Hypergraph Induced Graph Convolutional Network for Multi-Label Image Recognition

Yaqi Tan Zhongchen Ma Yongzhao Zhan Qirong Mao
Computer Science and Communication Engineering
Jiangsu University

JSIIAF

二等奖

Abstact

It is well known that exploiting label connections is important for multi-label image recognition. However, many of the existing methods utilize the connections between label pairs, ignoring high-order connections, thus may result in the degradation of recognition performance. A few methods can capture high-order label connections but have high complexity and low scalability. Inspired by the nature of hypergraph coding high-order connections, we propose a Hypergraph induced Graph Convolutional Network (HI-GCN), which can capture high-order label connections with high adaptivity and scalability. Specifically, we first build an adaptive hypergraph on labels in a data-driven manner, which allows the high-order label connections to be exploited adaptively and scalability. Then, by updating label embeddings via using Hypergraph induced Graph Convolutional network (HI-GCN), the label embeddings are mapped to label classifiers with high-order connections. Experimental results on MS-COCO, SUN and ESP datasets validate the effectiveness of our approach.

Method

1. Framework

The model consists of two parts: image feature extraction and learning of high-order connections classifier based on HI-GCN. In the image feature extraction stage, we uses pre-trained CNN to obtain image features, and converted into 2048-dimensional vectors after maximum pooling. The second part, learning of high-order connections classifier uses hypergraph-induced GCN to map the label embedded vectors to a higher correlated order classifier .

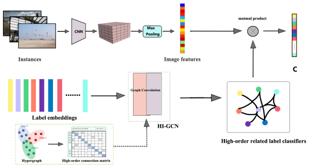


Fig. 4. An overview of our proposed model

2. The implementation process

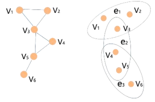
① Hypergraph Construction

For our hypergraph construction, we use a data-driven approach. The data-driven hypergraph is flexible and can be more relevant to particular applications. We regard each label in training set as a vertex. By treating each label as a central vertex, we can get its k-nearest neighbors through (k-Nearest Neighbor) for further processing. Now, we have built a k-uniform hypergraph, with the size of hyperedge is k+1.

However, if k is set too large, this may introduce negative connections and reduce learning performance. To obtain adaptive label connections, we reconstruct each label by its k-nearest neighbors. Specifically, for each label , we use k nearest neighbors to reconstruct $\mathbf{c}_i$ to obtain the reconstruction coefficient :

$$\min_{\boldsymbol{\alpha}_i} \left\|\mathbf{C}_i^*\boldsymbol{\alpha}_i - \mathbf{c}_i\right\|^2 + \lambda\left\|\boldsymbol{\alpha}_i\right\|^2$$

$$\mathbf{H}(v_j, e_i) = \begin{cases} 1 \text{ if } \alpha_{ij} \ge \tau \\ 0 \text{ otherwise} \end{cases} \qquad w(e_i) = \sum_{(i,j)\in e_i} \left|\alpha_{ij}\right|$$



	e1	e2	e3
V1	1	0	0
V2	1	0	0
V3	1	1	0
V4	0	1	1
V5	0	1	1
V6	0	0	1

(a) Simple graph (b) Hypergraph G (c) Incidence matrix

② Hypergraph induced of GCN

GCN is a convolutional network set up specifically for topological data. And the topological data tends to be graph that can only capture pairwise connection. How can hypergraph be used to induce GCN to capture high-order relationships will be the main content introduced in this section.

$$\mathbf{H}^{(l+1)} = \sigma\left(\hat{\mathbf{A}}\mathbf{H}^{(l)}\mathbf{W}^{(l)}\right)$$

The clique expansion algorithm can transform hypergraph into a simple graph. Thus, we get the connection matrix:

$$\hat{\mathbf{A}} = \mathbf{D_c}^{-\frac{1}{2}}\mathbf{H}\mathbf{W}\mathbf{H}^{\mathbf{T}}\mathbf{D_c}^{-\frac{1}{2}}$$

$$\mathbf{F}^{(l+1)} = \sigma\left(\hat{\mathbf{A}}\mathbf{F}^{(l)}\mathbf{W}^{(l)}\right)$$

③Loss function

$$L_{entropy} = \sum_{i=1}^{T}\sum_{j=1}^{n} y_{ij}\log p_{ij} + (1-y_{ij})\log(1-p_{ij})$$

Experiment

1.Datasets

1) SUN： covers 700 categories and 14,340 images (10038 training images and 4302 testing images). With 102 categories of distinguishing attributes, per image has about 6.7 labels.

