

一种基于增强结构嵌入的时序链路预测算法

杨育捷^{a,b}, 廖舒蕾^a, 王李明^a, 刘栋^{a,b}

(1.河南师范大学 a.计算机与信息工程学院;b.河南省教育人工智能与个性化学习重点实验室,河南 新乡 453007)

摘要:本文提出一种基于增强结构嵌入的时序链路预测算法(enhanced structural embedding algorithm for temporal link prediction, SE-TLP). SE-TLP设计双通道结构特征提取机制用于建模节点和连边特性,其联合4种加权节点中心性衡量节点地位,采用连边关系强度表示拓扑结构,并融合节点和连边特征以增强结构嵌入表示,提高预测准确性.同时,结合图卷积网络和门控循环单元,沿时间维度学习网络结构的演化.实验表明,SE-TLP在两类评价指标上均优于主流算法,性能分别提升7.47%和4.43%.

关键词:时序链路预测;网络嵌入;图卷积网络;门控循环单元(GRU);动态网络

中图分类号:O157.5;TP183

文献标志码:A

文章编号:1000-2367(2026)04-0075-08

动态网络通过刻画节点间的连接关系随时间的变化过程,揭示复杂系统中实体交互的动态特性^[1].时序链路预测作为动态网络分析的核心任务,旨在通过学习网络历史快照的结构和时序特征,预测网络的未来连接.网络的节点和连边信息是解析网络行为的重要补充,有助于深化对连接形成机制的理解.节点中心性^[2]是量化节点在网络中的重要性,中心性高的节点之间更容易形成连接.拓扑相似性则衡量节点间的结构关联强度,相似性高的节点对通常具有更强的连接倾向.现有研究借助节点中心性和拓扑相似性,在时序链路预测任务中取得进展.例如,Zhang等^[3]利用中心性区分目标节点对的目标邻居对链接形成的贡献,提出一种将节点中心性与时间序列相结合的时序链路预测方法.Choudhury等^[4]通过非连接节点对之间的时间相似性来构建动态相似性指标,并结合中心性挖掘节点的网络位置变化和邻域变化.因此,合理使用节点中心性和拓扑相似性能够提升链路预测模型对网络结构的建模能力,提高链路预测准确性.传统基于图论的方法在图学习能力和效率上存在局限,相关研究转向基于深度学习的图学习方法,通过神经网络和网络嵌入技术捕捉动态网络结构随时间的演化规律,生成低维嵌入表示以实现时序链路预测.但现有的基于深度学习的链路预测方法过于依赖神经网络模型对图数据的表征能力,忽略了节点和连边特征对链路预测的驱动作用,致使对动态网络结构的刻画过于片面.

针对上述问题,本文提出一种基于增强结构嵌入的时序链路预测算法(SE-TLP).SE-TLP算法利用节点中心性与拓扑相似性,设计一种双通道提取节点和连边特征机制实现网络结构的全面表征,具体而言:1)节点特征通道聚合4种加权中心性,动态表征网络节点地位演化;2)连边特征通道采用改进的资源分配(Resource allocation, RA)指标^[5],精准刻画连边关系强度变化. SE-TLP算法融合节点和连边特征,以增强网络结构的初始嵌入表示质量.同时,SE-TLP构建基于图卷积网络(Chebyshev graph convolutional network,

收稿日期:2025-04-10;**修回日期:**2025-05-06.

基金项目:国家自然科学基金(62072160);河南省科技攻关项目(252102210141);河南省国际科技合作项目(262102521055).

作者简介(通信作者):杨育捷(1978—),女,河南新乡人,河南师范大学副教授,博士,研究方向为链路预测、复杂网络, E-mail: yangyujie@htu.edu.cn.

引用本文:杨育捷,廖舒蕾,王李明,等.一种基于增强结构嵌入的时序链路预测算法[J].河南师范大学学报(自然科学版),2026,54(4):75-82.(Yang Yujie, Liao Shulei, Wang Liming, et al. An enhanced structural embedding algorithm for temporal link prediction[J]. Journal of Henan Normal University(Natural Science Edition), 2026, 54(4): 75-82. DOI:10.16366/j.cnki.1000-2367.2025.04.10.0003.)

ChebNet)、门控循环单元(gate recurrent unit, GRU)和全连接层神经网络(fully connected layer, FCL)的编码器-解码器框架,其中, ChebNet 负责学习网络空间结构特征, GRU 负责捕捉时间序列间的时间依赖性, FCL 层负责输出具有时空一致性的嵌入表示. SE-TLP 算法不仅有效保留网络节点中心性特征和连边关系强度特征还显著提升了动态网络嵌入的学习效果,进而实现更精准的时序链路预测.

1 相关工作

网络嵌入^[6]是一种将高维网络映射到低维向量空间的表示学习方法,可将复杂的网络分析任务转化为机器学习任务,常用于时序链路预测研究中.通常,网络嵌入方法可分为基于矩阵分解、随机游走和深度学习 3 类.基于矩阵分解的方法通过分解相似性矩阵生成节点嵌入实现链路预测.基于随机游走的方法利用随机游走序列捕获节点邻域拓扑信息,生成节点向量表示.尽管这两类方法具有框架简单、可解释性强的特点,但在大规模网络中面临扩展性不足和预测精度低的问题.基于深度学习的方法利用神经网络模型从历史快照中提取网络结构和时序信息,将高维信息映射到低维空间,提高动态图数据处理效率.图卷积网络(graph convolutional network, GCN)^[7]引入卷积核聚合节点及其邻域信息,生成节点嵌入表示,专门用于图数据分析. Mei 等^[8]采用 GCN 学习网络空间结构特征,解决动态网络的链接预测问题.然而, GCN 难以建模时间序列的动态变化,无法捕捉动态网络中结构演化规律.循环神经网络(recurrent neural network, RNN)及其变体,长短期记忆网络(long short-term memory, LSTM)和门控循环单元(GRU)^[9]通过循环结构和门控机制控制历史信息 and 当前信息的融合,擅长处理时间序列之间的依赖性,常用于学习动态网络的时序特性.例如, Pareja 等^[10]提出 EvolveGCN 模型,结合 GRU 动态调整 GCN 参数以捕捉图序列的动态性.这些方法有效解决了动态网络时空特征建模问题.

