基于城市轨道交通网络的广告营销站点选择

李 哲

(中铁第四勘察设计院集团有限公司, 武汉 430063)

摘要:城市轨道交通网络背景下的广告营销已成为一种流行的产品推广方式,能够快速、准确筛选出传播范围广、影响力高的站点,运营方和投资商对其投入大量广告可以产生较高的宣传力度和回报收益。影响力最大化为解决站点筛选提供了一种新的思路。首先,提出一种融合节点结构信息和网络拓扑结构信息的网络表示学习模型,并将其用于学习网络节点的特征向量;然后,将网络节点的特征向量输入聚类算法中有效筛选候选种子集,同时结合贪婪策略从候选种子集中筛选出种子节点集合,进而有效提升网络节点影响力最大化问题求解效果。最后,以武汉地铁网络为例进行验证。结果表明所提方法在传播范围上优于现有的影响力最大化方法,筛选高影响力站点具备合理性。

关键词:城市轨道交通;地铁广告营销;影响力最大化;网络表示学习;神经网络;聚类

中图分类号:TP391 **文献标志码**:A **文章编号**:1671-1807(2023)10-0245-08

近年来,城市轨道交通发展迅速,已在城市公共交通中占主导地位。其中,地铁作为一种速度快、运量大的轨道交通,受众相较于其他交通出行方式要高。另外,地铁内部是一种半封闭的空间,具有一定封闭性和规律性,其传播媒介具有传播范围广的特征。但是,由于地铁前期投入成本高、运营时间长、票价比较稳定,只依靠稳定的票价收入很难维持地铁运营方经营。目前,仅有部分城市地铁运营方能够实现盈利。因此,采用地铁广告收入反哺地铁运营是一种解决方案,对地铁投入相应的广告[1]可以产生较高的回报。同时,为了便于地铁运营方节约成本、增大收益,以利益最大化为目的,如何筛选出影响力大、传播范围广的地铁站,对其投入大量广告,而对普通地铁站投入少量广告?针对此问题,将地铁站筛选问题抽象为影响力最大化问题,其概念最早由 Domingos 和 Richardson[2]提出,旨在选择网络中最具影响力的节点,并确保这类节点的数量尽量少。

Kempe 等[3]将影响力最大化定义为一个优化问题,并且利用独立级联(IC)和线性阈值(LT)模型证明了影响力最大化是 NP 难问题。然后,以寻找到最佳种子集为目的,提出了一个近似因子为$(r-1/e)$的贪婪算法。虽然贪婪算法的准确性优于传统方法,但它仍然面临着大规模网络的可扩展性问题。目前,大量研究解决了贪婪算法的可扩展性问题,Leskovec 等[4]提出了 CELF(cost-effetive lazy forward selection)算法,该算法利用子模性对贪婪算法做了改进,通过减少蒙特卡罗模拟的次数,降低每个节点影响力传播的运行时间,评估表明 CELF 算法比贪婪算法快 700 倍。随后,Goyal 等[5]提出了 CELF++算法,该算法通过避免不必要的蒙特卡洛模拟来减少 CELF 算法中影响力的计算次数。为了减少了影响力传播的计算开销,研究者提出了不同的启发式算法,例如 Degree、PageRank 或其他中心性度量方法,这些启发式算法考虑了网络和传播模型中节点的特性。Chen 等[6]在度的基础上提出了 Degree Discount 算法,该算法的思想是在选择一个节点的邻居节点作为种子节点时,要对该节点的度进行折扣。然而,传统的影响力最大化问题方案都是采用不同的网络特性开展研究,因此大多数方法是不可扩展的。鉴于此,引入了网络表示学习(又称图嵌入)算法[7]。此算法将网络中的节点表示成低维、实值、稠密的向量,使得到的向量可以在向量空间中具有表示以及推理的能力,同时也保留了网络中节点的局部特征和全局特征,继而可以将得到的特征向量运用到复杂网络的常见应用中。

综上所述,主要针对地铁网络的影响力最大化问题展开研究。首先,通过网络表示学习方法对网络进行降维,以得到节点的特征向量。其次,将节点的特征向量进行聚类,得到聚类中心之后,使用其相对应

收稿日期:2022-12-26

作者简介:李哲(1996—),男,甘肃平凉人,中铁第四勘察设计院集团有限公司,助理工程师,硕士,研究方向为城市轨道通信信号设计、城轨云、地铁复杂网路等。

的节点作为候选种子节点集。最后，从候选种子节点集中利用贪婪策略选择最具影响力的节点，并将这类节点作为广告营销的重要站点，对其投入大量的广告，以达到回报收益最大化的目的。至此，提出一种基于地铁网络表示学习的影响力最大化模型，并以武汉地铁为例进行分析和验证，以期为城市轨道交通网络的广告营销站点选择提供借鉴。

