智能系统学报
CAAI TRANSACTIONS ON INTELLIGENT SYSTEMS

基于图卷积集成的网络表示学习

常新功, 王金珏

引用本文:
常新功,王金珏. 基于图卷积集成的网络表示学习[J]. 智能系统学报, 2022, 17(3): 547–555.
CHANG Xingong,WANG Jinjue. Network representation learning using graph convolution ensemble[J]. *CAAI Transactions on Intelligent Systems*, 2022, 17(3): 547–555.

在线阅读 View online: https://dx.doi.org/10.11992/tis.202107048

DOI: 10.11992/tis.202107048
网络出版地址: https://kns.cnki.net/kcms/detail/23.1538.TP.20220324.1345.002.html

基于图卷积集成的网络表示学习

常新功，王金珏
（山西财经大学 信息学院，山西 太原 030006）

摘　要：针对现有网络表示学习方法泛化能力较弱等问题，提出了将 stacking 集成思想应用于网络表示学习的方法，旨在提升网络表示性能。首先，将 3 个经典的浅层网络表示学习方法 DeepWalk、Node2Vec、Line 作为并列的初级学习器，训练得到三部分的节点嵌入拼接后作为新数据集；然后，选择图卷积网络 (graph convolutional network, GCN) 作为次级学习器对新数据集和网络结构进行 stacking 集成得到最终的节点嵌入，GCN 处理半监督分类问题有很好的效果，因为网络表示学习具有无监督性，所以利用网络的一阶邻近性设计损失函数；最后，设计评价指标分别评价初级学习器和集成后的节点嵌入。实验表明，选用 GCN 集成的效果良好，各评价指标平均提升了 1.47~2.97 倍。

关键词：网络表示学习；集成学习；图卷积网络；社交网络；深度学习；特征学习；节点嵌入；信息网络嵌入
中图分类号：TP301.6　　**文献标志码**：A　　**文章编号**：1673−4785(2022)03−0547−09

中文引用格式：常新功，王金珏. 基于图卷积集成的网络表示学习 [J]. 智能系统学报，2022，17(3)：547–555.
英文引用格式：CHANG Xingong, WANG Jinjue. Network representation learning using graph convolution ensemble[J]. CAAI transactions on intelligent systems, 2022, 17(3): 547–555.

Network representation learning using graph convolution ensemble

CHANG Xingong, WANG Jinjue
(School of Information, Shanxi University of Finance and Economics, Taiyuan 030006, China)

Abstract: Aimed at the weak generalization ability of the existing network representation learning methods, this paper proposes a stacking ensemble method that is applied to network representation learning to improve the performance of network representation. Three classical shallow network representation learning methods, namely, DeepWalk, Node2Vec, and Line, are used initially as parallel primary learners. After training, the embedded representation and three spliced node parts are obtained. Following this, the graph convolutional network (GCN) is selected as the secondary learner to stack and integrate the network structure and the new data set. This is done to obtain the final node embedded representation. GCN efficiently deals with semisupervised classification problems. As network representation learning is unsupervised, the first-order proximity of the network has been used to design the loss function. Finally, evaluation indices are designed to test the primary learner and integrated node eigenvector representation. The experimental results show that after the introduction of the integration method, each index improves by approximately 1.47–2.97 times.

Keywords: Network represents learning; Ensemble learning; Graph convolution network; Social network; Deep learning; Feature learning; Node embedding; Information network embedding

近年来，基于网络数据结构的深度学习十分流行，广泛应用于学术领域和工业领域。网络包括节点和边，其中节点表示实体，边表示节点之间的关系。现实世界中很多数据都可以表示为网络，例如社交网络[1-2]、生物–蛋白网络[3] 等。利用网络分析挖掘有价值的信息备受关注，因为高效的网络分析不仅处理节点分类[1]、链路预测[2]、网络可视化[4-5] 等下游任务时有着很好的效果，而且在金融欺诈、推荐系统等场景下都有实际的应用价值。例如，在社交网络中通过节点分类可以对不同的用户推荐不同的物品；在生物网络中，可以通过分析已知的疾病与基因关系预测潜在的致病基因等。

由于网络数据的非欧几里得结构，大多数传统的网络分析方法不适合使用机器学习技术解决。网络表示学习[6-8] 很好地解决了上述问题，通过将节点映射到低维空间中，节点用学习生成的

收稿日期：2021−07−23.　**网络出版日期**：2022−03−25.
基金项目：国家自然科学基金项目（61906110）；山西财经大学研究生创新项目（21cxxj088）.
通信作者：常新功. E-mail: c_x_g@126.com.

低维、稠密的向量重新表示，同时尽可能保留网络中包含的结构信息。因此，网络被映射到向量空间中就可以使用经典的机器学习技术处理很多网络分析问题。现有的网络表示学习方法主要分为以下 3 类：

1) 基于矩阵分解的网络表示学习。Roweis 等[9]提出的局部线性表示算法 (locally linear embeding, LLE) 假设节点和它的邻居节点都处于同一流形区域，通过它的邻居节点表示的线性组合近似得到节点表示；He 等[10]提出的保留局部映射算法 (locality preserving projections, LPP) 通过对非线性的拉普拉斯特征映射方法进行线性的近似得到节点表示；Tu 等[11]提出的图形分解算法 (max margin deep walk, MMDW) 通过对邻接矩阵分解得到节点表示。Cao 等[12]提出的 GraRep 算法通过保留节点的 k 阶邻近性保留全局网络结构。

2) 基于浅层神经网络的网络表示学习。Perozzi 等[13]提出的 DeepWalk 算法通过随机游走遍历网络中的节点得到有序的节点序列，然后利用 Skip-Gram 模型预测节点的前后序列学习得到节点的向量表示；Grover 等[14]提出的 Node2Vec 改进了 DeepWalk 的随机游走过程，通过引进两个参数 p 和 q 控制深度优先搜索和广度优先搜索；Tang 等[15]提出的 Line 算法能够处理任意类型的大规模网络，包括有向和无向、有权重和无权重，该算法保留了网络中节点的一阶邻近性和二阶邻近性。

