

基于多标签学习的创伤救治层链决策研究

赵鹏飞^{1,2}, 刘 华^{2,3}

(1. 昆明理工大学计算机重点实验室, 昆明 650500; 2. 昆明理工大学信息工程与自动化学院, 昆明 650500; 3. 昆明理工大学人事处, 昆明 650501)

摘要: 在现代创伤救治中, 根据患者伤情进行合理而准确的院前评估并制定相应的救治决策对降低患者伤残率与死亡率具有重要意义。为了改善人工制定决策的缺陷, 实现准确合理的标准化创伤救治决策制定, 本文利用多标签学习思想, 在对创伤救治决策进行深入分析与研究的基础上, 将整体救治决策进行子决策划分, 并提取出子决策对应的判定因素作为标签集。为了更好地考虑标签间的关联, 将Classifier Chains算法的链式思想与多标签K近邻(Multi-label K-nearest neighbor, ML-KNN)算法融合, 提出一种层链多标签学习算法, 称为层链多标签K近邻算法(Layer chain ML-KNN, LCML-KNN)。LCML-KNN算法将标签依特点划分为两个层链, 在第一层链的预测标签信息输出后对其进行独热编码, 转化后的标签看作新特征放入第二层链进行预测与判断。LCML-KNN算法不仅更好地考虑了标签间的关联性, 而且通过标签转化扩充了特征维数。在两个创伤类数据集上与现有各类多标签算法进行实验对比, 结果验证了LCML-KNN算法的鲁棒性和优越性。

关键词: 多标签学习; 创伤救治; 决策制定; 层链划分; 多标签K近邻

中图分类号: TN974 **文献标志码:** A

Layer Chains Decision of Trauma Treatment Based on Multi-label Learning

ZHAO Pengfei^{1,2}, LIU Hua^{2,3}

(1. Computer Key Laboratory, Kunming University of Science and Technology, Kunming 650500, China; 2. Faculty of Information Engineering and Automation, Kunming University of Science and Technology, Kunming 650500, China; 3. Personnel Department, Kunming University of Science and Technology, Kunming 650501, China)

Abstract: In modern trauma treatment, reasonable and accurate pre-hospital assessment based on the injury and making corresponding treatment decisions are of great significance for reducing the disability and mortality of patients. To improve the shortcomings of manual decision-making and achieve accurate and reasonable standardized trauma treatment decision-making, after in-depth analysis and research on the treatment decision, this study uses the multi-label learning method to divide the overall treatment decision into sub-decisions, and extracts judgment factors corresponding to the sub-decisions as a label sets. Next, to better consider the relationship between labels, this paper combines the chain idea of the Classifier Chains algorithm with the ML-KNN algorithm, and proposes a multi-label learning algorithm by improving the ML-KNN algorithm, named layer chains multi-label K-nearest neighbor (LCML-KNN). The LCML-KNN algorithm divides labels into two layer chains according to the characteristics. After the prediction label information of the first layer chain is output, it is uniquely encoded. And the transformed labels are

基金项目: 国家自然科学基金(61741206)。

收稿日期: 2021-06-16; **修订日期:** 2021-10-20

put into the second layer chain as new features for prediction and judgement. The LCML-KNN algorithm not only better takes into account the relationship between the labels but also expands the feature dimension through the label conversion. The experimental results with various existing multi-label learning algorithms on two trauma datasets verify the robustness and superiority of the LCML-KNN algorithm.

Key words: multi-label learning; trauma treatment; decision making; layer chains division; multi-label K-nearest neighbor (ML-KNN)

引 言

随着现代社会的不断进步与发展,各类创伤特别是高能创伤的发生率逐年递增^[1],且现代创伤中多发伤多、重伤多、严重并发症出现几率高,给患者的救治提出了新的挑战。为了更好地实现院前院内救治一体化^[2-3],不仅要在院内制定科学合理的救治方案,在院前对患者制定合理准确的救治决策同样具有重要意义。创伤救治决策与具体的伤情救治方案不同,它是在院前根据患者的创伤严重程度与即时症状表现等因素制定的综合救治建议,旨在缩短患者入院后的急诊停留时间,使患者尽快得到相应级别的院内医疗救治,有效降低患者的伤残率与死亡率。然而据相关调查显示^[4],由于不同地域和不同级别的院前急救人员与医生的专业领域知识、过往经验积累以及个体主观认知存在差异,给伤情评估与标准化救治决策的制定带来了一定阻碍。