1　基本概念

1.1　网络定义

给定一个网络 $G=(V,E)$，具有顶点集 V，边集 E，其中，用 $a_{xy}=1$ 表示节点 v_x 和 v_y 之间存在边。另外，在顶点嵌入中的任务是学习映射函数 $f:R^{|V|}\rightarrow R^d(d\leqslant|V|)$，其中 d 是向量空间的选定维数。

1.2　基于网络表示学习的影响力最大化策略

以地铁无向网络为研究对象，并且使用随机游走策略进行节点采样。一般的随机游走是对图中的一层进行一次遍历，在每一次遍历过程中，都以当前节点为起始节点进行一次随机游走采样。然后，将随机游走采样过程中所获得的节点序列输入到网络表示学习模型中进行相应的训练、学习。图 1 显示了基于网络表示学习的影响力最大化框架。首先，以游走序列（v_1,v_2,v_3,v_4,v_5）为例，先从节点 v_1 开始随机游走选择节点 v_3，接下来随机游走选择了节点 v_2，最后随机游走选择了节点 v_4。将随机游走形成的节点序列合并到一个语料库中。然后，将语料库输入嵌入模型中进行训练，得到每个节点的特征向量，并通过聚类算法将学习到的特征向量进行聚类，聚类后的中心向量相对应的节点将作为候选种子节点。最后，采用贪婪策略从候选种子节点集中筛选出种子节点，并在传播模型进行影响力传播。

1.3　引导随机游走的节点指标

在网络表示学习任务中，如果要使嵌入模型学习到丰富的网络信息，则输入嵌入模型中的节点序列需具有较高的相似性，且也要具有较高的信息传递性。因此，针对网络表示学习模型采样过程而言，应该优先选择与当前节点相似度和传递性较高的节点作为待采样节点。利用上述有引导的采样方式所产生的节点序列能充分地反映网络原始结构信息。

采用共同邻居指标来刻画节点之间的相似度，该方法是一种相对简单，而且效果比较好的相似度描述方法，计算公式为

$$\mathrm{CN}(i,j)=A_i\cap B_j \tag{1}$$

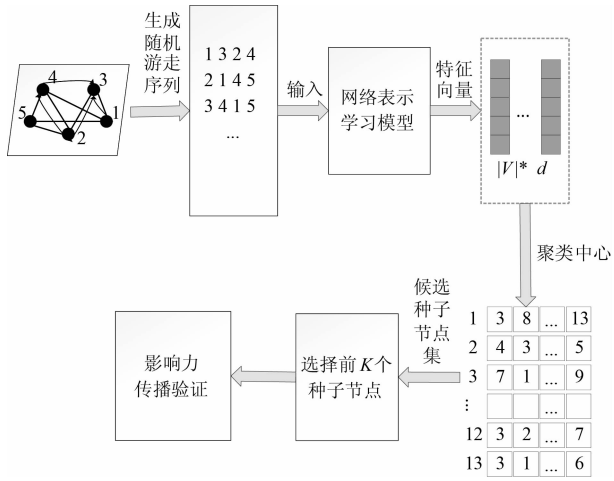

图 1　基于网络表示学习的影响力最大化框架

式中：$\mathrm{CN}(i,j)$ 表示节点 v_i 和节点 v_j 的共同邻居数，A_i 表示节点 v_i 的邻居节点集合，B_j 表示节点 v_j 的邻居节点集合。

采用集聚系数来描述待采样节点的传递性，集聚系数是用来描述一个图中的顶点之间集结成团程度的系数，计算公式为

$$\mathrm{CL}_j=\frac{2\{e_{pq}\mid p,q\in\Upsilon_j\}}{k_j(k_j-1)} \tag{2}$$

式中：CL_j 为节点 v_j 的集聚系数；Υ_j 为节点 v_j 的邻居节点集合；k_j 为节点 v_j 的度；e_{pq} 为节点 v_p 和 v_q 的连边。

总之，集聚系数是某节点的邻居节点之间相互连接的程度，利用 ClusterRank[8] 方法思想，节点的局部集聚系数与其传递能力一般呈负相关关系，即局部集聚系数越小，该节点传递能力越强。

