



嵌入双曲层的神经排序式图表示学习方法

唐素勤^{a,b}, 刘笑梅^b, 袁磊^a

(广西师范大学 a. 教育学部; b. 广西多源信息挖掘与安全重点实验室, 广西 桂林 541004)

摘 要: 为解决已有图表示学习方法复杂性较高的问题, 提出一种能在维持图特征表达力的同时提升学习效率的方法。通过在神经网络表示模型中设置适当的双曲几何结构捕获图数据的基本属性, 利用贝叶斯个性化排序目标最大化节点之间正确链接和错误链接的差距从而自动学习相似性信息, 在所设计的神经排序模型中使用双曲距离函数计算节点之间的层次距离。在此基础上, 基于黎曼梯度下降法学习节点的特征向量。实验结果表明, 相对 DNGR、HARP 等方法, 该方法能够高效地学习节点特征, 而且能获得更加紧凑、更具表达力的特征向量表示。

关键词: 图表示学习; 双曲几何; 双曲面模型; 神经网络; 贝叶斯个性化排序

开放科学(资源服务)标志码(OSID):



中文引用格式: 唐素勤, 刘笑梅, 袁磊. 嵌入双曲层的神经排序式图表示学习方法[J]. 计算机工程, 2020, 46(6): 81-87.

英文引用格式: TANG Suqin, LIU Xiaomei, YUAN Lei. Graph representation learning method based on neural ranking with embedded hyperbolic layer[J]. Computer Engineering, 2020, 46(6): 81-87.

Graph Representation Learning Method Based on Neural Ranking with Embedded Hyperbolic Layer

TANG Suqin^{a,b}, LIU Xiaomei^b, YUAN Lei^a

(a. Faculty of Education; b. Guangxi Key Lab of Multi-Source Information Mining and Security, Guangxi Normal University, Guilin, Guangxi 541004, China)

[Abstract] To address the high complexity of existing graph representation learning methods, this paper proposes a new graph representation learning method to improve the learning efficiency while maintaining the representation performance of graph features. The method captures the basic properties of graph data by establishing appropriate hyperbolic geometry structure in the neural network representation model. Then the Bayesian Personalized Ranking (BPR) target is used to maximize the gap between the correct links and the wrong links to automatically learn the similarity information. Moreover, the hyperbolic distance function is used to calculate the hierarchical distance between the nodes in the designed neural ranking model. Finally, the model uses the Riemannian gradient descent method to learn the feature vector of nodes. Experimental results show that the proposed method can efficiently learn node features, and can provide more compact and more expressive feature vector representations than DNGR, HARP and other methods.

[Key words] graph representation learning; hyperbolic geometry; hyperboloid model; neural network; Bayesian Personalized Ranking (BPR)

DOI: 10.19678/j.issn.1000-3428.0053895

0 概述

网络图包含一组相互连接的节点, 其中, 每对节点之间具有大量的关系信息^[1]。几乎每个领域都需要对图进行分析, 例如, 在生物网络中, 人们可能需要分类蛋白质的角色, 或者预测现有药物分子的新

应用; 在社交网络中, 需要有针对性地为用户投放广告或推荐新朋友等。因此, 研究人员开发了各种图数据挖掘算法应用于节点分类、标签推荐、异常检测和链接预测等任务。但是, 这些已有的图数据挖掘算法通常需要一组图特征信息作为算法的输入。例如, 为了从图中提取结构信息, 传统的方法通常依靠

基金项目: 国家自然科学基金(61967002, 61663004, 61662007, 61866004); 国家哲学社会科学基金教育学一般项目(BCA170081); 广西自然科学基金(2016GXNSFAA380146, 2017GXNSFAA198365)。

作者简介: 唐素勤(1972—), 女, 教授、博士, 主研方向为本体理论、知识图谱; 刘笑梅, 硕士研究生; 袁磊, 教授、博士。

收稿日期: 2019-02-07 修回日期: 2019-05-13 E-mail: sqtang@gxnu.edu.cn

图的相关统计(如度、聚集系数等)、核函数或手工特征来测量局部近邻结构^[2],然而这些手工设计的特征在学习过程中不能自适应,且需要大量的人力成本和专业知识的^[3]。因此,有学者提出了表示学习方法以避免耗时且成本昂贵的特征设计,该方法同时提高了特征的灵活性。

现有的图表示学习方法主要聚焦于为图数据设计新的神经网络模型或设计更加复杂的随机游走机制来探索网络结构,这些方法往往使图表示学习的复杂性大幅提高。最近,将数据嵌入非欧几里德空间的方法受到越来越多的关注,原因是欧几里德模型不能正确地反映复杂数据模式。受此启发,本文对神经网络表示设置适当的几何结构,特别是层次结构和聚集行为,以捕获图数据的基本属性。图数据的这些基本属性出现在许多遵循幂律分布的现实复杂网络场景中,可作为引入双曲几何的起点。双曲空间能够反映复杂网络的属性主要是由于其潜在的关键属性,即空间量随着距参考点的距离呈指数增长,这与欧几里德空间中较慢的多项式增长相反。因此,本文探索双曲空间是否有助于学习图数据的嵌入,即通过使用神经网络对相互作用或关系进行建模,并利用数据所存在的流形度量结构来学习得到节点的低维紧凑特征向量表示。

