Dynamic Graph Convolutional Networks by Manifold Regularization

Sichao Fu[1], Weifeng Liu[1,*[0000−0002−5388−9080]], Yicong Zhou[2[0000−0002−4487−6384]], Zheng-Jun Zha[3], and Liqiang Nie[4[0000−0003−1476−0273]]

[1] College of Information and Control Engineering, China University of Petroleum (East China), Qingdao 266580, China
liuwf@upc.edu.cn
[2] Faculty of Science and Technology, University of Macau, Macau 999078, China
yicongzhou@um.edu.mo
[3] School of Information Science and Technology, University of Science and Technology of China, Hefei 230026, China
zhazj@ustc.edu.cn
[4] School of Computer Science and Technology, Shandong University, Qingdao 266237, China
nieliqiang@gmail.com

Abstract. Recently, the most typical graph convolutional networks (GCN) has achieved excellent performance. GCN successfully generalizes traditional convolutional neural networks to encode arbitrary graphs. However, GCN only fuses the static structure information. It is difficult to guarantee that its structure information is optimal. To tackle the above problem, in this paper, we propose a manifold regularized dynamic graph convolutional network (MRDGCN). The proposed MRDGCN automatically updates the structure information by manifold regularization until model fitting. In particular, we build an optimization convolution layer formulation to acquire the optimal structure information. Thus, MRDGCN can automatically learn high-level sample features to improve the performance of data representation learning. To demonstrate the effectiveness of our proposed model, we apply MRDGCN on the semi-supervised classification tasks. The extensive experiment results on human activity dataset validate the performance of MRDGCN compared with GCN and other semi-supervised learning methods.

Keywords: Graph convolutional networks · Semi-supervised learning · Human activity recognition.

1 Introduction

With the development of virtual reality and augmented reality, human activity recognition (HAR) has attracted much attention in many areas including video surveillance and accident warning. However, traditional conventional shallow learning algorithms including support vector machine [8] cannot extract more

representative sample features and meet development needs, which directly affect the results of the image classification tasks.

To acquire high-level sample features from massive images, deep learning (DL) was introduced and has been demonstrated to be an effective method. The most successful method is semi-supervised learning with manifold regularization (MRSSL), which uses the manifold structure information of unlabeled and labeled samples distribution by regarding it as a regularization term of the objective function. MRSSL methods are only effective for regular Euclidean data. There exists vast amounts of non-Euclidean data or graph data of arbitrary structures. In recent years, spectral convolution methods [2] have received an increasing attention. Kipf and Welling [7] presented a graph convolutional networks to learn the sample features by fusing the direct neighbors relationships of each node.

However, the above method depend on the static samples distribution, which limit the range of its application. In this paper, we propose a dynamic graph convolutional network based on manifold regularization (MRDGCN) for semi-supervised classification. We introduce a manifold regularization term to the objective function, which can drive the objective function to change over the potential sample distribution manifold. When the objective function value cannot meet a specific threshold, MRGCN separately updates or optimizes its manifold structure information (except the first convolution layer) and network weight matrix until model fitting. After many times training iteration, our proposed MRGCN can acquire optimal structure information. In addition, we make an optimization and derivation for the convolution layer formulation of GCN, and then propose a general graph structure learning framework. Finally, MRDGCN can extract more high-level sample features by fusing its dynamic structure information to improve the performance of the semi-supervised classification. To verify the performance of our proposed MRDGCN model, we have tested on CAS-YNU-MHAD dataset for human activity recognition. The experiments results prove that the proposed MRDGCN algorithm shows better classification performance.

The remainder of this paper is arranged as follows: Section 2 details describes our proposed MRDGCN framework. Large-scale experimental results are presented in Section 3. At the end of this paper, we give the conclusion.

2 Manifold Regularized Dynamic Graph Convolutional Networks

We first introduce the traditional manifold regularized framework to original objective function of GCN, and then propose a dynamic graph structure learning (DGSL) method. Following, we give the optimization scheme of DGSL on each training iteration. Finally, we analyze the implement process of a two-layer MRDGCN. Figure 1 shows the general framework of our proposed MRDGCN.