结合共同邻居数和集聚系数两种方法，给出引导指标 I，计算公式为

$$I=f(\mathrm{CL}_j)[\mathrm{CN}(i,j)+1] \tag{3}$$

式中：CL_j 为式(2)中集聚系数；f 为集聚系数 CL_j 的线性负相关函数，f 一般取 $10^{-\mathrm{C}_j}$，CL_j 越大 f 越小；$\mathrm{CN}(i,j)$ 为式(1)中的共同邻居数。

2　基于网络表示学习的影响力最大化模型

2.1　改进的随机游走

结合 Node2Vec[9] 算法中引入的拓扑结构信息和 1.3 节所提出的节点结构信息在网络上进行随机游走，有

$$\delta(v_z,v_x)=\begin{cases}\dfrac{I}{p}, & d_{v_z,v_x}=0\\ I, & d_{v_z,v_x}=1\\ \dfrac{I}{q}, & d_{v_z,v_x}=2\end{cases} \tag{4}$$

式中:$\delta(v_z,v_x)$ 为引入节点结构信息指标 C_r 后的转移概率;I 为节点结构信息指标;v_z 为上一次访问过的节点;v_x 为下一次要访问的节点;d_{v_z,v_x} 为节点 v_z 和节点 v_x 的最短距离;v_y 为当前节点;参数 p 控制重复访问刚刚访问过的节点的概率。p 只在 $d_{v_z,v_x}=0$ 时起作用,$d_{v_z,v_x}=0$ 表示下一次要访问的节点正好是之前刚刚访问过的节点。因此,如果 p 较大,则访问刚刚访问过的节点的概率会变低,反之亦然。$d_{v_z,v_x}=1$ 表示下一次要访问的节点正好与之前刚刚访问过的节点相连。参数 q 则控制游走的方向是向外还是向内,q 只在 $d_{v_z,v_x}=2$ 时起作用,$d_{v_z,v_x}=2$ 表示下一次要访问的节点既不访问刚刚访问过的节点,也不与刚刚访问过的节点相连。若 q 值较大,则随机游走倾向于访问和 v_z 接近的顶点,游走路径偏向于广度优先搜索;反之,如果 q 较小,则随机游走过程倾向于访问远离 v_z 的顶点,游走路径偏向于深度优先搜索。

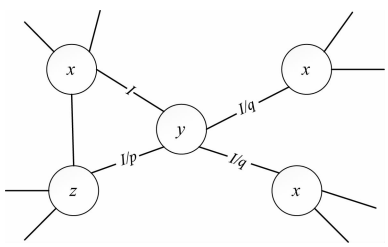

图 2　随机游走模型

2.2　Skip-Gram 模型

Skip-Gram 模型是 Word2Vec 中为了搞清楚单词与单词之间关系的一种模型。简而言之,它就是根据输入的一个单词来预测周围的上下文。如图 3 所示,Skip-Gram 模型是一个简单三层神经网络。

如果向 Skip-Gram 模型中输入的是一个特定的单词,则输出的就是 Softmax 概率排名靠前的几

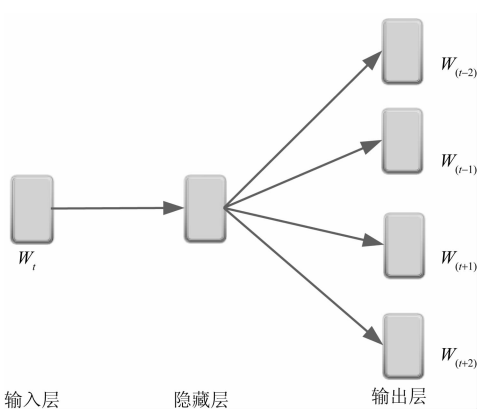

图 3　Skip-Gram 模型

个单词[10]。节点的游走序列类似于自然语言处理领域的文本语句,同时,节点类似于文本语句中的单词。因此,可以直接训练 Skip-Gram 模型得到节点的向量表示。如果给出一个单词序列,该序列为 $W=(w_0,w_1,\cdots,w_n)$,在训练要最大化概率 $Pr(w_0,w_1,\cdots,w_{n-1}\mid w_n)$。同样,给出一个随机游走序列 walk $=(v_0,v_1,\cdots,v_i)$,在训练时要最大化概率 $Pr(v_0,v_1,\cdots,v_{i-1}\mid v_i)$。如图 3 所示,假设滑动窗口 w 为 2,每次的训练语料为 $W_{(t-2)}$、$W_{(t-1)}$、$W_{(t+1)}$、$W_{(t+2)}$,Skip-Gram 模型的输入层为 $W_{(t)}$ 的 One-hot 编码向量,输出层为已知 $W_{(t)}$ 的条件下,计算节点序列中前后节点 $W_{(t-2)}$、$W_{(t-1)}$、$W_{(t+1)}$、$W_{(t+2)}$ 出现的概率。最终目标是学习一个节点的向量表示 ζ,优化函数 $J(\zeta)$ 的计算公式为

$$J(\zeta)=\max_{\zeta}\sum_{v\in l}\log_2 Pr[(v_{i-n},\cdots,v_{i-1},v_{i+1},\cdots,v_{i+n})\mid\zeta(v_i)] \tag{5}$$

