



基于知识图谱的潜在不适当用药预测*

林工钊¹, 滕飞^{1△}, 胡巧织², 金朝辉², 徐琰², Haibo Zhang³

1. 西南交通大学计算机与人工智能学院(成都 611756); 2. 四川大学华西医院 临床药学部(药剂科)(成都 610041);

3. 新西兰奥塔哥大学(达尼丁 9054)

【摘要】 目的 为提高潜在不适当用药(potentially inappropriate medication, PIM)预测的准确率,提出一种结合知识图谱和机器学习的PIM预测模型。方法 首先,基于2019版Beers标准,以知识图谱为基本结构,构建具有逻辑表达能力的PIM知识表示体系,实现从患者信息到PIM的推理过程。其次,利用分类器链算法建立每个PIM标签的机器学习预测模型,从数据中学习潜在特征关联。最后,根据样本量阈值,将知识图谱的部分推理结果作为分类器链上的输出标签,增加低频PIM预测结果的可靠性。结果 实验采用来自成都地区9家医疗机构的11 741份处方数据,对模型有效性进行评估。实验表明,该模型对于PIM数量预测的准确率为98.10%,F1值为93.66%,对于PIM多标签预测的汉明损失为0.06%,macro-F1为66.09%,与现有模型相比有着更高的预测精度。结论 该PIM预测模型具有更好的潜在不适当用药预测性能,并且对于低频PIM标签识别效果提升显著。

【关键词】 潜在不适当用药 机器学习 知识图谱 多标签分类

Knowledge Graph-Based Prediction of Potentially Inappropriate Medication LIN Gongchao¹, TENG Fei^{1△}, HU Qiaozhi², JIN Zhaohui², XU Ting², Haibo Zhang³. 1. School of Computing and Artificial Intelligence, Southwest Jiaotong University, Chengdu 611756, China; 2. Department of Pharmacy, West China Hospital, Sichuan University, Chengdu 610041, China; 3. University of Otago, Dunedin 9054, New Zealand

△ Corresponding author, E-mail: fteng@swjtu.edu.cn

【Abstract】 **Objective** To improve the accuracy of potentially inappropriate medication (PIM) prediction, a PIM prediction model that combines knowledge graph and machine learning was proposed. **Methods** Firstly, based on Beers criteria 2019 and using the knowledge graph as the basic structure, a PIM knowledge representation framework with logical expression capabilities was constructed, and a PIM inference process was implemented from patient information nodes to PIM nodes. Secondly, a machine learning prediction model for each PIM label was established based on the classifier chain algorithm, to learn the potential feature associations from the data. Finally, based on a threshold of sample size, a portion of reasoning results from the knowledge graph was used as output labels on the classifier chain to enhance the reliability of the prediction results of low-frequency PIMs. **Results** 11 741 prescriptions from 9 medical institutions in Chengdu were used to evaluate the effectiveness of the model. Experimental results show that the accuracy of the model for PIM quantity prediction is 98.10%, the F1 is 93.66%, the Hamming loss for PIM multi-label prediction is 0.06%, and the macroF1 is 66.09%, which has higher prediction accuracy than the existing models. **Conclusion** The method proposed has better prediction performance for potentially inappropriate medication and significantly improves the recognition of low-frequency PIM labels.

【Key words】 Potentially inappropriate medication Machine learning Knowledge graph Multi-label classification

随着社会经济的发展和医疗水平的提高,人均寿命的延长,老龄化问题日益加重。老年人作为一个特殊的群体,常伴有多种慢性疾病,多科就诊、多重用药现象普遍^[1],使得老年患者更容易发生药物不良反应^[2]。美国老年医学专家BEERS等^[3]在1991年首次提出潜在不适当用药(potentially inappropriate medication, PIM)的概念,主要指药物潜在不良风险超过预期收益,同时缺少较安全的可替代药物,并建立了用于评估PIM的Beers标准。但

由于各级医疗机构医务人员专业水平的差异,让大多数医生在开处方时便凭借自己的专业知识进行PIM判断并不现实^[4]。因此,研究PIM的自动化筛查具有重要意义。

目前的研究主要使用机器学习方法进行PIM预测。PATEL等^[4]使用逻辑回归和XGBoost模型来识别患有有关节炎老年人的潜在不适当用药。LIN等^[5]应用逻辑回归、CART树、随机森林和梯度提升树等模型,分析了不同处方药特征和阿片类药物相关的潜在不适当处方之间的关联。WU等^[6]基于404份患者处方,结合5种采样方式、3种特征筛选方式和18种机器学习算法建立了潜在不适当处方风险预警模型。然而,以上机器学习模型对数据有较强的依赖性^[7-9],在真实情况下,绝大部分PIM发生率低,

* 国家自然科学基金(No.62272398)、四川省科技厅项目(No. 2023 NSFSC1696)和四川大学华西医院学科卓越发展1.3.5工程项目(No. ZYJC 18028)资助

△ 通信作者, E-mail: fteng@swjtu.edu.cn

样本量少, 导致数据驱动的方法学习不充分且缺乏可靠性。

另一方面, 知识图谱作为一种以图模型描述事物关联的技术方法^[10], 能够有效整合医疗知识, 已在医疗知识问答^[11-12]、临床辅助决策^[13-14]、药物发现和推荐^[15-17]等方面显现出重要价值。因此, 本文提出了一种结合机器学习和知识图谱决策结果的PIM预测模型, 充分利用了数据模型的特征挖掘能力和知识模型的推理能力, 从而提升PIM识别效果。

