

一种基于层级信息优化的有向网络表示学习方法



李鑫超 李培峰 朱巧明

苏州大学计算机科学与技术学院 江苏 苏州 215006

江苏省计算机信息技术处理重点实验室 江苏 苏州 215006

(20175227013@stu.suda.edu.cn)

摘要 网络表示方法旨在将每个节点映射到低维向量空间,并保留节点在网络中的结构关系。有向网络的环中节点相互可达,破坏了非对称传递性,影响了模型对网络整体结构信息的学习。为削弱有向网络的环在表示学习中的影响,增强模型对全局结构信息的感知,文中提出了一种针对有向网络表示学习的优化方法。该方法借助 TrueSkill 方法获取节点的层级信息,将该信息转化为边权重并引入表示学习过程。文中将此方法应用到已有的多种有向网络表示学习方法中,多个有向网络数据集上的链接预测和节点分类任务的实验结果表明,所提方法的性能相比原有方法得到了明显提升。

关键词: 有向网络;表示学习;层级信息;链路预测

中图分类号 TP391.1

Directed Network Representation Method Based on Hierarchical Structure Information

LI Xin-chao, LI Pei-feng and ZHU Qiao-ming

School of Computer Sciences and Technology, Soochow University, Suzhou, Jiangsu 215006, China

Provincial Key Laboratory for Computer Information Processing Technology, Suzhou, Jiangsu 215006, China

Abstract Network embedding aims to embed each vertex into a low dimensional vector space and preserves certain structural relationships among the vertices in the networks. However, in the directed networks, vertexes can be reached by each other if they are in the same circle, which damages asymmetric transitivity preservation and makes representation learning model hard to capture global information of complex directed networks. This paper proposes an improved representation learning model for directed networks, which weakens the influence of circles in representation learning and enhances the ability of model to obtain global structure information. The proposed method uses TrueSkill to inference hierarchy of a directed graph and compute weight of each edge using hierarchy information. At last, this paper applies this method to some existing embedding models, and then conducts experiments on tasks of link prediction and node classification on several open source datasets. Experimental results show that the proposed method is highly scalable and effective.

Keywords Directed network, Representation learning, Hierarchical Structure, Link prediction

1 引言

随着大数据时代的到来,网络形式的数据得到越来越多研究者的关注,其通常用于表示事物之间的联系。常见的数据有社交网络、论文引用网络、网页链接网络等形式^[1]。一方面,这些信息网络包含的价值巨大,可用于个性化推荐^[2]、异常检测^[3]、节点分类^[4]和链接预测等任务^[5];另一方面,大规模的信息网络及其复杂的结构也给现有的信息挖掘技术带来了巨大挑战。在针对复杂信息网络的研究中,首先要解决的问题是网络表示方法,即如何将网络中高维度且稀疏的信息数据映射到低维连续的向量空间,从而减少处理网络数据的时间和降低空间复杂度。

近年来,基于深度学习的表示学习方法逐渐兴起,并在自然语言处理等领域取得了显著效果。其中最具代表性的 Word2Vec^[6]模型认为具有相似上下文的单词其语义也相似。受此启发,许多学者发现网络中具有相似拓补结构的节点包含的特征信息也较为相似。因此,网络表示问题可转化为词向量表示问题,可通过随机游走的策略从网络中抽取节点序列作为词向量模型的输入,从而获取节点的表示。当前,在此基础上的很多研究均取得了较好的效果,但绝大多数针对的是无向网络。在有向网络中,由于关系具有非对称性,一个节点往往具有两种角色:源端和目标端,故常用两个向量表征一个节点。因此,有向网络中边的有向性使得网络的结构和性质比无向网络更加复杂。

收稿日期:2019-12-03 返修日期:2020-04-28 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:国家自然科学基金(61836007,61772354,61773276)

This work was supported by the National Natural Science Foundation of China(61836007,61772354,61773276).

