

基于图卷积神经网络的中药方剂功效分类与自动推荐

赵丹宁¹, 卢海霞¹, 白 森^{2*}

1. 吉林省药品检验研究院, 吉林 长春 130033

2. 长春理工大学计算机科学技术学院, 吉林 长春 130013

摘 要: **目的** 实现海量中药方剂的功效自动化分类, 辅助建设更加健全完备的中药方剂信息化编码体系。**方法** 以中药方剂为样本, 以组成方剂的中药材和中药饮片为特征, 利用方剂主成分相似性对样本间关系进行表征, 设计一种图卷积神经网络模型, 并基于国家标准《中药方剂编码规则及编码》(GB/T31773—2015) 的方剂分类标准, 以其中 1 089 首经典古方剂的功效分类作为数据集进行模型训练。**结果** 在少量样本条件下, 可以实现中药方剂所在的 Top-6 功效分类 87.96% 的预测准确率, 与其他方法对比性能更占优势。以国标中的方剂功效信息进行验证, 可以很好地预测中药方剂的潜在功效。**结论** 基于图卷积神经网络的方法可以用于中药方剂的自动化分类, 帮助转变中药标准化工作中大量专业人力投入的工作模式, 并辅助中药研究者发现中药方剂的潜在临床功效。

关键词: 中药方剂; 自动分类; 功效预测; 图卷积; 神经网络

中图分类号: TP18; R289 文献标志码: A 文章编号: 0253-2670(2024)18-6298-07

DOI: 10.7501/j.issn.0253-2670.2024.18.019

Efficacy classification and automatic recommendation of traditional Chinese medicine formulas based on graph convolutional neural network

ZHAO Danning¹, LU Haixia¹, BAI Sen²

1. Jilin Institute for Drug Control, Changchun 130033, China

2. School of Computer Science and Technology, Changchun University of Science and Technology, Changchun 130013, China

Abstract: Objective To achieve automated classification of the therapeutic effects of a massive number of traditional Chinese medicine prescriptions and assist in the construction of a more robust and comprehensive information-based coding system for traditional Chinese medicine prescriptions. **Methods** Traditional Chinese medicine prescriptions were used as samples, with the traditional Chinese medicinal materials and traditional Chinese medicine decoction pieces comprising the features. The relationships between the samples were represented using the similarity of the main components of the prescriptions. A graph convolutional neural network model was designed, and based on the traditional Chinese medicine prescription classification standard in the national standard *Coding Rules and Coding of Traditional Chinese Medicine Prescriptions* (GB/T31773-2015), the therapeutic effect classification of 1 089 classic ancient traditional Chinese medicine prescriptions in the national standard was used as the dataset for model training. **Results** Under a limited number of samples, a prediction accuracy of 87.96% was achieved for the Top-6 therapeutic effect classifications of traditional Chinese medicine prescriptions. Comparative experiments with other methods demonstrated the significant performance advantages of the proposed method. Furthermore, a comparison with the therapeutic effect information in the national standard indicated that the proposed method can be used to predict the potential therapeutic effects of traditional Chinese medicine prescriptions. **Conclusion** The method based on graph convolutional neural networks can be used to automate the classification of traditional Chinese medicine prescriptions, helping to transform the work mode of extensive human resources in traditional Chinese medicine standardization and assisting traditional Chinese medicine researchers in discovering the potential clinical efficacy of traditional Chinese medicine prescriptions.

收稿日期: 2024-02-23

基金项目: 吉林省自然科学基金面上项目 (20240101349JC); 吉林省教育厅科学研究项目 (JJKH20240947KJ); 吉林省教育科学“十四五”规划 2023 年度课题 (GH23163)

作者简介: 赵丹宁 (1989—) 女, 医学信息学博士, 从事中医药数据分析研究。E-mail: zdn0525@126.com

*通信作者: 白 森 (1986—), 男, 副教授, 硕士生导师, 从事中医药数据分析研究。E-mail: baisenbx@126.com

Key words: traditional Chinese medicine formulas; automatic classification; efficacy prediction; graph convolution; neural network