通过对图表示学习方法和复杂网络中双曲几何理论进行分析,本文提出一种嵌入双曲层的神经排序式图表示学习方法 Neural-HRNE。该方法利用神经网络的无监督端到端方式以及双曲几何的分层自组织能力来自动抽取节点的相似性和层次结构信息。Neural-HRNE 方法没有使用过度复杂的节点交互机制,而是提出一种更小更快的神经排序架构来实现同样的性能。其中,利用贝叶斯个性化排序(Bayesian Personalized Ranking, BPR)^[4]目标来捕获节点表示向量之间的局部拓扑结构相似性。文献[5]研究表明,在 Lorentz 模型中学习嵌入比在 Poincaré 球中更加有效,其不仅更适用于上下层级关系的建模,而且能非常有效地执行黎曼优化并避免由 Poincaré 距离产生的数值不稳定性问题。因此,在 Neural-HRNE 方法中,本文将距离度量转换为所嵌入双曲面模型中的度量,图中的节点能利用双曲几何特性自组织为分层结构,从而使得 Neural-HRNE 方法可以高效地提取节点的潜在特征表示。在学习得到节点的潜在特征向量表示后,本文在多个不同尺度的数据集上分别进行节点推荐和节点分类任务,通过对比几种不同空间中的图表示学习方法来验证所提方法的有效性,并分析其维度敏感性和模型收敛性。

1 相关工作

早期的图表示学习可追溯到 2000 年,当时的算法主要将图表示学习作为降维技术的一部分,通过特征分解图的关联矩阵来得到节点的特征向量。但是,

就节点数量而言,多数算法通常至少具有二次时间复杂度,可扩展性较低限制了它们在大规模图上的应用^[6]。与特征工程中需要人工设计特征不同,深度学习会自动从数据中学习特征表示,这使得特征工程向特征学习转变。近期的图表示学习方法主要集中于为图数据设计新的神经网络模型。DeepWalk^[7]首先使用随机游走从输入图中采样一组路径,将采样的路径类比为来自语料库的句子,从而利用神经语言模型 Skip-Gram 学习得到节点的特征向量表示。DeepWalk 的成功激发了许多后续的研究,如 Node2vec^[8] 和 LINE^[9] 等。使用递归神经网络模型 GRU 来嵌入信息的级联路径以及将邻接矩阵输入到自编码器中重构邻域相似的节点也是较典型的方法。作为另一种流行的深度学习模型,卷积神经网络及其变体也被广泛应用于图表示学习。自 2008 年以来,研究人员聚焦于直接为复杂网络设计有效且可扩展的表示学习技术,且其表现出了良好的性能并能适用于各种应用^[6]。

对符号数据的表示学习已成为机器学习和人工智能领域的研究热点^[10],尽管这些方法在许多应用中已被证明可行,但它们中的大部分都是基于欧几里德嵌入来处理符号数据,而复杂网络分析中的相关研究表明,许多真实世界的网络具有潜在的树状层次结构,这是许多复杂数据集所固有的属性,如自然语言的幂律分布及社会和语义网络的无标度性。这种树状结构数据的几何形状随着距根的距离呈指数级扩展,数据可以在双曲空间中被精确地捕获,但不能在欧几里德空间中被捕获^[11]。因此,将数据嵌入双曲空间(一类具有恒定负曲率的非欧几里德空间)受到越来越多的关注。双曲空间可以看作树的连续版本,反之,树可以看作“离散的双曲空间”^[12],在 \mathbb{R}^2 中,类似的构造是不可能的,因为圆的周长($2\pi r$)和面积(πr^2)仅在欧几里德几何中分别关于 r 呈线性和二次方增长。受此启发,最近的一些研究聚焦于在各种领域开发有效的算法用于在双曲空间中嵌入数据,包括自然语言处理和网络科学等。例如,文献[12]开发了一个使用双曲空间建模复杂网络的几何框架,并通过假设网络具有双曲几何来呈现典型属性,如异构度分布和强聚类。此外,文献[10]将符号数据嵌入到一个 n 维 Poincaré 球中,用于学习分层表示,结果表明 Poincaré 嵌入优于欧几里德嵌入。文献[5]研究表明,在 Lorentz 模型中学习嵌入比在 Poincaré 球模型中学习更加有效,其将这种概念系统的层级关系建模为相关性和一般性的组合以学习偏序关系。文献[13]提出一种嵌入有向无环图的新方法,其使用测地凸锥在嵌入空间中引起偏序关系来扩展 Poincaré 嵌入从而建模不对称关系。文献[14]提出基于 Skip-Gram 的双曲空间中图的神经嵌入算法。文献[15]以无监督方式从文本语料库中学习双曲空间中的单词和句子的蕴涵嵌入。与欧几里德嵌入相比,使用具有少量维度的双