2) MS-COCO: is a widely used large multi-label dataset. It contains 122585 images (82081 training images and 40504 testing images) with 80 labels, and each image has about 2.9 labels.

3) ESP-GAME: consists of 100,000 images. We selected 60,000 images with 268 labels from it (20,847 training images and 10,169 testing images), each image with an average of 3.1 labels.

2.Results

We compare with state-of-the-art methods, including CNN-RNN, SRN, ResNet-101, ML-GCN on MS-COCO. As can be seen, our method reaches the best performance. Compared to RNN-CNN, our method improves about 21%, and to the resnet-101, improves about 5.1%. Compared with the he latest ML-GCN method, the MAP of our method on MS-COCO is increased by 0.9%. In addition, in order to further verify the effectiveness of our method, we compared with ML-GCN on ESP and SUN datasets, the result on ESP increased by 0.5%, and on SUN increased by 1.5%. The experimental results show that high-order label connection indeed help improve recognition performance.

METHOD	ALL						
	MAP	CP	CR	CF1	OP	OR	OF1
CNN-RNN[6]	61.2	—	—	—	—	—	—
SRN[24]	77.1	81.6	65.4	71.2	82.7	69.9	75.8
RESNET-101[25]	77.3	80.2	66.7	72.8	83.9	70.8	76.8
ML-GCN[1]	81.39	83.3	70.2	76.2	83.5	73.6	78.2
HI-GCN	82.34	84	71.3	77.1	84.1	73.5	78.4

We compared with ML-GCN [1] on ESP and SUN datasets, the result on ESP increased by 0.5%, and on SUN increased by 1.5%. The experimental results show that high-order label connection indeed help improve recognition performance.

DATASETS	METHOD	ALL						
		MAP	*CP*	*CR*	*CF1*	*OP*	*OR*	*OF1*
ESP[23]	ML-GCN[1]	30.96	46.8	22.7	30.6	**60.9**	30.4	**40.6**
	HI-GCN	**31.35**	**48.6**	**30.7**	**37.6**	60.0	**31.4**	**41.2**
SUN[21]	ML-GCN[1]	51.8	56.8	**38.6**	45.0	74.3	53.4	62.2
	HI-GCN	**53.3**	**64.6**	36.3	**46.4**	**74.8**	**55.1**	**63.4**

Conclusion

We propose a hypergraph-induced GCN to capture high-order connection between labels. In this method, we use a data-driven approach to build adaptive hypergraph, which allows hypergraph with variable-sized hyperedges. At the same time, based on the clique expansion algorithm, the hypergraph is transformed into simple graph which helps GCN to map embedding labels to high-order related label classifiers. Experimental results demonstrate the effectiveness of our method.

Main References

[1] Chen, Z. M., Wei, X. S., Wang, P., & Guo, Y. (2019). "Multi-label image recognition with graph convolutional networks." In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (pp. 5177-5186).
[2] Wang, Y., He, D., Li, F., Long, X., Zhou, Z., Ma, J., & Wen, S. (2019). "Multi-Label Classification with Label Graph Superimposing." arXiv preprint arXiv:1911.09243.
[3] Wang, J., Yang, Y., Mao, J., Huang, Z., Huang, C., & Xu, W. (2016). "Cnn-rnn: A unified framework for multi-label image classification." In Proceedings of the IEEE conference on computer vision and pattern recognition (pp. 2285-2294).
[4] Zhang, Z., Bai, L., Liang, Y., & Hancock, E. (2017). "Joint hypergraph learning and sparse regression for feature selection." Pattern Recognition, 63, 291-309.
[5] Wu, B., Jia, F., Liu, W., Ghanem, B., & Lyu, S. (2018). "Multi-label learning with missing labels using mixed dependency graphs." International Journal of Computer Vision, 126(8), 875-896.

Contact

The contact：Yaqi Tan
Mobile phone no.：17766231356
Email: Yaki1997@163.com