多标签学习是一类经典的监督学习方法。近年来,随着相关研究的不断深入推进,多标签学习不再局限于文本分类等传统任务的驱动,在诸多领域的复杂数据分析与疑难问题解决上也取得了优良的实际应用效果。例如,Folorunso等^[5]通过建立多标签分类模型并比较不同评估方法对尼日利亚精神疾病数据进行了深度剖析与探究;Wang等^[6]基于多标签学习思想提出一种层次认知结构学习模型(Hierarchical cognitive structure learning model, HCSM),通过在4个基准数据集上的实验证实了该模型性能优于最新的方法;Hou等^[7]提出了一种有效的多标签学习方法用于车辆的重新识别(re-ID),并通过对比实验证实了该方法优于现有的多种车辆重新识别方法。除此之外,多标签学习思想也逐渐应用于医学影像识别、生物基因分析和工业故障诊断等领域的研究^[8-10]。

对于创伤救治决策的制定,传统的人工制定方法存在不同急救人员与医生水平参差不齐、个人经验积累与主观认知差异对救治决策制定的客观合理性与准确性存在影响等缺陷。为了克服上述缺陷并对患者制定尽可能合理准确的标准化救治决策,本文基于多标签学习思想,将创伤救治整体决策划分为若干子决策,并将子决策对应的判定因素转化为多标签学习的标签集,从而将创伤救治决策制定的问题转化为一个多标签学习任务,通过客观的人工智能方法进行标准化创伤救治决策的预测与制定。

对于多标签算法的选择,ML-KNN是由KNN算法演变而来的一种优秀的惰性学习算法。与其他各类经典多标签学习算法相比,ML-KNN不仅无数据输入假设、准确度高,而且对异常点不敏感,是一种被广泛接受和采用的多标签学习算法。因此,本文在对各类多标签学习算法进行深入研究的基础上,将Classifier Chains算法^[11]的链式思想与多标签K近邻(Multi-label K-nearest neighbor, ML-KNN)算法^[12]融合,通过对ML-KNN算法进行改进,提出一种多标签学习算法,称为层链多标签K近邻算法(Layer chain ML-KNN, LCML-KNN)。LCML-KNN属于高阶算法,它根据专家小组的专业辅助与指导将标签划分为两个层链,第一层链模型预测完毕后将预测标签信息经独热编码后转化为新特征扩充数据集的样本维度,扩充后的特征信息用于第二层链模型的学习与预测。通过与若干经典的多标签学习算法以及集成学习方法的实验对比,证明了LCML-KNN算法的综合性能表现最佳,体现了该算法的有效性与优越性。

本研究的创新性有:(1)将创伤救治决策制定转化为多标签学习任务;对创伤救治整体决策进行子决策

的划分,提取出子决策对应的判定因素并转化为多标签学习的标签集,从而将创伤救治决策制定的问题转化为一个多标签学习任务,旨在通过客观的人工智能方法对不同创伤患者尽可能给出合理准确的标准化救治决策预测与建议。(2) 提出 LCML-KNN 算法:为了更好地考虑标签间的关联、挖掘不同标签间的关系,LCML-KNN 高阶算法基于链式思想将标签集中的标签根据专家小组的专业辅助与指导划分为两个层链,将第一层链的预测标签信息经独热编码后转化后为新特征对数据集进行样本维度的扩充,随后用扩充后的数据集继续进行第二层链模型的学习与预测,相较于现有多标签算法显示出了更为优越的整体性能。

1 材料与方 法

首先通过救治子决策的划分将子决策对应的判定因素转化为多标签学习的标签集,随后对不同标签的层链划分进行了细致说明;然后对多标签学习基本概念与常见算法进行了简要介绍,最后着重介绍了所提出的 LCML-KNN 算法的原理与特点。