曲嵌入能在诸多方面实现良好的性能,例如,回答词语的语义查询或复杂网络中的链接预测^[11]。上述结果表明,更好地考虑数据的固有几何形状可以改善下游预测效果。

2 模型构建

本文的图表示学习方法 Neural-HRNE 利用 BPR 来捕获节点之间的局部拓扑结构相似性,并通过双曲几何中的双曲面模型来有效探索图的拓扑结构信息,特别是层级结构。模型的整体架构如图 1 所示,该模型使用的神经网络结构主要包括输入层、嵌入层、隐藏层、双曲层、输出层和 BPR 层。具体而言,模型具有共享参数的 2 个部分,即一个接受正确的节点链接,另一个接受错误的节点链接,并旨在最大化正确链接和错误链接之间的差距。其中,嵌入层根据节点的局部邻居结构学习节点的统一矢量表示,隐藏层应用非线性降维嵌入节点的向量,双曲层使用双曲面模型来探索图的层次结构信息,即通过嵌入空间中的双曲面距离来建模节点对之间的关系,输出层和 BPR 层通过反向传播开启模型推断。

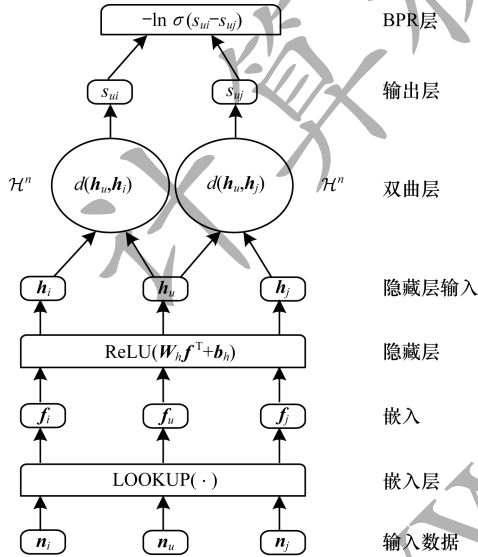


图 1 Neural-HRNE 模型整体架构

Fig. 1 Overall structure of the Neural-HRNE model

2.1 输入层和嵌入层

Neural-HRNE 的训练以节点三元组 (u, i, j) 作为输入,即给定节点 u 及其正确邻域 i 和随机采样的损坏邻域 j 作为样本。每个节点都表示为一个独热向量 \mathbf{n}_u (代表图 G 中的节点),然后通过嵌入层进行索引,因此,嵌入层是一个具有嵌入矩阵 $\mathbf{P} \in \mathbb{R}^{|\mathcal{V}| \times d}$ 的查找层, \mathbf{P} 的每一行嵌入节点的特征信息,其中, d 为嵌入维度。通过索引到节点,嵌入矩阵 \mathbf{P} 将每个节点转换为低维向量,如下:

$$\mathbf{f}_u = \text{LOOKUP}(\mathbf{P}, \mathbf{n}_u) \quad (1)$$

在具体实现过程中,本文使用预训练节点嵌入来初始化嵌入层,嵌入矩阵 \mathbf{P} 在学习过程中迭代更新。

2.2 隐藏层

给定图 G 中节点 u 的嵌入向量 $\mathbf{f}_u \in \mathbb{R}^{1 \times d}$,隐藏层的目的是将其嵌入向量转换为另一种表示 \mathbf{h}_u ,以实现非线性降维。给定 \mathbf{f}_u ,隐藏层通过式(2)产生 $\mathbf{h}_u \in \mathbb{R}^{1 \times h}$:

$$\mathbf{h}_u^T = \text{ReLU}(\mathbf{W}_h \mathbf{f}_u^T + \mathbf{b}_h) \quad (2)$$

激活函数使用线性整流函数 $\text{ReLU}(x)$,定义 $\max(0, x)$ 作为实现快速收敛的激活函数。参数 $\mathbf{W}_h \in \mathbb{R}^{h \times d}$ 和 $\mathbf{b}_h \in \mathbb{R}^{h \times 1}$ 分别是隐藏层的权值和偏置, h 表示隐藏层中神经元的数量。在隐藏层中,所有节点共享一组参数 $\{\mathbf{W}_h, \mathbf{b}_h\}$,从而实现跨不同节点的信息共享。

2.3 双曲层

双曲层采用双曲面模型来建模节点对之间的关系,这种 n 维双曲空间模型是 $n+1$ 维 Minkowski 空间中的一个流形。Minkowski 空间由公式 $\langle x, y \rangle_{\mathcal{H}} = -x_0 y_0 + \sum_{i=1}^n x_i y_i, x, y \in \mathbb{R}^{n+1}$ 所定义。 n 维双曲面模型 \mathcal{H}^n 位于 \mathbb{R}^{n+1} 内部,作为相对于 Minkowski 内积 $\mathcal{H}^n = \{x \in \mathbb{R}^{n+1} : \langle x, x \rangle_{\mathcal{H}} = -1, x_0 > 0\}$ 的单位“球体”的上半部分。因此,本文将 n 维双曲面模型定义为黎曼流形 $(\mathcal{H}^n, \mathbf{g}_{\ell})$,其中:

$$\mathbf{g}_{\ell}(x) = \begin{bmatrix} -1 & & & \\ & 1 & & \\ & & \ddots & \\ & & & 1 \end{bmatrix} \quad (3)$$

