计算机科学
COMPUTER SCIENCE

融合多特征的属性异质网络嵌入方法

汤启友, 张凤荔, 王瑞锦, 王雪婷, 周志远, 韩英军

引用本文

汤启友, 张凤荔, 王瑞锦, 王雪婷, 周志远, 韩英军融合多特征的属性异质网络嵌入方法[J]. 计算机科学, 2022, 49(12): 146-154.

TANG Qi-you, ZHANG Feng-li, WANG Rui-jin, WANG Xue-ting, ZHOU Zhi-yuan, HAN Ying-jun. Method of Attributed Heterogeneous Network Embedding with Multiple Features [J]. Computer Science, 2022, 49(12): 146-154.

计算机科学 COMPUTER SCIENCE
http://www.jsjkx.com
DOI:10.11896/jsjkx.211200082

融合多特征的属性异质网络嵌入方法

汤启友　张凤荔　王瑞锦　王雪婷　周志远　韩英军
电子科技大学信息与软件工程学院　成都 610054
(tangqiyou2018@163.com)

摘　要　网络嵌入旨在用低维、实值的向量表示非结构化网络中的节点,使节点嵌入尽可能地保留原始网络中的结构特征与属性特征。然而,当前研究主要集中于嵌入网络结构,对异质信息网络中具有丰富语义的关系属性和节点属性考虑得较少,可能导致节点嵌入语义缺失,从而影响下游应用的预测效果。针对该问题,设计了一种融合多特征的属性异质网络嵌入(Attributed Heterogeneous Network Embedding with Multiple Features,MFAHNE)方法。该方法通过序列采样、结构特征嵌入、属性特征嵌入、特征融合等步骤将网络中的关系属性、节点属性、结构语义等特征融合至最终节点嵌入。实验结果表明,该方法能兼顾结构特征与属性特征,实现两种特征信息的相互补充,优于传统的网络嵌入方法。

关键词:网络嵌入;异质信息网络;结构特征;属性特征;属性异质网络
中图法分类号　TP181

Method of Attributed Heterogeneous Network Embedding with Multiple Features

TANG Qi-you,ZHANG Feng-li,WANG Rui-jin,WANG Xue-ting,ZHOU Zhi-yuan and HAN Ying-jun
School of Information and Software Engineering,University of Electronic Science and Technology of China,Chengdu 610054,China

Abstract　Network embedding aims to represent nodes in unstructured network with low-dimensional,real-valued vectors,so that node embedding can retain the structural and attribute features of the original network as much as possible. However,current research mainly focuses on embedding the network structure. There are few researches considering relationship attributes and node attributes with rich semantics in heterogeneous information networks,which may result in semetic loss of node embedding and affect the prediction effect of downstream applications. To solve this problem,this paper designs a method of attributed heterogeneous network embedding with multiple features(MFAHNE). This method integrates the relationship attributes,node attributes and structural semantics in the network into the final node embedding through the steps of sampling sequence,embedding with structural feature,embedding with attribute feature and merging features. Experiment result shows that this method can take into account the structural feature and attribute features,realizes the mutual supplement of two kinds of feature information,and is better than the traditional network embedding methods.

Keywords　Network embedding,Heterogeneous information network,Structural feature,Attribute feature,Attributed heterogeneous network

1　引言

随着信息技术的不断发展,各类数字化系统记录了大量结构化数据和非结构化数据。许多研究者将这些数据抽象为信息网络[1],例如疾病症状网络[2]、道路时空网络[3]、文献网络[4]等。这些网络数据可被进一步挖掘分析,使网络节点采用向量表示,从而使网络数据更适合作为机器学习模型和深度学习模型的输入,提升了分类、聚类等任务的性能。

根据网络中边和节点的性质,例如节点类型数量与边类型数量,网络可分为同质网络与异质网络[5],同质网络中仅存在一类节点和一类关系,异质网络中存在多种类型的节点与关系;根据是否具有属性值网络可被分为属性网络与非属性

到稿日期:2021-12-07　返修日期:2022-03-28
基金项目:国家自然科学基金(61802033,61472064,61602096);四川省区域创新合作项目(2020YFQ0018);四川省科技计划重点研发项目(2021YFG0027,2020YFG0475,2018GZ0087,2019YJ0543);博士后基金项目(2018M643453);广东省国家重点实验室项目(2017B030314131);网络与数据安全四川省重点实验室开放课题(NDSMS201606)
This work was supported by the National Natural Science Foundation of China(61802033,61472064,61602096),Sichuan Regional Innovation Cooperation Project(2020YFQ0018),Sichuan Science and Technology Program(2021YFG0027,2020YFG0475,2018GZ0087,2019YJ0543),Chinese Postdoctoral Science Foundation(2018M643453),Guangdong Provincial Key Laboratory Project(2017B030314131) and Network and Data Security Key Laboratory of Sichuan Province Open Issue(NDSMS201606).
通信作者:张凤荔(fzhang@uestc.edu.cn)

网络，属性网络中保存能够描述节点更精细语义的属性信息，非属性网络中仅保存各节点间的结构关系。在对节点进行嵌入时，若要充分保留网络的结构信息与属性信息，则需要考虑网络中节点的异质性、结构的拓扑性和节点的属性值。因此，针对属性异质网络嵌入方法的研究是网络挖掘分析的一个重要研究内容。

当前已有较多关于异质网络的嵌入方法，然而这些嵌入方法通常仅考虑了网络结构信息。有部分关于属性网络的嵌入方法，但这些方法的嵌入对象通常为同质网络，无法适用于更加复杂的异质网络。与属性异质网络相关的嵌入研究仅考虑了节点属性，忽略了能够进一步区分同种关系中不同语义强弱的关系属性。

针对当前研究存在的问题，本文设计了一种能够同时保留属性异质网络结构信息与属性信息的嵌入方法。对于网络结构信息，首先得到采样过程中融合关系属性的元路径节点序列，再对节点进行嵌入；对于节点属性信息，将中心节点属性特征化并融合，使其特征分布与结构特征分布保持一致，避免分布不一致影响最终节点的嵌入能力，最后将二者的特征进行级联。该方法分别考虑了网络的结构特征和属性特征，形成了最终能够同时保留网络结构信息与属性信息的嵌入向量。

本文的主要贡献如下：

(1)定义了一种包含节点属性与关系属性的属性异质网络，该网络更符合现实生活，可为将来更多基于网络嵌入的应用落地提供理论支持。

(2)针对属性异质网络，设计了一种能够同时保留网络结构信息与属性信息的嵌入方法。该方法将包含节点与边的结构信息转换为结构特征向量，将关系属性转换为节点间的权重，将节点属性转换为属性特征向量。同时，该方法能够统一不同特征数值的分布情况，使最终节点嵌入兼顾结构特征和属性特征，实现二者信息的互补。

(3)将传统方法和本文方法进行对比实验，结果表明，本文方法能够更好地捕获属性异质网络中的各类信息特征，提升节点嵌入的分类效果和聚类效果。

2 相关工作

网络嵌入，又名网络表示学习[6]、图嵌入，是一种以低维向量表示高维复杂数据的数据压缩方法。根据节点类型与边类型，可将网络分为同质网络与异质网络；根据是否包含属性，可将网络分为属性网络与非属性网络。对于不同的网络有不同的嵌入方法，本节将从同质网络、异质网络、属性网络3个方面对部分网络嵌入方法进行介绍。

