似性;(2) 能够保留节点特征相似性;(3) 维度不能太大.接下来,我们将给出一个新的图表示学习方法 GeVI,该模型能够同时满足上述图表示学习的要求.

2.1 模型描述

记 $v_i \in V$ 为图 $G=(V,E)$中的第 i 个顶点,$v_j \in N(v_i)$为 v_i 的邻居,即,v_i 可以通过 E 中的若干条连接边到达 v_j.并记 $\Phi(v_i) \in R^d$ 为需要求取的 v_i 的低维表示.则类似于自然语言处理中的 skip-gram 模型[30],可以通过优化与图节点相关的联合概率(公式(1))获得$\Phi(v_i)$的表达:

$$\arg\max_{\Phi(v_i)} \prod_{v_i \in V} [\prod_{v_j \in N(v_i)^*} p(v_j \mid v_i; \Phi(v_i))] \tag{1}$$

为了将已知的节点特征融入到该模型中,可以将条件概率 $p(v_j|v_i;\Phi(v_i))$表示为公式(2)的形式:

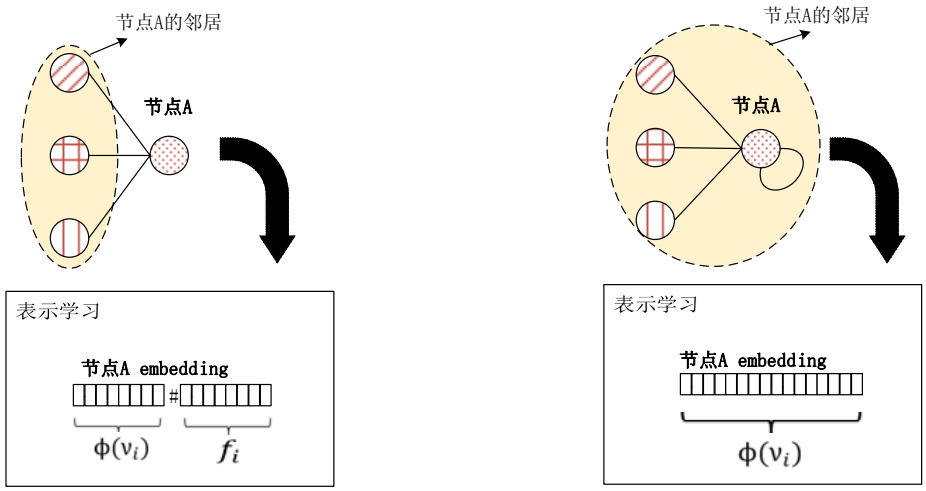
$$p(v_j \mid v_i; \Phi(v_i)) = \frac{\exp(f_j \cdot \Phi(v_i))}{\sum_{v_k \in V} \exp(f_k \cdot \Phi(v_i))} \tag{2}$$

则有:

$$p(v_j|v_i;\Phi(v_i)) \propto \exp(f_j \cdot \Phi(v_i)) \tag{3}$$

其直观的解释是:对于图 G 中每个节点 v_i,我们可以寻找一个低维表示$\Phi(v_i) \in R^d, d<<|V|$,使得$\Phi(v_i)$能够较好地解释邻居节点 v_j 的先验特征.将条件概率写成公式(2)的形式,事实上借鉴了 skip-gram 模型[30]的基本思想:如果两个单词(节点)有相似的上下文,则它们是相似的.在图的表示学习中,如果两个节点拥有共同或者特征相近的邻居节点(如图 1(a)所示),那么两个节点具有相似的低维表示.由于节点特征是已知的,可以要求学习到的节点的低维表示能够解释其邻居的节点特征,为此,将条件概率设计如公式(2).另一方面,如图 1(b)所示,还需要考虑当前节点特征的相似性,为此,需将当前节点特征融入到其中,我们分别设计了如图 2 所示的两种方案.

- 方案 1(GeVI.v1):拼接,即,表示学习后进行拼接.具体而言,在目标函数(1)中,令 $N(v_i)^*=N(v_i)$(如图 2(a)所示),并对公式(2)进行优化,将学习获得的$\Phi(v_i)$与 f_i 进行拼接,获得节点 v_i 的表示.在这种情况下,$\Phi(v_i)$满足了图 1(a)中所示的网络信息的刻画和学习,f_i 满足了图 1(b)所示的当前节点特征的刻画.
- 方案 2(GeVI.v2):融合,即,表示学习的过程中进行融合.具体而言,在目标函数(1)中,令 $N(v_i)^*=N(v_i) \cup \{v_i\}$(如图 2(b)所示).其直观含义是:使得所学习到的节点向量$\Phi(v_i)$既能解释自身节点特征,也能解释网络中邻居节点特征.



(a) 方案 1:GeVI.v1 (b) 方案 2:GeVI.v2

Fig.2 Two schemes to incorporate the vertices' features

图 2 融合节点特征的两种方案

具体实现上,可以通过有差别的采样方法分别实现方案 1 和方案 2 中的目标优化.

