

关于 MoleRec：具有子结构感知的分子表征学习的 组合药物推荐的复现报告

温嘉伟

2023 年 1 月 13 日

摘要

组合药物推荐是指根据患者的纵向病史向患者推荐个性化的药物组合，其本质目的是解决在安全约束下追求高精度的组合优化问题。药物对疾病的作用通常与其特殊子结构有关，但药物的子结构与目标疾病的重要关联通常被忽视。MoleRec 的作者提出了一种新颖的分子结构感知的关注方法，建模患者健康状况和药物子结构之间的关系此外，还设计了一种调整策略，在训练阶段重新加权药物相互作用 (DDI) 损失，帮助模型在推荐准确性和安全性之间找到更好的平衡点。但 MoleRec 模型仅提取了药物的分子结构作为药物的表征，而药物的共现信息 [26] 被认为是指导药物推荐的重要知识，对此，本人将药物的共现信息嵌入药物的表征，改进了 MoleRec 对药物表征的建模。在 MIMIC- 数据集上的实验证明，嵌入药物共现信息，模型的性能有所提高。

关键词：组合药物推荐；推荐系统；推荐系统；子结构感知；药物共现

1 引言

随着个体医疗数据（如纵向电子健康记录）的增加 [4, 7]，学习模型在改善临床决策的准确性和安全性方面显示出了良好的效果 [11, 14, 16]。其中，组合药物推荐旨在根据患者的健康状况为其提供适当的药物组合作为最终处方。然而，组合药物推荐面临两个主要挑战：1) 与传统推荐任务不同，药物推荐旨在返回一组药物而不是单个药物；2) 推荐的药物组合需要满足一定的安全要求，即抑制最终推荐药物之间的不良化学反应，即药物-药物相互作用 (DDI)。

组合药物推荐的早期工作 [20, 25] 仅根据病人当前的就诊情况来推荐药物，忽略了病人的纵向历史病历。后续的一些工作 [10, 17] 开始考虑加入病人纵向信息，结果证明纵向信息能够提高组合药物推荐的效果。杨等人 [23] 提出容纳整个药物分子的信息，但他们的模型未能识别每个药物分子内的子结构与目标疾病之间的关联，而最近许多基于学习的药物设计工作所证明，这种关联可以提供丰富的信息。最近，开发了一种基于症状的小而安全的药物推荐方法，称为 4SDrug[18]。但它往往会推荐相对较小的药品包装，以避免 DDI，进而满足安全原则，这可能会导致推荐准确性较低。

目前生物信息学、药学和数据挖掘领域的大量研究工作 [8, 9, 15, 27] 已经表明药物的生化活性通常与一些特殊的分子子结构相关。子结构通常不是孤立地发挥作用，而是一组子结构相互作用发挥作用，所以 MoleRec 作者希望通过对子结构之间的高阶交互进行建模，以捕

捉子结构与患者健康状况的关系。MoleRec 旨在通过建模分子亚结构之间的相互作用和患者健康状况与亚结构之间的关联性，提出一种新颖的分子亚结构感知的关注学习方法，用于顺序药物推荐。在 MIMIC-III 数据集上的实验的有效性使得 MoleRec 成为一种可行的方法，可以帮助医生和临床决策支持系统为患者提供个性化的药物组合推荐，提高临床决策的准确性和安全性。

2 相关工作

2.1 药物组合推荐

现有的药物组合推荐方法大多数分为基于实例的药物组合推荐方法和纵向药物推荐方法。基于实例的方法关注当前患者的健康状况，但忽略了患者纵向历史病历。一项早期工作参考了 LEAP[25]，它将药物推荐制定为多实例多标签顺序决策过程，而对纵向历史病历建模的方法利用临床就诊中的时间依赖性 [2, 12]。其中，RETAIN[3] 基于两级神经注意力模型，该模型检测过去有影响的就诊以及这些就诊中的重要临床变量。然而，RETAIN 却很少考虑安全性。因此，GAMENet[17] 通过将纵向患者记录联合建模为 EHR 图和将药物知识库建模为 DDI 图来考虑 DDI 冲突关系。此外，SafeDrug[23] 提取并编码丰富的分子结构信息，以提高药物推荐的安全性。最近的 4SDrug[18] 推荐少量药物以确保安全。本文提出的 MoleRec 旨在探索分子亚结构之间的相互作用以及患者健康状况与亚结构之间的相关性，与最先进的模型相比取得了显著的改进。并采用基于内容注意机制的序列到序列模型的变体来制定处方。

