

文章编号:1672-3961(2024)04-0035-07 DOI:10.6040/j.issn.1672-3961.0.2023.160

# 基于高斯分布和 Householder flow 的无监督图嵌入算法

刘国军, 范天祥, 王乃正, 张正达, 齐广智

(哈尔滨工业大学计算学部, 黑龙江 哈尔滨 150001)

**摘要:**为更好地表示节点,提出一种新的图嵌入方法,将节点表示为由均值和方差构成的高斯分布,通过应用一系列可逆 Householder 变换,将相对简单的分布转换为更灵活的分布,可以更好地捕获关于其表示的不确定性。为提高稳定性,采用 Wasserstein 距离进行分布之间的度量。试验结果表明,在多个基准数据集上,使用 Householder 变换的 Graph2Gauss (G2G) 算法比原始模型的链接预测表现更好。通过节点分类的效果可以看出,对于节点信息缺失的图,使用 Wasserstein 距离可以大幅增加节点分类的  $F_1$  分数。

**关键词:**无监督学习; 图嵌入; 高斯分布; Householder flow; Wasserstein 距离

**中图分类号:**TP181 **文献标志码:**A

**引用格式:**刘国军,范天祥,王乃正,等. 基于高斯分布和 Householder flow 的无监督图嵌入算法[J]. 山东大学学报(工学版), 2024, 54(4): 35-41.

LIU Guojun, FAN Tianxiang, WANG Naizheng, et al. Unsupervised graph embedding algorithm based on Gaussian distribution and Householder flow[J]. Journal of Shandong University (Engineering Science), 2024, 54(4):35-41.

## Unsupervised graph embedding algorithm based on Gaussian distribution and Householder flow

LIU Guojun, FAN Tianxiang, WANG Naizheng, ZHANG Zhengda, QI Guangzhi

(Faculty of Computing, Harbin Institute of Technology, Harbin 150001, Heilongjiang, China)

**Abstract:** In order to express nodes better, a novel graph embedding method was proposed, each node was represented as a Gaussian distribution composed of mean and variance. By applying a series of reversible Householder transformations, the relatively simple distribution was transformed into a more flexible distribution, so that the uncertainty about its representation could be better captured. To improve stability, Wasserstein distance was used to measure between distributions. The experimental results showed that on multiple benchmark datasets, Graph2Gauss (G2G) algorithm using Householder transformation performed better than the original model on link prediction. It could be seen from the effectiveness of node classification that using Wasserstein distance could significantly increase the  $F_1$  score for graphs with missing node information.

**Keywords:** unsupervised learning; graph embedding; Gaussian distribution; Householder flow; Wasserstein distance

## 0 引言

随着网络应用的不断普及,用户对数据的需求也不断提高。在新媒体,诸如购物资讯及短视频加入后,各领域每天都会产生庞大的数据,传统的检

索方法与数据存储方式开始面临巨大挑战。快速精确匹配数据和用户、现资讯、商品甚至传统电视节目、广播波段等的精准推送,满足人们个性化、多元化的需求是当前信息技术发展的主要方向。推荐信息定制化、资讯差异化具备重要的现实意义。更为高效的数据组织结构是支撑这些应用实现的

收稿日期:2023-07-05

基金项目:国家自然科学基金资助项目(61976071);黑龙江省联合基金资助项目(LH2020F012)

第一作者简介:刘国军(1979—),男,黑龙江哈尔滨人,副教授,博士生导师,博士,主要研究方向为机器学习、计算机视觉、图像处理、模式识别等。E-mail: hitliu@hit.edu.cn

重要基础。将单个数据与节点对应,标记合适的标签,同时将节点间的联系抽象为网络的边,形成的图网络结构是当前很多领域依赖的技术。图网络技术与精准推荐系统结合<sup>[1]</sup>,能够精准有效地检索信息<sup>[2]</sup>。图网络还能分析节点间关系的相似性和可靠性,自动提取图中隐含的信息。

实际的图网络中,节点间的联系极为稀疏。短视频平台拥有数亿用户,但对于每个用户,与其有关的账户少之又少。如何在压缩图网络的同时保持原有的属性,减少存储空间,成为一个重要的研究方向。随着神经网络研究日益引起人们的关注,无监督神经网络技术为图嵌入算法开辟了新的世界。对于节点嵌入的无监督学习,已提出许多不同的方法。

神经网络图嵌入算法将节点自身属性加入模型,演变为归纳式学习<sup>[3]</sup>,大大提高了模型的拟合能力。一些工作旨在处理图网络数据节点隐含的不确定性,例如由变分自编码器改进的变分图自动编码器<sup>[4]</sup>。GraphSAGE 通过从节点的局部邻域采样和聚合属性生成嵌入<sup>[5]</sup>。与此同时,统计关系学习<sup>[6]</sup>和自监督学习也被引入图嵌入中。