1 对象与方法

1.1 研究对象

本研究从中国药学会医院处方分析合作项目中, 随机抽取2018年成都地区9家医疗机构老年科门诊患者处方数据^[18-19], 包括年龄、性别、诊断、药品通用名等。

纳入标准: 2018年1-12月老年科65岁以上门诊患者处方。排除标准: 存在信息缺失或不完整的处方。

依据2019版Beers标准中的99条药物标准, 由两名专业医生对患者处方进行独立审查并标注存在的PIM编号, 存在分歧时由第三名医生做出决定。

利用正则表达式对诊断中的疾病名称进行抽取。而诊断和药品通用名属于多值类别特征, 具有不定长的特点, 本文对其进行one-hot编码。考虑到老年患者共病现象普遍, 而多重用药会显著增加药物的相互作用和药物不良反应风险^[20], 因此在特征中加入疾病数量和药物数量。

本研究经四川大学华西医院生物医学伦理审查委员会批准(2020年审651号)。

1.2 任务定义

本文将PIM预测视为一个多标签分类问题。PIM的判断涉及多个因素, 包括患者的基本信息(性别、年龄)、药物处方(如服用某种药物)、疾病诊断(如被诊断患有某种疾病)。将每个因素作为变量 x_k , 得到特征集合 $X = \{x_1, x_2, \dots, x_k\}$, 目标是预测Beers标准中对应的PIM编号集合 $y = \{y_1, y_2, \dots, y_M\}$, 最后进行累加得到处方中的PIM数量, 其中 K 和 M 分别表示特征数和标签数, $y_j \in \{0, 1\}$ 表示是否存在第 j 条PIM情况。

1.3 PIM知识表示体系

知识图谱是一种图结构的知识表示方法, 能够对客观事物及其之间的语义关系进行形式化描述。目前已有的知识图谱缺少对实体间逻辑关系的表示, 导致其推理能力弱、应用场景受限。本文基于美国老年医学会发布的2019版Beers标准, 构建面向逻辑表达的PIM知识表示体系, 并在此基础上实现具有复杂关系的PIM推理。见图1。

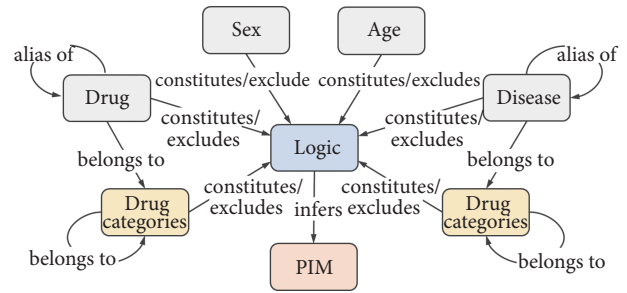


图 1 PIM知识表示体系结构

Fig 1 The structure of the PIM knowledge representation system

1.3.1 实体和关系表示

通过分析Beers标准内容和专业医生的判断流程, 发现PIM的评估不仅需要判断一些因素是否存在, 还可能排除另一些特定情况, 因此设计如图1所示的知识表示体系结构, 主要包括8类实体和5类关系。其中, “性别”“年龄”“药物”“疾病”“药物类别”“疾病类别”属于风险因素, 与“PIM”之间存在“构成”或“排除”关系。为了能够对逻辑关系进行表示, 在风险因素和PIM之间设立“logic”逻辑节点进行关联, “logic”和“PIM”之间形成“判断”关系。此外, 药物和疾病与其自身存在“别名”关系, 与其各自类别之间存在多级的“属于”关系。

1.3.2 逻辑表示

本文通过定义“logic”逻辑节点实现复杂关系的表示。每个逻辑节点的入边代表条件(风险因素), 出边代表结果(PIM编号)。针对不同的关系模式, 可以设置具有不同含义的逻辑节点进行表示。为了分析PIM评估中存在的逻辑关系, 我们将原有的每条标准内容进一步分解为更细粒度的判定单元, 发现可以将这些判定单元归纳为三类情况, 因此定义了三类逻辑节点(表1): 必要条

表 1 三类逻辑节点定义

Table 1 Definition of three types of logical nodes

Type	Definition	Case
Necessary condition node	All the constituent conditions exist.	+ Rivaroxaban + ≥ 75 yr. + Atrial fibrillation
Positive and negative condition node	All the constituent conditions exist and all the exclusion conditions do not exist.	+ Short- or rapid-acting insulin - Basal or long-acting insulin
Counting condition node	Any constituent condition occurs at least a specified number of times.	Anticholinergic ≥ 2

件节点、正负条件节点以及计数条件节点。例如,第45条PIM的内容为“达比加群或利伐沙班治疗 ≥ 75 岁患者VTE或房颤时慎用”,可以分解得到判定单元“+利伐沙班+ ≥ 75 岁+心房颤动”,其中“+”表示构成因素(与之相对,下文用“-”表示排除因素)。然后使用必要条件节点来表达该判定单元的逻辑,含义是如果同时出现以上三个因素则可以判断存在“PIM45”。除此之外,正负条件节点用来表示存在需要排除的情况,计数条件节点对风险因素的数量进行约束。

1.3.3 PIM推理

加入逻辑节点后,能够对实体之间的逻辑关系进行描述。本节给出基于前述知识表示体系设计的PIM搜索算法。当给定患者和处方的特征信息后,可以匹配从特征节点到PIM节点的子图,最后到达的节点即为处方中存在的PIM情况,其中逻辑节点起到了逻辑门的作用,能够对条件进行过滤。具体来说,首先匹配得到知识图谱中的起始节点集合start,搜索所有可以从start到达logic节点的路径path。然后根据path中logic节点类型可以分成两种情况:①若为计数条件节点且满足数量要求,则后继节点即为所求目标,加入Pims结果集合;②若入边为“构成”关系,则加入候选集合candidate并将前驱节点记录到pre集合,若入边为“排除”关系则进行剔除。最后,依次判断candidate中剩余的logic节点,如果其所有“构成”关系的前驱节点均包含于pre集合中,则将后继节点加入到