2.2 学习分子表示

现有的分子表示学习方法也可以分为两类：第一种是基于 SMILES 的方法，其中 SMILES 指的是简化分子输入线输入系统 [21]。这种方法使用 Transformer[19] 或 BERT[6] 等语言模型来处理分子的文本表示 (SMILES)。SMILES 是分子的线性编码，高度依赖于分子图的遍历顺序。因此，SMILES 的表达能力对于诸如药物推荐之类的问题是有限的，作者认为这些问题需要细粒度的分子结构提取。例如，基于这种现象和因果不变性原理 [13, 22]，最近一项名为 MoleOOD[24] 的工作提出利用这些环境不变的子结构来学习针对分布变化的鲁棒分子表示。作者提出的 MoleRec 的目标是学习用于药物组合推荐的子结构感知分子表示，以提供更安全、更准确的处方。

3 本文方法

3.1 本文方法概述

MoleRec 由三个部分组成：1) 患者表示模块，编码患者的纵向诊断和手术信息。2) 药物表示模块根据患者的不同状况生成药物的子结构感知表示。3) 预测模块负责仅使用药物的子结构感知表示来制定处方。MoleRec 的整体框架如图 1所示：

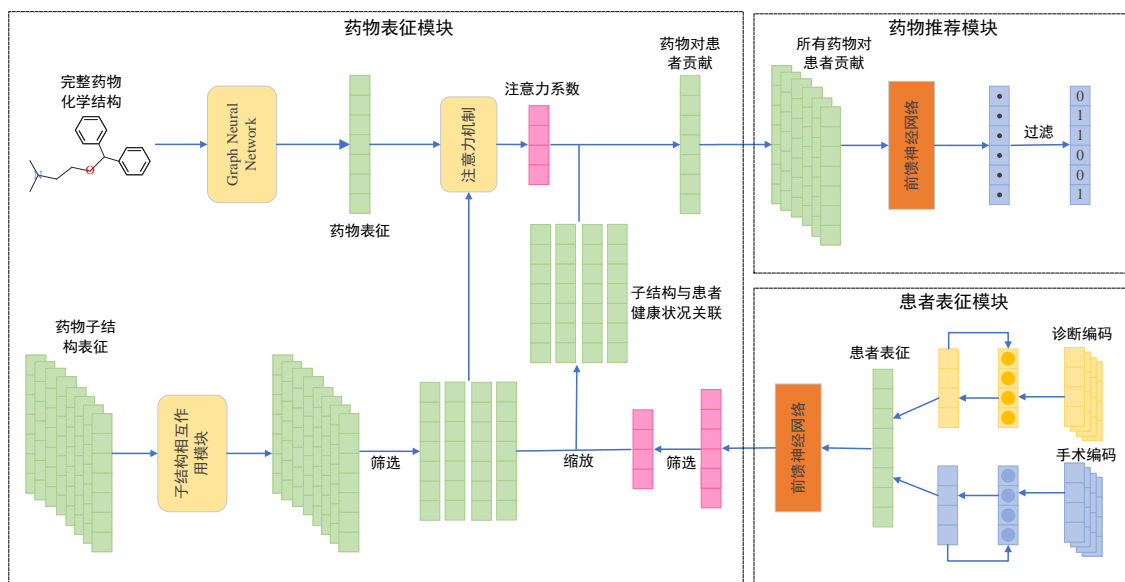


图 1. MoleRec 模型框架图

3.2 患者表征模块

患者的健康状况由诊断信息编码和手术信息编码组成，MoleRec 的作者通过两个可学习的嵌入表对患者的诊断信息编码和手术编码信息进行特征嵌入，得到每位患者每次就诊的特征向量。

受前人工作 [10, 17] 的启发，作者利用 Dual-RNN 对患者的纵向历史病历进行建模。Dual-RNN 是一种用于序列建模的神经网络模型，它由两个 RNN 分别处理患者诊断编码信息和手术编码信息的输入序列，并将两者的隐状态拼接起来作为最终的输出。在对患者历史病历进行建模时，将患者纵向历史病历中的每次就诊记录看作是一个输入序列，然后将就诊的诊断编码和手术编码的序列分别输入到两个 RNN 中进行处理，最终将两个方向的隐状态拼接起来得到患者健康状况的表征。

由于 Dual-RNN 可以同时考虑患者诊断编码和手术编码的信息，因此它可以更好地捕捉患者纵向历史病历中的长期依赖关系，从而有效地捕捉到患者健康状况的变化。此外，与单向 RNN 相比，Dual-RNN 能够提高模型的鲁棒性和泛化能力。