综上,基于矩阵分解方法和基于随机游走方法因时序建模能力不足,难以有效捕捉动态网络特性,预测精度较低且可扩展性差.基于深度学习的方法虽然能生成具有高表征的网络嵌入表示,但缺乏对网络节点地位和连边关系特性的动态建模,导致网络拓扑结构信息损失.本文提出的 SE-TLP 算法关注节点和连边特征变化对预测准确性的影响,设计双通道特征提取机制,量化网络节点中心性和连边关系强度以补充网络拓扑结构信息,加强对网络结构的动态表征.同时, SE-TLP 算法设计基于深度学习的编码器-解码器框架,强化对动态图的时空特性的学习能力.

2 基于增强结构嵌入的时序链路预测算法

2.1 定义

定义 1 动态网络(dynamic network)可以表示为一组快照集合,记作 $G = \{G_1, G_2, \dots, G_T\}$, 其中 $G_t = (V_t, E_t)$ 表示 t 时刻下的静态快照图, V_t 和 E_t 分别表示快照 G_t 的节点集合和连边集合.快照 G_t 使用邻接矩阵 A_t 表示节点间的连接关系, A_t 定义为:

$$A_{ij} = \begin{cases} 1, & \text{如果 } (i, j) \in E, \\ 0, & \text{如果 } (i, j) \notin E. \end{cases} \quad (1)$$

式(1)表示如果节点 i 和节点 j 之间存在连边,则 A_{ij} 为 1, 否则 A_{ij} 为 0.

定义 2 时序链路预测(temporal link prediction)定义为从前 T 个历史快照 $\{G_1, G_2, \dots, G_T\}$ 中学习网络特征,预测 $T+1$ 时刻的快照 G_{T+1} 中的连接.时序链路预测可以表示为:

$$G_{T+1} = f(G_1, G_2, \dots, G_T), \quad (2)$$

其中 $f(\cdot)$ 表示实现时序链路预测的算法.

2.2 SE-TLP 算法概述

SE-TLP 算法的整体框架如附录图 S1 所示,包括划分快照集合、双通道网络结构特征提取机制、特征输入、编码器、解码器和输出 6 个模块.首先, SE-TLP 算法将动态网络离散化为快照集合;其次,利用双通道特征提取机制提取快照的结构特征,具体为计算节点中心性和连边连接强度,并融合这两类特征;第三,采用由 ChebNet 和 GRU 组成的编码器,联合学习结构和时序演化特征,生成具有时空一致性的嵌入表示;最后,由

基于 FCL 的解码器将网络表示转化为预测概率矩阵 $\mathbf{P}_{T+1}$. 下面,对 SE-TLP 算法的各个模块做详细介绍.

2.3 双通道网络结构特征提取

由于节点和连边共同构成网络的拓扑表征,因此,特征提取应包含节点与连边. SE-TLP 算法设计双通道网络结构特征提取机制,通过加权节点中心性计算节点特征表示矩阵 $\mathbf{Z}$,基于拓扑相似性计算连边关系强度特征表示矩阵 $\bar{\mathbf{S}}$,并融合 $\mathbf{Z}$ 和 $\bar{\mathbf{S}}$ 生成包含完整网络结构信息的嵌入表示矩阵 $\mathbf{X}$. 融合特征公式如下:

$$\mathbf{X} = \bar{\mathbf{S}} \times \mathbf{Z}. \quad (3)$$

2.3.1 节点特征提取通道

节点中心性作为评价网络节点地位的核心指标,主要有度中心性(degree centrality)、接近中心性(closeness centrality)、介数中心性(betweenness centrality)和特征向量中心性(eigenvector centrality),从重要性、传播性和影响力等不同角度刻画节点中心性特征.由于单一指标对全面分析节点特征存在局限,SE-TLP 算法聚合了 4 种中心性综合评估节点地位.此外,不同中心性在不同的网络中对不同的节点会有评价差异,导致评估结果存在偏差.因此,SE-TLP 算法采用加权聚合策略,为 4 种中心性分配差异化权重.节点特征矩阵 $\mathbf{Z}$ 的计算方法为:

$$Z_{ii} = w_1 \cdot dc_i + w_2 \cdot bc_i + w_3 \cdot cc_i + w_4 \cdot ec_i, \quad (4)$$

其中, w_1, w_2, w_3, w_4 表示赋予中心性的权重值. dc_i, bc_i, cc_i, ec_i 分别表示节点 i 的度中心性、介数中心性、接近中心性和特征向量中心性.