结果集合Pims中。

以图2左上方的患者数据为例,根据本节提出的算法,能够匹配得到右上方显示的三种主要逻辑节点和路径。对应PIM的Beers标准内容则在下方给出。判定过程为:对于“logic_2”正负条件节点,路径中包含本应排除的“冠心病”节点,因此不属于“PIM44”;“logic_1”必要条件节点要求同时存在“老年痴呆症”和“苯二氮草类药物”两种因素,但是数据中缺少“苯二氮草类药物”构成条件,所以也不存在“PIM37”;“logic_3”计数条件节点仅当6类药物中同时出现3种以上时成立,处方中的“奥氮平”、“氟西汀”和“艾斯西酞普兰”恰好属于规定类别的药物,即说明存在“PIM54”。最终可以判定该患者处方只存在“PIM54”一种潜在不适当用药情况。

1.4 基于知识图谱的PIM预测模型

本节提出了一个知识数据双驱动的PIM预测模型,针对样本量充足的PIM标签使用机器学习进行自动特征学习,而对于小样本PIM则利用上述知识图谱模型进行推理,弥补机器学习模型的不足,其整体结构如图3所示。

1.4.1 机器学习模型构建

为了有效预测一份患者处方中的多种PIM,构建基于CatBoost的分类器链作为机器学习模型。

分类器链(classifier chains, CC)算法^[21]是处理多标签任务的一种问题转换方法,能够学习标签之间的相互关

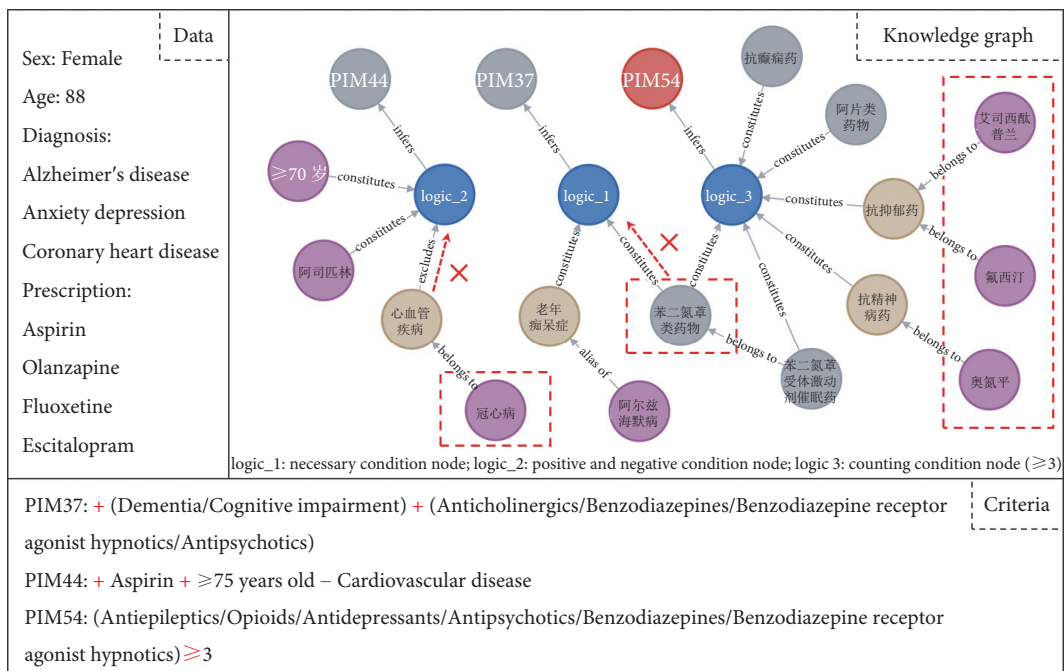


图 2 PIM推理案例

Fig 2 PIM inference case

The knowledge graph is constructed based on a Chinese corpus and is presented in the original Chinese form.

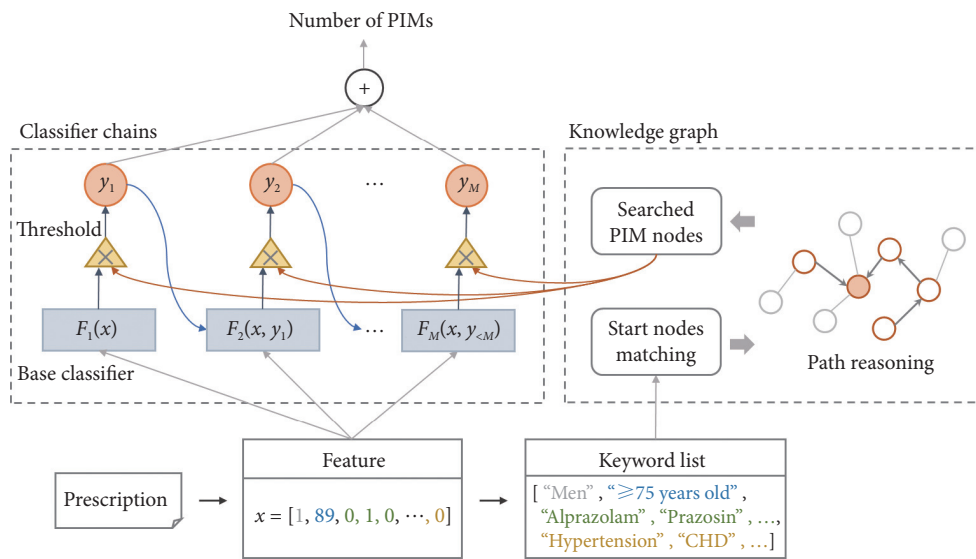


图 3 基于知识图谱的PIM预测模型
Fig 3 PIM prediction model based on knowledge graph

系。对于标签序列 $y = (y_1, y_2, \dots, y_M)$ ，分别构建 M 个分类器，训练过程中第 j 个分类器 F_j 的输入包括原始特征 x 和之前所有的标签 $(y_1, y_2, \dots, y_{j-1})$ 。预测阶段的标签序列 \hat{y} 计算过程如公式(1)所示。