\mathcal{H}^n 中两点之间的距离定义为连接两点的双曲面上的测地路径的长度。 \mathcal{H}^n 中的每个测地曲线(即双曲线)都是 \mathcal{H}^n 与 2D 平面之间的交线,该平面穿过外围欧几里德空间 \mathbb{R}^{n+1} 中的原点。 \mathcal{H}^n 上的相关距离函数为:

$$d_{\ell}(x, y) = \text{arccosh}(-\langle x, y \rangle_{\mathcal{H}}) \quad (4)$$

因此,双曲层将节点之间的双曲距离定义为:

$$d(\mathbf{h}_u, \mathbf{h}_i) = \text{arccosh}(-\langle \mathbf{h}_u, \mathbf{h}_i \rangle_{\mathcal{H}}) \quad (5)$$

2.4 输出层和 BPR 层

Neural-HRNE 方法通过式(6)的线性变换传递双曲距离:

$$s(\mathbf{h}_u, \mathbf{h}_i) = \mathbf{W}_s d(\mathbf{h}_u, \mathbf{h}_i) + \mathbf{b}_s \quad (6)$$

其中, $\mathbf{W}_s \in \mathbb{R}^1$ 和 $\mathbf{b}_s \in \mathbb{R}^1$ 是该层的标量参数。节点 u 和 i 之间的相似度分数定义为 s_{ui} ,通过式(6)计算。

BPR 层实现贝叶斯个性化排序目标。对于嵌入任务,排序目标是图中的相邻节点在嵌入空间中应该具有比非相邻节点更高的相似度向量表示。例如,2 个相邻节点 u 和 i 之间的相似度得分应该大于

2 个不相邻节点 u 和 j 之间的相似度得分。如图 1 所示,给定节点三元组 (u, i, j) , Neural-HRNE 使用 sigmoid 函数 $\sigma(x) = \frac{1}{1 + e^{-x}}$ 为保留排序顺序 $s_{ui} > s_{uj}$ 的概率建模,在数学上表示为:

$$P(s_{ui} > s_{uj}) = \sigma(s_{ui} - s_{uj}) = \frac{1}{1 + e^{-(s_{ui} - s_{uj})}} \quad (7)$$

如式(7)所示, s_{ui} 和 s_{uj} 之间的差异越大,排序顺序 $s_{ui} > s_{uj}$ 被保留的可能性就越大。假定从图 \mathcal{G} 生成的所有基于三元组的排序顺序是独立的,所有排序顺序被保留的概率定义如下:

$$\prod_{(u, i, j) \in \mathcal{D}} P(s_{ui} > s_{uj}) = \prod_{(u, i, j) \in \mathcal{D}} \sigma(s_{ui} - s_{uj}) \quad (8)$$

其中, \mathcal{D} 表示从图 \mathcal{G} 生成的训练三元组集合。

Neural-HRNE 方法的嵌入目标是最大化式(8)。为了便于计算,最小化负对数似然损失函数的和,如下:

$$\mathcal{L}(\Theta) = - \sum_{(u, i, j) \in \mathcal{D}} \ln \sigma(s_{ui} - s_{uj}) + \lambda \cdot \|\Theta\|_F^2 \quad (9)$$

其中, $\Theta = \{P, W_h, b_h, W_s, b_s\}$ 是在所有不同层中使用的模型参数, $\lambda \cdot \|\Theta\|_F^2$ 是防止模型过度拟合的正则项。

2.5 优化和学习

本文利用在 Lorentz 嵌入^[5]中基于黎曼随机梯度下降所推导的梯度下降算法(算法 1)来估算参数 θ 。在黎曼优化中,通常需要解决形式为 $\min_{\theta \in M} f(\theta)$ 的问题,在本文中即最小化 BPR 损失 $\mathcal{L}(\Theta)$,通过计算 $\theta_{t+1} = \exp_{\theta_t}(-\eta \text{grad} f(\theta_t))$ 更新参数 θ ,其中, $\text{grad} f(\theta_t) \in T_{\theta_t} M$ 表示黎曼梯度, η 表示学习率。算法的具体推导过程可参见文献[5]。

算法 1 Riemannian Stochastic Gradient Descent

Input learning rate η , number of epochs T

for $t = 1, 2, \dots, T$

$h_t \leftarrow g_t^{-1} \nabla f(\theta_t)$

$\text{grad} f(\theta_t) \leftarrow \text{proj}_{\theta_t}(h_t)$

$\theta_{t+1} \leftarrow \exp_{\theta_t}(-\eta \text{grad} f(\theta_t))$

Neural-HRNE 方法采用反向传播算法,利用 mini-batch 梯度下降来优化模型中的参数 $\Theta = \{P, W_h, b_h, W_s, b_s\}$ 。mini-batch 梯度下降的主要过程是:首先从图 \mathcal{G} 中抽取一批三元组,具体来说,给定一个任意节点 u ,采样它的邻居 i 中的一个节点,即 $i \in \mathcal{N}(u)$,其概率与边权重 w_{ui} 成比例,同时采样其非相邻节点 j ,即 $j \notin \mathcal{N}(u)$,概率与图中的节点度成比例;然后对于每个 mini-batch 训练三元组,通过使用链式规则,根据算法 1 来更新相应参数 Θ ,另外,给