2.3.2 连边特征提取通道

SE-TLP 算法提出一种基于 RA 指标的相似性度量方法用于刻画网络的连边关系强度特征,RA 引入公共邻居度的倒数作为权重,能够有效抑制大度节点对节点相似性的干扰.同时,由于节点对之间的路径越长越容易造成信息冗余.因此,为规避长路径信息冗余问题,SE-TLP 采用节点间最短路径计算节点对之间的关系强度,相似性矩阵 $\bar{\mathbf{S}}$ 的计算如式(5)所示:

$$\bar{S}_{ij} = \begin{cases} 0, & sp_{ij} > 2, \\ \sum_{v \in \Gamma(i) \cap \Gamma(j)} \frac{1}{k_v}, & sp_{ij} = 2, \\ 1, & sp_{ij} = 1, \end{cases} \quad (5)$$

其中, sp_{ij} 表示节点 i 和节点 j 之间的最短路径, $v \in \Gamma(i) \cap \Gamma(j)$ 指节点 v 为节点 i 和节点 j 的共同邻居, k_v 表示节点 v 的度.当节点 i 和 j 之间的最短路径长度大于 2 时,节点间关联较弱视为无相关性,两者的相似性得分为 0;当节点 i 和 j 之间的最短路径长度为 1 时,两者直接连接,则其相似性得分为 1;当节点 i 和 j 之间的最短路径长度为 2 时,节点间存在共同邻居,使用 RA 指标计算相似性得分.

2.4 基于神经网络的编码器-解码器结构

如图 1 所示,SE-TLP 算法使用基于神经网络的编码器-解码器结构,其中,ChebNet 作为编码器组件学习网络结构特征. ChebNet 结合切比雪夫(Chebyshev)多项式和图拉普拉斯算子,通过 Chebyshev 多项式逼近谱域卷积核,减少拉普拉斯矩阵分解造成的计算成本,在空间域实现图卷积运算,卷积核的数学表示为:

$$\mathbf{h}'_{ti} = \sum_{K=0}^K \theta_K T_K(\tilde{\mathbf{A}}), \quad (6)$$

其中, $\mathbf{h}'_{ti}$ 是 ChebNet 层的输出 $\mathbf{H}'_t$ 的第 i 行,表示节点 i 的向量. θ_K 是滤波器的参数, $T_K(\tilde{\mathbf{A}})$ 表示 K 阶切比雪夫多项式在图拉普拉斯算子上 $\tilde{\mathbf{A}}$ 的值, $\tilde{\mathbf{A}} = (2\mathbf{A}/\lambda_{\max}) - \mathbf{I}_n$ 是一个对角矩阵. SE-TLP 将快照 G_t 的结构特征表示 $\mathbf{X}_t$ 作为输入, ChebNet 聚合空间结构信息,生成嵌入表示 $\mathbf{H}'_t$, 计算过程表示为:

$$\mathbf{H}'_t = \text{ChebNet}^K(\mathbf{A}_t, \mathbf{X}_t). \quad (7)$$

此外,作为编码器的重要组成,GRU 采用门控机制,控制历史状态与当前状态信息流的结合,处理动态网络中的时序信息,有效建模网络结构的动态演化过程. $\{\mathbf{H}'_1, \mathbf{H}'_2, \dots, \mathbf{H}'_T\}$ 是快照集合 $\{G_1, G_2, \dots, G_T\}$ 经 ChebNet 处理后的嵌入表示集合,在 GRU 层作为输入,经 GRU 中的更新门(update gate)和重置门(reset gate)生成蕴含时序信息的网络嵌入表示矩阵 $\mathbf{H}_T$,

$$\mathbf{H}_T = \text{GRU}(\mathbf{H}'_1, \mathbf{H}'_2, \dots, \mathbf{H}'_T). \quad (8)$$

GRU 内部运行原理如下:

$$z_{h'_{t,i}} = \sigma(\mathbf{h}'_{t,i} \mathbf{W}^{(z)} + \mathbf{h}_{t-1,i} \mathbf{U}^{(z)}), \quad (9)$$

$$r_{h'_{t,i}} = \sigma(\mathbf{h}'_{t,i} \mathbf{W}^{(r)} + \mathbf{h}_{t-1,i} \mathbf{U}^{(r)}), \quad (10)$$

$$\tilde{\mathbf{h}}_{h'_{t,i}} = \tanh(\mathbf{h}'_{t,i} \mathbf{W}^{(h)} + (r_{t,i} \mathbf{e} \mathbf{h}_{t-1,i} \mathbf{U}^{(h)})), \quad (11)$$

$$\mathbf{h}_{h'_{t,i}} = (1 - z_{x_{t,i}}) \mathbf{e} \mathbf{h}_{t-1,i} + z_{x_{t,i}} \mathbf{e} \tilde{\mathbf{h}}_{t,i}, \quad (12)$$

其中, $z_{h'_{t,i}}, r_{h'_{t,i}}, \tilde{\mathbf{h}}_{h'_{t,i}}, \mathbf{h}_{h'_{t,i}}$ 分别是更新门、重置门、候选隐藏状态和隐藏状态, $\mathbf{W}^{(z)}, \mathbf{U}^{(z)}, \mathbf{W}^{(r)}, \mathbf{U}^{(r)}, \mathbf{W}^{(h)}$ 和 $\mathbf{U}^{(h)}$ 均是权重矩阵, $\sigma(\cdot)$ 是激活函数, $\mathbf{e}$ 表示 Hadamard 乘积. GRU 使用 $z_{h'_{t,i}}$ 和 $r_{h'_{t,i}}$ 控制隐藏状态 $\mathbf{h}_{h'_{t,i}}$ 中的信息流, 式(12)中 $(1 - z_{x_{t,i}}) \mathbf{e} \mathbf{h}_{t-1,i}$ 的表示使用更新门 $z_{h'_{t,i}}$ 选择性遗忘上一时刻隐藏状态 $\mathbf{h}_{t-1}$ 中的信息, $z_{x_{t,i}} \mathbf{e} \tilde{\mathbf{h}}_{t,i}$ 表示记忆当前信息, $\mathbf{h}_{h'_{t,i}}$ 表示经 GRU 处理后的向量表示.

