计算机科学 COMPUTER SCIENCE
http://www.jsjkx.com
DOI:10.11896/jsjkx.190900027

融合节点结构和内容的网络表示学习方法

张 虎 周晶晶 高海慧 王 鑫
山西大学计算机与信息技术学院 太原 030006

摘 要 随着神经网络技术的快速发展,面向复杂网络数据的网络表示学习方法受到越来越多的关注,其旨在学习网络中节点的低维度潜在表示,并将学习到的特征表示有效应用于基于图的各种分析任务。典型的浅层随机游走网络表示学习模型主要基于节点结构相似和节点内容相似,不能同时有效捕获节点结构和内容的相似信息,因此在结构和内容等价混合的网络数据上表现较差。为此,探索了节点结构相似和节点内容相似的融合特征,提出了一种基于无监督浅层神经网络联合学习的表示方法SN2vec。实验分别利用节点结构和内容等价混合的Brazilian air-traffic,American air-traffic,Wikipedia数据集在多标签分类和降维可视化任务上进行验证。结果显示,SN2vec在多标签分类任务中的*Micro-F*1值优于现有的浅层随机游走网络表示方法,并且可以较好地学习到潜在结构表示一致的节点。

关键词:网络表示学习;随机游走;复杂网络;浅层神经网络
中图法分类号 TP391

Network Representation Learning Method on Fusing Node Structure and Content

ZHANG Hu,ZHOU Jing-jing,GAO Hai-hui and WANG Xin
School of Computer and Information Technology,Shanxi University,Taiyuan 030006,China

Abstract With the rapid development of neural network technology,the network representation learning method for complex network has got more and more attention. It aims to learn the low-dimensional potential representation of nodes in the network and to apply the learned characteristic representation effectively to various analysis tasks for graph data. The typical shallow random walk network representation model is mainly based on two kinds of characteristic representation methods,which are the node structure similarity and node content similarity. However,the methods can't effectively capture similar information of node structure and content at the same time,and perform poorly on the network data with the equivalent structure and content. To this end,this paper explores the fusion characteristics of node structure and node content,and proposes a representation method called SN2vec,which is based on joint learning of unsupervised shallow neural networks. Further,in order to validate the effectiveness of the proposed model,this paper respectively conduct the multi-label classification and down-dimensional visualization tasks in Brazilian air-traffic,American air-traffic,and Wikipedia datasets. The results show that the *Micro-F*1 of using SN2vec in multi-label classification task is better than the existing shallow random walk network representation methods,and SN2vec can also learn better potential structural representation of consistent nodes.

Keywords Network representation learning,Random walk,Complex network,Shallow neural network

1 引言

网络表示学习旨在将高维稀疏的网络数据映射到低维稠密的实值向量空间中,以自适应地提取特征,便于机器学习等方法进行下游网络任务的分析,如节点分类、链路预测和降维可视化等[1-2]。在机器学习研究领域中,网络表示学习常被认为是表示学习技术集合的一个子集。

在表示学习中,有监督的机器学习方法需要一组区分节点的特征。典型的方法主要面向特定的专业领域采用手工提取特征的方式来学习节点表示,存在特征提取不准确且鲁棒性差的缺点[3]。另一种方法主要通过解决优化问题来学习节点表示,该方法提高了特征提取的准确性,但估计参数量较大,时间复杂度偏高[4]。

在无监督的学习[5]方法中,特征的表示需要定义一个独立于下游任务的目标函数来平衡准确性与计算效率。相比手工提取特征的方式,利用目标函数优化的表示方法可以获取更全面的特征,其与下游预测网络精度的任务密切相关[6]。

无监督的学习方式改进了有监督机器学习方法不易扩展及训练复杂性高的不足。在基于无监督浅层随机游走的传统方法中,DeepWalk 和 Struc2vec 是根据图的深度优先搜索

到稿日期:2019-09-09 返修日期:2020-04-03 本文已加入开放科学计划(OSID),请扫描上方二维码获取补充信息。
基金项目:国家社会科学基金(18BYY074);国家自然科学基金(61936012,61806117);山西省高等学校科技创新项目(201802012)
This work was supported by the National Social Science Fund of China(18BYY074),National Natural Science Foundation of China (61936012,61806117) and Scientific and Technological Innovation Programs of Higher Education Institutions in Shanxi (201802012).
通信作者:张虎(zhanghu@sxu.edu.cn)

(Deepth First Search, DFS)定义的结构相似性表示学习方法;LINE 和 Node2vec[7-8]是根据图的宽度优先搜索(Breadth First Search, BFS)定义的内容相似性表示学习方法。

