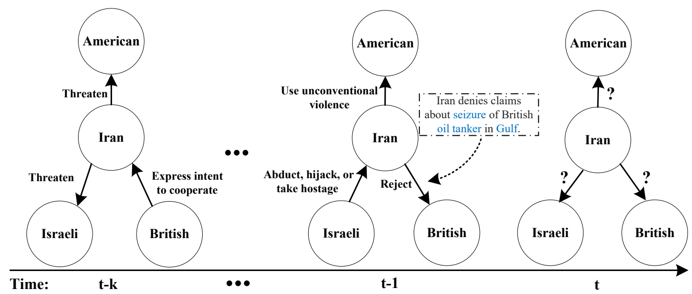


Fig. 1. An example of temporal event graph. It is composed of multiple co-occurring events of different types under different timestamps.

However, realizing such multi-event prediction problems in the real world faces many challenges:

- *C1*: Unstructured event content can enhance the semantic expression of event graphs. How to achieve heterogeneous data fusion is a challenging issue.
- *C2*: Multiple co-occurring events imply the structural dependence. How to adaptively model the local neighborhood information of event graphs to capture structural dependence remains a challenge.
- *C3*: Event types among actors in a temporal event graph change significantly over time. How to model the temporal dependence between temporally adjacent events with different types is also a key issue.

To address the aforementioned challenges, we proposed a novel **K**nowledge-aware **a**ttention based **t**emporal **G**raph **C**onvolutional **N**etwork (**KatGCN**) to predict multiple co-occurring events with different types in the future timestamp. Specifically, we model the social events from open source media as temporal event graph, and extract the background information and topic keywords from event content to enhance semantic. To capture the local structural dependence

*Corresponding author
DOI reference number:10.18293 / SEKE2021-089

of multiple co-occurring events, we design the knowledge-aware attention based graph aggregation method. Finally, we leverage the long-short term memory network to encode temporal dependence over temporal event graph for multi-event prediction.

Our contributions are summarized as follows:

- We design a novel knowledge-aware attention based graph aggregation method to capture the structural dependence of multiple co-occurring events.
- We develop a new model KatGCN for multi-event prediction, which integrates event content, structural dependences of event graphs and temporal dependence.
- We conduct extensive experiments on five-country datasets to verify the effectiveness of KatGCN and demonstrate the interpretability through a case study.

II. Related Work

Our work is closely related to many literatures on events prediction and knowledge graph learning.

A. Spatio-Temporal Event Prediction

Most existing machine learning methods for event prediction are only suitable for Euclidean or grid like data. For example, a linear regression model [4] utilized tweet frequency to predict the occurrence time of future events. Zhao et al [5] designed a new predictive model based on topic model that jointly characterizes temporal evolution in terms of both the semantics and geographical burstiness. Besides, more complex models, such as multi-task multi-class deep learning model (e.g., SIMDA [6], MITOR [1]), was proposed to predict the subtypes of future events and the scale of spatial events. Recently, Graph Convolutional Network (GCN [7]) has been proposed to address non-Euclidean data in many domains, such as social networks. For instance, DynamicGCN [2] was proposed to encode temporal text features into graphs for forecasting societal events and identifying their context graphs. Besides, REGNN [8] was proposed to learn the impact of historical actions and the surrounding environment on the current events for real-time event prediction.

B. Knowledge Graph Representation

Knowledge graphs (KG), which store real-world facts, is a form of multi-relation graphs. Since each fact changes over time, temporal knowledge graph (TKG) is generated.

Extensive studies have been done on modeling static, multi-relation graph data. For example, RGCN [9] was proposed to deal with the multi-relation graph directly by extending GCN, but it may face over-parameterization as the number of relations increases. Recently, attention mechanism has been applied to knowledge graph representation due to high efficiency and flexibility in modeling graph data. Wang et al [10] developed a novel model KGAT, which explicitly models the high-order connectivity in KG, propagating the embeddings from a node's neighbors to refine the node's embedding. Besides, RGHAT [11] was proposed to effectively utilize the neighborhood information of an entity. But the above methods aim to learn the embeddings of nodes and ignore the embeddings of relations. Therefore, CompGCN [12] was proposed to jointly embed both nodes and relations in a multi-relation graph by leveraging a variety of entity-relation composition operations from knowledge graph embedding techniques, which solves the over-parameterization problem.

