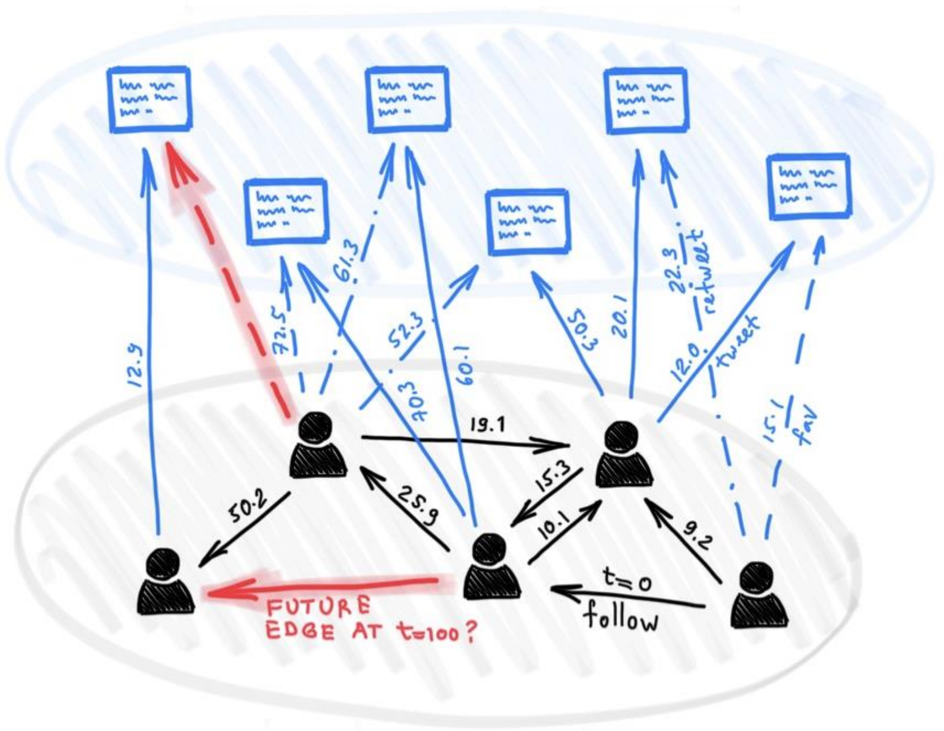
Towards Open Temporal Graph Neural Networks

Kaituo Feng[1], Changsheng Li[1], Xiaolu Zhang[2], Jun Zhou[2]

Beijing Institute of Technology[1]
Ant Group[2]

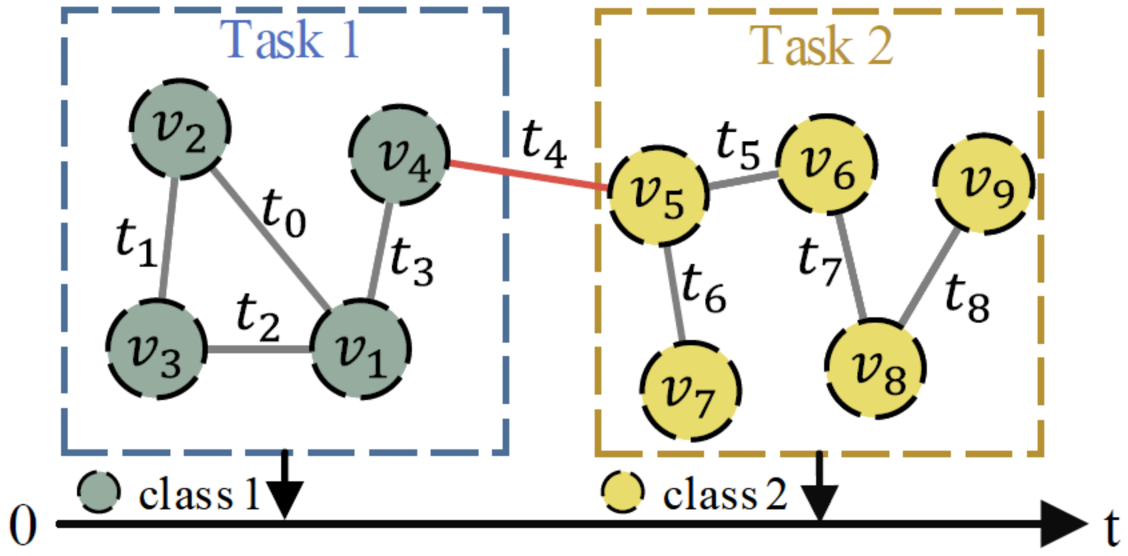
Background

- Temporal graph represents a sequence of time-stamped events.
- Recently, temporal graph neural networks (TGNs) have become powerful models for learning temporal graphs.



Background

- A basic assumption among TGNs is that the class set of nodes is always closed.
- However, in real-world scenarios, it often faces the open set problem with the dynamically increased class set as the time passes by.