图嵌入学习在深度学习的推动下有了一定的发展<sup>[7-11]</sup>。很多方法用低维空间中的点向量表征图中的节点,然而对于复杂的图数据,仅用一个向量描述节点会使其具有不确定性。基于无监督学习的 Graph2Gauss(G2G)算法把网络中的节点在低维嵌入空间中用高斯分布表示<sup>[12]</sup>,不再是在嵌入空间中用一个低维点向量表示,使得在嵌入空间中用一个区域表征图的节点成为一种可替代的思路。G2G 通过提出的  $K$  跳排序方法及锚点采样策略,保留了图的结构信息,大大降低了计算复杂度,又能取得好的结果。

但是,G2G 也存在 2 个问题:首先,G2G 算法在表征图中节点为高斯分布时,只关注了隐藏空间中高斯分布的对角协方差形式,可以在一定程度上降低计算复杂度,但同时使隐藏空间中的高斯分布缺少了一般性,不能足够灵活匹配真实的后验概率;其次,G2G 算法将图结构数据的节点嵌入一个高斯分布后,使用 KL(Kullback-Leibler)散度度量高斯分布之间的距离,严格来说,KL 散度并不能作为一个衡量距离的指标。KL 散度从形式上看并不具有对称性,且不满足三角不等式。在隐藏空间中度量对应图中节点的距离不能保持图的传递性,然而传递性是图和网络中非常重要的特性之一。

为解决以上问题,进一步提高算法的有效性和鲁棒性,本研究引入一系列可逆的 Householder 变换

对均值和方差进行变换,使用 Wasserstein 距离度量高斯分布之间的距离,在多个基准数据集上验证模型的效果。

## 1 G2G 无监督图嵌入算法

$G=(A, X)$  是带有附属信息的有向图,其中  $A \in \mathbf{R}^{N \times N}$  为图中边的邻接矩阵,  $X \in \mathbf{R}^{M \times D}$  为图中节点的附属信息,每个节点的附属信息是维度为  $D$  的向量。根据以上信息,在低维高斯空间  $h_i = N(\boldsymbol{\mu}_i, \boldsymbol{\sigma}_i)$  中,生成前后的节点附属信息和网络结构相似度应当尽可能不变,其中,  $\boldsymbol{\mu}_i$  和  $\boldsymbol{\sigma}_i$  分别为均值和方差。

为了保留图的结构信息,G2G 提出  $K$  跳排序方法。对于一个固定节点  $i$ ,离节点  $i$  有  $k$  步的节点集合

$$N_{ik} = \{j \in V | i \neq j, \min(s_p(i, j), K) = k\}, \quad (1)$$

式中:  $V$  为所有节点的集合;  $s_p(i, j)$  为返回  $i, j$  最短路径长度,若没有从  $i$  到  $j$  的路径,则为无穷;  $K$  为超参数,表示考虑的最远距离。得到对于节点  $i$  的所有  $K$  跳点集,如  $N_{i1}$  为距离节点  $i$  最近的节点集合,即与节点  $i$  直接有边的节点,  $N_{i2}$  表示距离节点  $i$  两跳的节点集合,但同时该集中的节点未曾出现在  $N_{i1}$  中,以此类推到  $N_{iK}$ 。

对于原始图,  $N_{i1}$  与节点  $i$  的距离小于  $N_{i2}$  与节点  $i$  的距离,  $N_{i2}$  与节点  $i$  的距离小于  $N_{i3}$  与节点  $i$  的距离,以此类推。原始图数据中距离近的节点更相似,更有可能有相似的特征,而对于节点在嵌入空间中,应保持一致,即在  $N_{i1}$  中的节点在隐藏空间的高斯分布与节点  $i$  的高斯分布的相似程度,应该比  $N_{i2}$  中的节点在隐藏空间的高斯分布与节点  $i$  的高斯分布相似程度更大,可表示为:

$$\Delta(h_i, h_{k_1}) < \Delta(h_i, h_{k_2}) < \dots < \Delta(h_i, h_{k_K}), \quad (2)$$

式中,  $\forall k_1 \in N_{i1}, \forall k_2 \in N_{i2}, \dots, \forall k_K \in N_{iK}$ 。在隐藏空间中的 2 个高斯分布  $h_i$  和  $h_j$  的差异性

$$\Delta(h_i, h_j) = D_{\text{KL}}(N_j, N_i) = \frac{1}{2} \left[ \text{tr}(\boldsymbol{\sigma}_i^{-1} \boldsymbol{\sigma}_j) + (\boldsymbol{\mu}_i - \boldsymbol{\mu}_j) \boldsymbol{\sigma}_i^{-1} (\boldsymbol{\mu}_i - \boldsymbol{\mu}_j) - L - \ln \frac{d_{\text{et}}(\boldsymbol{\sigma}_j)}{d_{\text{et}}(\boldsymbol{\sigma}_i)} \right], \quad (3)$$

