结构深层聚类网络

作者：凯鲁嘎吉 - 博客园 http://www.cnblogs.com/kailugaji/

本博文是对Structural Deep Clustering Network这篇文章的展开与叙述。了解这篇文章的前提需要知道GCN图卷积神经网络，DEC深度嵌入聚类，自编码器等。

这是GCN用于聚类的一篇文章。聚类是数据分析的一项基本任务。近年来，从深度学习方法中获得灵感的深度聚类获得了最先进的性能，并引起了广泛的关注。目前的深度聚类方法通常利用深度学习强大的表示能力来提高聚类结果，例如自编码器，这表明学习一种有效的聚类表示是一个至关重要的要求。深度聚类方法的优点是从数据本身中提取有用的表征，而不是从数据的结构中提取有用的表示，这在表示学习中很少受到关注。基于图卷积网络(GCN)在图结构编码方面取得的巨大成功，本文提出了一种结构化深度聚类网络(SDCN)，将结构信息整合到深度聚类中。具体来说，设计了一个传递算子，将自动编码器学习到的表示转换到相应的GCN层，并设计了一个双自监督机制来统一这两种不同的深层神经结构，引导整个模型的更新。通过这种方式，从低阶到高阶的多种数据结构自然地与自动编码器学习到的多种表示相结合。此外，从理论上分析了传递算子，即通过传递算子，GCN将自编码器特有的表示改进为高阶图正则化约束，而自编码器有助于缓解GCN中的过平滑问题。通过全面的实验，证明所提出的模型可以始终比最先进的技术表现得更好。