SE-TLP 算法的解码器由全连接层 FCL 组成, 该层通过可训练权重矩阵建立输入与输出之间的全局映射关系. 在 SE-TLP 算法中, 邻接矩阵集合 $\{\mathbf{A}_1, \mathbf{A}_2, \dots, \mathbf{A}_T\}$ 和网络结构特征矩阵集合 $\{\mathbf{X}_1, \mathbf{X}_2, \dots, \mathbf{X}_T\}$ 作为初始输入, 在编码过程中获得具有时空一致性的嵌入表示矩阵 $\mathbf{H}_T$. 然后, $\mathbf{H}_T$ 作为 FCL 层的输入, 被转换成概率矩阵 $\mathbf{P}_{T+1}$, 计算过程如下:

$$\mathbf{P}_{T+1} = \sigma(\mathbf{W}\mathbf{H}_T + \mathbf{b}), \quad (13)$$

其中 $\mathbf{W}$ 和 $\mathbf{b}$ 是解码器中的权重矩阵和偏置向量.

SE-TLP 具体如算法 1 所示:

算法 1 基于增强结构嵌入的时序链路预测算法(SE-TLP)

输入 快照集合 $\{G_1, G_2, \dots, G_T\}$, 邻接矩阵集合 $\{\mathbf{A}_1, \mathbf{A}_2, \dots, \mathbf{A}_T\}$

输出 概率矩阵 $\mathbf{P}_{T+1}$

- 1 for $t=1$ to T do;
- 2 分别获取快照 G_t 的节点特征矩阵 $\mathbf{Z}_t$ 和连边特征矩阵 $\bar{\mathbf{S}}_t$
- 3 融合 $\mathbf{Z}_t$ 和 $\bar{\mathbf{S}}_t$ 生成网络结构特征 $\mathbf{X}_t$
- 4 return $\{\mathbf{X}_1, \mathbf{X}_2, \dots, \mathbf{X}_T\}$
- 5 end for
- 6 根据式(7)获取空间结构嵌入表示 $\{\mathbf{H}'_1, \mathbf{H}'_2, \dots, \mathbf{H}'_T\}$
- 7 根据式(8)获取时序嵌入表示 $\mathbf{H}_T$
- 8 根据式(13)生成 $T+1$ 时刻的概率矩阵 $\mathbf{P}_{T+1}$
- 9 end for

3 实验分析

3.1 实验设置

实验在 6 种不同类型不同规模的动态网络数据集中进行, 包括 Enron^[11], CollegeMsg(CM)^[12], Autonomous Systems(AS)^[13], Email-Eu-Core(Eu)^[14], Bitcoin^[15] 和 Wikipedia^[16].

SE-TLP 算法中 ChebNet 和 GRU 的数量均为 1, 输入维度为网络结构特征矩阵维度, 隐藏层嵌入维度设为 128, K 值取 3 算法通过阈值法将概率矩阵 $\mathbf{P}_{T+1}$ 转换为邻接矩阵 $\mathbf{A}_{T+1}$, 阈值取 0.5. 实验采用 Adam 优化器, 学习率设置为 0.001, 数据集按照 8:2 比例划分训练集和测试集, 训练迭代次数为 150, 每种方法重复运行 5 次取均值以确保结果的稳定性.

SE-TLP 算法与 6 种基准算法进行对比, 分别是: 共同邻居(common neighbors, CN)^[17]、Katz 指标^[18]; GCN^[7]和图注意力网络(graph attention network, GAT)^[19]; GC-LSTM^[20]和 STP-GCN^[21]. 为真实评估预测性能, 实验选用曲线下面积(area under the curve, AUC)和精确度(precision)作为评价指标.

3.2 中心性对链路预测性能的影响和分析

节点中心性用于刻画节点在网络中的作用,是分析网络结构特征的重要工具.实验分别采用度中心性、接近中心性、介数中心性和特征向量中心性提取节点特征进行链路预测,不同中心性在各类动态网络数据集上的 AUC 性能存在显著差异,表明不同中心性对链路预测性能的贡献程度随网络结构特性而变化,单一中心性难以全面表征网络中节点的整体结构特征.SE-TLP 利用 4 种中心性的互补性,根据它们在链路预测任务中的 AUC 性能表现分配差异化权重,并聚合 4 种中心性,中心性 c_n 加权公式如下:

$$w_{c_n} = \frac{AUC_{c_n}}{\sum_{p=1}^4 AUC_{c_p}}, \quad (14)$$

其中, w_{c_n} 表示节点中心性 c_n 的权值, AUC_{c_n} 是基于单一中心性 c_n 的算法的 AUC 值, $\sum_{p=1}^4 AUC_{c_p}$ 是 4 种基于单一节点中心性的算法的 AUC 性能值之和.

SE-TLP 算法在 6 种动态网络数据集集中的 AUC 性能均优于 4 种基于单一中心性的算法,这说明,通过加权聚合 4 种中心性指标,能够更全面量化节点特征信息,从而深化网络结构表征并提升链路预测性能.

