

基于图滤波器的符号属性图链路关系预测算法

马 华,姜 伟,陈 明,钟世杰

(湖南师范大学 信息科学与工程学院,湖南 长沙 410081)

摘 要:带节点属性的符号网络在信息学、生物学等多个领域应用广泛,链路符号预测是该类数据分析中的一个热点问题。基于符号图神经网络的模型是该问题的最新有效解决方案,但现有方法几乎均基于社会平衡理论,且未充分利用节点属性。针对以上问题,从图信号处理角度设计了一个符号图神经网络,提出了一种端到端的符号属性图链路预测算法。首先,给出了基于低频和高频信号的带通滤波器的符号图神经网络,用于获得基于符号拓扑图的节点嵌入;其次,构造属性相似性图,利用图卷积网络得到属性相似性图节点嵌入;最后,引入注意力机制,融合符号拓扑图与属性相似性图两种节点表达,并将其输入符号判别器,通过 Adam 优化器训练模型。在三个药物数据集上进行了对比实验与模型设置的影响分析。与典型的符号图卷积网络与符号图谱嵌入,以及最近提出的基于图滤波的符号卷积网络的对比结果表明,该模型在 AUC 与 F1 指标上比最好的基线方法提升了 8.68% 与 10.04%。

关键词:图滤波器;符号属性图;图神经网络;节点嵌入;链路关系预测

中图分类号:TP311

文献标识码:A

文章编号:1673-629X(2023)09-0126-07

doi:10.3969/j.issn.1673-629X.2023.09.019

Link Relationship Prediction Algorithm for Attributed and Signed Graphs Based on Graph Filter

MA Hua,JIANG Wei,CHEN Ming,ZHONG Shi-jie

(School of Information Science and Engineering,Hunan Normal University,Changsha 410081,China)

Abstract:Signed networks with node attributes are widely used in many fields such as informatics and biology. Link sign prediction is a hot topic in this kind of data analysis. The model based on signed graph neural networks is the latest effective solution to this problem, but the existing methods are almost based on social balance theory, and do not make full use of node attributes. Aiming at the above problems, a signed graph neural network was designed from the view of graph signal processing and an end-to-end link relationship prediction algorithm for attributed and signed graphs was proposed based on graph filter. Firstly, based on low frequency and high frequency signals, a signed graph neural network with band-pass filter was proposed to obtain node embedding based on signed topology graph. Secondly, an attributed similarity graph was constructed and node representation was obtained by graph convolution networks. Finally, the attention mechanism was introduced to fuse two kinds of node representation of signed topology graph and attributed similarity graph. Its input was used to train symbol discriminator, and the model was trained by Adam optimizer. Experiments are conducted on three drug datasets, in order to analyze the comparison to baselines and the effect of model settings. Compared with typical models such as signed graph convolution network and signed spectral embedding models, and the newly proposed signed graph filtering-based convolutional network, the AUC and F1 metric values are superior to the best baseline method up to 8.68% and 10.04%, respectively.

Key words:graph filter; attributed and signed graphs; graph neural network; node embedding; link relationship prediction

0 引 言

带属性的符号网络(或称符号属性图)是指节点含有属性信息,且边权带正、负符号的一类网络。链路关系与符号预测是这类网络上的热点问题,在许多领域中

均有应用,如在医学领域中的药物相互作用关系预测、蛋白质复合物检测。图神经网络(graph neural network, GNN)与图信号处理(graph signal processing, GSP)是当前流行的图数据挖掘技术。传统的非图深度学习方

收稿日期:2022-11-14

修回日期:2023-03-15

基金项目:国家自然科学基金面上项目(71971221,62077014);湖南省教育厅科学研究项目(22B0097);长沙市自然科学基金项目(kq2202248)

作者简介:马 华(1979-),男,CCF 高级会员(20482S),教授,研究方向为智能推荐、图数据挖掘;通信作者:陈 明(1983-),女,CCF 会员(19125M),讲师,研究方向为图信号处理。

法,例如,随机游走、张量分解、标签传播和符号谱嵌入(signed spectral embedding, SSE)等,与 GNN 与 GSP 相比,GNN 与 GSP 具备充分的整合图结构和节点特征的强大能力^[1]。GNN 具有端对端的学习框架、随机梯度优化等集成优化器和 API 库等优势,加速了其在图数据挖掘中的成功^[2]。GSP 将节点信息视为图上的信号,利用图论和信号处理理论,融合图结构对信号进行变换与分析^[3]。典型的图卷积网络^[4](graph convolutional networks, GCN)可以解释为 GSP 和卷积网络的结合体。GSP 的迅速发展,为 GNN 这类深度学习技术提供了强大的理论支撑。