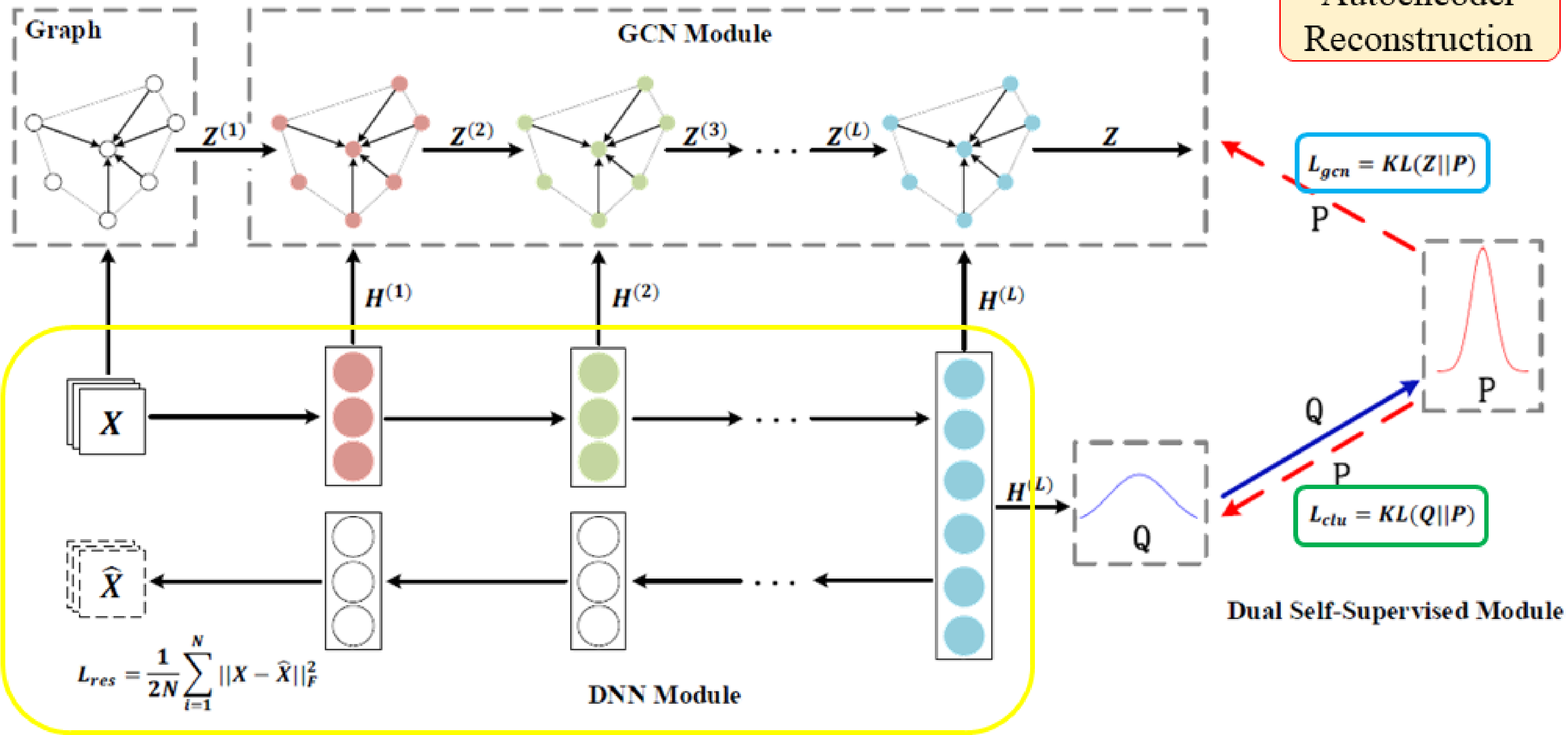
➢ SDCN总体架构与目标函数

$$\mathcal{L} = \mathcal{L}_{res} + \alpha\mathcal{L}_{clu} + \beta\mathcal{L}_{gcn}$$

Autoencoder Reconstruction

DEC KL

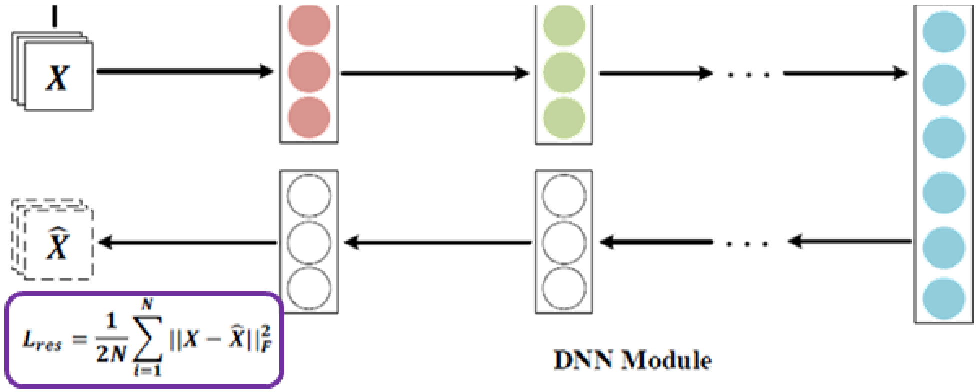
GCN KL



Deyu Bo, Xiao Wang*, Chuan Shi, Meiqi Zhu, Emiao Lu, Peng Cui. Structural Deep Clustering Network. WWW 2020. (CCF-A)

Structural Deep Clustering Network

➢ DNN Module



- 学习一种有效的数据表示对于深度聚类是非常重要的。对于不同类型的数据，有几种可供选择的无监督方法来学习表示。例如去噪自编码器，卷积自编码器，LSTM编解码器和对抗式自编码器。它们是基本自动编码器的变种。
- 在本文中，为了通用性，使用基本的自动编码器来学习原始数据的表示，以适应不同类型的数据特征。

$$\mathcal{L} = \underbrace{\boxed{\mathcal{L}_{res}}}_{\text{Autoencoder Reconstruction}} + \alpha \underbrace{\mathcal{L}_{clu}}_{\text{DEC KL}} + \beta \underbrace{\mathcal{L}_{gcn}}_{\text{GCN KL}}$$

- 编码

$$\mathbf{H}^{(\ell)} = \phi\left(\mathbf{W}_e^{(\ell)}\mathbf{H}^{(\ell-1)} + \mathbf{b}_e^{(\ell)}\right)$$