3.3 快照数量对 SE-TLP 算法性能和计算效率的影响分析

动态网络基于时间被划分为一系列快照,快照数量决定时间信息的粒度,直接影响时序链路预测性能.本节使用 AUC 和 Precision 评估 SE-TLP 算法在不同快照数量下的预测性能,并记录不同网络在不同快照数量下训练一个轮次(epoch)数据所耗费的时间,研究快照数量影响链路预测性能的规律,结果如表 1 所示.结果表明,SE-TLP 算法的 AUC 和 Precision 性能随快照数量的增加呈稳定上升趋势,这是因为更多的快照数量能够提供更细粒度的时序演化信息,从而提升 SE-TLP 算法学习网络动态特征的效果.当快照数量达到一定值后,增幅减缓,表明新增快照所包含的时序信息逐渐冗余,模型学习到的特征减少,导致性能提升趋缓.该结果表明,SE-TLP 算法在不同时间粒度下均能保持较高的预测性能.

此外,快照数量对 SE-TLP 算法的整体计算效率影响显著.随着快照数量的增加,双通道网络结构特征提取步骤的调用频次、ChebNet 图卷积操作的执行次数以及 GRU 处理的时间序列长度均随之上升,导致 SE-TLP 算法处理动态网络的计算开销上升.表 S1 展示了不同网络在不同快照数量下训练一个 epoch 数据所需的平均时间,结果表明,训练时间随快照数量增加而增长,但是整体训练过程在可接受时间范围内完成.因此,本文综合考虑预测性能和计算效率的平衡,为各数据集选取合适的快照数量,分别是 45、80、60、80、60 和 80,表 S1 中加粗的数值表示数据集在相应快照数量下的性能值和训练时间.

3.4 SE-TLP 算法在不同训练集比例下的性能分析

图 1 为 SE-TLP 算法在不同网络、不同训练集比例 μ 下的 AUC 和 Precision 性能对比结果.

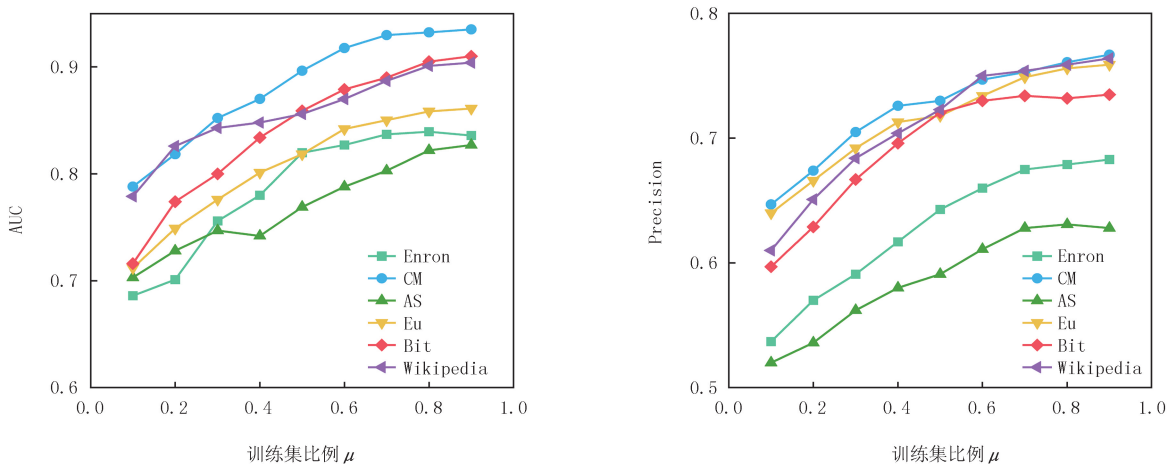


图1 SE-TLP在不同训练集比例下的AUC和Precision性能图

Fig.1 AUC and Precision performance of SE-TLP under different training set proportions

如图 1 所示,随着训练集比例 μ 的逐步增加,SE-TLP 算法的 AUC 和 Precision 整体呈上升趋势,表明

更多的训练数据有助于模型更充分地学习网络结构与演化规律.此外,当 μ 超过 0.5 时,SE-TLP 性能提升速度放缓,当 μ 大于 0.7 后,性能差距缩小,且最优和次优性能多集中在 μ 大于 0.8 区间.即随着训练数据增加,性能稳健提升,并在训练数据到达一定比例时,性能逐渐有收敛趋势,这种趋势表现符合一般实验规律.然而,虽增加训练数据可以提高模型的泛化能力,但数据量过大并不会显著提升性能,甚至有过拟合的风险,因此,后续实验将训练集比率设置为 0.8,在不增加计算资源的同时确保算法最佳性能.

3.5 实验对比与分析

SE-TLP 算法与 6 种基准算法在不同类型动态数据集集中的 AUC 性能对比.结果如表 1 所示,其中加粗的数字和标有下划线的数值分别为最优和次优性能值.与 6 种基准算法相比,SE-TLP 算法在 6 个网络中均取得最佳 AUC 性能,性能值最高提升 28.3%,最低提升 7.47%.综合分析,SE-TLP 算法保留了网络节点和连边的特性,增强了网络结构的初始嵌入表示能力,提高了链路预测整体性能,在 6 种数据集中均表现优越.而其他基准算法仅依赖节点连通性信息学习网络整体结构,对动态网络中信息的提取匮乏,整体预测准确性不如 SE-TLP.

