Denoised and Self-supervised Graph Neural Networks for Text Classification

Juan Shu
May 2023

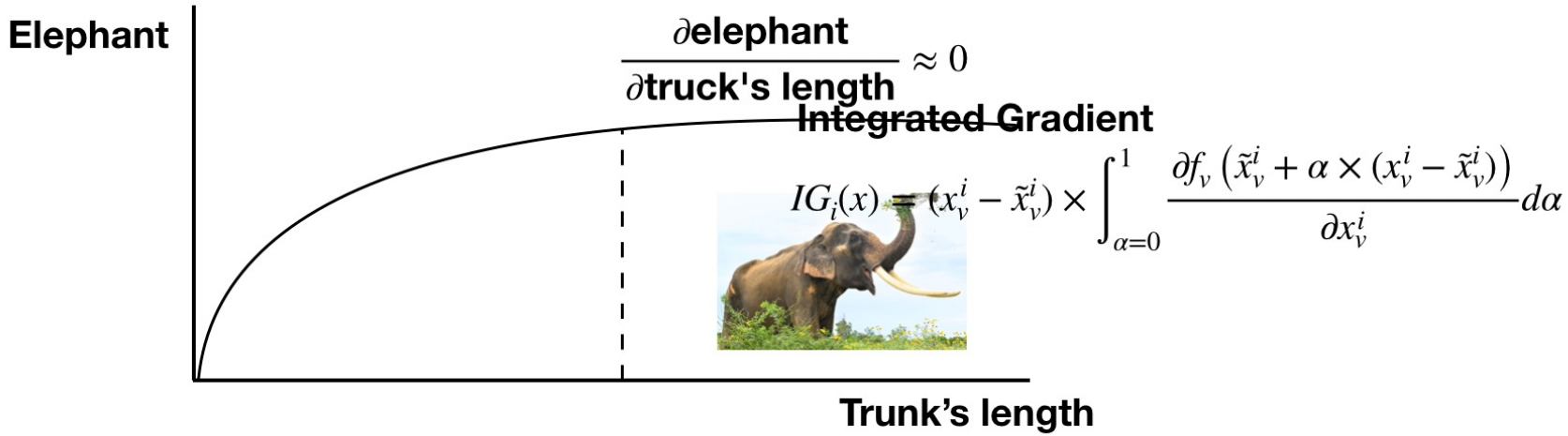
Motivation

- **GNN is powerful in text classification** (captures syntactic, semantic, and long-range dependencies among words and phrases)

- **Manually created text graphs can have many noise or missing features** (wrong edges, missing edges)