定训练所需的三元组集合 \mathcal{D} 和嵌入维度 d ,计算和更新 P 中相应嵌入向量的总成本是 $O(d)$ 。类似地,计算和更新隐藏层中参数 $\{W_h, b_h\}$ 的总成本是 $O(hd + h)$, h 为隐藏层神经元个数。因此,Neural-HRNE 方法迭代一次的总计算复杂度为 $|\mathcal{D}| \cdot (O(d) + O(hd + h))$ 。

3 实验结果与分析

将本文方法与现有学习方法进行比较,其中,选取一些常见的具有明确上下级层次结构的概念网络以及未明确编码对象层次结构关系的基准图数据集,以评估各方法的节点推荐和节点分类任务效果,最后分析各方法在不同表示学习空间中的性能,即在欧几里德、Poincaré 和双曲面模型中分析各方法对维度的敏感性和模型的收敛性。

3.1 实验设置

本文考虑具有明确上下级层次结构的数据和非树状的有向无环图结构,数据集的相关统计信息如表 1 所示。其中:

1) WordNet^[16] 是一个庞大的英文词汇数据库,在 WordNet 中,名词、动词、形容词和副词各自被组成一个同义词网络,每个同义词集合都代表一种基本的语义概念,本文嵌入 WordNet 的名词和动词层次结构。

2) ACM 计算分类系统 CCS^[17] 是由 ACM 计算机协会设计的用于分类计算机主题的系统,其可看作一种分层本体,各种 ACM 期刊使用该系统来按领域组织主题。

3) DBLP 是来自 DBLP 数据集^[18]的作者网络,其中包括共同作者、作者引用和文本相似性视图。本文抽取 DBLP 中的研究人员的共同作者图,其标签表明研究人员发表其研究成果的领域,本文选择其中的“数据库”“数据挖掘”“信息检索”和“机器学习”4 个不同的研究领域作为标签。

4) PPI 是蛋白质分子之间构建的蛋白质-蛋白质相互作用网络^[19]。本文使用 PPI 网络的智人诱导子图,只有人类基因被保留为节点。Hallmark 基因集中提供的基因组被视为节点类别并代表蛋白质生物状态。

5) Wikipedia^[20] 是一个维基百科中储存前 100 万字节中单词的共现网络。在该数据集中,每个节点都是一个单词,每条边是单词之间的共现关系,每个节点都有一个标签,表示单词的 POS 词性。

表 1 数据集的相关统计信息
Table 1 Relevant statistics of dataset

层次集				网络			
相关统计	WordNet Nouns	WordNet Verbs	CCS	相关统计	DBLP	PPI	Wikipedia
节点数	82 115	13 542	2 299	节点数	29 199	3 890	4 777
连边数	769 130	35 079	6 526	连边数	133 664	76 584	184 812
深度	19	12	5	标签数	4	50	40

在对比实验中,所有方法均只使用节点的拓扑结构特征,选取一些基于欧几里德的神经网络嵌入和基于双曲空间的图嵌入方法。本文将 Neural-HRNE 双曲层中的双曲面度量替换为欧几里德和 Poincaré 度量并以此来作为其中的对比基线。对比方法具体如下:

1) DNGR^[21]采用随机冲浪策略来捕获图结构信息,并进一步将这些结构信息转换为正点互信息矩阵,然后训练堆栈降噪自编码器以学习节点表示。

2) HARP^[22]递归地合并原始图中的节点和边以获得具有相似性的一系列较小的连续图结构体,合并的图均具有不同的粒度,从而提供了原始图的全局结构视图。HARP 可作为一种通用的元策略来改进图嵌入算法,本文选取文献[22]中表现较好的 HARP(N2V)方法,其结合了 Node2vec 算法用于加强节点嵌入。

3) 文献[14]提出了一种双曲空间中图的神经嵌入算法,其采用与 DeepWalk 类似的方法,不同的是该算法不再使用欧几里德度量,而使用 Poincaré 度量并在双曲空间中通过反向传播来学习节点的向量表示。

4) 文献[13]提出一种基于黎曼流形的“测地线凸锥”模型来学习层次嵌入,其有效解决了 Order 嵌入^[23]和 Poincaré 嵌入^[10]中嵌入空间维度灾难、不能编码不对称关系和 Poincaré 球边界坍塌问题。

5) 为了研究本文双曲面度量对节点嵌入的影响,首先将 Neural-HRNE 方法中的距离度量(式(4))改为直接计算欧几里德距离,即 $d = \|u - v\|^2$,并将其命名为 EuclideanEmb。然后将 Neural-HRNE 方法中的距离度量改为文献[10]中的 Poincaré 距离,并使用其中的黎曼梯度下降来学习节点嵌入,将其命名为 PoincaréEmb。其他设置则保持原样。