由于公式(2)中归一化项的时间复杂度很高,本文利用文献[30]提出的负采样方法进行优化.具体的,我们不是直接优化公式(1)中的目标函数,而是通过最大化如下转换函数来学习节点的表示向量:

$$\Delta_{i,j}=\log\sigma(\Phi(v_i)^T f_j)+\sum_{t=1}^{s}E_{v_t\sim P_n(v)}[\log\sigma(-\Phi(v_i)^T f_t)] \tag{4}$$

其中,$\sigma(x)=\dfrac{1}{1+\exp(-x)}$ 是 sigmoid 函数.$v_t\sim P_n(v)$说明按照分布 $P_n(v)$抽取节点 v_i 的负样本 v_t.记 v_t 在图中的度数为 $\deg(v_t)$,与文献[22]的设定一样,我们定义 $P_n(v)\propto\deg(v)^{3/4}$.对于每个目标节点,都有 s 个负样本点.记 C_i 为节点 v_i 的邻居集合,其中,$C_i\subset V$,最终,融合网络信息和节点特征的目标函数定义为

$$L=-\sum_{v_i\in V}\sum_{v_j\in C_i}\Delta_{i,j} \tag{5}$$

为了简化模型,我们对公式(5)进行改写,最终形式如下:

$$L=-\sum_{(v_i,v_j,\gamma)\in D}\log\sigma(\gamma\Phi(v_i)^T f_j) \tag{6}$$

其中,样本集合 D 包含了所有的正负样本,里面的每一个元素含义如下:若γ=1,则 v_j 是 v_i 的正样本,即 v_j 是 v_i 的邻居;若γ=−1,则 v_j 是 v_i 的负样本,即 v_j 是通过分布 $P_n(v)$抽样得到的.每个正样本都对应 s 个负样本.

2.2 学习算法及复杂度分析

针对第 2.1 节所述的模型,学习算法的设计主要涉及两个关键部分:邻居节点的采样以及目标函数的优化.以下我们进行详细阐述.

本文借鉴文献[22,23]的思路,采用邻居的推广定义.具体地,在 DeepWalk 模型中,给定窗口大小 k,在 G 上随机生成的游走路径中,如果节点 v_j 出现在节点 v_i 的邻居窗口中,则节点 v_j 是节点 v_i 的邻居节点.即 v_j 不必是 v_i 的直接邻居,比如$(v_i,v_j)\notin E$,但只要 v_i 能够在 k 步内到达 v_j 即可.实施上,类似于 DeepWalk 的设定,首先利用截断的随机游走从图 G 中生成大量的路径,然后从这些生成的路径中获得每个节点 $v_i\in V$ 的邻居.在 DeepWalk 中,节点的邻居不包括其本身,而由于我们将节点特征作为其先验信息,因此可以将节点本身作为其邻居,即:要求当前节点不但能够解释周围节点,还要能够解释其本身.本文采用了两种方案:对于 GeVI.v1,节点的邻居不包括其本身,最终的节点表示为学习到的向量$\Phi(v_i)$与其自身的节点特征向量 f_i 拼接得到;而 GeVI.v2 则将节点本身作为其邻居,将学习到的输入向量$\Phi(v_i)$直接作为图表示.

目标函数中样本集合 D 的生成过程见算法 1.首先生成所有的正样本集合 D_+.对于每个节点,以该节点为起始节点进行随机游走,生成长度为α_1 的路径 p.将集合$\{(p_i,p_j,1)|0<i,j\leqslant\alpha_1;|j-i|<\alpha_2\}$添加到正样本集合 D_+中,其中,p_i 为路径 p 的第 i 个元素.然后生成所有的负样本集合 D_-.对于 D_+中的每个样本$(p_i,p_j,1)$,从分布 p_n 中抽取 s 个节点,组成正样本$(p_i,p_j,1)$对应的负样本集合$\{(p_i,p_k,-1)\}^s$,然后将其添加到负样本集合 D_-中.最后,$D=D\cup D_+\cup D_-$.重复上述过程α_4 次.

算法 1. 负采样.

Input:图 $G(V,E)$,随机游走长度α_1,滑动窗口大小α_2,负样本比例α_3,以每个节点作为开始节点游走的次数α_4;

Output:训练数据集:D.

```
D+←0; D-←0; D←0;
i←0
for k<α4 do:
    for v in |V| do:
        通过随机游走算法从节点 v 开始,生成一条长度为α1 的路径 p.
        for i<α1 do:
            将正样本集合{(pi,pj,1)||i−j|<α2,i≠j}放进 D+中(GeVI.v1) or
            将正样本集合{(pi,pj,1)||i−j|<α2}放进 D+中(GeVI.v2)
        endfor
```

9: **end for**

10: **end for**

11: **for** $(p_i,p_j,1)\in D_+$ **do**:

12: 将负样本集合 $\{(p_i,p_k,-1)\,|\,k\sim p_n\}^{\alpha_2}$ 放进 D_-中.