1.1 标签集的建立

创伤救治决策是在院前根据患者的创伤严重程度与即时症状表现等因素制定的综合救治建议。为了构建多标签学习的标签集,对整体救治决策进行救治子决策的划分。对于子决策划分,首先要对患者是否具有救治意义做出准确判断,即患者的伤情是否具有生还可能,对于判定为无生还可能的患者可终止后续的决策制定与建议。然后对于具有救治意义的患者应对其伤情严重程度做出客观准确的评估并制定相应决策:无大碍的轻伤患者可就近送往社区医院;评估为严重创伤以及多发伤的患者,应尽快联系医院急救中心开辟绿色通道,尽可能在“黄金时间窗口”内使其入院救治,降低患者死亡率^[13-14]。除此之外,由于创伤是“时间敏感性”疾病^[15],所以整体救治决策中还包含以下两个重要子决策:(1) 对患者是否存在并发症或严重症状进行判断,根据具体情况制定相应决策并开展必要的院前急救,典型的并发症与症状包括气道堵塞、大出血、气胸和低体温等^[16-17];(2) 对创伤患者特别是严重创伤患者在入院途中和入院后潜在的伤情恶劣演化与并发症发生的风险进行合理预判,并提前制定好相应的应对决策与措施。

基于上述4个子决策的分析,将4个子决策对应的判定因素提取出来,结果如表1所示。

最后,将上述4个判定因素转化为4个标签,表示为 L_1 、 L_2 、 L_3 和 L_4 ,分别对应判定因素1至4,并由此建立了包含4个标签的标签集。标签集的建立流程如图1所示。

表 1 子决策对应的判定因素

Table 1 Judgment factors corresponding to sub-decisions

子决策	判定因素
1	是否具有救治意义(潜在可存活的伤情)
2	创伤严重程度
3	是否存在并发症或严重症状
4	有无伤情恶劣演化与并发症发生的风险

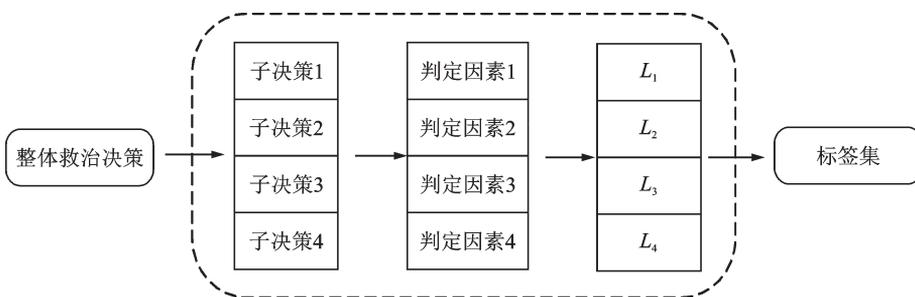


图 1 标签集的建立流程

Fig.1 Flow of establishment of label sets

1.2 标签的层链划分

为了更好地考虑与运用标签间关联,将标签集中的4个标签进行了科学合理的层链划分。具体方法为首先在专家小组的专业辅助与指导下考虑了不同标签(判定因素)的医学意义及其潜在的相互关联与影响。为了使第一层链的标签信息在转化为新特征并扩充特征维度后能够对第二层链标签的准确预测起到促进作用即实现信息层级递进,将与患者创伤严重性相关的判定因素1(L_1)和判定因素2(L_2)放入第一层链,将判定因素3(L_3)与判定因素4(L_4)放入第二层链。这样划分的原因其一是对于患者来说,创伤严重程度越重,存在并发症或严重症状的概率越大,即使当前没有并发症出现的重伤患者,其在送院救治过程中出现伤情恶劣演化的风险概率较大;其二是这样的划分也符合创伤救治的实际逻辑,即患者是否具有救治意义及创伤严重程度是需要首先考虑和判断的,在此判断的基础上再进一步对患者是否存在并发症或严重症状以及有无伤情恶劣演化风险进行判断。综上所述,在层链划分环节将标签 L_1 与 L_2 放入第一层链,将标签 L_3 与 L_4 放入第二层链,旨在将原本并行的分类器模型进行分层处理,目的是充分挖掘标签相关性。

另外需要说明的是,除了不同层链之间存在先后顺序之外,处于同一层链中的不同标签并没有先后顺序,分类器对同一层链中的标签是同时进行分类预测的,即两个层链的分类器分别对各自层链中的标签进行了多标签分类。

1.3 多标签学习

在一般的有监督学习中,样本包含若干特征向量并对应一个唯一的标签,称为单标签学习。而在实际的许多复杂任务与问题中,一个样本可能同时与多个标签相关联,通过对标签集中多个标签的同时学习来对新样本进行分类与预测,这就是多标签学习。