3.3 药物表征模块

3.3.1 完整药物表征模块

将药物的化学分子结构转换成图，并使用 GNN 进行建模，从而获得完整药物的表征。在这种方法中，药物的化学结构被表示为一个图，其中原子和化学键被视为图的节点和边。值得注意的是作者对完整药物进行建模的作用与前人工作有所不同，MoleRec 对完整药物的建模并不直接用于最终推荐药物，而是为了计算子结构与其对应完整药物的相关系数，以根据子结构对患者健康状况的贡献程度，推荐完整的药物。

3.3.2 子结构感知模块

MoleRec 作者假设药物分子的性质取决于其特定的子结构，并且对患者疾病的治疗也依赖于某些子结构的功能，作者的目标是充分利用药物的子结构感知信息。首先 MoleRec 通过 BRICS[5] 将完整的药物分子分解多个子结构，然后为了建模子结构之间的高阶相互作用，作者设计了一个子结构相互作用模块 (SIM)。SIM 以 Set Attention Block (SAB) 作为药物子结构的交互模块。首先通过 GNN 对所有的子结构进行编码，然后再子结构之间执行自注意力机制，从而得到新的子结构特征表示。

$$E_s^* = SIM(E_s) = LN(H + FF_1(H)) \quad (2-1)$$

$$where H = LN(E_s + \mathfrak{R}_{vanilla}(E_s)) \quad (2-2)$$

FF_1 代表馈神经网络， LN 是正则化层。 $\mathfrak{R}_{vanilla}$ 代表普通自注意力机制，其定义为：

$$\mathfrak{R}_{vanilla} = softmax(\frac{Q_k K_x^T}{\sqrt{d_k}}) V_x \quad (2-3)$$

MoleRec 作者还设计了一个子结构相关性模块 SRM，SRM 负责显式地建模患者对每个子结构的依赖程度。

3.4 药物推荐模块

每一种药物通过拆分成子结构，并评估子结构对患者健康状况的贡献程度，而子结构对药物的注意力系数反应药物是患者的疾病的治疗是否具有重要作用。最后药物推荐模块通过前馈神经网络，把所有子结构感知的药物对患者疾病治疗重要性的表征转换为每种药物的推荐概率，然后挑选出概率大于作者预定的阈值的药物作为推荐的药物组合。

3.5 DDI 损失权重退火策略

作者设计了一种调整策略，用于在训练过程中重新加权药物-药物相互作用损失，以在推荐准确性和安全性之间找到更好的平衡点。

$$\alpha = \begin{cases} 1 & \rho < \phi \\ \min \left\{ 1, \exp \left(\tau \left(1 - \frac{\rho}{\phi} \right) \right) \right\} & \rho \geq \phi \end{cases} \quad (2-4)$$

当药物-药物相互作用的差距较大时，调整策略会使 DDI 损失缓慢下降，以保持 DDI 损失的权重较高。当差距较小时，调整策略会使 DDI 损失急剧下降，模型将更加关注优化多标签预测损失。通过这种调整策略，作者能够在推荐准确性和安全性之间取得更好的平衡，并提高了模型的性能。

4 复现细节

4.1 与已有开源代码对比

本人复现工作中引用了 MoleRec 作者的基本框架的代码：1) MoleRec 的作者提出了一种新颖的分子结构感知的关注方法代码首先通过 GNN 网络对药物分子结构进行建模，得到每

种药物子结构的特征向量。然后再通过一个集合注意力模块，对药物子结构之间的相互作用进行建模。2) MoleRec 作者代码通过 GNN 网络对药物整体分子结构进行建模。3) MoleRec 作者代码通过双 RNN 网络对患者纵向历史病历进行建模，通过前馈神经网络捕捉患者健康状况与药物子结构之间的相关性。

作者仅通过对药物完整结构建模来获取药物的表征，而药物的共现信息和药相互作用信息也是药物重要的特征信息，本人在对药物结构进行建模的过程中，把药物共现信息和药物相互作用信息融入到药物的整体表征中，以增强对药物完整水平的表征。新增模块如下：1) 药物共现信息提取，在数据预处理阶段，从数据集中提取药物共现信息。2) 通过 GCN 网络对药物共现信息和药物相互作用信息的建模。3) 设计三种融合组合：同时融入药物共现信息和药物相互作用信息、仅融入药物共现信息、仅融入药物相互作用信息，对三种融合组合进行实验。

