DGCN-rs: A Dilated Graph Convolutional Networks Jointly Modelling Relation and Semantic for Multi-event Forecasting

Xin Song[1], Haiyang Wang[1], and Bin Zhou[2](✉)

[1] College of Computer, National University of Defense Technology, Changsha, China
{songxin,wanghaiyang19}@nudt.edu.cn

[2] Key Lab. of Software Engineering for Complex Systems, School of Computer, National University of Defense Technology, Changsha, China
binzhou@nudt.edu.cn

Abstract. Forecasting *multiple co-occurring events of different types* (a.k.a. *multi-event*) from open-source social media is extremely beneficial for decision makers seeking to avoid, control related social unrest and risks. Most existing work either fails to jointly model the entity-relation and semantic dependence among multiple events, or has limited long-term or inconsecutive forecasting performances. In order to address the above limitations, we design a Dilated Graph Convolutional Networks (DGCN-rs) jointly modelling relation and semantic information for multi-event forecasting. We construct a temporal event graph (TEG) for entity-relation dependence and a semantic context graph (SCG) for semantic dependence to capture useful historical clues. To obtain better graph embedding, we utilize GCN to aggregate the neighborhoods of TEG and SCG. Considering the long-term and inconsecutive dependence of social events over time, we apply dilated casual convolutional network to automatically capture such temporal dependence by stacked the layers with increasing dilated factors. We compare the proposed model DGCN-rs with state-of-the-art methods on five-country datasets. The results exhibit better performance than other models.

Keywords: Multi-event forecasting · Relation and semantic · GCN · Dilated casual convolutional network

1 Introduction

Social events initiated by different organizations and groups are common in reality, and they are usually *co-occurring and of different types*, a.k.a. *multi-event*, such as *rally broke out at different cities simultaneously*, or *protests happened periodically within several weeks or months*, etc. The ability to successfully forecast multi-event thus be extremely beneficial for decision makers seeking to avoid, control, or alleviate the associated social upheaval and risks. Traditional methods tend to focus on the prediction of some single type of event or event scale

T. Mantoro et al. (Eds.): ICONIP 2021, LNCS 13110, pp. 666–676, 2021.
https://doi.org/10.1007/978-3-030-92238-2_55

[2,4]. The most intuitive way for *multi-event* forecasting is that a single event type forecasting method is repeated for different types. Although this approach simplifies the multi-event forecasting model, it ignores the potential relation and semantic dependence among multi-event.

Recently, researchers [3] attempt to model *multi-event* as *temporal event graph* (TEG), which can be viewed as a sequence of event graphs splited in time ascending order. Figure 1 shows an example subgraph of a TEG, the event (*Protester, Demonstrate or rally, Isfahan Gov, 2018-04-23*) mentioned *Escalation of Protests in Iran's Drought-Ridden Isfahan Province*. The previous events contain some useful information to provide historical clues for future prediction: (1) potential relation and semantic information in periodically and correlated events, such as rallies on 2018-04-11 and 2018-04-23; (2)temporal dependence, including the long temporal range (from *2018-04-05* to *2018-04-23*) and inconsecutive temporal dependence across different time intervals. How to identify such information to help predict multi-event in the future is still challenging.

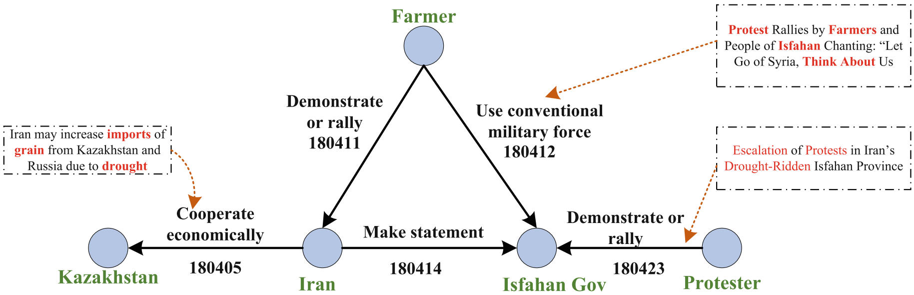


Fig. 1. An example subgraph of temporal event graph, where entities denote event actors and relations (edges) with a timestamp denote event types between two entities.