Brazilian air-traffic, American air-traffic 和 Wikipedia 为节点结构相似和节点内容相似等价混合的数据集,根据 DFS 和 BFS 定义的传统方法通常不能兼顾网络的两类相似性。如图 1 所示,从节点 u 和节点 v 出发的 BFS 策略无法捕获在宽度搜索路径上的节点 h,DFS 策略无法捕获在深度搜索路径上的节点 a,c,p。

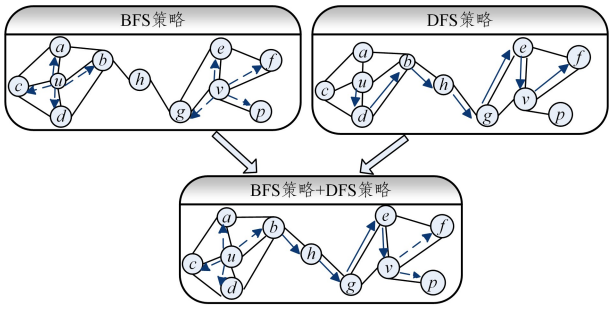

图 1 不同搜索策略

Fig. 1 Different search strategies

针对现有方法的不足,本文面向同质网络探索融合节点结构和节点内容相似的联合学习网络表示方法,根据 Struc2vec 和 Node2vec 在浅层神经网络上捕获节点结构和节点内容的特性,提出融合节点结构和节点内容的网络表示学习方法 SN2vec。SN2vec 可以更好地表示结构和内容等价混合的同质网络,如图 1 所示,融合 DFS 和 BFS 的搜索策略可以捕获两种方法各自无法捕获的节点 a,c,h 和 p。

2 相关工作

传统的网络表示方法是在降维的背景下描述的。在线性代数视角下,无监督特征学习方法通常利用图的各种矩阵进行表示,尤其是拉普拉斯算子和邻接矩阵[9]。降维的经典技术包括主成分分析(Principal Component Analysis, PCA)[10]和多维尺度(Multi-Dimensional Scale, MDS)[11]。这两种方法都将图表示为 $n*m$ 矩阵,并转换为 $n*k$ 矩阵,其中 $k \ll m$。两种方法都能够捕获线性结构信息,但无法发现输入数据中的非线性结构。已经提出的几种线性(例如 PCA)和非线性(例如 IsoMap)维数降低技术[12]存在计算和统计性能缺陷。就计算效率而言,数据矩阵的特征分解是昂贵的。

网络表示方法有不同的分类体系[13]。本文根据表示方法的结构将网络表示方法分为深层网络表示学习和浅层网络表示学习。在深层网络表示学习上,主流的方法利用深度学习模型来捕捉节点间的非线性关系。SDNE(Structural Deep Network Embedding)[14]通过无监督学习方法来自动捕捉节点的局部关系,将节点的二阶近邻作为输入来学习二阶近邻的低维表示。TADW(Network Representation Learning with Rich Text Information)[15]不仅考虑网络的结构信息,还利用节点产生的文本信息,在矩阵分解的基础上用节点的文本表示矩阵来进行约束。深层网络模型虽然能够捕获非线性结构,但忽略了网络结构中节点的潜在表示。大多数深层网络模型采用深度学习方法,而深度神经网络存在因训练数据较少而出现过拟合,计算量较大且需要高内存的缺点[16],不适用于较少规模的网络数据集。

在浅层网络表示学习上,主流方法很大程度上受 skip-gram[17]在自然语言和网络挖掘中的应用的启发。一系列基于 skip-gram 的模型用于将网络结构编码成连续的空间向量表示,如 DeepWalk, LINE 和 Node2vec 等。DeepWalk[18]通过从每个图节点开始随机游走生成节点序列来模拟句子,这一系列节点序列组成了"语料库"。DeepWalk 设定了背景窗口的大小,然后将随机游走得到的"语料库"输入 skip-gram 模型,得到每个图节点的图嵌入表示。LINE[19]首先分别优化直接相连的一阶相似性和共享邻居节点的二阶相似性,然后将这两种相似性进行合并作为节点的输出。LINE 还进一步定义了节点内容相似性,但是忽略了网络结构中节点的高阶邻近性。Node2vec[20]在 DeepWalk 的"语料库"的生成步骤上,额外使用了两个参数来控制随机游走的去向。Node2vec 虽然对探索不同种类的网络结构具有灵活性,但没有顾及高阶非线性结构中具有相同结构角色的节点的潜在表示。Struc2vec[21]通过对全局存在的潜在网络结构进行表示,发现了在全局结构上不相邻却有相同结构角色的节点,但忽略了节点的内容相似性。NE-FLGC(Network Embedding Based on Fusing Local and Global Network Structure with node Content)[22]是一个联合表示属性网络的模型,它结合了节点、一阶邻域信息和二阶邻域信息生成的文本信息,然后将节点和上下文的向量拼接在一起。GraRep 和 HOPE 是基于矩阵分解提出的网络表示学习模型[23-24]。

