What nodes vote to? Graph classification without readout phase

[1]Yuxing Tian, [2]Zheng Liu, [1]Weiding Liu, [1]Zeyu Zhang, [1]Yanwen Qu

[1]School of Computer Information and Engineering, Jiangxi Normal University

[2]Jiangsu Key Laboratory of BDSIP, Nanjing University of Posts and Telecommunications

Introduction

- Graph Neural Networks(GNNs)

  -be constructed to deal with graph related tasks.

- Message Passing Neural Networks (MPNNs)

  -a Message Passing phase used for updating node embeddings .

  -a Readout phase used for extracting graph feature.

- Graph classification algorithms

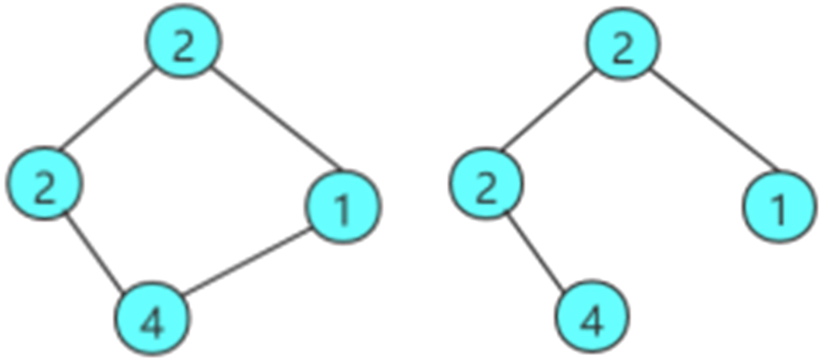
  -reasearch on graph convolutional operators.

  -research on graph feature extraction.

Problem

- Graph feature extraction

  -may obscure the effect of the node embedding of each node.



➢ Fig. 1 :An example to show two graphs with different structures. Node embeddings are given in the node circles. If the global mean pooling, which is widely used in many graph classification models, is adopted in the Readout phase, the graph features extracted from the two graphs will be the same.

Contribution

- We propose a new phase named Node Voting phase based on the attention mechanism.

- We propose a new model called NVnet which includes a Message Passing phase, a Node Voting phase, and a decoder.

- Experiments on 4 graph classification benchmark datasets and results show that it performs well.

- We use 4 different graph convolutional operators to construct four different Message Passing phases respectively. The experimental results show that NVnet performs better than MPNNs in most cases.

