
QB期刊 基于疾病子图位置编码的图Transformer用于改进共病预测

作者：writer 来源：科学网

本文原地址：<https://www.iikx.com/news/progress/38593.html>

本文仅供学习交流之用，版权归原作者所有，请勿用于商业用途！

QB期刊 基于疾病子图位置编码的图Transformer用于改进共病预测。论文标题：Graph transformer with disease subgraph positional encoding for improved comorbidity prediction

期刊：Quantitative Biology

作者：Xihan Qin, Li Liao

发表时间：28 Aug 2025

DOI：10.1002/qub2.70008

微信链接：点击此处阅读微信文章

共病（多种疾病同时发生）预测是精准医疗的关键环节，但现有基于人类蛋白质相互作用组（HI）的方法面临难以逾越的技术瓶颈，如位置编码局限：传统方法（如拉普拉斯位置编码LPE）仅捕捉聚类信息，忽略疾病亚图的标签关联，无法完整表征节点的疾病属性；特征捕捉不足：未能有效融合蛋白质节点连接性与疾病亚图关联，难以挖掘共病背后的分子机制；预测性能有限：面对不平衡的临床数据集，现有模型的区分能力不足，ROC AUC与准确率有待提升；泛化能力薄弱：对不同共病定义阈值（如RR0、RR1）的适配性差，难以满足多样化临床场景需求。这些痛点导致共病预测结果不够精准，无法为临床决策提供可靠支撑，成为疾病管理的拦路虎。

近期，Quantitative Biology 发表了一篇来自美国特拉华大学Xihan Qin和Li Liao团队的研究性论文Graph transformer with disease subgraph positional encoding for improved comorbidity prediction，提出含疾病亚图位置编码的图 Transformer 模型（TSPE），通过整合聚类信息与疾病关联特征，解决传统共病预测中位置编码不足的核心问题，在临床基准数据集上实现性能大幅突破，为疾病管理与治疗方案优化提供精准工具。

DOI: 10.1002/qub.2.70008

RESEARCH ARTICLE



Graph transformer with disease subgraph positional encoding for improved comorbidity prediction

Xihan Qin | Li Liao

全文概要

针对传统模型编码信息不全、特征融合不足、预测精度有限的三大瓶颈，团队构建了 TSPE 创新框架——以 Transformer 注意力机制捕捉蛋白质节点交互，提出亚图位置编码（SPE）整合拉普拉斯位置编码（LPE）的聚类信息与图编码器嵌入位置编码（GPE）的疾病关联信息，通过 Node2Vec 生成节点嵌入，结合无掩码多头注意力实现共病二元分类。经实证，该模型在 RR0 和 RR1 基准数据集上表现突出：RR0 数据集 ROC AUC 达 0.9489，较现有最优方法（BSE_SVM）提升 28.24%，准确率提升 3.04%；RR1 数据集 ROC AUC 达 0.8009，提升 15.40%，准确率提升 4.93%。消融实验证实，SPE 显著优于 LPE 和无编码（NoPE），Matthews 相关系数（MCC）提升 2.38% 以上。该模型为共病预测提供了更全面的特征表征方案，未来将拓展至更大规模 HI 数据集，优化泛化能力并探索多疾病交互预测场景。

核心架构与流程

TSPE 采用 节点嵌入—亚图编码—Transformer 编码—分类输出 的四步流程（图1）：

- 1.节点嵌入生成：通过 Node2Vec 对 HI 中的蛋白质节点进行嵌入，捕捉网络结构特征；
- 2.SPE 编码整合：将 LPE（聚类信息）与 GPE（疾病亚图标签信息）融合，生成最终位置编码，公式如下： $E = [(M + LPE), GPE]$ 其中 M 为节点嵌入矩阵，通过奇异值分解（SVD）降低 GPE 维度，确保编码维度一致性；
- 3.Transformer
编码：编码器与解码器均采用无掩码多头注意力机制，捕捉疾病亚图间的节点交互；
- 4.分类输出：通过列 - wise L2 范数加权求和与 sigmoid 激活函数，实现共病二元分类，损失函数采用二元交叉熵（BCE）。

图示呈现从疾病 A、B 的相关蛋白质节点，经 Node2Vec 嵌入、SPE 编码，到 Transformer 编码器 - 解码器处理，最终输出共病预测结果的完整流程，核心是 SPE 编码与注意力机制的结合