通信作者:李培峰(pfli@suda.edu.cn)

通常,基于路径的有向网络表示方法需要学习两个节点 N_1 和 N_2 间的 Rooted PageRank 值(从其中一点出发能到达另一点的概率)。影响该值的主要因素是 N_1 和 N_2 两节点间路径上的节点,而其他指向 N_1 和 N_2 的节点对该值的影响较小。如图 1 所示,由于节点 B 和 D 处于同一个环内,两点间不同方向的 Rooted PageRank 值相同,即 $B \rightarrow D$ 和 $D \rightarrow B$ 两种情况出现的概率相同。但考虑到指向 B 点的边较多且 D 点射出的边较多,可以得出 $D \rightarrow B$ 的概率大于 $B \rightarrow D$ 。例如,在微博这类社交网络中,粉丝较少的人群通常会关注粉丝较多的人群,即层级高的节点指向层级低的节点的概率更大,而反过来的概率就会很小。因此,Rooted PageRank 值在面对有向图中的环时很难挖掘到这种全局信息。

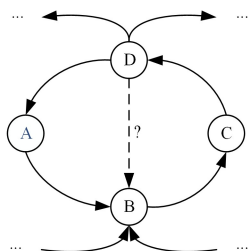


图 1 有向网络中边的预测

Fig. 1 Link prediction in directed network

由于 Rooted PageRank 值在面对环结构时并不能很好地反映环中两点间的关系,反而引入了噪声,加大了有向网络的表示学习难度。为此,本文为节点引入层级信息,根据节点层级信息获取边的权重,从而对环中的边进行差异化处理;然后,缩小违反层级顺序边的权重,从而减小环结构对模型学习全局结构信息的影响。在链接预测和节点分类任务的评测上,本文验证了所提方法在获取网络结构信息和节点属性信息上的优越性。

2 相关工作

网络表示学习是通过学习两个节点之间的拓扑近似信息实现将网络映射到低维空间。已有的表示学习算法可分为 3 类:基于矩阵分解的方法、基于深度神经网络的方法、基于随机游走并使用语言模型的方法。

节点之间的近似信息可通过邻接矩阵或拉普拉斯矩阵表示。因此,早期的网络表示学习主要是通过矩阵分解降维来获取节点向量。其中,谱聚类方法^[7]将特征分解应用于网络的拉普拉斯矩阵,以获得有用的低维表示;GraRep^[8]算法考虑了 k 步的邻接矩阵,通过奇异值分解(Singular Value Decomposition, SVD)获取 k 取不同值时网络的向量表示,最后将不同 k 值对应的网络表示拼接起来,组成维度更高、表达能力更强的节点表示;HOPE^[9]算法主要针对有向网络设计,着眼于保存原始网络中的非对称性信息和构建不同的非对称关系矩阵,最后使用 SVD 算法进行矩阵降维得到节点的网络表示。总之,这类方法较为直观,但也存在空间占用率高、计算复杂度大、无法在线学习等问题。

另一类模型使用深度学习学习非线性的网络结构信息,这类方法往往计算效率高,且具有较强的鲁棒性。其中,

SDNE^[10]算法通过结合深度自编码器和拉普拉斯矩阵来对一级和二级相似度建模;SAE^[11]法使用堆栈自编码器来重构正则化的相似度矩阵。

基于随机游走的方法使用网络中节点与节点间的共现关系来学习节点的向量表示。DeepWalk^[12]算法第一次将随机游走方法引入网络表示学习方法中,其作者在实验中验证了随机游走序列中的节点和文档中的单词一样遵从指数定律,从而将 Word2Vec 算法应用到随机游走策略上。若将 DeepWalk 视为使用深度优先遍历(Depth First Search, DFS)构造邻域,则 Line^[13]算法可以看作使用广度优先遍历(Breadth First Search, BFS)构造邻域,引入二阶相似度,并对所有的一阶相似度和二阶相似度节点对进行概率建模,最小化概率分布和经验分布之间的相对熵。Node2vec^[14]算法是将以上两种方法结合,通过超参数将 BFS 和 DFS 同时引入游走序列生成的过程中,从而同时获取节点的一阶和二阶相似度信息。