式中,  $\text{tr}(\cdot)$  为矩阵的迹,  $d_{\text{et}}(\cdot)$  为矩阵的行列式,  $L$  为均值向量的维度。对于 2 个高斯分布,其对应的 KL 散度越大,表明 2 个分布差异性越大,越不相似。

G2G 定义两点之间的能量  $E_{ij}$  为其对应的 KL 散度,即  $E_{ij} = D_{\text{KL}}(N_j \| N_i)$ ,则基于能量的损失函数

$$L_{\text{oss}} = \sum_i \sum_{k < l} \sum_{j_k \in N_{ik}} \sum_{j_l \in N_{il}} (E_{ijk}^2 + \exp(-E_{ijl})) = \sum_{(i,j_k,j_l) \in D_t} (E_{ijk}^2 + \exp(-E_{ijl})), \quad (4)$$

式中,  $D_t$  为所有有效的三元组集合,  $D_t = \{(i, j_k, j_l) \mid s_p(i, j_k) < s_p(i, j_l)\}$ 。

通过  $K$  跳排序对图的邻居节点排序, 并对隐空间中得到的分布也有相应的排序限制, 但很难找到一个合适的方法对图中所有的节点对限制都满足, G2G 提出相应的采样策略, 并调整对应的基于能量的损失函数, 使采样策略调整前后一致。

## 2 基于 Householder 变换的 G2G 算法

G2G 经过非线性变换得到与嵌入节点分布相关的均值与方差, 用神经网络进行参数化可以大幅提高计算效率, 但是产生的结果始终是标准正态分布, 即协方差矩阵是对角化的。在多维向量空间中, 真实的分布往往是任意而又复杂多变的, 因此, G2G 无法灵活拟合真实的情况。

对此, 本研究考虑用标准化流 (normalizing flow, NF) 思想解决这一问题<sup>[13-14]</sup>。对于一个简单的后验分布, 使用一系列的可逆矩阵作用在简单分布上, 可以对分布进行变换, 逼近真正的后验分布。变换一次后得到一个分布, 再使用多次可逆变换得到更加真实的分布。基于上述理论, 文献[15]提出的 Householder 流 (Householder flow, HF) 是一种保体积流, 使雅可比行列式等于 1, 大大降低计算的复杂度。此外, 文献[16]还给出了 Householder 变换的几何解释。

在通常情况下, 任何满秩矩阵  $B$  由特征值进行分解<sup>[17]</sup>:

$$B = UDU^T, \quad (5)$$

式中,  $U$  为正交特征向量,  $D$  为对角矩阵。由文献[18]的推导表明, 本研究可以使用 Householder 变换表示任何正交矩阵, 并且 Householder 矩阵也是正交矩阵。因此, Householder 矩阵可以在保留矩阵形状的前提下近似一个满秩协方差矩阵。HF 的定义如下: 对于一个给定的向量  $z^{t-1}$ , 变换其反射镜像平面可以由一个 Householder 向量  $v_t$  定义, 并且  $v_t$  垂直于超平面, 向量  $z^{t-1}$  的镜像点

$$z^t = (I - 2 \frac{v_t v_t^T}{\|v_t\|_2^2}) z^{t-1} = H_t z^{t-1}, \quad (6)$$

式中  $H_t = I - 2 \frac{v_t v_t^T}{\|v_t\|_2^2}$  为 Householder 矩阵。

图的特征矩阵和邻接矩阵作为输入, 经过编码层, 输出学习得到的均值和方差, 经过 Householder 变换得到更接近真实后验分布的均值和方差, 如图 1 所示。

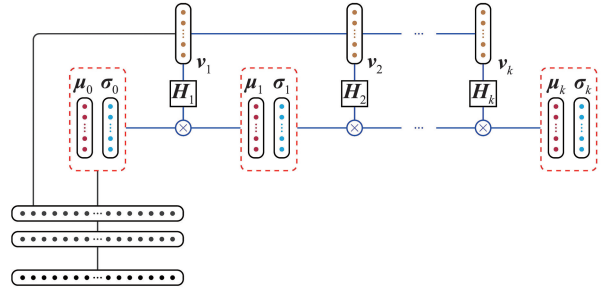


图 1 基于 Householder 变换的 G2G 示意图  
Fig.1 The diagram of G2G based on Householder transformation

## 3 基于 Wasserstein 距离的 G2G 算法

G2G 模型的最终目的是减小嵌入分布和真实分布之间的差异, 即缩小损失函数的散度。最优结果为两分布完全相等, KL 散度为 0。但是根据物理学定义, 参考系选取不同, 结果也会产生偏差。参考系分别选取真实分布与选择嵌入分布时, 计算结果互不相同。计算 2 个数据量相似度的方式不会因参考系的不同而发生改变, 所以本研究考虑用 Wasserstein 距离代替 KL 散度进行计算<sup>[19]</sup>。