式中:$\zeta(v_i)\in R^d$ 为节点 v_i 的向量表示;n 为滑动窗口,表示每次训练时选择当前节点的前后 n 个邻居节点。条件独立假设概率的计算公式为

$$J(\zeta)=\max_{\zeta}\sum_{v\in l}\sum_{-n\leqslant j\leqslant n,j\neq 0}\log_2 Pr[v_{i+j}\mid\zeta(v_i)] \tag{6}$$

条件概率 $Pr[v_j\mid\zeta(v_i)]$ 的计算公式为

$$Pr[v_j\mid\zeta(v_i)]=\frac{\exp[\zeta'(v_j)\zeta(v_i)]}{\sum_{v\in V}\exp[\zeta'(v)\zeta(v_i)]} \tag{7}$$

式中:$\zeta'(v)\in R^d$ 为节点 v 的上下文节点的向量表示。选用随机梯度下降法[11]更新模型参数,更新式为

$$\zeta=\zeta-\alpha\frac{\partial J(\zeta)}{\partial\zeta} \tag{8}$$

式中:α 为步长,表示随机梯度下降的速率,α 越大下降速度越快,反之亦然。

2.3　种子集选择

上述几节介绍了如何通过网络表示学习模型学习节点的特征向量。为了筛选出种子节点集,首先采用 K-means[12]算法对网络中的节点特征向量进行聚类,这样能得到相应数量的聚类中心,并将其作为候选种子节点集合。但通过实验得出,采用 K-means 算法得到的聚类中心一般是不存在的节点,也就是说找到的大部分节点不是数据集中对应的节点,这是因为 K-means 算法在计算聚类中心时只考虑最小化目标函数,而忽略了它是否真实存在。由于在影响力最大化问题研究中节点必须是真实存在的,因此选择欧氏距离来计算与聚类中心

相距最近的节点，该类节点作为候选种子节点，其计算公式为

$$d(\boldsymbol{A},\boldsymbol{B}) = \sqrt{\sum_{i=1}^{n}(\boldsymbol{A}_i - \boldsymbol{B}_i)^2} \tag{9}$$

式中：d 为欧氏距离；$\boldsymbol{A}$ 和 $\boldsymbol{B}$ 分别为节点特征向量；$\boldsymbol{A}_i$ 和 $\boldsymbol{B}_i$ 表示向量 $\boldsymbol{A}$ 和 $\boldsymbol{B}$ 的各分量。

接下来，为了提高算法的效率，引入了 KD 树[13]算法产生候选种子节点集。KD 树算法的核心思想是划分 K 维空间。在划分的过程中，期望两个子空间中被超平面划分的数据量是相同的。KD 树算法可以分成两个阶段，第一个阶段是构建 KD 树，第二阶段是在 KD 树中进行搜索。

最后，当候选种子节点集筛选完成后，要从中选出影响力最大的节点作为种子节点，所以使用改进后的贪婪算法 CELF 进行种子节点筛选，该算法充分利用了子模性，这种特性简单来说就是节点的边际影响力呈递减的趋势，因此节省了大量的计算时间，提高了算法的效率。

3　实例分析

3.1　数据集

以武汉地铁为实例进行验证。截至目前，武汉轨道交通已建成运营轨道交通 1 号线、2 号线一期、机场线、2 号线南延线、4 号线、3 号线、5 号线、6 号线一期、6 号线二期、7 号线、8 号线一期、8 号线二期、8 号线三期、阳逻线、11 号线东段（光谷火车站－左岭）、11 号线三期葛店段、16 号线、纸坊线、蔡甸线，总运营里程达 435 km，车站总数达 282 座。国家发改委所批复的第四期建设方案中明确表示，到 2024 年需建成 14 条运营线路，其总长为 606 km 的交通网，完全形成“主城成网、新城通线”。武汉地铁交通网络如图 4 所示。