Our Model

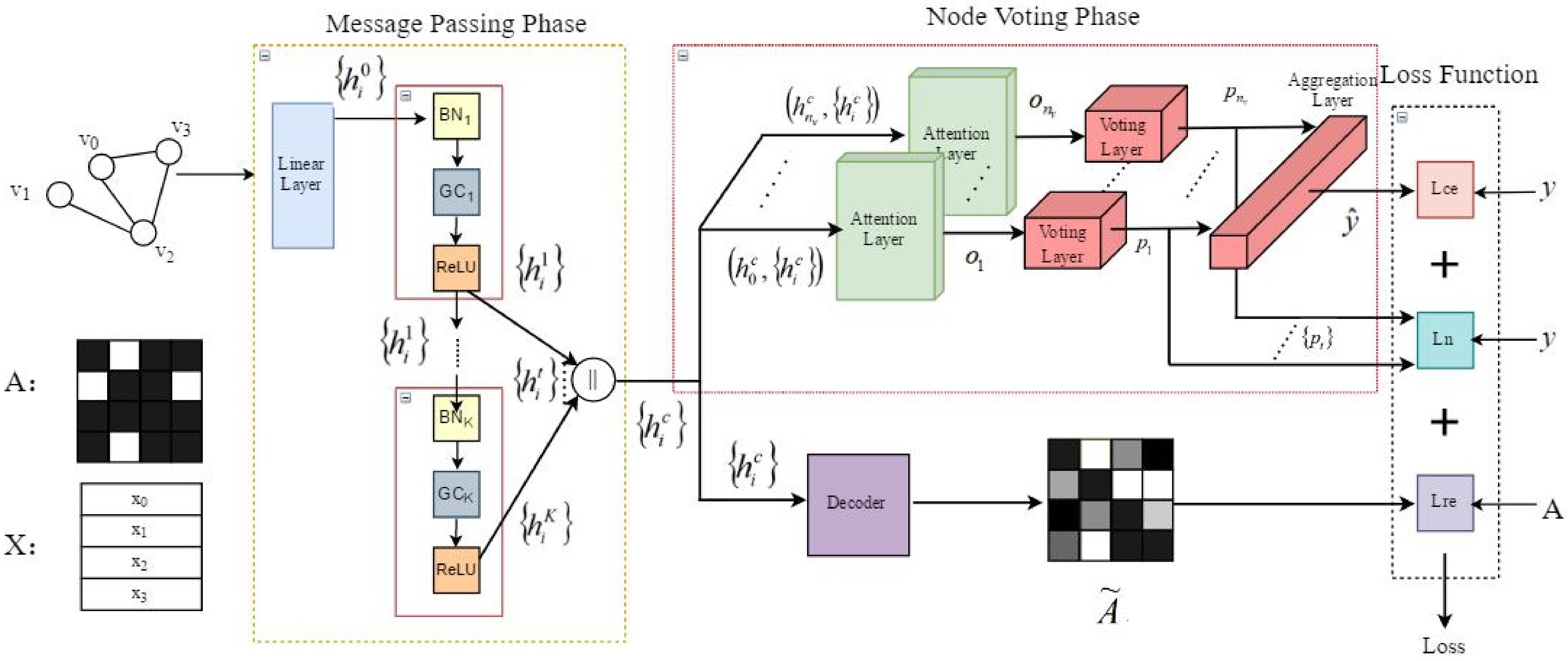


Fig. 2. The left part is the input sample. The middle part is the framework of the NVnet. It consists of three components: a Message Passing phase for node embeddings updating, a decoder block for graph reconstruction, and a Node Voting phase for graph classification. The calculation diagram of the loss function is shown on the right.

Node Voting Phase

Attention Mechanism

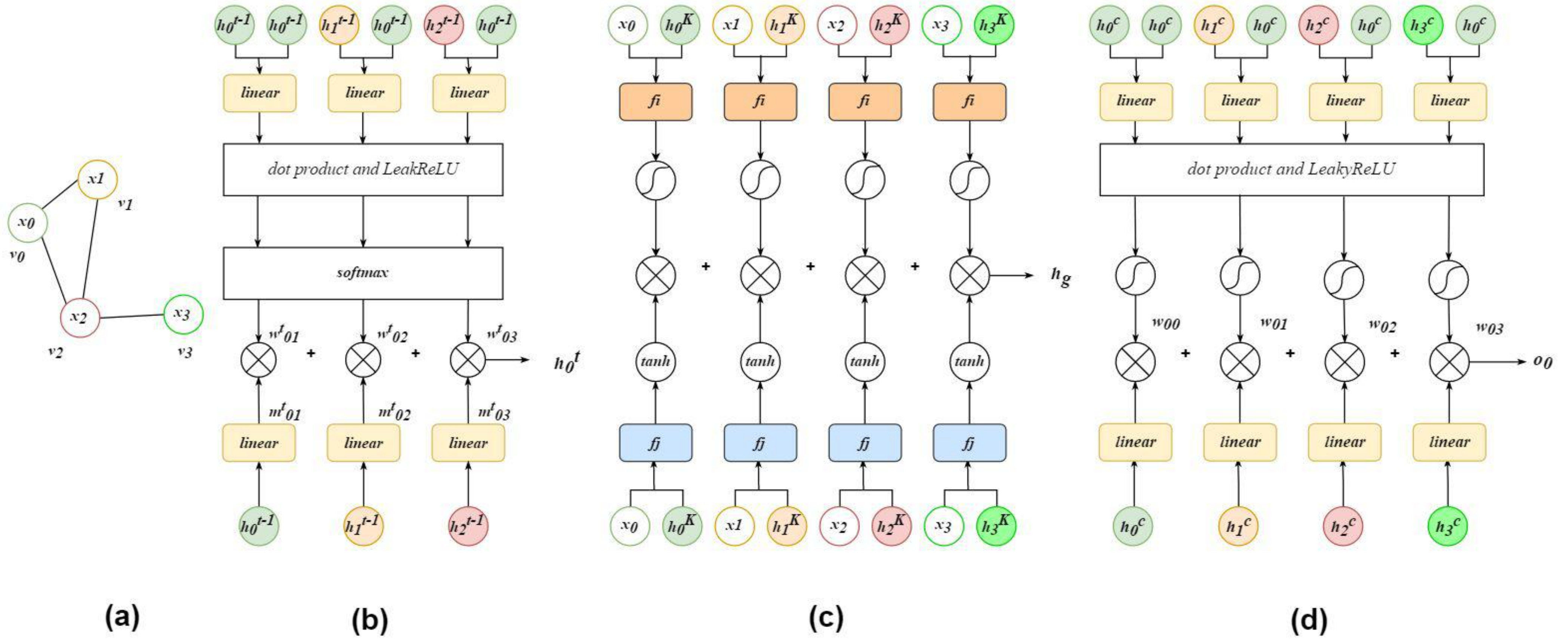


Fig. 3. An example to show three different attention mechanisms. Fig. 3 (b) shows the calculation process of GATconv. Fig. 3 (c) shows the calculation process of GA. Fig. 3 (d) shows the calculation process of NVnet.