这些技术在符号图上有相应的延伸,如符号图卷积网络^[5](signed graph convolutional networks, SGCN)及在此基础上改进的符号图注意力网络^[6](signed graph attention networks, SiGAT)、符号图扩散网络^[7](signed graph diffusion network, SGDNET)等^[8]。但是,它们大多是 GNN 与社会平衡理论结合的产物,用于解决符号图上节点信息的聚合与传播问题。平衡性的判断建立于边符号的累计奇偶性,其使用场景有限,且并未考虑边的权重属性。这些模型的测试案例通常不存在节点属性,因而,不能很好地适用于带节点属性的应用场景。事实上,节点属性(图信号)可能是影响形成图拓扑结构的关键因素。

针对上述问题,该文提出基于图滤波器的符号属性图卷积网络模型(graph filtering-based convolutional network for attributed and signed graphs, ASGFCN)用于链路符号预测。首先,提出基于带通滤波器的符号图卷积网络,利用带通滤波器在谱域上实现符号图上节点信息的聚合与传播,用于获得基于符号图拓扑的节点嵌入。在设计符号图滤波器时,同时考虑了低频与高频信号,并通过参数化频率响应函数,提高了模型的自适应学习能力。其次,考虑到节点属性可能是图拓扑的关键来源,构造属性相似性图并利用图卷积网络得到另一种节点表达。最终利用注意力机制融合这两种表达,将其输入符号判别器,通过 Adam 优化器训练模型。在多个数据集上与基线方法进行的对比分析表明,基于 ASGFCN 的链路预测算法具有更好的性能,其 AUC 与 F1 指标分别比最好的基线方法提升了 8.68% 与 10.04%。

1 符号属性图的链路关系预测问题

用 $G = \{V, E, X\}$ 表达符号属性图,其中, $V = \{v_1, v_2, \dots, v_n\}$ 是由 n 个节点组成的集合; E 表示边集,连接节点 $v_i, v_j \in V$ 的边记为 e_{ij} ,且 $e_{ij} \in \{+, -\}$; $X \in R^{n \times d}$ 为节点的特征矩阵,其中 n 表示特征维度。由边集 E 可以得到邻接矩阵 A ,矩阵元素 A_{ij} 表示节点对 v_i

和 v_j 的带符号边权。

该文考虑以下应用场景中的符号图链路或符号预测问题:药物相互作用(drug-drug interactions, DDIs)符号预测与药物副作用识别预测。

(1)符号 DDIs 预测。已知一批药物以及它们之间的一些积极或消极关系,对于关系未知的药物对,确定其链接关系。如图 1 所示, $? \in \{+, -, 0\}$,正号和负号分别表示两种药物组合时的积极或消极作用。

(2)药物副作用识别。已知一批药物以及它们之间的副作用,对于关系未知的药物对,确定药物节点之间是否存在副作用,如图 1 所示, $? \in \{-, 0\}$ 。

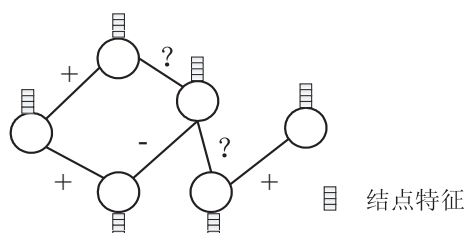


图 1 符号图上的链路关系预测问题

2 相关工作

符号属性图上的链路预测方法一般可以分为基于图深度学习的方法和基于非图深度学习的方法。近十年来,符号属性图上的链路预测方法主要以非图深度学习方法(特别是谱方法)为主导,如符号拉普拉斯矩阵方法^[9]及其变式^[10]、平衡正则割^[11]与 SSE^[12]和一些其他的方法^[13-15]。这些方法通常是二阶段式的:首先抽取特征,然后是关系判别。

与这些浅层模型相比,图深度学习方法构建了联合求解图节点嵌入及下游任务的端到端学习框架,能实现更好的预测效果。GNNs 是当前流行的用于图结构数据分析的深度学习方法。该类方法能较好地利用节点特征与网络结构,是端到端训练模型,且拥有成熟的训练工具箱^[16]。

