

基于多标记深度森林的膝骨关节炎智能辅助诊断方法

龙锦益^{1,2,3}, 杨宇¹, 张子龙¹, 叶倩云^{2,4}, 吴汉瑞¹, 张荣华^{2,5,6}, 张佳^{1,2,*}

(¹暨南大学信息科学技术学院, 广州 510632; ²暨南大学广东省中医药信息技术重点实验室, 广州 510632; ³广州琶洲实验室, 广州 510335; ⁴暨南大学中医学院, 广州 510632; ⁵暨南大学药学院, 广州 510632; ⁶暨南大学癌症研究所, 广州 510632)

摘要: 目的: 利用多标记深度森林算法构建退行性膝骨关节炎智能辅助诊断模型, 并探索多标记方法在中医数据集上的优势。方法: 基于 1421 例临床样本, 使用多标记深度森林算法构建分类模型, 在 6 个评价指标上与其他 5 种多标记算法进行对比; 使用多标记算法 Rank-SVM、ML-kNN 和单标记算法 SVM、kNN 建模并对比。结果: 使用多标记深度森林构建的分类模型的汉明损失为 0.0458、1-错误率为 0.3521, 覆盖率为 4.077、排序损失为 0.1582、平均精度为 0.7394、准确率为 0.9542, 在六个评价指标上的表现优于其他 5 种对比算法, 其中在 KOA 上的 AUC 为 0.8123 远高于对比算法。结论: 在多证兼夹的中医数据集上, 多标记分类算法性能优于其对应的单标记算法, MLDF 算法在 KOA 的诊断结果上与真实的诊断的一致性较好, 具有较好的推广和应用前景。

关键词: 退行性膝骨关节炎; 人工智能; 机器学习; 多标记学习

基金资助: 国家自然科学基金项目(No. 62106084 & No. 61773179); 广东省国家自然科学基金项目(No. 2022a1515010468 & No. 2019A1515012175); 广州市科技计划项目(No.202201010498), 广东省中医药信息化重点实验室(No.2021B1212040007)

Intelligent assisted diagnosis of knee osteoarthritis based on multi-label deep forest

LONG Jin-yi^{1,2,3}, YANG Yu¹, ZHANG Zi-long¹, YE Qian-yun^{2,4}, WU Han-rui¹, ZHANG Rong-hua^{2,4,5},
ZHANG Jia^{1,2,*}

(¹College of Information Science and Technology, Jinan University, Guangzhou 510632, China; ²Guangdong Key Lab of Traditional Chinese Medicine Information Technology Jinan University, Guangzhou 510632, China; ³Pazhou Lab, Guangzhou 510335, China; ⁴College of Traditional Chinese Medicine, Jinan University, Guangzhou 510632, China; ⁵College of Pharmacy, Jinan University, Guangzhou 510632, China; ⁶Cancer Research Institution, Jinan University, Guangzhou 510632, China)

Abstract: Objective: To construct an intelligent assisted diagnosis model for degenerative knee osteoarthritis using a multi-label deep forest algorithm and explore the advantages of the multi-label method in the TCM data set. Methods: Based on 1421 clinical samples, a classification model was constructed using a multi-label deep forest algorithm and compared with five other multi-marker algorithms on six evaluation indexes. Multi-label algorithms Rank-SVM and ML-kNN and single-label algorithms SVM and kNN were used for modeling and comparison. Results: The Hamming loss(0.0458)、Oneerror(0.3521)、Coverage(4.077)、Ranking loss (0.1582)、Average precision (0.7394) and Accuracy (0.9542) of the KOA classification model constructed by multi-label deep forest are better than the other five comparison algorithms in six evaluation metrics. The AUC of KOA (0.8123) is much higher than that of the comparison algorithm. Conclusion: On the TCM data set with multiple syndromes and clips, the performance of the multi-label classification algorithm is better than that of its corresponding single-label algorithm. The diagnosis results of the MLDF algorithm in KOA are in good agreement with the confirmed diagnosis, and it has good promotion and application prospects.