$$\hat{y} = (F_1(x), F_2(x, F_1(x)), \dots) \quad (1)$$

本文选择以 CatBoost^[22] 模型作为基分类器。CatBoost 是一种以对称决策树 (oblivious trees) 为弱学习器的 Boosting 算法，训练过程为依次构建多个弱学习器 f ，每个学习器重点关注上一轮学习中误差较大的样本，最后对多个学习器进行线性组合，得到强学习器 F ，如公式(2)所示。

$$F(x) = \sum_{t=1}^T a_t f_t(x; \theta_t) \quad (2)$$

其中， T 为迭代次数， θ_t 为第 t 个弱学习器 f_t 的学习参数， a_t 为对应的权重。

与其他机器学习模型相比，CatBoost 能够更高效地处理类别特征。数据中的药物和疾病经过二值编码后形成高维稀疏的类别特征，而 PIM 的判断主要取决于这些风险因素的组合关系。CatBoost 能够在每次子树分裂时，将已选择的特征和数据集中其他特征进行组合，构造新的特征，从而使模型可以捕获风险因素之间的高阶依赖关系。

1.4.2 知识数据双驱动的 PIM 预测

如图 3 所示，在分类器链模型部分，输入是经过数值编码的患者和处方信息。所以在使用知识图谱模型之前首先将输入数据转换为特征词列表，例如当存在某种疾病时，列表中添加对应疾病名称。然后，分别使用机器学习模型和知识图谱推理模型得到预测结果。下一步，设定可调整的样本量阈值 λ ，当待预测标签 y_j 在数据集中的样本量低于 λ 时，使用知识图谱推理结果作为最终输出以及分类器链中后续基分类器 F_{j+1}, \dots, F_M 的输入，使后续基分类器准确率能够进一步提升。最后，对所有标签值 $(\hat{y}_1, \hat{y}_2, \dots, \hat{y}_M)$ 进行累加，得到 PIM 数量。该过程通过将知识图谱路径推理结果作为机器学习模型的补充，增加了结果可靠性。

习模型和知识图谱推理模型得到预测结果。下一步，设定可调整的样本量阈值 λ ，当待预测标签 y_j 在数据集中的样本量低于 λ 时，使用知识图谱推理结果作为最终输出以及分类器链中后续基分类器 F_{j+1}, \dots, F_M 的输入，使后续基分类器准确率能够进一步提升。最后，对所有标签值 $(\hat{y}_1, \hat{y}_2, \dots, \hat{y}_M)$ 进行累加，得到 PIM 数量。该过程通过将知识图谱路径推理结果作为机器学习模型的补充，增加了结果可靠性。

1.5 实验设置

由于当一份处方中 PIM 数量大于等于 5 时，属于严重的潜在不适当用药情况，发生概率较小，因此在评估时归为一类。数据集按照 7 : 3 的比例划分为训练集和测试集，在训练集上进行五折交叉验证选择最佳模型参数，在测试集上进行外部验证，评估模型的泛化性能。样本量阈值 λ 设为 30。

1.6 评价指标

本研究从 PIM 数量和 PIM 多标签两方面评价模型性能。

对于 PIM 数量，将取值视为不同的类别，然后使用多分类评价指标，包括准确率 (accuracy) 和宏平均计算的精确率 (precision)、召回率 (recall)、F1 值。

对于 PIM 多标签，分别使用基于样本的评价指标：子集准确率 (subset accuracy, SA)、汉明损失 (hamming loss, HL)，和基于标签的评价指标：macro-F1、micro-F1。SA 要求所有预测标签与真实标签完全一致。HL 衡量预测错误的标签在所有标签中的占比，越小越好。计算公式如下：

$$SA = \frac{1}{N} \sum_{i=1}^N [\widehat{y}_i = y_i] \quad (3)$$

$$HL = \frac{1}{N} \sum_{i=1}^N |\widehat{y}_i \Delta y_i| \quad (4)$$

其中, N 是样本量, \widehat{y}_i 是第*i*个样本的预测标签集合, y_i 是第*i*个样本的真实标签集合, $[\ast]$ 运算表示当 \ast 成立时为1, 否则为0, Δ 表示两个集合的对称差。

macro-F1倾向于小样本, 因此能够反映模型在低频PIM标签上预测效果, 而micro-F1反映模型在所有标签上的表现。计算公式如下:

$$macro-F1 = \frac{1}{M} \sum_{j=1}^M \frac{2TP_j}{2TP_j + FN_j + FP_j} \quad (5)$$

$$micro-F1 = \frac{2 \sum_{j=1}^M TP_j}{2 \sum_{j=1}^M TP_j + \sum_{j=1}^M FN_j + \sum_{j=1}^M FP_j} \quad (6)$$

其中, M 是标签数, TP_j 、 FP_j 、 FN_j 分别表示第*j*个标签真阳性、假阳性、假阴性的样本数量。

2 结果

2.1 数据集详情

构建的数据集中共有处方信息11 741条, 涉及41种PIM情况, 其中39种PIM情况发生率低于2.64%, 共包含2787维特征, 其中药物特征526维, 疾病特征2 257维。数据集各字段信息和PIM分布情况分别如表2和图4所示。