基于目前的研究,多标签学习算法按照不同的思想可以分为3类:问题转换、算法自适应和集成学习方法。除此之外,多标签学习还可以划分成如图2所示的三大策略:其中一阶算法包括 Binary Relevance 算法、ML-DT 算法^[18]和 ML-KNN 算法;二阶策略的代表算法包括 Rank-SVM 和 Calibrated Label Ranking^[19];高阶算法考虑了多个标签间的内在联系,包含 Classifier Chains、Label Powerset^[20]、ECC 和 RAkEL^[21]等算法。

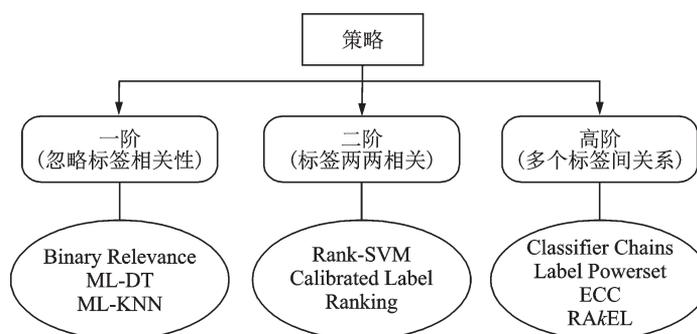


图2 多标签学习的不同策略

Fig.2 Different strategies for multi-label learning

1.4 Classifier Chains 与 ML-KNN

分类器链(Classifier Chains)算法是高阶算法,同时也是一类问题转换方法,其基本思想是将多标签学习转化为若干二分类问题。假设标签集中的标签总数为 q ,分类器链生成 q 个二进制的分类器,链中每个分类器的特征空间用先前分类器的预测标签进行扩充,即链中的后续二进制分类器是建立在前置

预测基础之上的。图3对 Classifier Chains 的原理进行了举例说明,其中 X 代表输入空间, Y 代表标签。作为高阶算法的 Classifier Chains 以随机链序的方式考虑了标签之间的潜在关联性,虽然算法简单,但在一定程度上挖掘了不同标签之间的关系。

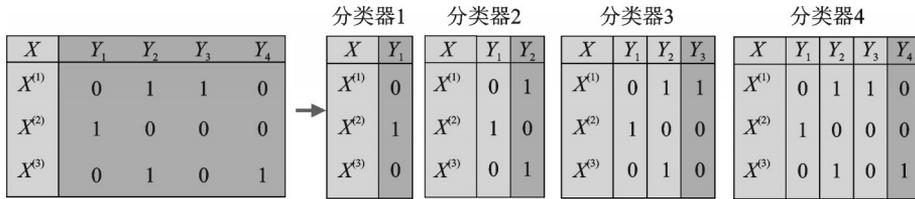


图3 Classifier Chains 算法的原理

Fig.3 Principle of Classifier Chains algorithm

ML-KNN 算法是一类惰性学习算法也是一阶策略算法,针对多标签学习问题由 KNN 算法改进而来。ML-KNN 算法的思想是通过计算欧几里得距离来寻找 K 个最近邻样本,然后利用最大后验概率 (Maximum a posteriori, MAP) 准则和贝叶斯准则进行结果的判断与预测。ML-KNN 算法的描述如下:

假设已知样本序列 X 和与其对应的标签集 y_X , $N(X)$ 代表 X 的 K 个近邻样本的集合。 H_l^1 定义为 X 选中 l 的情形, H_0^l 为未选中标签 l 的情形,事件 $E_i^l (i \in \{0, 1, 2, \dots, K\})$ 代表 X 的 K 个最近邻中包含标签 l 的数量为 i 。然后根据 MAP 准则得到分类函数

$$y_X(l) = \arg \max_{b \in \{0, 1\}} P(H_b^l) P(E_i^l | H_b^l) \tag{1}$$

从而对样本 X 是否含有标签 l 做出判定,并对训练集样本的标签属性进行预测。

ML-KNN 算法考虑了先验概率,其优点是准确率相对优良且对异常点不敏感,适用于处理类不平衡的数据集。但该算法作为一阶策略的算法,其不足之处是完全忽略了标签信息的关联性。

1.5 LCML-KNN

LCML-KNN 算法是一种高阶算法,由 ML-KNN 算法改进而来,其原理示意图如图 4 所示。LCML-KNN 首先将不同标签根据专家指导划分为两个层链,第一层链的标签包括 L_1 和 L_2 ,第二层链的标签包括 L_3 和 L_4 ,且每个层链的分类器模型都采用 ML-KNN 算法;随后,第一层链分类器对 L_1 和 L_2 预测并得到两个预测标签,将预测标签组合经独热编码后转化为新样本特征扩充数据集样本维度;然后用扩充后的数据集继续进行第二层链分类器的学习,并得到 L_3 与 L_4 的预测标签;最后将两个层链的 4 个预测标签进行合并就得到了最终的多标签预测结果。