3) 基于深度学习的网络表示学习。Wang 等[16]提出的 SDNE 算法利用深度神经网络对网络表示学习进行建模，将输入节点映射到高度非线性空间中获取网络结构信息。Hamilton 等[17]提出的 GraphSAGE 是一种适用于大规模网络的归纳式学习方法，通过聚集采样到的邻居节点表示更新当前节点的特征表示。Wang 等[18]提出的 GraphGAN 引入对抗生成网络进行网络表示学习。上述研究方法大多是设计一种有效的模型分别应用不同的数据集学习得到高质量的网络表示，但是单一模型的泛化能力较弱。为了解决此问题，目前有学者提出使用集成思想学习网络表示，Zhang 等[19]提出的基于集成学习的网络表示学习，其中 stacking 集成分别将 GCN 和 GAE 作为初级模型，得到两部分节点嵌入拼接后作为节点特征，其与原始图数据构成新数据集，最后将三层 GCN 作为次级模型处理新数据集，使用部分节点标签进行半监督训练。

本文引入了 stacking 集成方法学习网络表示。集成方法是对于同一网络并行训练多个较弱的个体学习器，每个个体学习器的输出都是网络表示，然后采用某种结合策略集成这些输出进而得到更好的网络表示。stacking 集成方法是集成方法的一种，结合策略是学习法，即选用次级学习器集成个体学习器的输出。次级学习器的选择是影响结果的重要因素，现有工作证明 Kipf 等[20]提出的图卷积神经网络[21](graph convolutional network, GCN) 在提升网络分析性能上有着显著的效果，GCN 通过卷积层聚合网络中节点及邻居的信息，根据归一化拉普拉斯矩阵的性质向邻居分配权重，中心节点及邻居信息加权后更新中心节点的特征表示。

综上所述，本文的贡献有以下几点：

1) 提出了基于 stacking 集成学习的网络表示学习，并行训练多个较弱的初级学习器，并将它们的网络表示拼接，选用 GCN 作为次级学习器，聚合中心节点及邻居信息得到最终的网络表示，这样可得到更好的网络表示。

2) 利用网络的一阶邻近性设计了损失函数；

3) 设计了评价指标 MRR、Hit@1、Hit@3、Hit@10，分别评价初级学习器和集成后的网络表示，验证了提出的算法具有较好的网络表示性能，各评价指标平均提升了 1.47~2.97 倍。

1　问题定义

定义 1　给定网络 $G=<V,E>$，其中 V 表示节点集合，E 表示节点之间的边集合，记 $v_i \in V$ 表示一个节点，$e_{i,j}=(v_i,v_j) \in E$ 表示一条边，由 E 构建邻接矩阵 $\boldsymbol{A} \in \mathbf{R}^{n\times n}$ 表示网络的拓扑结构，$n=|V|$，若 $e_{i,j} \in E$，则 $A_{i,j}>0$，若 $e_{i,j} \notin E$，则 $A_{i,j}=0$。

定义 2[6]　给定网络 G，每个节点的属性特征是 m 维，G 有 n 个节点，则网络 G 对应的节点特征矩阵 $\boldsymbol{H} \in \mathbf{R}^{n\times m}$。网络表示学习的目标是根据网络中任意节点 $v_i \in V$ 学习得到低维向量 $\boldsymbol{Z} \in \mathbf{R}^{n\times d}$，其中 $d \ll n$。学习到的低维向量表示可客观反映节点在原始网络中的结构特性。例如，相似的节点应相互靠近，不相似的节点应相互远离。

定义 3　一阶邻近性[15]。网络中的一阶邻近性是指两个节点之间存在边，若节点 v_i 和 v_j 之间存在边，这条边的权重 $w_{i,j}$ 表示 v_i 和 v_j 之间的一阶邻近性，若节点 v_i 和 v_j 之间没有边，则 v_i 和 v_j 之间的一阶邻近性为 0。

定义 4　二阶邻近性[15]。网络中一对节点 v_i 和 v_j 之间的二阶邻近性是指它们的邻域网络结构之间的相似性，令 $l_i=(w_{i,1},w_{i,2},\cdots,w_{i,|V|})$ 表示节点

v_i 与其他所有节点的一阶邻近性, v_i 和 v_j 的二阶邻近性由 l_i 和 l_j 的相似性决定。

定义 5　集成学习[22]。集成学习是构建多个个体学习器 $\ell_1, \ell_2, \cdots, \ell_n$, 再用某种结合策略将它们的输出结合起来, 结合策略有平均法、投票法和学习法。给定网络 G, 定义 2 中的网络表示学习方法可作为个体学习器, 其结构如图 1。若个体学习器是同种则是同质集成, 否则是异质集成。

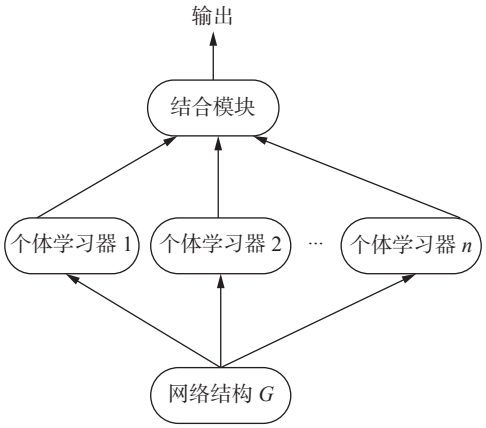


图 1　集成学习结构

Fig. 1　Structure of ensemble learning

定义 6　stacking 集成学习[22]。stacking 集成学习的结合策略是学习法, 对于同一网络通过 k 个初级学习器 $\ell_1, \ell_2, \cdots, \ell_k$ 学习得到 k 部分节点嵌入的特征向量 $z_0, z_1, \cdots, z_{k-1}$, 其嵌入维度均为 d 维, 然后按节点将 z_i, $i \in [0, k-1]$ 对应拼接得到嵌入 z, 其嵌入维度是 $k \times d$ 维, 最后使用次级学习器 ℓ 得到最终的嵌入 z', 为了方便对比设置其嵌入维度也是 d 维。

2　基于 GCN 集成的网络表示学习方法

本文将 stacking 集成思想引入网络表示学习, 对于同一网络数据基于 3 个初级学习器生成 3 部分嵌入并将其拼接, 然后选取 GCN 作为次级学习器得到最终的嵌入, 最后使用评价指标进行评价, 具体流程如图 2 所示。