2.2 对比实验

将本文所提出方法与以下模型进行对比: Random Forest^[23]、XGBoost^[24]、CatBoost^[22]为3种常用于PIM预测研究的机器学习模型, AutoInt^[25]适合处理本任务中的高维稀疏数据, FT-Transformer^[26]是针对表格数据设计的

表 2 数据集字段信息

Table 2 Field information in the dataset

Number	Description	Range	Type
1	Sex	0, 1	Categorical
2	Age	65-119	Numeric
3	Number of diseases	1-19	Numeric
4	Number of drugs	1-23	Numeric
5-2261	Taking a certain kind of medicine	0, 1	Categorical
2262-2787	Suffering from a certain disease	0, 1	Categorical
2788-2828	A certain PIM exists (Target)	0, 1	Categorical
2829	Number of existing PIMs (Target)	0-10	Categorical

Transformer模型, DANets^[27]和T2G-FORMER^[28]分别构建了抽象网络模块和特征关系图, 在表格数据任务上具有突出的性能表现。实验中, 对比模型同样采用分类器链进行多标签的预测, 并将标签结果累加得到PIM数量。实验结果如表3所示。

在PIM数量预测方面, 机器学习模型F1值最高为89.93%, 而本文方法在结合知识图谱后F1值达到93.66%, 提升了3.73%, 并且准确率达到98.10%, 具有较高的临床应用价值。

在PIM标签预测方面, 本文方法能够准确预测97.98%的完整标签序列, 并且macro-F1达到66.09%, 与CatBoost相比提升1.36%, 与其他模型相比平均提升23.23%。其中, 编号为24、34、37的PIM标签预测结果如表4所示。在利用知识图谱进行路径推理后, F1值平均提高了18.56%, 说明模型能够有效提升低频PIM的预测效果。

2.3 消融实验

2.3.1 不同多标签分类策略对模型性能的影响

在本文模型中, 构建了分类器链作为多标签分类策略。为了验证其对模型的影响, 分别使用标签幂集(label powerset, LP)和二元关联(binary relevance, BR)^[21]进行替

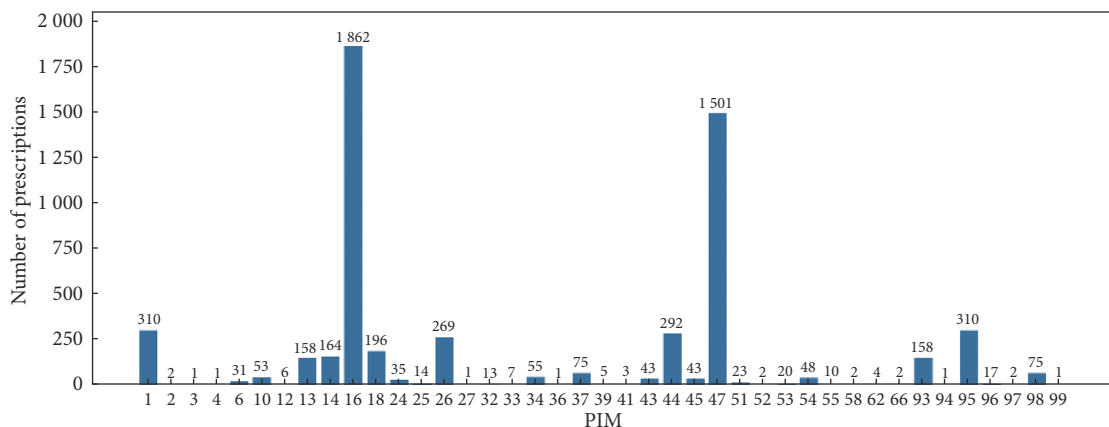


图 4 PIM分布情况

Fig 4 The distribution of PIMs

表 3 模型预测性能对比

Table 3 Comparison of prediction performance of the models

Model	Number of PIMs				PIM labels			
	Acc/%	Pre/%	Rec/%	F1/%	SA/%	HL/%	Macro-F1/%	Micro-F1/%
Random Forest	92.96	84.90	60.94	68.68	92.93	0.25	38.49	88.76
XGBoost	96.57	92.51	84.67	88.28	96.48	0.10	51.96	95.72
CatBoost	<u>97.76</u>	<u>92.97</u>	<u>87.37</u>	<u>89.93</u>	<u>97.62</u>	<u>0.07</u>	<u>64.73</u>	<u>97.27</u>
AutoInt	86.74	72.52	72.99	71.62	86.46	0.37	45.43	85.44
DANets	92.59	77.45	65.62	69.54	92.17	0.25	34.95	89.52
FT-Transformer	94.32	87.53	72.82	78.85	94.18	0.18	40.23	92.06
T2G-FORMER	95.32	81.03	73.50	76.83	95.23	0.14	46.72	93.97
Model proposed in the study	98.10	94.60	92.83	93.66	97.98	0.06	66.09	97.69

Acc: accuracy; Pre: precision; Rec: recall; SA: subset accuracy; HL: hamming loss. The top performances are marked in bold, and the second best results are underlined.

表 4 部分PIM标签预测结果比较

Table 4 Comparison of the prediction results for partial labels

PIM	Model	Acc/%	Pre/%	Rec/%	F1/%
24	CatBoost	99.80	87.50	53.85	66.67
	CatBoost+KG	99.94	86.67	100.00	92.86
34	CatBoost	99.86	100.00	64.29	78.26
	CatBoost+KG	99.89	100.00	71.43	83.33
37	CatBoost	99.69	100.00	52.17	68.57
	CatBoost+KG	99.91	100.00	86.96	93.02

Acc: accuracy; Pre: precision; Rec: recall.