本文研究的 SN2vec 网络表示方法可在无向、无权的同质网络上弥补深层神经网络模型不适用于较小规模网络的不足,同时又能解决主流的浅层表示方法不能同时捕获节点结构和节点内容相似的问题。在节点结构和节点内容等价混合的 Brazilian air-traffic, American air-traffic 和 Wikipedia 等数据集上,SN2vec 不仅可以捕获同质网络上的一阶邻近和二阶邻近,还可以在全局网络上捕捉同一结构的潜在表示。

3 融合节点结构和内容的网络表示学习模型

3.1 Struc2vec 表示方法

Struc2vec[21]是一种基于对节点结构相似表示的无监督学习方法,主要通过捕获不相邻、不"共享"邻居,但却有相同结构角色的节点来刻画结构的相似性。如图 2 所示,节点 u 和节点 v 的度分别为 4 和 3,与 u 连接的邻居结构是 1 个三角形,与 v 连接的邻居结构是 2 个三角形,并且 u 和 v 的邻居节点与其他网络连接的个数都为 2,则该方法认为节点 u,v 在全局的结构上高度相似。

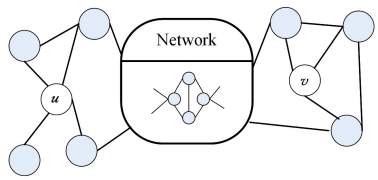

图 2 全局的结构潜在相似图

Fig. 2 Global structural potential similarity graph

3.2 Node2vec 表示方法

Node2vec[20]通过带权重的随机游走构建一系列线性序

列，并将这些线性序列输入 skip-gram 模型训练向量，可以对局部节点的内容相似进行较好地刻画。为了解决网络节点采样的灵活性问题，Node2vec 尝试利用控制参数来调整搜索空间。

如图 3 所示，边缘标签 α 表示搜索偏差，参数 p 和 q 控制随机游走序列跳转的概率和步行探测起始节点 u 的邻域的速度。参数 p 控制重复访问刚刚访问过的顶点的概率，其中，$p>\max(q,1)$ 表示在连续的两步中不会遍历已经遍历过的节点；$p<\min(q,1)$ 表示遍历节点接近起始节点 u。$q>1$ 偏向遍历邻近 t 节点的 X_1 节点，倾向于 BFS 搜索策略；$q<1$ 偏向遍历远离 t 节点的 X_2 节点，倾向于 DFS 搜索策略。

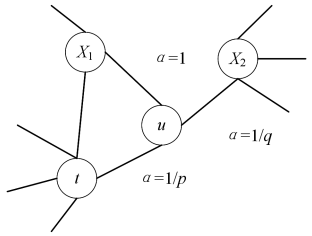

图 3　Node2vec 中的随机游走过程

Fig. 3　Random walk of Node2vec

3.3　SN2vec 联合表示方法

受 Struc2vec 模型和 Node2vec 模型的启发，本文提出基于节点内容相似和结构相似的联合表示网络模型 SN2vec，具体如图 4 所示。以相邻结构为例，SN2vec 利用 Struc2vec 模型表示得到全局的结构潜在表示，通过 Node2vec 模型表示得到局部的内容相似表示，最终拼接这两种表示得到节点 v 的联合表示。

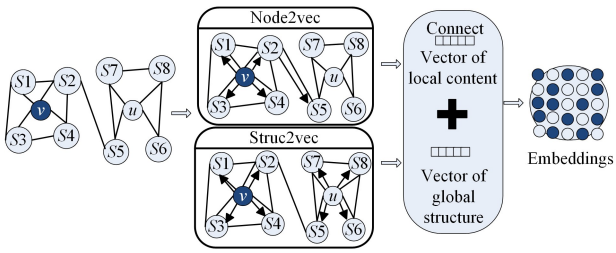

图 4　SN2vec 模型

Fig. 4　SN2vec model

3.3.1　SN2vec 的结构相似表示

假设已知图 $G=(V,E)$，在结构表示中，首先定义顶点 u 和 v 的距离如下：

$$f_k(u,v)=f_{k-1}(u,v)+\frac{\max(s(R_k(u)),s(R_k(v))}{\min(s(R_k(u)),s(R_k(v))} \tag{1}$$