2.1　初级学习器

初级学习器选择 DeepWalk[13]、Node2Vec[14] 和 Line[15]。DeepWalk[13] 发现在短的随机游走中出现的节点分布类似于自然语言中的单词分布, 于是采用广泛使用的单词表示学习模型 Skip-Gram 模型学习节点表示; Node2Vec[14] 认为 DeepWalk 的表达能力不足以捕捉网络中连接的多样性, 所以设计了一个灵活的网络邻域概念, 并设计随机游走策略对邻域节点采样, 该策略能平滑地在广度优先采样 (BFS) 和深度优先采样 (DFS) 之间进行插值; Line[15] 是针对大规模的网络嵌入, 可以保持一阶和二阶邻近性。图 3 给出了一个说明示例, 节点 6 和节点 7 之间边的权重较大, 即节点 6 和节点 7 有较高的一阶邻近性, 它们在嵌入空间的距离应很近; 虽然节点 5 和节点 6 没有直接相连的边, 但是它们有很多共同的邻居, 所以它们有较高的二阶邻近性, 在嵌入空间中距离也应很近。一阶邻近性和二阶邻近性都很重要, 一阶邻近性可以用两个节点之间的联合概率分布度量, v_i 和 v_j 的一阶邻近性如式 (1):

$$p_1(v_i, v_j) = \frac{1}{1 + e^{-z_i^{\mathrm{T}} z_j}} \tag{1}$$

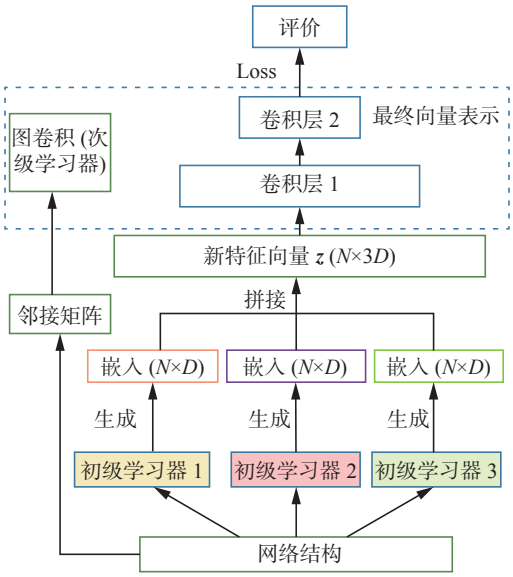


图 2　基于 GCN 集成的网络表示学习结构

Fig. 2　Network representation learning structure based on GCN ensemble method

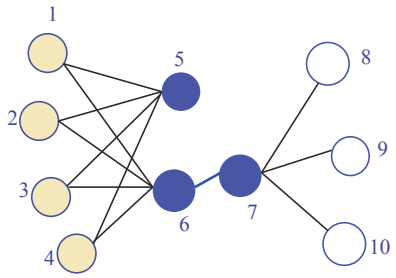


图 3　网络简单示例

Fig. 3　Simple example of network

二阶邻近性通过节点 v_i 的上下文节点 v_j 的概率建模, 即

$$p_2(v_j|v_i) = \frac{e^{z_j^{\mathrm{T}} z_i}}{\sum_k e^{z_k^{\mathrm{T}} z_i}}$$

条件分布意味着在上下文中具有相似分布的节点彼此相似, 通过最小化两种分布和经验分布的 KL 散度, 可以得到既保持一阶邻近性又保持

二阶邻近性的节点表示。

2.2　次级学习器

引入 stacking 集成方法学习网络表示，选择 DeepWalk[13]、Node2Vec[14] 和 Line[15] 作为初级学习器。若初级学习器是同种的则为同质集成，否则为异质集成。3 个初级学习器学习得到的嵌入分别是 z_1、z_2、z_3，且维数均设为 d，并将 z_1、z_2、z_3 拼接得到嵌入 z'，维数为 $3\times d$。这个过程中不使用节点的辅助信息，仅利用网络的拓扑结构学习节点的特征表示。选用 GCN 图卷积网络模型[21] 作为 stacking 的次级学习器，学习得到最终的嵌入 z，维数是 d。

GCN 模型的输入有两部分，若网络 G 有 N 个节点，则一部分是嵌入 z'，每个节点有 H 维，其大小为 $N\times H$，另一部分是网络 G 的邻接矩阵 $\boldsymbol{A}$，其大小为 $N\times N$。首先，通过计算得到归一化矩阵 $\hat{\boldsymbol{A}}\in\mathbf{R}^{n\times n}$，如式 (2)：

$$\hat{\boldsymbol{A}}=\boldsymbol{D}^{-\frac{1}{2}}\tilde{\boldsymbol{A}}\boldsymbol{D}^{-\frac{1}{2}} \tag{2}$$

式中：$\tilde{\boldsymbol{A}}=\boldsymbol{A}+\boldsymbol{I}$，$\boldsymbol{I}\in\mathbf{R}^{n\times n}$是单位矩阵；$\boldsymbol{D}$ 是$\tilde{\boldsymbol{A}}$的度矩阵。拉普拉斯矩阵有良好的性质，下面针对对称归一化拉普拉斯矩阵作详解。$\hat{\boldsymbol{D}}^{-1}\hat{\boldsymbol{A}}$即根据邻居个数取平均，$\hat{\boldsymbol{D}}^{-1/2}\hat{\boldsymbol{A}}\hat{\boldsymbol{D}}^{-1/2}$是对称归一化拉普拉斯矩阵，其与$\hat{\boldsymbol{D}}^{-1}\hat{\boldsymbol{A}}$的对比示例如图 4 所示。