以上算法大多仅适用于无向网络,将其直接应用到有向网络中无法保留非对称传递性。相比对于无向网络表示学习的研究,对有向网络表示学习的研究较少,但近年来有了初步的探索。APD^[15]算法使用随机游走策略获取节点间非对称的高阶相似度信息,并证明了该方法实际上是隐式地保留了节点间的 Rooted PageRank。Sami 等^[16]引入对边的表示作为节点的投影矩阵,降低了节点的表示维度,更有利于学习网络的结构信息。HOPE 通过 Adamic Adar(AA),Katz Index(KI),Common Neighbors(CN) 等高阶相似性矩阵获取网络的非传递对称性。文献[17]证明了 DeepWalk 等同于邻接矩阵分解,随机游走中的采样过程对应的目标是 k 阶转移概率,可以通过改变 k 阶转移矩阵前的系数来控制采样过程,因此引入注意力机制自动学习这些系数可以达到最好的采样效果。

3 有向网络表示学习方法

本节首先介绍基于网络结构信息的基础表示学习方法,然后介绍推导节点层级信息的方法,最后给出将层级信息融入模型训练中的优化过程。

3.1 基础模型

给定有向网络 $G=(V, E)$,其中 $V=v_1, v_2, \dots, v_n$ 表示节点集合, n 是节点个数, E 是有向边集合, $e_{ij}=(v_i, v_j) \in E$ 代表从节点 i 指向节点 j 的一条边。网络表示旨在将网络中的节点映射为一个低维稠密的特征向量表示,主要分为采样和训练两个过程。采样阶段使用随机游走策略捕捉网络结构信息,从任意节点 v_i 出发,随机游走固定长度 l ,得到随机游走序列 $S=v_i, v_{i+1}, v_{i+2}, \dots, v_{i+l}$ 作为训练集;在训练阶段,对序列中的任意节点 u 按照窗口大小 b 抽取上下文节点集合,针对有向网络的特点,只使用当前节点的右窗口节点集合 $context(v_i)=v_{i+1}, v_{i+2}, \dots, v_{i+b}$ 。针对有向网络的非对称性,每个节点 v 处在源端和目标端位置时有不同的向量表示 s_v 和 t_v 。采样的节点对 (u, v) 的共现概率定义为:

$$p(v|u) = \frac{\exp(s_u \cdot s_v)}{\sum_{n \in V} \exp(s_u \cdot t_n)} \quad (1)$$

网络中的节点数量通常很多,式(1)计算量很大,因此常

采用负采样的方法提升训练效率,即为每个节点 u 随机采样出 k 个负样本,构成负样本集 $Sample(k, u)$,然后采用极大似然法定义目标函数:

$$\mathcal{L} = \sum_{u, v \in E_{\text{train}}} (\log \sigma(g(u, v)) + \sum_{v^- \in Sample(k, u)} \log(1 - \sigma(g(u, v^-)))$$
 (2)

其中, E_{train} 是基于随机游走采样的节点对, $g(u, v)$ 用于计算节点对间的余弦相似度, σ 表示 sigmoid 激活函数。

3.2 推导节点层级信息

有向网络的方向性由节点属性的差异造成。例如,在社交网络中,关注者与非关注者的影响力存在较大差异,被关注者往往具有较大的影响力,而关注者的影响力往往低于被关注者。抽象到一般的有向网络,源端节点往往处于比目标端节点更高的层级。但是,有向网络的环中的任意节点间相互可达,环结构会模糊节点间的层级次序,导致环内部的节点在向量表示上也更趋于相近,网络表示学习模型很难学到这类信息。因此,需要一种能在有向网络中快速推导节点所属层级的方法。