其中，$R_k(u)$ 为到顶点 u 的距离为 k 的顶点集合；$s(\cdot)$ 为顶点集合的有序度序列；$k=0,\cdots,k-1$，且当 $k=0$ 时，令 $f_{-1}=0$。

其次，根据顶点对的顶点距离构建层次带权图，其中 w 为带权图的权函数，具体定义如下。

1）当顶点对 u,v 位于不同层时，将位于 k 层的当前节点 v 与其对应的上层节点 v_{k-1} 和下层节点 v_{k+1} 相连以构建层次图，此时权重定义如下：

$$w(v_k,v_{k-1})=1 \tag{2}$$

$$w(v_k,v_{k+1})=\log(\sum_{v\in v}1+\mathrm{e}) \tag{3}$$

当顶点对 u,v 位于同一层时，权重定义如下：

$$w_k(u,v)=\mathrm{e}^{-f_k(u,v)} \tag{4}$$

其中，k 表示节点所在层。

当 u,v 同在 k 层时，以 $p_k(u,v)=\frac{w_k(u,v)}{\sum_{v\in V,v\neq u}w_k(u,v)}$ 的概率采样；当 v 在 k 层，u 在 $k+1$ 层时，以 $p_k(v_k,v_{k+1})=1-\frac{1}{1+w(v_k,v_{k+1})}$ 的概率采样；当 v 在 k 层，u 在 $k-1$ 层时，以 $p_k(v_k,v_{k-1})=1-p_k(v_k,v_{k+1})$ 的概率采样。最后，得到的顶点序列如式(5)所示：

$$L_1(R_{k+1}(v),R_k(v),R_{k-1}(v)) \tag{5}$$

3.3.2　SN2vec 的内容相似表示

在内容相似表示中，SN2vec 通过对局部节点的特性分析来调整参数 p 和 q，以更好地表示节点内容相似性。对于源节点 v 的内容相似性表示，SN2vec 首先使用式(6)的概率分布产生固定长度为 L_2 的节点 v 的邻居序列，其次通过对节点 v 的邻居序列进行向量化学习，获得节点的内容相似性。

$$p(c_i=x|c_{i-1}=v)=\begin{cases}\frac{\pi_{vx}}{z}, & \text{if } (v,x)\in E\\ 0, & \text{otherwise}\end{cases} \tag{6}$$

其中，c_i 表示随机游走序列中的第 i 个节点，π_{vx} 表示节点 v 到节点 x 的转移概率，Z 为归一化因子。随机游走中的转移概率定义为 $\pi_{vx}=\alpha_{pq}(t,x)$，具体表示如式(7)所示：

$$\alpha_{pq}(t,x)=\begin{cases}\frac{1}{p}, & d_{tx}=0\\ 1, & d_{tx}=1\\ \frac{1}{q}, & d_{tx}=2\end{cases} \tag{7}$$

其中，t 代表前一节点，x 为下一个可能的节点；p,q 为第 3.2 节 Node2vec 表示方法中控制随机游走序列跳转的概率和步行探测起始节点邻域速度的参数；d_{tx} 为节点 t 和节点 x 之间存在的最短路径距离，且 d_{tx} 必须为 $\{0,1,2\}$ 其中之一。

当 $d_{tx}=0$ 时，若采样的节点 x 与前一个采样的节点为同一节点，则式(6)中 $\pi_{vx}=\frac{1}{p}$；当 $d_{tx}=1$ 时，若采样的节点 x 与前一个采样的节点直接相连，则 $\pi_{vx}=1$；当 $d_{tx}=2$ 时，若采样的节点 x 与前一个采样的节点不直接相连，则 $\pi_{vx}=\frac{1}{q}$。

3.3.3　SN2vec 的联合表示

由式(1)－式(5)得到结构相似的网络表示函数 $f(v)$，具体表示如式(8)所示：

$$f(v)=skip\text{-}gram(L_1(R_{k+1}(v),R_k(v),R_{k-1}(v))) \tag{8}$$

由式(6)和式(7)得到节点内容相似函数 $g(v)$，具体定义如式(9)所示：

$$g(v)=skip\text{-}gram(L_2(p(c_i=x|c_{i-1}=v))) \tag{9}$$

进一步融合结构相似和内容相似，得到联合表示函数 $F(v)$ 和节点向量空间表达 $\boldsymbol{X}_v$，分别如式(10)和式(11)所示：

$$F(v)=f(v)\oplus g(v) \tag{10}$$

其中，节点 $v\in V$，$\oplus$ 表示向量拼接运算。

$$\boldsymbol{X}_v=\begin{pmatrix}\boldsymbol{X}_{sv}\\ \boldsymbol{X}_{nv}\end{pmatrix} \tag{11}$$