Wasserstein 距离可以计算多维空间中 2 个节点间的距离, 也可以得出 2 个分布之间的差异。Wasserstein 距离也称推土机距离<sup>[20]</sup>, 推土机是形容该算法的计算过程: 考虑 2 个离散分布  $P$  和  $Q$ , 为了让  $P, Q$  变成相同的分布, 挨个观察每个节点并计算距离; 此时  $P, Q$  好比两堆土, 期望为将  $P$  移动至  $Q$  上, 使形状即分布完全一致, 移动代价表示为移动的土量乘以移动的距离, 则 Wasserstein 距离可以直观地看成 2 个土堆的移动问题。土堆移动存在代价最小的方案, 此方案即 2 个分布的 Wasserstein 距离。2 个高斯分布间的 Wasserstein 距离<sup>[21]</sup>

$$d_{\text{ist}} = W_2(N(\mu_1, \sigma_1), N(\mu_2, \sigma_2)) = \|\mu_1 - \mu_2\|_2 + t_r(\sigma_1 + \sigma_2 - 2(\sigma_1^{1/2} \sigma_2 \sigma_1^{1/2})^{1/2}). \quad (7)$$

而高斯分布的  $\sigma_i$  为对角阵, 即  $\sigma_1 \sigma_2 = \sigma_2 \sigma_1$ , 则式(7)简化为:

$$W_2(N(\mu_1, \sigma_1), N(\mu_2, \sigma_2))^2 = \|\mu_1 - \mu_2\|_2^2 + \|\sigma_1^{1/2} - \sigma_2^{1/2}\|_F^2. \quad (8)$$

Wasserstein 距离相比 KL 散度优势在于, 即使 2 个分布的支撑集没有重叠或重叠非常少, 仍然能反映 2 个分布的距离, 而 KL 散度可能无意义。此外, 从公

式形式上看, Wasserstein 距离满足包括对称性和三角不等式等特性, 使 Wasserstein 距离成为一个测量距离的指标, 其中满足三角不等式保证了 Wasserstein 距离能够使节点保持传递性。因此, Wasserstein 距离非常适合作为测量隐空间内两点差异性的方法。

## 4 算法总结

**算法 1** 基于 Householder 变换的 G2G 算法

**输入:** 图的邻接矩阵  $A \in \mathbf{R}^{N \times N}$ 、节点数  $N$ 、最大采样跳数  $K$ 。

**输出:** 损失  $L_{\text{oss}}$ 。

采样  $\Omega \in \mathbf{R}^{N \times K} \leftarrow U(N_{ik})$ ;

抽取三元组  $D_t \leftarrow D_{\text{append}}(\cup_{s_p(i,j_k) < s_p(i,j_l)} (i, j_k, j_l))$ ;

搭建神经网络  $G2G()$ , 初始化化参数  $W, b, v_0$ , 遍历所有序列  $n \in \Omega$ ;

$$\sigma, \mu \leftarrow G2G(A, N) \begin{cases} \mu = W_\mu X + b_\mu \\ \sigma = W_\sigma + b_\sigma \end{cases};$$

$$\mathbf{Z}^0 \leftarrow \mu, \sigma;$$

$$\mathbf{Z}^t \leftarrow \mathbf{H}_t \mathbf{H}_{t-1} \cdots \mathbf{H}_1 \mathbf{Z}^0;$$

$$E_{ij} = \text{mean}(\sum D_{\text{KL}}(N_j || N_i));$$

由  $E_{ij}$  求得损失  $L_{\text{oss}}$ ;

end。

算法采样时, 将每一个原点  $i \in V$  的采样序列  $N_{ik}$  做并集 ( $U$ ), 得到目的序列  $\Omega$ , 采样序列  $N_{ik}$  由 G2G 提出  $K$  跳排序方法得到。构造三元组, 为计算损失函数提供输入。由神经网络得到嵌入参数  $\sigma$  和  $\mu$  后进行 HF 变换。计算基于能量的损失函数, 得到结果。算法的进一步改进是针对损失函数的改变, 引入 Wasserstein 距离, 将 KL 散度的计算换成更加稳定的推土机距离。

## 5 试验

### 5.1 数据集

数据集的节点数、边数及类别数如表 1 所示。Cora 是机器学习论文引文网络数据集, 包含 70 个

领域, 整个网络包含 19 793 篇论文 (节点) 和 65 311 条边, 边表示论文之间的引用关系<sup>[22]</sup>。但大多数方法使用的是其子集 Cora-ML, 从 Cora 中提取出来, 包含 7 个领域的 2 995 篇文章, 同时包含 8 416 条边。其他数据集为常用引文关系中的常见数据集。Citeseer 数据集包含 6 个类别的 4 230 篇论文和 5 358 条边<sup>[23]</sup>。DBLP 是关于计算机领域参考书目的数据集, 包含 4 个领域的 17 716 篇论文和 105 734 条边<sup>[24]</sup>。Pubmed 包含 18 230 个节点和 79 612 条边, 划分为 3 个类别<sup>[25]</sup>。