中医作为一种古老且完整的治疗体系,经过数千年的发展与积累,形成了以中药方剂为核心的丰富的中药理论知识。以中国方剂数据库^[1]为例,其中收录了 710 余种古籍及现代文献中的 84 464 首方剂。海量的中药方剂给中药的临床合理用药、科学监督管理带来了挑战。为此,国家标准《中药方剂编码规则及编码》(GB/T31773—2015)等相关文献提出了以中药方剂功能属性及对应的中医证候为依据的分类及编码标准。然而,目前国标 GB/T31773—2015 仅从 1 800 余部中医经典著作的 9 万余首方剂中选择并完成了 1 089 首经典名方的分类工作。中药方剂数量之巨,为人工编纂并全面完成中药方剂功效分类和标准化任务带来了巨大的挑战。以人工方式完成中药方剂的功效分类面临着 2 个主要的困难:(1) 中药方剂数量巨大,人工标记成本过高;(2) 国标 GB/T31773—2015 总结和归纳了中药方剂的 20 种功效分类(解表方、泻下方、和解方、清热方、祛暑方、温里方、补益方、固涩方、安神方、开窍方、理气方、理血方、治风方、治燥方、祛湿方、祛痰方、消食方、驱虫方、涌吐方、痈疡方),将中药方剂准确归纳到所属分类需要具备丰富的中药专业知识,由专业技术人员完成,否则有可能影响到分类结果的可靠性。

系统地总结中药方剂所属的功能分类类目不仅是中医药标准化任务的一项重要课题,对于依据临床证候实现最终的中药用药自动推荐^[2]也十分重要,因此,中药方剂的自动分类具有较高的研究价值。近年来,图卷积神经网络^[3]被证明可以有效在图结构上对数据进行表征,相比于传统的机器学习方法,图卷积神经网络模型能够在对节点特征进行学习的同时,更好地捕获网络的拓扑信息,因此在很多分类任务上取得了良好的结果。为了更好地学习和表示功效分类-中药方剂-药味组成和功效分类-中药方剂-药味成分之间的关联关系,本研究对中药方剂进行图结构建模,并采用图卷积神经网络对中药方剂功效分类进行自动预测。

本研究依据国标 GB/T31773—2015 提供的数据集,研究中药方剂在 20 个功效上的分类。由于分类的类目较多,而带有标注的样本量较少(国标 GB/T31773—2015 中目前给出了 1 089 首中药方剂的分类),通过实验发现小样本条件下预测中药方剂

的唯一功效分类能够达到的准确率较低。因此,本研究采用了中药方剂的 Top-k 置信度分类,即最有可能的 k 个分类。预测中药方剂的 Top-k 分类还有以下 2 方面的价值:(1) 减少分类标注人员需要关注的分类类目范围,降低人工标注工作量;(2) 预测中药方剂的潜在辅助功效。中药方剂具有主要功效外,往往具有其他辅助功效。如国标 GB/T31773—2015 收录的《伤寒六书》中“柴葛解肌汤”具有解肌清热的功效,按其主要功效应当分类为“解表方”大类下的“辛凉解表”小类。但是该方剂同时具有“清热方”的辅助功效,中药方剂的辅助功效可以作为其临床用药的参考。根据推理结果给出中药方剂 Top-k 置信度的功效分类,可以在预测中药方剂主要功效分类的同时,预测中药方剂的潜在功效。

1 中药分类与图卷积神经网络

1.1 中药分类体系

中药分类体系的研究有着悠久的历史,中国最早的中药学著作《神农本草经》就按照药性、功效、毒性等将 365 种药物进行了归类。后世的中药分类方法逐步完善,分类理论也逐步丰富,传统方法可以概括为药性分类法和自然分类法。其中药性分类法根据中药的药性、功效进行分类;自然分类法根据中药的自然属性进行分类。近年来,一些现代的科学方法也被引入中药分类^[3-6],如基于药用部位、植物学分类、化学分类等分类方法。对中药分类体系的研究形成了一些关于中药分类的标准,如国家标准《全国主要产品分类与代码》中药部分《中药分类与编码》(GB/T 7635.1-2002)、《中药方剂编码规则集编码》(GB/T31773—2015)、《中药编码规则及编码》(GB/T31774—2015),以及地方标准《中药饮片与中药方剂编码规则》(SZJG 38—2011)等。其中国标 GB/T31773—2015 和 GB/T31774—2015 是最新的权威标准。国标 GB/T31774—2015 主要采用自然属性分类法对中药材进行分类,其体系涵盖了中药材的药用来源、药用部位、药材种类、炮炙方法等;国标 GB/T31773—2015 主要采用药性分类法,按照功效对中药方剂进行了分类。