2.1 同质网络嵌入

同质网络嵌入方法主要用于处理节点类型单一的网络。DeepWalk[7]使用随机游走算法得到节点序列，将节点视为单词，使用word2vec[8-9]实现对节点的嵌入，该方法首次基于采样的方式实现网络节点嵌入，嵌入效率较高，训练所需空间较小。Node2vec[10]对DeepWalk进行改进，通过控制返回概率和进出概率来实现结合DFS和BFS的有偏随机游走，该方法可适用于具有不同拓扑结构的网络，但设置两个参数时存在部分主观问题。LINE[11]考虑了节点的一阶结构与二阶结构，实现了局部结构相似节点KL散度距离最小的嵌入目标，该方法对于大规模网络具有较好的处理速度。GraphSAGE[12]不断聚合邻居节点的信息，该方法可使节点嵌入同时具有自身结构与局部结构信息，但无法处理带权网络，邻居节点仅能实现等权聚合。GAT[13]在聚合相邻节点时使用了注意力机制，该方法通过计算注意力分数来对不同节点实现非等权聚合，但无法利用边上的权重进行进一步的聚合。

以上同质网络嵌入方法能够处理节点类型单一的网络，相比矩阵分解，该类嵌入方法简洁，灵活度较高。但现实生活中各种网络组成对象类型复杂，同质网络嵌入方法无法很好地适应异质网络节点嵌入。

2.2 异质网络嵌入

异质网络嵌入方法主要用于处理具有多种类型的节点与边的网络，更适用于复杂的现实网络。metapath2vec[14]基于元路径对网络进行采样后再嵌入，该方法在嵌入节点时考虑了节点的异质性，但在选择下一个节点时没有区分同类关系中不同节点关系实例的差异，损失了部分语义信息。HIN2vec[15]考虑了网络中边的异质性，基于一个神经网络模型可以同时得到节点和元路径的嵌入向量，该方法进一步区分了边类型，但没有使用节点间的关系属性。HAN[16]多次利用注意力机制，得到节点级和元路径级的注意力系数，该方法可充分利用节点与元路径间的关系，针对不同层次的对象实现选择性聚合。

异质网络嵌入方法能够进一步适应现实生活中的网络，在嵌入时能够考虑不同节点的类型，使最终嵌入能够保留原始网络中的信息。但异质网络嵌入方法通常仅考虑网络结构信息，忽略了进一步区分不同节点差异的属性信息。

2.3 属性网络嵌入

网络中的节点与边通常附有相关属性，这些属性能更精细地反映节点语义与关系语义，因此部分研究对带有各种属性的网络嵌入方法进行了研究。ASNE[17]在嵌入时加入了社交网络中的节点属性，实现了对网络中的人物节点的嵌入，该方法得到的最终嵌入融合了节点属性信息，但无法适用于异质网络嵌入。MIRand[18]将节点属性视为另一类节点，通过一种新随机游走算法生成节点序列，得到节点嵌入，该方法可捕获节点间无边但属性相似的节点对，实现了深层信息的嵌入，但在节点分层时仍无法适用于异质网络。HANE[19]将不同类别邻居的属性特征转换到同一空间，通过聚合邻居特征与组合节点自身特征得到最终的节点嵌入，该方法适用于异质网络，使用了节点属性，嵌入效果较好，但未利用到节点间的关系属性。

当前这些属性网络嵌入方法能够进一步考虑节点属性，使最终嵌入保留更多的原始网络信息。但这些嵌入方法的研究对象通常为节点类型单一的同质网络，忽略了网络的异质性。在考虑网络节点异质性时仅使用了节点属性，忽略了能够进一步区分节点关系强弱的关系属性。

综上，目前尚无研究同时使用网络结构信息与包含节点属性、关系属性的嵌入方法。因此，本文拟研究同时保留属性异质网络结构信息与各类属性信息的嵌入方法，该方法能够充分利用网络中的多种特征，可为未来具有更加复杂的结构和语义的大规模网络挖掘任务提供参考。

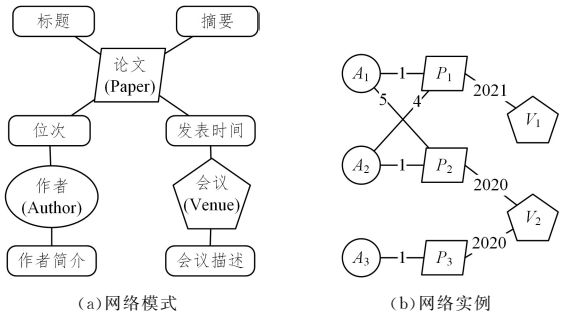
3 相关概念的定义

定义 1(网络嵌入,Network Embedding) 对于一个抽象网络 $G=(V,E)$,V 表示网络中的节点集合,$|V|$ 为节点数量,E 表示网络中的边集合,$|E|$ 为边数量。将网络 G 使用低维、实值、稠密的 m 维包含 $|V|$ 个向量的集合表示网络节点的方法称作网络嵌入,其中 $m \ll |V|$,该方法又被称为图嵌入、网络表示学习。

定义 2(属性异质网络,Attributed Heterogeneous Network) 对于一个网络 $G=(V,E,A,T,R)$,V 表示网络中的节点集合,E 表示网络中的边集合,A 表示网络中的属性集合,T 表示节点类型集合,R 表示边类型集合。集合 V 中包含 $|V|$ 个节点,对于任意节点 $v \in V$,存在一个映射 Ψ 使得 $\Psi(v)=t, t \in T$,$|T|$ 为节点类型数。集合 E 中包含 $|E|$ 条边,对于任意一条边 $e \in E$,存在一个映射 Ω 使得 $\Omega(e)=r, r \in R$,$|R|$ 为边类型数。集合 A 中包含 $|A|$ 个属性,所有属性分为节点属性与关系属性两类,对于任意节点属性 $att^v \in A$,存在一个映射 $\Phi(att^v) \in V$,即每一个节点属性必属于某个节点,不会单独存在;对于任意关系属性 $att^e \in A$,存在一个映射 $X(att^e) \in E$,即每一个关系属性必属于某条边,不会单独存在。若 $|T|=1$ 且 $|R|=1$,则网络 G 为同质网络;若 $|T|+|R|>2$,则网络 G 为异质网络;若 $A=\varnothing$,则网络 G 为非属性网络;若 $A \neq \varnothing$,则网络 G 为属性网络。若 $|T|+|R|>2$ 且 $A \neq \varnothing$,则网络 G 为属性异质网络。

图 1(a)为属性异质学术网络模式图,其中包含作者(Author)、论文(Paper)、会议(Venue)3 类节点,节点之间的边表示两个节点具有联系。A-P 表示作者 A 发表过论文 P,存在关系属性作者位次;P-V 表示论文 P 发表在会议 V 上,存在关系属性发表时间。同时节点也存在与之关联的属性,作者节点有作者简介,论文节点有标题、摘要等属性,会议节点有会议描述。这些属性信息能对节点语义进行补充,在嵌入时保留这些属性信息,能够提升节点保留原始网络信息的能力。