其中，$\boldsymbol{X}_{sv}\in R^{n*d}$ 为节点 v 的结构相似性向量表示空间；$\boldsymbol{X}_{nv}\in R^{n*d}$ 为节点 v 的内容相似性向量表示空间；$\boldsymbol{X}_v\in R^{n*2d}$ 为节点 v 联合表示后的串行融合特征；d 为表示网络后的向量空间的维数，n 为网络顶点 V 的个数。

4 实验结果与分析

4.1 数据集

为了验证模型的有效性，本文选择了 3 个无向无权网络图数据集作为实验数据，不同数据集的规模如表 1 所列。

表 1 实验数据集

Table 1 Experimental datasets

	Wikipedia	Brazilian air-traffic	American air-traffic
Node	2405	103	1190
Edge	17981	1038	13599

Wikipedia[25]数据集来源于网络表示学习开源框架 OpenNE，共包含 2405 个不同类别的网页和 17981 条网页之间的链接关系，其中网页为网络图的节点，网页间的链接为边。

Brazilian air-traffic[21]为 2016 年的巴西空中交通网络数据集，包含 103 个节点和 1038 条边，节点表示机场，边表示机场与机场间的往来关系。每个机场指定为与机场活跃度相对应的 4 个可能标签中的一个，所有机场标签均分为 4 组。

American air-traffic 数据集[21]为 2016 年的美国空中交通网络，包含 1190 个节点和 13599 条边，具体标签划分与 Brazilian air-traffic 相同。

4.2 基线方法

DeepWalk：通过模拟均匀随机游走来学习 d 维特征表示。DeepWalk 中的采样策略可以看作带有 $p=1$ 和 $q=1$ 的 Node2vec 的特例。

LINE：在两个独立的阶段中学习 d 维特征表示。在第一阶段，通过 BFS 样式模拟在节点的一阶邻居上学习 $d/2$ 维度；在第二阶段，通过严格地从源节点的二阶邻居上采样节点，来学习下一个 $d/2$ 维度。

Struc2vec：通过多层次图的随机游走来学习具有结构等价特性的网络结构，生成 d 维特征结构。

Node2vec：通过控制参数 p,q 来实现对网络使用不同的搜索方法，进而通过对网络进行有偏置的随机游走来学习 d 维的特征表示。Node2vec 是 Deepwalk 方法的扩展。

4.3 节点多标签分类实验及结果分析

为了与 OpenNE 的实验结果进行对比，本文实验在 Wikipedia 数据集上采取了同样比例的训练与测试数据，即 80%的数据用于训练，20%的数据进行测试[25]。本文利用实验学习到的内容相似参数为 $p=0.25, q=4$，选择与 Struc2vec 相同的步长设置和窗口大小的结构相似参数。SN2vec 模型在 Wikipedia 数据集上得到的多标签分类结果优于基线模型，结果如表 2 所列。

表 2 多标签分类实验的 *Micro-F*1 结果

Table 2 *Micro-F*1 results of multi-label classification

Algorithm	Wikipedia	Brazilian air-traffic	American air-traffic
DeepWalk	0.658	0.249	0.523
LINE(2nd)	0.661	0.444	0.409
Node2vec	0.655	0.462	0.510
Struc2vec	0.178	0.714	0.515
SN2vec	**0.667**	**0.769**	**0.528**

在 Brazilian air-traffic 和 American air-traffic 数据集上，SN2vec 在使用结构相似性表示时采取与 Struc2vec 相同的随机游走超参数设置，即每个节点的游走步数为 10，步长为 80，滑动窗口大小为 10。根据 Brazilian air-traffic 和 American air-traffic 数据集均分为 4 组这一特定条件，在使用内容相似性表示时，SN2vec 侧重于对宽度和深度同时进行搜索，即 $p=q$。结合数据集的大小及实验学习参数情况，当 $p=q=0.25$ 时，SN2vec 取得的训练效果较好。

本文采用 Micro-F1 和平均 Micro-F1 作为多标签分类的评价指标，使用随机样本重复实验 10 次(80%的节点用于训练)，各方法的平均性能如表 2 和图 5 所示。为了与 OPenNE 中的基线模型进行比较，Wikipedia 数据集的节点多标签分类结果的评价指标是 Micro-F1；Brazilian air-traffic 和 American air-traffic 数据集上节点多标签分类结果的评价指标为平均 Micro-F1。