因为地铁车辆拥有成对运行的特征，且不考虑

图 4　武汉地铁网络

每个路段的方向、运载能力及线路等级,所以将武汉地铁网络构建为无向无权网络。考虑到武汉地铁网络体系模式较多、规模庞大,因此在研究过程中,采用 Space-L 方法构建网络的拓扑结构,该方法的主要的思想为:首先对网络中节点先编号,然后去除重复节点,最后根据从左到右、从上到下的原则确定节点之间的连边。为确保实验结果的可靠性,实验数据从武汉轨道交通集团官方网站(http://wh.bendibao.com/ditie/linemap.shtml)提取了武汉地铁 1～8 号线、11 号线、16 号线、阳逻线等多条线路信息,整合处理后得到 248 个地铁站点和 268 条运营线路,并以此为基础,利用复杂网络分析工具 Gephi 中生成武汉地铁复杂网络拓扑模型,如图 5 所示。

如表 1 所示,采用 Space-L 方法对武汉地铁站点进行编号,共计 248 个站点,节点编号为 1、2、…、248。

如表 2 所示,对武汉地铁网络进行特性计算,分析后得到,网络的平均度为 2.161,度小于 2.161 以下的节点数占整个网络的 70%以上,说明网络中重要性节点占少数;网络的平均加权度为 2.161,由于武汉地铁网络定义为无权无向网络,所以其含义与

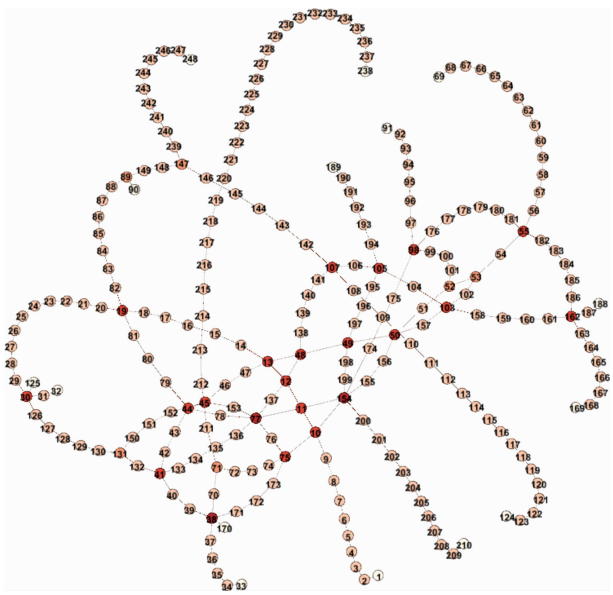

图 5　武汉地铁网络拓扑图

表 1　武汉地铁站点编号

编号	站点名称
1	汉口北
2	滠口新城
…	…
247	湾湖
248	周家河

表 2　武汉地铁网络特性

网络特征指标	计算值
平均度	2.161
平均加权度	2.161
网络直径	58
平均集聚系数	0.004
平均路径长度	17.154

平均度相同;网络的直径为 58,平均路径为 17.154,约有 44%节点的最短路径大于平均距离,说明网络中节点之间的连接不紧密,网络较稀疏;平均集聚系数为 0.004,说明网络的连通性不高,相对不紧密。

3.2　对比算法

CELF[13]:该算法利用影响力函数的子模性对贪婪算法做了改进,通过减少蒙特卡罗模拟的次数,降低每个节点影响力传播的运行时间。

DC(Degree Centrality)[14]:该算法首先计算网络中每个节点的度值,然后根据其大小对节点进行排序,最后在排序后的节点中选择前 K 节点作为该网络的种子节点。

BC(Betweenness Centrality)[15]:该算法计算经过一个节点的最短路径的数量。经过一个节点的最短路径的数量越多,就说明它的 BC 值越高。

CC(Closeness Centrality)[16]:该算法计算出一个节点到所有其他可到达节点的最短距离的倒数,并将归一化值累加起来。一个节点的 CC 值越大越接近图的中心。

DD(Degree Discount)[6]:该算法的思想是在选择一个节点的邻居节点作为种子节点时,要对该节点的度进行折扣。

PR(Page Rank)[17]:该算法认为在影响力最大化问题中,一个节点其邻居的数量和指向其邻居的重要性决定了该节点的重要性。

3.3　实验设置

对于学术界而言,传播模型广泛用于评估节点影响力最大化。因此,遵循类似任务中常用的评估标准,使用 IC 模型[3](independent cascade model)做验证,根据传播实验得到的传播范围最优的前 K 个节点,作为广告营销大力投入广告的站点。根据前文实验分析得出,在设置候选种子集数量时比种子集多 20 个,使用蒙特卡罗模拟了 1 000 次实验。对于提出的网络表示学习模型,设置嵌入维数设置为 200,窗口宽度设置为 10,游走长度设置为