SE-TLP 算法与 6 种基准算法的 Precision 对比结果如表 2 所示,其中加粗的数字和标有下划线的数值分别为最优和次优性能值.如表 2 所示,SE-TLP 算法在 Enron,CM, Eu, Bit 和 Wikipedia 网络中均取得最佳 Precision 性能值,SE-TLP 算法在多数网络中展现出显著优势, Precision 性能值最高提升 21.59%,最低提升 4.43%,表现出更强的准确预测阳性样本的能力.这是因为 SE-TLP 联合了 ChebNet 和 GRU 模型学习网络时空特性,更深入地挖掘动态网络中潜在的结构和时序信息.在 AS 这类稀疏且具有高节点中心性的网络中,网络演化由核心节点主导,而 SE-TLP 算法从节点和连边两个层面考虑网络结构演化.相比之下, GAT 的注意力机制能自适应聚焦核心节点影响力, STP-GCN 提出节点强度中心性关注网络核心节点的作用,因此 SE-TLP 算法的性能略低于 GAT 和 STP-GCN.

表 1 SE-TLP 与 6 种基准算法的 AUC 性能对比

Tab. 1 AUC performance comparison between SE-TLP and six baseline algorithms						
算法	Enron	CM	AS	Eu	Bit	Wikipedia
CN ^[17]	0.701 6	0.750 8	0.697 1	0.747 2	0.632 1	0.739 3
Katz ^[18]	0.702 7	0.775 8	0.671 9	0.672 7	0.660 9	0.730 1
GCN ^[7]	0.717 9	0.811 6	0.755 4	0.656 6	0.838 3	0.671 6
GAT ^[19]	0.740 9	0.793 1	0.744 4	0.698 9	0.790 7	0.726 1
GC-LSTM ^[20]	0.736 2	<u>0.851 7</u>	0.752 6	0.752 7	0.831 3	0.810 9
STP-GCN ^[21]	<u>0.777 8</u>	0.832 0	<u>0.772 9</u>	<u>0.783 1</u>	<u>0.856 3</u>	<u>0.851 4</u>
SE-TLP	0.839 5	0.932 3	0.822 3	0.858 5	0.904 6	0.901 0

表 2 SE-TLP 与 6 种基准算法的 Precision 性能比较

Tab. 2 Precision performance comparison between SE-TLP and six baseline algorithms						
算法	Enron	CM	AS	Eu	Bit	Wikipedia
CN ^[17]	0.564 0	0.709 5	0.576 9	0.540 3	0.652 1	0.611 2
Katz ^[18]	0.602 3	0.698 4	0.628 2	0.656 9	0.663 0	0.654 0
GCN ^[7]	0.631 3	<u>0.724 1</u>	0.606 9	0.697 5	0.636 2	0.675 9
GAT ^[19]	0.631 1	0.702 1	0.688 9	0.632 7	<u>0.687 2</u>	0.662 0
GC-LSTM ^[20]	<u>0.650 6</u>	0.700 9	0.598 3	<u>0.720 2</u>	0.671 4	0.729 3
STP-GCN ^[21]	0.647 3	0.722 3	<u>0.634 1</u>	0.706 3	0.660 8	<u>0.748 9</u>
SE-TLP	0.678 8	0.760 6	0.630 8	0.755 8	0.731 5	0.759 1

3.6 消融实验

为进一步验证 SE-TLP 算法中双通道网络结构特征提取机制的有效性,本节设计了一种 SE-TLP 变体,只保留了由 ChebNet,GRU 和 FCL 组成的编码器-解码器框架,记作 model1,与 SE-TLP 算法进行性能对比,对比结果如表 3 所示,其中加粗数字表示最佳性能.实验结果显示,SE-TLP 算法在 6 个数据集集中的 AUC 性能平均提升 2.87%,且在 5 个数据集集中的 Precision 性能平均提升 2.44%.这表明,SE-TLP 算法设计双通道特征提取机制,通过计算节点中心性和连边关系强度保留完整网络结构信息,对提升预测性能具有关键作用.

此外,为验证 SE-TLP 算法中聚合 4 种不同权重值的中心性以准确捕捉节点特征方法的有效性,本节设计了一种聚合 4 种具有相同权重中心性的方法,记作 model2.表 3 中的性能对比结果表明,SE-TLP 算法在 6 种数据集集中的 AUC 性能平均提升为 1.51%,同时,在 5 种数据集中 Precision 性能的平均提升为 1.65%.这充分说明,SE-TLP 算法为不同中心性为其分配不同权重的设计,能够更准确地反映网络节点特征和连边特征的演化,从而提升了链路预测性能.

表 3 SE-TLP 与 model1,model2 的 AUC 和 Precision 性能对比
Tab. 3 AUC and Precision performance comparison among SE-TLP and model1 and model2

指标	算法	Enron	CM	AS	Eu	Bit	Wikipedia
AUC	SE-TLP	0.839 5	0.932 3	0.822 3	0.858 5	0.904 6	0.901 0
	model1	0.813 0	0.912 0	0.796 5	0.834 1	0.876 8	0.879 7
	model2	0.817 7	0.931 0	0.812 1	0.841 0	0.881 9	0.898 0
Precision	SE-TLP	0.678 8	0.760 6	0.630 8	0.755 8	0.731 5	0.759 1
	model1	0.662 7	0.732 4	0.638 3	0.742 2	0.719 7	0.740 9
	model2	0.682 4	0.741 8	0.623 7	0.745 8	0.724 0	0.742 6