13: **endfor**

14: **return** $D=D_+\cup D_-$

我们使用批量梯度下降的方法优化目标函数(4),由于 f_j 是已知的,因此只需要更新节点的表示向量 $\Phi(v_i)$.为方便表达,下式中将 $\Phi(v_i)$简写为 Φ_i.对于样本(v_i,v_j,γ),变量 Φ_i 的梯度计算见公式(7):

$$\frac{\partial L}{\partial \Phi_i}=\frac{\partial \log\sigma(\gamma\Phi_i^T f_j)}{\partial \Phi_i}=\frac{1}{\sigma(\gamma\Phi_i^T f_j)}\sigma(\gamma\Phi_i^T f_j)(1-\sigma(\gamma\Phi_i^T))\gamma\Phi_i=\gamma\Phi_i(1-\sigma(\gamma\Phi_i^T)) \tag{7}$$

由于目标函数大于 0,存在下界,所以最小化过程最终会收敛.目标函数优化过程见算法 2.第 1 行,我们首先随机初始化节点的输入向量 Φ.在第 2 行,根据算法 1 生成样本集合 D.在第 4 行,则将样本集合 D 划分成 b 个互不相交的集合.然后在第 8 行,累加样本点关于输入向量的梯度.在第 10 行,利用批量随机梯度下降方法更新参数 Φ.重复上述过程,直到算法收敛.

算法 2. 目标函数最优化.

Input:图 $G(V,E)$,节点特征向量 f_i,随机游走长度α_1,滑动窗口大小α_2,负样本比例α_3,以每个节点为起始节点游走的次数α_4,节点表示向量维度 d,批量大小 b,梯度更新步长η;

Output:图表示$\Phi\in R^{|V|\times d}$.

1: 随机初始化节点表示向量Φ

2: $D\leftarrow NegSample(G,\alpha_1,\alpha_2,\alpha_3,\alpha_4)$

3: **while** 不收敛 **do**:

4: $Batches\leftarrow ConstgructBatch(D,b)$

5: **for** each *batch B* in *Batchs* do:

6: $\nabla\Phi=0$

7: **for** each sample (v_i,v_j,γ) in B do:

8: $\nabla\Phi_i\leftarrow\nabla\Phi_i+\frac{\partial L}{\partial\Phi_i}$

9: **end for**

10: $\Phi\leftarrow\Phi-\eta\nabla\Phi$

11: **end for**

12: **end while**

- 时间复杂度分析

在算法 1 中,一共生成了$|V|\times\alpha_4$ 条路径,对于每条路径,需要生成 $2\alpha_1\alpha_2$ 个正样本,生成每个样本的时间复杂度为 $O(1)$,因此,生成所有正样本的时间复杂度为 $O(2\alpha_1\alpha_2\alpha_4|V|)$(算法 1 中第 3 行~第 10 行).对于每个正样本,需要采样出α_3个负样本,采样每个负样本的时间复杂度为 $O(1)$,因此,生成负样本的时间复杂度为 $O(2\alpha_1\alpha_2\alpha_3\alpha_4|V|)$.综上,算法 1 的时间复杂度为 $O(2\alpha_1\alpha_2\alpha_3\alpha_4|V|)$,正比于节点数量$|V|$.

在算法 2 中,首先根据算法 1 生成所有的样本点,其时间复杂度为 $O(2\alpha_1\alpha_2\alpha_3\alpha_4|V|)$.然后对于每次循环,需要计算每个样本点的梯度,梯度计算的时间复杂度为 $O(1)$,一共有 $2\alpha_1\alpha_2\alpha_3\alpha_4|V|$个样本点,因此每次更新的时间复杂度为 $O(2\alpha_1\alpha_2\alpha_3\alpha_4|V|)$.因此,算法 2 的时间复杂度为 $O(4\alpha_1\alpha_2\alpha_3\alpha_4|V|)$,正比于节点数量$|V|$.

综上,GeVI 模型的时间复杂度为 $O(|V|)$.说明本文提出的算法能够应用于大规模图表示学习任务中.

3 对比实验

为了验证本文提出算法的有效性,我们在 3 个公开数据集(Citeseer,Cora,PubMed)上与几个具有代表性的图表示学习方法进行对比.

3.1 实验设定

实验方案与文献[22]类似,首先利用无监督的方法学习图表示,然后将其应用在多分类任务中.本文采用 *MicroF*1 和 *MacroF*1 两个指标作为模型性能的评测标准.计算公式如下:

$$MicroR=\frac{\sum_{i=1}^{k}TP_i}{\sum_{i=1}^{k}TP_i+FN_i} \tag{8}$$

$$MicroP=\frac{\sum_{i=1}^{k}TP_i}{\sum_{i=1}^{k}TP_i+FP_i} \tag{9}$$

$$MicroF1=2\frac{MicroR\cdot MicroP}{MicroR+MicroP} \tag{10}$$

$$MacroR=\frac{1}{k}\sum_{i=1}^{k}\frac{TP_i}{TP_i+FN_i} \tag{11}$$

$$MacroP=\frac{1}{k}\sum_{i=1}^{k}\frac{TP_i}{TP_i+FP_i} \tag{12}$$

$$MacroF1=2\frac{MacroR\cdot MacroP}{MacroR+MacroP} \tag{13}$$

其中,k 表示类别数,TP_i 表示在类别 i 上的预测正确的正类数,FN_i 表示在类别 i 上预测错误的负类数,FP_i 表示在类别 i 上预测错误的正类数.