10，迭代次数设置为 10，p 和 q 分别设置为 2、0.5，然后选择 5 个负采样。

实验硬件环境为：Intel（R） Core（TM） i7 — 7 700 CPU 3.60 GHz，8 G 内存。

实验软件环境为：Windows10 系统，使用 Python 3.6版本。

3.4　实验结果分析

图 6 和图 7 中的影响力范围是指种子集在网络中传播影响的节点个数。种子集的影响范围越大，说明算法性能越好，继而对该类节点投入大量广告能带来广泛传播宣传，获得较高的回报收益。

文中所提出的网络表示学习模型与其他模型之间的主要区别是引入了节点结构信息，且所提出模型能捕获网络中节点的相似性和传递性特征。实验结果表明，上述特征引入的重要性，如图 6 所示，当传播概率取 0.1 时，在武汉地铁数据集中，所提方法 Our_Method 的传播范围优于 CELF、DD、BC、CC、DC、PR 等算法，且 Our_Method 相比于 CELF、DD、BC、CC、DC、PR 等算法，分别提高了 1.2%、44.8%、38.2%、42.3%、33.1%、15.6%。从结果分析得出，所提方法 Our_Method 表现出了较好的性能，同时在筛选影响力较高的站点时具有高效性，且这类站点的传播性能较强，所以通过对该类站点投入较多广告能获得相对较高收益，以达到投资方的利益最大化。

如图 7 所示，当传播概率取 0.05 时，在武汉地铁数据集中，所提方法 Our_Method 的传播范围仍优于 CELF、DD、BC、CC、DC、PR 等算法，分析得出所提方法 Our_Method 对站点选择的扩展性较强，适用于不同传播率。

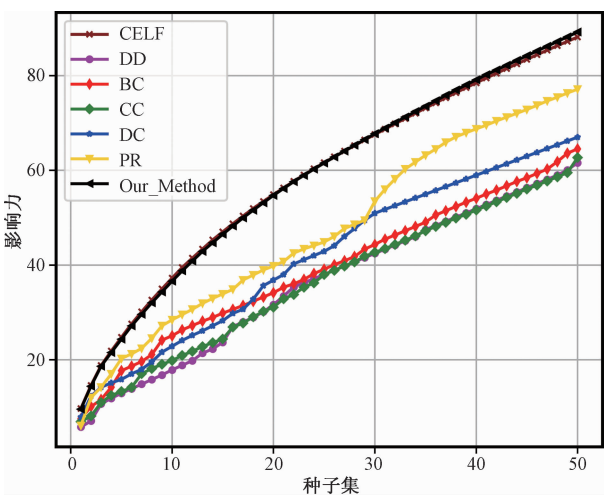

图 6　p=0.1 时在武汉地铁网络下的对比实验

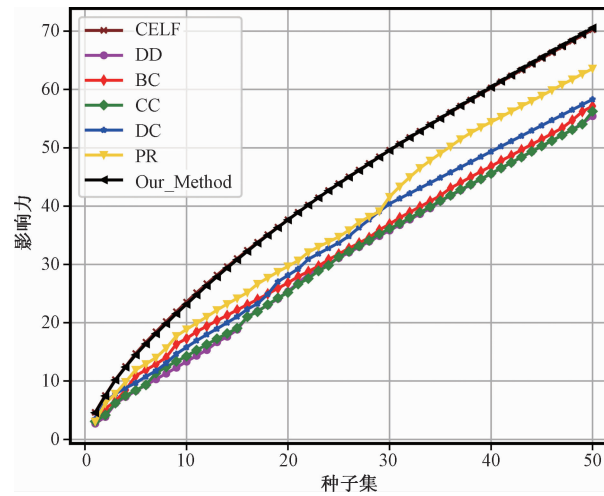

图 7　p=0.05 时在武汉地铁网络下的对比实验

如表 3 所示，当 p=0.1 时，各对比方法下的武汉地铁网络前 50 站点中部分站点传播范围。分析得出，Our_Method 方法筛选出的站点传播范围比其他方法的传播范围都高，其筛选出的站点为司门口黄鹤楼、湖北大学、工人村、国博中心南、头道街、广埠屯、铁机路、光谷五路、汉口火车站、园林路。以传播范围最大的站点为例，Our_Method 方法筛选出的站点传播范围为 89.215，比 CELF、DD、BC、CC、DC、PR 方法高出 1.059、27.606、24.67、26.51、22.2、12.072。