- **Denoise and reconstruct (remove edges or add edges) the input adaptively!**

Methods : Integrated Gradient



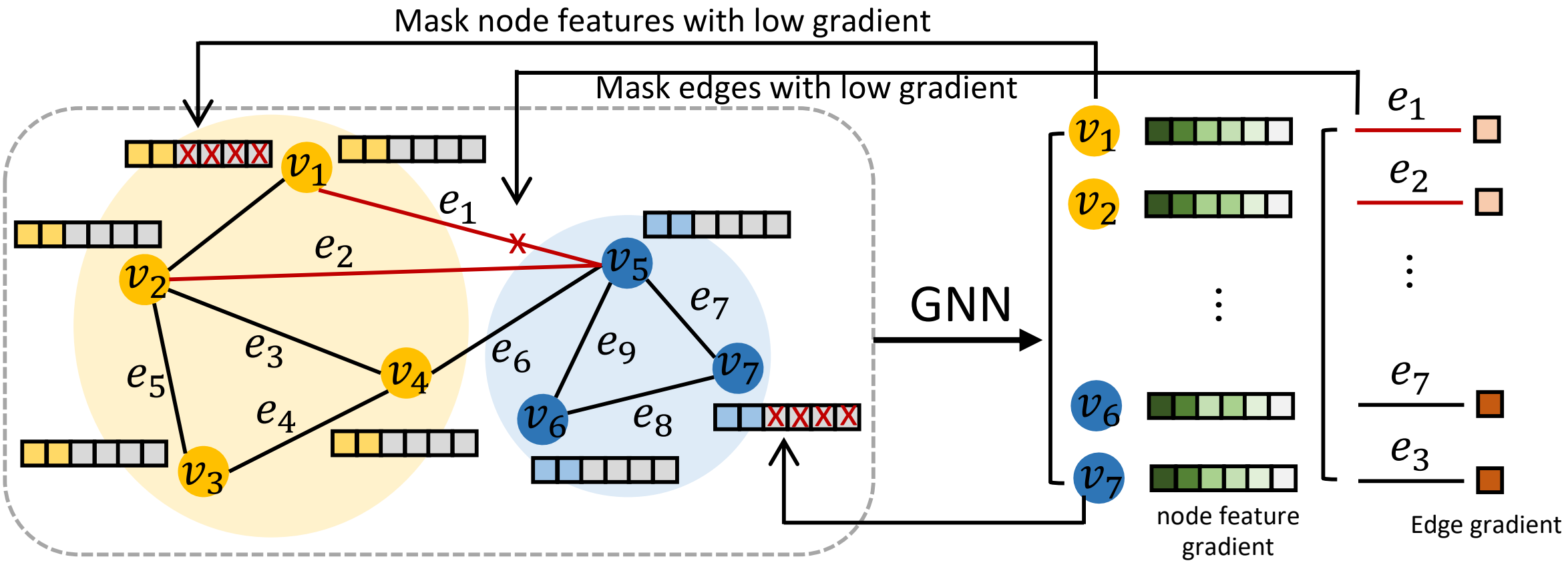
Integrated gradient of node features

$$IG(x_v^i) = (x_v^i - \tilde{x}_v^i) \times \sum_{j=1}^{J} \frac{\partial f_v\left(\tilde{x}_v^i + \frac{j}{J} \times (x_v^i - \tilde{x}_v^i\right)}{\partial x_v^i} \times \frac{1}{J}$$