4.2 改进与创新

MoleRec 提出的新颖的分子结构感知的关注的方法在药物组合推荐任务中取得了良好的效果。与传统的基于序列生成的方法相比，该方法将药物推荐任务转化为多标签预测问题，更加合理。与其他基线方法相比，该方法考虑了药物分子的子结构之间的相互作用和与患者健康状况之间的依赖关系。但该方法在提取药物完整水平的特征时仅使用了药物的化学结构，而药共现信息和药物相互作用信息也是进行组合药物推荐的重要信息。对此本人在药物完整水平的表征上进行了改进，融入了药物共现信息，以增强药物在整体水平上对患者健康状况的影响，改进后的 MoleRec 如图 2 所示。本人进行了以下三次实验以证明药物共现信息和药物相互作用信息对药物完整水平的增强的有效性：1) 同时融入药物共现信息和药物相互作用信息；2) 仅融入药物共现信息；3) 仅融入药物相互作用信息。

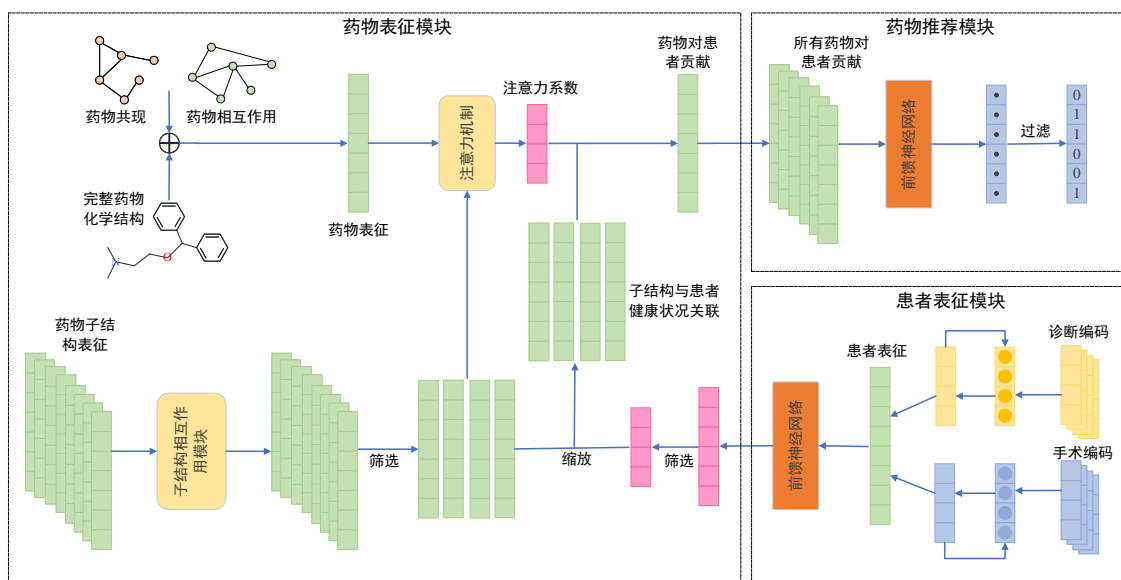


图 2. MoleRec 改进框架图

5 实验结果分析

5.1 对论文实验复现

5.1.1 定量分析

MoleRec 通过 DDI、Jaccard、F1-score、PRAUC 等指标评估该模型的效果，结果如表 1 所示，该表格汇总了所有方法的预测性能。从该表格可以看出本人的复现 MoleRec 的结果与作者的结果无明显差异，在 DDI 率、Jaccard 和 F1-score 这些指标略低于作者，但 PRAUC 指标略高于作者的实验结果。实验结果证明 MoleRec 能够有效的实现组合药物推荐，并且该方法以显著的优势优于所有基线。