3.7 计算复杂度分析

算法的计算复杂度是衡量其有效性的重要指标,SE-TLP 的复杂度主要由编码器-解码器中神经网络模型的参数规模、节点和连边特征的计算方法、网络规模以及网络快照数量决定,可以表示为 $O(T(N^2 + m_1(Nm_3)^2 + m_2(Nm_3)^2))$,其中, m_1 表示编码器中 ChebNet 和 GRU 的层数, m_2 表示解码器中 FCL 的层数, m_3 指嵌入维度大小, N 是动态网络节点数, T 是网络快照数量. N^2 表示每个快照中计算结构特征的开销, $m_1(Nm_3)^2$ 是编码器中每层的嵌入计算复杂度, $m_2(Nm_3)^2$ 是解码器中每层的嵌入计算复杂度.基准算法中, CN 和 Katz 是基于网络拓扑结构的算法,复杂性主要集中在计算节点对的公共邻居上;而 GCN, GAT,GC-LSTM,STP-GCN 和 SE-TLP 是基于图神经网络的算法,复杂性主要取决于图的规模和神经网络层上的计算.实验结果表明,SE-TLP 在保证可执行的前提下,性能显著优于 CN,Katz,GCN 和 GAT 算法.相比框架相近的 GC-LSTM 和 STP-GCN 算法,SE-TLP 在不显著增加复杂度的情况下,性能值提升明显,在时序链接预测任务上展现出更强的预测效果.

4 结论

本文提出的 SE-TLP 算法设计双通道网络结构特征提取机制,计算 4 种加权节点中心性和拓扑相似性以表示节点和连边特征,并融合这两类特征增强网络结构表示.同时设计基于 ChebNet 和 GRU 的编码器-解码器框架,实现对动态网络时空特征的协同学习与高效表征.在 6 种动态网络数据集中的实验结果表明,SE-TLP 的 AUC 和 Precision 性能均优于基准算法,验证了其在时序链路预测任务中的优越表现.然而,SE-TLP 算法在特征融合过程中对特征间复杂关联的建模仍存在不足,可能会导致网络信息损失或冗余.未来的工作将重点研究网络特征间的深层关联,引入自适应特征融合方法,最大化保留图结构中的关键信息,以解决现有方法的网络结构信息表征不充分问题.

附录见电子版(DOI:10.16366/j.cnki.1000-2367.2025.04.10.0003).

参 考 文 献

[1] Artime O,Grassia M,De Domenico M,et al.Robustness and resilience of complex networks[J].Nature Reviews Physics,2024,6(2):114-131.

[2] Bloch F,Jackson M O,Tebaldi P.Centrality measures in networks[J].Social Choice and Welfare,2023,61(2):413-453.

[3] Zhang T,Zhang K,Lv L S,et al.Temporal link prediction using node centrality and time series[J].International Journal of Future Computer and Communication,2020,9(3):62-65.

[4] Choudhury N,Uddin S.Evolutionary features for dynamic link prediction in social networks[J].Applied Sciences,2023,13(5):2913.

[5] Maritan C A,Lee G K.Resource allocation and strategy[J].Journal of Management,2017,43(8):2411-2420.

[6] Xue G T,Zhong M,Li J X,et al.Dynamic network embedding survey[J].Neurocomputing,2022,472:212-223.

[7] Defferrard M,Bresson X,Vandergheynst P.Convolutional neural networks on graphs with fast localized spectral filtering[C]//Proceedings of the 30th International Conference on Neural Information Processing Systems.[S.l.]:ACM,2016:3844-3852.

[8] Mei P,Zhao Y H.Dynamic network link prediction with node representation learning from graph convolutional networks[J].Scientific Reports,2024,14(1):538.

[9] NOSOUHIAN S,NOSOUHIAN F,KHOSHOUEI A K.A review of recurrent neural network architecture for sequence learning:Comparison between LSTM and GRU[EB/OL].[2025-05-13].<https://www.preprints.org/manuscript/202107.0252>.

[10] Pareja A,Domeniconi G,Chen J,et al.EvolveGCN:evolving graph convolutional networks for dynamic graphs[J].Proceedings of the AAAI Conference on Artificial Intelligence,2020,34(4):5363-5370.

[11] Hardin J S,Sarkis G,Urc P C.Network analysis with the enron email corpus[J].Journal of Statistics Education,2015,23(2):1-27.

[12] Panzarasa P,Opsahl T,Carley K M.Patterns and dynamics of users'behavior and interaction:Network analysis of an online community[J].Journal of the American Society for Information Science and Technology,2009,60(5):911-932.

[13] Leskovec J,Kleinberg J,Faloutsos C.Graphs over time:densification laws,shrinking diameters and possible explanations[C]//Proceedings of the Eleventh ACM SIGKDD International Conference on Knowledge Discovery in Data Mining.[S.l.]:ACM,2005:177-187.

[14] Yin H,Benson A R,Leskovec J,et al.Local higher-order graph clustering[C]//Proceedings of the 23rd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining.[S.l.]:ACM,2017:555-564.

[15] Kumar S,Spezzano F,Subrahmanian V S,et al.Edge weight prediction in weighted signed networks[C]//2016 IEEE 16th International Conference on Data Mining(ICDM).[S.l.]:IEEE,2016:221-230.

[16] Rozemberczki B,Allen C,Sarkar R.Multi-Scale attributed node embedding[J].Journal of Complex Networks,2021,9(2):cnab014.

[17] Papadopoulos F,Aldecoa R,Krioukov D.Network geometry inference using common neighbors[J].Physical Review E,2015,92(2):022807.

[18] Lü L Y,Jin C H,Zhou T.Similarity index based on local paths for link prediction of complex networks[J].Physical Review E,2009,80(4):046122.

[19] Jung J W,Heo H S,Yu H J,et al.Graph attention networks for speaker verification[C]//ICASSP 2021-2021 IEEE International Conference on Acoustics,Speech and Signal Processing(ICASSP).[S.l.]:IEEE,2021:6149-6153.

