VCHN with Snowball GCN Architechture

Zhai Mingshu
Department of Computer Science
Shanghai Jiao Tong University
zhaimingshuzms@sjtu.edu.cn

Abstract

View-Consistent Heterogeneous Network (VCHN) is a successful architechture of transductive learning on graph with very few labeled data.It combines spectral view aggregation and spatial view aggregation. The most common way to deploy spectral view aggregation is to use Graph Convolutional Network (GCN) and VCHN adapts it as well.My work is the changement of the original GCN architechture to snowball.

1 Introduction

1.1 Background

Transductive learning on graph aims to predict the labels of nodes.Transductive learning on graph with very few labeled data is even more challenging because the lack of supervision. VCHN [1] is one of the most successful architecture in this area.Its key idea is designing the loss function based on the view consistency of different view.

Snowball architechture of GCN is combine features of different layers like snowballing. The features of previous layers are concatenated together and multiplied by learnable parameter matrix. After the activation function,it turns into the input feature of the new layer.

1.2 Motivation

The original implementation of VCHN uses two GNN frameworks:GCN and GAT. In the standard implementation of VCHN `https://github.com/kunzhan/VCHN` , a two layer basic GCN is implemented.

If we use them all alone,the snowball architechture GCN performs better than basic GCN in this task. The idea is that we can replace the basic GCN with snowball GCN in ensemble learning to get better performance.

2 Related Works

2.1 View-Consistent Heterogeneous Network

View Consistency View-Consistent Heterogeneous Network is an ensemble learning model which use soft cross-entropy loss function to construct the consistency loss function between two views.

$$l_v = -\frac{1}{n}\sum_{i=1}^{n}\sum_{j=1}^{k} y_{ij}^{(2)} \ln y_{ij}^{(1)} \tag{1}$$

Where $y_{ij}^{(1)}$ is the predictive probability of the spectral view of VCHN(GCN) and $y_{ij}^{(2)}$ is the predictive probability of the spatial view of VCHN(GAT). VCHN use adam optimizer to train super

36th Conference on Neural Information Processing Systems (NeurIPS 2022).

parameter and the goal is minimize overall loss

$$l_{overall} = l_1 + l_2 + l_v \tag{2}$$

Where l_1 is the sum of loss of the first model,including the loss on labeled data and pseudo data,so as l_2

Graph Convolutional Network Graph Convolutional Network (GCN) [2] is the spectral view model VCHN used. We consider a multi-layer Graph Convolutional Network (GCN) with the following layer-wise propagation rule:

$$H^{(l+1)} = \sigma(\hat{D}^{-\frac{1}{2}}\hat{A}\hat{D}^{-\frac{1}{2}}H^{(l)}W^{(l)}) \tag{3}$$

Here,$\hat{A} = A + I_n$ is sum of the adjacency matrix of undirected graph G and the identity matrix I $\hat{D}^{-\frac{1}{2}}\hat{A}\hat{D}$ means row normalization.

Graph Attention Network Graph Attention Network (GAT) [3] uses attention machanism to construct graph attentional layer.

$$\alpha_{ij} = \frac{\exp(\mathrm{LeakyReLU}(\vec{\alpha}^T[\mathbf{W}\vec{h_i}||\mathbf{W}\vec{h_j}]))}{\sum_{k\in\mathcal{N}(i)}\exp(\mathrm{LeakyReLU}(\vec{\alpha}^T[\mathbf{W}\vec{h_i}||\mathbf{W}\vec{h_k}]))} \tag{4}$$

$$\vec{h_i}' = \sigma(\sum_{j\in\mathcal{N}(i)}\alpha_{ij}\mathbf{W}\vec{h_j}) \tag{5}$$

VCHN performs multi-head attention so the layer is slightly different to it.