图 2(b)为一个属性异质网络学术实例图,作者 $A1$ 和 $A2$ 合作了论文 $P1$ 和 $P2$,因此与会议 $V1$ 和 $V2$ 均有语义联系。但两位作者对两篇论文的贡献不同,$A1$ 是 $P1$ 的主要贡献者,$A2$ 是 $P2$ 的主要贡献者,因此 $A1$ 的研究方向与会议 $V1$ 的领域更相关,$A2$ 的研究方向与会议 $V2$ 的领域更相关。



(a)网络模式 (b)网络实例

图 1 属性异质学术网络模式与实例图

Fig. 1 Network scheme and instance diagram of attributed heterogeneous academic network

定义 3(属性异质网络嵌入 Attributed Heterogeneous Network Embedding) 对于一个属性异质网络 $G=(V,E,A,T,R)$,将其转换成一个具有 $|V|$ 个低维、实值、稠密的 m 维向量集合,称为属性异质网络嵌入,其中 $m \ll |V|$。该方法能够让非结构化网络数据转换为结构化向量数据,为机器学习、深度学习模型提供输入数据,且能充分保留原始网络中的结构信息与属性信息,提升下游任务的预测性能。

4 方法介绍

本节对融合多特征的属性异质网络嵌入方法从总体流程到方法细节进行了介绍,其中包含总体流程、序列采样、结构特征嵌入、属性特征嵌入、特征融合 5 个部分。

4.1 总体流程

图 2 给出了整个嵌入流程,主要分为 4 个模块:序列采样模块、结构特征嵌入模块、属性特征嵌入模块、特征融合模块。

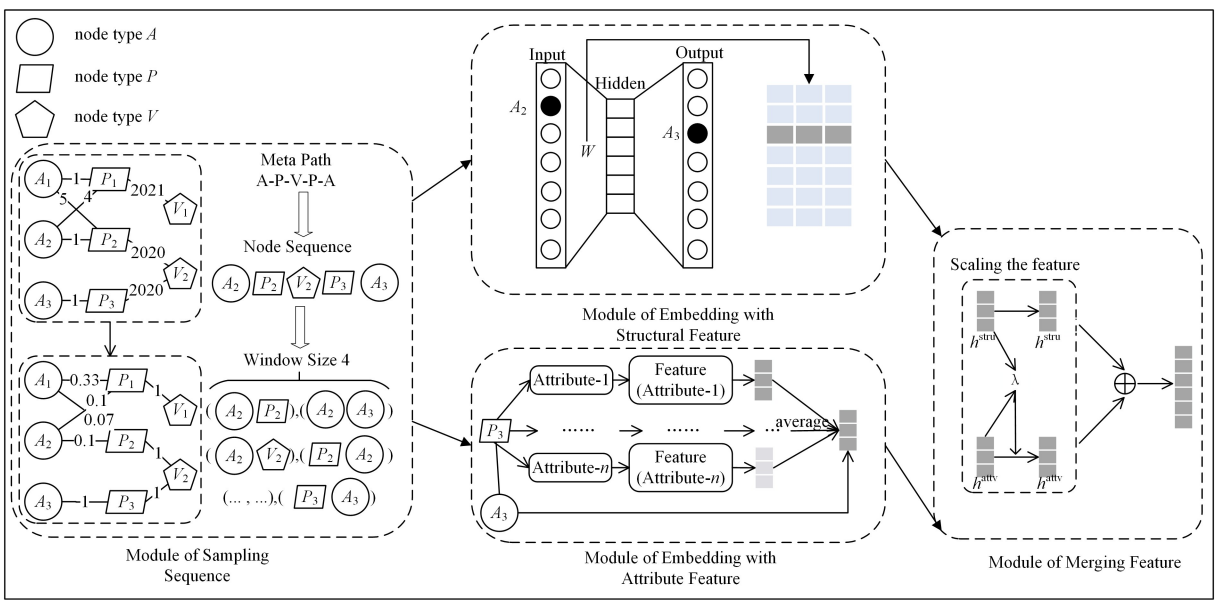


图 2 总体流程

Fig. 2 Overall process

序列采样模块主要负责对网络进行采样，为后续模块提供融合关系属性的元路径[13]节点序列实例。

结构特征嵌入模块主要负责对节点的结构信息进行嵌入，采用异质 Skip-Gram[14]算法对所得采样序列中的节点进行训练，得到保留网络结构信息的结构特征。

属性特征嵌入模块主要负责对节点的属性信息进行嵌入。该模块首先基于不同的特征化方法将节点属性向量化，再将不同属性特征向量进行融合，以得到保留节点属性信息的属性特征。

特征融合模块主要负责将同一节点的结构特征与属性特征进行融合。独立得到的结构特征与属性特征通常不具有相同的语义与数值量纲，需要避免直接相加引起语义混淆的问题，使所有网络信息得到充分保留。

4.2 序列采样

基于序列采样的嵌入方法是网络嵌入的一种主要方式，其核心在于选择适当的方法对网络生成大量节点序列。网络中关系属性能够区分同类关系中的不同关系实例，元路径能够区分节点的异质性，二者皆可影响最终的节点嵌入效果。因此，对于具有关系属性的异质网络，可以先将关系属性权重化，再结合元路径实现基于关系属性的序列采样。

4.2.1 关系属性权重化

关系属性可分为数值型属性与非数值型属性两类，本文仅对数值型关系属性进行讨论。对于数值型属性，又可按照是否共享语义空间进行分类，即同一数值在所有同种关系中是否具有相同语义。

第一类属性具有同一数值在同类关系中有不相同语义的特征，例如作者在论文中的排名顺序，不同论文中作者总数不同时，相同的位次数值具有不同的语义。对于这些属性，需要按语义空间分别单独计算其权重。

$$w_i^{r,k}=\frac{att_i^{r,k}}{\sum_{j=1}^{n} att_j^{r,k}} \tag{1}$$

其中，$att_j^{r,k}(j=1,2,\cdots,n)$为关系 r 上的第 k 个独立语义空间中的第 j 个属性，该空间的属性个数为 n；$att_i^{r,k}$，$w_i^{r,k}$ 分别为原始属性和权重化值。

若关系属性存在数值越小节点联系越紧密的性质，则需要在权重化前进行转换处理，以保持权重值大小与关系紧密程度的一致性。

$$att_i^{r,k}=att_{\max}^{r,k}-att_{i,o}^{r,k}+att_{\min}^{r,k} \tag{2}$$

其中，$att_i^{r,k}$，$att_{\max}^{r,k}$，$att_{i,o}^{r,k}$，$att_{\min}^{r,k}$分别为转换后的属性值、原始属性中的最大值、原始属性值、原始属性中的最小值。在式(2)中加上最小值是为了避免当原始属性值为最大值时造成结果为零的情况，这里仅考虑所有属性值均为正数的情况。

第二类属性数值在同一关系中具有相同语义，例如用户对电影的评分、用户与商品的购买时间等属性，相同数值都具有相同语义。对于该类属性，可对所有关系属性进行统一计算。

$$w_i^r=\frac{att_i^r}{\sum_{j=1}^{n} att_j^r} \tag{3}$$

其中，$att_j^r(j=1,2,\cdots,n)$为关系 r 上的第 j 种属性，该类关系属性值的种类数为 n；att_i^r 和 w_i^r 分别为原始属性与权重值。