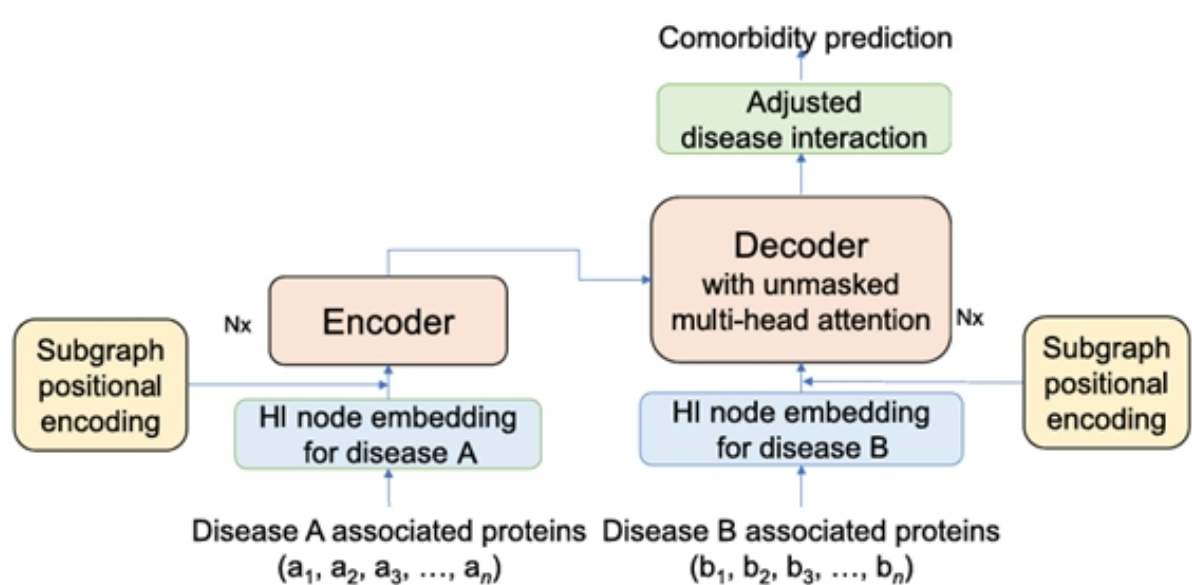


图1. TSPE 模型架构示意图

关键编码方法对比

团队创新提出的 SPE，解决传统编码信息不全的问题，三类编码方法差异如下(图2)：

- LPE（拉普拉斯位置编码）：仅捕捉图聚类信息，无法识别节点的疾病亚图归属；
- GPE（图编码器嵌入位置编码）：基于已知亚图信息生成权重矩阵，突出疾病关联，但缺乏全局聚类特征；
- SPE（亚图位置编码）：通过拼接（M+LPE）与 GPE，同时整合聚类信息与疾病标签关联，实现更全面的节点表征。