通讯作者: 张佳, 广东省广州市天河区黄埔大道西 601 号暨南大学, 邮编: 510632, 电话: 020-85220010

E-mail: jiazhang@jnu.edu.cn

Keywords: degenerative knee osteoarthritis; Artificial intelligence; Machine learning; Multi-label learning

Funding: National Natural Science Foundation of China (No. 62106084 & No. 61773179); National Natural Science Foundation of Guangdong, China (No.2022A1515010468 & No. 2019A1515012175); the Science and Technology Project in Guangzhou (No.202201010498); Guangdong Provincial Key Laboratory of Traditional Chinese Medicine Informatization (No.2021B1212040007)

膝骨关节炎(knee osteoarthritis, KOA)是退行性骨关节病中常见的类型,是老年人中常见疾病,其中又以女性、肥胖者居多,其临床的主要表现为关节疼痛、肿胀、活动受限等,严重时会使患者致残^[1]。随着我国人口老龄化进程逐渐加快,膝骨关节炎严重威胁着人们的身体健康,给社会经济带来巨大负担。随着信息技术高速发展,人工智能广泛应用于图像识别、辅助诊断等多个医疗领域,为许多疾病诊治提供新的解决方法^[2]。近10年来,机器学习算法开始应用于KOA预测领域^[3],但大部分都侧重于影像学或者生物指标,而二者往往因为价格昂贵,采集困难导致临床转化应用受限。中医认为“有诸内者,必形诸外”,通过中医四诊捕获的临床信息对于预测疾病的发生、发展和预后具有重要价值,且中医指标易于采集,成本低。因此本文提出结合患者中医四诊信息,利用辨证论治理论,使用多标记机器学习算法构建KOA辅助诊断模型。

退行性膝骨关节炎诊断与多标记分类算法

退行性膝骨关节炎的临床诊断通常基于患者的既往病史、体征和症状以及影像学检查。磁共振成像(MRI)、关节镜检查、高频彩超和热构图是接受程度最高的四种影像学检查方式^[4]。然而,这些基于图像的KOA诊断方式对老年人来说既麻烦又昂贵。既往研究表明,中医四诊信息有利于各类疾病的诊断和预后,如慢性乙型肝炎综合征^[5]、冠心病等^[6]。中医治疗膝骨关节炎具有疗效可靠、价格合理,不良反应少等优点,是中医诊治的优势病种^[7],对老年人来说更方便、经济和安全。

多标记分类方法是机器学习中一个重要研究方向,被广泛用于多语义分割问题。例如在文本分类中,每个文本都可能属于多个主题。在对这样的文本进行分类时,每一个主题可以看作是一个标签,如果用一个标签代表一篇文章的主题,可能过于笼统,多标记学习框架则专门用于解决这类问题。在中医临床实际情况中,简单证型较少出现,一个病例往往有多个症状,可能出现多种相关证候,多个证型兼夹的情况较为常见。从信息学角度来看,多证型兼夹的中医诊断过程属于典型的多标记学习问题^[8]。

根据算法对标记相互关系利用的策略,多标记学习又分为低阶策略和高阶策略。例如一个患者可能同时存在气虚、气滞血瘀和心脾不交等多个中医证型,低阶策略下的多标记学习往往把这种多证兼夹的情况分为单个证型或者两两组合进行处理,这样的方式忽略了其他证型对目标

证型的影响。而高阶策略的多标记算法则可以避免忽略标签相关性这个问题,更符合中医临床出现的多症状兼夹情况。

基于此,我们提出了一种基于多标记深度森林(Multi-Label Learning with Deep Forest)^[9]的膝骨关节炎智能辅助诊断方法。MLDF是一种基于多标记树的非神经网络风格的深层模型,底层使用高阶策略实现,其标签相关性则是通过逐层训练,重用上层表现好的特征实现。且相比于传统的神经网络分类算法,MLDF在训练过程中以依赖数据结果的方式控制模型增长,这使得我们的算法适用于不同规模的训练数据,而不仅限于大规模数据。因此我们认为多标记深度森林特有的结构以及在训练过程中对数据的处理方式可以有效挖掘中医数据中的复杂的证型兼夹关系,从而得到更优的分类结果。

实验部分

1 数据集介绍

1.1 一般资料 本研究的患者均为2011-2021年暨南大学附属第一医院、佛山市中医院、五邑中医院骨关节科、湖南中医药大学附属第一医院的门诊和住院患者。共纳入5013例样本,其中健康样本601例,膝骨关节炎患病样本4412例。数据集包含556个变量,其中膝骨关节炎、平和质、气虚质、阳虚质、阴虚质、痰湿质、湿热质、血瘀质、气郁质以及特禀质等10个变量作为标记。