可见,虽然对中药分类已经有了丰富的研究并形成了国家标准,但是现有研究基本集中在如何构建中药分类体系上,而利用机器学习的模型方法,对中药进行自动化分类的方法研究仍然较少。

1.2 图卷积神经网络

图卷积神经网络是一种用于处理图数据的深度学习模型。与传统的卷积神经网络主要应用于处理图像数据相比,图卷积神经网络可以看作卷积神经网络在图数据上的推广。目前对图卷积网络的主要研究大多是以 Kipf 等^[3]提出的图卷积网络模型的改进或变种,该模型在每层神经网络上对图中节点及其相邻节点的特征进行卷积计算并嵌入到新的向量中。近年来对图卷积网络的研究可以大致分为 2 个方向:(1)从模型性能、模型结构上对图卷积网络进行改进和优化,如 Chiang 等^[7]研究了大规模图上的图卷积算法,Khan 等^[8]研究了当图中具有异构的边时的图卷积网络,Jiang 等^[9]研究了二分图上的图卷积网络。(2)研究图卷积方法在特定领域中的应用。例如,Zhao 等^[10]、戴俊明等^[11]研究了图卷积神经网络用于交通预测,Zhang 等^[12]、蒋宗礼等^[13]研究了基于图卷积网络在知识图谱等领域进行链接预测,Jia 等^[14]在图像处理中使用图卷积神经网络进行图像识别。

在医学、药学相关领域中,也有很多基于图卷积神经网络的应用研究,如 Guo 等^[15]基于图卷积神经网络进行药物相互作用的预测,Tian 等^[16]将图卷积网络运用于三维医学图像处理中。在中医药领域,Jin 等^[17]将多图卷积神经网络应用到中药推荐应用中。虽然目前将机器学习方法与中医药研究结合正逐渐成为新的研究趋势^[17-19],但基于图卷积网络与深度学习的药物分类方法和科学研究仍然较少。

2 研究方法

2.1 数据采集

本研究使用的中药方剂与药味组成、药味主成分数据主要来源于国标 GB/T31773—2015 及国标 GB/T31774—2015。

2.1.1 数据来源

(1) 国标 GB/T31773—2015: 收录了 1 089 首中药方剂并将其归纳入 20 个功效分类中,中药方剂由多种中药材(中药饮片)组成。同一味中药材经过不同的炮制方法可以生成不同的中药饮片,以半夏(中药材)为例,为了对其进行加工炮制以降低其自身的毒性,在中医临床应用中半夏分为生半夏、清半夏、姜半夏和法半夏等中药饮片,且其功效与临床应用也不同。

(2) 国标 GB/T31774—2015: 进一步对中药饮片进行归类,通过分类编码对中药饮片进行标准化。因此,可以依据不同中药饮片对应的编码确定其是否由同种中药材炮制而成。如图 1 所示,一种中药饮片基于一种中药材制成,一种中药材可以制成多种中药饮片;一种中药方剂由多种中药饮片组成。国标 GB/T31774—2015 共收录了 1 219 种中药材及 1 603 味中药饮片,本研究提出的基于图卷积神经网络的模型训练和验证主要以上述国标中的中药方剂、中药饮片以及中药材作为源数据。

2.1.2 数据特征 由于国标中的中医药数据仅以文本形式提供,因此本研究采用了基本的自然语言处理方法对源数据进行了信息抽取与结构化,并以图 1 所示的图结构的形式进行存储。该数据集中还包