与第一类属性一致，若某类关系属性数值越小，表示节点联系程度越紧密，需要进行转换。

$$att_i^r=att_{\max}^r-att_{i,o}^r+att_{\min}^r \tag{4}$$

其中，att_i^r，$att_{\max}^r$，$att_{i,o}^r$，$att_{\min}^r$分别为转换后的属性值、原始属性中的最大值、原始属性值、原始属性中的最小值。

第三类属性的不同数值在同一类关系中具有相同语义，例如论文与所属会议的发表时间，虽然各论文发表时间不同，但其所属领域不变。对于该类属性，可把其权重统一视为 1。

4.2.2 基于关系属性和元路径的序列采样

将关系属性转换为边权重后，即可结合元路径实现概率采样。元路径通常被定义为一种序列模式 $Path:V_1\xrightarrow{(r_1)}V_2\xrightarrow{(r_1)}\cdots V_{n-2}\xrightarrow{r_{n-2}}V_{n-1}\xrightarrow{r_{n-2}}V_n$。每个节点序列中的节点类型均按照元路径定义方式排列，元路径中的节点类型通常对称，即$\Psi(v_1)=\Psi(v_n)$，$\Psi(v_2)=\Psi(v_{n-1})$。

传统的元路径采样方法在确定下一节点类型后，再在指定类型节点中随机选择一个节点作为下一个节点。该方法没有考虑节点的联系紧密程度，无法区分不同节点对的关系强弱，忽略了具体的关系语义。为了更好地保留与利用原始网络中的关系属性，将利用关系属性转换后的权重实现结合元路径的有偏采样。对于节点 v_i，下一个节点 v_j 的选择概率为：

$$prob(v_j\mid v_i)=\begin{cases}\dfrac{w_{i,j}^r}{\sum_{l}^{N_{t+1}(v_i)} w_{i,l}^r}, & (v_i,v_j)\in E,\psi(v_j)=t+1\\ 0, & (v_i,v_j)\in E,\psi(v_j)\neq t+1\\ 0, & (v_i,v_j)\notin E\end{cases} \tag{5}$$

其中，$prob(v_j\mid v_i)$为从节点 v_i 跳转到 v_j 的概率，当 v_i 和 v_j 间不存在边或 v_j 不是目标类型节点时，跳转概率为零，当 v_i 和 v_j 间存在边且 v_j 是目标类型节点时，根据权重计算相应的跳转概率；$w_{i,j}^r$ 为节点 v_i 与节点 v_j 间的权重；$w_{i,l}^r$ 为节点 v_i 与其邻居类型为 $t+1$ 的节点间的权重。

4.3 结构特征嵌入

对于异质网络，在对结构信息嵌入时使用异质 Skip-Gram 算法，算法流程如图 2 所示，分为输入层、隐藏层、输出层以及中间权重。模型训练完成后，输入层与隐藏层间的权重即为节点嵌入 W，$W\in R^{|V|\times dim}$，$|V|$为网络节点数，dim 为嵌入向量长度(即隐藏层神经元数)。该算法将基于窗口得到的节点对作为输入与输出，即输入层为中心节点，输出层为窗口内的共现节点。最终实现目标节点共现概率最大化，优化目标如下：

$$\arg\max \prod_{v\in V}\prod_{t\in T}\prod_{c_t\in N_t(v)} p(c_t\mid v,\theta) \tag{6}$$

即实现对所有节点每种类型邻居中的所有邻居节点共现概率最大化，其中 V 为网络中的节点集合，T 为节点类型集合，$N_t(v)$为节点 v 邻居中类型为 t 的节点，θ 为模型中的各类参数，最终目标使所有点的共现节点出现概率最大。过多的概率连乘会导致目标结果过小，难以训练模型。为了保证模型正常训练，需要做以下变化：

$$\arg\max \sum_{v\in V}\sum_{t\in T}\sum_{c_t\in N_t(v)} \log p(c_t\mid v,\theta) \tag{7}$$

通过对数函数将乘积结果转换为累加结果，实现模型的可训练性。节点共现概率可表示为：

$$p(c_t|v,\theta)=\frac{e^{h_{c_t}^{stru}\cdot h_v^{stru}}}{\sum_{u\in V}e^{h_u^{stru}\cdot h_v^{stru}}} \tag{8}$$

其中，V 为所有节点的集合，h_{ct}^{stru}，h_u^{stru}，h_v^{stru} 分别为待预测共现节点、其他节点、中心节点对应的嵌入向量。

在具体计算时，式(8)中的分母部分需要与所有节点进行计算，当节点数量过多时，计算复杂度将大大增加，影响模型训练的性能。因此，基于负采样[9]的优化方法被用于解决分母部分计算量大的问题。负采样方法的主要思想为，通过较少负样本模拟整体负样本分布，从而实现固定的计算复杂度，即：

$$p(c_t|v,\theta)=\frac{e^{h_{c_t}^{stru}\cdot h_v^{stru}}}{\sum_{i=1}^{m}e^{h_{u_i}^{stru}\cdot h_v^{stru}}} \tag{9}$$

其中，m 表示负样本数量，u_i 为原始负样本中随机挑选的节点，最终损失函数为：

$$loss=-\log\sigma(X_{c_t}\cdot X_v)+\sum_{i=1}^{m}E_{u_i}\sim P_t(u)[\log\sigma(-X_{u_i}\cdot X_v)] \tag{10}$$

其中，σ 为激活函数，$\sigma(x)=1/(1+e^{-x})$，该函数可引入非线性因素，提高模型的表达能力。经过迭代训练，可实现目标节点对的共现概率最大化，输入层与隐藏层间的权重矩阵 $\boldsymbol{W}$ 即为节点嵌入向量集合 H^{stru}，$H^{stru}\in R^{|V|\times dim}$，该嵌入向量集合保留了原始网络中的结构信息，可为节点的最终嵌入提供重要语义。

4.4 属性特征嵌入

属性信息能从个体角度描述节点更具体的语义，使其与其他同类型节点区分开。然而，节点属性与节点间的关系表示方法不同，难以直接联系。同时，节点属性存在种类多、类型不一致等问题，导致难以同时融合各类原始属性信息。为了将各类属性信息融入至最终节点嵌入，需要对属性信息进行嵌入与融合。

对于原始节点属性，首先需要将各类属性转换为特征向量：

$$\boldsymbol{h}_{v,i}^{attv}=feature(att_{v,i}) \tag{11}$$

其中，$att_{v,i}$ 表示节点 v 的第 i 个属性，$feature()$ 表示将节点属性选择对应的方法进行特征化，$\boldsymbol{h}_{v,i}^{attv}$ 表示属性 $att_{v,i}$ 特征化后的向量。对于文本属性，可以采用 doc2vec[20]，fastText[21] 等方法进行处理；对于图像属性，可以采用 CNN、池化等方法进行处理；对于数值属性，可以在原始数值的基础上进行归一化处理；对于类别属性，可以使用 one-hot 方法。根据节点属性的不同特点，选择不同的属性特征化方法。

对于得到的独立属性特征，可以进行级联拼接、相加等融合方式。但各节点属性数量不一致，若采用级联拼接的融合方式，则会引起属性特征维度不固定，因此采用相加取平均的方式。

$$\boldsymbol{h}_v^{attv}=\frac{1}{n}\sum_{i=1}^{n}h_{v,i}^{attv} \tag{12}$$

其中，n 为节点属性数量，$\boldsymbol{h}_v^{attv}$ 表示节点的属性特征向量，该特征向量可融合节点自身的各种属性，是节点最终嵌入语义的重要来源。