当前一些 GNNs 在符号图上进行了扩展,它多为现有模型与社会平衡理论的结合。例如,SGCN^[5]基于社会平衡理论提出一种新的符号图节点嵌入方法,符号图注意力网络^[6](signed graph attention network, SiGAT)与基于注意力的符号网络嵌入^[17](signed network embedding based on attention, SNEA)也是在 SGCN 基础上基于平衡理论所提出的方法。另一类模型则通过扩展符号图 GSP 理论获得启发。例如,Chen 等人提出以基于图滤波的符号卷积网络(signed graph filtering-based convolutional network, SGFCN)来表示药物^[18-19],节点特征被视为图信号通过专门的谱滤波器进行传递和聚合,该图滤波器可捕捉药物对的同配性和异配性。

相较非深度学习方法而言,符号图 GNNs 方法集成了节点特征与网络结构,且是端到端模型,因此对符号图数据的挖掘能力有较大提升。但仍具备一些不足:(1)它们所基于的平衡理论在许多场景中并不适用,且该理论并未考虑边的权重属性。(2)这些模型的测试场景并未考虑到节点的多通道属性,对节点属性的利用程度有限。事实上,节点属性(图信号)可能是边的符号属性的关键性潜在影响因素。

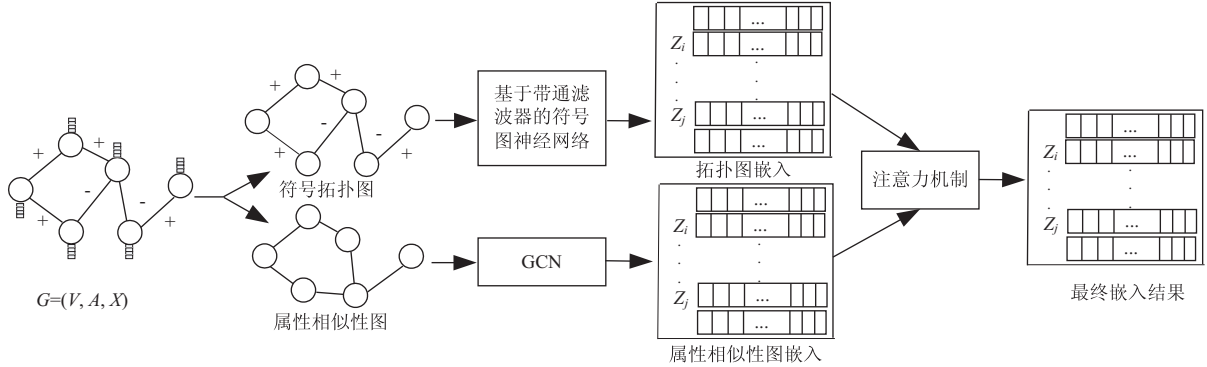


图 2 基于图滤波器的符号图卷积网络

3.1 基于带通滤波器的符号图神经网络

3.1.1 GCN 以及其 GSP 视角

GNN 通过多层叠加得到节点嵌入结果,每层包含特征转换、消息聚合两个算子。不同的 GNN 模型采用不同的聚合操作。

图 G 上的消息聚合过程可视为对图信号的滤波操作。用 H 表示图滤波器, $y = Hx$ 将定义在 n 个节点上的图信号 $x \in R^n$ 变换为另一个图信号。 H 可通过在整个图的谱域内调整频率强度得到。首先,选择一个图矩阵进行谱特征分解。设 $M \in R^{n \times n}$ 是定义在 G 上的实对称半正定矩阵,对其进行分解: $M = U \text{diag}(\lambda_1, \lambda_2, \dots, \lambda_n) U^T$, 特征值 $\lambda_i \geq 0$ 被称为图频率, $U = \{U_1, U_2, \dots, U_n\}$ 包含相应的特征向量作为列。在 GSP 中 U 用作傅里叶基,将一个顶点域图信号映射至谱域。 U^T 则用于重建信号。然后,定义以下的图滤波器:

$$H = U \text{diag}(h(\lambda_1), \dots, h(\lambda_n)) U^T \quad (1)$$

这里需选择频率响应函数 $h(\lambda)$, 调整谱域信号。以 GCN 为例,每层的计算方法为:

$$X^{(l+1)} = \sigma(\hat{A} X^{(l)} \theta_{GCN}^{(l)}) \quad (2)$$

其中, l 为层数。 $\hat{A} = \bar{D}^{1/2} \bar{A} \bar{D}^{1/2}$, 负责节点消息的聚合, $\bar{A} = A + I$ 为加入自循环的邻接矩阵, \bar{D} 为 \bar{A} 的度矩阵。 $\theta_{GCN}^{(l)}$ 是第 l 层的权重矩阵,负责对节点消息特征变换。 $H = \hat{A}$ 解释为由 $\bar{L} = I - \bar{A}$ 特征分解,由此定义频率响应函数为 $h(\lambda) = 1 - \lambda$ 的图滤波器。