文献[18]在去除有向网络中环的算法中使用了网络层级结构推导,其通过 TrueSkill^[19] 和 SocialAgony^[20] 两种打分算法首先计算出每个节点的分数并推导出层级结构;然后,通过投票机制将两种打分算法和多种删边策略结合,从而去除网络中的错误边。受此启发,本文采用 TrueSkill 作为学习全局层级信息的算法,并将该信息作为先验信息引入表示学习中。TrueSkill 采用一种软打分机制,每个节点的得分并不是一个固定的值,而是服从正态分布 $N(\mu, \sigma^2)$, 这种设计具有更强的合理性和灵活性。其中, μ 表示该节点的平均分数, σ 表示节点分数的不稳定性。为了便于比较各个节点之间的层级顺序,通常将 $\mu - 3\sigma$ 作为衡量次序的标准。

在推导层级结构的过程中,网络中每个节点都被打上一个用于衡量所属层级的分数,层级越高,分数就越高。假设存在一条有向边 $e_{i,j}$, 由节点 v_i 指向节点 v_j , 通常情况下, v_i 的层级高于 v_j , 即 $score(v_i) - score(v_j) > 0$, 因此定义有向边的得分 $score(e_{i,j}) = score(v_i) - score(v_j)$, 用于衡量有向边的重要程度。当边的方向符合层次顺序时,其得分为正数,在网络中的重要性较大;反之,得分为负数,会影响网络的整体层级结构。为网络中的所有边计算分数时,可以发现每个环中至少存在一条边的得分为负数,从而可以找到网络中影响层级结构的边。

对于得分为负数的边,本文并没有使用直接删除的硬规则方式,而是采用相对灵活的加权方式来降低环对学习整体网络信息的影响。直接删除得分为负数的边存在以下缺点: 1) 在环结构较多的网络中, TrueSkill 获取层级信息的能力会受到影响,因此得分的可信度会下降,直接删除边反而会受影响模型获取全局结构信息; 2) 层级信息主要包括网络全局结构特征,而有时某些局部特征会与这类特征相冲突,例如在社交网络中,社会影响力是决定关注顺序的主要特征,但是高影响力的人也有可能因为朋友或同学等其他社交因素关注了低影响力的人,因此这些信息有助于模型学习到局部的某些特征信息。

根据 Trueskill 得到每个节点的层级分数,将该信息转变为有向边的权重,边的权重计算方式定义如下:

$$p_{i,j} = \text{sigmoid}(w * (\text{score}(e_{i,j}) + b))$$
 (3)

其中,为了防止在训练过程中出现梯度爆炸,设 w 在 0-1 之间。最后,将边权重引入表示学习的目标函数中,从而在模型中引入全局层级特征这一先验信息。改进后的目标函数定义如下:

$$\mathcal{L}_{\text{new}} = \sum_{u, v \in E_{\text{train}}} [\log \sigma(g(u, v)) * p_{i,j} + \sum_{v^- \in Sample(k, u)} \log(1 - \sigma(g(u, v^-)))]$$
 (4)

4 实验

本节主要通过链接预测和节点分类任务在多个数据集和有向网络表示学习模型上验证本文提出的改进方法的有效性。

4.1 实验设置

本文共选取 4 个公开的有向网络数据集作为实验语料,具体信息如下。

(1) Wiki-Vote: 该网络包含了自维基百科成立到 2008 年 1 月该站所有的投票数据。网络中的节点代表该网站用户,有向边 $e_{i,j}$ 表示用户 i 向用户 j 投票。

(2) Jdk-dependency: 针对 JDK 1.6.0.7 版本框架的类依赖框架,每个节点代表框架中的类,每条边代表两个类之间的调用关系。因为类依赖大多为单向引用,该网络的特点是节点和环的数量较少,层级结构相对清晰,但边数量较大。