本文嵌入模型中有若干用户定义的超参数。其中,隐藏层中的神经元数量设置为 150,嵌入模型中的正则化系数 λ 设置为 0.000 05。在模型学习和优化过程中,设置学习率 η 为 0.5,批量大小定为 100。对于其他对比方法,使用网格搜索从集合 $\{0.01, 0.001, 0.000\ 1\}$ 中选择正则化系数,并从集合 $\{0.01, 0.05, 0.1, 0.5\}$ 中选择学习率。对于其余的超参数,本文使用各方法在原文中所建议的默认参数值。设置所有节点的表示维度为 50 维。

3.2 节点推荐

节点推荐可用于相似性搜索等领域,其任务是根据自身的上下文向用户推荐感兴趣的对象。在现实场景中,推荐的节点有各种类型,如用户兴趣、社交朋友和查询文档等。使用表示学习的低维矢量通常比原始表示密集得多,这减轻了较多数据的稀疏性问题,使得查询任务更加简单和准确。

根据多数图表示学习中的评估方法,本文对于给定的查询节点,计算目标节点与查询节点间的距离并对目标节点进行排序。在实验中,评估本文所提方法在嵌入明确或隐含层级结构数据上的有效性。在评估过程中,本文分别使用上述 5 个数据集来评估嵌入质量,并将数据视为无向传递闭包,这样的分层结构不能从观察到的边直接得出,而需要被推断出来。为了测量嵌入质量,本文计算每个观察到的边 (u, v) 在嵌入空间中的相应距离 $d = (u, v)$,并按升序排列 u 的所有未观察到的边的距离,即 $\{d = (u, v') : (u, v') \notin D\}$,得到原始的正确元组排名(越低越好),然后计算所有正确节点的前 50 平均精度均值 MAP@50,结果如表 2 所示,其中,最优结果加粗表示。

表 2 MAP@50 结果对比

Table 2 Comparison of MAP@50 results

方法	WordNet Nouns	WordNet Verbs	CCS	DBLP	PPI	Wikipedia
DNGR 方法	0.458	0.398	0.432	0.505	0.402	0.324
HARP 方法	0.427	0.403	0.452	0.581	0.413	0.435
文献[14]方法	0.662	0.572	0.795	0.607	0.574	0.561
文献[13]方法	0.705	0.641	0.846	0.703	0.751	0.780
EuclideanEmb 方法	0.384	0.384	0.401	0.479	0.326	0.315
PoincaréEmb 方法	0.726	0.635	0.803	0.789	0.760	0.756
Neural-HRNE 方法	0.728	0.657	0.820	0.754	0.761	0.759

从表 2 可以看出,因为双曲几何的分层自组织能力,其所选择的使用双曲空间的表示学习方法比欧几里德空间中的方法更加有效,特别是当学习的特征维度较低时,双曲几何能得到更加紧凑的表示,所以能够更好地在有限空间中表示复杂函数。相比于双曲几何强层次嵌入方法,本文方法也表现出了较好的表达力,显示出其性能优势。

3.3 节点分类

节点分类通过在标记的节点嵌入集上应用分类器来进行训练,即给定未标记节点的特征向量表示,训练的分类器可以预测其类标签。由于 WordNet 是一个词

典且 CCS 通常作为分类法使用,没有可利用标签,因此本文仅使用 DBLP、PPI 和 Wikipedia 数据集来评估嵌入表示的监督学习任务效果,即节点分类的有效性。

本文从数据集中随机抽取不同比例的标记节点,并将它们用作训练数据,其余的作为测试数据。对于欧几里德模型,本文使用 LibLinear 库中的 one-vs-rest SVM 分类器预测每个节点最可能的标签,而对于双曲模型,本文使用基于双曲距离的内核训练 SVM 分类模型。重复上述过程 10 次,表 3 和表 4 所示为 DBLP、PPI 和 Wikipedia 数据集中的各方法平均性能表现,最优结果加粗表示。

表 3 DBLP 数据集上的准确率对比
Table 3 Accuracy comparison on DBLP dataset

方法	被标记节点的比例/%								
	10	20	30	40	50	60	70	80	90
DNGR 方法	38.51	40.58	41.62	45.09	50.21	51.53	53.47	57.36	57.94
HARP 方法	35.83	39.68	42.56	47.02	49.27	50.02	54.38	56.74	57.04
文献[14]方法	46.26	49.50	50.54	52.61	53.78	55.09	56.46	58.23	60.71
文献[13]方法	48.20	51.45	53.76	54.05	56.52	58.47	59.39	62.28	62.59
EuclideanEmb 方法	35.65	34.72	37.65	40.56	42.07	43.11	45.62	47.52	47.30
PoincaréEmb 方法	45.34	48.23	52.96	54.64	55.72	57.15	60.04	62.38	62.77
Neural-HRNE 方法	48.56	50.97	52.85	53.53	57.07	58.78	59.45	62.60	63.50

表 4 PPI 和 Wikipedia 数据集上的节点多标签分类 Macro-F1 结果
Table 4 Macro-F1 results of node multi label classification on PPI and Wikipedia datasets