3.1.2 基于带通滤波器的符号图神经网络

无符号图的谱域分析已被广泛用于开发 GNN 体

3 基于 ASGFCN 的符号属性图节点嵌入

针对符号属性图上的链接预测任务所提出的 ASGFCN,其整体流程见图 2。分别在符号拓扑图与属性相似图上进行节点嵌入,引用注意力机制将两者融合用于下游链路预测。针对符号图节点嵌入,设计了基于带通滤波器的符号图神经网络,并讨论低高频的作用。对于节点属性,构建相似性网络,利用 GCN 节点嵌入。下面分别阐述节点嵌入过程与链路关系预测。

系结构。许多众所周知的 GNN,如 GCN、适应性图卷积神经网络^[20] (adaptive graph convolutional neural networks, AGCN) 等,都依赖于谱域分析。这些设计基于从规范化无符号图拉普拉斯特征分解得出的频率解释。然而,将现有的谱域 GNN 设计直接应用于符号图存在局限性,主要包括:(1)度矩阵中可能的零对角项使得拉普拉斯函数的规范化变得不可行。(2)图拉普拉斯可能出现负特征值,使得频率排序时难以准确决定将特征值的最小负值、最小正值或最小绝对值视为低频值。

为此,考虑无符号图拉普拉斯修正的定义: $\bar{L} = \bar{D} - A$ 。这里的 \bar{D} 为符号度,定义为: $d_i = \sum_j |A_{ij}|$, d_i 为符号网络第 i 个节点的度。与 GCN 一致,文中所提方法使用半正定的对称规范化符号拉普拉斯矩阵进行特征分解,该矩阵定义为: $\bar{L}_{sym} = \bar{D}^{-1/2} \bar{L} \bar{D}^{-1/2}$ 。

符号图谱聚类通过 \bar{L}_{sym} 特征分解得到前 K 个最小特征值对应的特征向量捕获符号图上节点间的正负边关系。受此启发,可设计一个使较小特征值发挥更大作用的低通滤波器。低通滤波器与谱聚类相似,它仅考虑前 K 个特征值对应的特征向量,理想型的低通滤波器的频率响应函数定义如下:

$$h_L(\lambda_i) = \begin{cases} > 0, & i \leq K \\ 0, & \text{others} \end{cases} \quad (3)$$

无符号图上较大特征值对应的特征向量为高频信号,在节点嵌入时也能发挥作用,用于获取异质关系所隐含的信息^[21]。在后面的实验中尝试同时考虑低频

与高频信号的组合。用以下函数来表示这种带通滤波器:

$$h_B(\lambda_i) = \begin{cases} 0, & k_l < i < k_h \\ \theta_\lambda \geq 0, & \text{others} \end{cases} \quad (4)$$

其中, θ_λ 是可学习的参数, Relu 函数确保其非负性。

将节点原始矩阵 X 视为广义图信号, 堆叠两层 GNN 得到符号拓扑图节点嵌入结果, 计算方法为:

$$Z_A = H_{th} \sigma(H_{th} X \theta_A^{(0)}) \theta_A^{(1)} \quad (5)$$

其中, $\theta_A^{(\cdot)}$ 表示每层的可训练权重矩阵, $\sigma(\cdot)$ 表示非线性激活函数, 如双曲正切函数等。

3.2 基于属性相似性图的节点嵌入

符号属性图上的节点可能具有多样化属性, 如药物相互作用网络上药物节点具有蛋白质靶标特征与副作用特征等多种属性。这些节点属性是产生链接关系的潜在因素。为了挖掘节点属性的潜力, 通过计算节点间的特征相似性来构造属性相似性图, 利用 GCN 产生一个辅助的节点嵌入结果。

首先, 采用药物网络中常用的杰卡德系数产生一个相似性图, 计算节点 i 与节点 j 间的特征相似性得分 S_{ij} 。然后, 保留阈值 μ 之上的连边。最后, 对相似性矩阵 S 采用 GCN, 堆叠两层得到无符号的属性相似图节点嵌入结果, 方法如下:

$$Z_{AF} = \sigma(\hat{S} \tanh(\hat{S} X \theta_{AF}^{(0)}) \theta_{AF}^{(1)}) \quad (6)$$