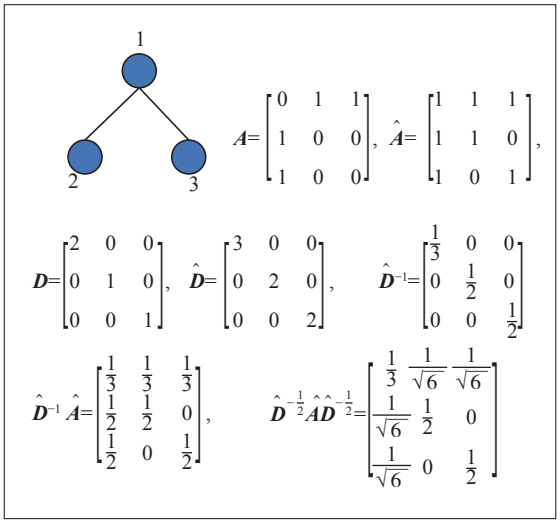


图 4　拉普拉斯矩阵示例

Fig. 4　Example of Laplacian matrix

与$\hat{\boldsymbol{D}}^{-1}\hat{\boldsymbol{A}}$相比，$\hat{\boldsymbol{D}}^{-1/2}\hat{\boldsymbol{A}}\hat{\boldsymbol{D}}^{-1/2}$是对称矩阵，改变的值含义如下：节点 1 有两个邻居节点 2、3，但是节点 2、3 分别只有一个邻居，所以节点 1 对节点 2、3 影响要大一点，对应矩阵中第一行第二列和第三列，由1/3变为$1/\sqrt{6}$，其值变大了；节点 3 的邻居只有节点 1，但是节点 1 有两个邻居，所以节点 3 对节点 1 的影响要小一点，对应矩阵中第三行第一列，由1/2变为$1/\sqrt{6}$，其值变小了。相比节点 i 的邻居数量，节点 i 的邻居 j 的邻居数量多，所以节点 i 对邻居 j 的影响小。

然后，GCN 的整体结构如图 5 所示，用式 (3)、(4) 描述：

$$f^{(0)}(z',\boldsymbol{A})=\mathrm{ReLU}(\hat{\boldsymbol{A}}z'\boldsymbol{W}^{(0)}) \tag{3}$$

$$z=f^{(1)}(f^{(0)},\boldsymbol{A})=\tanh(\hat{\boldsymbol{A}}f^{(0)}\boldsymbol{W}^{(1)}) \tag{4}$$

式中：$\hat{\boldsymbol{A}}$如式 (3) 所示计算，设计了两层卷积层，第一层使用 ReLU 激活函数，第二层使用 tanh 激活函数，$\boldsymbol{W}^{(0)}$ 和 $\boldsymbol{W}^{(1)}$ 是需要训练的权重矩阵，最终得到输出 z。对于网络$G=<V,E>$，这里的 GCN 是基于空域的图卷积网络，其时间复杂度为$O(|E|HTF)$，H 是输入特征维数 384，T 是中间层维数 256，F 是输出层维数 128。

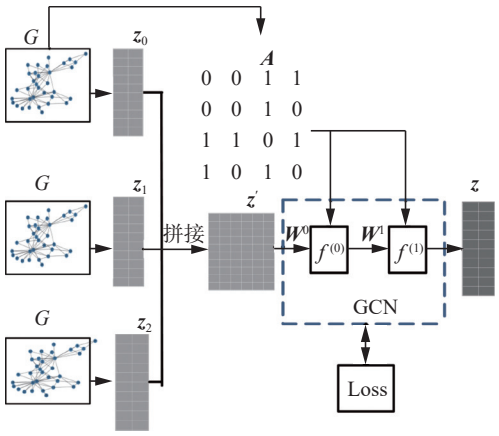


图 5　图卷积集成网络模型结构

Fig. 5　Structure of GCN ensemble model

2.3　损失函数

利用网络的一阶邻近性设计损失函数，根据噪声分布对边采样负边，任意边的损失函数为

$$-\log\sigma(\mathrm{z}_i^{\mathrm{T}}\cdot z_j)+\frac{1}{K}\sum_{i=1}^{K}E_{v_n\sim P_n(v)}[\log\sigma(\mathrm{z}_i^{\mathrm{T}}\cdot z_n)] \tag{5}$$

式中：第一项是根据观测到的边即正例的 loss；第二项是为正例采样的负例的 loss；K 是负边的个数；$\sigma(x)=1/(1+\exp(-x))$是 sigmoid 函数；设置$P_n(v)\propto d_v^{3/4}$，其在文献 [23] 中提出，d_v是节点v的出度。

边采样根据边的权重选用 alias table[15] 方法进行，从 alias table 中采样一条边的时间复杂度是 $O(1)$，负采样的时间复杂度是$O(d(K+1))$，d 表示出度，K 表示 K 条负边，所以每步的时间复杂度是 O(dK)，步数的多少取决于边的数量$|E|$，因此计算损失的时间复杂度为$O(dK|E|)$，与节点数量 N 无关。此边采样策略在不影响准确性的情况下提高了效率。

2.4　评价指标

通过 2.3 节损失函数影响模型的训练学习，得到最终的网络嵌入表示 z，对于网络表示学习

的无监督性, 设计评价指标[24] 评价网络表示学习的好坏。对于节点 v_i 和 v_j 之间的边即一个正例, 由一对节点 (v_i, v_j) 表示, 一个正例对应采样 K 条负边, 即采样 K 个点$(n_1,n_2,\cdots,n_k)$, 其中$i,j\notin(1,K)$, 构成负例集合$\{(v_i,n_1),(v_i,n_2),\cdots,(v_i,n_k)\}$。

衡量一对节点的相似度可计算它们网络表示的内积, 正例(v_i,v_j)的相似度$s=z_i\cdot z_j^{\mathrm{T}}$, 负例的相似度$s_p=z_i\cdot z_{n_p}^{\mathrm{T}}$, $p=(1,2,\cdots,K)$, 相似值越大越好, 所以将 s_p 的值由大到小排序, 记录 s 插入$\{s_p\}$的索引 ranking, 索引是从 0 开始的, 衡量指标需要的是排名位置, 所以令 ranking=ranking+1, ranking 越小说明网络表示学习的嵌入越有效。

上文针对一个正例计算得到了一个 ranking, 对于整个网络设计指标如表 1 所示。

表 1　评价指标

Table 1　Evaluating indicator

评价指标	含义
MRR	所有ranking的倒数的平均值
Hit@1/3/10	ranking在负例相似度排名的前1/3/10个

评价数据边的数量为$|E^{'}|$, 时间复杂度为$O(K|E^{'}|)$。

2.5　算法描述

基于图卷积集成的网络表示主要包括 3 个步骤, 首先得到初级学习器的网络表示, 然后用 stacking 集成, 其中次级学习器选用 GCN。对于网络表示学习的无监督性在 GCN 模型中设计了损失函数, 也设计了其测试指标, 相关算法如算法 1 所示。