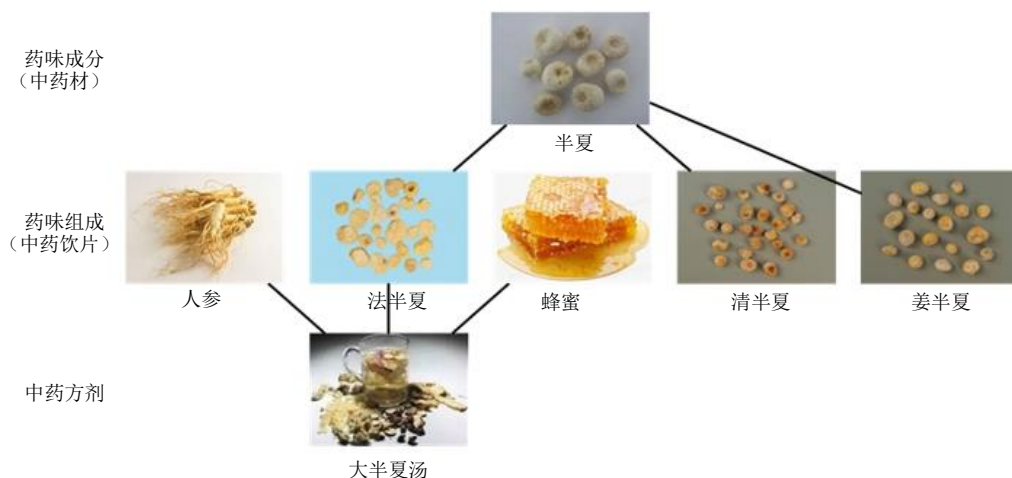


图 1 中药材、中药饮片、中药方剂关系示意图

Fig. 1 Relationship diagram of traditional Chinese medicinal materials, traditional Chinese medicine decoction pieces and traditional Chinese medicine formulas

含中药方剂相关的其他类型特征，如国标 GB/T31773—2015 给出了其收录的中药方剂的来源著作、功效描述、主治症状描述等。本研究仅采用中药方剂药味组成和药味成分相关的中药饮片、中药材作为特征进行建模，主要基于以下原因：（1）相比于功效描述等信息，中药方剂的组成对于不同文献来源具有较高的一致性；（2）中药方剂的药味组成是实现其功效的核心要素，功效、主治症状等信息内在蕴含于药味组成信息中；（3）中药方剂组成信息容易获取，基于其进行分类可以保证方法的实用性。

需要说明的是，训练样本的数量是影响预测模型准确率的关键因素之一。然而，由于以下原因，本研究仅采用了国标中千余首中药方剂作为样本：（1）国标给出的标注权威准确，其他数据源的数据难以保证权威性，会影响预测结果的可信度；（2）中药分类问题专业性强，标注困难且标注成本高。例如，《金匱要略》中给出“麻黄杏仁薏苡甘草汤”的功用为“发汗解表，祛风除湿”，按照国标分类，非专业人

员难以判断应将其分到“解表方”还是“祛湿方”的类别，难以提供如国标中权威的标注数据。

2.2 中药方剂分类算法模型

模型构建可以分为3部分：（1）根据图1中所示的方剂特征，构建包括中药饮片和中药材信息的中药方剂特征矩阵；（2）构建以中药方剂为节点的网络拓扑；（3）设计节点分类器。通过对卷积神经网络进行训练，最终使每个节点对应的特征向量收敛为低维稠密向量表示，并通过分类器输出得到方剂所属类别。

2.2.1 中药方剂特征矩阵构建 如图2所示，中药方剂对应的节点特征由2部分组成，分别是药味组成（中药饮片）和药味成分（中药材）。假设共有 m_1 个中药饮片和 m_2 个中药材，对中药饮片和中药材分别进行 Multi-Hot 编码，则对于1个中药方剂 v_i 可以将其特征向量表示为 $x_i \in \mathbb{R}^{m_1+m_2}$ 。对于1组中药方剂 $\{v_1, v_2, \dots, v_n\}$ ，其特征矩阵可以表示为 $X \in \mathbb{R}^{n \times (m_1+m_2)}$ 。本研究使用的数据集中， $X \in \mathbb{R}^{1089 \times (1603+1219)}$ 。