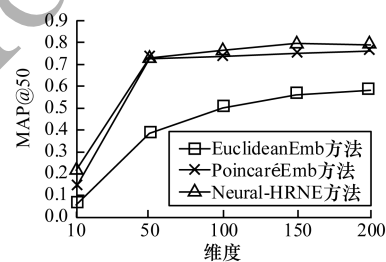
方法	PPI			Wikipedia		
	被标记节点的比例为 20%	被标记节点的比例为 50%	被标记节点的比例为 80%	被标记节点的比例为 20%	被标记节点的比例为 50%	被标记节点的比例为 80%
DNGR 方法	18.8	21.5	23.8	12.0	14.1	17.2
HARP 方法	19.6	22.7	25.0	13.9	17.2	21.5
文献[14]方法	20.2	23.3	28.3	17.5	22.4	24.8
文献[13]方法	21.4	25.2	30.2	17.2	25.4	30.8
EuclideanEmb 方法	14.7	18.0	21.5	8.7	12.0	16.5
PoincaréEmb 方法	20.3	25.7	32.1	17.3	24.7	30.5
Neural-HRNE 方法	21.7	26.1	32.5	17.2	25.0	30.0

从表 3、表 4 可以看出,相比于经典的基于欧几里德空间的方法,双曲几何对于数据的层次结构特征抽取更加有效,这有助于轻松高效地处理各种下游图数据分析任务。随着分类训练数据的增加,各分类器的性能均在不断提高,而对比于同等情况下使用欧几里德嵌入的 EuclideanEmb,本文提出的图表示学习方法在 Poincaré 和双曲面模型中性能均有所提高。在同时使用了双曲几何自组织能力的方法中,Neural-HRNE 的结果基本持平或略微降低,这可能是由于 Neural-HRNE 中的节点采样策略过于简单,未探索到节点之间诸如同质性或结构等价性的关系,或者所设计的神经网络较为浅层并直接使用了 One-hot 来初始化该神经网络,导致方法学习性能有所降低。因此,下一步考虑探索结合复杂网络和机器学习的更加高效的采样策略和更加强健的神经网络,以提高 Neural-HRNE 方法的性能。

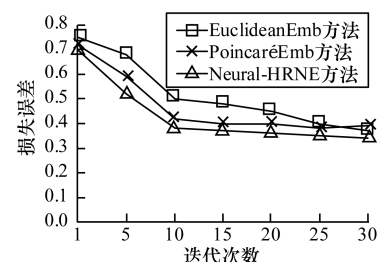
3.4 模型性能分析

为了进一步说明双曲面模型对图表示学习性能的影响,本文选取仅改变双曲层设置的表示学习方法进行比较。具体地,分别在 EuclideanEmb、PoincaréEmb 和 Neural-HRNE 方法中测试模型对表示学习维度的敏感性和模型的收敛性。图 2 所示为不同空间学习方法在 WordNet 名词层的维度敏感性和收敛性。图 2(a)为仅改变表示学习的维度后对 WordNet 名词层的节点推荐结果 MAP@50 的影响,从中可以看出,相比于欧几里德几何,双曲几何对空间的使用效率更高,欧几里德嵌入在维度较低时对特征的表达能力较弱,而双曲几何能在较低的维度时依然具有较好的表现性能,双曲几何大约在 50 维时就趋于稳定,其比欧

几里德更能提供紧凑的特征向量表示。图 2(b)为 EuclideanEmb、PoincaréEmb 和 Neural-HRNE 模型在 50 维时对 WordNet 名词层中节点特征表示学习的收敛情况。从中可以看出,本文提出的神经排序网络大约在 10 个时期内收敛,与欧几里德嵌入相比,双曲几何的嵌入方法收敛更快,损失误差更低。



(a) 维度敏感性分析



(b) 模型收敛性分析

图 2 3 种方法在 WordNet 名词中的维度敏感性和模型收敛性结果对比

Fig. 2 Comparison of dimensional sensitivity and model convergence results of three methods in WordNet terms

对不同空间中图表示学习方法的维度敏感性和模型收敛性进行分析,结果表明,相比于欧几里德嵌

入,双曲几何能提供更高质量的嵌入,特别是在学习维度较低的情况下。本文提出的双曲面模型表现出与 Poincaré 模型相当的性能,且其神经网络模型具有高效的学习能力。

4 结束语

如何对图中的节点进行有效的特征表示一直是图挖掘领域的研究热点。本文提出一种嵌入双曲层的神经排序式图表示学习方法,以提取节点的相似性和层次结构特性。该方法利用贝叶斯个性化排序作为其目标函数,并在其中加入一层双曲层来度量节点对之间的局部拓扑结构相似性,利用黎曼梯度下降来学习更高效的节点特征向量表示。对比使用欧几里德、Poincaré 和双曲面模型的不同表示学习方法的性能,结果表明,本文所提方法能够更高效地学习节点特征,而且可以获得更加紧凑、更具表达力的特征向量表示。嵌入双曲空间中的层次结构能很好地获取数据的基础语义,下一步将探索双曲空间的优化方法以提高嵌入质量并获得更快的收敛速度。此外,将双曲嵌入有效地整合到下游的任务和应用中,以及在多关系数据或图像等更复杂的数据嵌入中应用双曲几何理论,也是今后的研究重点。