There are also attempts to model TKG. RE-NET [13] was designed to predict future interactions. EvolveGCN [14] has been proposed for link prediction to capture the dynamism of the graph sequence through using an RNN to evolve the GCN parameters. In addition, a graph learning framework Glean [3] based on event knowledge graphs was developed to incorporate both relational and word contexts.

III. Methodology

We provide the technical details of our proposed model KatGCN. Fig. 2 shows an overview of KatGCN. The key objectives are (1) integrating the semantic features of event content into event graphs; (2) utilizing neighborhood information to capture the structural dependencies between multiple co-occurring events; (3) encoding temporal dependence over different timestamps for multi-event prediction.

A. Problem Definition

Temporal Event Graph (TE graph). TE graph is built on a sequence of event graphs in ascending time order [3]. Each event graph is a multi-relation directed graph with a timestamp, where entities represent event actors and relations represent event types. Let $\mathcal{E}$ be a finite set of entities (nodes) and $\mathcal{R}$ be a finite set of relations (edges). An event can be defined as a quadruple $(subject\ entity, event\ type, object\ entity)_t$, represented as $(s, r, o)_t$, where $s, o \in \mathcal{E}$ and $r \in \mathcal{R}$. We denoted a set of events at time t as $\mathcal{G}_t = \{(s, r, o)_t\}$. A TE graph can be presented as $\mathcal{G} = \{\mathcal{G}_{t-k}, \mathcal{G}_{t-k+1}, \ldots, \mathcal{G}_t\}$.

Problem Formulation. We transform the task of multi-event prediction into a multi-label classification problem to model the occurrence probability of different events at $t+1$:

$$\{\mathcal{G}_{t-k}, \mathcal{G}_{t-k+1}, \ldots \mathcal{G}_t\} \stackrel{\text{model}}{\rightarrow} \mathrm{P}\left(Y_{t+1} \mid \mathcal{G}_{t-k}, ..., \mathcal{G}_t\right) \quad (1)$$

Where $Y_{t+1} \in \mathbb{R}^{|\mathcal{R}|}$ is a vector of event types.

B. Semantic Enhancement

For challenge *C1* of Section I, we introduce the semantic enhancement module. We use the pre-trained model sent2vec [15] to get the initial embedding vector $\mathrm{h}_{(\bullet)} \in \mathbb{R}^d$ of entities, relations and keywords. However, entities, relations and event content are always closely related. Therefore, we extract background and topic keywords from the event content to enhance the semantic expression of event graphs.

1) Entity Semantic Enhancement: We introduce the entity semantic enhancement to incorporate backgrounds into event graphs. For instance, as shown in fig. 1, an event (*Iran Reject British*) mentioned that *Iran denies claims about seizure of British oil tanker in Gulf*. Words such as *oil tanker*, *Gulf*, show the event background, which can further enhance the semantic integrity of event graphs to improve prediction results.