表 1 试验数据集

Table 1 Experimental dataset

数据集	节点数/个	边数/条	类别数/个
Cora-ML	2 995	8 416	7
Cora	19 793	65 311	70
Citeseer	4 230	5 358	6
DBLP	17 716	105 734	4
Pubmed	18 230	79 612	3

本试验使用以上多个基准数据集, 与多个传统的图嵌入算法<sup>[26-27]</sup>及原始 G2G 算法进行对比, 对本研究提到的改进分别进行测试, 对比 G2G 的原始试验数据, 比较算法性能。本研究采用曲线下面积  $A_{\text{UC}}$  与平均精度  $A_p$ <sup>[28]</sup> 2 个维度对边链接预测的准确性进行评价。

### 5.2 Householder 变换对链接预测的影响

链接预测是图结构嵌入评测中常用的测试方法。本研究在评测时将隐藏一部分边的信息, 将处理后的图网络结构交给神经网络处理, 即部分节点之间是否链接对神经网络是不可知的。为了方便比较, 本研究将数据集的边集分为 3 份, 验证集占 5%, 测试集占 10%, 剩余为训练集。在验证集和测试集中, 非边的数量相同。

本试验根据 G2G 中评估的参数进行测试, 并在相同环境中对原始 G2G 算法进行测试。使用 Householder 矩阵对分布进行变换, 分别测试针对均值变换及同时针对均值和方差进行变换。所有变换次数为 1。试验结果如表 2 所示。

表 2 链接预测的结果对比

Table 2 Comparison of link prediction performance

方法	Cora-ML		Cora		Citeseer		DBLP		Pubmed	
	$A_{\text{UC}}$	$A_p$	$A_{\text{UC}}$	$A_p$	$A_{\text{UC}}$	$A_p$	$A_{\text{UC}}$	$A_p$	$A_{\text{UC}}$	$A_p$
Logistic Regression	90.01	89.75	86.58	86.51	81.70	79.10	82.04	81.91	90.50	90.99
node2vec <sup>[26]</sup>	76.80	75.26	79.95	78.98	83.04	83.74	95.42	95.33	95.42	95.33
TADW <sup>[27]</sup>	81.26	81.34	76.56	78.06	70.14	72.93	65.67	59.85	62.72	68.02
TRIDNR <sup>[24]</sup>	84.51	85.69	81.61	81.08	87.23	88.87	92.01	91.62	—	—
GAE <sup>[29]</sup>	96.65	96.67	97.91	98.07	92.31	93.88	95.78	96.67	96.07	96.12

表2(续)

方法	Cora-ML		Cora		Citeseer		DBLP		Pubmed	
	$A_{UC}$	$A_P$	$A_{UC}$	$A_P$	$A_{UC}$	$A_P$	$A_{UC}$	$A_P$	$A_{UC}$	$A_P$
G2G oh	95.93	95.31	98.45	98.50	96.11	95.47	<b>98.06</b>	<b>98.47</b>	97.01	97.12
G2G oh $\mu+\sigma$	<b>99.04</b>	<b>99.05</b>	<b>99.40</b>	<b>99.43</b>	99.10	<b>99.20</b>	97.73	98.12	<b>97.22</b>	<b>97.50</b>
G2G oh $\sigma$	98.76	98.60	99.11	99.18	99.16	98.86	97.17	97.56	96.77	97.06
G2G	96.54	96.10	98.74	98.63	95.02	94.50	98.36	98.37	96.42	95.66
G2G $\mu+\sigma$	<b>98.77</b>	<b>98.80</b>	<b>99.44</b>	<b>99.45</b>	97.76	96.18	<b>98.53</b>	<b>98.60</b>	<b>96.55</b>	<b>95.94</b>
G2G $\sigma$	98.71	98.56	99.27	99.20	<b>97.93</b>	<b>97.57</b>	98.41	98.39	96.50	95.82

注:“—”表示无试验结果。 $\sigma$ 表示仅针对方差进行变换。 $\mu+\sigma$ 表示同时变换正态分布的均值与方差。G2G oh表示不带附属信息的图结构,G2G表示携带附属信息的图结构,这两部分为本研究在pytorch环境配置下的运行结果。

由表2可知,不同的变换策略对不同数据集有不同的效果,对于Cora、Cora-ML及Pubmed数据集,同时变换正态分布的均值与方差的效果最好;对于Citeseer数据集,2种变换策略的成绩相差无几,但比不变换的结果要好;对于DBLP数据集,文献[12]中不变换的算法评测结果要略好于变换策略。总体来说,对于不带附属信息的图,对编码层生成的分布需要进行变换,而使用何种变换策略是需要根据图网络自身决定,才能达到最好的边预测结果。