Obviously, the relation information between entities and the semantic information of events are inseparable, and jointly modeling them is vital for multi-event forecasting. Existing methods [7] consider relation information but ignore semantic information. Besides, some methods [18] using the pre-defined feature of events without dynamic development of events over time. Recently, graph features [2] constructed from text data have proven to be advantageous. In the paper, we extract event keywords under each timestamp to jointly construct a semantic context graph (SCG). Then, we introduce knowledge-aware attention based GCN [15] to capture the relation dependence of TEG at each timestamp for better relation embedding. Meanwhile, we utilize GCN to capture the semantic relevance of SCG at each timestamp for better semantic embedding.

In addition to preserve relation and semantic dependence, temporal dependence is also important, such as long-term or inconsecutive dependence with different temporal intervals. However, traditional methods [9,12] based on RNN, usually learn all features of historical timestamps equally, which is limited to capture the inconsecutive dependence across different time intervals and long-term dependence. On the contrary, CNN-based methods, such as dilated convolution, enjoy the advantages of parallel computing and can capture very long sequences by stacking layers. To ensure the sequence of data, some methods [14] combine dilated convolution with causal convolution. Hence, we introduce a dilated casual convolution(DCC) to capture the inconsecutive dependency by setting different dilated factors and capture the long-term dependencies by stacked convolution layers. Our contributions are summarized as follows:

1 We construct the TEG and SCG to jointly model the relation and semantic dependence of social events, which is helpful to capture the dynamic changes of events over time and provide historical clues for future forecasting.
2 We design a dilated casual convolution network to automatically capture temporal dependence among events, including long-term and inconsecutive dependence by stacked the layers with increasing dilated factors.
3 We evaluate the proposed model on five-country datasets with state-of-the-art models and demonstrate the effectiveness of our model by ablation test and sensitivity analysis.

2 Related Work

There has been extensive research to predict social events. Linear regression models use the frequency of tweets to predict when future events will occur [1]. More sophisticated features (e.g., topic-related keywords [13]) and multi-task multi-class deep learning models (for example, SIMDA [5], MITOR [4]) to predict sub-types and the scale of spatial events. Recently, GCN-based methods have also been studied. For example, DynamicGCN [2], REGNN [10], etc. Recently, the temporal knowledge graph modeling methods have applied in event forecasting. Existing work mainly follows two directions: The first is a combination of GNN and RNN, for example, glean [3], Renet [7], etc. Then, some more complex methods are studied, such as DCRNN [9], GaAn [17]. However, the RNN-based method is inefficient for long sequences and its gradient may explore when they are combined with GCNs. Therefore, CNN-based methods are proposed to encode temporal information, which expand the fields of neural network by stacking many layers or using global pooling. For example, STGCN [16], Graph WaveNet [14], etc.

3 Methodology

Figure 2 provides the system framework of DGCN-rs. Firstly, we construct TEG and SCG to jointly model social events. Then we utilize the knowledge-aware attention based GCN and GCN to aggregating neighborhood information of TEG and SCG at each timestamp, respectively, for better relation and semantic embedding. To capture the temporal dependence in TEG and SCG, such as the long-term or inconsecutive dependence, we design a dilated casual convolution for temporal embedding and obtain the global historical embedding X_t. We finally add a (MLP) to predict the probability of co-occurring events at $t + 1$.