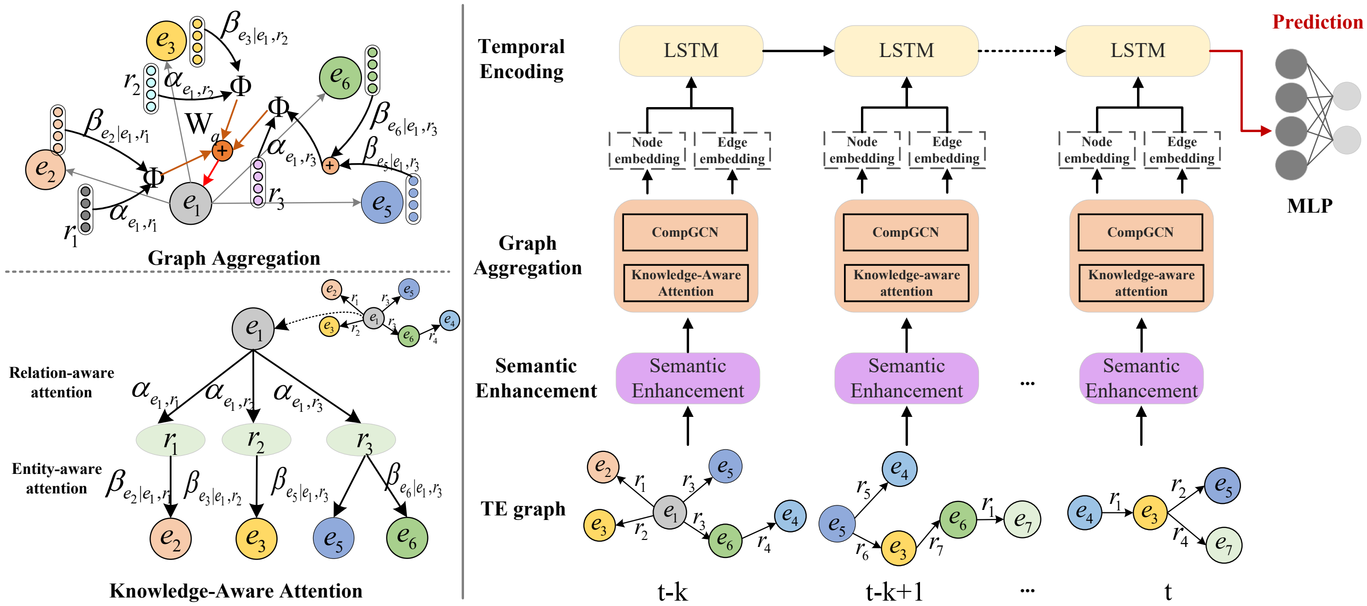


Fig. 2. System framework of the proposed model KatGCN for multi-event prediction. Input data consists of temporal event graph based on social events; We introduce semantic enhancement module to enrich the semantic information of event graph. Then, we design a knowledge-aware attention based graph aggregation method to capture the structural dependencies between multiple co-occurring events. Finally we feed the sequence of TE graph embedding into LSTM to capture the temporal dependence, and add a multi-layer perceptron (MLP) to predict the probability of co-occurring events at $t+1$.

For a given entity e in $\mathcal{G}_t$, we obtain the top ten relevant background words based on TF-IDF algorithm from all the event content at t to enhance the semantic of e, as follows:

$$\mathrm{h}'_{e,t} = \tanh\left(W_f \cdot \left[\mathrm{h}_{e,t} \ ; \sum_{\text{word } \in Top^{e}_{10,t}} \mathrm{h}_{\text{word}}\right]\right) \in \mathbb{R}^d \quad (2)$$

Where $W_f \in \mathbb{R}^{d\times 2d}$ is a learnable weight matrix, and ; is the concatenation operator. If e has no related words, we use zero vector to represent semantic. Then we can get the new entity embedding vector $\mathrm{h}'_{e,t}$ (includeing $\mathrm{h}'_{s,t}$ or $\mathrm{h}'_{o,t}$) at t.

2) Relation Semantic Enhancement: Event graphs contain many edges, which represent event types. Obviously, events with the same event type have similar topic keywords. For example, protest events usually include such words as *demonstrate*, *strike*, *disturbance*, etc. But Yield events usually contain such words as *surrender*, *ousted*, etc. To expand the difference between different event types for better relations embedding, we extract topic words for each event type based on LDA model [16] to enhance the semantic of relation r, as follows:

$$\mathrm{h}'_{r,t} = \tanh\left(W_k \cdot \left[\mathrm{h}_{r,t} \ ; \sum_{\text{word } \in L_{i,t} \& r \rightarrow L_i} \mathrm{h}_{\text{word}}\right]\right) \in \mathbb{R}^d \quad (3)$$

Where $\mathrm{L}_{i,t} = [l_{i,1}, \ldots, l_{i,n}]_t$ is i-th row of topic keywords matrix L_t generated from LDA model at t, which represents a set of keywords of n topics for i-th event type in G_t. Thus, we can get a new relation (edge) embedding vector $\mathrm{h}'_{r,t}$ at t.

C. Knowledge-Aware Attention based Graph Aggregation

For challenge *C2* of Section I, we design a novel knowledge-aware attention based graph aggregation method to fully capture the structural dependence between multiple co-occurring events.