表 1. 模型各指标性能表

Method	DDI ↓	Jaccard ↑	F1-score ↑	PRAUC ↑	Avg.# of Drugs
Logistic	0.0816 ± 0.0007	0.4924 ± 0.0030	0.6509 ± 0.0027	0.7589 ± 0.0026	16.7474 ± 0.1131
Regression	0.0817 ± 0.0007	0.4856 ± 0.0031	0.6438 ± 0.0028	0.7590 ± 0.0026	16.2578 ± 0.0992
ECC	0.0817 ± 0.0007	0.4856 ± 0.0031	0.6438 ± 0.0028	0.7590 ± 0.0026	16.2578 ± 0.0992
RETAIN	0.0871 ± 0.0013	0.4866 ± 0.0034	0.6471 ± 0.0032	0.7593 ± 0.0035	18.5941 ± 0.2186
LEAP	0.0760 ± 0.0008	0.4540 ± 0.0027	0.6158 ± 0.0025	0.6598 ± 0.0026	18.6739 ± 0.0661
DMNC	0.0801 ± 0.0011	0.4550 ± 0.0031	0.6160 ± 0.0031	0.6757 ± 0.0029	20.0000 ± 0.0000
GAMENet	0.0859 ± 0.0005	0.5037 ± 0.0015	0.6601 ± 0.0014	0.7673 ± 0.0024	27.2603 ± 0.1929
SafeDrug	0.0773 ± 0.0006	0.5126 ± 0.0028	0.6691 ± 0.0023	0.7655 ± 0.0022	20.8940 ± 0.1086
4SDrug	0.0703 ± 0.0011	0.4800 ± 0.0027	0.6404 ± 0.0024	0.7611 ± 0.0026	16.1684 ± 0.1280
MoleRec	0.0724 ± 0.0008	0.5305 ± 0.0033	0.6843 ± 0.0029	0.7736 ± 0.0027	21.0893 ± 0.1788
Ours	0.0726 ± 0.0009	0.5296 ± 0.0031	0.6837 ± 0.0025	0.7767 ± 0.0021	21.2022 ± 0.1639

5.1.2 定性分析

MoleRec 还定性评估了患者的健康状况与 MoleRec 学到的子结构之间的相关性。图 3 显示了患者的健康状况和一些子结构之间的学习相关性。由于 MoleRec 作者并没有详细说明作者选取的示例药物为何种药物，所有本人随机选取了一种药物，评估了子结构与患者健康的相关性。结果与 MoleRec 作者实验结果趋于一致，子结构与患者相关程度高，则该药物在推荐的组合药物的肯能性高。

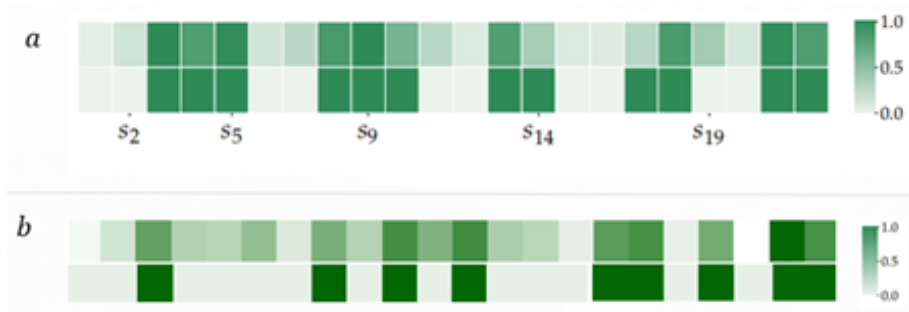


图 3. 药物子结构与患者健康状况的关联程度

但是实验结果中存在部分异常，例如 MoleRec 作者的实验中子结构虽然与患者相关低，但其对于药物在推荐的组合药物中。作者通过对比子结构与患者健康状况的相关程度和子结构对药物的注意力系数可以解释该问题。图 4展示了子结构与患者健康状况的相关程度与药物注意力系数存在差异，说明了仅从药物完整水平与患者健康状况来推荐组合药物存在着不足，即进一步证明了从子结构水平推荐药物的有效性。本人随机选取一种药物，评估了其子结构与患者健康状况相关程度和药物本身的注意力系数，实验结果显示与作者的结果存在相同的现象。



图 4. 子结构对完整药物和患者健康状况相关性差异

该论文还进行了一系列的消融实验来评估其方法的有效性和性能。1) 时间依赖性：将每个访问视为单个训练样本，而不是使用循环神经网络来建模它们的时间依赖性。结果表明，相比于纵向模型，该方法的性能较差。2) 子结构交互模块：移除子结构交互模块，并进行了实验。结果表明，在安全性方面，性能下降明显，这表明建模子结构之间的相互作用可以在很大程度上提高推荐的安全性。3) 子结构相关性模块：舍弃了提出的子结构相关性模块，并评估其影响。结果表明，没有子结构相关性模块时，给定药物分子子结构感知表示被重新计算，这表明子结构相关性模块可以有效地建模患者和分子子结构之间的相关性。4.) 多标签边界损失：将多标签边界损失从模型中移除，并将最终目标函数减少为仅包含二元交叉熵损失和 DDI 损失的加权和。结果表明，没有多标签边界损失时，模型推荐的药物数量较少，从而减少了 DDI 的发生，并且在最终处方中推荐的真实药物数量较少。这表明多标签边界损失可以有效地使预测的真实标签的概率至少比其他标签大。5) 分子分割策略：使用 BRICS 方法将分子分解为子结构。为了评估模型对分子分割策略的敏感性，采用了另一种方法称为反合成组合分析过程 (RECAP)。结果表明，RECAP 和 BRICS 在模型上表现相当的性能，并且都优于基线方法。本人复现结果如表 2所示：