如表 4 所示，当 p=0.1 时，所提方法 Our_Method 筛选武汉地铁网络前 50 站点传播范围。分析得出，大智路站、王家墩东站、光霞站、武汉站东广场、野芷湖站等站点为换乘站，与其他普通站相比，具备较大客流量，对这类站点投入大量广告会产生较大的宣传效果，进而得到的较高回报收益；武汉站东广场、汉口火车站等交通枢纽站也具有较大客流量，对其投入大量广告，也能产生较大的回报收益；长岭山站、昙华林武胜门站、新庙村站等传播范围增长速度较快的站点，由于本身具备较强的传播性，对其投入大量广告，也能产生较大的回报收益。通过上述实验结果分析，证明所提方法 Our_Method 具备合理性，对广告营销站点选择具有指导性。

4　结论

以城市轨道交通网络的广告营销站点选择为方向进行了研究分析，尝试将广告营销站点选择问题转化为影响力最大化问题，综合考虑网络中节点结构和网络拓扑结构信息对节点影响力最大化的影

表 3　$p=0.1$ 时，各对比方法下的武汉地铁网络前 50 站点中部分站点传播范围

Our_Method		CELF		DD		BC		CC		DC		PR	
编号	传播范围	编号	传播范围	编号	传播范围	编号	传播范围	编号	传播范围	编号	传播范围	编号	传播范围
196	80.215	32	79.511	51	52.779	155	55.006	102	52.491	17	59.788	67	69.561
155	81.215	156	80.508	57	53.665	156	55.824	195	53.37	18	60.591	208	70.441
207	82.215	170	81.491	59	54.551	109	56.734	43	54.249	20	61.394	3	71.244
146	83.215	125	82.471	61	55.437	223	57.644	140	55.149	21	62.197	167	72.047
9	84.215	41	83.446	63	56.323	51	58.447	201	56.059	22	63.000	93	72.850
56	85.215	44	84.421	65	57.209	110	59.357	53	56.862	23	63.803	191	73.750
97	86.215	195	85.376	67	58.095	224	60.267	73	57.748	25	64.606	31	74.660
231	87.215	139	86.331	73	58.981	38	61.834	172	58.634	26	65.409	35	75.463
43	88.215	149	87.246	80	60.081	158	63.635	135	59.513	27	66.212	235	76.340
95	89.215	90	88.156	83	61.609	56	64.545	152	62.705	28	67.015	121	77.143

表 4　$p=0.1$ 时，采用 Our_Method 方法筛选武汉地铁网络前 50 站点传播范围

编号	站点	传播范围	编号	站点	传播范围
128	海口三路	9.599	12	大智路	62.824
235	长岭山	14.422	138	六渡桥	64.070
197	昙华林武胜门	18.660	85	四新大道	65.298
123	新庙村	21.538	171	塔子湖	66.523
213	幸福湾	24.364	183	文治街	67.745
191	光霞	27.164	45	王家墩东	68.920
245	马影河	29.629	30	码头潭公园	70.092
186	马湖	32.047	162	野芷湖	71.264
178	中南医院	34.341	159	建安街	72.416
210	武汉站东广场	36.601	72	兴业路	73.564
169	青龙山地铁小镇	38.788	134	石桥	74.710
66	金融港北	40.858	79	云飞路	75.856
77	香港路	42.892	188	军运村	76.966
2	滠口新城	44.722	156	新河街	78.063
88	体育中心	46.513	24	古田一路	79.145
81	双墩	48.254	196	司门口黄鹤楼	80.215
92	杨春湖	49.973	155	湖北大学	81.215
34	航空总部	51.59	207	工人村	82.215
203	建设二路	53.164	146	国博中心南	83.215
150	园博园	54.718	9	头道街	84.215
75	赵家条	56.192	56	广埠屯	85.215
144	前进村	57.608	97	铁机路	86.215
199	三角路	58.96	231	光谷五路	87.215
136	三眼桥	60.266	43	汉口火车站	88.215
241	小军山	61.562	95	园林路	89.215

响，提出了结合节点结构信息和网络拓扑结构信息的指标。进一步提出具备多元信息的网络表示学习影响力最大化方法，使最终学习到的节点特征向量能准确反映其网络原始结构信息，同时使节点的影响力传播范围更广。以武汉地铁为实例进行了验证，结果表明所提方法比传统的影响力最大化方法筛选出的站点传播范围广、客流量大，具备广告投入的合理性。因此，地铁运营方和投资商可以采用所提方法筛选出传播范围前 K 的站点，对其投入大量广告，以达到回报收益最大化的目的。

参考文献

[1] 赵嘉祺，杨松. 地铁广告营销的一些思考[J]. 现代商贸工业，2017(27)：61-62.