1) Knowledge-Aware Attention: Considering that the event graph is a multi-relation graph, the embedding of relations (edges) cannot be ignored. Motivated by GAT [17], we propose a new knowledge-aware attention mechanism, including entity-aware attention and relation-aware attention, to distinguish the importance of neighboring entities and relations.

Relation-Aware Attention. Considering that different relations have different weights when expressing the same entity, we design relation-aware attention. For entity s, the relation-aware score represents the weight of each outgoing relation connected to the entity, defined as:

$$\mathbf{a}^t_{s,r} = Attention\left(W_1 \mathrm{h}'_{s,t}, W_1 \mathrm{h}'_{r,t}\right) \quad (4)$$

$$\alpha^t_{s,r} = \frac{\exp\left(\text{LeakyReL U}\left(m^{\mathrm{T}} \cdot \mathbf{a}^t_{s,r}\right)\right)}{\sum_{r_j \in N_s} \exp\left(\text{LeakyReLU}\left(m^{\mathrm{T}} \cdot \mathbf{a}^t_{s,r_j}\right)\right)} \quad (5)$$

Where $h'_{s,t},\ h'_{r,t} \in \mathbb{R}^d$ are the embedding vectors of entity s and relation r at t, respectively. W_1 and m are training parameters. N_s is a set of relations with s as the subject entity. The relation-aware attention score $\alpha_{s,r}$ represents the weights of outgoing relations r when representing the entity s.

Entity-Aware Attention The weights of neighboring entities under the same relation may also be different, which inspires the entity-aware attention. We design entity-aware attention to capture the difference in importance between different entities based on the same relation. We regard the object entities based on the same relation as a group, then we calculate the entity-aware attention score, which is defined as:

$$\mathbf{b}^t_{o|s,r} = Attention\left(W_2 \mathbf{a}^t_{s,r}, W_2 \mathrm{h}'_{o,t}\right) \quad (6)$$

$$\beta_{o|s,r}^{t}=\frac{\exp\left(\operatorname{LeakyReLU}\left(n^{\mathrm{T}}\cdot\mathbf{b}_{o|s,r}^{t}\right)\right)}{\sum_{o_j\in N_{s,r}}\exp\left(\operatorname{LeakyReLU}\left(n^{\mathrm{T}}\cdot\mathbf{b}_{o_j|s,r}^{t}\right)\right)} \tag{7}$$

Where $h'_{o,t}$ is the embedding of the entity o under relation r and entity s. $N_{s,r}$ represents a set of object entities of s under relation r. W_2 and n are training parameters. The entity-aware attention score $\beta_{o|s,r}^{t}$ shares all the object entities information of the same subject entity under the same relation at t, which is beneficial to capture the association between different co-occurring events under the same relation.

2) Graph Aggregation: The event graph is multi-relation directed graph. We need to get the embedding of entities and relations to get the event graph representation. Inspired by CompGCN [12], we design a novel knowledge-aware attention based CompGCN to learn the the event graph representation.

Specifically, We leverage the entity-relation composition operation [18] based on the knowledge-aware attention to incorporate the embedding of entities and relations into the GCN. For an entity s in $\mathcal{G}_t$, we apply the knowledge-aware attention based CompGCN to update its embedding vector:

$$\mathbf{h}_{s,t}^{\prime,(l+1)}=f\left(\sum_{(r,o)\in N(s)}W_q^{(l)}\Phi\left(\alpha_{s,r}^{t}\ \mathrm{h}_{r,t}^{\prime,(l)},\beta_{o|s,r}^{t}\ \mathrm{h}_{o,t}^{\prime,(l)}\right)\right) \tag{8}$$

Here, $\Phi:\mathbb{R}^d\times\mathbb{R}^d\to\mathbb{R}^d$ is a composition operator. We choose multiplication as Φ. $\mathrm{h}_{r,t}^{\prime,(l)}$ and $h_{o,t}^{\prime,(l)}$ denote feature embedding in l-th aggragation layer for relation r and entity o, respectively. W_q is a relation-specific parameter. $f(\bullet)$ is the ReLU activation function. Next, we update the embedding vector of relation r in $\mathcal{G}_t$:

$$\mathbf{h}_{r,t}^{\prime,(l+1)}=W_{rel}^{(l)}h_{r,t}^{\prime,(l)} \tag{9}$$

Where, $W_{rel}^{(l)}$ is a learnable transformation matrix in the l-th layer, which can project all the relations to the same embedding space as entities, so that the prediction task can perform operations on the nodes and edges uniformly.