对于带有附属信息的图的评测结果,算法在不同数据集之间的表现差异较大,但变换后的效果都比原来好。在Cora、Cora-ML及Citeseer数据集中,变换后的评测效果大幅提升。对于DBLP与

Pubmed数据集,3种策略对边预测的影响相差无几,文献[12]不变换算法的评测结果略低于变换策略。总体来看,对分布进行变换的策略,在带有附属信息的图结构中也得到了较好的结果。对于一个新数据集,使用对均值进行变换的算法得到良好分类分数的可能性更大。

### 5.3 不同损失函数对预测的影响

使用不同的距离度量函数对训练过程的参数更新有不同的影响,本研究针对原始的G2G及G2G oh算法分别替换了损失函数,并与原损失函数KL散度进行对比,结果如表3所示。

从表3中可以看出,在不进行Household变换的算法中,使用Wasserstein距离与KL散度的效果相似。

表3 不同距离度量函数的链接预测分数对比

Table 3 Comparison of link prediction for different distance metric functions

方法	Cora-ML		Cora		Citeseer		DBLP		Pubmed	
	$A_{UC}$	$A_P$	$A_{UC}$	$A_P$	$A_{UC}$	$A_P$	$A_{UC}$	$A_P$	$A_{UC}$	$A_P$
G2G oh W	96.09	95.65	98.45	98.48	95.75	95.42	98.04	98.47	96.93	97.09
G2G W	96.52	96.08	98.78	98.66	95.04	94.49	98.32	98.34	96.55	95.72
G2G oh KL	95.93	95.31	98.45	98.50	96.11	95.47	98.06	98.47	97.01	97.12
G2G KL	96.54	96.10	98.74	98.63	95.02	94.50	98.36	98.37	96.42	95.66

注:W表示使用Wasserstein距离作为距离度量函数,KL表示使用KL散度作为距离度量函数。

### 5.4 Householder 矩阵变换次数的影响

Householder变换属于线性复杂度的变换,所以考虑多次HF变换研究其对试验结果的影响。本研究对编码层输出的均值与方差分别进行不同次数的变换,在其他参数相同的情况下比较边预测的结果,用Citeseer数据集进行评估,如图2所示。由图2可以看出,不同模型在不同HF变换次数的对比试验中, $A_{UC}$ 呈现出不同的变化趋势。但是总体来说,多次变换并没有给试验结果带来很大改善,有的预测结果甚至出现较大降幅。考虑Citeseer数据集的结构特点,作为一个引文网络,过多地变换图网络节点的映射关系,可能会使模型过度复杂,增加模型预测的难度。考虑HF变换是线性变换,结合试

验结果的实际预测和算法运行的开销,当一次变换足够精准,能够等价多次变换时,考虑采用单次HF变换进行相应试验。

在图结构数据中,节点分类同样是较为常用的一种分类任务,常用于数据分析与预测,例如视频网站对视频内容的分类,或是社交网站利用节点分类算法对用户打标签进行分类。因此对于图嵌入算法,比较常用节点分类评估分数。本试验使用Citeseer数据集进行评估。将整个训练集删除节点分类标签,通过无监督学习的方式进行图嵌入。根据文献[30]的评估方案,随机选择节点进行标签标记,总共标记的节点数占比从2%到10%,共评估5组。将被标记的节点集合作为新的训练集合,使用

逻辑回归对剩余的节点进行预测,将预测结果作为评估分数。

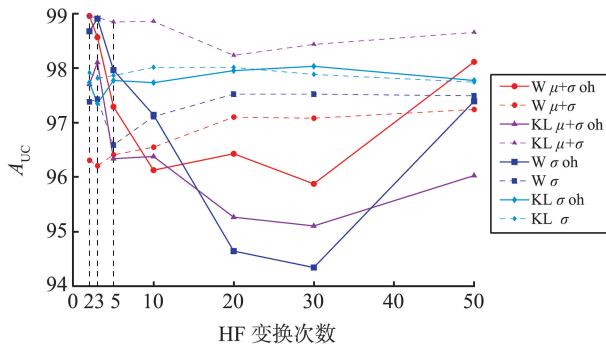
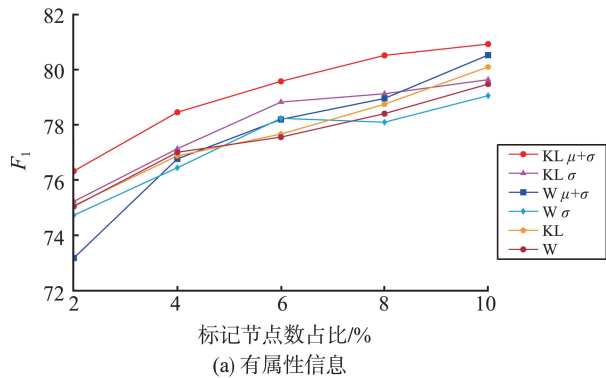