对于缺少节点属性的节点，可将其邻居节点的属性特征向量作为自己的属性特征向量，最终得到属性特征集合 H^{attv}，$H^{attv}\in R^{|V|\times dim}$，该集合大小与结构特征集合大小一致。

4.5 特征融合

经过结构特征嵌入与属性特征嵌入，可得到保留结构信息的结构特征与保留属性信息的属性特征，两种特征都可为最终的节点嵌入提供丰富语义，本节需要对两种特征进行融合。

$$h_v=fuse(h_v^{stru},h_v^{attv}) \tag{13}$$

常见的融合方式有取平均值、级联拼接等操作。取平均值可保证最终的嵌入维度大小保持不变，但存在着特征语义不一致的问题；级联拼接可以避免相加引起特征语义混淆的问题，但拼接后会引起特征大小发生变化。由于维度大小变化仅会引起计算量的变化，不会引起语义混淆，因此最终选用级联拼接的方式。

$$\boldsymbol{h}_v=h_v^{stru}\oplus h_v^{attv} \tag{14}$$

其中，h_v^{stru}，$h_v^{attv}\in R^{dim}$，$\oplus$ 表示级联拼接，$\boldsymbol{h}_v$ 表示拼接后的向量，$\boldsymbol{h}_v\in R^{2\cdot dim}$。

但简单的拼接忽略了两种特征向量的数值分布问题。由于结构特征与属性特征分别通过计算获得，两种特征中的具体数值存在分布差异，即一类特征中数值均值较大、标准差较大，另一类特征中数值均值较小、标准差较小。若不对两种特征进行处理而直接级联拼接，则会导致部分特征对后续分类、聚类等任务预测结果的贡献度不足。因此，在拼接前需要将两种特征的数值分布进行统一，使得：

$$\begin{cases}mean(H^{stru})=mean(H^{attv})\\std(H^{stru})=std(H^{attv})\end{cases} \tag{15}$$

其中，$mean()$，$std()$ 分别表示计算平均值、标准差，H^{stru}，H^{attv} 分别为结构特征向量集合、属性特征向量集合。为了统一两个特征集合的数值分布，需要计算缩放系数：

$$\lambda=std(H^{stru})/std(H^{attv}) \tag{16}$$

即缩放系数由两个特征集合的标准差决定。对于两个特征集合，选择结构特征集合不变，对属性特征集合进行缩放。对属性特征集合中所有特征向量的所有数值进行以下变换：

$$x'=mean(H^{stru})+\lambda(x-mean(H^{attv})) \tag{17}$$

其中，x 为属性特征向量的原始数值，x' 为经过缩放后的数值。通过式(17)的处理，可实现结构特征与属性特征的数值分布一致性，且保证了属性特征数值的相对偏差不发生变化。再将处理后的结构特征与属性特征进行级联拼接以及激活函数处理，从而得到最终的节点嵌入。

$$h_v^{final}=\mathrm{sigmoid}(h_v) \tag{18}$$

其中，$\mathrm{sigmoid}(x)=1/(1+e^{-x})$，经过激活函数处理的嵌入向量可让嵌入数值处于区间(0,1)中，可防止出现过大或过小的极值影响特征语义，最终得到节点嵌入集合 H^{final}，$H^{final}\in R^{|V|\times 2\cdot dim}$，即最终节点嵌入的维度大小为单一特征的两倍。该嵌入同时保留了结构特征与属性特征，最终语义由二者共同决定。当二者的语义相同时，最终节点语义与两种特征的语义一致；当二者语义相反时，最终语义由特征更显著、贡献

度更大的特征决定。

整个嵌入算法的流程如算法 1 所示。

算法 1 MFAHNE

输入：$G=(V,E,A,T,R)$

输出：$H^{final}\in R^{|V|\times dim}$

1. ＃序列采样
2. for v_i in V：
3. for v_j in V：
4. $w_{i,j}$←weighted($att_{i,j}$)(式(1)－式(4))
5. Sequence←sample(W)(式(5))
6. ＃特征嵌入模块
7. for batch in batches：
8. score←计算相似分数(H^{stru}_{batch})(式(8)、式(9))
9. loss←计算损失(score，H^{stru}_{batch})(式(6)、式(7)、式(10))
10. 更新参数(loss)
11. ＃属性嵌入模块
12. for v in V：
13. for i in num_{attv}(v)：
14. $h^{attv}_{v,i}$←feature($att_{v,i}$)(式(11))
15. h^{attv}_{v}←average($h^{attv}_{v,i}$)(式(12))
16. ＃特征融合
17. λ←std(H^{stru})/std(H^{attv})(式(15)、式(16))
18. for h^{attv}_{v} in H^{attv}：
19. for x in h^{attv}_{v}：
20. x'←mean(H^{stru})＋λ(x－mean(H^{attv}))(式(17))
21. H^{final}←sigmoid($H^{stru}\oplus H^{attv}$)(式(13)、式(14)、式(18))

整个算法的流程主要包括 4 个部分：1)序列采样，首先结合式(1)－式(4)将关系属性权重化，然后以得到的权重集合 W 为基础，基于元路径实现对异质网络的序列采样，得到节点序列集合 Sequence；2)结构特征嵌入，根据得到的节点序列集合，将整个训练集按照参数批处理数量划分成多个子训练集 batches，对于每个批次的训练集 batch，采用异质 Skip-Gram 算法进行处理，每个批次的节点通过计算相似分数、计算损失、更新参数等步骤实现对节点的嵌入，得到结构特征集合 H^{stru}；3)属性特征嵌入，首先对单个节点的所有属性进行特征化，再将单个节点的所有属性进行平均，得到所有节点的属性特征集合 H^{attv}；4)特征融合，基于得到的结构特征和属性特征，根据两个集合标准差平均值得到缩放系数，将所有属性特征的所有数值缩放后，再将结构特征与属性特征级联拼接与归一化，得到最终节点嵌入集合 H^{final}。

最终得到的节点嵌入由结构特征向量与属性特征向量组成，通过特征缩放，避免了不同类特征向量数值大小不一致的问题。由于没有进行进一步融合，这些特征向量在计算过程中可分别做出贡献，对最终结果起到相互补充信息的作用，从而提升节点分类、聚类任务的性能。

5 实验分析

本节将本文的 MFAHNE 方法与部分传统方法进行实验比较，将从数据集、基线模型、分类实验、聚类实验 4 个部分进行介绍。

5.1 数据集

AMiner[1)] 是一个提供科技情报大数据挖掘与服务的系统平台，其收集了大量科研人员、科研论文、学术活动等数据。本文从其数据库中选择了近年来计算机领域不同方向的部分作者、论文、会议信息，构成了学术网络，网络相关属性如表 1 所列。

表 1 AMiner 学术网络属性

Table 1 Attributes of AMiner academic network

网络属性	属性值
作者数	115437
论文数	57585
会议数	253
A-P 关系数	222612
P-V 关系数	57585
作者属性	无
论文属性	标题、摘要
会议属性	无
A-P 关系属性	作者位次
P-V 关系属性	发表年份
作者节点平均度数	1.928
论文节点平均度数	4.866
会议节点平均度数	227.609
节点类别数	9
时间范围	[2015，2021]