To summarize, the advantage of our graph aggregation lies in distinguishing the importance of different neighboring entities and relations. We apply two layers to realize the aggregation of two-hop neighborhoods. For $\mathcal{G}_t$, we obtain the embedding matrix H_t^e of entities and H_t^r of relations.

D. Event Prediction

1) Temporal Encoding: For challenge *C3* of Section I, we utilize a temporal encoding module to capture temporal dependence between temporally adjacent events. Given a sequence of embedding matrix of entities and relations, i.e., $\left\{\mathrm{H}_{t-k:t}^{e},\mathrm{H}_{t-k:t}^{r}\right\}$, we apply LSTM to encode historical information in the TE graph, aiming to model temporal dependence from the graph sequence. To reduce the spatial of feature embeddings and obtain salient feature, we employ the max pooling operation over the embedding matrix of entities and relations, respectively. Then, we feed them into the LSTM model to get the historical global embedding X_t:

$$\mathrm{X_t}=\operatorname{LSTM}\left(\left[p\left(\mathrm{H}_t^e\right);p\left(\mathrm{H}_t^r\right)\right],\mathrm{X_{t-1}}\right) \tag{10}$$

Where, $p(\bullet)$ represents the max pooling operation applied element-wise over all nodes or edges.

2) Multi-event Prediction: Through temporal encoding, we have obtained the historical embedding X_t up to time t. Then, we model the probability of multiple co-occurring events in the future timestamp $t+1$ based on TE graph:

$$\mathrm{P}\left(Y_{t+1}\mid\mathcal{G}_{t-k},\ldots,\mathcal{G}_t\right)=\sigma\left(\mathrm{W}_\mu\mathrm{X}_t\right) \tag{11}$$

We feed the X_t into a MLP to calculate the probability of different event types. We define the MLP as a linear softmax classifier parameterized by W_μ. σ is a nonlinear function.

Next, we adopt the categorical cross-entropy [19] loss:

$$\mathcal{L}=-\frac{1}{|\mathcal{R}|}\sum_{i\in\mathcal{R}}y_i\ln\left(\frac{\exp\left(\hat{y}_i\right)}{\sum_{j\in\mathcal{R}}\exp\left(\hat{y}_j\right)}\right) \tag{12}$$

Where $\hat{y}_i$ is the model prediction for event type i before the nonlinear function (σ) in (11) .

IV. Experiments and Results

We evaluate the performance of KatGCN for multi-event prediction. We aim to answer the following key questions: (1) Whether KatGCN achieve satisfactory predicting results compared with other baselines; (2) Whether different modules in KatGCN can improve the experimental results better; (3) Whether the results of KatGCN have better interpretability.

A. Datasets and Evaluation Metrics

The experimental evaluation was conducted on the Global Database of Events, Language, and Tone event data (GDELT[1]). It contains political events designed to assess national and international crisis events. These events are divided into 20 main types and 220 subtypes such as Appeal, Yield, Protest etc. Each event is coded into 58 fields including date, actor attributes (actor1, actor2), event type, source (event URL) etc. In this paper, we focus on all subtypes of events and select country-level datasets from five countries (Iran, Iraq, Saudi Arabia, Syria, and Turkey) from January 1, 2018 to June 20, 2020. We split the dataset of each country into three subsets, i.e., train(80%), valid(10%), test(10%). The time granularity is one day. We use the three metrics to evaluate the results of the experiment, including F1-score, F2-score and Recall.