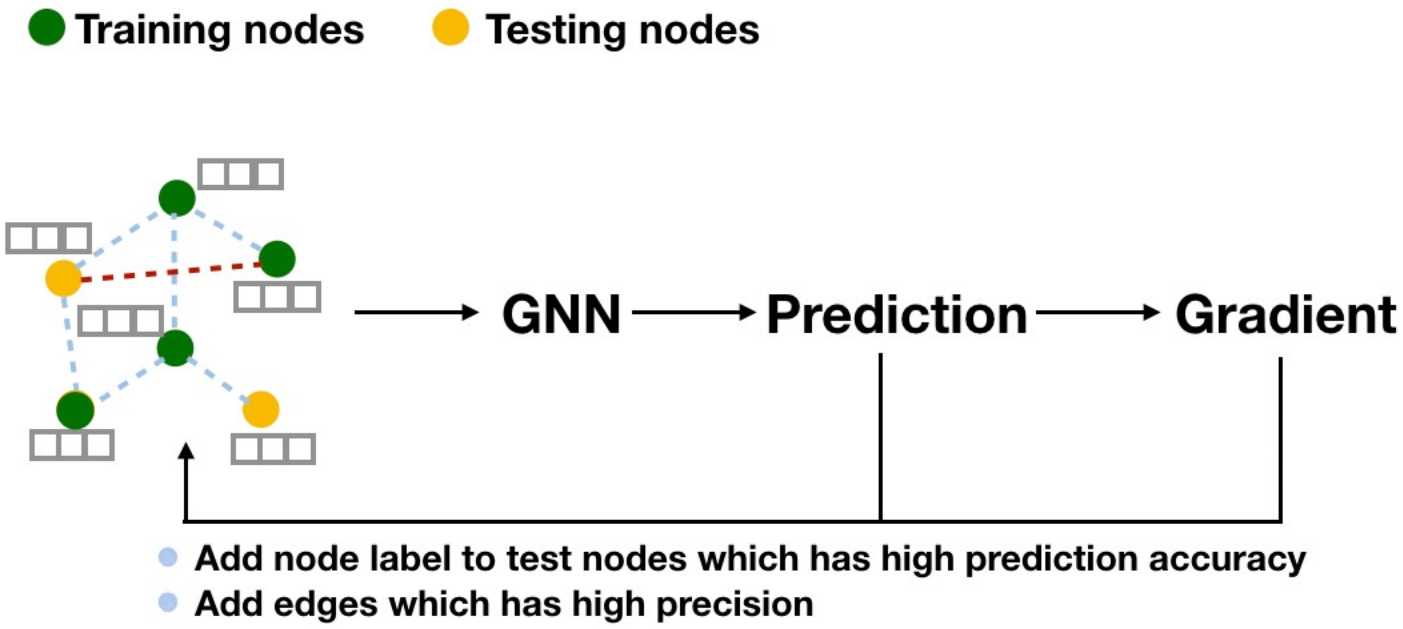
Integrated gradient of edges

$$IG(e_{vu}) = (e_{vu} - \tilde{e}_{vu}) \times \sum_{j=1}^{J} \frac{\partial f_e\left(\tilde{e}_{vu} + \frac{j}{J} \times (e_{vu} - \tilde{e}_{vu}\right)}{\partial e_{vu}} \times \frac{1}{J}$$

Denoised procedure



Self-supervised training procedure



! Self-supervised learning

Experimental Results on text classification

Model	20NG	R8	R52	Ohsumed	MR
TF-IDF + LR	0.8319 ± 0.0000	0.9374 ± 0.0000	0.8695 ± 0.0000	0.5466 ± 0.0000	0.7459 ± 0.0000
CNN-rand	0.7693 ± 0.0061	0.9402 ± 0.0057	0.8537 ± 0.0047	0.4387 ± 0.0100	0.7498 ± 0.0070
CNN-non-static	0.8215 ± 0.0052	0.9571 ± 0.0052	0.8759 ± 0.0048	0.5844 ± 0.0106	0.7775 ± 0.0072
LSTM	0.6571 ± 0.0152	0.9368 ± 0.0082	0.8554 ± 0.0113	0.4113 ± 0.0117	0.7506 ± 0.0044
LSTM (pretrain)	0.7543 ± 0.0172	0.9609 ± 0.0019	0.9048 ± 0.0086	0.5110 ± 0.0150	0.7733 ± 0.0089
Bi-LSTM	0.7318 ± 0.0185	0.9631 ± 0.0033	0.9054 ± 0.0091	0.4927 ± 0.0107	0.7768 ± 0.0086
PV-DBOW	0.7436 ± 0.0018	0.8587 ± 0.0010	0.7829 ± 0.0011	0.4665 ± 0.0019	0.6109 ± 0.0010
PV-DM	0.5114 ± 0.0022	0.5207 ± 0.0004	0.4492 ± 0.0005	0.2950 ± 0.0007	0.5947 ± 0.0038
PTE	0.7674 ± 0.0029	0.9669 ± 0.0013	0.9071 ± 0.0014	0.5358 ± 0.0029	0.7023 ± 0.0036
fastText	0.7938 ± 0.0030	0.9613 ± 0.0021	0.9281 ± 0.0009	0.5770 ± 0.0049	0.7514 ± 0.0020
fastText (bigrams)	0.7967 ± 0.0029	0.9474 ± 0.0011	0.9099 ± 0.0005	0.5569 ± 0.0039	0.7624 ± 0.0012
SWEM	0.8516 ± 0.0029	0.9532 ± 0.0026	0.9294 ± 0.0024	0.6312 ± 0.0055	0.7665 ± 0.0063
LEAM	0.8191 ± 0.0024	0.9331 ± 0.0024	0.9184 ± 0.0023	0.5858 ± 0.0079	0.7695 ± 0.0045
Graph-CNN-C	0.8142 ± 0.0032	0.9699 ± 0.0012	0.9275 ± 0.0022	0.6386 ± 0.0053	0.7722 ± 0.0027
Graph-CNN-S	–	0.9680 ± 0.0020	0.9274 ± 0.0024	0.6282 ± 0.0037	0.7699 ± 0.0014
Graph-CNN-F	–	0.9689 ± 0.0006	0.9320 ± 0.0004	0.6304 ± 0.0077	0.7674 ± 0.0021
Text GCN	0.8634 ± 0.0009	0.9707 ± 0.0010	0.9356 ± 0.0018	0.6836 ± 0.0056	0.7674 ± 0.0020
Text dsGCN	**0.8810 ± 0.0011**	**0.9712 ± 0.0009**	**0.9608 ± 0.0018**	**0.7052 ± 0.0088**	**0.7729 ± 0.0018**

Discussion

- **By denoising and reconstructing the input adaptively, we can achieve better text classification results with GNN**
- **Ongoing work1: test the explanation accuracy of the model**
- **Ongoing work2: make the current model more computationally efficient**