- 解码

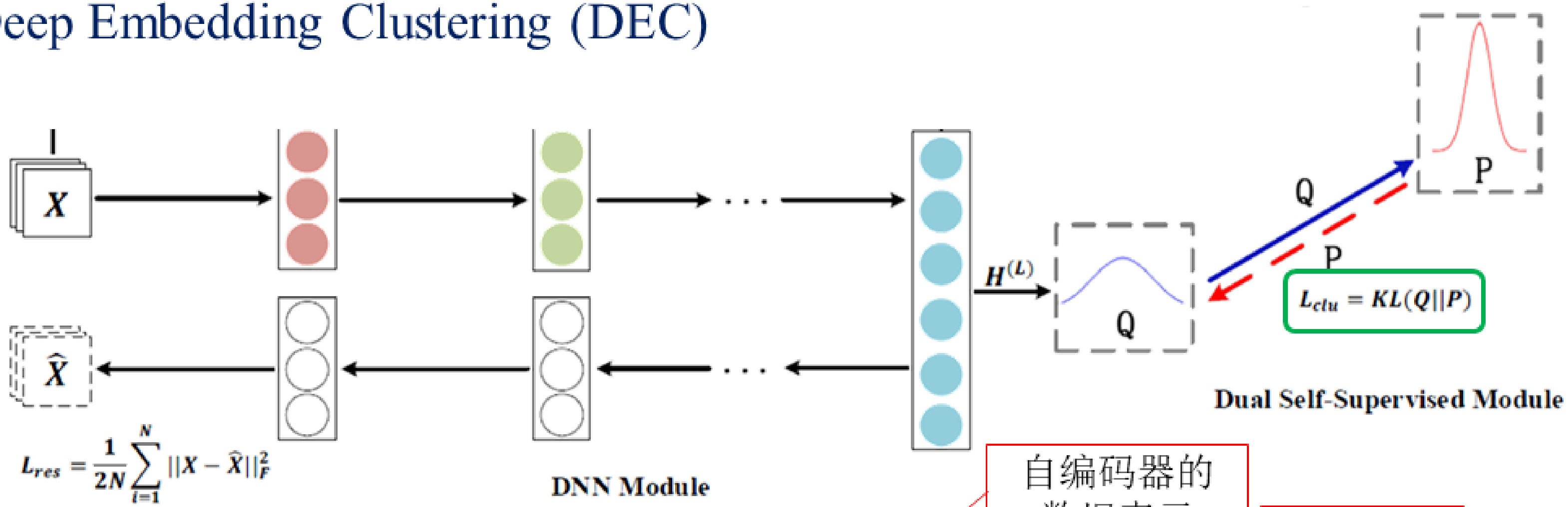
$$\mathbf{H}^{(\ell)} = \phi\left(\mathbf{W}_d^{(\ell)}\mathbf{H}^{(\ell-1)} + \mathbf{b}_d^{(\ell)}\right)$$

- 自编码器重构损失

$$\mathcal{L}_{res} = \frac{1}{2N}\sum_{i=1}^{N}\|\mathbf{x}_i - \hat{\mathbf{x}}_i\|_2^2 = \frac{1}{2N}\|\mathbf{X} - \hat{\mathbf{X}}\|_F^2$$

Deyu Bo, Xiao Wang*, Chuan Shi, Meiqi Zhu, Emiao Lu, Peng Cui. Structural Deep Clustering Network. WWW 2020. (CCF-A)

➢ Deep Embedding Clustering (DEC)



自编码器的数据表示

聚类中心

- 低维学生t分布

$$q_{ij} = \frac{(1 + \|\mathbf{h}_i - \boldsymbol{\mu}_j\|^2 / v)^{-\frac{v+1}{2}}}{\sum_{j'}(1 + \|\mathbf{h}_i - \boldsymbol{\mu}_{j'}\|^2 / v)^{-\frac{v+1}{2}}}$$

$$\mathcal{L} = \mathcal{L}_{res} + \alpha \boxed{\mathcal{L}_{clu}} + \beta \mathcal{L}_{gcn}$$

Autoencoder Reconstruction

DEC KL

GCN KL

- 聚类KL散度损失

$$\mathcal{L}_{clu} = KL(P||Q) = \sum_i \sum_j p_{ij} log \frac{p_{ij}}{q_{ij}}$$

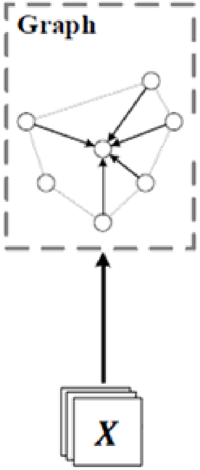
- 高维Shape操作

$$p_{ij} = \frac{q_{ij}^2 / f_j}{\sum_{j'} q_{ij'}^2 / f_{j'}}$$

Deyu Bo, Xiao Wang*, Chuan Shi, Meiqi Zhu, Emiao Lu, Peng Cui. Structural Deep Clustering Network. WWW 2020. (CCF-A)

Structural Deep Clustering Network

➤ KNN Graph



- 计算相似度矩阵S(N*N的矩阵)
  - Heat Kernel(连续数据)

$$S_{ij} = e^{-\frac{\|x_i - x_j\|^2}{t}}$$

  - Dot-product(离散数据)

$$S_{ij} = x_j^T x_i$$

- 取前K个最相似的点作为该数据的邻居
- 构建K近邻无向图
- 得到邻接矩阵A(N*N的矩阵)