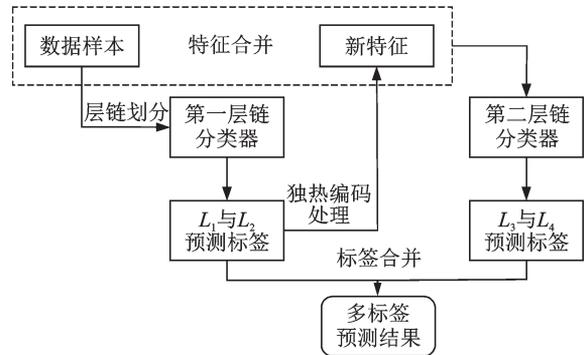


图4 LCML-KNN 算法的原理

Fig.4 Principle of the LCML-KNN algorithm

LCML-KNN 算法的特点如下:

(1) 标签层链划分:将不同标签根据专家小组的辅助与指导划入两个层链,在第一层链分类器模型预测的标签组合输出后经独热编码转化为新特征,并用扩充的数据集进行第二层链分类器模型的学习与预测。层链划分处理对预测准确率的提升起到了关键作用。

(2) 新特征独热编码处理:通过对第一层链的预测标签组合进行独热编码转化的操作,有效防止了过拟合问题的出现。

(3) 更好地考虑标签关联性:LCML-KNN算法作为一种高阶算法,与ML-KNN等一阶算法相比,通过划分层链更好地考虑了标签间的关联性,进一步挖掘了全局的标签信息,加深了分类器模型对数据的理解。

2 数据分析

2.1 数据基本情况

为了能够使分类器得到更多数据样本的充分学习与训练,将某军医院校分多次提供的小规模外科创伤数据集整合为了一个包含3 514条数据的大数据集用于研究。该数据集包含了各类创伤患者的伤情数据,从数据种类的丰富性与多样性来看具有良好的创伤种类、伤情情况的数据覆盖性。将数据样本按照8:2的比例进行了训练集与测试集的划分,经特征提取后样本共包含特征24个,特征中包括了对救治决策制定起重要参考作用的受伤部位、组织伤型、伤类、并发症和包括呼吸频率与收缩压之内的简单生理体征等,且每个特征都转化为了数字属性并经过了标准的离散化处理。多标签学习的标签集共包含4个标签,分别对应4个判定因素 L_1 、 L_2 、 L_3 和 L_4 。

2.2 数据相关分析

为了对本文所使用的数据进行更为深入的挖掘与了解,接下来从多角度对数据的相关统计展开计算与分析。在分析之前预先定义以下符号: D 代表数据集; p 为数据集样本总数; q 为标签总数; Y_i 是与样本实例 x_i 相关的标签序列; d 表示特征的数量。

(1) 标签基数

标签基数(Label Cardinality)是数据集中每个实例的标签平均数量,是一个标准度量,在一定程度上反映了总体样本的标签稀疏情况^[22],其表达式为

$$\text{Label Cardinality} = \frac{1}{p} \sum_{i=1}^p |Y_i| \quad (2)$$

(2) 标签密度

标签密度(Label Density)是基于标签总数 q 的归一化标签基数^[22],其表达式为

$$\text{Label Density} = \frac{1}{pq} \sum_{i=1}^p |Y_i| \quad (3)$$

(3) 标签多样性

标签多样性(Label Diversity)是数据集中存在的不同标签序列的数量^[23],其表达式为

$$\text{Label Diversity} = \left| \left\{ Y \mid \exists x: (x, Y) \in D \right\} \right| \quad (4)$$

(4) 数据复杂度

数据复杂度(Complexity)定义为样本总数 p 、特征数量 d 和标签总数 q 三者的乘积,反映了多标签数据集的整体复杂程度,计算公式为

$$\text{Complexity} = p \times d \times q \quad (5)$$

数据的相关统计分析结果在表2中进行了详细展示,其中 n 表示特征属性为数字属性。

表2 数据的相关统计分析

Table 2 Relevant statistical analysis of data

p	d	q	Label Cardinality	Label Density	Label Diversity	Complexity
3 514	$24n$	4	2.157	0.539	7	3.373E5

