国际人工智能会议
AAAI 2021 论文北京预讲会

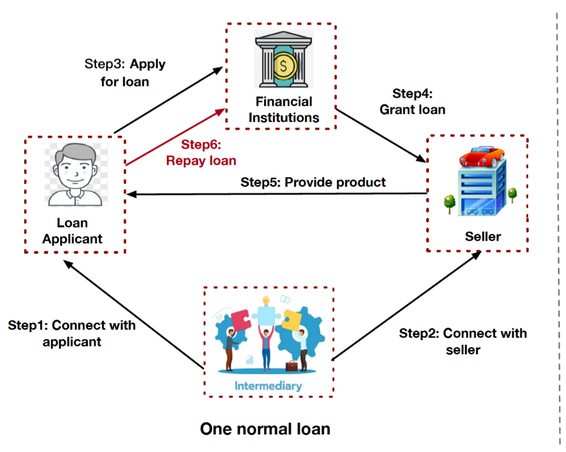
Towards Consumer Loan Fraud Detection: Graph Neural Networks with Role-Constrained Conditional Random Field

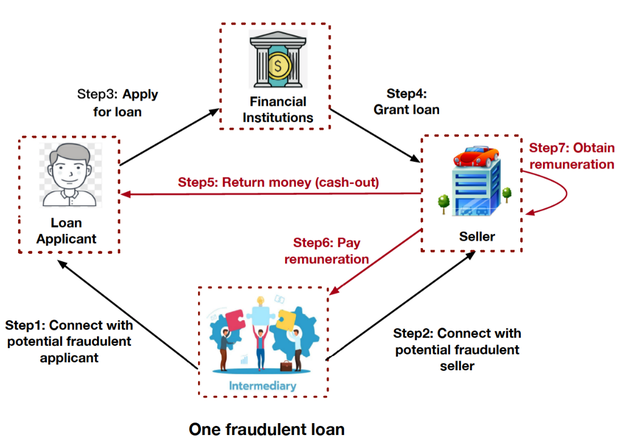
Bingbing Xu, Huawei Shen, Bingjie Sun, Rong An, Qi Cao, Xueqi Cheng
Institute of Computing Technology, Chinese Academy of Sciences
University of Chinese Academy of Sciences, Beijing, China
Ant Financial Services Group

INTRODUCTION

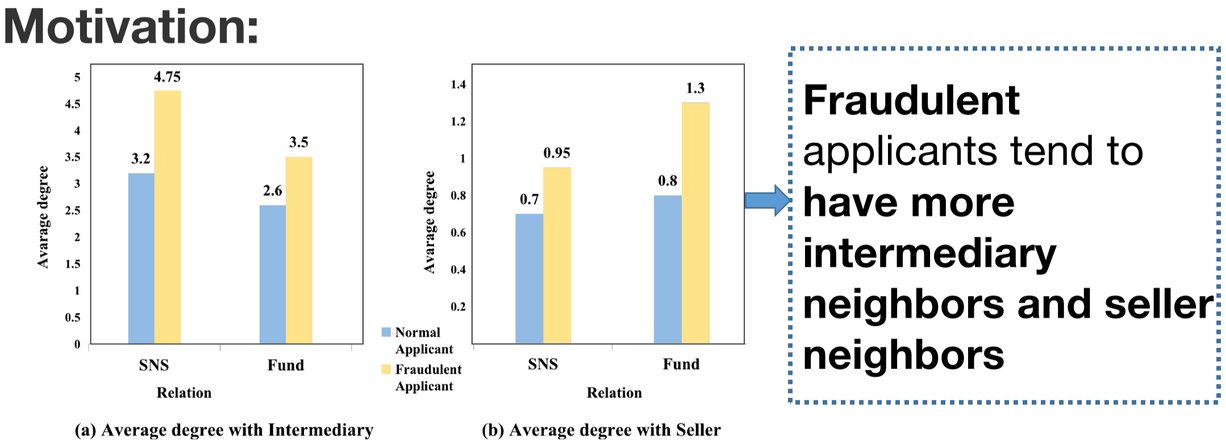
Consumer Loan Detection:

- **Consumer loans**, i.e., loans given to consumers to finance certain types of expenditures, is increasingly popular.
- Loan fraud detection is formulated as a **node classification** problem, i.e., predict whether an applicant node is fraudulent given node attribute and social relationships among nodes.
- Graph neural network with a **Role-constrained** Conditional random field, namely **GRC**, to learn the representation of applicants.