Node Voting Phase

- Attention layer

-Supposing the attention layer contains M heads.first,the attention weights $\{w^m_{ij} | j=1,...,n_v\}$ related to node v_i in the m-th head is calculated according to (9).

$$w_{ij}^{h\_m} = \mathrm{sigmoid}(\mathrm{LeakyReLU}(\mathrm{concat}(\{h_i^c \Phi_m, h_j^c \Phi_m\})\theta_m^T)) \quad (9)$$

-then,the output feature a_i^m of the m-th head is calculated according to (10)

$$a_i^m = \sum_{j=1,\ldots n_v} w_{ij}^{h\_m}(h_j^c \Phi_m) \quad (10)$$

$$o_i = \mathrm{concat}(\{a_i^m \mid m = 1, \ldots M\}) \quad (11)$$

Node Voting Phase

- Voting layer

  -In this layer, v_i will predict the class distribution p_i of the graph according to (12)

$$p_i = \mathrm{softmax}(\mathrm{Ln}_2(\mathrm{Dropout}(\mathrm{LeakyReLU}(\mathrm{Ln}_1(o_i))))) \qquad (12)$$

- Aggregation layer

  -the graph classification result will be calculated according to (13)

$$\hat{y} = \mathrm{mean}(\{p_i \mid i = 1, \ldots, n_v\}) \qquad (13)$$

Loss function

- *The loss function consists of three parts. λ_{re} and λ_n are hyper-parameters. L_{ce} represents the cross entropy loss between the targets and predictions.*

$$Loss = L_{ce} + \lambda_{re} L_{re} + \lambda_n L_n \qquad (14)$$

- *The regularization term L_n is used to constrain the crossentropy loss between the predictions of the nodes and the real class distribution of the graphs the nodes belong to.*

$$L_{\mathrm{n}} = \frac{1}{\sum_{s=1,..,S} n_v^s} \sum_{s=1,\ldots,S} \sum_{i=1,\ldots,n_v^s} \mathrm{CrossEntropy}(p_i^s, y^s) \qquad (17)$$

Loss function

- *The regularization term L_{re} is used to constrain the reconstruction loss between the adjacency matrix of the input graph and that of the reconstructed graph.*

$$L_{re} = \frac{1}{S} \sum_{s} L_{s,re} \qquad (15)$$

$$L_{s,re} = \frac{\sum_{i=1,\ldots,n_v, j=1,\ldots,n_v, e_{ij} \in E} -\log(\widetilde{A}_{ij})}{\sum_{i=1,\ldots,n_v, j=1,\ldots,n_v, e_{ij} \in E} 1} + \frac{\sum_{i=1,\ldots,n_v, j=1,\ldots,n_v, e_{ij} \notin E} -\log(1-\widetilde{A}_{ij})}{\sum_{i=1,\ldots,n_v, j=1,\ldots,n_v, e_{ij} \notin E} 1} \qquad (16)$$

Experiments

- Dataset

-The benchmark datasets include 2 bio-informatics datasets: MUTAG and PROTEINS, and 2 social network datasets: IMDB-BINARY (IMDB-B) and COLLAB.

- Configuration

-we compared NVnet with 10 baselines,all of the results are obtained under 10 fold cross validation. We implement NVnet based on pytorch geometric (version 1.3.2) .

-we use Adam for optimization training, batch-size is 20, the number of training epochs is 100. The initial value of the learning rate is 0.001, and the learning rate is decayed by half every 20 epochs.

Experiments

Performance

- NVnet(GIN-εconv) achieves the highest accuracy rates on MUTAG, PROTEINS and IMDB-BINARY. Meanwhile, NVnet(GCNconv) achieves the highest average accuracy on COLLAB.

- when the graph convolutional operators used by NVnet and MPNNs belong to the same category, compared with MPNNs using global mean pooling as the Readout function, NVnet can achieve higher average accuracy on 15 of the 16 comparisons.

Method	Dataset			
	MUTAG	PROTEINS	COLLAB	IMDB-B
GCN[12]	74.6±7.7	73.1±3.8	***80.6±2.1***	72.6±4.5
GraphSAGE[12]	74.9±8.7	73.8±3.6	79.7±1.7	72.4±3.6
GIN-0[12]	***85.7±7.7***	72.1±5.1	79.3±2.7	***72.8±4.5***
GIN-ε[12]	83.4±7.5	72.6±4.9	79.8±2.4	72.1±5.1
Graclus[12]	77.1±7.2	73.0±4.1	79.6±2.0	72.2±4.2
top_k pooling[12]	76.3±7.5	72.7±4.1	79.7±2.2	72.5±4.6
DiffPool[12]	85.0±10.3	***75.1±3.5***	78.9±2.3	72.6±3.9
Global Attention[12]	74.6±8.0	72.5±4.5	79.6±2.2	72.3±3.8
Set2Set[12]	73.3±6.9	73.6±3.7	79.6±2.3	72.2±4.2
SortPool[12]	77.3±8.9	72.4±4.1	77.7±3.1	72.4±3.8
NVnet(GCNconv)	80.2±9.7	74.8±3.7	***81.8±1.0***	73.7±5.7
NVnet(SAGEconv)	79.8±6.2	75.7±3.5	81.3±2.3	74.0±5.5
NVnet(GIN-0conv)	85.4±9.6	76.1±3.1	80.6±1.0	73.4±5.3
NVnet(GIN-εconv)	***87.3±6.1***	***76.3±3.0***	80.5±1.8	***74.0±3.7***

Conclusion

- A node voting-based graph classification model without Readout phase named NVnet is proposed.