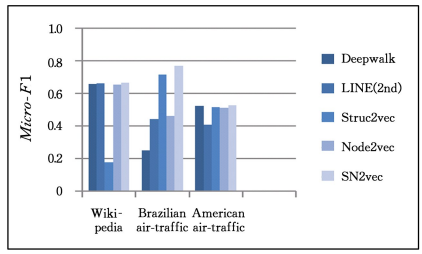

图 5 多标签分类的 *Micro-F*1 结果

Fig. 5 *Micro-F*1 results of multi-label classification

本文融合结构和节点嵌入的网络表示方法，采取 Struc2vec 和 Node2vec 模型的最优参数进行训练。表 2 结果显示，与 OpenNE 开源框架的基线模型在 Wikipedia 数据集上的结果相比，SN2vec 的 *Micro-F*1 值比经典的 DeepWalk，LINE(2nd)和 Node2vec 分别提高了 1%，0.6%和 1.2%；比传统的 Struc2vec 提高了约 48.9%，这可能与数据集中含有较少的远距离相似结构有关。SN2vec 弥补了 Struc2vec 不能同时捕获节点结构相似和节点内容相似的不足。

在节点数和边数较少的 Brazilian air-traffic 数据集上，与 Struc2vec，Node2vec 和 LINE 相比，SN2vec 在分类精度上分别提高了 5.5%，30.7%和 32.5%，这说明在相对较小的数据集上，SN2vec 表示网络结构的性能大幅提高；同时 Deepwalk 的分类精度表现较差，仅为 0.249，原因可能为 Deepwalk 基线模型采用原文献的最优参数进行训练，忽略了 Deepwalk 的随机游走参数设置较大，而 Brazilian air-traffic 数据集规模较小这一事实。

在 American air-traffic 数据集上，LINE 表现较差，原因可能是该数据集的数据是根据机场活跃度进行均分的，在一定程度上数据比较均匀且共享公共邻居节点较少。与 Struc2vec，Node2vec 和 Deepwalk 相比，SN2vec 在分类精度上分别提高了 1.3%，1.8%和 0.5%，原因可能是在网络数据进行低维嵌入时，基线模型方法采取了最佳的嵌入维度 128 维，而 SN2vec 方法联合表示嵌入的维度为 256 维，故对后续的分类精度有一定的影响。

4.4 降维可视化(t-SNE)实验及结果分析

通过 SN2vec 方法对 American air-traffic 和 Wikipedia 数据集进行网络学习表示后，运用 t-SNE 进行降维可视化，采用 Deepwalk，LINE 和 Node2vec 作为基线实验进行对比。实验结果如图 6、图 7 所示，其中不同颜色表示不同节点类别。

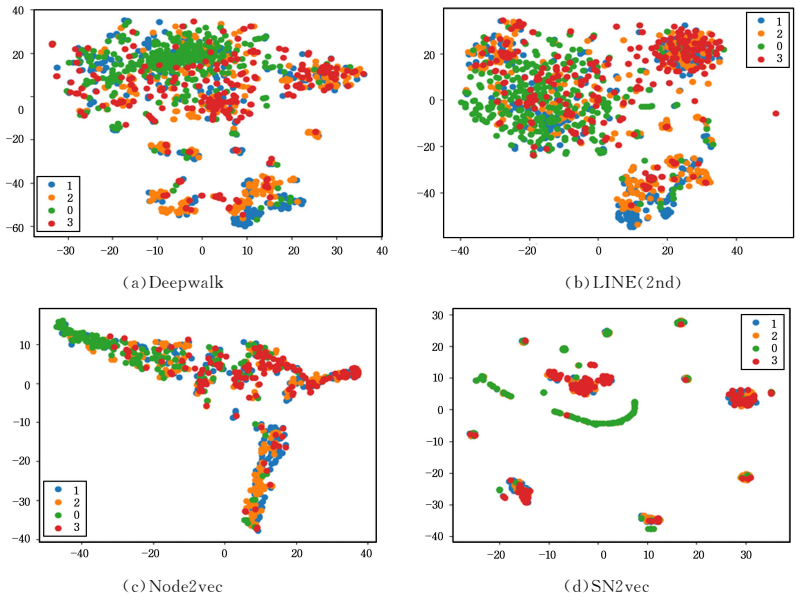

图 6　American air-traffic 的降维可视化(电子版为彩色)