表 2. 消融实验结果

Method	DDI ↓	Jaccard ↑	F1-score ↑	PRAUC ↑	Avg.# of Drugs
SafeDruge	0.0773 ± 0.0006	0.5126 ± 0.0028	0.6691 ± 0.0023	0.7655 ± 0.0022	20.8940 ± 0.1086
MoleRec	0.0726 ± 0.0009	0.5296 ± 0.0031	0.6837 ± 0.0025	0.7767 ± 0.0021	21.2022 ± 0.1639
visits	0.0695 ± 0.0008	0.5162 ± 0.0034	0.6729 ± 0.0027	0.7596 ± 0.0025	21.5612 ± 0.1534
- (SIM)	0.0727 ± 0.0012	0.5328 ± 0.0026	0.6868 ± 0.0029	0.7771 ± 0.0023	21.8491 ± 0.1434
- (SRM)	0.0787 ± 0.003	0.4967 ± 0.0031	0.6549 ± 0.0030	0.7491 ± 0.0026	18.7157 ± 0.1764
- multi-label	0.0678 ± 0.0015	0.4311 ± 0.0027	0.5897 ± 0.0033	0.7509 ± 0.0031	11.9592 ± 0.1542
RECAP	0.0767 ± 0.0010	0.4819 ± 0.0029	0.6420 ± 0.0025	0.7093 ± 0.0014	19.6292 ± 0.1356
GNN	0.0706 ± 0.0011	0.5141 ± 0.0022	0.6707 ± 0.0028	0.7545 ± 0.0016	22.3506 ± 0.2243
linear adjusting	0.0753 ± 0.008	0.5308 ± 0.0033	0.6850 ± 0.0029	0.7793 ± 0.0025	22.7889 ± 0.1732

5.2 对 MoleRec 改进后实验结果

为增强药物完整水平的表征，陈钱钰 [1] 等人认识到 MR 任务中分子图难以区分的问题，并开发了一种上下文感知 GNN，为分子图配备从 EHR 中提取的上下文信息，从而学习更多可区分的药物表示。受其启发，本人融入药物共现信息和药物相互作用信息，以概述学习到的药物的嵌入，从而带来更准确的推荐。同时为了证明药物共现信息和药物相互作用信息的有效性，分别进行了只融入 DDI 和只融入共现信息的实验，实验结果如表 3 所示：

表 3. 改进 MoleRec 各指标性能

Method	DDI ↓	Jaccard ↑	F1-score ↑	PRAUC ↑	Avg.# of Drugs
MoleRec	0.0724 ± 0.0008	0.5305 ± 0.0033	0.6843 ± 0.0029	0.7736 ± 0.0027	21.0893 ± 0.1788
Ours	0.0726 ± 0.0009	0.5296 ± 0.0031	0.6837 ± 0.0025	0.7767 ± 0.0021	21.2022 ± 0.1639
EHR & DDI	0.0675 ± 0.0007	0.5096 ± 0.0028	0.6666 ± 0.0025	0.7513 ± 0.0027	21.5445 ± 0.1382
EHR Only	0.0730 ± 0.0007	0.5294 ± 0.0034	0.6835 ± 0.0029	0.7716 ± 0.0028	20.5620 ± 0.1326
DDI Only	0.0690 ± 0.0008	0.5099 ± 0.0031	0.6668 ± 0.0028	0.7527 ± 0.0029	21.7876 ± 0.1453

从实验结果可以看出，融入药物共现信息和药物相互作用信息后，推荐的药物组合的 DDI 指标有显著改善，DDI 指标指示着推荐药物的相互作用率，在推荐药物时安全性时至关重要的目标，DDI 指标的改善，表明改进后的模型能为患者提供更加安全的药物组合推荐。虽然其他指标略低于 MoleRec，但仍优于基线方法。本人通过融入药物共现信息和药物相互作用信息从而改进模型，使得 MoleRec 具有更加均衡的性能，实验证明本人对模型的改进使得模型的性能得到优化。