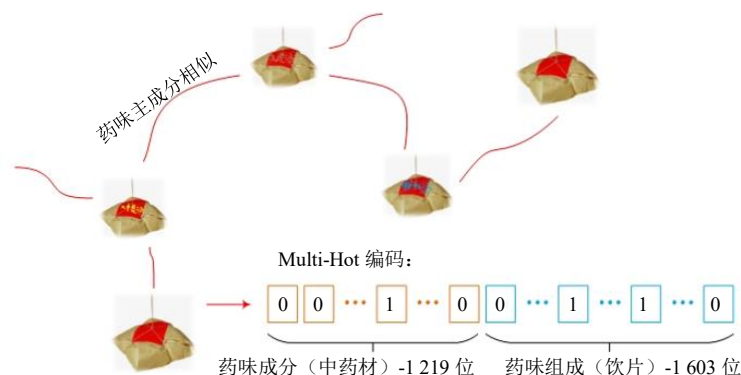


图2 中药方剂的拓扑关系构建与节点特征初始化

Fig. 2 Topological relationship construction and node feature initialization of traditional Chinese medicine formulas

2.2.2 中药方剂节点网络拓扑构建 如图2所示，本研究基于中药方剂的药味组成相似性构建节点间的网络拓扑。将组成中药方剂的前3个中药饮片作为中药方剂的主要成分，当2个中药方剂至少有2个主要成分相同时，认为2个中药方剂为的主要成分相似，并将图2中2个中药方剂设置为邻居节点。判断中药方剂是否为邻居节点的示例见表1，中药方剂五虎汤和华盖散的主要成分中均含有麻黄和苦杏仁，因此它们在图2中为相邻节点。

2.2.3 节点分类器设计 本研究采用2层的图卷积网络，其中第2层神经网络将节点的特征向量嵌入

表1 主要成分相似的中药方剂示例

Table 1 Examples of traditional Chinese medicine formulas with similar main ingredients

方名	药味组成
五虎汤	麻黄、苦杏仁、石膏、甘草、细辛、桑白皮、生姜
华盖散	麻黄、苦杏仁、炙甘草、桑白皮、茯苓、炒紫苏子、陈皮

1个长度与中药方剂类目数量相等的向量中，并输入 softmax 分类器。本研究使用的图卷积网络可以表示为： $Z = \text{softmax}(\tilde{A} \text{ReLU}(\tilde{A}XW_0)W_1)$ ，其中 $\tilde{A} = D^{-1/2}AD^{1/2}$ ， $\text{softmax}(x_i) = \frac{\exp(x_i)}{\sum_j \exp(x_j)}$ 。

本研究使用的交叉熵损失函数：

$L = -\frac{1}{n} \sum_{i=1}^n \sum_{j=1}^c y_{ij} \ln(p_{ij})$ ，其中 n 为中药方剂样本数， c 为中药方剂类目数量， y_{ij} 为符号函数（当样本 i 真实类目为 j 时取 1，否则为 0）， p_{ij} 为预测样本 i 属于类目 j 的概率。

2.3 中药方剂的 Top-k 分类与潜在功效预测

国标 GB/T31773—2015 将中药方剂划分为 20 个类目，较多的类目数量和较少的带标注的样本导致自动预测中药方剂所在的单一类目准确率较低。因此，本研究提出可以对中药方剂所在的 Top-k 功效分类进行预测，即最有可能的 k 个分类功效类别。

本研究通过 softmax 分类器输出的每个类目的预测概率进行 Top-k 类目的提取。形式化地，对于 1 个中药方剂 v ，经过 softmax 分类器推理得到的 v 的各类目预测概率为 $\{p_1, p_2, \dots, p_c\}$ ， c 为类目数量，若其中 Top-k 最大值为 $\{p_{t_1}, p_{t_2}, \dots, p_{t_k}\}$ ，则该中药方剂的 Top-k 分类为 {类目 t_1 ，类目 t_2 ，…，类目 t_k }。

3 结果

通过对比图卷积神经网络与随机森林、支持向量机、XGBoost 等传统机器学习方法在中药方剂功效分类上的性能差异，验证图卷积神经网络在该分类预测问题上的优势。

3.1 参数设置

采用国标 GB/T31773—2015 给出的 1 089 味中药方剂作为样本。对于图卷积网络，采用 2 层的神经网络，设置隐层的嵌入向量的长度为 256，边的权重设置为 0.05，学习率（learning rate）为 0.01，丢弃率（dropout rate）为 0.5。采用 Adam 优化器^[20]，设置最大训练周期数（epoch）为 200，并当验证集的损失值连续超过 10 轮不再下降时终止训练。