算法 1　基于图卷积集成的网络表示

输入　网络 $G=(V,E)$, 窗口大小 w, 节点表示向量维度 d, 每个节点随机游走次数γ, 随机游走序列长度 t, 节点个数 size, 邻近性阶数 order, 超参数 p 控制节点随机游走回溯概率, 超参数 q 控制 BFS 和 DFS, 数据集文件 datafile, 训练轮数 epoch, GCN 参数Θ;

输出　网络表示 output。

1) z_1 = DeepWalk(G,w,γ,t,d)

2) z_2 = Node2Vec(G,w,γ,t,d,p,q)

3) z_3 = Line(size,d,order)

4) $\boldsymbol{X}$ = Concatenate(z_1,z_2,z_3)

5) $\hat{\boldsymbol{A}}$ = load_date(datafile)

6) for each epoch do:

7) $\boldsymbol{X}_1$ = GraphConvolution$_1(\boldsymbol{X},\hat{\boldsymbol{A}},\Theta)$

8) $\boldsymbol{X}_2$ = ReLU$(\boldsymbol{X}_1)$

9) $\boldsymbol{X}_3$ = GraphConvolution$_2(\boldsymbol{X}_2,\hat{\boldsymbol{A}},\Theta)$

10) output = tanh$(\boldsymbol{X}_3)$

11) loss(output)

12) Backpropagation(loss)

13) 更新 output,Θ

14) end for

15) 返回 output

训练阶段进行模型计算和损失计算, 所以训练阶段的时间复杂度是$O(|E|HTF+dK|E|)$, 测试阶段的时间复杂度是$O(K|E'|)$, 其中 H 是特征输入维数 384, T 为中间层维数 256, F 为输出层维数 128, 数据边的数量是$|E|$, 测试数据边的数量是$|E'|$。综上所述, 总体时间复杂度是$O(|E|HTF+dK|E|)$。

3　实验和结果分析

在 6 个数据集上分别对比 DeepWalk、Node2Vec、Line 这 3 个经典的网络表示学习方法和 stacking 集成后的实验效果, 验证 GCN 作为 stacking 集成次级学习器的有效性。实验环境为: Windows10 操作系统, Intel i7-6700 2.6 GHz CPU, nvidia GeForce GTX 950M GPU, 8 GB 内存。编写 Python 和 Pytorch 实现。

3.1　实验设定

1) 数据集

实验使用 6 个真实数据集, 即 Cora、Citeseer、Pubmed、Wiki-Vote、P2P-Gnutella05 和 Email-Enron, 详细信息见表 2。Cora 是引文网络, 由机器学习论文组成, 每个节点代表一篇论文, 论文根据论文的主题分为 7 类, 边代表论文间的引用关系。Citeseer 也是引文网络, 是从 Citeseer 数字论文图书馆中选取的一部分论文, 该网络被分为 6 类, 边代表论文间的引用关系。Pubmed 数据集包括来自 Pubmed 数据库的关于糖尿病的科学出版物, 被分为 3 类。Wiki-Vote 是社交网络, 数据集包含从 Wikipedia 创建到 2008 年 1 月的所有 Wikipedia 投票数据。网络中的节点表示 Wikipedia 用户, 从节点 i 到节点 j 的定向边表示用户 i 给用户 j 的投票。P2P-Gnutella05 是因特网点对点网络, 数据集是从 2002 年 8 月开始的 Gnutella 点对点文件共享网络的一系列快照, 共收集了 9 个 Gnutella 网络快照。节点表示 Gnutella 网络拓扑中的主机, 边表示 Gnutella 主机之间的连接。Email-Enron 是安然公司管理人员的电子邮件通信网络, 覆盖了大约 50 万封电子邮件数据集中的所有电子邮件通信, 这些数据最初是由联邦能源管理委员会在调查期间公布在网上的, 网络的节点是电子邮件地址, 边表示电子邮件地址之间的通信。

表 2　数据集信息

Table 2　Dataset information

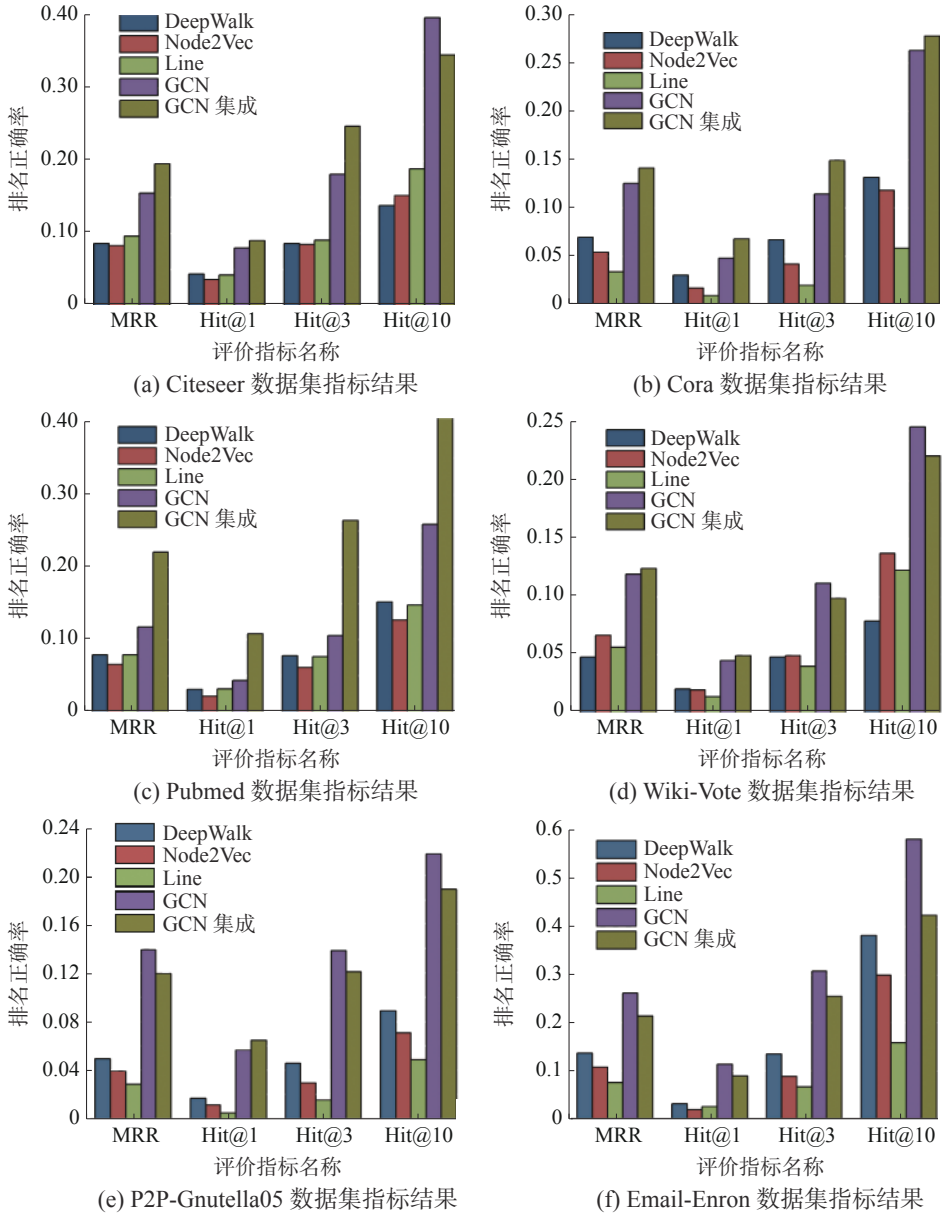
数据集	节点数	边数
Cora	2 708	5 278
Citeseer	3 312	4 660
Pubmed	19 717	44 338
Wiki-Vote	7 118	103 747
P2P-Gnutella05	8 846	31 839
Email-Enron	36 692	183 831