B. Comparative Methods

We compare KatGCN with some state-of-the-art baselines:

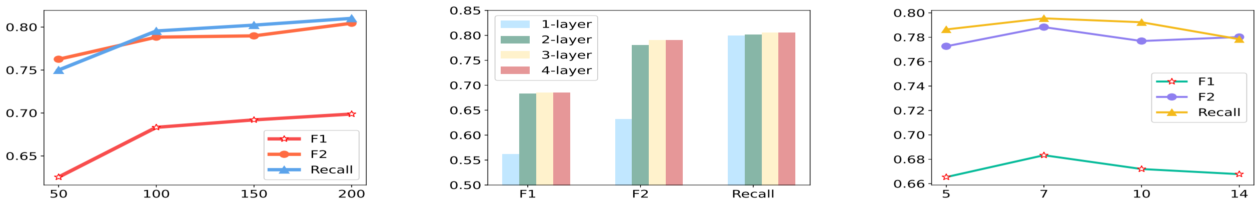
- **DNN:** We feed TF-IDF text features to a deep neural network for events prediction.
- **RE-NET [13]:** It contains a recurrent event encoder and a neighborhood aggregator to infer future facts.
- **Glean [3]:** This is a temporal graph learning method with heterogeneous data fusion for predicting multi-event.

Next, we conduct ablation studies:

[1]https://www.gdeltproject.org/

TABLE I
PREDICTION RESULTS OF KATGCN AND BASELINES OVER ALL DATASETS.

Method	Iran			Iraq			Saudi Arabia			Syria			Turkey		
	F1	F2	Recall	F1	F2	Recall	F1	F2	Recall	F1	F2	Recall	F1	F2	Recall
DNN	49.08	54.71	59.62	53.07	58.44	65.57	47.21	51.72	55.81	54.86	59.07	60.59	57.71	59.46	65.67
RE-NET	56.20	60.21	62.99	55.46	62.04	68.82	54.82	59.06	66.25	56.08	63.58	68.97	60.04	64.77	70.52
Glean	57.20	64.88	73.06	59.05	70.25	74.60	56.18	63.24	70.16	58.65	65.47	73.21	61.55	67.04	73.47
KatGCN-semantic	66.15	77.04	79.62	66.04	72.05	76.50	62.88	67.01	72.99	59.94	67.70	77.03	67.02	70.67	77.52
KatGCN-attention	60.54	68.02	71.93	64.09	69.03	75.29	59.01	62.55	69.55	57.90	62.54	71.20	63.05	68.34	74.37
KatGCN	**68.33**	**78.83**	**79.95**	**67.34**	**72.37**	**77.59**	**64.07**	**68.67**	**73.69**	**61.27**	**68.75**	**77.25**	**67.66**	**72.84**	**78.26**



(a) Embedding Dimensions(d)

(b) Layers of Aggregation(l)

(c) Time Step of History(k)

Fig. 3. Sensitivity Analysis.

- **KatGCN-semantic:** without the semantic enhancement module. We only consider the structure-graph information and temporal dependency in the TE graph.
- **KatGCN-attention:** without the knowledge-aware attention module. We only use the classical CompGCN to achieve event graph aggregation.

C. Experiments Results

We evaluate the prediction performance of our proposed model across five datasets. To avoid errors caused by randomness, we obtain an average of 10 experiments on each dataset. Table. I presents comparison and ablation results.

1) Prediction Performance: Our model KatGCN outperforms all other baselines on five datasets. From the above comparison results, we have the following observations:

- The difference of F-score and Recall across different datasets may be due to the different distribution of event types in each country.
- The DNN has the weakest performance than other methods, which shows that simple static features ignore the potential relations between events with different types and are less effective in multi-event prediction.
- KatGCN presents the best performance on the five datasets. The reason is that we introduce a knowledge-aware attention mechanism to make full use of the neighborhood information of the event graph.
- Graph based methods (RE-NET, Glean, KatGCN) are obviously better than static features based methods (DNN), which shows that the graphs can model structural dependence between different events (e.g., sharing actors).

2) Ablation Experiments: From the results of ablation experiments, we can observe the following findings:

- Overall, the results of the variant methods show poorer performance than KatGCN.
- The performance of KatGCN-attention drops significantly, which suggests knowledge-aware attention plays an important role in the performance improvement.
- The semantic enhancement is also essential, which can slightly improve the performance of multi-event prediction by enriching the semantics of event graphs.

D. Sensitivity Analysis

We study the parameter sensitivity analysis of KatGCN, mainly including: the embedding dimensions (d), layers of graph aggregation (l), and the time step of history (k):

1) Embedding Dimensions: We study how the embedding dimensions affect the model performance. As shown in Fig. 3(a), the performance improves obviously with d increases when d is below 100. Higher d will not bring significant performance improvement and may cost more training time.