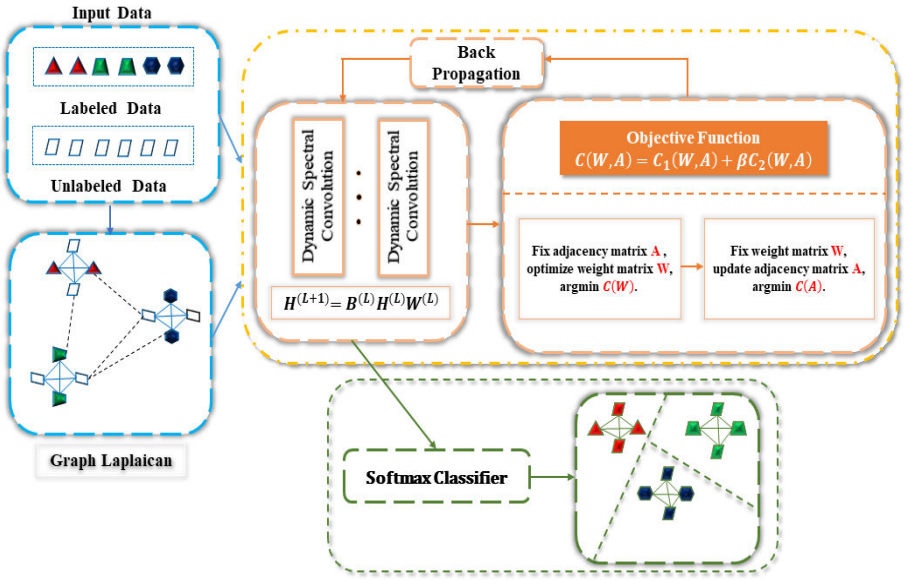

Fig. 1. The framework of the manifold regularized dynamic graph convolutional networks.

2.1 Dynamic Graph Structure Learning

To build a linear and deep model, Kipf and Welling limited the order of the Chebyshev polynomials ($K = 1$), in other words, it only considered the direct relationships between any two samples. Finally, it acquired a linear convolution layer formulation, i.e. $H^{(L+1)} = \sigma(\widetilde{B}H^{(L)}W^{(L)})$. $\widetilde{B}$ is equal to $\widetilde{D}^{-\frac{1}{2}}(A+I_N)\widetilde{D}^{-\frac{1}{2}}$. $W^{(L)}$ is weight parameter matrix to be learned in the training iteration process. The detailed evolution process can be found in [7].

However, GCN relied on a static sample manifold distribution information $\widetilde{B}$ that is computed according to the input data. In real life, the local geometry distribution between data always in changing, thus GCN cannot get the most effective sample features. To tackle this issue, we apply traditional manifold regularization framework to the cross entropy loss function of GCN, and then propose a DGSL model to encode the complex data structure. Therefore, our proposed DGSL can be denoted as the following optimization issue, i.e.

$$\begin{aligned} C(W,A) &= C_1(W,A) + \beta C_2(W,A) \\ &= -y \log Y + \beta tr(Y^T M Y) \end{aligned} \tag{1}$$

In DGSL, $C_1(W,A)$ denotes the cross entropy loss. $tr(Y^T MY)$ represents the manifold regularization term. M denotes that how to use more effective method to better preserve the local geometric distribution of samples. In this paper, we use universally used graph Laplacian matrix. y is the label information matrix

of the input sample. Y denotes the probability distribution matrix of samples, i.e. we can get Y by taking the extracted sample features $H^{(L+1)}$ of the last convolution layer as the input of the Softmax function. Thus, the above objective function can be written as:

$$C(W, A) = -y \log Y + \beta tr(Y^T L Y) \tag{2}$$

We can regard the objective function of DGSL as a dual optimization issue, i.e. we need to update or optimize two variable W and A. β is a balance parameter of the objective function.

2.2 Optimization Scheme

In this paper, we use the alternative optimization strategy to update adjacency matrix and weight matrix. We first fix the adjacency matrix A and optimize weight matrix W, thus the above objective function can be simplified as the following form:

$$C(W) = -y \log Y + \beta tr(Y^T L Y) \tag{3}$$

Where the initial A of each convolution layer is calculated according the input samples by k-NearestNeighbor with the Euclidean distance. In this paper, this problem can be solved by the gradient descent method [1].

Following, we fix weight matrix W, the above objective function on updating adjacency matrix A (except for the first convolution layer) can be denoted as

$$C(A) = -y \log Y + \beta tr(Y^T L Y) \tag{4}$$

The A can be updated according to output sample features of final convolution layer of the last training iteration. With the update of adjacency matrix A, we also further make an optimization for the weight matrix W. Our proposed DGSL will repeat this process until the objective function value do not come down for many epochs.