(1) 数据集

我们在 3 个公开数据集上进行评测,包括 Citeseer,Cora,PubMed.去掉孤立点后,各个数据集的相关统计信息说明如下,见表 2.

- CiteSeer:包含 3 264 篇出版物、4 591 条边、6 个类别,每个类别表示出版物的细分领域.其中,每篇出版物均由长度为 3 703 的二值向量表示,每一维表示该出版物是否出现相应的单词;
- Cora:包含 2 708 篇机器学习领域的论文、5 429 条边、7 个类别,每个类别表示论文的细分领域,边表示了论文之间的引用关系.其中,每篇论文均由长度为 1 433 的二值向量表示,每一维表示该文档是否出现相应的单词;
- PubMed:包含 19 717 篇生物医学领域的文章、44 338 条边、3 个类别,每个类别表示文章的细分领域,边表示论文之间的引用关系.其中,每篇论文均由长度为 500 的 TF-IDF 向量表示.

(2) 对比方法

我们将对比方法分成 3 类:(1) 仅利用节点特征;(2) 仅利用网络信息;(3) 同时利用节点特征和网络信息.对于第 3 类,我们进一步分成两小类,分别是:(a) 原始模型不可以利用两种信息,扩展原有算法,使得它们可以同时利用网络信息和节点特征;(b) 本来就可以直接利用网络信息和节点特征的方法.

Table2 Data sets

表 2 数据集

Dataset	节点数量	边数量	类别数
Citeseer	3 264	4 591	6
Cora	2 708	5 429	7
PubMed	19 717	44 338	3

以下是对比方法的简要介绍.

- SVD:直接通过 SVD 分解将节点特征信息进行降维,降维后的向量作为图表示.属于第 1 类对比方法,只利用了节点特征;
- DeepWalk[22]:DeepWalk 只利用了网络信息学习图表示;
- DeepWalk#SVD:通过将 DeepWalk 和 SVD 得到的图表示进行拼接,扩展了原有算法,使得到的图表示同时包含了图结构信息和节点特征;
- TADW[25]:TADW 通过矩阵分解的形式,直接利用了网络信息和节点特征学习得到图表示.

(3) 参数设定

基于 SVD 分解方法的维度统一取 200(与文献[25]保持一致).TADW 按照文献[25]中提供的最优参数进行设定,对于新增的数据集 PubMed,设置其 k=128(取 64,128,256 中最好的).由于 GeVI 模型要求节点特征向量的长度等于节点输入向量,因此我们首先通过 SVD 分解,将节点特征向量进行降维,将降维后的向量作为节点特征向量.表 3 中详细列举了我们在实验中用到的参数.由于 DeepWalk 与 GeVI 共享全部的参数,因此 DeepWalk 的参数设定与 GeVI 一致.

Table 3 Parameter setting

表 3 参数设定

数据集	α_1	α_2	α_3	α_4	d
Citeseer	40	5	9	10	128
Cora	40	5	9	10	128
Pubmed	10	2	1	10	128

3.2 结果及分析

在所有模型上,我们首先利用无监督的方法学习图表示.然后将其运用于节点分类任务,通过分类的效果判断图表示的质量.对于节点分类任务,我们首先将图节点随机等分成两部分,分别是测试集和训练集.然后在训练集上训练一个分类器,在测试集上进行测试.为了比较不同训练数据对模型性能的影响,我们进一步对训练集进行采样,分别取训练样本的 6%~100%(即整体样本量 3%~50%)的数据训练分类器,然后在测试集上进行测试.这样可以保证基于不同规模训练数据训练的分类器都在同一个测试集上进行测试.我们使用 Softmax 分类器,重复上述实验 10 次,报告平均结果.

表 4~表 6 报告了在 Citeseer,Cora,PubMed 数据集上的分类 *Micro-F*1 值和 *Macro-F*1 值(粗体表示所有方法中最好的,斜体表示基线方法中最好的,τ%(↑)表示 GeVI 相对于最好的基线算法的提升比例).实验结果显示:本文提出的方法比所有基线方法效果要好,证明了将节点特征看做是节点的先验信息的有效性.GeVI 模型比只利用了网络信息的模型(DeepWalk)或只利用了节点特征的模型(SVD)的效果好,这说明了融合节点特征的必要性.此外,GeVI 模型也比通过将两种信息进行简单拼接的方法(DeepWalk#SVD)及基于矩阵分解的信息融合方法 TADW 好,这说明了本文提出的信息融合方法的有效性.而 GeVI.v1 和 GeVI.v2 之间则不相上下:在 Citeseer 和 Cora 数据集上,GeVI.v2 效果最好;在 PubMed 数据集上,当训练数据小于 20%时,GeVI.v2 取得最优的效果,当训练数据大于 20%时,GeVI.v1 取得最优的效果.