2) Layers of graph aggregation: The number of layers l represents the hops of neighbors that nodes aggregate. Fig. 3(b) shows the impact of different l. Compared with 1-layer, 2-layer significantly improves the performance. But the performance is almost unchanged when l increases. We infer that there is overfitting due to the increase in parameters.

3) Time Step of History: We need to encode events information of past k time step. Fig. 3(c) illustrates the performance with different k. The performance reaches the best when k is 7. But larger k is not likely to go higher performance.

E. Case Study

We present a case to show how the proposed model identifies historical event information to predict multi-event in the future. Then we verity the interpretability of knowledge-aware attention based graph aggregation method via an example.

1) Identify historical events: We select a series of social events from the Iran datasets as a case. We utilize the TE graph of past week to successfully predict multiple events on January 16, 2020. As shown in Fig. 4, we describe a series of social events about *the shooting down of Ukraine International Airlines Flight*. We find that *student* initiated an event of *demonstrate or rally* on January 11. Then, *government criticize or denounce Ukrain* on January 12 and *citizen* try to *express intent or negotiate* to *Iran* and *threaten* the *Tehran* on January

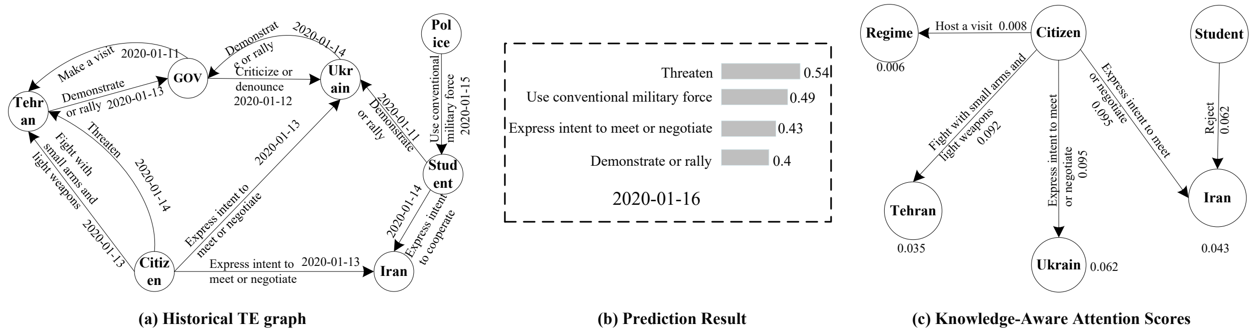


Fig. 4. An example temporal event graph about the social event *the shooting down of Ukraine International Airlines Flight* for our case study.

13 and 14 respectively. The events that occurred between different actors and temporal dependence were successfully captured by our model. In the prediction result, our model correctly predicted the possible events on January 16.

2) Interpretability: Benefiting from the knowledge-aware attention, we show the interpretability of our model. As shown in Fig. 4, We take some co-occurring events on January 13, 2020 as an example. we chose *citizen* as the central actor, and calculated different attention scores of neighboring relations and entities during aggregation. We observe that the event of *Fight with small arms and light weapons* has larger attention score. Besides, for the event of *Express intent to meet or negotiate*, the entity *Ukrain* has a larger attention score than Iran. This is the result we expected, and also consistent with the historical events and our prediction result.

V. Conclusion

In the paper, we propose a **K**nowledge-aware **a**ttention based **t**emporal **G**raph **C**onvolutional **N**etwork (KatGCN) for multi-event prediction. Specifically, we first model social events as TE graph and extract event background and topic keywords from event content to enhance semantic expression of event graphs. Then, we design a knowledge-aware attention based graph aggregation method to fully use the neighborhood information and capture structural dependency between co-occurring events. Finally, we utilize temporal encoding to capture temporal dependence between temporally adjacent events. Experiments on five-country datasets show KatGCN significantly outperforms the state-of-the-art baselines and has interpretability. Future work will consider predict event actors of different events to infer a complete social event.

Acknowledgment

This work is supported by the National Key R&D Program of China under Grants No.2018YFC0831703 and Postgraduate Scientic Research Innovation Project of Hunan Province CX20200015.