Fig. 6　t-SNE of American air-traffic

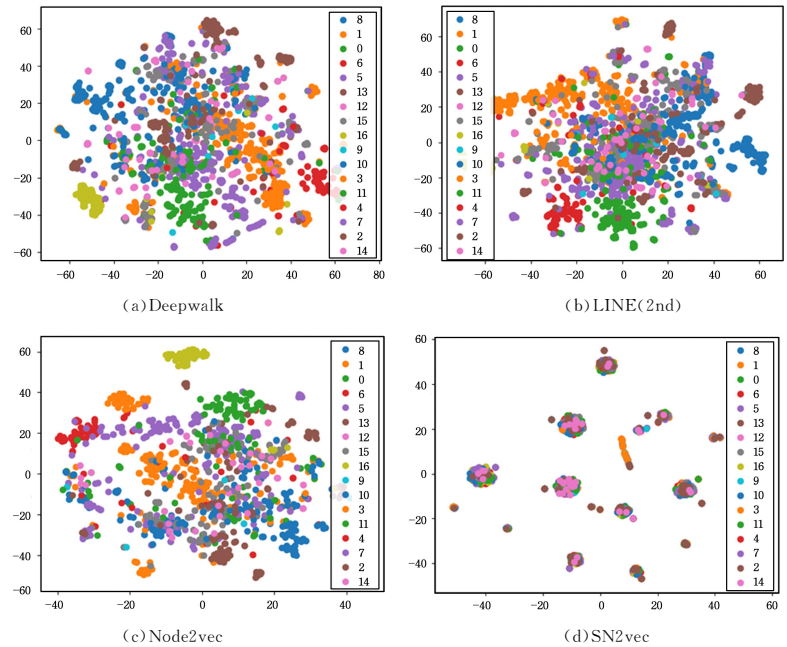

图 7　Wikipedia 的降维可视化(电子版为彩色)

Fig. 7　t-SNE of Wikipedia

在 t-SNE 降维可视化实验中，基线模型 Deepwalk，LINE 和 Node2vec 将高维的网络表示向量降为 2 维可视化发现，这 3 种方法均侧重于节点内容的表达。在全局结构上，图 6(a)、图 6(b)、图 7(a)和图 7(b)显示，Deepwalk 和 LINE 不能将在结构上潜在表示相同的节点很好地可视化；图 6(c)、图 7(c)显示，Node2vec 加入偏置游走参数后，在一定的局部结构上可学习到结构潜在表示相同的节点，但相比 SN2vec，Node2vec 仅学习到局部的结构潜在表示，并不能在全局的更大范围内进行学习；图 6(d)、图 7(d)的降维可视化显示，在全局结构上 SN2vec 可以很好地学习到结构潜在表示一致的节点，并且在这些结构中可以较好地根据节点内容来分辨其类别。

4.5　不同联合表示方式的实验及结果分析

本文提出的方法在联合表示时采取了向量拼接方式，在多标签分类结果上优于现有的浅层随机游走网络表示方法，但直接拼接的方式会扩大网络的嵌入维度。因此，本文分别通过加运算和乘运算的方式将局部和全局表示结合，提出 SN2vec 的变体 SN2vec^{+} 和 SN2vec*，以此分析不同联合表示方式对多标签分类精度的影响。具体实验结果如表 3 和图 8 所示。表 3 和图 8 的结果显示，在 Brazilian air-traffic 和 American air-traffic 数据集上，SN2vec^{+} 模型的多标签分类结果优于 SN2vec 和 SN2vec* 模型。在 Wikipedia 数据集上，SN2vec^{+} 的分类结果较差，猜想其可能与数据集的节点结构与内容不是严格意义上的等价有关。SN2vec* 表现较差的原因可能是网络表示后的向量多为小数类型，若在局部和全局联合表示上采取乘运算，则会影响嵌入表示向量的精度。

表 3 不同联合表示方式的多标签分类实验的 *Micro-F*1 结果

Table 3 *Micro-F*1 results of different joint representations in multi-label classification

Algorithm	Wikipedia	Brazilian air-traffic	American air-traffic
SN2vec	**0.667**	0.769	0.528
SN2vec^{+}	0.583	**0.792**	**0.715**
SN2vec*	0.408	0.560	0.504

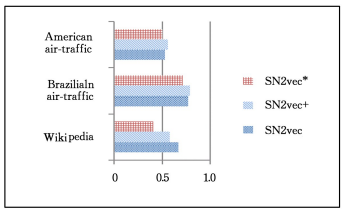

图 8 不同联合表示方式的多标签分类的 *Micro-F*1 结果比较

Fig. 8 Comparison of *Micro-F*1 of multi-label classification results for different joint representations

结束语 本文面向同质网络探索融合节点结构和节点内容相似的联合学习网络表示方法，提出了融合节点结构和节点内容的网络表示学习方法 SN2vec。实验表明，SN2vec 在 3 个数据集上的节点多标签分类结果全部优于基线模型，并且可以较好地学习到结构潜在表示一致的节点。然而，该模型依然存在浅层随机游走网络表示方法普遍存在的缺点，即过度强调度信息，且性能高度依赖超参数。因此，接下来将专注于解决 SN2vec 中超参数的依赖性问题，并继续探索模型融合表示后的维度空间对表示结果的影响。

参 考 文 献

[1] GOYAL P, FERRARA E. Graph Embedding Techniques, Applications, and Performance: A Survey[J]. Knowledge-Based Systems, 2017, 2(17): 155-164.