2.3 A Two-layer MRDGCN

To increase the scale adaptability of model, GCN let $\lambda_{max} = 2$. With the diversification of adjacency matrix A computing methods and the differences between different datasets, $\lambda_{max} = 2$ has the limitation. To solve this problem, we make an optimization again, and then the spectral convolution with one-order polynomial can be simplified as the following form:

$$\begin{aligned} g_\theta(L) \star X &= \sum_{k=0}^{k=1} \theta_k T_k(\widetilde{L}) X \\ &= \theta_0 X + \theta_1 (\frac{2}{\lambda_{max}} L - I_N) X \end{aligned} \tag{5}$$

Where θ_0 and θ_1 denote the filter parameters, i.e. weight matrix $W^{(L)}$ of each convolution layer. To further avoid the overfitting issue of the above formula, we can get the following expression by using a single parameter θ, i.e.

$$g_\theta(L) \star X = \theta \frac{2}{\lambda_{max}}(I_N + D^{-\frac{1}{2}} A D^{-\frac{1}{2}})X \tag{6}$$

To further avoid vanishing gradients and instabilities by stacking the above form to build a deep network model, we also introduce the renormalization trick that is suggested in [7]. And then we get an optimization convolution layer formulation for a signal $X \in R^{n \times m}$ with n samples and m dimensional sample features, which is named the manifold regularized dynamic graph convolutional networks (MRDGCN), i.e.

$$\begin{aligned} g_\theta(L) \star X &= \frac{2}{\lambda_{max}} \widetilde{D}^{-\frac{1}{2}} \widetilde{A} \widetilde{D}^{-\frac{1}{2}} X\theta \\ &= \frac{2}{\lambda_{max}} B^{(L)} H^{(L)} W^{(L)} \end{aligned} \tag{7}$$

Where $\theta \in R^{m \times z}$ or $W^{(L)}$ is filter parameter of each network layer. $\widetilde{A} = A + I_N$ and $B^{(L)} = \widetilde{D}^{-\frac{1}{2}} \widetilde{A} \widetilde{D}^{-\frac{1}{2}}$.

In this paper, we use a two-layer MRDGCN for semi-supervised classification to demonstrate the effectiveness of our proposed method. And then, we can get the following formulation by stacking two-layer Eq. (7), i.e.

$$H^{(2)} = \frac{2}{\lambda_{max}} B^{(1)} RELU(\frac{2}{\lambda_{max}} B^{(0)} X W^{(0)}) W^{(1)} \tag{8}$$

Where X denotes the initial sample features. After two-layer convolution operation, MRDGCN can efficiently learn the sample features with n samples and Q dimensional (The number of Q is equal to classes). Following, we can get the probability distribution matrix Y, and then can get the optimal structure information of the second layer and weight matrix by our proposed optimization scheme after many training iteration. Finally, MRDGCN can extract more representative sample features by fusing the optimal structure information and give the recognition accuracy by Sofmax classifier.

3 Experiments

To demonstrate the effectiveness of our proposed algorithm, in this section, we conduct substantial experiments to test the proposed MRDGCN, MRDGCN-1 and other semi-supervised learning methods, such as the representative GCN [7] and Chebyshev ($K = 2$) [2], on several human activity datasets. The baseline MRFDGCN-1 denotes an optimized GCN variation, i.e. $Eq.(7)$, which still use a static sample structure information. MRDGCN is constructed by the optimized convolution layer formulation and our proposed objective function ($Eq.(2)$). Following, we detailed present the used datasets, experiment parameters setting and comparison results, respectively.

3.1 Human Activity Datasets

CAS-YNU-MHAD dataset [4] is composed of 1086 human actions with 10 different classes. We also use the same feature extraction method that is suggested in [4] to acquire the STCP features [5] of each depth human action image, i.e. we can get a feature matrix with 1086 samples and 81648 dimensional features.

3.2 Experiment Parameters

In our semi-supervised classification experiments, for CAS-YNU-MHAD dataset, 200 samples are randomly selected as test set, 200 samples are randomly choose as validation set, and the remaining part for training set . In its training set, we randomly chose 10%, 20%, 30%, 40% and 50% samples as labeled data and the rest samples for unlabeled data.

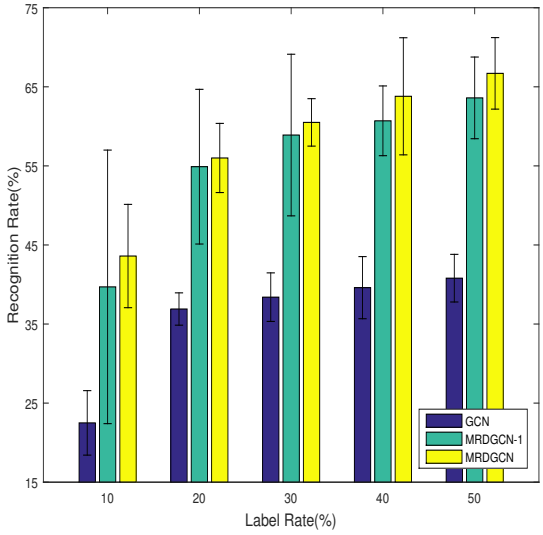

Fig. 2. Recognition rates of all categories on the CAS-YNU-MHAD database.