3 实验结果及分析

在实验环节,首先确定了评价各类算法性能的多标签学习评价指标。随后,设置了3组对比实验,分别令LCML-KNN算法与经典的多标签算法、经典的集成学习算法RAkEL和ML-KNN算法进行实验对比来探究LCML-KNN算法的整体性能表现。接下来,为了进一步验证算法的鲁棒性和优越性,在同属于创伤类数据集的一个战争创伤数据集上进行了不同算法的对比实验。最后,对两个数据集上的实验结果进行了总结分析与讨论。

3.1 评价指标

为了对各类算法的性能进行科学的综合评价,针对该多标签学习任务的特点,选取了以下5个整体评价指标:汉明损失(Hloss)、平均准确率(Average_acc)、平均精度(Average_pre)、 F_1 -score以及AUC值。评价指标的部分公式如下

$$\text{Hloss}(h) = \frac{1}{p} \sum_{i=1}^p \frac{1}{q} |h(x_i) \Delta Y_i| \quad (6)$$

$$\text{Average_acc}(h) = \frac{1}{p} \sum_{i=1}^p \frac{|Y_i \cap h(x_i)|}{|Y_i \cup h(x_i)|} \quad (7)$$

$$\text{Average_pre}(h) = \frac{1}{p} \sum_{i=1}^p \frac{|Y_i \cap h(x_i)|}{|h(x_i)|} \quad (8)$$

$$F^\beta(h) = \frac{(1 + \beta^2) \cdot \text{precision}(h) \cdot \text{recall}(h)}{\beta^2 \cdot [\text{precision}(h) + \text{recall}(h)]} \quad (9)$$

式中: F^β 为平均精度和平均召回率的加权调和平均值,precision(h)代表平均精度,recall(h)代表平均召回率。为了求取 F_1 -score,参数 β 的值设定为1。

3.2 与经典多标签学习算法对比

在第1组对比实验中,将LCML-KNN与常见的经典多标签学习算法进行对比。具体来说,选取Binary Relevance(BR)、Label Powerset(LP)和Classifier Chains(CC)3种算法作为对照,对比实验结果如表3所示。

表3 LCML-KNN与各类算法的实验结果对比

Table 3 Comparison of experimental results between LCML-KNN and various algorithms

算法	Hloss	Average_acc	Average_pre	F_1 -score	AUC
BR	0.012 9	0.957 7	0.978 3	0.978 9	0.960
LP	0.013 7	0.956 4	0.977 1	0.977 3	0.957
CC	0.010 7	0.964 9	0.980 7	0.980 9	0.968
RAkEL	0.009 2	0.971 1	0.988 4	0.989 6	0.979
ML-KNN	0.008 6	0.972 4	0.986 0	0.991 9	0.977
LCML-KNN	0.007 8	0.980 4	0.992 7	0.988 2	0.983

从表3中可以看出,Classifier Chains的各项评价指标均优于Binary Relevance和Label Powerset。这是因为Classifier Chains通过链式思想考虑了标签之间的潜在关联性,即每预测完一个标签便将其作为新特征对特征空间进行扩充。但是,这3种算法的评价指标表现均不如LCML-KNN,即使是考虑了

标签关联性的 Classifier Chains 算法,其平均准确率与平均精度较 LCML-KNN 也分别低了 1.55% 和 1.2%。

3.3 与 RAKEL 算法对比

在第 2 组对比实验中,将 LCML-KNN 与 RAKEL 集成学习算法进行实验对比。

RAKEL 即随机 k 标签集(Random k -labelsets)算法,是一种性能优良的集成学习方法。它的原理是将标签集随机划分为若干个标签子集,然后选择 n 个随机的 k 标签集并学习 n 个 LP 分类器,每个 LP 分类器为自己对应的 k 标签集中的每个标签提供分类预测,最后集成这 n 个 LP 分类器得到最终预测。对比实验结果如表 3 所示。

从表 3 的各评价指标表现可以看出,虽然 LCML-KNN 在汉明损失、平均准确率、平均精度和 AUC 值的性能表现更佳,但是 F_1 -score 指标的表现与 RAKEL 相比有细微差距。但从整体来看,LCML-KNN 的综合性能表现是更优的,其原因是虽然 RAKEL 是一种优良的集成学习算法,但相较于 LCML-KNN 而言并没有考虑标签信息,没有挖掘标签间的潜在关联性。