3.3 基于注意力机制的节点嵌入融合

对 Z_A 与 Z_{AF} 进行加权求和, 将两者融合:

$$Z = \alpha_A Z_A + \alpha_{AF} Z_{AF} \quad (7)$$

其中, $\alpha_A, \alpha_{AF} \in R^{n \times 1}$ 为对应于 Z_A 和 Z_{AF} 的注意力系数。节点 i 在 Z_A 的嵌入结果为 Z_{Ai} 。首先通过一层神经网络学习 Z_{Ai} 初始注意力, 然后使用共享的注意力向量 $q \in R^{p \times 1}$ 获得注意力值 W_{Ai} , 计算方法如下:

$$W_{Ai} = q^T \cdot \tanh(\theta_{ATT} \cdot (Z_{Ai})^T + b_A) \quad (8)$$

其中, θ_{ATT} 为权重矩阵, b_A 为对应 Z_A 的偏置向量。同理可得 Z_{AF} 对应节点 i 的注意力值 W_{AFi} 。然后, 使用 softmax 函数将注意力值 W_{Ai} 和 W_{AFi} 整合获得最终权重, 方法如下:

$$\alpha_{Ai} = \text{softmax}(W_{Ai}) = \frac{\exp(W_{Ai})}{\exp(W_{Ai}) + \exp(W_{AFi})} \quad (9)$$

同理:

$$\alpha_{AFi} = \text{softmax}(W_{AFi}) \quad (10)$$

最后, 将式(9)与式(10)推广至任意节点即得式(7)。

4 基于 ASGFCN 的链路预测算法

本节提出了基于 ASGFCN 的端到端链路关系预测算法, 并将其应用于药物网络链路预测问题中。

为采用端到端的方式训练整个预测框架, 采用回归型判别器对 v_i 与 v_j 的嵌入结果 (Z_i, Z_j) 进行分类。 $\theta^R = \{\theta_1^R, \theta_2^R, \dots, \theta_{|C|}^R\}$ 包含回归系数, 其中 θ_k^R 表示第 k 类边类型的系数。ASGFCN 网络参数包含符号拓扑图节点嵌入的权重矩阵 $\theta_A^{NN} = (\theta_A^{(0)}, \theta_A^{(1)})$ 与属性相似性图节点嵌入的权重矩阵 $\theta_{AF}^{NN} = (\theta_{AF}^{(0)}, \theta_{AF}^{(1)})$ 。使用 $t_{ij} \in T$ 表示节点 v_i 和 v_j 之间的连边类型, T 取决于具体的预测任务。损失函数定义如下:

$$L(\theta_A^{NN}, \theta_{AF}^{NN}, \theta^R) = \sum_{ij} - \theta_{C_{ij}} \sum_{t \in C} \Pi(t_{ij} = t) \log \frac{\exp([Z_i \parallel Z_j] \theta_t^R)}{\sum_{q=1}^{|T|} ([Z_i \parallel Z_j] \theta_q^R)} \quad (11)$$

其中, $\theta_{C_{ij}}$ 表示与连边类型 t_{ij} 相关联的权重, 如果给定的预测为真, 则 $\Pi(\cdot)$ 返回 1, 其余则返回 0。

对第2节介绍的两个预测问题, T 的设定不同。对 DDIs 符号预测任务, 使用“0”边作为负样本训练模型, $T \in \{+, -, 0\}$ 。对第2节介绍的药物副作用判别任务, 使用“+”边作为负样本, $T \in \{-, 0\}$ 。所提算法流程如图3所示。

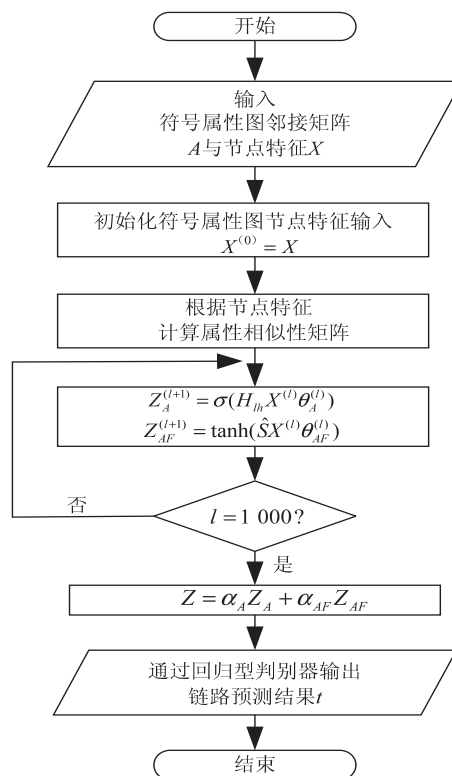


图3 算法流程

5 实验设计与分析

将所提算法与多个基线进行了对比分析。

5.1 数据集

使用了3种数据集, 如表1所示。数据集简介如下:(1)D-1562。Shi 等人^[22]从 DrugBank V4 数据库收集得到。包括 180 576 个已知的 DDIs, 其中包含