1.2 诊断标准 KOA诊断标准参照《膝骨关节炎中医诊疗指南(2020年版)》^[10] Kellgren—Lawrence 影像分级。

1.3 纳入标准 ①男女不限 ②签署知情同意书。

1.4 数据预处理 为了降低数据的稀疏性,控制正负样本比例,加快模型收敛速度,对数据集进行降采样、空缺值填补、特征分解、对频数为零的特征进行剔除以及归一化处理。最后得到1421例样本作为训练集,其中KOA患者样本1111例,男性患者309例,女性患者802例,平均年龄为66.40岁,平均体重59.51千克。

2. 方法

由多层次级联森林组成的多标记深度森林结构如图1所示,其每一层级联森林由基于预测聚类树的随机森林(蓝色)和完全随机森林(黑色)两类森林构成,每一个森林模块都有100棵预测聚类树^[11]。MLDF的深度由其内部度量机制决定,在模型训练前设置最大层数max,在每一层训练结束时模型内部评价当前层效果是否提升,如果连续

三层没有提升则模型停止增长,最后选取效果最好的层作为 MLDF 的输出层,也就是说模型最终输出层数不会大于初始设定的 max。

MLDF 训练时,患者数据经过第 t 层级联森林得到特征表示 H_t ,测量感知特征重用部分根据其内部的性能评价机制使用上一层的表征 G^{t-1} 来更新 H_t ,得到新的表征 G^t ,再与原始的患者数据拼接在一起作为下一层的输入,直到模型收敛停止增长。

本文实验由两部分组成:①使用 MLDF 构建辅助诊断模型,与 Rank-SVM、ML-kNN、BP-MLL、ML-ARAM、Binary Relevance(BR)等 5 种低阶多标记学习算法^[12]做对比,在 6 个评价指标上评估我们提出的方法的有效性。②

探究单标记算法与多标记算法在中医多证兼夹的数据集上的表现,选取单标记算法 kNN, SVM 与其对应的多标记算法 ML-kNN 和 Rank-SVM 进行比较。

为了保证实验结果的无偏性,本研究训练时采用五折交叉验证。具体地,把数据集随机分为 5 等份,每次选取一份作为测试集,其余部分作为训练集,重复该过程 5 次,最终得到 5 次结果的均值,需要说明的是,测试集之间是互斥的,所以能有效覆盖整个数据集。

3 评价指标

本文使用汉明损失、1-错误率、排序损失、覆盖率、平均精度等 5 个多标记学习中常用评价指标以及准确率用以评估方法效果^[12]。

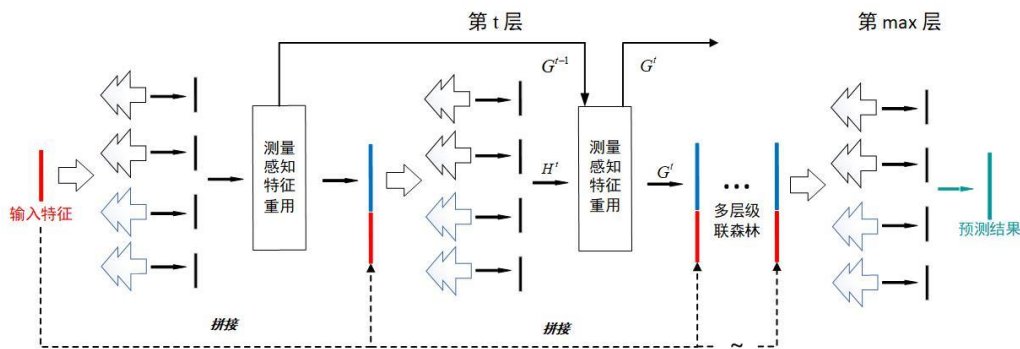


图 1 多标记深度森林结构图

4 实验结果

4.1 结果 1 从图 2 可以看出在单标记算法和其对应的多标记算法在 10 个标记（即膝骨关节炎、平和质、气虚质、阳虚质、阴虚质、痰湿质、湿热质、血瘀质、气郁质以及特禀质）上的准确率, Rank-SVM 和 ML-kNN 在退行性膝骨关节炎上的分类准确率分别为 0.7746 和 0.7787, 均高于其对应的单标记算法 SVM(0.7641)和 kNN(0.7570), 并且在其他 9 种中医体质的分类预测中多标记算法普遍优于其对应的单标记算法。

表 1 展示了多标记算法与其对应的单标记算法在 6 个评价指标上的结果,加粗的为对比算法中表现最佳的结果,可以看到在所有评价指标中, Rank-SVM 均优于 SVM, ML-kNN 仅在 Coverage 和 One-Error 上要略低于 kNN。可以得出,在 KOA 的中医数据中,基于单标记算法改进的