该学术网络包含作者(Author)、论文(Paper)、会议(Venue)3 类节点，*A-P* 关系表示作者发表了论文，*P-V* 关系表示论文在会议发表。论文存在标题、摘要等属性，构成该网络时，由于没有考虑作者、会议的节点属性，因此将与其联系的论文属性特征视为自身属性特征。对于关系属性，*A-P* 关系上的属性为作者位次，位次越小，表示作者与论文的联系紧密程度越大；*P-V* 关系上的关系属性为发表年份，虽然不同的论文发表年份不一致，但其研究方向固定，可忽略年份不一致引起的语义不一致问题。该网络包含了 9 个计算机领域的相关方向：计算机体系结构、计算机网络、网络与信息安全、软件工程、数据库、计算机科学理论、计算机图形学与多媒体、人机交互与普适计算、交叉/综合/新兴。会议标签为所属领域，论文标签为所属会议标签，作者标签为发表论文数最多的论文类别标签。

5.2 基线模型

(1)Deepwalk。该方法基于随机游走对原始网络节点采样，得到节点序列，再使用 word2vec 对节点序列进行处理，得到节点嵌入向量。

(2)Node2vec。该方法在采样的同时考虑了 DFS 与 BFS 的特性，可通过参数控制采样的方式来适应不同的网络结构，对于采样得到的节点序列同样使用 word2vec 得到节点嵌入。

(3)LINE。该方法对节点的一阶和二阶相似度同时进行建模，最小化节点间的 KL 散度。

(4)metapath2vec_n。该方法中的 n 表示嵌入维度，整个方法基于元路径对网络采样，可以设定不同类型节点的出现顺序，从而实现对不同关系语义特征的捕获。基于得到的元路径节点序列，采用异质 Skip-Gram 算法得到节点嵌入，适用于

1) https://www.aminer.cn/data/?nav=openData

异质网络。

(5)MFAHNE。该方法将关系属性转换为权重,结合元路径实现有偏采样,再基于异质 Skip-Gram 算法得到节点结构特征,然后将节点的各类属性特征化与融合得到属性特征,最后将两种特征统一分布后进行级联拼接,得到最终的节点嵌入。

对于 MFAHNE,训练结构特征时采用均匀分布随机初始化参数,节点重复采样次数为 500,元路径重复次数为 25,维度为 64,窗口长度为 4,批处理数量为 50,学习率为 0.05,节点最小出现次数为 5,负样本数量为 5,采用的元路径为 APVPA。对于属性特征,使用 fastText 对标题和摘要进行训练,嵌入维度与结构特征相同,因此 MFAHNE 的最终嵌入维度为 128。对于其他方法,默认嵌入维度为 128。

同时,本文也设计了部分变体来验证 MFAHNE 方法中部分操作的有效性。

(1)$\text{Feature}_{\text{stru}}$:该方法仅使用结构特征进行实验,除采样节点序列不一致外,其他参数与 metapath2vec_{64} 一致。

(2)$\text{Feature}_{\text{attv}}$:该方法仅使用属性特征进行实验。

(3)$\text{MFAHNE}_{\text{cat}}$:该方法不统一特征分布,直接对原始结构特征与属性特征进行级联拼接。

(4)$\text{MFAHNE}_{\text{ave}}$:该方法不采用拼接策略,统一特征分布后,再将结构特征与属性特征取平均值。

5.3 分类任务

对于分类任务,本文采用 KNN 分类算法对所得节点进行验证。KNN 直接利用待分类数据与训练集数据的距离作为判定标准,无须训练参数,分类结果可直接反映嵌入向量效果,实验中节点参考邻居数 $k=5$。整个数据集被划分为训练集和测试集,训练集比例分别为 10%,20%,30%,40%,50%,60%,70%,80%,90%,将相应的剩余节点作为测试集。由于原始网络中各类别节点数量不平衡,为了避免随机划分数据集导致训练集与测试集节点类型数量出现偏差,在随机选择训练集时,每个类别分别选取相应比例的节点。评估分类结果指标采用 Micro-F1 和 Macro-F1,其值越高表示结果越好。实验重复进行 10 次,最终取所有结果的平均值。

表 2 列出了传统嵌入方法、本文方法及变体方法对节点采用不同比例的训练集进行训练后的分类结果。MFAHNE 在 Micro-F1 与 Macro-F1 指标上均优于其他方法,这表明该方法能使节点嵌入充分保留原始网络的结构信息与属性信息,提升了节点的分类效果。在 Deepwalk,Node2vec,LINE 这 3 种同质网络嵌入算法中,LINE 的效果最差,原因可能是该测试数据集为异质网络,具有明显的类型拓扑结构,LINE 只考虑节点的一阶信息和二阶信息,不足以捕获节点特征。metapath2vec 普遍优于 3 种同质网络嵌入方法,表明对于异质网络,需要充分考虑节点的异质性。$\text{Feature}_{\text{stru}}$ 优于 metapath2vec_{64},这说明融合关系属性的结构特征保留了更多的原始网络语义。$\text{MFAHNE}_{\text{cat}}$ 与 $\text{Feature}_{\text{stru}}$ 的结果近乎一致,说明没有经过特征分布处理的属性特征向量对最终节点嵌入效果的贡献不足。MFAHNE 同时优于单独的 $\text{Feature}_{\text{stru}}$ 与 $\text{Feature}_{\text{attv}}$,这说明结构特征信息与属性特征信息能够相互补充,共同提高最终节点嵌入的分类能力。$\text{MFAHNE}_{\text{ave}}$ 同时优于 $\text{Feature}_{\text{stru}}$ 与 $\text{Feature}_{\text{attv}}$,但弱于 MFAHNE,这说明结构特征与属性特征在融合时发生了部分语义混淆,降低了最终节点的嵌入能力。