To minimize the loss value of our proposed objective function, the Adam optimizer [6] with an initialized learning rate of 0.01 is used to train a two-layer MRDGCN. The maximum training iteration up to 200 epochs and MRDGCN will early stopping training if the loss value of objective function do not descend for 10 times continuously on its validation set. We use the weights initialization method that is suggested in [3] during the training process. The L2 regularization with a regularization parameter of 5×10^{-4} and the dropout [9] with a dropout rate of 0.5 is used to solve the overfitting problem of MRDGCN. The

balance parameter β of objective function is 0.001. In addition, on the CAS-YNU-MHAD dataset, the output feature dimension of the first convolution layer is 128.

3.3 Results

Figure 2 show the histogram of the recognition rates for all classes on the CAS-YNU-MHAD dataset. Reported numbers denote the mean recognition accuracy with five run experiments in percent. The y-axis denotes the mean recognition accuracy over all categories for GCN, MRDGCN-1 and MRDGCN model. The x-axis denotes the number of labeled samples randomly chose from training samples. From Figure 2, we can find that the proposed MRDGCN achieves a higher recognition rate than other models, especially only a few labeled samples.

Table 1. Comparison of the different algorithms.

Method	CAS-YNU-MHAD(30%)
Chebyshev($K = 2$)	47.9
MRDGCN	60.5

To further demonstrate the effectiveness of dynamic graph structure learning method and the optimization of convolution layer formulation, we compare other semi-supervised learning algorithms. In addition, we use the 30% labeled samples on the CAS-YNU-MHAD training samples. As shown in Table 1, we can find that our proposed MRDGCN model performs better than the state-of-the-art method.

4 Conclusion

In this paper, we present a dynamic GCN model (MRDGCN) for human activity recognition. During the training process, we employ the alternative solution method to learn the optimal filter parameter and structure information, leading to a dynamic graph structure learning model. In addition, we further make a generalization for the layer-wise propagation rule of GCN, and then MRDGCN can extract richer sample features. Substantial experiment results on human activity dataset including CAS-YNU-MHAD demonstrate that MRDGCN model outperforms the typical GCN.

Acknowledgments

This work was supported in part by the National Natural Science Foundation of China under Grant 61671480, in part by the Fundamental Research Funds for the Central Universities, China University of Petroleum (East China) under Grant 18CX07011A and YCX2019080.

References

1. Burges, C., Shaked, T., Renshaw, E., Lazier, A., Deeds, M., Hamilton, N., Hullender, G.N.: Learning to rank using gradient descent. In: Proceedings of the 22nd International Conference on Machine Learning. pp. 89–96 (2005)
2. Defferrard, M., Bresson, X., Vandergheynst, P.: Convolutional neural networks on graphs with fast localized spectral filtering. In: Advances in Neural Information Processing Systems. pp. 3844–3852 (2016)
3. Glorot, X., Bengio, Y.: Understanding the difficulty of training deep feedforward neural networks. In: Proceedings of the 13th International Conference on Artificial Intelligence and Statistics. pp. 249–256 (2010)
4. Guo, Y., Tao, D., Liu, W., Cheng, J.: Multiview cauchy estimator feature embedding for depth and inertial sensor-based human action recognition. IEEE Transactions on Systems, Man, and Cybernetics: Systems **47**(4), 617–627 (2016)
5. Ji, X., Cheng, J., Feng, W.: Spatio-temporal cuboid pyramid for action recognition using depth motion sequences. In: 2016 Eighth International Conference on Advanced Computational Intelligence. pp. 208–213. IEEE (2016)
6. Kingma, D.P., Ba, J.: Adam: A method for stochastic optimization. In: International Conference of Learning Representation (2014)
7. Kipf, T.N., Welling, M.: Semi-supervised classification with graph convolutional networks. In: International Conference of Learning Representation (2017)
8. Mangasarian, O.L., Wild, E.W.: Multisurface proximal support vector machine classification via generalized eigenvalues. IEEE Transactions on Pattern Analysis and Machine Intelligence **28**(1), 69–74 (2005)
9. Srivastava, N., Hinton, G., Krizhevsky, A., Sutskever, I., Salakhutdinov, R.: Dropout: a simple way to prevent neural networks from overfitting. The Journal of Machine Learning Research **15**(1), 1929–1958 (2014)