[2] TANG L, LIU H. Leveraging social media networks for classification[J]. Data Mining and Knowledge Discovery, 2011, 23(3): 447-478.

[3] HAMILTON W L, YING R, LESKOVEC J. Representation learning on graphs: Methods and applications[J]. arXiv, 2017: 1709.05584.

[4] BENGIO Y, COURVILLE A, VINCENT P. Representation learning: A review and new perspectives[J]. IEEE TPAMI, 2013, 35(8): 1798-1828.

[5] OU M D, CUI P, PEI J, et al. Asymmetric transitivity preserving graph embedding[C]// 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. 2016: 1105-1114.

[6] PENNINGTON J, SORCHER R, MANNING C D. GloVe: Global vectors for word respresentation[C]// Conference on Empirical Methods in Natural Language Processing. 2014.

[7] CAI H Y, ZHENG V W, CHANG K. A Comprehensive Survey of Graph Embedding: Problems, Techniques and Applications [J]. arXiv: 2017: 1709.07604.

[8] WILLIAM L H, YING R, LESKOVEC J. Inductive Representation Learning on Large Graphs[J]. arXiv: 2017: 1706.02216.

[9] BELKIN, MIKHAIL, PARTHA N. Laplacian eigenmaps and spectral techniques for embedding and clustering[J]. Advances in Neural Information Processing Systems, 2001, 14(6): 585-591.

[10] WOLD S, ESBENSEN K, GELADI P. Principal component analysis[J]. Chemometrics and Intelligent Laboratory Systems, 1987, 2(1/3): 37-52.

[11] JOSEPH B, KRUSKAL, WISH M. Multidimensional scaling [M]// Methods, 1978: 116.

[12] TU C C, YANG C, LIU Z Y. A summary of network represents learning[J]. Chinese Science: Information Science, 2017, 47(8): 980-996.

[13] ZHOU J, CUI G Q, ZHANG Z Y, et al. Graph Neural Networks: A Review of Methods and Applications[J]. arXiv: 1812.08434.

[14] WANG D X, CUI P, ZHU W W. Structural Deep Network Embedding[C]// The 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. 2016: 1225-1234.

[15] YANG C, LIU Z Y ZHAO D L, et al. Network representation learning with rich text information[C]// International Joint Conference on Artificial Intelligence (IJCAI). 2015: 2111-2117.

[16] LI Q H, LI C P, ZHANG J, et al. Survey of Compressed Deep Neural Network[J]. Computer Science, 2019, 46(9): 1-14.

[17] MIKOLOV T, SUTSKEVER I, CHEN K, et al. Distributed Representations of Words and Phrases and their Compositionality[C]// Annual Conference on Neural Information Processing Systems (NIPS). 2013: 3111-3119.

[18] PEROZZI B, AL-RFOU R, SKIENA S. DeepWalk: Online Learning of Social Representations[C]// The 20th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD). 2014: 701-710.

[19] TANG T, QU M, WANG M, et al. LINE: Large-scale Information Network Embedding[C]// The 24th International Conference on World Wide Web (WWW). 2015: 1067-1077.

[20] GROVER A, LESKOVEC J. node2vec: Scalable Feature Learning for Networks[C]// 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. 2016: 855-864.

[21] LEONARDO F, RIBEIRO R, PEDRO H P, et al. struc2vec: Learning Node Representations from Structural Identity[C]// The 23rd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining(KDD). 2017: 385-394.

[22] XU H, LIU H, WANG W, et al. NE-FLGC: Network Embedding Based on Fusing Local (First-Order) and Global (Second-Order) Network Structure with Node Content[C]// Pacific-Asia Conference on Knowledge Discovery and Data Mining (PAKDD). 2018: 260-271.

[23] CAO S H, LU W, XU Q K. GraRep: Learning Graph Representations with Global Structural Information[C]// The 24th ACM International Conference on Knowledge Discovery and Data Mining (KDD). 2015: 1105-1114.

[24] KAZEMI SEYED M, GOEL R, JAIN K, et al. Relational representation learning for dynamic (knowledge) graphs: a survey [J]. arXiv: 2019: 1905.11485.

[25] Tsinghua University built on open source framework OpenNE [EB/OL]. http://tech.ifeng.com/a/20171028/44733568_0.shtml.

ZHANG Hu, born in 1979, Ph.D, associate professor, is a member of China Computer Federation. His main research interests include Natural Language Processing and representation learning.