换。其中, LP将不同的标签集合整体作为一个类别进行多分类任务, 然后根据样本量阈值, 用知识图谱推理得到的PIM标签进行替换; BR先为每个PIM标签训练一个独立的分类器, 再与知识图谱结合。实验结果如表5所示。

从表5可以看出, 使用分类器链处理多标签后, 模型在各项指标上均表现更好, 尤其是相比二元关联方法, 子集准确率进一步提高了0.20%, macro-F1提高了4.49%, 说明了分类器链通过引入其他标签, 使模型能够捕获标签

之间的关联信息, 从而有助于性能提升。

2.3.2 不同样本量阈值对模型性能的影响

本文模型根据样本量阈值选择性地加入知识图谱推理结果。为了验证其对模型的影响, 在不同的样本量阈值 λ 下进行实验, 结果如表6所示。其中 λ 取0时表示不加入知识图谱。

从表6可以看出, 在加入知识图谱后, PIM标签和数量的预测性能都有了一定提升。当 λ 取值为30时, PIM标签的macro-F1值达到最大, 此时模型整体预测效果最佳。随着 λ 增加, 模型性能出现下降, 说明机器学习和知识图谱之间存在一定的平衡状态。

2.4 知识图谱推理分析实验

为了说明预测结果的可靠性, 以PIM34为例进行分析, 其具体内容为: 心脏衰竭患者避免和谨慎使用西洛他唑、决奈达隆、非二氢吡啶类钙通道阻滞药、NSAIDs、COX-2抑制药、噻唑烷二酮, 共涉及至少37种具体药物。而数据中只有60份处方存在PIM34, 机器学习模型难以

表 5 不同多标签分类策略时的模型性能对比

Table 5 Comparison of model performance with different multi-label classification strategies

Strategy	Number of PIMs				PIM labels			
	Acc/%	Pre/%	Rec/%	F1/%	SA/%	HL/%	Macro-F1/%	Macro-F1/%
LP	96.96	<u>93.74</u>	<u>87.79</u>	<u>90.48</u>	96.74	00.10	54.15	95.72
BR	<u>97.90</u>	90.77	86.91	88.71	<u>97.76</u>	<u>00.06</u>	<u>61.60</u>	<u>97.47</u>
CC	98.10	94.60	92.83	93.66	97.98	00.06	66.09	97.69

LP: label powerset; BR: binary relevance; CC: classifier chains; Acc: accuracy; Pre: precision; Rec: recall; SA: subset accuracy; HL: hamming loss. The top performances are marked in bold, and the second best results are underlined.

表 6 不同样本量阈值时的模型性能对比

Table 6 Comparison of model performance at different sample size thresholds

λ	Number of PIMs				PIM labels			
	Acc/%	Pre/%	Rec/%	F1/%	SA/%	HL/%	Macro-F1/%	Micro-F1/%
0	97.76	92.97	87.37	89.93	97.62	0.07	64.73	97.27
10	97.81	<u>94.07</u>	87.50	90.44	97.76	0.06	<u>64.94</u>	97.32
30	98.10	94.60	92.83	93.66	97.98	<u>0.06</u>	66.09	<u>97.69</u>
50	<u>98.07</u>	92.70	<u>89.80</u>	<u>90.93</u>	<u>97.96</u>	0.05	64.71	97.75
100	97.28	88.25	82.54	84.98	97.19	0.07	61.67	96.99

Acc: accuracy; Pre: precision; Rec: recall; SA: subset accuracy; HL: hamming loss. The top performances are marked in bold, and the second best results are underlined.

充分学习重要特征,因此召回率较低。

利用SHAP^[29]计算某个样本的特征重要性,结果如图5所示。该样本存在PIM34但未能被CatBoost识别。其中,阿司匹林和洛索洛芬都属于PIM34的构成因素,但在数据集中,76%的PIM34与阿司匹林有关,而仅有5%的情况与洛索洛芬相关。由此可以解释为什么该样本中“未使用阿司匹林”比“使用洛索洛芬”具有更高的重要性,且对预测结果呈负贡献。相对地,利用知识图谱则可以显式推理出“洛索洛芬”属于NSAIDs药物,与“心衰”共同构成PIM34,推理路径如图6所示。

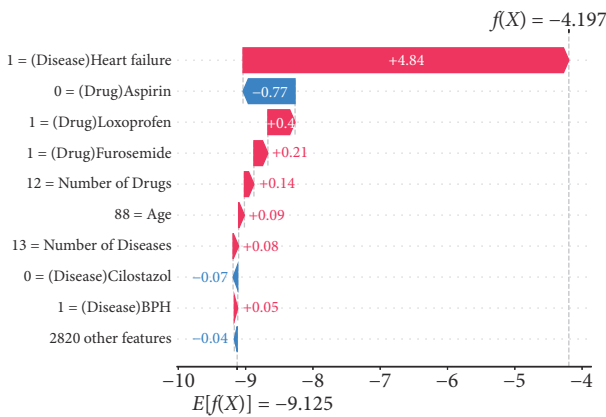


图 5 SHAP计算特征重要性

Fig 5 Feature importance calculated by SHAP

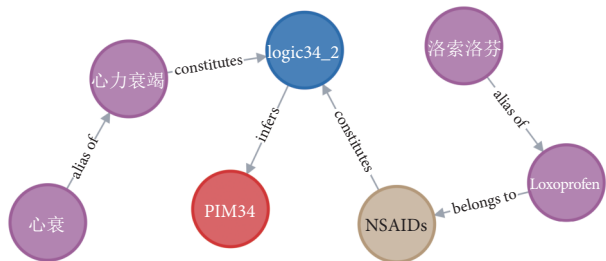


图 6 PIM34的路径查找过程

Fig 6 The path finding process of PIM34

The knowledge graph is constructed based on a Chinese corpus and is presented in the original Chinese form.