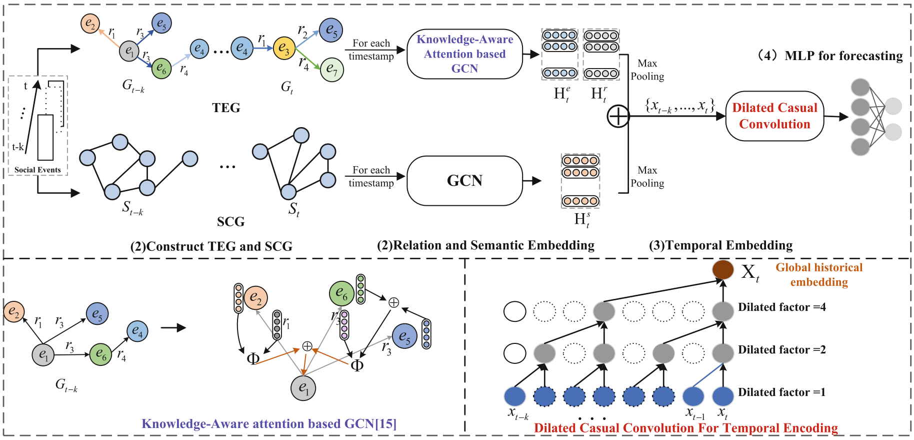


Fig. 2. System framework of our proposed model DGCN-rs.

3.1 Preliminaries

Temporal Event Graph (TEG). We construct a TEG based on event entities, event types and time to model the relation information between events. A TEG is defined as $\mathcal{G} = \{\mathcal{G}_{t-k}, \dots, \mathcal{G}_{t-1}, \mathcal{G}_t\}$, where $\mathcal{G}_t$ is a multi-relation, directed graph with a time-stamped edges (event types) between nodes (entities). Let $\mathcal{E}$ and $\mathcal{R}$ be the finite set of nodes and edges. Then, a social event is defined as a quadruple (s, r, o, t) or $(s, r, o)_t$. where $s, o \in \mathcal{E}$ and $r \in \mathcal{R}$ and the direction of edge is pointing from s to o. We denoted a set of events at time t as $\mathcal{G}_t = \{(s, r, o)_t\}$.

Semantic Context Graph (SCG). We construct a SCG based on the event content text to model the context semantic information. A SCG is defined as $\mathcal{S} = \{\mathcal{S}_{t-k}, \dots, \mathcal{S}_{t-1}, \mathcal{S}_t\}$, where each semantic graph $\mathcal{S}_t$ is a simple undirected graph, where nodes denote keywords extracted by removing very common words and rare words from the content text, the edges are based on word co-occurrence in the collection of content texts at timestamp t. The edges weights between two nodes are calculated by PMI [6] to measure the semantic relevance between words. We define there is no edge between word i and j when the $PMI(i, j) < 0$.

Problem Formulated. Given an observed TEG and SCG, we aim to encode the historical social events and learn a function f which can forecast the co-occurring probability of different event types at the future timestamp $t + 1$, denoted as:

$$\{\mathrm{G}_{t-k:t}, S_{t-k:t}\} \overset{encode}{\longrightarrow} \mathrm{X}_t \overset{model}{\longrightarrow} \mathrm{P}\left(\mathrm{Y}_{t+1} \mid \mathrm{G}_{t-k:t}, S_{t-k:t}\right) \tag{1}$$

where $Y_{t+1} \in R^{|\mathcal{R}|}$is a vector of event types. X_t is a global historical embedding based on all social events of k historical consecutive time steps.

3.2 GCN for Relation and Semantic Embedding

After constructing the TEG and SCG, we utilize the knowledge-aware attention based GCN [15] and GCN [8] to aggregate the local neighborhood information of TEG and SCG, respectively, so that we can capture the relation dependence between entities and the semantic relevance between words.

Relation Embedding. Each event graph $\mathcal{G}_t$ expresses relation information among *multi-event*. Based on our previous work [15], we utilize knowledge-aware attention based GCN to effectively aggregate neighborhood information of an entity s, which can distribute different importance scores to neighboring relations and neighboring entities:

$$\mathrm{h}_s^{(l+1)} = f\left(\sum_{(r,o)\in N(s)} W_q^{(l)} \varPhi\left(\alpha_{s,r}\mathrm{h}_r^{(l)}, \beta_{o|s,r}\mathrm{h}_o^{(l)}\right)\right) \tag{2}$$