➤ GCN Module

网络连接权重

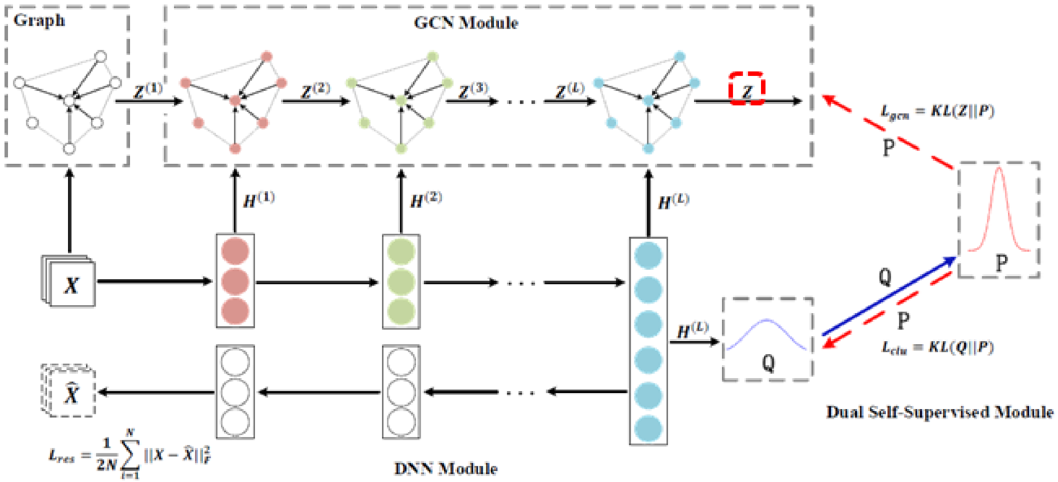
$$Z^{(\ell)} = \phi(\boxed{\widetilde{D}^{-\frac{1}{2}}\widetilde{A}\widetilde{D}^{-\frac{1}{2}}} Z^{(\ell-1)} \boxed{W^{(\ell-1)}})$$

归一化的邻接矩阵L
(半正定对称归一化的Laplacian矩阵L)

Laplacian矩阵:
L=D-A
L: Laplacian矩阵(行/列元素之和为0)
D: 顶点的度矩阵/对角矩阵(对角线上的元素为各顶点的度)
A: 图的邻接矩阵

Deyu Bo, Xiao Wang*, Chuan Shi, Meiqi Zhu, Emiao Lu, Peng Cui. Structural Deep Clustering Network. WWW 2020. (CCF-A)

➢ Dual Self-Supervised GCN Module



- 考虑到自动编码器H学习到的表示能够重构数据本身，并且包含不同有价值的信息，本文将Z和H这两种表示结合在一起，得到一个更完整、更强大的表示。
- 新的Z是每一层GCN得到的Z与每一层DNN得到的H两者的加权求和后的结果。

因为GCN学习到的表示包含两种不同的信息，因此最终的聚类结果由Z得到。 $r_i = \arg\max_j \ z_{ij}$

- GCN的KL散度损失

聚类损失中的P

$$\mathcal{L}_{gcn} = KL(P||Z) = \sum_i \sum_j p_{ij} \log \frac{p_{ij}}{\boxed{z_{ij}}}$$

$$Z^{(1)} = \phi(\widetilde{D}^{-\frac{1}{2}}\widetilde{A}\widetilde{D}^{-\frac{1}{2}}XW^{(1)})$$

$$Z^{(\ell)} = \phi(\widetilde{D}^{-\frac{1}{2}}\widetilde{A}\widetilde{D}^{-\frac{1}{2}}Z^{(\ell-1)}W^{(\ell-1)})$$

$$\boxed{\widetilde{Z}^{(\ell-1)} = (1-\epsilon)Z^{(\ell-1)} + \epsilon H^{(\ell-1)}}$$

$$Z^{(\ell)} = \phi\left(\widetilde{D}^{-\frac{1}{2}}\widetilde{A}\widetilde{D}^{-\frac{1}{2}}\widetilde{Z}^{(\ell-1)}W^{(\ell-1)}\right)$$

$$Z = softmax\left(\widetilde{D}^{-\frac{1}{2}}\widetilde{A}\widetilde{D}^{-\frac{1}{2}}Z^{(L)}W^{(L)}\right)$$

Deyu Bo, Xiao Wang*, Chuan Shi, Meiqi Zhu, Emiao Lu, Peng Cui. Structural Deep Clustering Network. WWW 2020. (CCF-A)

参考文献：

[1] Deyu Bo, Xiao Wang*, Chuan Shi, Meiqi Zhu, Emiao Lu, Peng Cui. Structural Deep Clustering Network. WWW 2020. (CCF-A)

[2] 王啸老师个人主页：https://wangxiaocs.github.io/

深度学习中的拓扑美学：图神经网络报告题目：Dive into the Message Passing Mechanism of Graph Neural Networks

[3] 自编码器、DEC相关：Deep Clustering Algorithms

[4] 如何理解 Graph Convolutional Network（GCN）？