Table 4 *Micro-F*1 and *macro-F*1 score on Citeseer dataset

表 4 在 Citeseer 数据集上的 *micro-F*1 值和 *macro-F*1 值

训练样本比例(η)		3%	5%	7%	10%	20%	30%	40%	50%
*Micro-F*1	SVD	0.564	0.597	0.618	0.627	0.650	0.667	0.676	0.687
	DeepWalk	0.437	0.455	0.470	0.479	0.512	0.532	0.541	0.548
	DeepWalk#SVD	0.557	0.582	0.617	0.632	0.663	0.681	0.683	0.696
	TADW	*0.639*	*0.656*	*0.668*	*0.677*	*0.696*	*0.708*	*0.712*	*0.718*
	GeVI.v1	0.648	0.674	0.682	0.694	0.702	0.713	0.718	0.725
	GeVI.v2	**0.678**	**0.697**	**0.703**	**0.709**	**0.726**	**0.733**	**0.742**	**0.742**
	τ%(↑)	6.10%	6.25%	5.24%	4.73%	4.31%	3.53%	4.21%	3.34%

Table 4 *Micro-F*1 and *macro-F*1 score on Citeseer dataset (Continued)
表 4 在 Citeseer 数据集上的 *micro-F*1 值和 *macro-F*1 值(续)

训练样本比例(η)		3%	5%	7%	10%	20%	30%	40%	50%
*Macro-F*1	SVD	0.499	0.538	0.568	0.575	0.606	0.624	0.635	0.645
	DeepWalk	0.394	0.412	0.426	0.436	0.462	0.479	0.489	0.494
	DeepWalk#SVD	0.496	0.529	0.566	0.584	0.618	0.639	0.642	0.655
	TADW	*0.564*	*0.583*	*0.609*	*0.621*	*0.640*	*0.656*	*0.662*	*0.667*
	GeVI.v1	0.574	0.610	0.628	0.641	0.655	0.670	0.678	0.687
	GeVI.v2	**0.598**	**0.625**	**0.641**	**0.654**	**0.670**	**0.682**	**0.693**	**0.692**
	τ%(↑)	6.03%	7.20%	5.26%	5.31%	4.69%	3.96%	4.68%	3.75%

Table 5 *Micro-F*1 and *macro-F*1 score on Cora dataset
表 5 在 Cora 数据集上的 *micro-F*1 值和 *macro-F*1 值

训练样本比例(η)		3%	5%	7%	10%	20%	30%	40%	50%
*Micro-F*1	SVD	0.525	0.587	0.611	0.647	0.688	0.707	0.721	0.731
	DeepWalk	0.631	0.669	0.695	0.722	0.748	0.766	0.775	0.783
	DeepWalk#SVD	*0.665*	*0.717*	*0.742*	*0.765*	*0.792*	*0.814*	*0.820*	*0.826*
	TADW	0.601	0.653	0.677	0.703	0.734	0.747	0.75	0.756
	GeVI.v1	0.688	0.757	0.769	0.797	0.823	0.838	0.845	0.855
	GeVI.v2	**0.745**	**0.8**	**0.813**	**0.829**	**0.846**	**0.86**	**0.865**	**0.87**
	τ%(↑)	12.03%	11.58%	9.57%	8.37%	6.82%	5.65%	5.49%	5.33%
*Macro-F*1	SVD	0.442	0.525	0.563	0.613	0.657	0.680	0.697	0.709
	DeepWalk	0.593	0.652	0.681	0.709	0.737	0.758	0.767	0.776
	DeepWalk#SVD	*0.619*	*0.693*	*0.723*	*0.749*	*0.778*	*0.802*	*0.809*	*0.813*
	TADW	0.519	0.601	0.641	0.674	0.709	0.725	0.728	0.735
	GeVI.v1#SVD	0.621	0.724	0.745	0.778	0.807	0.823	0.831	0.840
	GeVI.v2	**0.701**	**0.782**	**0.802**	**0.817**	**0.836**	**0.852**	**0.855**	**0.861**
	τ%(↑)	13.25%	12.84%	10.93%	9.08%	7.46%	6.23%	5.69%	5.90%

Table 6 *Micro-F*1 and *macro-F*1 score on PubMed dataset
表 6 在 Pubmed 数据集上的 *micro-F*1 值和 *macro-F*1 值