125 298 个正链接和 55 278 个负链接。连边的符号是根据 DrugBank 数据库中的语义描述来划分的。药物自身属性特征包含药物结合蛋白和化学结构。(2) D-548。药物副作用数据集,由 Liu 等人^[23]收集。它包括 548 种药物,只有化学结构属性。其中 48 584 种药物副作用(负连接)是从 TWOSIDES 得到的,仅表示两

种药物间是否存在副作用,其 8 299 种正连接是通过特征相似性计算得到。(3) D-597。该数据集的提取方式与 D-1562 相同,但来自于 DrugBank V5。其原始数据包括 Wang 等人^[24]收集的 614 种小分子药物。删除了几个孤立的无邻居节点药物,得到了包含 597 个药物的数据集。

表 1 实验使用的数据集

数据集	节点数量	连边数 (总数/+/-)	平均度 (总数/+/-)
D-1562	1 562	180 576/125 298/55 278	231.2/60.4/70.8
D-548	548	56 883/8 299/48 584	207.6/30.3/177.3
D-597	597	80 702/40 202/40 500	270.4/134.7/135.7

5.2 实验设置

每次实验运行都采用 5 折交叉验证,所有结果为 5 次运行的平均值,并选取两种常用的评价指标受试者工作特性曲线(AUC)下面积和 F1 度量,值越大,说明模型性能表现越好。

设定的基线方法如下:(1) SGCN^[5],它是一种基于社会平衡理论的符号图节点嵌入方法。(2) SGFCN^[18-19],它将药物关系网络看作是同配图和异配图的重叠,节点特征被当作图信号通过专门的谱滤波器进行传递和聚合。(3) SSE^[12],它是用于符号网络嵌入问题的基线方法,它将符号网络映射到一个由符号图拉普拉斯特征向量组成的空间中,它仅集成了网络结构,且是二阶段式。(4) A-LHW-C。它是 ASGFCN 的变体,与 ASGFCN 相比,它仅利用单通道化学结构特征构成属性相似性图。(5) A-L。ASGFCN 的变体,它仅考虑低频信号。(6) A-LH。ASGFCN 的变体,它同时考虑低频与高频信号。

所有算法都基于 PyTorch 实现,且均使用两个网络层和一个学习率为 0.01 的 Adam 优化器。为了保证带通滤波器同等考虑低通与高通信号,式(4)的 k_l

与 k_h 皆设为 $K/2$ 。使用不同的迭代次数训练模型,发现迭代 1 000 次足以获得良好的结果。

5.3 实验结果和分析

本节旨在回答四个关键问题:(1) ASGFCN 与基线方法对比结果;(2) 高频信号是否影响模型预测效果;(3) 属性相似性图的作用;(4) 带通滤波器中傅里叶基数量是否影响 ASGFCN 性能。具体内容如下:

(1) ASGFCN 与基线对比实验。

实验设置文中所提方法在三个数据集上与其他基线的对比,得到表 2 的统计结果。表 2 中指标纵向对比,较大者以粗体突出表示。从表中可以看出,SSE 的指标在三个数据集上最低,ASGFCN 的指标在两个数据集上明显高于其它基线。在 D-548 上与 SGFCN 和 SGCN 几乎相当。通过分析实验结果,发现 ASGFCN 性能显著,在多个数据集上都优于基线。具体而言,在 AUC 指标上优于基线 11.19%,在 F1 指标上优于基线 12.11%。在 D-548 上,ASGFCN 与 SGFCN 和 SGCN 效果几乎相当,这是由于 D-548 数据集本身未统计多样化属性。SSE 效果最差,这是其本身未利用网络结构所导致的。

表 2 ASGFCN 与基线对比结果

模型	D-1562		D-548		D-597	
	AUC	F1	AUC	F1	AUC	F1
SGCN	0.91	0.833	0.956	0.957	0.842	0.758
SGFCN	0.91	0.834	0.958	0.959	0.841	0.757
SSE	0.87	0.799	0.922	0.904	0.822	0.743
ASGFCN	0.953	0.884	0.956	0.957	0.914	0.833

(2) ASGFCN 变体 A-L 与变体 A-LH 对比实验。

实验设置基于 ASGFCN 变体 A-L 或 A-LH 的链路预测算法在三个数据集的性能比较,A-L 仅考虑低频信号,A-LH 考虑低频与高频信号。统计结果如表 3 所示。表 3 展示了基于 A-L 或 A-LH 的链路预测算法的 AUC 指标与 F1 指标在三个数据集对比情况。