2) 参数设定

对于 stacking 集成方法中的 GCN 模型，使用 RMSProp 优化器更新训练参数，学习率设为 0.001，训练次数设为 200，卷积层为 2 层。对于 DeepWalk 和 Node2Vec 共同参数，节点游走次数设为 10，窗口大小设为 10，随机游走的长度设为 40。Node2Vec 的超参数 p=0.25、q=4。对于 Line，负采样数设为 10，学习率设为 0.025。为了方便比较，上述方法的节点表示维度均设为 128。

3.2　异质集成实验结果

实验选择 4 个领域的数据集，包括 Cora、Citeseer、Pubmed、Wiki-Vote、P2P-Gnutella05 和 Email-Enron。对于同一数据集，对比各初级学习器、GCN 和 stacking 异质 GCN 集成的特征表示的质量。GCN 的参数设定同 stacking 集成方法中的 GCN 模型参数。GCN 集成过程中仅使用网络结构，GCN 使用网络结构和数据集的属性特征，数据集没有的使用单位阵代替属性特征。图 6 展示了各数据集上的评价指标 MRR、Hit@1、Hit@3、Hit@10 的比较，各评价指标平均提升了 1.47~2.97 倍。



(a) Citeseer 数据集指标结果　(b) Cora 数据集指标结果

(c) Pubmed 数据集指标结果　(d) Wiki-Vote 数据集指标结果

(e) P2P-Gnutella05 数据集指标结果　(f) Email-Enron 数据集指标结果

图 6　各数据集异质集成评价指标结果

Fig. 6　Heterogeneous integration of evaluation index results of all datasets

实验结果显示, 在各数据集上 stacking 集成后的效果明显优于各初级学习器, 仅使用网络结构的 GCN 集成与使用网络结构和属性特征的 GCN 效果相当。这一方面归功于初级学习器的“好而不同”, 即初级学习器有一定的网络表示学习能力, 并且学习器之间具有差异性, 会有互补作用; 另一方面归功于 GCN 作为 stacking 集成次级学习器的有效性, GCN 根据对称归一化拉普拉斯矩阵的性质为邻居分配权重, 然后聚合邻居信息。

3.3　损失函数有效性验证

本文根据网络的一阶邻近性设计了损失函数, 通过设计使用损失函数和未使用损失函数的实验来验证损失函数的有效性。表 3 展示了各数据集评价指标的比较, 图中数据集名称的表示未使用损失函数, 数据集名称中的“-loss”表示使用了损失函数。实验结果表明, 使用损失函数的评价指标与未使用损失函数的相比平均提升了 0.44~1.79 倍, 验证了本文损失函数的有效性。

表 3　损失函数有效性验证指标结果

Table 3　Results of validation index of loss function

数据集	MRR	Hits@1	Hits@3	Hits@10
Cora	0.084	0.021	0.057	0.181
Cora-loss	**0.142**	**0.068**	**0.149**	**0.278**
Citeseer	0.140	0.045	0.127	0.248
Citeseer-loss	**0.195**	**0.09**	**0.247**	**0.346**
Pubmed	0.114	0.035	0.091	0.259
Pubmed-loss	**0.220**	**0.108**	**0.264**	**0.407**
Wiki	0.095	0.024	0.071	0.183
Wiki-loss	**0.124**	**0.049**	**0.098**	**0.221**
P2P	0.076	0.022	0.059	0.154
P2P-loss	**0.119**	**0.064**	**0.121**	**0.188**
Email	0.102	0.025	0.074	0.245
Email-loss	**0.213**	**0.089**	**0.255**	**0.424**

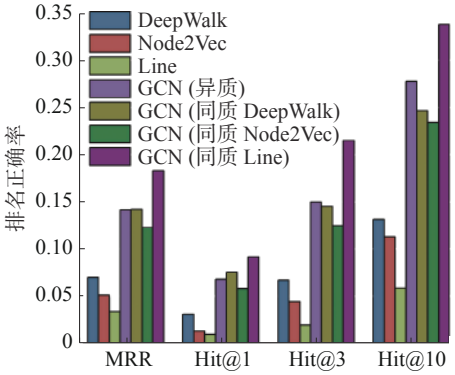
3.4　同质集成实验对比

本节对比算法分别进行同质 stacking, 对比设计如表 4 所示, 第 1~3 行是同质集成, 第 4 行是 3.2 节的实验设定。图 7 展示了 Cora、Citeseer 和 P2P-Gnutella05 数据集同质、异质集成及 3 个初级学习器对比的实验结果。