3 讨论

本文针对机器学习模型难以有效预测低频PIM的问题,提出了一种结合知识图谱的分类器链模型,实现知识数据双驱动的PIM预测过程。该模型首先在知识图谱中引入逻辑节点进行复杂逻辑关系的表示,并提出PIM路径搜索算法,将PIM判断过程转换为知识图谱中的路径推理,增加结果的可靠性。其次,构建基于CatBoost的分类器链模型捕获特征和标签的关联信息,实现数据驱动的PIM预测。最后,通过样本量阈值将知识图谱推理结果选择性加入到数据模型中,提升模型预测性能,并且减

少了机器学习方法对样本数据的依赖。

在对比实验中,本文方法相比未结合知识图谱的CatBoost模型,在macro-F1值上有明显提升,说明对于低频PIM标签,使用显式知识推理方式,可以解决因样本量少而导致的机器学习模型学习不充分的问题,从而提升预测效果。同时可以发现,深度学习模型与XGBoost、CatBoost等GBDT类机器学习模型存在一定差距,其可能原因是本任务中的特征具有高维稀疏的特性,而待预测PIM标签通常只与其中少数特征因素相关,导致深度学习模型更容易产生过拟合。

通过对样本量阈值λ进行消融实验可以发现,当λ持续增加时,更多的PIM标签通过知识图谱推理得到,此时子集准确率出现下降。可能原因是知识不够完善,而随着样本量增加,出现的疾病和药物种类增多,更容易导致初始节点匹配不准确,产生错误的推理结果。由此表明,利用机器学习的自动特征学习能力,也可以与知识图谱形成有效互补。

本文方法也存在一定局限性。一是知识图谱所包含的疾病和药物信息有限,需要随着药物品种和循证医学证据的更新而不断完善。二是纳入的处方中缺少一些PIM情况,导致无法评估本文模型在这些PIM上的表现。后续研究考虑扩大样本量,进一步对更多低频PIM的预测进行评估和分析。

本研究针对传统机器学习模型在PIM预测中的数据依赖问题,通过引入知识图谱推理,增加预测结果的准确性和可靠性。实验结果表明,该模型在PIM数量预测上准确率达到98.10%,F1值达到93.66%,在PIM多标签预测上macro-F1值达到66.09%,均优于当前主流模型。因此,本研究提出的方法可以有效预测PIM,并且显著提升低频PIM标签识别效果。

* * *

作者贡献声明 滕飞、金朝辉、徐琰和Haibo Zhang负责论文构思和监督指导,林工钊和胡巧织负责数据审编和正式分析,滕飞、金朝辉和徐琰负责经费获取,胡巧织负责调查研究,林工钊、滕飞和Haibo Zhang负责研究方法,林工钊负责软件和初稿写作,林工钊和滕飞负责阅读与编辑写作。所有作者已经同意将文章提交给本刊,且对将要发表的本刊进行最终定稿,并同意对工作的所有方面负责。

利益冲突 所有作者均声明不存在利益冲突

参 考 文 献

[1] 廖世莉,田方圆,张莹,等. 2018年成都9家医院老年科住院患者潜在不适当用药与相关风险因素分析. *药物流行病学杂志*, 2021, 30(4): 245-250. doi: 10.19960/j.cnki.issn1005-0698.2021.04.006.