通过分析实验结果,发现额外考虑高频信号的 A-LH 与 A-L 相比,有更好的性能。A-LH 与 A-L 之间的效果差距在 D-548 达到最大,D-1562 次之,D-597 差距最小。对 D-548 数据集,A-LH 的 AUC 指标优于 A-L 1.32%,F1 指标则优于 A-L 1.80%。对 D-597 数据集,A-LH 的 AUC 指标相对 A-L 也达到了

0.53%的提升,F1指标上A-LH也有1.05%的提升。这表明在无符号图中表征节点间差异性的高频信号在

符号图中仍可表征节点间某种特殊关系,且此关系影响信号在符号图上的传播。

表3 基线对比结果高频信号对ASGFCN的影响

模型	D-1562		D-548		D-597	
	AUC	F1	AUC	F1	AUC	F1
A-L	0.946	0.879	0.939	0.951	0.911	0.833
A-LH	0.953	0.884	0.944	0.961	0.923	0.848

(3) ASGFCN与变体A-LHW-C对比实验。

实验比较所提方法与A-LHW-C在两个数据集的性能,结果如表4所示。A-LHW-C与所提方法相比,仅考虑单通道化学结构特征构成属性相似性图。

表4 属性相似性图的影响

模型	D-1562		D-597	
	AUC	F1	AUC	F1
A-LHW-C	0.902	0.831	0.839	0.757
ASGFCN	0.953	0.884	0.914	0.833

表4展示了所提方法与A-LHW-C的AUC指标与F1指标在两个数据集的对比情况,由于D-548无多样化属性,故不将其加入对比。从表中可以看出,该文提出的算法在两个数据集上指标都明显大于对比的变体,在AUC指标上,ASGFCN比仅利用化学结构作为特征属性的A-LHW-C提高了5.65%。通过分析实验结果,发现利用多通道节点属性的ASGFCN在性能上明显优于仅利用单通道节点属性的A-LHW-C。这表明节点属性会影响边的符号属性,而ASGFCN模型更好地利用了节点多通道属性,也就捕获到了边符号属性的影响因素。

(4) ASGFCN的带通滤波器中K取值对比实验。

将D-1562作为图数据放入基于ASGFCN的链路预测框架,决定傅里叶基数量的K值分别取值为{8, 16, 32, 64, 128},设置对比实验,结果如图4所示。

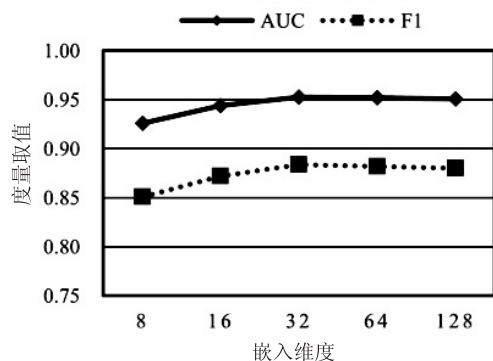


图4 不同傅里叶基数量的影响

图4表示基于ASGFCN模型的链路预测算法性能随带通滤波器傅里叶基数量变化的结果折线图。实线与虚线分别代表AUC指标与F1指标。从图中可以看出折线总体波动情况不大,呈现先上升后缓慢下降

的趋势,K=32时,折线处于最高点。分析实验结果得到ASGFCN带通滤波器中傅里叶基数量对模型性能影响较小的结论。为了充分发挥ASGFCN性能,ASGFCN模型及变体设定K=32。

6 结束语

针对现有的符号图神经网络模型主要基于社会平衡理论且未充分利用节点属性的问题,提出一种基于图滤波器的符号属性图链路预测算法。从新的视角设计了兼顾低高频信号的符号图滤波器,并通过参数化频率响应函数,提高了ASGFCN模型的自适应学习能力。构造属性相似性图并利用图卷积网络得到另一种节点表达,并通过注意力机制融合符号拓扑图与属性相似性图节点嵌入,进一步挖掘节点属性所蕴涵的信息。基于多个数据集的对比实验验证了ASGFCN链路预测算法具有良好的性能。未来将进一步尝试合适的符号图拉普拉斯矩阵,为符号属性图定制图谱滤波器,并丰富实验集种类。

参考文献:

- [1] HE Y X. GNNs for node clustering in signed and directed networks[C]//Proc of ACM international conference on web search and data mining (WSDM). New York: ACM, 2022:1547-1548.
- [2] 吴 博,梁 循,张树森,等. 图神经网络前沿进展与应用[J]. 计算机学报,2022,45(1):35-68.
- [3] ORTEGA A, FROSSARD P, KOVACEVIC J, et al. Graph signal processing: overview, challenges, and applications[J]. Proc of the IEEE, 2018, 106(5):808-828.
- [4] ZITNIK M, AGRAWAL M, LSKOVEC J. Modeling polypharmacy side effects with graph convolutional networks[J]. Bioinformatics, 2018, 34(13):i457-i466.
- [5] DERR T, MA Y, TANG J L. Signed graph convolutional networks[C]//Proc of IEEE international conference on data mining. [s. l.]: IEEE, 2018:929-934.
- [6] HUANG J J, SHEN H W, HOU L, et al. Signed graph attention networks[C]//Proc of international conference on artificial neural networks. Berlin: Springer, 2019:566-577.
- [7] JUNG J, YOO J, KANG U. Signed graph diffusion network[EB/OL]. (2020-12-28)[2022-07-07]. <https://arxiv.org/abs/2012.12345>.

- org/abs/2012.14191.
- [8] 刘先锋,石 静,陈 明,等. 基于符号规范化割的半监督学习与图像分割[J]. 小型微型计算机系统,2020,41(7): 1547–1552.
 - [9] KUNEGIS J, SCHMIDT S, LOMMATZSCH A, et al. Spectral analysis of signed graphs for clustering, prediction and visualization[C]//Proc of SIAM international conference on data mining. Philadelphia: SIAM, 2010: 559–570.
 - [10] MERCADO P, TUDISCO F, HEIN M. Clustering signed networks with the geometric mean of Laplacians[C]//Proc of advances in neural information processing systems. New York: Curran Associates Inc, 2017: 4421–4429.
 - [11] CHIANG K Y, WHANG J J, DHILLON I S. Scalable clustering of signed networks using balance normalized cut[C]//Proc of ACM international conference on information and knowledge management. New York: ACM, 2012: 615–624.
 - [12] ZHENG Q, SKILLICORN D B. Spectral embedding of signed networks[C]//Proc of SIAM international conference on data mining. Philadelphia: SIAM, 2015: 55–63.
 - [13] CUCURINGU M, DAVIES P, GLIELMO A, et al. Sponge: a generalized eigen problem for clustering signed networks[C]//Proc of the 22nd international conference on artificial intelligence and statistics. Brookline: Microtome Publishing, 2019: 1088–1098.
 - [14] WANG S H, AGGARWAL C, TANG J L, et al. Attributed signed network embedding[C]//Proc of ACM on conference on information and knowledge management. New York: ACM, 2017: 137–146.
 - [15] YUAN S H, WU X T, XIANG Y. SNE: signed network embedding[C]//Proc of Pacific Asia conference on knowledge discovery and data mining. Berlin: Springer, 2017: 183–195.
 - [16] 马 帅, 刘建伟, 左 信, 等. 图神经网络综述[J]. 计算机研究与发展, 2022, 59(1): 47–80.
 - [17] LI Y, TIAN Y, ZHANG J W, et al. Learning signed network embedding via graph attention[C]//Proc of AAAI conference on artificial intelligence. New York: AAAI, 2020: 4772–4779.
 - [18] CHEN M, PAN Y, JI C Y. Predicting drug drug interactions by signed graph filtering – based convolutional networks[C]//Proc of international symposium on bioinformatics research and applications. Berlin: Springer, 2021: 375–387.
 - [19] CHEN M, JIANG W, PAN Y, et al. SGFNNs: signed graph filtering–based neural networks for predicting drug – drug interactions[J]. Journal of Computational Biology, 2022, 29(10): 1104–1116.
 - [20] LI R Y, WANG S, ZHU F Y, et al. Adaptive graph convolutional neural networks[C]//Proc of AAAI conference on artificial intelligence. New Orleans: AAAI, 2018.
 - [21] BO D Y, WANG X, SHI C, et al. Beyond low–frequency information in graph convolutional networks[C]//Proc of AAAI conference on artificial intelligence. California: AAAI, 2021: 3950–3957.
 - [22] SHI J Y, MAO K T, YU H, et al. Detecting drug communities and predicting comprehensive drug–drug interactions via balance regularized semi – nonnegative matrix factorization[J]. Journal of Cheminformatics, 2019, 11(1): 1–16.
 - [23] LIU T R, CUI J C, ZHUANG H, et al. Modeling polypharmacy effects with heterogeneous signed graph convolutional networks[J]. Applied Intelligence, 2021, 51(11): 8316–8333.
 - [24] WANG F, LEI X J, LIAO B, et al. Predicting drug–drug interactions by graph convolutional network with multi–kernel[J]. Briefings in Bioinformatics, 2022, 23(1): bbab511.