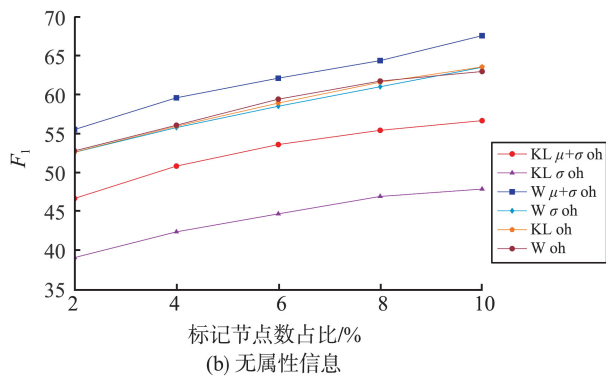
图2 各模型在 Citeseer 数据集进行多次 Householder 变换的表现

Fig.2 Performance of each model in multiple Householder transformations on Citeseer

节点分类结果分为2组,其中一组是带有附属信息的 G2G 模型,如图3(a)所示,另一组比较分散的是不带附属信息的 G2G 模型,如图3(b)所示。



(a) 有属性信息



(b) 无属性信息

图3 各模型在 Citeseer 数据集节点分类任务的评测结果

Fig.3 Evaluation results of each model in node classification task on Citeseer

在附带信息的 G2G 模型中可以看到,经过 HF 变换和替换损失函数后,节点的预测性能略有上升,得益于模型优化带来的图嵌入的准确性上升,因此每一种类别的节点分离更加清晰。此外,替换损失函数后,在有标签节点比例较低的部分提高更加明显,可以看出 Wasserstein 距离更适用于训练集数据较少的情况。随着有标签节点比例上升,各模

型之间的差距下降,接近统一。

在不附带信息的 G2G 模型中,由于节点本身不带有附属信息,节点的分类全部依靠其相邻节点的链接情况和节点在图中的位置,所以  $F_1$  结果比附带信息的图低很多。在这种情况下,Wasserstein 距离使评价分数大幅提升。

## 6 结论

本研究针对无监督图嵌入模型 G2G 存在的 2 个问题进行了优化,在图的边链接预测和节点分类测试中获得了较好的结果,在一定程度上提高了嵌入的准确性。通过引入一系列可逆的 Householder 变换对均值和方差进行变换,获得更加灵活的近似后验分布,更好地捕获节点的不确定性。在变换时使用 2 种策略,一种是针对均值和方差同时进行变换,另一种是仅变换方差。经过边链接预测试验对比发现,在大多数情况下,使用 Householder 变换的 G2G 比原始模型的链接预测表现更好。测试多次变换对试验结果的影响,经过对比发现,大多数模型在经过多次变换后, $A_{UC}$  呈现出不同的变化趋势,有时模型过度复杂并不能带来更好的结果。通过节点分类效果可以看出,对于节点信息缺失的图(即仅有结构信息,没有节点附属内容),使用 Wasserstein 距离可以大幅增加节点分类的  $F_1$  分数。

### 参考文献:

- [1] HE X, DENG K, WANG X, et al. LightGCN: simplifying and powering graph convolution network for recommendation[C]// Proceedings of the 43rd International ACM SIGIR Conference on Research and Development in Information Retrieval. New York, USA: ACM, 2020: 639-648.
- [2] WAIKHOM L, PATGIRI R. Graph neural networks: methods, applications, and opportunities[EB/OL]. (2021-08-24) [2023-07-01]. <https://arxiv.org/abs/2108.10733>.
- [3] NIEPERT M, AHMED M, KUTZKOV K. Learning convolutional neural networks for graphs[C]// Proceedings of the 33rd International Conference on Machine Learning. New York, USA: PMLR, 2016: 2014-2023.
- [4] YANG L, CHEUNG N M, LI J, et al. Deep clustering by gaussian mixture variational autoencoders with graph embedding[C]// Proceedings of the 2019 IEEE/CVF International Conference on Computer Vision. Seoul, Korea: IEEE, 2019: 6440-6449.