(3) Cora-citation: 为文献引用网络,其中节点表示科技文献,两个节点间的边表示文献间的引用关系。该网络的特点是节点数量较大,但是节点间的关系较少,能获取的信息相对较少。

(4) Cit-HepPh: 为文献引用网络,用于表示 arxiv 上论文间的引用关系,但与 Cora-citation 相比,该网络规模大,关系数量多,节点间的联系更为密切。

这些语料在网络规模、网络密度及所属领域等方面有较大差别,因而能较全面地体现所提方法的有效性。具体统计数据如表 1 所列。

表 1 数据集统计

Table 1 Statistics of datasets

数据集	节点数	边数	平均度数
Wiki-Vote	7 115	103 689	14.57
Jdk-dependency	6 434	150 985	23.47
Cora-citation	23 166	91 500	3.95
Cit-HepPh	34 546	421 578	12.20

本实验将每个数据集按照 1:1 的比例随机划分成训练集和测试集,且在抽取训练集时需保持其所有的边具有连通性。构建完训练集网络后,须通过随机游走策略构建节点序列并采样节点对,本文参考 Node2Vec 中使用的采样方法,参数设置如下:每个节点作为起点游走的次数 $n=80$, 每条节点序列长度 $l=100$, 节点对采样窗口大小 $w=2$, Node2Vec 的参数 $p=1, q=1$ 。在模型训练前,预先对每个节点设置一个负例节点集合,对于每个正例节点对,从头节点的负例节点集合中随机抽取 k 个节点构成负例训练样本, k 值设为 5。模型训练

时,使用 Adam 方法作为统计优化方法,学习率设置为 0.0005,并设置超参数为 $\beta_1=0.9, \beta_2=0.999, \epsilon=1 \times 10^{-8}$ 。节点向量表示维度 $d=64$ 。

本文将提出的改进方法应用在多个有向网络表示学习模型上,并通过链接预测和节点分类实验证明所提方法具有普遍适用性。具体的有向网络表示学习模型如下。

(1)APP^[15]:一种保留有向网络的非对称结构并基于随机游走的表示学习方法。文献[15]通过理论推导证明了该嵌入方法实际上保留了 Routed PageRank 值。

(2)AsymProj(AP)^[16]:该方法显式地引入边的表示 M_r ,节点相似度计算由 $v_i \cdot v_j$ 转变为 $v_i^T M_r v_j$,从而自然地引入有向网络表示的非对称性。

(3)Watch Your Step(WYS)^[17]:在以往随机游走的过程中,步长通常作为超参数需要人为调试,该方法通过引入图注意力机制,自动学习适合当前网络的最佳步长。

4.2 链接预测

链接预测在现实生活中有很多重要应用场景,如信息挖掘、推荐系统和社交网络,因此大多数网络表示模型采用链接预测任务评估其学习网络信息的能力,并将 AUC 作为评测指标。链接预测的实验结果如表 2 所列,其中,APP, AP 和 WYS 是文献[15-17]对应的系统,APP⁺, WYS⁺ 和 AP⁺ 是在以上系统中加入了本文提出的层级信息后的系统。

表 2 链接预测结果

Table 2 Link prediction results

Datasets	APP		WYS		Asym Proj	
	APP	APP ⁺	WYS	WYS ⁺	AP	AP ⁺
Wiki_vote	86.3	93.8	93.8	95.9	91.7	93.6
Jdk-dependency	97.0	98.1	96.4	95.9	95.1	96.0
Cora-citation	85.1	89.4	—	—	80.2	85.7
Cit-HepPh	88.8	93.4	—	—	87.3	91.9