表 2 节点分类结果

Table 2 Results of node classification

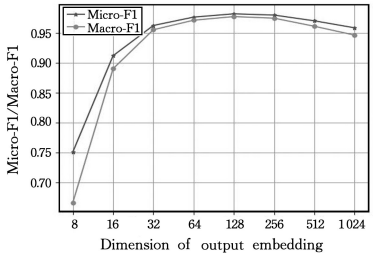
		10%	20%	30%	40%	50%	60%	70%	80%	90%
Micro-F1	Deepwalk	0.8219	0.8494	0.8659	0.8795	0.8945	0.9154	0.9238	0.9298	0.9277
	Node2vec	0.8559	0.8657	0.8797	0.8921	0.9136	0.9266	0.9339	0.9443	0.9440
	LINE(1st+2nd)	0.2417	0.2598	0.2702	0.2971	0.3209	0.3255	0.3368	0.3478	0.3297
	$\text{metapath2vec}_{128}$	0.7902	0.8368	0.8793	0.9004	0.9173	0.9310	0.9392	0.9483	0.9508
	metapath2vec_{64}	0.8317	0.8601	0.8955	0.9125	0.9269	0.9411	0.9472	0.9559	0.9573
	$\text{Feature}_{\text{stru}}$	0.8399	0.8636	0.9000	0.9161	0.9333	0.9469	0.9549	0.9626	0.9649
	$\text{Feature}_{\text{attv}}$	0.9105	0.9099	0.9140	0.9136	0.9154	0.9130	0.9146	0.9084	0.9027
	$\text{MFAHNE}_{\text{cat}}$	0.8406	0.8639	0.9006	0.9167	0.9331	0.9466	0.9551	0.9629	0.9654
	$\text{MFAHNE}_{\text{ave}}$	0.9108	0.9154	0.9349	0.9480	0.9609	0.9689	0.9746	0.9775	0.9728
	MFAHNE	**0.9353**	**0.9362**	**0.9519**	**0.9603**	**0.9701**	**0.9756**	**0.9800**	**0.9820**	**0.9799**
Macro-F1	Deepwalk	0.8073	0.8275	0.8543	0.8637	0.8852	0.9100	0.9185	0.9259	0.9243
	Node2vec	0.8343	0.8553	0.8614	0.8785	0.9022	0.9179	0.9253	0.9377	0.9382
	LINE(1st+2nd)	0.1692	0.1857	0.1966	0.2195	0.2432	0.2502	0.2580	0.2695	0.2531
	$\text{metapath2vec}_{128}$	0.7650	0.8107	0.8674	0.8910	0.9112	0.9250	0.9344	0.9474	0.9474
	metapath2vec_{64}	0.8120	0.8424	0.8890	0.9044	0.9223	0.9339	0.9424	0.9542	0.9539
	$\text{Feature}_{\text{stru}}$	0.8219	0.8473	0.8903	0.9084	0.9285	0.9421	0.9515	0.9608	0.9634
	$\text{Feature}_{\text{attv}}$	0.8845	0.8807	0.8873	0.8868	0.8896	0.8864	0.8912	0.8829	0.8764
	$\text{MFAHNE}_{\text{cat}}$	0.8227	0.8479	0.8911	0.9093	0.9282	0.9418	0.9518	0.9613	0.9638
	$\text{MFAHNE}_{\text{ave}}$	0.8917	0.8958	0.9210	0.9383	0.9539	0.9615	0.9693	0.9735	0.9684
	MFAHNE	**0.9187**	**0.9195**	**0.9396**	**0.9504**	**0.9627**	**0.9680**	**0.9745**	**0.9774**	**0.9749**

MFAHNE 利用网络中的结构信息、节点属性与关系属性,能够充分保留网络中的各类特征信息,同时在实现特征融合时,通过统一分布处理,能够避免特征贡献不平均的问题,最终得到一个较好的嵌入,提升了节点的分类效果。

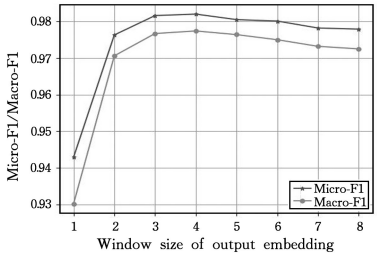
为了解方法参数对分类实验结果的影响,本文对 MFAHNE 的节点嵌入维度、采样窗口长度两个参数进行分析。分类的训练集、测试集比例分别为 80%和 20%,除嵌入维度与窗口长度外,其余参数保持不变。

图 3(a)给出了节点分类结果随嵌入维度变化而变化的情况。当节点嵌入维度较小时,分类效果较差,随着嵌入维度

的增加，分类效果快速提升；当嵌入维度达到 32 时，提升效果趋于平缓；当嵌入维度达到 128 时，节点嵌入效果最佳；当嵌入维度进一步增大时，效果降低。嵌入维度越大，表示嵌入能从更多空间角度描述节点语义，但嵌入维度过大时，可能由于数据量不足，不足以训练所有维度特征。过大的嵌入维度也需要更多的训练资源与计算资源，因此需要根据网络大小选择合适的嵌入维度。



(a)嵌入维度



(b)窗口长度

图 3 分类参数敏感性分析

Fig. 3 Parameter sensitivity analysis for classification

图 3(b)给出了节点分类结果随窗口长度的变化情况。当窗口长度为 1 时，分类结果明显差于其他情况；当窗口长度从 2 增加到 4 时，分类结果成正比增加，窗口长度为 4 时分类结果最优；当窗口长度超过 4 时，分类结果与窗口长度成反比减小。窗口长度决定着采样序列中心节点的邻居范围，当窗口长度过小时，难以体现节点之间的共现性，当窗口长度过大时，容易将非共现节点错误地视为邻居节点，因此窗口长度的大小需要根据网络节点的结构特征来确定。

5.4 聚类任务

对于聚类任务，本文采用 K-means 聚类算法对所得节点进行聚类验证。K-means 是一种无监督聚类算法，通过迭代聚类中心确定各节点的类别，类簇数设为节点类型数。采用 NMI 和 AMI 指标来评估聚类结果，值越大表示聚类效果越好。实验重复进行 10 次，最终取所有结果的平均值。

传统嵌入方法、本文方法及变体方法的聚类结果如表 3 所列，MFAHNE 优于其他所有方法，表明了本文方法能够对原始网络实现更好的嵌入，提升了节点的聚类效果。具体来看，Node2vec，LINE 的节点嵌入效果难以实现一个较好的聚类结果，针对异质网络的嵌入方法结果优于同质网络方法。$\text{Feature}_{\text{stru}}$ 优于 metapath2vec_{64}，说明融合关系属性可提升节点的聚类效果。$\text{MFAHNE}_{\text{cat}}$ 仅对 $\text{Feature}_{\text{stru}}$ 有微弱提升，说明未统一数值分布的属性特征对聚类结果的贡献较小。MFAHNE 同时优于 $\text{Feature}_{\text{stru}}$ 与 $\text{Feature}_{\text{attv}}$，说明 MFAHNE 能够实现结构特征与属性特征的信息相互补充。$\text{MFAHNE}_{\text{ave}}$ 弱于 MFAHNE，说明结构特征与属性特征在相加时出现了语义丢失。

表 3 节点聚类结果

Table 3 Results of node clustering

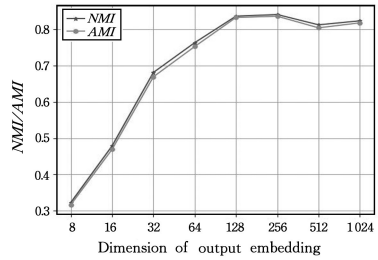
	NMI	*AMI*
Deepwalk	0.5004	0.4945
Node2vec	0.0678	0.0578
LINE(1st+2nd)	0.1201	0.1167
$\text{metapath2vec}_{128}$	0.4991	0.4990
metapath2vec_{64}	0.5518	0.5501
$\text{Feature}_{\text{stru}}$	0.5977	0.5945
$\text{Feature}_{\text{attv}}$	0.7211	0.7147
$\text{MFAHNE}_{\text{cat}}$	0.5860	0.5844
$\text{MFAHNE}_{\text{ave}}$	0.7491	0.7428
MFAHNE	0.8369	0.8332

对于聚类任务，MFAHNE 能够得到更好的聚类结果，说明 MFAHNE 对于网路中各类特征的嵌入能力较好，语义丢失较少。

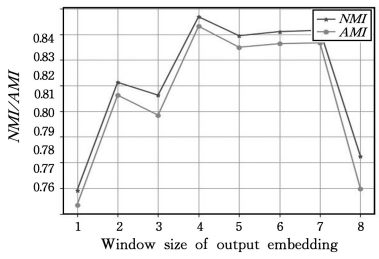
为了解方法参数对聚类实验结果的影响，本文对 MFAHNE 的节点嵌入维度、采样窗口长度两个参数进行了分析。聚类实验中，除嵌入维度与窗口长度外，其余参数保持不变。