Here, $\varPhi : R^d \times R^d \rightarrow R^d$ is a composition operator. We choose multiplication as $\varPhi$. $\mathrm{h}_r^{(l)}$ and $\mathrm{h}_o^{(l)}$ denote feature embedding in l-th layer for relation r and entity o, respectively. W_q is a relation-specific parameter. $f(\bullet)$ is the ReLU activation function. The $\alpha_{s,r}$ represents the weight of relation r and $\beta_{o|s,r}$ represents the weight of object entity o under the same s and r, which are calculated as follows:

$$\alpha_{s,r} = \frac{\exp\left(\sigma\left(m^{\mathrm{T}} \cdot \mathrm{W}_1\left(h_s, h_r\right)\right)\right)}{\sum_{r_j\in N_s} \exp\left(\sigma\left(m^{\mathrm{T}} \cdot \mathrm{W}_1\left(h_s, h_{r_j}\right)\right)\right)} \tag{3}$$

$$\beta_{ols,r} = \frac{\exp\left(\sigma\left(n^{T} \cdot \mathrm{W}_2\left(\left(h_s, h_r\right), h_o\right)\right)\right)}{\sum_{o_j\in N_{s,r}} \exp\left(\sigma\left(n^{\mathrm{T}} \cdot \mathrm{W}_2\left(\left(h_s, h_r\right), h_{o_j}\right)\right)\right)} \tag{4}$$

where σ is LeakyRelu and W_1 and W_2 are the trainable parameter. Next, we update the embedding vector of relation r:

$$\mathrm{h}_r^{(l+1)} = W_{rel}^{(l)}\mathrm{h}_r^{(l)} \tag{5}$$

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. By calculating the embedding of all entities and relations at each timestamp in TEG, we can get a matrix sequence $\left\{\left[H_{t-k}^{e}, H_{t-k}^{r}\right], \ldots, \left[H_t^e, H_t^r\right]\right\}$.

Semantic Embedding. In addition to relation embedding, the content text of social events are also indispensable when describing social events. Generally, each event is attached with a paragraph of text describing more semantic information. For example, in the introduction example Fig. 1, the event content of the quadruple (*Iran, Cooperate economically, Kazakhstan, 2018-04-05*) mentioned *Iran may increase imports of grain from Kazakhstan and Russia due to drought*, which contains more semantic information from keywords, such as *drought*, *imports*, *grain* and the event semantic are changing with time. Thus, we construct a SCG and utilize traditional GCN to capture the semantic relevance between words more deeply and obtain a better semantic context of events, as follows:

$$H^{(l+1)} = g\left(\hat{A}H^{(l)}W^{(l)} + b^{(l)}\right) \tag{6}$$

where l denotes the layer number. $\hat{A} = \tilde{D}^{-\frac{1}{2}}\tilde{A}\tilde{D}^{-\frac{1}{2}}$ is the normalized symmetric adjacency matrix and $\tilde{A} = A + I_N$. $\tilde{D}$ is the degree matrix. $H_t^{(l)}$ is the node embedding matrix passing by GCN layers. Intuitively, nodes aggregate semantic information from their local neighbors. As the number of layers increases, nodes can receive the semantics of further words from SCG. Finally, we can get a matrix sequence $\left\{H_{t-k}^s, \ldots, H_{t-1}^s, H_t^s\right\}$ to describe the event semantic context.

Relation and Semantic Fusion. We have obtained a sequence of embeddings of entities, event types, as well as the semantic embedding, represented as $\left\{\left[H_{t-k}^e, H_{t-k}^r, H_{t-k}^s\right], \ldots, \left[H_t^e, H_t^r, H_t^s\right]\right\}$. We first integrate above information at each timestamp t, as follows:

$$x_t = \left[p\left(H_t^e\right) : p\left(H_t^e\right) : p\left(H_t^e\right)\right] \tag{7}$$

where: denotes the concatenate operation. p represents the max pooling operation to reduce the dimension of the feature embedding matrix.

3.3 Dilated Casual Convolutions for Temporal Embedding

In addition to the relation and semantic information of social events, events also preserve temporal dependence, such as long-term or inconsecutive dependence with different tintervals. For the example of introduction in Fig. 1, a series of social event caused by *drought and water shortage* have two characteristics in temporal dependence: (1) The involved temporal range is relatively long, from *2018-04-05* to *2018-04-23*, and may continue until the end of the drought. (2) The time of social events is not consecutive and there are different time intervals.