多标记算法总体性能更优。

4.2 结果 2 表 2 中显示的是 MLDF 与 BP-MLL、ML-ARAM、Binary Relevance、ML-kNN 和 Rank-SVM 等其他 5 种多标记算法对比实验结果,从表 2 可以看到,使用高阶策略的 MLDF 算法在六个评价指标上均以较大优势优于其他算法,可以得出,MLDF 同基于低阶策略设计的多标记算法相比,MLDF 能更有效地挖掘出 KOA 与其他标记间的数据关系,从而提高模型分类效果。

4.3 结果 3 表 3 给出了 MLDF 与 BP-MLL、ML-ARAM、Binary Relevance、ML-kNN、Rank-SVM 等 5 种算法在 KOA 和 9 种中医体质下的 AUC 对比。从表中可以看出 MLDF 在各个标记下的 AUC 结果均高于 0.8,在阳虚质的分类上达到最大值 0.9991,10 个标记平均 AUC 为 0.9723,其表现全面优于其他 5 种对比算法。

表 1 多标记算法与单标记算法对比

评价指标	kNN	ML-kNN	SVM	Rank-SVM
准确率	0.9176	0.9279	0.9299	0.9394
汉明损失	0.0736	0.0720	0.0730	0.0605
1-错误率	0.4084	0.4295	0.4514	0.4084
覆盖率	4.2971	4.3330	5.0450	4.4330
排序损失	0.2087	0.2063	0.2720	0.2020
平均精度	0.6113	0.6167	0.5312	0.7143

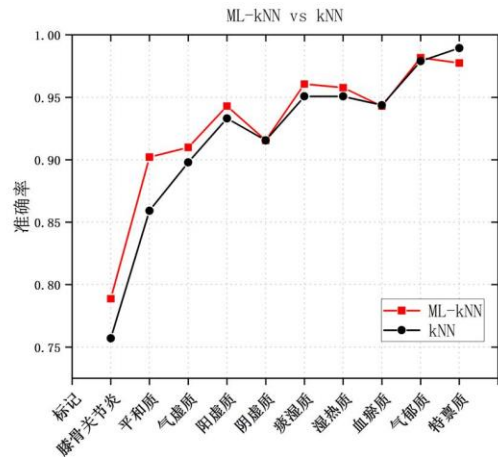
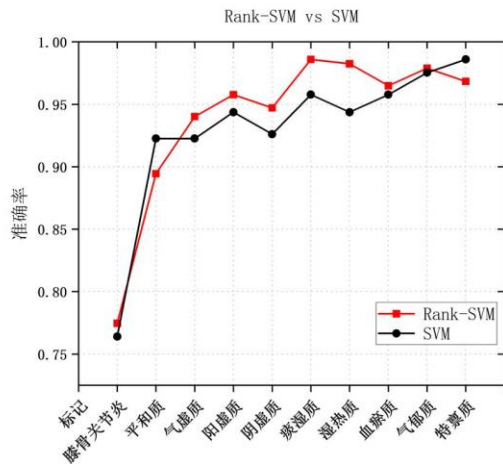


图 2 多标记算法与单标记算法在 10 个标记上的对比

表 2 MLDF 与 5 种多标记分类算法对比

评价指标	MLDF	BP-MLL	ML-ARAM	BR	ML-kNN	Rank-SVM
准确率	0.9514	0.5354	0.9191	0.9263	0.9279	0.9394
汉明损失	0.0485	0.4645	0.0808	0.0736	0.0720	0.0605
1-错误率	0.3591	0.8732	0.4429	0.4084	0.4295	0.4084
覆盖率	4.0985	7.6739	4.5021	4.2971	4.3330	4.4330
排序损失	0.1641	0.6455	0.2248	0.2087	0.2063	0.2020
平均精度	0.7312	0.2937	0.5883	0.6113	0.6167	0.7143

表 3 6 种多标记算法在 10 个标记上的 AUC 结果

标记	MLDF	BP-MLL	ML-ARAM	BR	ML-kNN	Rank-SVM
膝骨关节炎	0.8122	0.5526	0.7241	0.7556	0.6889	0.7746
平和质	0.9685	0.4857	0.9267	0.9473	0.9485	0.8943
气虚质	0.9859	0.5288	0.9220	0.9435	0.9485	0.9401
阳虚质	0.9991	0.6268	0.9597	0.9677	0.9642	0.9577
阴虚质	0.9957	0.5104	0.9172	0.9513	0.9523	0.9471
痰湿质	0.9981	0.5552	0.8371	0.8981	0.9757	0.9859
湿热质	0.9978	0.5239	0.9096	0.9263	0.9762	0.9823
血瘀质	0.9926	0.5378	0.8596	0.9028	0.9652	0.9647
气郁质	0.9934	0.4338	0.7587	0.8518	0.9561	0.9788
特禀质	0.9800	0.6499	0.7272	0.7410	0.7011	0.9683