3.2 性能测试

首先从 2 个维度测试基于图卷积网络的中药方剂自动分类方法性能：（1）对于不同的 k ，算法在预测 Top-k 分类时的准确率；（2）图卷积网络方法和 XGBoost、随机森林、支持向量机等其他分类方法的性能比较。

其他算法使用了与图卷积网络中节点相同的特征矩阵，并对模型参数进行了调优。XGBoost、随机森林、支持向量机基于 python 工具包 scikit-learn 实现。（1）XGBoost 参数：学习率为 0.1；n_estimators（树的个数）为 1 000；max_depth（树的深度）为 6；subsample（样本子采样）为 0.8；colsample_btree（特征子采样）为 0.8；objective（目标函数）为 multi:softmax。（2）随机森林参数：学习率为 0.1；criterion（评估准则）为 gini；n_estimators 为 1 000；max_depth 为 6；min_samples_leaf 为 1；bootstrap（是否使用自助法）为 true。（3）支持向量机参数：kernel（核函数）为 linear；probability（是否启用概率估计）为 true。

将中药方剂样本按照 80%、10%、10% 的比例划分为训练集、验证集、测试集。针对每个 Top-k 分类作为预测目标，表 2 列出图卷积网络、XGBoost、随机森林、支持向量机等机器算法的预测结果。

本研究还测试了训练样本数量对中药方剂分类预测准确率的影响。固定验证集中样本数量为 10%，训练集中样本数量分别为 80%、70%、60%，方法的预测准确率如表 3 所示。

3.3 功效分类结果分析

从表 2 可以看出：（1）由于当前训练样本数量较少，且分类类目较多，基于图卷积网络的方法对 Top-1 分类的准确率仅能达到 0.541 2。但是，随着 k 的增大，Top-k 分类的预测准确率显著上升，Top-

表 2 中药方剂 Top-k 分类预测准确率 ($\bar{x} \pm s, n=10$)

Table 2 Top-k classification accuracy of traditional Chinese medicine formulas ($\bar{x} \pm s, n=10$)

Top-k	预测准确率			
	图卷积网络	XGBoost	随机森林	支持向量机
Top-1	0.541 2±0.010 6	0.490 8±0.024 4	0.489 0±0.039 8	0.506 4±0.040 5
Top-2	0.681 9±0.011 5	0.664 7±0.021 7	0.661 9±0.032 7	0.670 6±0.027 9
Top-3	0.754 6±0.011 3	0.749 1±0.023 5	0.752 8±0.023 4	0.759 7±0.026 7
Top-4	0.809 2±0.010 9	0.805 5±0.023 9	0.802 3±0.019 2	0.805 3±0.024 8
Top-5	0.845 4±0.009 8	0.848 2±0.017 2	0.841 0±0.014 8	0.846 3±0.018 3
Top-6	0.879 6±0.010 4	0.878 4±0.018 1	0.879 0±0.010 8	0.883 3±0.018 3
Top-7	0.900 9±0.010 0	0.900 0±0.017 7	0.901 9±0.016 4	0.906 7±0.015 5

表3 不同训练样本数量下分类预测准确率

Table 3 Prediction accuracy with different numbers of training samples

Top-k	不同训练样本数量下的预测准确率		
	80%	70%	60%
Top-1	0.541 2	0.466 7	0.450 9
Top-2	0.681 9	0.627 2	0.608 6
Top-3	0.754 6	0.714 8	0.695 8
Top-4	0.809 2	0.773 8	0.760 4
Top-5	0.845 4	0.814 7	0.803 7
Top-6	0.879 6	0.844 1	0.838 4
Top-7	0.900 9	0.879 0	0.862 5

表4 预测得到的 Top-3 功效与国标提供的功效信息对比

Table 4 Comparison of predicted Top-3 effects with effect information provided by national standards