6 总结与展望

6.1 总结

该论文是关于药物组合推荐的研究，提出了一种基于分子亚结构感知的注意力学习方法，称为 MoleRec。该方法通过建模分子亚结构之间的相互作用和患者健康状况与亚结构之间的依赖关系，实现了个性化的药物组合推荐。论文还设计了一种调整策略，用于在训练阶段重新加权药物-药物相互作用损失，以在推荐准确性和安全性之间找到更好的平衡点。

MoleRec 在药物完整水平的表征有所欠缺，本人通过融入药物共现信息和药物相互作用信息增强对药物完整水平的表征。在 MIMIC-III 数据集上进行的实验表明，MoleRec 方法在准确性和安全性方面都取得了良好的效果，而本人对模型的改进使得 MoleRec 在安全性方面取得了更好的效果，改进后的 MoleRec 具有更均衡的准确性和安全性。最后融入药物相互作用信息的实验结果中，组合药物推荐的安全性与准确率的变化，启发了本人未来研究工作的方向，在未来设计一种更加合理的药物相互作用信息对药物推荐的影响策略，在保证安全性的前提下，释放安全性对模型准确度的限制

6.2 展望

实验结果表明当融入药物相互作用信息 DDI 时，最终推荐药物组合的 DDI 有显著降低，而仅融入药物共现信息 HER 时，对结果影响并不明显。此结果表明组合药物推荐的 DDI 与药物的准确率是成反比的，当 DDI 结果改善时，准确率会有所下降。这个现象对本人的启发是：未来可以通过设计一种更加合理的 DDI 对组合药物推荐影响的策略，根据患者疾病的重要程度设置不同强度的 DDI 作用，然后依次为患者的每一种疾病，在不同 DDI 强度的影响下推荐药物组合。通过这种策略，既能释放 DDI 对组合药物推荐的性能的限制，又能确保安全性不发生显著改变。

References

- [1] Qianyu Chen, Xin Li, Kunnan Geng, and Mingzhong Wang. Context-aware safe medication recommendations with molecular graph and ddi graph embedding. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 37, pages 7053–7060, 2023.
- [2] Edward Choi, Mohammad Taha Bahadori, Andy Schuetz, Walter F Stewart, and Jimeng Sun. Doctor ai: Predicting clinical events via recurrent neural networks. In *Machine learning for healthcare conference*, pages 301–318. PMLR, 2016.
- [3] Edward Choi, Mohammad Taha Bahadori, Jimeng Sun, Joshua Kulas, Andy Schuetz, and Walter Stewart. Retain: An interpretable predictive model for healthcare using reverse time attention mechanism. *Advances in neural information processing systems*, 29, 2016.
- [4] Martin R Cowie, Juuso I Blomster, Lesley H Curtis, Sylvie Duclaux, Ian Ford, Fleur Fritz, Samantha Goldman, Salim Janmohamed, Jörg Kreuzer, Mark Leenay, et al. Electronic health records to facilitate clinical research. *Clinical Research in Cardiology*, 106:1–9, 2017.
- [5] Jörg Degen, Christof Wegscheid-Gerlach, Andrea Zaliani, and Matthias Rarey. On the art of compiling and using ‘drug-like’ chemical fragment spaces. *ChemMedChem: Chemistry Enabling Drug Discovery*, 3(10):1503–1507, 2008.
- [6] Jacob Devlin, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova. Bert: pre-training of deep bidirectional transformers for language understanding. google ai language. *arXiv preprint arXiv:1810.04805*, 2019.