训练样本比例(η)		3%	5%	7%	10%	20%	30%	40%	50%
*Micro-F*1	SVD	*0.709*	*0.755*	*0.781*	*0.801*	*0.820*	*0.829*	*0.834*	0.836
	DeepWalk	0.626	0.639	0.645	0.652	0.662	0.665	0.667	0.668
	DeepWalk#SVD	0.678	0.712	0.738	0.762	0.805	0.822	0.833	*0.839*
	TADW	0.653	0.709	0.749	0.779	0.806	0.819	0.826	0.830
	GeVI.v1	0.797	0.803	0.811	0.820	**0.839**	**0.847**	**0.854**	**0.858**
	GeVI.v2	**0.815**	**0.818**	**0.825**	**0.831**	0.838	0.842	0.844	0.845
	τ%(↑)	14.95%	8.34%	5.63%	3.75%	2.32%	2.17%	2.40%	2.27%
*Macro-F*1	SVD	*0.672*	*0.739*	*0.775*	*0.799*	*0.820*	*0.830*	*0.835*	0.837
	DeepWalk	0.582	0.595	0.600	0.610	0.622	0.623	0.626	0.628
	DeepWalk#SVD	0.651	0.693	0.726	0.755	0.803	0.821	0.832	*0.838*
	TADW	0.559	0.662	0.730	0.772	0.805	0.818	0.825	0.829
	GeVI.v1	0.787	0.795	0.804	0.813	**0.833**	**0.843**	**0.850**	**0.855**
	GeVI.v2	**0.807**	**0.812**	**0.819**	**0.824**	0.832	0.836	0.838	0.839
	τ%(↑)	20.09%	9.88%	5.68%	3.13%	1.59%	1.57%	1.80%	2.03%

记η为使用到的训练样本的比例.从表 4~表 6 中可以观察到.

1) 相对于基线方法,GeVI 模型在训练数据少的情况下具有更大的优势.大部分基线方法随着η的下降,性能迅速变差.因为这些方法学习到的图表示中含有大量的噪声,且训练数据与测试数据之间存在不一致性;相反地,由于 GeVI 模型同时利用了网络信息和节点特征,因此学到的图表示存在更少的噪声,数据之间的一致性更强;
2) 在 Citeseer 和 Cora 数据集上,GeVI.v2 模型性能最好.在 Citeseer 数据集上,以 *Micro-F*1 为评价标准,GeVI.v2 模型比最好的基线方法相对提高了 3.34%(η=50%)~6.25%(η=5%);以 *Macro-F*1 为评价标准,相对提高了 3.75%(η=50%)~7.20%(η=5%);
3) 在 Cora 数据集上,以 *Micro-F*1 为评价标准,GEVI.V2 模型比最好的基线方法相对提高了 5.33%

(η=50%)~12.03%(η=3%);以 *Macro-F*1 为评价标准,相对提高了 5.69%(η=40%)~13.25%(η=3%).这充分表明本文提出算法的稳定性和有效性;

4) 在 PubMed 数据集上,虽然 GeVI 模型均比所有对比方法好,但相比于数据集 Citeseer 和 Cora,GeVI 模型在 PubMed 上的优势没这么明显.随着η的提高,GeVI 模型与对比方法之间的性能差别迅速减小.特别地,对于 GeVI.v2 模型,以 *Micro-F*1 为性能评价指标,性能提高的幅度从 14.95%(η=3%)下降到 0.72%(η=50%);以 *Macro-F*1 为性能评价指标,性能提高的幅度从 20.09%(η=3%)下降到 0.12%(η=50%).此外,当η>20%的时候,GeVI.v1 模型的性能比 GeVI.v2 的好.这是因为对于 Pubmed 而言,节点特征在图分类任务中比网络信息更重要(在对比方法中,SVD 分解在大多数情况下均比其他方法好),相对于 GeVI.v2 方法,GeVI.v1 方法在对内容信息提取方面的损失更少;同时,由于 pubmed 数据量足够(当训练样本达到 20%时,样本总量可以达到 3 400),因此虽然 GeVI.v1 方法得到的节点向量维度较高,但随着数据量的增加,GeVI.v1 方法还是获得了较稳定、一致的提升.

5) 综上,我们可以发现如下规律:当节点数量比较少的时候,应该将节点本身作为其邻居,然后学习出统一的图表示(GeVI.v2);当节点数量很多,而且节点特征很重要的时候,为了最大限度地保留节点特征,我们将节点本身从其邻居中去掉,然后将学习到的输入向量与其节点特征进行拼接,作为图表示(GeVI.v1).

4 总 结

通过将节点特征作为先验知识,并令当前节点解释邻居节点,本文提出了融合网络信息和节点特征的图表示学习方法.为了将当前结点特征更好地融合到节点表示中,我们提出了两种具体实现方案,即:

(1) GeVI.v1:不把节点本身作为其邻居,最后通过将学习到的表示向量与其本身的特征向量进行拼接作为图表示;

(2) GeVI.v2:将节点本身作为其邻居,学习到的表示向量直接作为图表示.

实验结果表明了这两种方法均比对比的基线方法效果好.具体地,当节点数量不多,网络信息与节点特征的重要性差不多的时候,第 2 种融合方案 GeVI.v2 更好;当节点数量比较多,节点特征很重要的时候,第 1 种融合方案 GeVI.v1 更优.在未来的工作中,我们将进一步考虑更高效的网络信息的利用方法,比如通过 Node2vec 获取网络信息,Node2vec 通过结合深度优先和广度优先两种游走方法,以更充分地利用网络信息.此外,还将进一步研究在某些图中可能存在部分节点特征缺失的情况.本文提出的方法也可以扩展到其他的表示学习领域,比如自然语言处理中的词嵌入表示学习,例如,利用类似于本文提出的方法将单词的附加属性嵌入到其表示中.