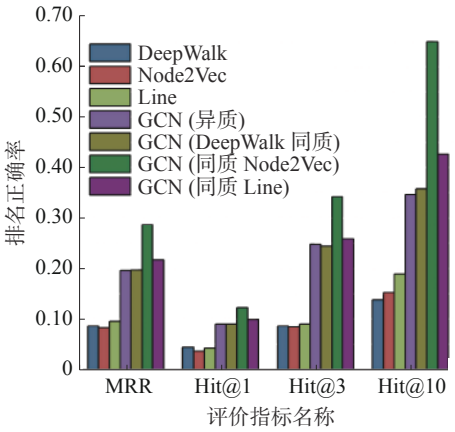
表 4　对比算法设计

Table 4　Design of contrast algorithms

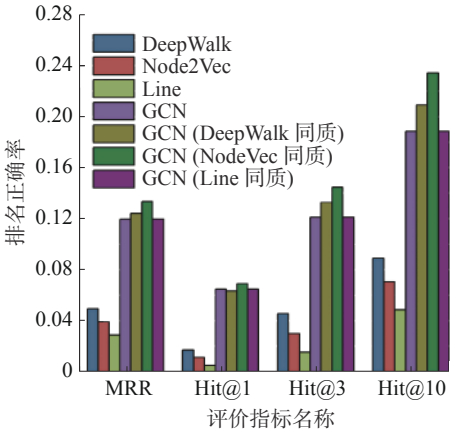
初级学习器1	初级学习器2	初级学习器3	次级学习器
DeepWalk	DeepWalk	DeepWalk	GCN
Node2Vec	Node2Vec	Node2Vec	GCN
Line	Line	Line	GCN
DeepWalk	Node2Vec	Line	GCN



(a) Cora 数据集指标结果



(b) Citeseer 数据集指标结果



(c) P2P-Gnutella05 数据集指标结果

图 7　各数据集同质/异质集成对比

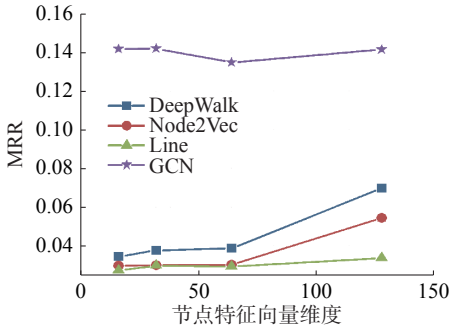
Fig. 7　Comparison of homogeneous / heterogeneous integration among datasets

实验结果表明, 在不同数据集上不同的同质集成各评价指标的表现不同。但是同质集成效果均明显优于初级学习器的效果, 平均提升了 1.46~1.91 倍, 所以异质集成的效果平均优于同质集成。在 Cora 数据集上, DeepWalk 和 Node2Vec 同质集成的效果略差于异质集成, Line 同质集成略好于异质集成; 在 Citeseer 数据集上, DeepWalk 同质集成效果与异质集成相当, Line 和 Node2Vec 同质集成略好于异质集成; 在 P2P-Gnutella05 数据集上, Line 同质集成效果与异质集成相当, Node-

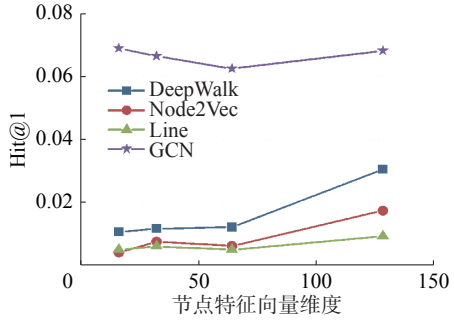
2Vec 和 DeepWalk 同质集成略好于异质集成。因为数据集网络结构具有多样性和复杂性，所以在不同数据集上表现效果不同，有的同质集成效果略优于异质集成。GCN 不仅可以作为集成器，本身也是学习器，有一定的学习能力。

3.5　参数敏感性分析

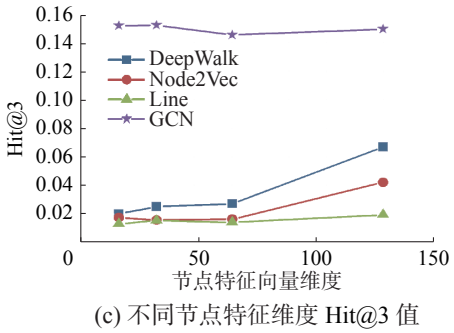
本节进行参数敏感性实验，主要分析不同特征维度对性能的影响。实验选用 Cora 数据集，图 8 分别展示了 MRR 和 Hit@1、Hit@3、Hit@10 的实验结果。



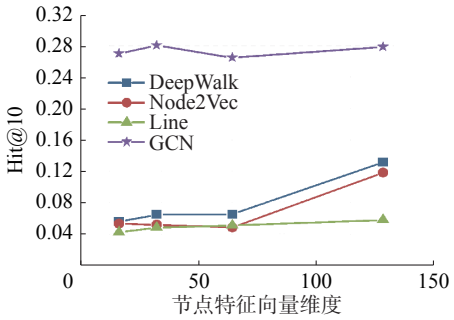
(a) 不同节点特征维度 MRR 值



(b) 不同节点特征维度 Hit@1 值



(c) 不同节点特征维度 Hit@3 值



(d) 不同节点特征维度 Hit@10 值

图 8　参数敏感性分析

Fig. 8　Parametric sensitivity analysis

实验结果表明，节点特征向量维度增加到 128 时，初级学习器的效果没有明显提升；但是 GCN 异质集成的效果却没有大幅受节点特征向量维度的影响，说明节点特征维度不是实验结果的重要影响因素。

4　结束语

在网络表示学习中，如何设计算法学习到高质量的节点表示仍是一个重要的研究课题。本文引入了 stacking 集成方法学习网络表示。首先并行训练多个简单的初级学习器，并将它们的嵌入拼接，选用 GCN 作为次级学习器，聚合得到最终的网络表示，然后对网络表示学习的无监督性，利用网络的一阶邻近性设计损失函数；最后改进了评价指标 MRR、Hit@1、Hit@3、Hit@10，分别测试初级学习器和集成后的节点特征向量表示，验证了提出算法具有较好的网络表示性能。