左图为 LPE 仅覆盖连通组件内节点，中图为 GPE 突出疾病亚图关联，右图为 SPE 融合两者优势，同时捕捉聚类与疾病归属信息

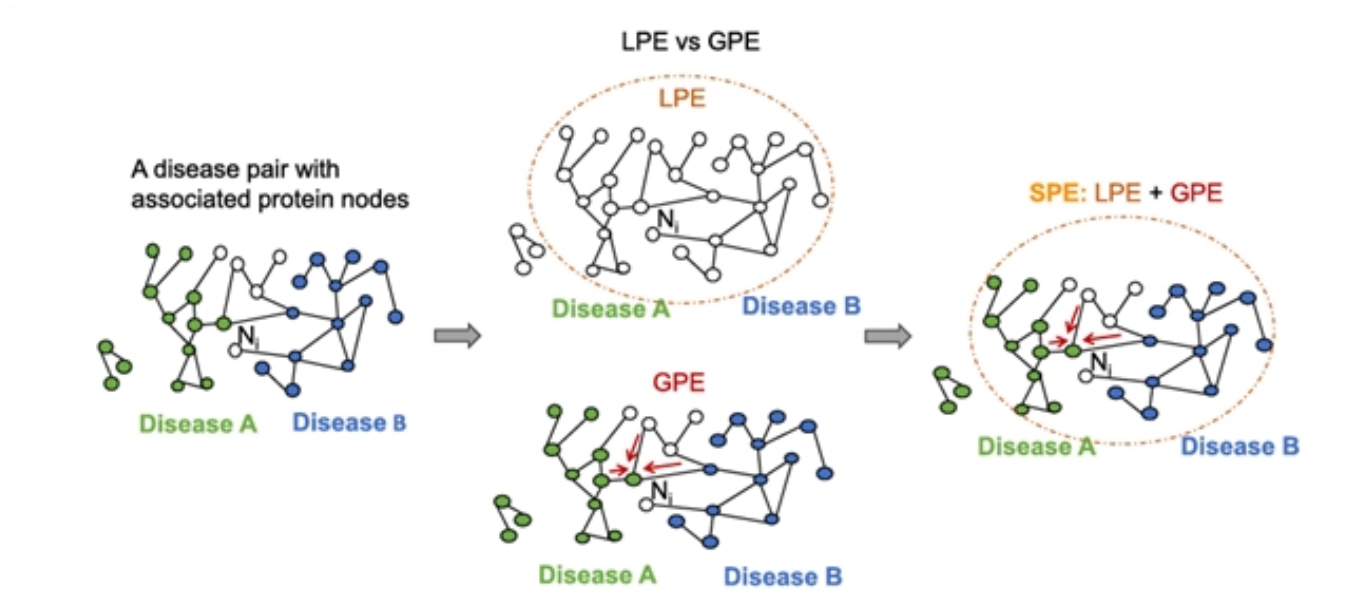


图 2. 不同位置编码方法对节点 N_i 的表征对比

核心数学方程

拉普拉斯矩阵与归一化：

$$L = D - A, \quad \bar{L} = D^{-\frac{1}{2}} L D^{-\frac{1}{2}}$$

其中D为度矩阵，A为邻接矩阵，用于 LPE 的特征分解；

GPE 生成（基于 SVD 降维）：

$$Z = AW, \quad Z = U \Sigma V^T, \quad GPE = U_d$$

其中W为疾病亚图权重矩阵， U_d 为前 d 个左奇异向量；

分类输出加权：

$$s = \text{softmax} \left(\|X\|_{2,col}^2 \right), \quad \hat{y}' = \sum_{j=1}^n (X \text{diag}(s))_j, \quad \hat{y} = \sigma (W \hat{y}' + b)$$

实证效果

团队在 RR0 (RR>0 为共病) 和 RR1 (RR ≥ 1 为共病) 两个临床基准数据集上进行 10 折交叉验证, 结果显著优于现有方法。

总结

本研究创新性地将亚图位置编码 (SPE) 与 Transformer 注意力机制结合, 构建 TSPE 模型, 突破传统共病预测中位置编码的局限。通过整合人类蛋白质相互作用组的结构特征与疾病亚图的关联信息, TSPE 在临床数据集上实现性能大幅提升, 为共病预测提供了更精准、全面的技术方案, 彰显了图深度学习在生物医学领域的应用潜力。

QB期刊介绍

Quantitative Biology (QB) 期刊是由清华大学、北京大学、高教出版社联合创办的全英文学术期刊。QB主要刊登生物信息学、计算生物学、系统生物学、理论生物学和合成生物学的最新研究成果和前沿进展, 并为生命科学与计算机、数学、物理等交叉研究领域打造一个学术水平高、可读性强、具有全球影响力的交叉学科期刊品牌。

《前沿》系列英文学术期刊

由教育部主管、高等教育出版社主办的《前沿》(Frontiers) 系列英文学术期刊, 于2006年正式创刊, 以网络版和印刷版向全球发行。系列期刊包括基础科学、生命科学、工程技术和人文社会科学四个主题, 是我国覆盖学科最广泛的英文学术期刊群, 其中12种被SCI收录, 其他也被AHCI、Ei、MEDLINE或相应学科国际权威检索系统收录, 具有一定的国际学术影响力。系列期刊采用在线优先出版方式, 保证文章以最快速度发表。

中国学术前沿期刊网

<http://journal.hep.com.cn>



来源: Quantitative Biology

更多 科学进展 请访问 <https://www.iikx.com/news/progress/>

本文版权归原作者所有，请勿用于商业用途，[爱科学iikx.com](https://www.iikx.com)转发