图 4(a)给出了聚类结果随嵌入维度变化而变化的情况。可以发现，当嵌入维度较小时，随着嵌入维度的增加，聚类结果均匀增加；当维度达到 128 时，效果增加的情况趋于平缓；当维度达到 256 时，效果最佳；当维度大于 256 时，效果降低。这说明嵌入维度能够明显影响聚类效果，嵌入维度的提升通常能够提升聚类效果，但过大的嵌入维度会导致聚类效果降低。

图 4(b)给出了聚类结果随窗口长度变化而变化的情况。可以看出，当窗口长度为 4 时，聚类效果最佳。窗口长度过大会引入较大范围内的邻居，增加语义不一致邻居节点的出现概率，窗口长度过小时无法体现节点的局部结构特征，两种情况均会影响聚类效果。选择一个合适的窗口长度既能提升聚类效果，又能减少共现节点对数量，从而加快训练过程，因此需要根据具体网络选择合适的窗口长度。



(a)嵌入维度



(b)窗口长度

图 4 聚类参数敏感性分析

Fig. 4 Parameter sensitivity analysis for clustering

分析嵌入维度、窗口长度对分类、聚类任务的影响可以发现，过大或过小的参数都将导致节点嵌入效果不佳，需要根据

实际的网络特征确定合适的参数。

结束语 本文定义了一种包含节点属性与关系属性的属性异质网络。为了充分融合该网络中的多种特征，设计了一种同时保留网络结构信息与节点属性信息的嵌入方法，该方法分别对网络结构信息与节点属性信息嵌入后再实现融合，融合时设计了一种统一特征数值分布的方法。实验结果表明，关系属性能够区分节点间的联系强度，提高节点结构特征的嵌入效果，通过统一特征分布，实现了结构信息与属性信息的相互补充，提高了最终节点的嵌入效果。

对于异质网络嵌入研究，网络中的结构信息与属性信息都可为节点嵌入提供重要语义，关键在于同时充分保存二者的特征。已有部分研究将异质网络嵌入方法应用于实际系统，但网络中的节点种类多、网络结构复杂、属性模态多等问题影响着研究更进一步落地应用。由于标准数据获取困难，本文仅对学术网络进行了研究，网络中的论文节点缺少其他模态的属性，因此在属性嵌入的过程中仅考虑了标题、摘要两种文本信息。

更多现实生活中的各类网络总是处于动态变化之中，不同节点在不断建立连接、断开连接，节点属性、关系属性同样也会发生变化。如何捕获这些变化特征以实现节点嵌入的动态更新，使其更符合结构、属性变化后的语义，减小训练节点嵌入的计算代价，也是异质网络嵌入的一个重要研究课题。

参 考 文 献

[1] SUN Y Z, HAN J W. Mining heterogeneous information networks: a structural analysis approach[J]. SIGKDD Explorations, 2012, 14(2): 20-28.

[2] WANG Z F, WEN R, CHEN X, et al. Online Disease Diagnosis with Inductive Heterogeneous Graph Convolutional Networks [C]// Proceedings of the 30th International Conference on World Wide Web. 2021: 3349-3358.

[3] HONG H T, LIN Y C, YANG X Q, et al. HetETA: Heterogeneous Information Network Embedding for Estimating Time of Arrival[C]// Proceedings of the 26th ACM SIGKDD Conference on Knowledge Discovery and Data Mining. 2020: 2444-2454.

[4] SEBASTIAN Y, SIEW E, ORIMAYE S O. Learning the heterogeneous bibliographic information network for literature-based discovery[J]. Knowledge Based Systems, 2017, 115: 66-79.

[5] ZHOU H, ZHAO Z Y, LI C. Survey on Representation Learning Methods Oriented to Heterogeneous Information Network[J]. Journal of Frontiers of Computer Science and Technology, 2019, 13(7): 1081-1093.

[6] DING Y, WEI H, PAN Z S, et al. Survey of network representation learning[J]. Computer Science, 2020, 47(9): 52-59.

[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. 2014: 701-710.

[8] MIKOLOV T, CHEN K, CORRADO G, et al. Efficient Estimation of Word Representations in Vector Space[C]// Proceedings of the 1st International Conference on Learning Representations. 2013.

[9] MIKOLOV T, SUTSKEVER I, CHEN K, et al. Distributed Representations of Words and Phrases and their Compositionality[C]// Proceedings of the 26th International Conference on Neural Information Processing Systems. 2013: 3111-3119.

[10] 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. 2016: 855-864.

[11] TANG J, QU M, WANG M Z, et al. LINE: Large-scale Information Network Embedding[C]// Proceedings of the 24th International Conference on World Wide Web. 2015: 1067-1077.

[12] HAMILTON W, YING Z T, LESKOVEC J. Inductive Representation Learning on Large Graphs[C]// Advances in Neural Information Processing Systems 30: Annual Conference on Neural Information Processing Systems. 2017: 1024-1034.

[13] VELICKOVIC P, CUCURULL G, CASANOVA A, et al. Graph Attention Networks[C]// Proceedings of the 6th International Conference on Learning Representations. 2018.

[14] DONG Y X, CHAWLA N V, SWAMI A. metapath2vec: Scalable Representation Learning for Heterogeneous Networks[C]// Proceedings of the 23rd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. 2017: 135-144.

[15] FU T Y, LEE W C, LEI Z. HIN2Vec: Explore Meta-paths in Heterogeneous Information Networks for Representation Learning[C]// Proceedings of the 2017 ACM on Conference on Information and Knowledge Management. 2017: 1797-1806.

[16] WANG X, JI H Y, SHI C, et al. Heterogeneous Graph Attention Network[C]// The 28th World Wide Web Conference. 2019: 2022-2032.

[17] LIAO L Z, HE X N, ZHANG H W, et al. Attributed Social Network Embedding[J]. IEEE Transactions on Knowledge and Data Engineering, 2018, 30(12): 2257-2270.

[18] BANDYOPADHYAY S, BISWAS A, KARA H, et al. A Multilayered Informative Random Walk for Attributed Social Network Embedding[C]// Proceedings of the 24th European Conference on Artificial Intelligence. 2020: 1738-1745.

[19] WANG Y Y, DUAN Z H, LIAO B B, et al. Heterogeneous Attributed Network Embedding with Graph Convolutional Networks[C]// Proceedings of the 33rd AAAI Conference on Artificial Intelligence. 2019: 10061-10062.

[20] LE Q, MIKOLOV T. Distributed Representations of Sentences and Documents[C]// Proceedings of the 31th International Conference on Machine Learning. 2014: 1188-1196.

[21] JOULIN A, GRAVE E, BOJANOWSKI P, et al. Bag of Tricks for Efficient Text Classification[C]// Proceedings of the 15th Conference of the European Chapter of the Association for Computational Linguistics. 2017: 427-431.

TANG Qi-you, born in 1996, postgraduate, is a member of China Computer Federation. His main research interests include data mining and network embedding.

ZHANG Feng-li, born in 1963, Ph. D, professor, Ph.D supervisor, is a member of China Computer Federation. Her main research interests include network security, cloud computing and big data analysis.

（责任编辑：喻藜）