References:

[1] Sen P, Namata G, Bilgic M, Getoor L, Galligher B, Eliassi-Rad T. Collective classification in network data. AI Magazine, 2008, 29(3):93. [doi: 10.1609/aimag.v29i3.2157]

[2] Zhou DY, Huang JY, Scholkopf B. Learning from labeled and unlabeled data on a directed graph. In: Raedt LD, Wrobe S, eds. Proc. of the 22nd Int'l Conf. on Machine Learning. 2005. 1036−1043. [doi: 10.1145/1102351.1102482]

[3] Bhagat S, Cormode G, Muthukrishnan S. Node classification in social networks. In: Aggarwal C, ed. Proc. of the Social Network Data Analytics. Boston: Springer-Verlag, 2011. 115−148. [doi: 10.1007/978-1-4419-8462-3_5]

[4] Tu C, Liu Z, Sun M. Inferring correspondences from multiple sources for microblog user tags. In: Huang H, Liu T, Zhang HP, Tang J, eds. Proc. of the Chinese National Conf. on Social Media Processing. Berlin, Heidelberg: Springer-Verlag, 2014. 1−12. [doi: 10.1007/ 978-3-662-45558-6_1]

[5] Xing QL, Liu L, Liu YQ, Zhang M, Ma SP. Study on user tags in Weibo. Ruan Jian Xue Bao/Journal of Software, 2015,26(7): 1626−1637 (in Chinese with English abstract). http://www.jos.org.cn/1000-9825/4655.htm [doi: 10.13328/j.cnki.jos.004655]

[6] Bhuyan MH, Bhattacharyya DK, Kalita JK. Network anomaly detection: Methods, systems and tools. IEEE Communications Surveys & Tutorials, 2014,16(1):303−336. [doi: 10.1109/SURV.2013.052213.00046]

[7] Lü L, Zhou T. Link prediction in complex networks: A survey. Physica A: Statistical Mechanics and Its Applications, 2011,390(6): 1150−1170. [doi: 10.1016/j.physa.2010.11.027]

[8] Liben−Nowell D, Kleinberg J. The link−prediction problem for social networks. Journal of the Association for Information Science and Technology, 2007,58(7):1019−1031. [doi: 10.1002/asi.20591]

[9] Yu X, Ren X, Sun Y, Gu Q, Sturt B, Khandelwal U, Norick B, Han J. Personalized entity recommendation: A heterogeneous information network approach. In: Carterette B, ed. Proc. of the 7th ACM Int'l Conf. on Web Search and Data Mining. New York: ACM Press, 2014. 283−292. [doi: 10.1145/2556195.2556259]

[10] Meng XW, Liu SD, Zhang YJ, Hu X. Research on social recommender systems. Ruan Jian Xue Bao/Journal of Software, 2015, 26(6):1356−1372 (in Chinese with English abstract). http://www.jos.org.cn/1000-9825/4831.htm [doi: 10.13328/j.cnki.jos.004831]

[11] Liu ZY, Sun MS, Lin YK, Xie RB. Knowledge representation learning: A review. Ji Suan Ji Yan Jiu Yu Fa Zhan/Journal of Computer Research and Development, 2016,53(2):247−261 (in Chinese with English abstract). [doi: 10.7544/issn1000-1239.2016. 20160020]

[12] Roweis ST, Saul LK. Nonlinear dimensionality reduction by locally linear embedding. Science, 2000,290(5500):2323−2326. [doi: 10.1126/science.290.5500.2323]

[13] Belkin M, Niyogi P. Laplacian eigenmaps and spectral techniques for embedding and clustering. In: Suzanna B, Sebastian T, Klaus O, eds. Proc. of the Advances in Neural Information Processing Systems. Cambridge: MIT Press, 2002. 585−591.

[14] Chen M, Yang Q, Tang X. Directed graph embedding. In: Rajeev S, Mehta H, Bagga RK, eds. Proc. of the IJCAI. San Francisco: Morgan Kaufmann Publishers Inc., 2007. 2707−2712.

[15] Tang L, Liu H. Relational learning via latent social dimensions. In: Osmar RZ, ed. Proc. of the 15th ACM SIGKDD Int'l Conf. on Knowledge Discovery and Data Mining. New York: ACM Press, 2009. 817−826. [doi: 10.1145/1557019.1557109]

[16] Le TM, Lauw HW. Probabilistic latent document network embedding. In: Kumar R, ed. Proc. of the IEEE Int'l Conf. on Data Mining (ICDM). Shenzhen: IEEE, 2014. 270−279. [doi: 10.1109/ICDM.2014.119]

[17] Jacob Y, Denoyer L, Gallinari P. Learning latent representations of nodes for classifying in heterogeneous social networks. In: Carterette B, ed. Proc. of the 7th ACM Int'l Conf. on Web Search and Data Mining. New York: ACM Press, 2014. 373−382. [doi: 10.1145/2556195.2556225]