2.2 Snowball Architechture of GCN

Snowball architecture of GCN [4] combines all previous layers' features to get the new layer's feature.This helps to get a richer representation for each node.

$$\begin{aligned} \mathbf{H_0} = \mathbf{X} \quad \mathbf{H_{l+1}} &= f(L[H_0, \cdots, H_l]W_l), l = 0, 1, 2, \cdots, n-1 \\ C &= g([H_0, \cdots, H_n]W_n) \\ output &= \mathrm{softmax}(L^p C W_C) \end{aligned} \tag{6}$$

Here,L is the row normalized Laplace matrix,H_i is the feature of ith layer, W is the learnable parameter matrix and f, g is the activation function. Snowball architecture is efficient because the primitive GCN finally generates a matrix with low rank if the number of layers increases to infinity.

3 Method

3.1 The Changement of Hyperparameter

The primitive GCN that VCHN used has fixed two inner layers of convolution. Snowball architecture needs adaptive number of layers to enhance the ability of GCN, so a new hyperparameter "–layer" has been added to hyperparameter lists.

Also we need lower learning rate and weight decay because the snowball GCN is more complex and needs to train longer.

3.2 Use Module List to Contain Multiple Layers

Snowball architechture needs adaptive number of layers so moudle list is used to contain variable number of layers in a module.

4 Experiments

4.1 Few Labeled Data

Performance			
Dataset	Cora 0.005	Cora 0.01	Cora 0.03
VCHN	74.9%	81.3%	83.1%
Snowball(tanh)	71.36%	74.78%	80.72%
GCN	50.9%	63.3%	76.5%
GAT	41.4%	48.6%	56.8%
VCHN-snowball	75.25%	77.24%	83.27%
Truncated Krylov	74.89%	78.15%	81.92%

Due to the constraints of author's calculation device,only rough hyperparameters are available and the model's performance still has space to improve. To get similar performance of VCHN-snowball as the following figure, You need to use recommended hyperparameter in code `https://github.com/zhaimingshuzms/modified_VCLN`. The data of VCHN-snowball are all with validations and are the average accuracy of 10 independent experiments with fixed random seed.

4.2 The Whole Cora Dataset

The Whole Cora Dataset can't be provided to VCHN-snowball directly because the lack of space of pseudo labels. Author provides 50% labeled data to VCHN-snowball and gets an incredible high prediction accuracy. During 30 independent experiments divided into 3 partitions. $90.66\%, 90.23\%, 90.78\%$ accuracies are acquired,which is close to or even better than the best model for Cora(SSP 90.16%) [5]

The potential capacity of VCHN-snowball is huge because you can apply bagging technique on Cora dataset and send the data to multiple VCHN-snowball models to get a even better performance.

5 Conclusion

5.1 VCHN with Snowball GCN Architechture

In Cora 0.005% and Cora 0.03%,VCHN-snowball evinces the effection of snowball architechture to promote accuracy.

VCHN-snowball is a potential effective model for the whole Cora dataset as well because it owns the capacity of right prediction using only a few data.

5.2 Further Exploration

Note that snowball GCN is not the best architechture of GCN and we can use Block Krylov Subspace Method to get better performance. [4] The GAT network VCHN used is also a primitive GAT and can be improved by some tricks. [6]

VCHN is a method of ensemble learning and better basic model is hopefully to get a better ultimate VCHN model.

References

[1] Zhuolin Liao, Xiaolin Zhang, Wei Su, and Kun Zhan. View-consistent heterogeneous network on graphs with few labeled nodes. *IEEE Transactions on Cybernetics*, 2022.

[2] Thomas N. Kipf and Max Welling. Semi-supervised classification with graph convolutional networks. *CoRR*, abs/1609.02907, 2016.

[3] Petar Veličković, Guillem Cucurull, Arantxa Casanova, Adriana Romero, Pietro Liò, and Yoshua Bengio. Graph attention networks. *arXiv e-prints*, page arXiv:1710.10903, October 2017.

[4] Break the ceiling: Stronger multi-scale deep graph convolutional networks. *CoRR*, abs/1906.02174, 2019. Withdrawn.

[5] Mohammad Rasool Izadi, Yihao Fang, Robert Stevenson, and Lizhen Lin. Optimization of Graph Neural Networks with Natural Gradient Descent. *arXiv e-prints*, page arXiv:2008.09624, August 2020.

[6] Yangkun Wang. Bag of tricks of semi-supervised classification with graph neural networks. *CoRR*, abs/2103.13355, 2021.