Inspired by WaveNet [14], we apply a CNN-based method to capture such temporal dependence, named dilated casual convolution (DCC). The receptive field of DCC increases exponentially with the increase of the hidden layer, so we can capture the long-term dependence in social events by increasing the number of hidden layers. Based on x, we utilize the dilated causal convolution with filter f to encode temporal dependence:

$$X_t = \sum_{k=0}^{K-1} x(t - d \times k)f(k) \tag{8}$$

where d is the dilated factor, which means the input is selected every d time steps (or $d-1$ time intervals). s is the kernel size of the filter. To capture the long-term dependence and the periodic dependence across different time intervals, we stack the dilated causal convolutional layers in the order of increasing the dilation factor, which enable our model to capture longer sequences with less layers and save computation resources. Finally, we get the global historical embedding X_t, which includes three aspects of information: relation information of multi-event, semantic context information, long-term and periodic temporal dependence across different intervals

3.4 Multi-event Forecasting

Forecasting. We have obtained the historical embedding X_t up to time t. Then, we transform the task of multi-event forecasting into a multi-label classification problem to model the occurrence probability of different events at $t+1$:

$$\mathrm{P}\left(Y_{t+1} \mid G_{t-k:t}\right) = \sigma\left(\mathrm{W}_{\mu}\mathrm{X}_{t}\right) \tag{9}$$

where $Y_{t+1} \in R^{|\mathcal{R}|}$ is a vector of event types. $\mathrm{X}_{t-k:t}$ is the global historical embedding up to t. We further feed $\mathrm{X}_{t-k:t}$ into a multi-layer perceptron (MLP) to calculate the probability of different event types. σ is a nonlinear function.

Optimization. We optimize the categorical cross-entropy loss [11]:

$$\mathcal{L} = -\frac{1}{L}\sum_{i\in L} y_i \ln\left(\frac{\exp\left(\hat{y}_i\right)}{\sum_{j\in L}\exp\left(\hat{y}_j\right)}\right) \tag{10}$$

where $\hat{y}_i$ is the model prediction for event type i before the nonlinear function σ. $L \in N_+$ represents the total number of Event types.

4 Experiments and Results

We evaluate the performance of DGCN-rs for multi-event forecasting. Our results indicate that our model achieves significant performance gains.

4.1 Datasets and Settings

The experimental evaluation was conducted on the Global Database of Events, Language, and Tone event data (GDELT). These events are coded using 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 types of events and select country-level datasets from five countries (Iran, Iraq, Saudi Arabia, Syria, and Turkey). We choose a wide range time period social events from January 1, 2018, to June 20, 2020. In our experiments, The time granularity is one day.

In our experiments, we pre-train a 100-dimensional sent2vec [6] for entities, relations and keywords in the vocabulary using the event content from each country. Then, we split the dataset of each country into three subsets, i.e., train(80%), valid(10%), test(10%). For hyper-parameter setting, the historical time step (day) k is set to 14 but for Glean and RE-NET,the k is 7. The layers of GCN are set to 2. All the parameters are trained using the Adam optimizer. For dilated casual convolution, dilated factor d is set 1,2,4.

4.2 Baselines

We compare our methods with several state-of-the-art baselines as follows:

- **DNN:** It consist of three dense layers. We use non-temporal TF-IDF text features extracted from all the event contents in historical time steps.
- **RE-NET** [7]**:** It contains a recurrent event encoder and a neighborhood aggregator to infer future facts.
- **Glean** [3]**:** A temporal graph learning method with heterogeneous data fusion for predicting co-occurring events of multiple types.

We also conduct some ablation tests:

- −**attention**: we remove the knowledge-aware attention, and only use the basic CompGCN to get the TEG embedding matrix at each timestamp.
- −**semantic**: we remove the SCG, only utilizing the relation embedding based on TEG as the global historical information.
- **+GRU**: we utilize GRU instead of DCC to capture temporal dependence.
- **+RNN**: similar to +GRU, we change the GRU module to a RNN module.
- **+LSTM**: we replace the GRU module with a simpler LSTM module.