- [5] HAMILTON W, YING Z, LESKOVEC J. Inductive representation learning on large graphs [C]//Proceedings of the 31st International Conference on Neural Information Processing Systems. Long Beach, USA: Curran Associates Inc., 2017: 1024-1034.
- [6] QU M, BENGIO Y, TANG J. GMNN: graph Markov neural networks [C]//Proceedings of the 36th International Conference on Machine Learning. Long Beach, USA: PMLR, 2019: 5241-5250.
- [7] WANG X, BO D, SHI C, et al. A survey on heterogeneous graph embedding: methods, techniques, applications and sources [J]. IEEE Transactions on Big Data, 2023, 9(2): 415-436.
- [8] GOYAL P, FERRARA E. Graph embedding techniques, applications, and performance: a survey [J]. Knowledge-Based Systems, 2018, 151: 78-94.
- [9] VELICKOVIC P, CUCURULL G, CASANOVA A, et al. Graph attention networks [EB/OL]. (2018-02-04) [2023-07-01]. <https://arxiv.org/abs/1710.10903>.
- [10] YING C, CAI T, LUO S, et al. Do transformers really perform badly for graph representation? [EB/OL]. (2021-11-24) [2023-07-01]. <https://arxiv.org/abs/2106.05234v1>.
- [11] WANG Z, YU D, LI Q, et al. SR-HGN: semantic-and relation-aware heterogeneous graph neural network [J]. Expert Systems with Applications, 2023, 224: 119982.
- [12] BOJCHEVSKI A, GÜNNEMANN S. Deep gaussian embedding of graphs: unsupervised inductive learning via ranking [EB/OL]. (2018-02-27) [2023-07-01]. <https://arxiv.org/abs/1707.03815v2>.
- [13] REZENDE D, MOHAMED S. Variational inference with normalizing flows [C]//Proceedings of the 32nd International Conference on Machine Learning. Lille, France: PMLR, 2015: 1530-1538.
- [14] KOBYZEV I, PRINCE S J D, BRUBAKER M A. Normalizing flows: an introduction and review of current methods [J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2020, 43(11): 3964-3979.
- [15] HOUSEHOLDER A S. Unitary triangularization of a nonsymmetric matrix [J]. Journal of the ACM, 1958, 5(4): 339-342.
- [16] LIU G, LIU Y, GUO M, et al. Variational inference with Gaussian mixture model and householder flow [J]. Neural Networks, 2019, 109: 43-55.
- [17] 王梅. 几类特殊对称矩阵的分解 [D]. 天津: 天津工业大学, 2019.  
WANG Mei. Decomposition of some special symmetric matrices [D]. Tianjin: Tianjin University of Technology, 2019.
- [18] SUN X, BISCHOF C. A basis-kernel representation of orthogonal matrices [J]. SIAM Journal on Matrix Analysis and Applications, 1995, 16(4): 1184-1196.
- [19] ZHU D, CUI P, WANG D, et al. Deep variational network embedding in Wasserstein space [C]//Proceedings of the 24th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining. London, UK: ACM, 2018: 2827-2836.
- [20] 刘泽华. Wasserstein 空间上的积分及其在分布式鲁棒优化中的应用 [D]. 南京: 南京大学, 2020.  
LIU Zehua. Calculus on Wasserstein space and its application in distributed robust optimization [D]. Nanjing: Nanjing University, 2020.
- [21] ÇELIK T Ö, JAMNESHAN A, MONTÚFAR G, et al. Wasserstein distance to independence models [J]. Journal of Symbolic Computation, 2021, 104: 855-873.
- [22] MCCALLUM A K, NIGAM K, RENNIE J, et al. Automating the construction of internet portals with machine learning [J]. Information Retrieval, 2000, 3: 127-163.
- [23] GILES C L, BOLLACKER K D, LAWRENCE S. CiteSeer: an automatic citation indexing system [C]//Proceedings of the Third ACM Conference on Digital Libraries. Pittsburgh, USA: ACM, 1998: 89-98.
- [24] PAN S, WU J, ZHU X, et al. Tri-party deep network representation [C]//Proceedings of the Twenty-Fifth International Joint Conference on Artificial Intelligence. New York, USA: IJCAI, 2016: 1895-1901.
- [25] SEN P, NAMATA G, BILGIC M, et al. Collective classification in network data [J]. AI Magazine, 2008, 29(3): 93-106.
- [26] GROVER A, LESKOVEC J. Node2vec: scalable feature learning for networks [C]//Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery & Data Mining. San Francisco, USA: ACM, 2016: 855-864.
- [27] YANG C, LIU Z, ZHAO D, et al. Network representation learning with rich text information [C]//Proceedings of the Twenty-Fourth International Joint Conference on Artificial Intelligence. Buenos Aires, Argentina: AAAI, 2015: 2111-2117.
- [28] FAWCETT T. An introduction to ROC analysis [J]. Pattern Recognition Letters, 2006, 27(8): 861-874.
- [29] KIPF T N, WELING M. Variational graph auto-encoders [EB/OL]. (2016-11-26) [2023-07-01]. <https://arxiv.org/abs/1611.07308>.
- [30] PEROZZI B, AL-RFOU R, SKIENA S. Deepwalk: online learning of social representations [C]//Proceedings of the 20th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining. New York, USA: ACM, 2014: 701-710.