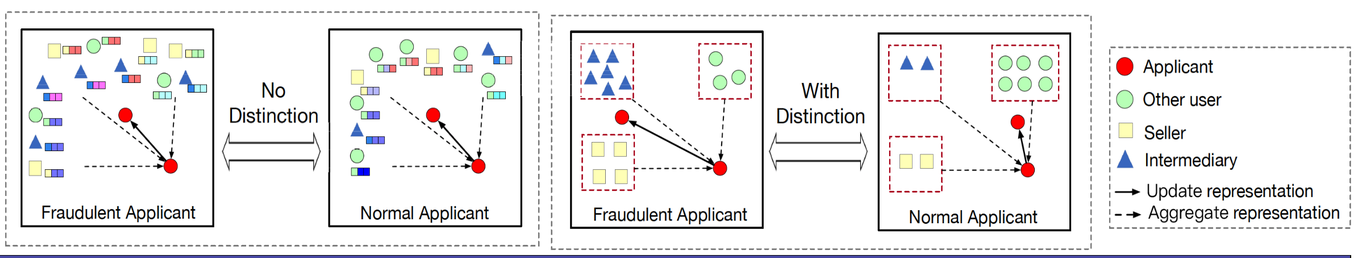


Motivation:



Fraudulent applicants tend to **have more intermediary neighbors and seller neighbors**

- GRC is equiped with a conditional random field component to learn **role-aware node represenations**,thus distinguishing the fraudulent applicants from normal applicants and enhancing the classification.



METHOD

GRC: Graph neural network with a Role-constrained Conditional random field

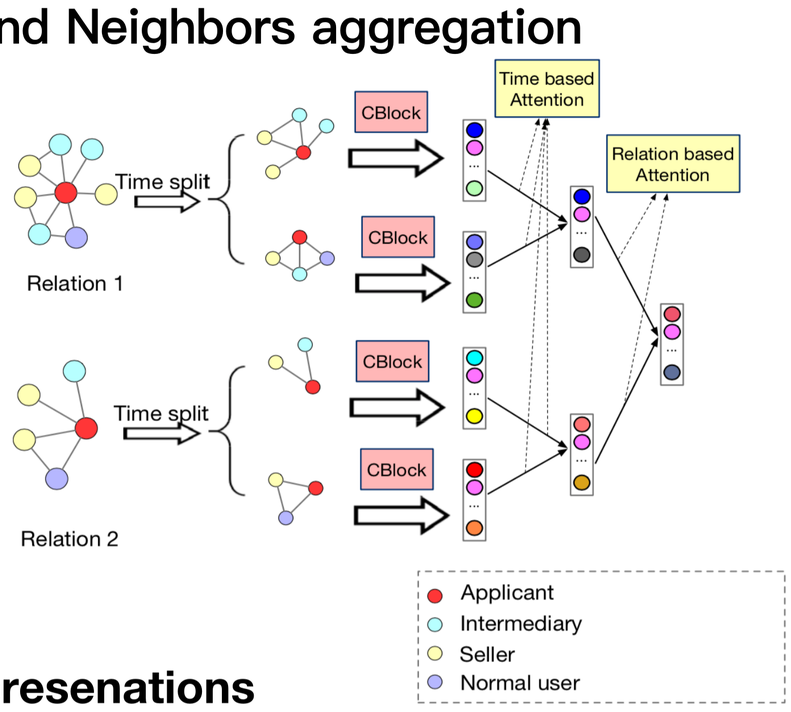
➢ Feature transformation and Neighbors aggregation

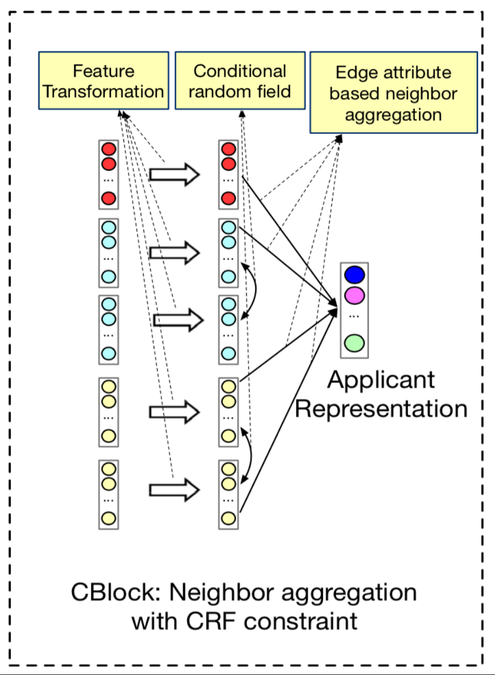
$$X^{'} = XW,$$
$$e_{uv} = a^{i} X^{i}_{uv},$$
$$h_v = \sum_{u \in N(v)} e_{uv} H^K_u + H^K_v$$