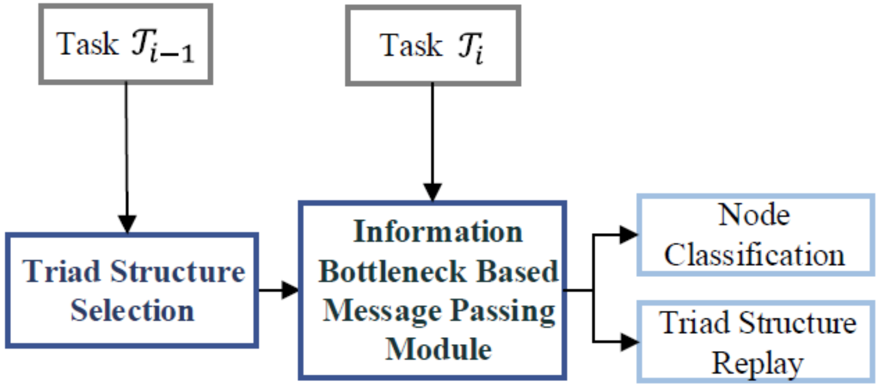
Motivation

- We attempt to investigate open temporal graph neural network.

- Challenge 1:
  - ➢What knowledge should be transferred between connected new class node and old class node on open temporal graph neural network?
- Challenge 2:
  - ➢How to address the catastrophic forgetting over old classes on open temporal graph neural network?

Method

- To prevent catastrophic knowledge forgetting, we propose to select representative and diverse triads to replay.
- To mitigate the issue of heterophily propagation, We design a new message passing mechanism to only propagate class-agnostic knowledge between nodes of different classes.



Method

- Based on information bottleneck, we could extract class-agnostic representations from node embeddings.
- For connected nodes with different classes, we only transfer class-agnostic knowledge between them.
- We assume the pseudo-label of a test node to be the one that appears the most times among its neighbor nodes in the training set.

$$J_{IB} = \min_{Z(t)} I(Z(t), Y) - \beta I(Z(t), X(t))$$

$$\begin{aligned} J_{IB} \leq \mathcal{L}_{IB} = & \mathbb{E}_{p(Z(t),Y)}[\log q_\mu(y|z(t))] - \mathbb{E}_{p(Z(t))}\mathbb{E}_{p(Y)}[\log q_\mu(y|z(t))] \\ & - \beta(\sup_\psi \mathbb{E}_{p(X(t),Z(t))}[T_\psi(x(t), z(t))] - \log(\mathbb{E}_{p(X(t))p(Z(t))}[e^{T_\psi(x(t),z(t))}])) \end{aligned}$$

Method

- We intend to select both representative and diverse triads to replay for overcoming catastrophic forgetting.
- We develop a greedy algorithm to find its approximate solution, and give a theoretical guarantee to the lower bound of the approximation ratio.

$$\begin{aligned}\mathcal{I}_{loss}(g_k^c, \theta) &= \left.\frac{\mathrm{d}\,\mathcal{L}_(G_k, \theta_{\varepsilon, g_k^c})}{\mathrm{d}\,\varepsilon}\right|_{\varepsilon=0} = \nabla_\theta \mathcal{L}(G_k, \theta)^\top \left.\frac{\mathrm{d}\,\hat{\theta}_{\varepsilon, g_k^c}}{\mathrm{d}\,\varepsilon}\right|_{\varepsilon=0} \\ &= -\nabla_\theta \mathcal{L}(G_k, \theta)^\top H_\theta^{-1} \nabla_\theta \mathcal{L}(g_k^c, \theta)\end{aligned}$$

$$S_k^c = \arg \max_{\{g_{k,1}^c, \cdots, g_{k,M}^c\}} F(S_k^c) = \arg \max_{\{g_{k,1}^c, \cdots, g_{k,M}^c\}} \left(\sum_{i=1}^{M} \mathcal{R}(g_{k,i}^c) + \gamma \frac{\left|\bigcup_{i=1}^{M} \mathcal{C}(g_{k,i}^c)\right|}{|N_k^c|} \right)$$

Experiment

- Our method outperform other baselines by a large margin.

Table 2: Comparisons (%) of our method with baselines. The bold represents the best in each column.