[2] DOMINGOS P，RICHARDSON M. Mining the network value of customers[C]//Proceedings of the Seventh ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. San Francisco，California：Association for Computing Machinery，2001：57-66.

[3] KEMPE D，KLEINBERG J，TARDOS É. Maximizing the spread of influence through a social network[C]//Proceedings of the Ninth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. Washington，D. C.：Association for Computing Machinery，2003：137-146.

[4] LESKOVEC J，KRAUSE A，GUESTRIN C，et al. Cost-effective outbreak detection in networks[C]//Proceedings of the 13th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. San Jose，California：Association for Computing Machinery，2007：420-429.

[5] GOYAL A，LU W，LAKSHMANAN L V S. CELF++：optimizing the greedy algorithm for influence maximization in social networks[C]//Proceedings of the 20th International Conference Companion on World Wide Web. Hyderabad，India：Association for Computing Machinery，2011：47-48.

[6] CHEN W，WANG Y，YANG S. Efficient influence maximization in social networks[C]//Proceedings of the 15th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. Paris，France：Association for Computing Machinery，2009：199-208.

[7] 涂存超，杨成，刘知远，等. 网络表示学习综述[J]. 中国科学：信息科学，2017，47(8)：980-996.

[8] MALLIAROS F D，VAZIRGIANNIS M. Clustering and community detection in directed networks：a survey[J]. Physics Reports，2013，533(4)：95-142.

[9] GROVER A，LESKOVEC J. Node2vec：scalable feature learning for networks[C]//San Francisco：Association for Computing Machinery，2016：855-864.

[10] BROWN P F, PIETRA V J D, SOUZA P V D, et al. Class-based n-gram models of natural language[J]. Computational Lingus, 1992, 18(4): 467-479.

[11] ROBBINS H, MONRO S. A stochastic approximation method[J]. The Annals of Mathematical Statistics, 1951, 22(3): 400-407.

[12] HARTIGAN J A, WONG M A. Algorithm AS 136: a *k*-means clustering algorithm[J]. Journal of the Royal Statistical Society. Series C (Applied Statistics). 1979, 28(1): 100-108.

[13] BENTLEY J L. Multidimensional binary search trees used for associative searching[J]. Communications of the ACM, 1975, 18(9): 509-517.

[14] YOU K, TEMPO R, QIU L. Distributed algorithms for computation of centrality measures in complex networks [J]. IEEE Transactions on Automatic Control, 2016, 62 (5): 2080-2094.

[15] FREEMAN L C. A set of measures of centrality based on betweenness[J]. Sociometry, 1977: 35-41.

[16] COHEN E, DELLING D, PAJOR T, et al. Computing classic closeness centrality, at scale[C]//Proceedings of the Second ACM Conference on Online Social Networks, 2014: 37-50.

[17] LUO Z, CAI W, LI Y, et al. A pagerank-based heuristic algorithm for influence maximization in the social network[M]. Berlin: Springer, 2012: 485-490.

Research on the Selection of Advertising Marketing Stations Based on Urban Rail Transit Network

LI Zhe

(China Railway Siyuan Survey and Design Group Co. Ltd., Wuhan 430063, China)

Abstract: Advertising marketing in the context of urban rail transit network has become a popular way of product promotion. With quickly and accurately screening out stations with a wide range of communication and high influence, operators and investors can invest a large amount of advertising in them to generate higher publicity and returns. The proposal of influence maximization provides a new way to solve the problem of station selection. Firstly, a network representation learning model was proposed, which integrates node structure information and network topology structure information, and was used to learn the feature vectors of network nodes. Further, the feature vector of network nodes was input into the clustering algorithm to effectively screen the candidate seed set, and at the same time, the greedy strategy was combined to screen the seed node set from the candidate seed set, so as to effectively improve the effect of the problem of maximizing the influence of network nodes. Taking Wuhan metro network as an example, the results show that the proposed method is superior to the existing influence maximization method in terms of its spread scope, and it is reasonable to select high-influence stations.

Keywords: urban rail transit; subway advertising marketing; influence maximization; network representation learning; neural network; clustering