讨论

中医药在防治重大慢性疑难疾病中具有独特突出优势。人工智能和中医药的交叉融合为中医药数字化防控重大疑难疾病和慢性疾病提供了突破口。多标记学习是人工智能的一个重要研究方向,符合中医辨证论治思维,有利于解决多证兼夹问题。在 kNN 和 SVM 两个单标记算法中,我们将 KOA 和 9 种中医体质分解为多个单标记问题进行预测,然后将结果进行组合。在这个过程中丢失了标记间的潜在关系。然而,在 ML-kNN 和 Rank-SVM 两种多标记算法中,保留了标记间的潜在关系,因此,ML-kNN

和 Rank-SVM 的预测精度理论会优于 kNN 和 SVM,我们的实验也证实了这一点。同时,通过单标记算法与其对应的多标记算法的对比,揭示了使用多标记算法考虑 KOA 与体质间相关性以及症状兼夹性,可以提高对 KOA 的诊断精度。

不同多标记算法其底层策略有所不同,MLDF 作为较新提出的一种高阶多标记算法,充分利用有限数据,挖掘 KOA 与其他证候间的内在关系,从而提高整体预测效果。在本研究中,我们将 MLDF 用于构建 KOA 辅助诊断模型,同时与其他五种低阶多标记算法作对比,从测评结果可以

看出本文采用的 MLDF 算法在 KOA 的诊断结果与真实的诊断的一致性较好,且冗余诊断较少,能够很好地解决兼夹证型诊断问题,为构建 KOA 智能辅助诊断模型提供了一个有效的思路与方法。

最后我们在此基础上使用机器学习构建了 KOA 辅助诊断小程序,可以预测患有 KOA 的可能性。该实现可在线获取得到: <https://ihealthapp.jnu.edu.cn/ms/#/tool-ggjb>

参 考 文 献

- [1] 徐颖鹏,谢利民,王文岳.中药外治与西药治疗膝骨关节炎疗效及安全性 Meta 分析[J].中国中药杂志,2012,37(19):2977-2984.
- [2] 李灿东,辛基梁,雷黄伟等.中医健康管理与人工智能[J].中华中医药杂志,2019,34(08):3586-3588.
- [3] Jamshidi A, Pelletier J P, Martel-Pelletier J. Machine-learning-based patient-specific prediction models for knee osteoarthritis[J]. Nature Reviews Rheumatology, 2019, 15(1): 49-60.
- [4] Xu H, Zhao G, Xia F, et al. The diagnosis and treatment of knee osteoarthritis: A literature review[J]. Int J Clin Exp Med, 2019, 12(5): 4589-4599.
- [5] 於仁俊,杜斌,尤松鑫辨治慢性乙型肝炎经验[J].中华中医药杂志,2022,37(02):874-878.
- [6] 沈绍功,姚金城.中医心病之心绞痛诊断与疗效标准[J].中国医药学报,2004(10):581-583.
- [7] 辛基梁,张佳,李绍滋,等.中医健康状态辨识中的多标记分类方法研究[J].中华中医药杂志,2019,34(09):3952-3955.
- [8] Xia S, Zhang J, Du G, et al. A microcosmic syndrome differentiation model for metabolic syndrome with multilabel learning[J]. Evidence-Based Complementary and Alternative Medicine, 2020, 2020.
- [9] Yang L, Wu X Z, Jiang Y, et al. Multi-label learning with deep forest [C]. In Proceedings of the 24th European Conference on Artificial Intelligence, 2020, pp. 1634-1641.
- [10] 陈卫衡.膝骨关节炎中医诊疗指南(2020 年版)[J].中医正骨,2020,32(10):1-14.
- [11] Kocev D, Vens C, Struyf J, et al. Tree ensembles for predicting structured outputs[J]. Pattern Recognition, 2013, 46(3): 817-833.
- [12] Gibaja E, Ventura S. A tutorial on multilabel learning[J]. ACM Computing Surveys (CSUR), 2015, 47(3): 1-38.