[20] Chen J Y,Wang X K,Xu X H.GC-LSTM:graph convolution embedded LSTM for dynamic network link prediction[J].Applied Intelligence,2022,52(7):7513-7528.

[21] Sserwadda A,Ozcan A,Yaslan Y.Structural and topological guided GCN for link prediction in temporal networks[J].Journal of Ambient Intelligence and Humanized Computing,2023,14(7):9667-9675.

An enhanced structural embedding algorithm for temporal link prediction

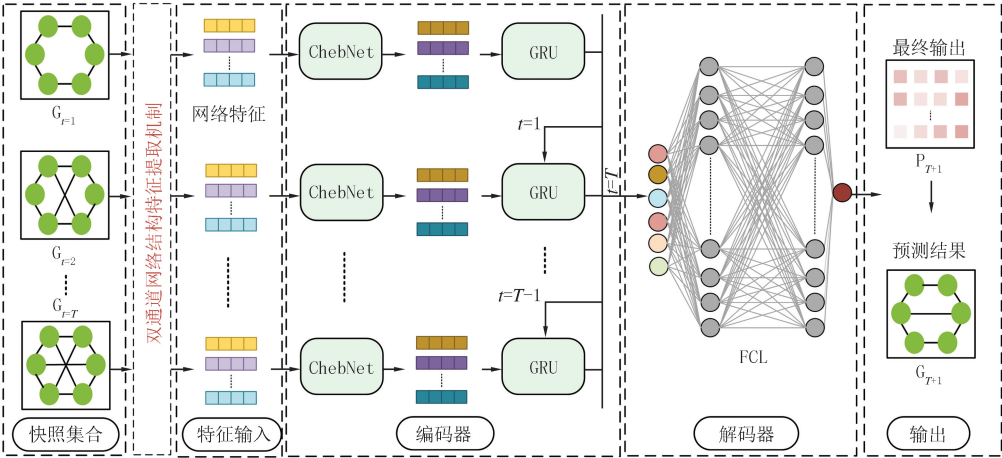
Yang Yujie^{a,b}, Liao Shulei^a, Wang Liming^a, Liu Dong^{a,b}

(a. College of Computer and Information Engineering; b. Key Laboratory of Artificial Intelligence and Personalized Learning in Education of Henan Province, Henan Norma University, Xinxiang 453007, China)

Abstract: In this paper, an enhanced structural embedding algorithm for temporal link prediction(SE-TLP)is proposed. SE-TLP employs a dual-channel structural feature extraction mechanism to separately model node and edge features. It integrates four weighted node centralities to assess node status, designs edge relation strength to represent topological structure, and fuses node and edge features to enhance structural embedding representations, thereby improving prediction accuracy. Meanwhile, SE-TLP combines a graph convolutional network with a gated recurrent unit to learn the network's structural evolution over time. Experiments show that SE-TLP outperforms mainstream algorithms on both evaluation metrics, achieving performance improvements of 7.47% and 4.43% respectively.

Keywords: temporal link prediction; network embedding; graph convolutional network; gated recurrent unit(GRU); dynamic network

[责任编辑 陈留院 杨浦]



图S1 SE-TLP算法整体框架图
Fig.S1 Overall framework of the SE-TLP algorithm

表 S1 SE-TLP 在不同快照数量下的性能和训练时间比较表

Tab. S1 Performance and training time comparison of SE-TLP at different snapshots

快照数量	指标	数据集					
		Enron	CM	AS	Eu	Bitcoin	Wikipedia
7	AUC	0.741 3	0.863 1	0.780 0	0.690 9	0.854 4	0.806 1
	Precision	0.666 7	0.717 6	0.605 4	0.698 4	0.677 1	0.681 0
	训练时间/s	4.3	21.3	20.2	77.1	27.4	23.5
10	AUC	0.750 9	0.876 7	0.792 5	0.713 1	0.867 8	0.819 0
	Precision	0.663 0	0.725 9	0.600 0	0.706 6	0.685 6	0.689 9
	训练时间/s	4.88	25.7	24.3	81.3	30.8	28.2
15	AUC	0.773 4	0.880 9	0.809 5	0.749 5	0.885 8	0.834 2
	Precision	0.669 7	0.733 3	0.610 0	0.708 3	0.709 5	0.694 3
	训练时间/s	5.27	30.4	27.6	90.3	35.6	33.1
30	AUC	0.815 9	0.895 1	0.812 3	0.790 6	0.897 8	0.859 5
	Precision	0.671 1	0.739 9	0.619 9	0.725 9	0.716 6	0.740 7
	训练时间/s	7.41	46.9	31.4	115.5	48.1	64.2
45	AUC	0.839 5	0.918 4	0.819 0	0.822 9	0.907 1	0.877 7
	Precision	0.678 8	0.746 5	0.625 6	0.733 5	0.723 8	0.752 0
	训练时间/s	8.36	68.5	48.9	128.7	66.5	95.4
60	AUC	0.847 4	0.926 7	0.822 3	0.841 1	0.904 6	0.887 1
	Precision	0.685 6	0.754 8	0.630 8	0.740 2	0.731 5	0.748 8
	训练时间/s	9.33	86.4	67.5	157.1	74.3	123.2
80	AUC	0.855 5	0.932 3	0.825 5	0.858 5	0.902 2	0.901 0
	Precision	0.689 1	0.760 6	0.632 3	0.755 8	0.737 2	0.759 1
	训练时间/s	13.54	112.1	83.4	174.4	96.7	150.6