方名	Top-3 功效预测	国标 GB/T31773—2015 的功效信息		
		分类	功效	主治
麻黄汤	解表, 理气, 祛痰	解表	发汗解表, 宣肺平喘	外感风寒表实证; 恶寒发热, 头痛身疼, 无汗而喘; 舌苔薄白, 脉浮紧
桂枝汤	解表, 补益, 温里	解表	解肌发表, 调和营卫	外感风寒表虚证; 恶风发热, 汗出头痛, 鼻鸣干呕, 口不渴; 舌淡苔白, 脉浮缓或浮弱
九味羌活汤	解表, 治风, 祛湿	解表	发汗祛湿, 兼清里热	外感风寒湿邪, 内有蕴热证。恶寒发热, 无汗, 头痛项强, 肢体酸楚疼痛, 口苦而渴; 舌苔白或微黄, 脉浮
大承气汤	泻下, 理气, 祛湿	泻下	峻下热结	阳明腑实证; 症见大便不通, 频转矢气, 脘腹痞满, 腹痛拒按, 按之硬, 日晡潮热, 神昏谵语, 手足濇然汗出, 舌苔黄燥起刺或焦黑燥裂, 脉沉实, 或热结旁流, 下得清水, 色纯青, 脐腹疼痛, 按之坚硬有块, 或里热实证之热厥, 疮病或发狂等
大黄牡丹汤	泻下, 理血, 清热	泻下	泻热破瘀, 散结消肿	肠痈初起; 右少腹疼痛拒按, 甚则局部有痞块, 发热恶寒, 自汗出, 或右足屈而不伸; 苔黄腻, 脉滑数
小承气汤	泻下, 理气, 消食	泻下	轻下热结, 除满消痞	伤寒阳明腑实证, 谵语潮热, 大便秘结, 胸腹痞满舌苔黄, 脉滑数; 痢疾初起, 腹中痞痛, 或脘腹胀满, 里急后重者

对于“解表”类方剂,以麻黄汤为例,除了主要功效“解表”外,预测表明其可能具有理气、祛痰的功效,而国标中给出的功效“宣肺平喘”证明麻黄汤具有一定的理气祛痰的功效。桂枝汤具有“调和营卫”的功效,这与预测得到的补益、温里功效相吻合。九味羌活汤预测得到治风、祛湿的潜在功效,与国标给出的“发汗祛湿”功效及主治“外感风寒湿邪”症状相印证。对于“泻下”类方剂,以大黄牡丹汤为例,预测得到理血、清热功效,与“泻热破瘀,散结消肿”的功效相吻合;小承气

汤预测的理气、消食功能,也与“轻下热结,除满消痞”的功效相吻合。

7 准确率可以达到 90%以上。(2) 相比于 XGBoost、随机森林、支持向量机等其他分类方法,在 k 值较小时,图卷积方法的分类准确率具有一定优势。

从表 3 可以看出,训练样本数量的增加可以显著提升预测准确率。因此提升权威的标注数据数量、质量应当作为优化中药自动化分类的后续重点工作。

3.4 潜在功效预测结果分析

本研究选择了国标 GB/T31773—2015 中的前 2 个功效分类类目“解表”和“泻下”,以及每个功效下的前 3 个中药方剂进行潜在功效预测结果的分析,结果见表 4。

对其他类目的功效相关的中药方剂也进行了上述人工分析,分析结果表明本研究提供的中药方剂 Top-k 分类方法可以用于预测中药方剂主要功效外的潜在功效,可以帮助中药从业者更好地进行用药决策。

4 结论

本研究采用图卷积神经网络建立了中药方剂自动分类方法,基于中药方剂所含的中药材、中药饮

片信息,对中药方剂的 Top-k 功能分类进行预测。Top-k 功能分类的预测或可用于提升分类人员的标注效率,以及预测中药方剂的潜在辅助功效,以帮助临床用药决策。虽然由于带标注样本量较少的现状,本方法在 Top-1 分类预测上准确率仍然较低,但是随着未来国标中权威分类标注数据的扩充,本方法性能仍然有较大的提升空间。由于本研究使用的中药材、中药饮片特征信息在不同的资料中一致性、通用性较强,易于在实际中使用,同时也易于未来在更多数据集上推广。