Method	Reddit		Yelp		TaoBao	
	AP(↑)	AF(↓)	AP(↑)	AF(↓)	AP(↑)	AF(↓)
ContinualGNN	52.17 ± 2.46	25.59 ± 5.39	49.73 ± 0.27	28.76 ± 1.52	58.39 ± 0.24	47.03 ± 0.50
ER-GAT	52.03 ± 2.59	22.67 ± 3.30	62.05 ± 0.70	18.91 ± 1.09	70.09 ± 0.88	23.24 ± 0.36
TWC-GAT	52.88 ± 0.53	19.60 ± 3.64	60.90 ± 3.74	16.92 ± 0.63	59.91 ± 1.71	42.78 ± 1.39
TGAT	48.47 ± 1.81	31.03 ± 4.48	64.89 ± 1.27	27.31 ± 3.99	60.62 ± 0.23	43.35 ± 0.77
TGAT+EWC	50.16 ± 2.45	28.27 ± 4.00	66.58 ± 3.11	25.48 ± 1.75	64.03 ± 0.62	38.26 ± 1.20
TGAT+iCaRL	54.50 ± 2.04	27.66 ± 1.11	71.71 ± 2.48	17.56 ± 2.46	73.74 ± 1.40	23.90 ± 2.04
TGAT+BiC	54.61 ± 0.89	25.42 ± 2.72	74.73 ± 3.54	16.42 ± 4.41	74.05 ± 0.48	23.27 ± 0.65
TGN	47.49 ± 0.48	32.06 ± 1.91	56.24 ± 1.65	41.27 ± 2.30	65.89 ± 1.20	36.15 ± 1.55
TGN+EWC	49.45 ± 1.45	31.74 ± 1.11	60.83 ± 3.55	35.73 ± 3.48	68.89 ± 2.09	32.08 ± 3.88
TGN+iCaRL	50.86 ± 4.83	31.01 ± 2.78	73.34 ± 1.99	15.43 ± 0.93	77.42 ± 0.80	19.57 ± 1.29
TGN+BiC	53.16 ± 1.53	26.83 ± 0.95	73.98 ± 2.07	16.79 ± 2.90	77.40 ± 0.80	18.63 ± 1.69
TREND	49.61 ± 2.92	28.68 ± 4.20	57.28 ± 2.83	37.48 ± 3.26	61.02 ± 0.16	42.44 ± 0.14
TREND+EWC	53.12 ± 3.30	25.70 ± 3.08	65.45 ± 4.79	26.80 ± 4.98	62.72 ± 1.18	40.00 ± 2.09
TREND+iCaRL	52.53 ± 3.67	30.63 ± 0.18	69.93 ± 5.55	15.81 ± 7.48	74.49 ± 0.05	23.27 ± 0.25
TREND+BiC	54.22 ± 0.56	22.42 ± 3.15	71.15 ± 2.42	12.78 ± 5.12	75.13 ± 1.06	21.70 ± 0.63
OTGNet (Ours)	**73.88** ± 4.55	**19.25** ± 5.10	**83.78** ± 1.06	**4.98** ± 0.46	**79.92** ± 0.12	**12.82** ± 0.61

Experiment

- From ablation study, we can find that each component in our method is effective.

Table 3: Ablation study of our proposed information bottleneck based propagation mechanism.

Setting	Reddit		Yelp		TaoBao	
	AP(↑)	AF(↓)	AP(↑)	AF(↓)	AP(↑)	AF(↓)
OTGNet-w.o.-IB	54.10 ± 2.01	34.00 ± 1.63	76.93 ± 5.14	14.96 ± 5.61	79.00 ± 0.37	13.41 ± 0.57
OTGNet-w.o.-prop	54.67 ± 2.05	28.73 ± 2.63	75.67 ± 1.69	12.87 ± 1.19	79.07 ± 0.02	14.48 ± 0.34
OTGNet-GBK	58.79 ± 1.08	25.22 ± 2.22	77.03 ± 2.99	9.79 ± 1.15	77.73 ± 0.27	15.49 ± 0.34
OTGNet	**73.88** ± 4.55	**19.25** ± 5.10	**83.78** ± 1.06	**4.98** ± 0.46	**79.92** ± 0.12	**12.82** ± 0.61

Table 4: Results of triad selection strategy on the three datasets.

Setting	Reddit		Yelp		TaoBao	
	AP(↑)	AF(↓)	AP(↑)	AF(↓)	AP(↑)	AF(↓)
OTGNet-w.o.-triad	60.81 ± 4.46	34.94 ± 4.73	69.28 ± 1.73	23.79 ± 1.75	67.05 ± 0.44	31.44 ± 0.41
OTGNet-random	69.66 ± 3.81	23.24 ± 3.83	78.76 ± 2.62	9.19 ± 1.65	79.09 ± 0.36	13.89 ± 0.45
OTGNet-w.o.-diversity	71.06 ± 5.73	22.96 ± 6.91	80.76 ± 2.60	9.91 ± 3.83	78.84 ± 0.46	13.87 ± 1.18
OTGNet	**73.88** ± 4.55	**19.25** ± 5.10	**83.78** ± 1.06	**4.98** ± 0.46	**79.92** ± 0.12	**12.82** ± 0.61

Table 5: Results of evolution pattern preservation on the three datasets.

Setting	Reddit		Yelp		TaoBao	
	AP(↑)	AF(↓)	AP(↑)	AF(↓)	AP(↑)	AF(↓)
OTGNet-w.o.-pattern	70.23 ± 5.56	23.10 ± 7.44	81.44 ± 1.38	6.97 ± 3.10	79.01 ± 0.19	14.05 ± 0.46
OTGNet	**73.88** ± 4.55	**19.25** ± 5.10	**83.78** ± 1.06	**4.98** ± 0.46	**79.92** ± 0.12	**12.82** ± 0.61

Thanks for your attention