Table 1. Forecasting results of our proposed model DGCN-rs

Method	Iran		Iraq		Turkey		Syria		Saudi Arabia	
	F1	Recall	F1	Recall	F1	Recall	F1	Recall	F1	Recall
DNN	49.08	59.62	53.07	65.57	57.71	65.67	54.86	60.59	47.21	55.81
RE-NET	56.20	62.99	55.46	68.82	60.04	70.52	56.08	68.97	54.82	66.25
Glean	57.20	73.06	59.05	74.60	61.55	73.47	58.65	73.21	56.18	70.16
−attention	70.31	82.38	62.45	80.65	**69.96**	**87.72**	62.76	85.46	**69.62**	**87.85**
−semantic	69.21	80.64	63.24	79.56	67.98	82.08	63.33	83.95	67.71	83.53
+GRU	65.53	73.07	60.04	74.41	68.04	81.72	61.73	80.07	67.59	84.58
+RNN	65.58	75.31	60.69	75.52	68.64	83.98	61.29	78.00	67.07	82.77
+LSTM	65.50	73.25	60.02	73.94	67.03	79.21	61.24	77.64	66.58	81.65
DGCN-rs	**71.60**	**84.95**	**64.87**	**83.07**	69.41	85.26	**64.56**	**85.42**	69.00	86.93

4.3 Experiments Results

We evaluate forecasting performance of our proposed model across datasets of five countries. Table 1 presents comparison and ablation results.

Forecasting Performance. From the comparison results, we observe:

- Overall, our model outperforms other baselines across all datasets in terms of F1-score and recall. The result difference of different datasets may be due to the different distribution of event types in each country, leading to strength differences in relation and semantic dependence between multi-event.

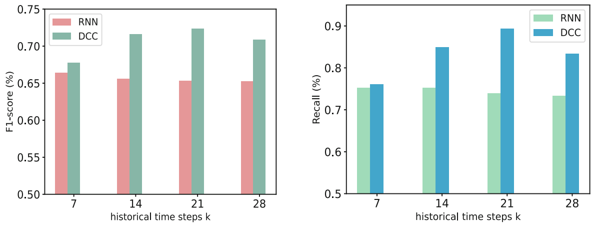
– The classic deep learning model DNN perform poorly, which shows ignoring graph structure and time dependence, and only considering summation of static historical text features is less effective for multi-event forecasting.
– The RE-NET model utilizes the temporal event graph and produces better performance than DNN, which indicates the relation dependence and temporal dependence of TEG can be helpful for multi-event forecasting.
– The Glean model achieves better results than RE-NET, because glean adds event semantic, which indicates that both relation and semantic information are indispensable for multi-event forecasting.
– Our model achieves the best performance. There may be three reasons: 1) we introduce the knowledge attention when capturing the relation dependence of TEG; 2) We built SCG, which can capture the semantic context of social event for better forecasting; 3) we utilize a DCC instead of RNN to capture the temporal dependence. Next, we will use the results of ablation tests to verify the above reasons.

Ablation Tests. From above ablation results, we have following observations:

– The difference in F-score and Recall of different datasets may be due to the different distribution of event types in each country, leading to strength differences in the relation and semantic dependence of multi-event.
– The results of -attention drop slightly, especially in the Turkey and Saudi Arabie dataset. The reason is that capturing unique temporal dependence is more important than the aggregation ways of local neighbors for relation embedding when forecasting long-term sequence data.
– The semantic is also essential, which can improve the performance of multi-event forecasting by enriching the context semantic of social events.
– The results of +GRU, +RNN, or +LSTM drops significantly and the performance of GRU and LSTM is not better than a simple RNN. The reason may be they increase the number of parameters leading to more complexity. Besides, we find the performance of our method is better than traditional temporal encoding methods. The likely reason is that it can capture long-term and inconsecutive temporal dependence with different intervals.

4.4 Sensitivity Analysis

We mainly conduct sensitivity analysis on the important parameter k (Historical Time Step) in temporal embedding module to prove that our model performs the advantages of our model for long-term forecasting. We evaluated the difference in forecasting performance of historical time steps from 7 to 28 days. The results of F1-score and Recall are shown in Fig. 3. We observe DCC is more suitable for capturing long-term dependence among events. The performance improves with k becoming larger. However, when we use RNN, the performance drop slightly as the k becomes larger. This result indicates that social events hide potential long-term temporal dependence and utilizing DCN is more conducive to improving the forecasting performance of such events.