通过在不同数据集上的实验,对比 3 种表示学习模型在加入层级信息前后的性能。可以发现:1)总体来说,加入由 TrueSkill 获取的层级信息后,各个系统的链接预测性能有了较大的提升(WYS⁺ 在 Jdk-dependency 的实验除外);2)在面向 Cora-citation 这类较为稀疏的网络时,直接通过随机游走策略获取网络结构信息往往较为困难,因而在该数据集上模型表示性能弱于其他数据集,但是在改进后,APP 和 Asym-Proj 模型都有较大的性能提升,这表明在局部信息不充分的情况下,全局获取的层级信息对模型学习网络结构仍有重要作用;3)在面对复杂网络时,层级信息依然有较好的提升效果,提高了符合层级次序的边在训练中的权重,从而降低了有向网络中的环带来的噪声影响;4)在 Jdk-dependency 这类环较少的网络中,改进前模型的性能就较高,加入层级信息后性能并没有很明显的提升,这一现象从反面证明了有向网络中环的存在会在很大程度上影响表示学习模型获取网络整体结构信息,因为该类网络本身的层级结构就较为清晰,所以加入 TrueSkill 获取的层级结构信息后性能提升有限。

4.3 节点分类

本节通过节点分类任务来评估优化方法的有效性,选用 cora 和 citeseer 作为实验数据,其中 cora 数据集有 7 类节点,共包含 2708 个节点,5429 条边,citeseer 数据集有 6 类节点,

共包含 3312 个节点,4732 条边。实验中训练节点向量时,采用与链接预测相同的实验设置,得到节点向量表示后,使用神经网络对分类任务进行训练;同时采用 marco-F1 作为评测指标,通过设置不同的训练集和测试集划分比例,来全面地比较模型改进前后的性能差异。对于改进前后的 APP 模型,本文只选取节点的源端向量作为分类模型的输入。

实验结果如图 2 所示,其中纵坐标表示节点分类任务的 marco-F1,横坐标表示数据集中训练集的比例。改进后的 APP 模型的性能在两个数据集上均有提升,说明该改进方法对节点分类任务是有帮助的。数据集 cora 比 citeseer 的节点数量少,但是其网络中的边更多,结构也更加复杂。由图 2 可以发现,改进后的 APP 方法在 cora 数据集上提升更明显,表明本文提出的改进方法能充分利用有向网络结构信息,使得改进后的模型有更好的网络表示能力。

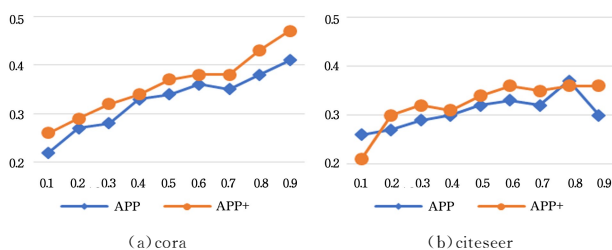


图 2 在 cora 和 citeseer 上的节点分类性能比较

Fig. 2 Performance comparison between APP and APP⁺ on cora and citeseer

结束语 本文主要提出了一种针对有向网络表示学习的优化方案。在有向网络中,使用随机游走策略获取网络表示时,学习到的是节点间的 Routed PageRank 值。但是,由于环的存在,该值通常无法正确表示环中两点间的关系,反而引入了噪声信息,并且环结构破坏了有向网络中的非对称传递性,导致对有向网络进行表示学习的难度加大。为解决这些问题,本文引入 TrueSkill 方法获取网络的层级结构信息,根据这些全局信息学习每条边的权重,从而使环内的边权重差异化,削弱环带来的负面影响。实验数据表明,改进后的网络表示学习方法与原方法相比取得了显著的性能提升。

下一步的工作将考虑以下几个方向:1)采用更优的方式将全局层级信息融入到基于随机游走策略的网络表示学习方法中;2)将节点的属性信息和网络的结构信息相融合,进一步提升网络表示的能力;3)针对动态网络,使网络表示实现在线学习。