参考文献

- [1] YANG Jinji, HU Bo, WANG Xinming, et al. Personalized recommendation algorithm for learning to rank by knowledge graph [J]. Journal of Chinese Computer Systems, 2018, 39(11): 2419-2423. (in Chinese)
杨晋吉, 胡波, 王欣明, 等. 一种知识图谱的排序学习个性化推荐算法[J]. 小型微型计算机系统, 2018, 39(11): 2419-2423.
- [2] HAMILTON W L, YING R, LESKOVEC J. Representation learning on graphs: methods and applications [EB/OL]. [2019-01-10]. <https://arxiv.org/abs/1709.05584>.
- [3] CHEN Weizheng, ZHANG Yan, LI Xiaoming. Network representation learning [J]. Big Data Research, 2015, 1(3): 8-22. (in Chinese)
陈维政, 张岩, 李晓明. 网络表示学习 [J]. 大数据, 2015, 1(3): 8-22.
- [4] RENDLE S, FREUDENTHALER C, GANTNER Z, et al. BPR: Bayesian personalized ranking from implicit feedback [C]// Proceedings of the 25th Conference on Uncertainty in Artificial Intelligence. Montreal, Canada: AUAI Press, 2009: 452-461.
- [5] NICKEL M, KIELA D. Learning continuous hierarchies in the lorentz model of hyperbolic geometry [C]// Proceedings of the 35th International Conference on Machine Learning. Stockholm, Sweden: JMLR Press, 2018: 3776-3785.
- [6] QI Jinshan, LIANG Xun, LI Zhiyu, et al. Representation learning of large-scale complex information network: concepts, methods and challenges [J]. Chinese Journal of Computers, 2018, 41(10): 2394-2420. (in Chinese)
齐金山, 梁循, 李志宇, 等. 大规模复杂信息网络表示学习: 概念、方法与挑战 [J]. 计算机学报, 2018, 41(10): 2394-2420.
- [7] 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, USA: ACM Press, 2014: 701-710.
- [8] GROVER A, LESKOVEC J. Node2vec: scalable feature learning for networks [C]// Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. San Francisco, USA: ACM Press, 2016: 855-864.
- [9] 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. Florence, Italy: ACM Press, 2015: 1067-1077.
- [10] NICKEL M, KIELA D. Poincaré embeddings for learning hierarchical representations [EB/OL]. [2019-01-10]. <https://arxiv.org/pdf/1705.08039.pdf>.
- [11] CHO H, DEMEO B, PENG J, et al. Large-margin classification in hyperbolic space [EB/OL]. [2019-01-10]. <https://arxiv.org/abs/1806.00437>.
- [12] KRIOUKOV D, PAPADOPOULOS F, KITSACK M, et al. Hyperbolic geometry of complex networks [J]. Physical Review E, 2010, 82(3): 036106.
- [13] GANEVA O E, BÉCIGNEUL G, HOFMANN T. Hyperbolic entailment cones for learning hierarchical embeddings [C]// Proceedings of the 35th International Conference on Machine Learning. Stockholm, Sweden: JMLR Press, 2018: 1632-1641.
- [14] CHAMBERLAIN B P, CLOUGH J, DEISENROTH M P. Neural embeddings of graphs in hyperbolic space [EB/OL]. [2019-01-10]. <https://arxiv.org/abs/1705.10359>.
- [15] DHINGRA B, SHALLUE C J, NOROUZI M, et al. Embedding text in hyperbolic spaces [C]// Proceedings of the 12th Workshop on Graph-Based Methods for Natural Language Processing. New Orleans, USA: ACL Press, 2018: 59-69.
- [16] MILLER G. WordNet: an electronic lexical database [M]. Cambridge, USA: MIT Press, 1998.
- [17] MIRKIN B G, NASCIMENTO S, PEREIRA L M. Representing a computer science research organization on the ACM computing classification system [C]// Proceedings of the 16th International Conference on Conceptual Structures. Toulouse, France: CEUR-WS Press, 2008: 57-65.
- [18] TANG Jie, ZHANG Jing, YAO Limin, et al. Arnetminer: extraction and mining of academic social networks [C]// Proceedings of the 14th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. Las Vegas, USA: ACM Press, 2008: 990-998.
- [19] STARK C, BREITKREUTZ B J, REGULY T, et al. BioGRID: a general repository for interaction datasets [J]. Nucleic Acids Research, 2006, 34(S): 535-539.
- [20] MAHONEY M. Large text compression benchmark [EB/OL]. [2019-01-10]. <http://mattmahoney.net/dc/text.html>.
- [21] CAO Shaosheng, LU Wei, XU Qiangkai. Deep neural networks for learning graph representations [C]// Proceedings of the 30th AAAI Conference on Artificial Intelligence. Phoenix, USA: AAAI Press, 2016: 1145-1152.
- [22] CHEN H, PEROZZI B, HU Y, et al. HARP: hierarchical representation learning for networks [C]// Proceedings of the 32nd AAAI Conference on Artificial Intelligence. New Orleans, USA: AAAI Press, 2018: 2127-2134.
- [23] VENDROV I, KIROS R, FIDLER S, et al. Order-embeddings of images and language [EB/OL]. [2019-01-10]. <https://arxiv.org/abs/1511.06361>.