[18] Bourigault S, Lagnier C, Lamprier S, Denoyer L, Gallinari P. Learning social network embeddings for predicting information diffusion. In: Carterette B, ed. Proc. of the 7th ACM Int'l Conf. on Web Search and Data Mining. New York, 2014. 393−402. [doi: 10.1145/2556195.2556216]

[19] Nallapati RM, Ahmed A, Xing EP, Cohen WW. Joint latent topic models for text and citations. In: Li Y, Liu B, Sarawagi S, eds. Proc. of the 14th ACM SIGKDD Int'l Conf. on Knowledge Discovery and Data Mining. New York: ACM Press, 2008. 542−550. [doi: 10.1145/1401890.1401957]

[20] Chang J, Blei D. Relational topic models for document networks. In: Dyk DV, Welling M, eds. Proc. of the Int'l Conf. on Artificial Intelligence and Statistics. Florida: PMLR, 2009. 81−88.

[21] Le T, Lauw HW. Probabilistic latent document network embedding. In: Kumar R, ed. Proc. of the 2014 IEEE Int'l Conf. on Data Mining (ICDM). Shenzhen: IEEE, 2014. 270−279. [doi: 10.1109/ICDM.2014.119]

[22] Perozzi B, Al-Rfou R, Skiena S. Deepwalk: Online learning of social representations. In: Perlich C, Saka E, Shen D, Lee K, Li Y, eds. Proc. of the 20th ACM SIGKDD Int'l Conf. on Knowledge Discovery and Data Mining. New York: ACM Press, 2014. 701−710. [doi: 10.1145/2623330.2623732]

[23] Grover A, Leskovec J. node2vec: Scalable feature learning for networks. In: Proc. of the 22nd ACM SIGKDD Int'l Conf. on Knowledge Discovery and Data Mining. ACM Press, 2016. 855−864. [doi: 10.1145/2939672.2939754]

[24] Tang J, Qu M, Wang M, Zhang M, Yan J, Mei Q. Line: Large-Scale information network embedding. In: Gangemi A, ed. Proc. of the 24th Int'l World Wide Web Conf. on Steering Committee Republic and Canton of Geneva. 2015. 1067−1077. [doi: 10.1145/ 2736277.2741093]

[25] Yang C, Liu Z, Zhao D, Sun M, Chang EY. Network representation learning with rich text information. In: Yang Q, Wooldridge M, eds. Proc. of the IJCAI. 2015. 2111−2117.

[26] Yang Z, Cohen WW, Salakhutdinov R. Revisiting semi-supervised learning with graph embeddings. In: Balcan MF, Weinberger KQ, eds. Proc. of the 33rd Int'l Conf. on Machine Learning (ICML 2016). New York: JMLR.org, 2016. 40−48.

[27] Chojnacki W, Brooks MJ. A note on the locally linear embedding algorithm. Int'l Journal of Pattern Recognition and Artificial Intelligence, 2009,23(8):1739−1752. [doi: 10.1142/S0218001409007752]

[28] Newman ME. Modularity and community structure in networks. Proc. of the National Academy of Sciences, 2006,103(23): 8577−8582. [doi: 10.1073/pnas.0601602103]

[29] Iwata T, Saito K, Ueda N, Stromsten S, Griffiths TL, Tenenbaum JB. Parametrice mbedding for class visualization. Neural Computation, 2007,19(9):2536−2556. [doi: 10.1162/neco.2007.19.9.2536]

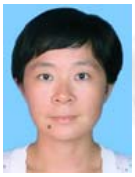
[30] Mikolov T, Sutskever I, Chen K, Corrado GS, Dean J. Distributed representations of words and phrases and their compositionality. In: Burges CJC, Bottou L, Welling M, Ghahramani Z, Weinberger KQ, eds. Proc. of the Advances in Neural Information Processing Systems. MIT Press, 2013. 3111−3119.

附中文参考文献:

[5] 邢千里,刘列,刘奕群,张敏,马少平.微博中用户标签的研究.软件学报,2015,26(7):1626−1637. http://www.jos.org.cn/1000-9825/4655.htm [doi: 10.13328/j.cnki.jos.004655]

[10] 孟祥武,刘树栋,张玉洁,胡勋.社会化推荐系统研究.软件学报,2015,26(6):1356−1372. http://www.jos.org.cn/1000-9825/4831.htm [doi: 10.13328/j.cnki.jos.004831]

[11] 刘知远,孙茂松,林衍凯,谢若冰.知识表示学习研究进展.计算机研究与发展,2016,53(2):247−261. [doi: 10.7544/issn1000-1239.2016.20160020]



温雯(1981－),女,江西赣州人,博士,副教授, CCF 专业会员,主要研究领域为数据挖掘,机器学习.

郝志峰(1968－),男,博士,教授,博士生导师,主要研究领域为数据挖掘,机器学习.

黄家明(1992－),男,硕士,CCF 学生会员,主要研究领域为数据挖掘,机器学习.

王丽娟(1978－),女,博士,副教授,CCF 专业会员,主要研究领域为数据挖掘,机器学习.

蔡瑞初(1983－),男,博士,教授,博士生导师,CCF 高级会员,主要研究领域为因果关系,数据挖掘,机器学习.