[2] 王思蒙,张晨,孙雪,等. 社区老年人潜在不适当用药及其应对模式的

- 研究进展. *中国全科医学*, 2022, 25(13): 1551–1556. doi: 10.12114/j.issn.1007-9572.2022.0077.
- [3] BEERS M H, OUSLANDER J G, ROLLINGHER I, *et al.* Explicit criteria for determining inappropriate medication use in nursing home residents. *Arch Inter Med*, 1991, 151(9): 1825–1832. doi: 10.1001/archinte.1991.00400090107019.
- [4] PATEL J, LADANI A, SAMBAMOORTHY N, *et al.* A machine learning approach to identify predictors of potentially inappropriate non-steroidal anti-inflammatory drugs (NSAIDs) use in older adults with osteoarthritis. *Int J Environ Res Public Health*, 2021, 18(1): 155. doi: 10.3390/ijerph18010155.
- [5] LIN H C, WANG Z, HU Y H, *et al.* Characteristics of statewide prescription drug monitoring programs and potentially inappropriate opioid prescribing to patients with non-cancer chronic pain: a machine learning application. *Prev Med*, 2022, 161: 107116. doi: 10.1016/j.ypmed.2022.107116.
- [6] WU X, CHANG H, LI M, *et al.* A machine learning-based risk warning platform for potentially inappropriate prescriptions for elderly patients with cardiovascular disease. *Front Pharmacol*, 2022, 13: 804566. doi: 10.3389/fphar.2022.804566.
- [7] 潘崇煜, 黄健, 郝建国, 等. 融合零样本学习和小样本学习的弱监督学习方法综述. *系统工程与电子技术*, 2020, 42(10): 2246–2256. doi: 10.3969/j.issn.1001-506X.2020.10.13.
- [8] 欧阳宵, 陶红, 范瑞东, 等. 利用标签相关性先验的弱监督多标签学习方法. *软件学报*, 2023, 34(4): 1732–1748. doi: 10.13328/j.cnki.jos.006703.
- [9] 徐玲玲, 迟冬祥. 面向不平衡数据集的机器学习分类策略. *计算机工程与应用*, 2020, 56(24): 12–27. doi: 10.3778/j.issn.1002-8331.2007-0120.
- [10] 赵杨, 刘成翰, 杨涛, 等. 新型冠状病毒肺炎知识图谱研究现状与展望. *中华医学图书情报杂志*, 2022, 31(5): 32–44. doi: 10.3969/j.issn.1671-3982.2022.05.005.
- [11] XIA F, LI B, WENG Y, *et al.* MedConQA: medical conversational question answering system based on knowledge graphs//Proceedings of the The 2022 Conference on Empirical Methods in Natural Language Processing: System Demonstrations. Abu Dhabi, UAE: Association for Computational Linguistics, 2022: 148–158. doi: 10.18653/v1/2022.emnlp-demos.15.
- [12] HUANG J, CHEN Y, LI Y, *et al.* Medical knowledge-based network for Patient-oriented Visual Question Answering. *Inform Process Manag*, 2023, 60(2): 103241. doi: 10.1016/j.ipm.2022.103241.
- [13] ZHU C, YANG Z, XIA X, *et al.* Multimodal reasoning based on knowledge graph embedding for specific diseases. *Bioinformatics*, 2022, 38(8): 2235–2245. doi: 10.1093/bioinformatics/btac085.
- [14] YANG R, YE Q, CHENG C, *et al.* Decision-making system for the diagnosis of syndrome based on traditional Chinese medicine knowledge graph. *Evid Based Complement Alternat Med*, 2022, 2022: 8693937. doi: 10.1155/2022/8693937.
- [15] SU X, HU L, YOU Z, *et al.* Attention-based knowledge graph representation learning for predicting drug-drug interactions. *Brief Bioinform*, 2022, 23(3): bbac140. doi: 10.1093/bib/bbac140.
- [16] GAO Z, DING P, XU R. KG-predict: a knowledge graph computational framework for drug repurposing. *J Biomed Inform*, 2022, 132: 104133. doi: 10.1016/j.jbi.2022.104133.
- [17] ZHOU G, HAIHONG E, KUANG Z, *et al.* Clinical decision support system for hypertension medication based on knowledge graph. *Comput Methods Programs Biomed*, 2022, 227: 107220. doi: 10.1016/j.cmpb.2022.107220.
- [18] TIAN F, LI H, CHEN Z, *et al.* Potentially inappropriate medications in Chinese older outpatients in tertiary hospitals according to Beers criteria: a cross-sectional study. *Int J Clin Pract*, 2021, 75(8): e14348. doi: 10.1111/ijcp.14348.
- [19] HU Q, TIAN F, JIN Z, *et al.* Developing a warning model of potentially inappropriate medications in older Chinese outpatients in tertiary hospitals: a machine-learning study. *J Clin Med*, 2023, 12(7): 2619. doi: 10.3390/jcm12072619.
- [20] 郭和坚, 朱亚兰, 胡晓霞, 等. 基于Beers标准和STOPP/START标准评价神经内科老年患者出院带药处方潜在不适当用药情况. *临床药物治疗杂志*, 2022, 20(11): 65–71. doi: 10.3969/j.issn.1672-3384.2022.11.014.
- [21] ZHANG M L, ZHOU Z H. A review on multi-label learning algorithms. *IEEE Transact Knowl Data Engin*, 2014, 26(8): 1819–1837. doi: 10.1109/TKDE.2013.39.
- [22] VERONIKA DOROGUSH A, ERSHOV V, GULIN A. CatBoost: gradient boosting with categorical features support. arXiv e-prints, 2018: arXiv: 1810.11363. doi: 10.48550/arXiv.1810.11363.
- [23] BREIMAN L. Random forests. *Machine Learn*, 2001, 45: 5–32. doi: 10.1023/A:1010933404324.
- [24] CHEN T, GUESTRIN C. Xgboost: a scalable tree boosting system//Proceedings of the 22nd acm sigkdd international conference on knowledge discovery and data mining. San Francisco, USA, 2016: 785–794. doi: 10.1145/2939672.2939785.
- [25] SONG W, SHI C, XIAO Z, *et al.* AutoInt: automatic feature interaction learning via self-attentive neural networks//Proceedings of the 28th ACM international conference on information and knowledge management. Beijing, China, 2019: 1161–1170. doi: 10.1145/3357384.3357925.
- [26] GORISHNIY Y, RUBACHEV I, KHRULKOV V, *et al.* Revisiting deep learning models for tabular data. *Advance Neur Informat Process Syst*, 2021, 34: 18932–18943.
- [27] CHEN J, LIAO K, WAN Y, *et al.* DANets: deep abstract networks for tabular data classification and regression. *Proceed AAAI Conference Artif Intellig*, 2022, 36(4): 3930–3938. doi: 10.1609/aaai.v36i4.20309.
- [28] YAN J, CHEN J, WU Y, *et al.* T2g-former: organizing tabular features into relation graphs promotes heterogeneous feature interaction. *Proceed AAAI Conference Artif Intellig*, 2023, 37(9): 10720–10728. doi: 10.1609/aaai.v37i9.26272.
- [29] LUNDBERG S M, LEE S I. A unified approach to interpreting model predictions. *Advance Neur Informat Process Syst*, 2017, 30: 4765–4774.

(2023-06-05收稿, 2023-09-16修回)

编辑 余琳



开放获取 本文遵循知识共享署名—非商业性使用4.0国际许可协议(CC BY-NC 4.0), 允许第三方对本刊发表的论文自由共享(即在任何媒介以任何形式复制、发行原文)、演绎(即修改、转换或以原文为基础进行创作), 必须给出适当的署名, 提供指向本文许可协议的链接, 同时标明是否对原文作了修改; 不得将本文用于商业目的。CC BY-NC 4.0许可协议访问<https://creativecommons.org/licenses/by-nc/4.0/>。

© 2023 《四川大学学报(医学版)》编辑部 版权所有