在 6 个数据集上进行实验，在各数据集上 stacking 集成后的效果明显优于各初级学习器，因为 GCN 作为 stacking 异质集成次级学习器的有效性及初级学习器的“好而不同”。对比算法选择 stacking 同质集成进行比较，实验结果表明同质集成的效果均明显优于初级学习器，且异质集成的效果平均优于同质集成，有的数据集同质集成效果由于异质集成是由于 GCN 不仅是集成器，更是学习器，有一定的学习能力。对于参数敏感性分析，实验结果表明节点向量维度不是实验结果的重要影响因素。

未来研究工作包括探索其他算法作为初级学习器、次级学习器对集成的影响和探索如何提高不同网络结构的适应性去处理归纳性任务。

参考文献：

[1] BHAGAT S, CORMODE G, MUTHUKRISHNAN S. Node Classification in Social Networks[M]// Social Network Data Analytics. Boston, MA: Springer, 2011: 115−148.

[2] LIBEN-NOWELL D, KLEINBERG J. The link-prediction problem for social networks[J]. Journal of the American society for information science and technology, 2007, 58(7): 1019–1031.

[3] COŞKUN M, KOYUTÜRK M. Node similarity based graph convolution for link prediction in biological networks[J]. Bioinformatics (Oxford, England), 2021, 37(23): 4501–4508.

[4] DER MAATEN L V, HINTON G. Visualizing data using t-SNE[J]. Journal of machine learning research, 2008: 2579–2605.

[5] TANG Jian, LIU Jingzhou, ZHANG Ming, et al. Visualizing Large-Scale and High-Dimensional Data[C]//Proceedings of the 25th International Conference ompanion on World Wide Web. Canada, New York, 2016: 287−297.

[6] ZHANG Daokun, YIN Jie, ZHU Xingquan, et al. Network representation learning: a survey[J]. IEEE transactions on big data, 2020, 6(1): 3–28.

[7] BENGIO Y, COURVILLE A, VINCENT P. Representation learning: a review and new perspectives[J]. IEEE transactions on pattern analysis and machine intelligence, 2013, 35(8): 1798–1828.

[8] 尹赢, 吉立新, 黄瑞阳, 等. 网络表示学习的研究与发展[J]. 网络与信息安全学报, 2019, 5(2): 77–87.
YIN Ying, JI Lixin, HUANG Ruiyang, et al. Research and development of network representation learning[J]. Chinese journal of network and information security, 2019, 5(2): 77–87.

[9] ROWEIS S T, SAUL L K. Nonlinear dimensionality reduction by locally linear embedding[J]. Science, 2000, 290(5500): 2323–2326.

[10] HE Xiaofei, Niyogi P. Locality preserving projections [J]. In advances in neural information processing systems, 2004, 16: 153–160.

[11] TU Cunchao, ZHANG Weicheng, LIU Zhiyuan, et al. Max-margin DeepWalk: discriminative learning of network representation[C]//Proceedings of the 25th International Joint Conference on Artifificial Intelligence. New York: ACM, 2016: 3889–3895.

[12] CAO Shaosheng, LU Wei, XU Qiongkai. Grarep: Learning graph representations with global structural informat- ion[C]// Proceedings of the 24th ACM International on Conference on Information and Knowledge Management. Melbourne, Australia, 2015: 891–900.

[13] 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 and Data Mining. New York : ACM, 2014: 701–710.

[14] GROVER A, LESKOVEC J. node2vec: scalable feature learning for networks[EB/OL]. (2016−07−03)[2021−07−23]https://arxiv.org/abs/1607.00653.

[15] TANG Jian, QU Meng, WANG Mingzhe, et al. LINE: large-scale information network embedding[C]//Proceedings of the 24th International Conference on World Wide Web. New York: ACM, 2015: 1067–1077.

[16] WANG Daixin, CUI Peng, ZHU Wenwu. Structural deep network embedding[C]//Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. New York, USA, 2016: 1225−1234.

[17] HAMILTON W, YING Z, LESKOVEC J. Inductive repres- entation learning on large graphs[C]//Advances in Neural Information Processing Systems. Long Beach , USA, 2017: 1024−1034.

[18] WANG Hongwei, WANG Jia , WANG Jialin, et al. Graph- GAN: graph representation learning with generative adve- rsarial nets[C]// Proceedings of the 32th AAAI Conference on Artificial Intelligence, New Orleans, USA, 2018: 2508−2515.

[19] ZHANG Boyu, IANG Ji, WANG Xin. Network representation learning with ensemble methods[J]. Neurocomputing, 2020, 380: 141–149.

[20] 徐冰冰, 岑科廷, 黄俊杰, 等. 图卷积神经网络综述 [J]. 计算机学报, 2020, 43(5): 755–780.
XU Bingbing, CEN Keting, HUANG Junjie, et al. A survey on graph convolutional neural network[J]. Chinese journal of computers, 2020, 43(5): 755–780.

[21] KIPF T N, WELLING M . Semi-supervised classifific-ati- on with graph convolutional networks[EB/OL]. (2016−09−09)[2021−07−23]https://arxiv.org/abs/1609.02907.

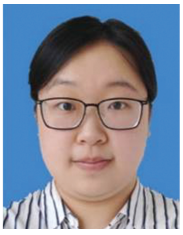
[22] 周志华. 机器学习 [M]. 北京: 清华大学出版社, 2016.

[23] MIKOLOV T, SUTSKEVER I, CHEN KAI, et al. Distributed representations of words and phrases and their compositionality[EB/OL]. (2013−19−16)[2021−07−23]https://arxiv.org/abs/1310.4546v1.

[24] SUN ZHIQING, DENG ZHI-HONG, NIE JIAN-YUN, et al. RotatE: knowledge graph embedding by relational rotation in complex space[EB/OL]. (2019−02−26)[2021−07−23]https://arxiv.org/abs/1902.10197v1.

作者简介:

常新功, 教授, 博士, CCF 高级会员, 主要研究方向为图神经网络、数据挖掘、进化算法。主持 10 项山西省重点课题。发表学术论文 30 余篇。



王金珏, 硕士研究生, 主要研究方向为图神经网络、数据挖掘。