参考文献

- [1] CAI H Y, ZHENG V W, CHANG K. A Comprehensive Survey of Graph Embedding: Problems, Techniques and Applications [J]. IEEE Transactions on Knowledge and Data Engineering, 2017, 30(9): 1616-1637.
- [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] CHANDOLA V, BANERJEE A, KUMAR V. Anomaly Detec-

- tion: A Survey[J]. *ACM Computing Surveys*, 2009, 41(3):1-72.
- [4] WANG X, CUI P, WANG J, et al. Community preserving network embedding[C]// *The 31st AAAI Conference on Artificial Intelligence*. 2017.
- [5] WEI X K, XU L CH, CAO B K, et al. Cross view link prediction by learning noise-resilient representation consensus[C]// *The 26th International Conference on World Wide Web*. International World Wide Web Conferences Steering Committee. 2017: 1611-1619.
- [6] TOMAS M, LLYA S, CHEN K, et al. Distributed representations of words and phrases and their compositionality[C]// *NIPS 2013*. Cambridge, MA: MIT Press, 2013: 3111-3119.
- [7] TANG L, LIU H. Leveraging social media networks for classification[J]. *Data Mining and Knowledge Discovery*, 2011, 23(3): 447-478.
- [8] CAO S H, LU W, XU Q K. GraRep: Learning Graph Representations with Global Structural Information[C]// *ACM International Conference on Information & Knowledge Management*. ACM, 2015.
- [9] OU M D, CUI P, PEI J, et al. Asymmetric Transitivity Preserving Graph Embedding[C]// *The 22nd ACM SIGKDD International Conference*. ACM, 2016.
- [10] WANG D, CUI P, ZHU W. Structural Deep Network Embedding[C]// *The 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*. ACM, 2016: 1225-1234.
- [11] TIAN F, GAO B, CUI Q, et al. Learning deep representations for graph clustering[C]// *The Twenty-Eighth AAAI Conference on Artificial Intelligence*. 2014.
- [12] 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*. ACM, 2014: 701-710.
- [13] TANG J, QU M, WANG M, et al. Line: Large-scale information network embedding[C]// *The 24th International Conference on World Wide Web*. International World Wide Web Conferences Steering Committee. 2015: 1067-1077.
- [14] GROVER A, LESKOVEC J. node2vec: Scalable Feature Learning for Networks[C]// *The 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*. ACM, 2016: 855-864.
- [15] ZHOU C, LIU Y, LIU X, et al. Scalable graph embedding for asymmetric proximity[C]// *The 31st AAAI Conference on Artificial Intelligence*. 2017.
- [16] ABU-EL-HAIJA S, PEROZZI B, AL-RFOU R. Learning edge representations via low-rank asymmetric projections[C]// *The 2017 ACM Conference on Information and Knowledge Management*. ACM, 2017: 1787-1796.
- [17] ABU-EL-HAIJA S, PEROZZI B, AL-RFOU R, et al. Watch Your Step: Learning Node Embeddings via Graph Attention[C]// *Neural Information Processing Systems (NIPS)*. 2018: 9180-9190.
- [18] SUN J, AJWANI D, NICHOLSON P K, et al. Breaking Cycles in Noisy Hierarchies[C]// *the 2017 ACM on Web Science Conference*. ACM, 2017: 151-160.
- [19] HERBRICH R, MINKA T, GRAEPEL T. TrueSkill: A Bayesian Skill Rating System[C]// *Neural Information Processing Systems (NIPS)*. 2007.
- [20] GUPTE M, SHANKAR P, LI J, et al. Finding Hierarchy in Directed Online Social Networks[C]// *The 20th International Conference on World Wide Web (WWW 2011)*. Hyderabad, India, DBLP, 2011.



LI Xin-chao, born in 1995, postgraduate. His main research interests include natural language processing and representation learning.



LI Pei-feng, born in 1971, Ph.D, professor, Ph.D supervisor, is a member of China Computer Federation. His main research interests include natural language processing and machine learning.