➢ CRF constraint

$$\psi_u(H_u, X^{'}_u) = ||H_u - X^{'}_u||^2_2 \to 0,$$
$$\psi_p(H_u, H_v) = f_{uv}||H_u - H_v||^2_2 \to 0$$
$$H^{k+1}_u = \frac{\alpha X^{'}_u + \beta \sum_{v \in R(u)} f_{uv} H^k_v}{\alpha + \beta \sum_{v \in R(u)} f_{uv}}$$

- **Learn role-aware node represenations**





➢ Time slot aggregation

$$w_{r,ti} = \frac{\exp(w_{r,ti})}{\sum_{i \in 1,2} \exp(w_{r,ti})}$$
$$h^r_v = w_{r,t1} h^r_{v,t1} + w_{r,t2} h^r_{v,t2}.$$

➢ Relation aggregation

$$w_r = \frac{\exp(w_r)}{\sum_{r \in R} \exp(w_r)},$$
$$h_v = \sum_{r \in R} w_r h^r_v.$$

EXPERIMENTS

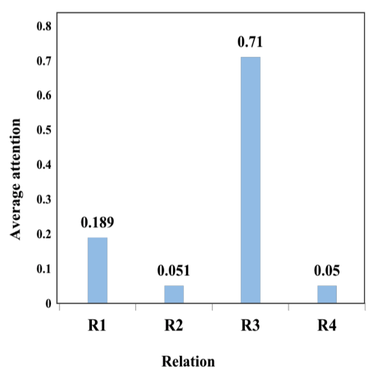
Experimental results:

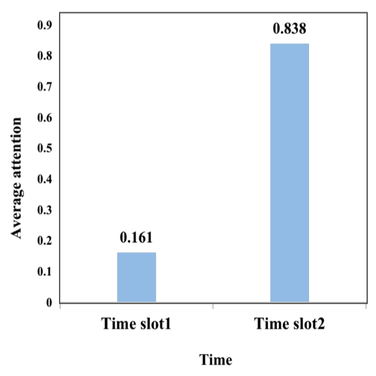
Table 1: Results of loan fraud detection

Size	Method	SVM	MLP	GCN	GRC
1,000	Precision	66.32%	72.80%	69.78%	72.28%
	Recall	51.00%	44.20%	61.20%	82.40%
	F1-Score	57.66%	55.00%	65.21%	**77.01%**
2,000	Precision	65.80%	71.25%	68.59%	74.38%
	Recall	51.60%	46.60%	69.00%	89.70%
	F1-Score	57.84%	56.35%	68.79%	**81.32%**
3,000	Precision	66.50%	69.25%	70.35%	77.09%
	Recall	47.10%	51.80%	71.60%	89.50%
	F1-Score	55.14%	59.27%	70.97%	**82.83%**
12,000	Precision	68.65%	66.82%	74.06%	77.50%
	Recall	54.10%	59.20%	84.80%	95.50%
	F1-Score	60.51%	62.78%	79.07%	**85.56%**

Attntion analysis:

- “Relation3” and “Time slot2” are of higher importance
- Modeling the impacts of relations and time slots is necessary.





Ablation study:

- Each component achieves the positive effect.
- The edge attributes are important to identify the important neighbors.

Table 2: Ablation study

Method	precision	Recall	F1 score
GRC(w/o CRF)	77.71%	91.00%	83.83%
GRC(w/o EA)	79.60%	88.20%	83.68%
GRC(w/o RA)	77.58%	92.40%	84.34%
GRC(w/o TA)	76.54%	91.70%	83.44%
GRC	77.50%	95.50%	**85.56%**

主办方：中国中文信息学会青年工作委员会
承办方：智源社区