(a) (b)

Fig. 3. Sensitivity analysis.

5 Conclusion

In the paper, we propose a novel dilated graph convolutional networks jointly modelling relation and semantic for multi-event forecasting. We construct a temporal event graph (TEG) and semantic context graph (SCG) based on social events to model the entity-relation and semantic information among multi-event. We employ our previous work of knowledge-aware attention based GCN to capture the relation dependence in TEG for better relation embedding. We use GCN to capture the semantic dependence in SCG for better semantic embedding. Considering the limitations of traditional temporal (such as RNN) encoding methods in capturing long-term dependence, we use dilated casual convolutional network to automatically capture the long-term and inconsecutive temporal dependence with different intervals between social events by stacking layers with increased dilated factors. Finally, we conduct extensive experiments to prove that our model outperforms other baselines. We also verify the effectiveness of our model through ablation experiments and sensitivity analysis. In the future, we will study more scalable methods to apply our proposed model on larger datasets in other domains.

Acknowledgement. This work was supported by the National Key Research and Development Program of China No. 2018YFC0831703.

References

1. Bollen, J., Mao, H., Zeng, X.: Twitter mood predicts the stock market. J. Comput. Sci. **2**(1), 1–8 (2011)
2. Deng, S., Rangwala, H., Ning, Y.: 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. Deng, S., Rangwala, H., Ning, Y.: 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. Gao, Y., Zhao, L.: Incomplete label multi-task ordinal regression for spatial event scale forecasting. In: Proceedings of the AAAI Conference on Artificial Intelligence, vol. 32 (2018)

5. Gao, Y., Zhao, L., Wu, L., Ye, Y., Xiong, H., Yang, C.: 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)
6. Gupta, P., Pagliardini, M., Jaggi, M.: Better word embeddings by disentangling contextual n-gram information. In: NAACL-HLT (1), pp. 933–939. Association for Computational Linguistics (2019)
7. Jin, W., Qu, M., Jin, X., Ren, X.: Recurrent event network: autoregressive structure inference over temporal knowledge graphs (2019)
8. Kipf, T.N., Welling, M.: Semi-supervised classification with graph convolutional networks (2016)
9. Li, Y., Yu, R., Shahabi, C., Liu, Y.: Diffusion convolutional recurrent neural network: data-driven traffic forecasting (2017)
10. Luo, W., et al.: 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)
11. Menon, A.K., Rawat, A.S., Reddi, S., Kumar, S.: Multilabel reductions: what is my loss optimising? (2019)
12. Pareja, A., et al.: EvolveGCN: evolving graph convolutional networks for dynamic graphs. In: Proceedings of the AAAI Conference on Artificial Intelligence, vol. 34, pp. 5363–5370 (2020)
13. Wang, X., Gerber, M.S., Brown, D.E.: Automatic crime prediction using events extracted from twitter posts. In: Yang, S.J., Greenberg, A.M., Endsley, M. (eds.) SBP 2012. LNCS, vol. 7227, pp. 231–238. Springer, Heidelberg (2012). https://doi.org/10.1007/978-3-642-29047-3_28
14. Wu, Z., Pan, S., Long, G., Jiang, J., Zhang, C.: Graph wavenet for deep spatial-temporal graph modeling (2019)
15. Song, X., Wang, H., Zeng, K., Liu, Y., Zhou, B.: KatGCN: knowledge-aware attention based temporal graph convolutional network for multi-event prediction. In: SEKE, pp. 417–422 (2021)
16. Yu, B., Yin, H., Zhu, Z.: Spatio-temporal graph convolutional networks: a deep learning framework for traffic forecasting (2017)
17. Zhang, J., Shi, X., Xie, J., Ma, H., King, I., Yeung, D.Y.: Gaan: gated attention networks for learning on large and spatiotemporal graphs (2018)
18. Zhao, L., Sun, Q., Ye, J., Chen, F., Lu, C.T., Ramakrishnan, N.: Multi-task learning for spatio-temporal event forecasting. In: Proceedings of the 21th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, pp. 1503–1512 (2015)