利益冲突 所有作者均声明不存在利益冲突

参考文献

- [1] 中国中医科学院中医药信息研究所. 中国方剂数据库 [EB/OL]. (2023-07-23) [2023-07-23]. <https://www.ncmi.cn/project/project-showProjectList.html?type=data&id=d5,2023-7-23/2023-7-23>.
- [2] Jin Y Y, Zhang W, He X N, *et al.* Syndrome-aware herb recommendation with multi-graph convolution network [A] // 2020 IEEE 36th International Conference on Data Engineering (ICDE) [C]. Dallas: IEEE, 2020: 145-156.
- [3] Kipf T N, Welling M. Semi-supervised classification with graph convolutional networks [J/OL]. *arXiv* (2016-09-09) [2023-07-26]. <http://arxiv.org/abs/1609.02907>.
- [4] 张建华, 刘玲, 高军, 等. 中药分类编码及其应用价值 [J]. 中国中药杂志, 2003, 28(6): 583-585.
- [5] 刘玲, 高军, 沈绍武. 《中药分类与代码》研究 [J]. 中国中医药信息杂志, 2002, 9(9): 4-5.
- [6] 王红刚, 刘基柱, 潘利明. 基于植物分类系统对《中药分类与编码》改进初探 [J]. 中草药, 2013, 44(11): 1514-1516.
- [7] Chiang W L, Liu X Q, Si S, *et al.* Cluster-GCN: An efficient algorithm for training deep and large graph convolutional networks [J/OL]. *arXiv* (2019-05-20) [2023-08-01]. <http://arxiv.org/abs/1905.07953>.
- [8] Khan M R, Blumenstock J E. Multi-GCN: Graph convolutional networks for multi-view networks, with applications to global poverty [J]. *Proc AAAI Conf Artif Intell*, 2019, 33(1): 606-613.
- [9] Jiang Y L, Lin H J, Li Y, *et al.* Exploiting node-feature bipartite graph in graph convolutional networks [J]. *Inf Sci*, 2023, 628: 409-423.
- [10] Zhao L, Song Y J, Zhang C, *et al.* T-GCN: A temporal graph convolutional network for traffic prediction [J]. *IEEE Trans Intell Transp Syst*, 2020, 21(9): 3848-3858.
- [11] 戴俊明, 曹阳, 沈琴琴, 等. 基于多时空图卷积网络的交通流预测 [J]. 计算机应用研究, 2022, 39(3): 780-784.
- [12] Zhang J R, Huang J, Gao J L, *et al.* Knowledge graph embedding by logical-default attention graph convolution neural network for link prediction [J]. *Inf Sci*, 2022, 593: 201-215.
- [13] 蒋宗礼, 张文婷, 张津丽. 异质网中基于图卷积神经网络的链路预测方法 [J]. 计算机工程与设计, 2022, 43(1): 150-156.
- [14] Jia S, Jiang S G, Zhang S Y, *et al.* Graph-in-graph convolutional network for hyperspectral image classification [J]. *IEEE Trans Neural Netw Learn Syst*, 2022, 35(1): 1157-1171.
- [15] Guo L, Lei X J, Chen M, *et al.* MSResG: Using GAE and residual GCN to predict drug-drug interactions based on multi-source drug features [J]. *Interdiscip Sci*, 2023, 15(2): 171-188.
- [16] Tian F R, Tian Z Q, Chen Z, *et al.* Surface-GCN: Learning interaction experience for organ segmentation in 3D medical images [J]. *Med Phys*, 2023, 50(8): 5030-5044.
- [17] Jin Y Y, Ji W D, Zhang W, *et al.* A KG-enhanced multi-graph neural network for attentive herb recommendation [J]. *Transact Comput Biol Bioinform*, 2021, 19(5): 2560-2571.
- [18] Zhao W, Lu W, Li Z, *et al.* Traditional Chinese medicine herbal prescription recommendation model based on multi-graph convolutional network [J]. *J Ethnopharmacol*, 2022, 297: 115109.
- [19] 赵书言, 王书睿, 申镇华, 等. 基于人工智能算法的中药主归经络判别研究 [J]. 中草药, 2024, 55(16): 5573-5582.
- [20] Kingma D, Ba J. Adam: A method for stochastic optimization [J/OL]. *arXiv* (2014-12-22) [2023-08-26]. <https://d-o-i.org/10.48550/arXiv.1412.6980>.

[责任编辑 潘明佳]