References

[1] Y. Gao and L. Zhao, "Incomplete label multi-task ordinal regression for spatial event scale forecasting," in *Proceedings of the AAAI Conference on Artificial Intelligence*, vol. 32, 2018.

[2] S. Deng, H. Rangwala, and Y. Ning, "Learning dynamic context graphs for predicting social events," in *Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining*, pp. 1007–1016, 2019.

[3] S. Deng, H. Rangwala, and Y. Ning, "Dynamic knowledge graph based multi-event forecasting," in *Proceedings of the 26th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining*, pp. 1585–1595, 2020.

[4] J. Bollen, H. Mao, and X. Zeng, "Twitter mood predicts the stock market," *Journal of computational science*, vol. 2, no. 1, pp. 1–8, 2011.

[5] L. Zhao, F. Chen, C.-T. Lu, and N. Ramakrishnan, "Spatiotemporal event forecasting in social media," in *Proceedings of the 2015 SIAM international conference on data mining*, pp. 963–971, SIAM, 2015.

[6] Y. Gao, L. Zhao, L. Wu, Y. Ye, H. Xiong, and C. Yang, "Incomplete label multi-task deep learning for spatio-temporal event subtype forecasting," in *Proceedings of the AAAI Conference on Artificial Intelligence*, vol. 33, pp. 3638–3646, 2019.

[7] T. N. Kipf and M. Welling, "Semi-supervised classification with graph convolutional networks," 2016.

[8] W. Luo, H. Zhang, X. Yang, L. Bo, X. Yang, Z. Li, X. Qie, and J. Ye, "Dynamic heterogeneous graph neural network for real-time event prediction," in *Proceedings of the 26th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining*, pp. 3213–3223, 2020.

[9] M. Schlichtkrull, T. N. Kipf, P. Bloem, R. Van Den Berg, I. Titov, and M. Welling, "Modeling relational data with graph convolutional networks," in *European semantic web conference*, pp. 593–607, Springer, 2018.

[10] X. Wang, X. He, Y. Cao, M. Liu, and T.-S. Chua, "Kgat: Knowledge graph attention network for recommendation," in *Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining*, pp. 950–958, 2019.

[11] Z. Zhang, F. Zhuang, H. Zhu, Z. Shi, H. Xiong, and Q. He, "Relational graph neural network with hierarchical attention for knowledge graph completion," in *Proceedings of the AAAI Conference on Artificial Intelligence*, vol. 34, pp. 9612–9619, 2020.

[12] S. Vashishth, S. Sanyal, V. Nitin, and P. Talukdar, "Composition-based multi-relational graph convolutional networks," 2019.

[13] W. Jin, M. Qu, X. Jin, and X. Ren, "Recurrent event network: Autoregressive structure inference over temporal knowledge graphs," in *Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, pp. 6669–6683, 2020.

[14] A. Pareja, G. Domeniconi, J. Chen, T. Ma, T. Suzumura, H. Kanezashi, T. Kaler, T. Schardl, and C. Leiserson, "Evolvegcn: Evolving graph convolutional networks for dynamic graphs," in *Proceedings of the AAAI Conference on Artificial Intelligence*, vol. 34, pp. 5363–5370, 2020.

[15] P. Gupta, M. Pagliardini, and M. Jaggi, "Better word embeddings by disentangling contextual n-gram information," in *NAACL-HLT (1)*, pp. 933–939, Association for Computational Linguistics, 2019.

[16] D. Blei, A. Ng, and M. I. Jordan, "Latent dirichlet allocation," *J. Mach. Learn. Res.*, vol. 3, pp. 993–1022, 2003.

[17] P. Veličković, G. Cucurull, A. Casanova, A. Romero, P. Lio, and Y. Bengio, "Graph attention networks," *arXiv preprint arXiv:1710.10903*, 2017.

[18] M. Nickel, K. Murphy, V. Tresp, and E. Gabrilovich, "A review of relational machine learning for knowledge graphs," *Proceedings of the IEEE*, vol. 104, no. 1, pp. 11–33, 2015.

[19] A. K. Menon, A. S. Rawat, S. Reddi, and S. Kumar, "Multilabel reductions: what is my loss optimising?," 2019.