- [7] R Scott Evans. Electronic health records: then, now, and in the future. *Yearbook of medical informatics*, 25(S 01):S48–S61, 2016.
- [8] Yu-Ting Kao, Shu-Fen Wang, Meng-Hsiu Wu, Shwu-Huey Her, Yi-Hsuan Yang, Chung-Hsien Lee, Hsiao-Feng Lee, An-Rong Lee, Li-Chien Chang, and Li-Heng Pao. A substructure-based screening approach to uncover n-nitrosamines in drug substances. *Journal of Food and Drug Analysis*, 30(1):150, 2022.
- [9] Justin Klekota and Frederick P Roth. Chemical substructures that enrich for biological activity. *Bioinformatics*, 24(21):2518–2525, 2008.
- [10] Hung Le, Truyen Tran, and Svetha Venkatesh. Dual memory neural computer for asynchronous two-view sequential learning. In *Proceedings of the 24th ACM SIGKDD international conference on knowledge discovery & data mining*, pages 1637–1645, 2018.
- [11] Cecilia S Lee and Aaron Y Lee. Clinical applications of continual learning machine learning. *The Lancet Digital Health*, 2(6):e279–e281, 2020.
- [12] Xuan Lin, Zhe Quan, Zhi-Jie Wang, Tengfei Ma, and Xiangxiang Zeng. Kgnn: Knowledge graph neural network for drug-drug interaction prediction. In *IJCAI*, volume 380, pages 2739–2745, 2020.
- [13] Krikamol Muandet, David Balduzzi, and Bernhard Schölkopf. Domain generalization via invariant feature representation. In *International conference on machine learning*, pages 10–18. PMLR, 2013.
- [14] Ziad Obermeyer and Ezekiel J Emanuel. Predicting the future—big data, machine learning, and clinical medicine. *The New England journal of medicine*, 375(13):1216, 2016.
- [15] Chuleeporn Phanus-Umporn, Watshara Shoombuatong, Veda Prachayasittikul, Nuttapat Anuwongcharoen, and Chanin Nantasenamat. Privileged substructures for anti-sickling activity via cheminformatic analysis. *RSC advances*, 8(11):5920–5935, 2018.
- [16] Farah Shamout, Tingting Zhu, and David A Clifton. Machine learning for clinical outcome prediction. *IEEE reviews in Biomedical Engineering*, 14:116–126, 2020.
- [17] Junyuan Shang, Cao Xiao, Tengfei Ma, Hongyan Li, and Jimeng Sun. Gamenet: Graph augmented memory networks for recommending medication combination. In *proceedings of the AAAI Conference on Artificial Intelligence*, volume 33, pages 1126–1133, 2019.
- [18] Yanchao Tan, Chengjun Kong, Leisheng Yu, Pan Li, Chaochao Chen, Xiaolin Zheng, Vicki S Hertzberg, and Carl Yang. 4sdrug: Symptom-based set-to-set small and safe drug recommendation. In *Proceedings of the 28th ACM SIGKDD Conference on Knowledge Discovery and Data Mining*, pages 3970–3980, 2022.

- [19] Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N Gomez, Łukasz Kaiser, and Illia Polosukhin. Attention is all you need. *Advances in neural information processing systems*, 30, 2017.
- [20] Meng Wang, Mengyue Liu, Jun Liu, Sen Wang, Guodong Long, and Buyue Qian. Safe medicine recommendation via medical knowledge graph embedding. *arXiv preprint arXiv:1710.05980*, 2017.
- [21] David Weininger. Smiles, a chemical language and information system. 1. introduction to methodology and encoding rules. *Journal of chemical information and computer sciences*, 28(1):31–36, 1988.
- [22] Qitian Wu, Hengrui Zhang, Junchi Yan, and David Wipf. Handling distribution shifts on graphs: An invariance perspective. *arXiv preprint arXiv:2202.02466*, 2022.
- [23] Chaoqi Yang, Cao Xiao, Fenglong Ma, Lucas Glass, and Jimeng Sun. Safedrug: Dual molecular graph encoders for recommending effective and safe drug combinations. *arXiv preprint arXiv:2105.02711*, 2021.
- [24] Nianzu Yang, Kaipeng Zeng, Qitian Wu, Xiaosong Jia, and Junchi Yan. Learning sub-structure invariance for out-of-distribution molecular representations. *Advances in Neural Information Processing Systems*, 35:12964–12978, 2022.
- [25] Yutao Zhang, Robert Chen, Jie Tang, Walter F Stewart, and Jimeng Sun. Leap: learning to prescribe effective and safe treatment combinations for multimorbidity. In *proceedings of the 23rd ACM SIGKDD international conference on knowledge Discovery and data Mining*, pages 1315–1324, 2017.
- [26] Zhi Zheng, Chao Wang, Tong Xu, Dazhong Shen, Penggang Qin, Baoxing Huai, Tongzhu Liu, and Enhong Chen. Drug package recommendation via interaction-aware graph induction. In *Proceedings of the Web Conference 2021*, pages 1284–1295, 2021.
- [27] Jiajing Zhu, Yongguo Liu, Chuanbiao Wen, and Xindong Wu. Dgdfs: Dependence guided discriminative feature selection for predicting adverse drug-drug interaction. *IEEE Transactions on Knowledge and Data Engineering*, 34(1):271–285, 2020.