3.4 与 ML-KNN 算法对比

由于本文提出的 LCML-KNN 算法是基于分类器链的链式思想对 ML-KNN 改进而来的,因此在第 3 组对比实验中,将 LCML-KNN 与 ML-KNN 进行综合性能的比较。实验结果如表 3 所示。

从表 3 的各项评价指标结果可知,属于高阶算法的 LCML-KNN 在 5 个评价指标中有 4 个要比 ML-KNN 好,特别是重要的平均准确率的结果要高于 ML-KNN 0.8%。而 ML-KNN 算法仅在 F_1 -score 这一项指标上的表现要略高于 LCML-KNN 0.37%。因此,所提出的 LCML-KNN 算法在综合性能表现上要优于传统的 ML-KNN 算法,体现了 LCML-KNN 的优越性。

LCML-KNN 的整体性能之所以要比 ML-KNN 好是因为 LCML-KNN 充分考虑与挖掘了标签间的关联性,并根据标签集中不同标签的特点将其进行了层链划分处理,而 ML-KNN 则是完全忽略标签间的关联。

3.5 战争创伤数据集上的算法对比情况

为了进一步验证 LCML-KNN 算法的鲁棒性和优越性,本文在同属于创伤类数据集的另一个战争创伤数据集上同样进行了不同算法的对比实验。

该战争创伤数据集共包含可用数据 1 216 条,样本特征维数为 26,按照同样 8:2 的比例对数据样本进行训练集与测试集的划分,每个特征都转化为数字属性并经过标准的离散化处理。对于标签集,因为战争创伤救治考虑的判定因素与外科创伤救治的判定因素不尽相同,因此标签集包含同样的 4 个标签,分别对应 4 个重要判定因素 L_1 、 L_2 、 L_3 和 L_4 。对比实验结果如表 4 所示。

表 4 LCML-KNN 与各类算法在战争创伤数据集上的实验结果对比

Table 4 Comparison of experimental results between LCML-KNN and various algorithms on the war trauma dataset

算法	Hloss	Average_acc	Average_pre	F_1 -score	AUC
BR	0.016 9	0.946 6	0.959 1	0.960 1	0.947
LP	0.017 6	0.943 7	0.956 2	0.957 1	0.943
CC	0.014 2	0.953 8	0.970 8	0.969 2	0.956
RAKEL	0.011 8	0.960 2	0.973 4	0.973 8	0.967
ML-KNN	0.011 6	0.961 3	0.972 9	0.974 5	0.963
LCML-KNN	0.010 1	0.968 2	0.979 0	0.976 4	0.971

从表4的实验结果中可以看出,所提出的LCML-KNN算法在各项评价指标的性能表现上均优于其他传统的多标签学习算法,从而进一步有力验证了LCML-KNN算法的鲁棒性和优越性。

3.6 实验分析与讨论

通过上述两个数据集的不同算法的对比实验,全面而客观地验证了本文所提出的LCML-KNN算法的鲁棒性与优越性。LCML-KNN通过链式思想将子决策对应的判定因素进行了层链划分,充分挖掘了不同标签间的关联性,对标签信息有了更为详实的了解,是准确率提升的关键。同时,第一层链预测标签的特征转化不仅扩充了原始数据的特征维度、丰富了可用的数据信息,而且数据标签信息的层级递进预测也使分类器模型能够更为透彻地学习与理解数据特征,提升模型性能。其次,对第一层链的预测信息进行独热编码操作能够有效地避免过拟合风险的发生,保证了第二层链分类器模型的预测质量。

最后需要说明的是,层链划分操作虽然可以使分类器整体性能得到有效提升,但与Classifier Chains算法一样,由于分层处理对算法准确率要求更高,为了避免第一层链分类器模型的预测错误沿链向后传播而影响第二层链分类器模型的性能,在层链划分时需尽可能地对所采用数据集的标签信息以及标签关联性进行深入挖掘与专业分析,以确保层链的科学划分。对于本研究,在层链划分时通过专家小组介入的方式确保了层链划分的合理性和专业性,进而保证了整体性能特别是准确率的有效提升。

4 结束语

为了改善传统人工制定创伤救治决策的缺陷与不足,本文基于多标签学习思想,在对创伤救治决策进行深入分析与研究的基础上,对救治决策进行了子决策划分,并提取出子决策对应的判定因素转化为多标签学习的标签集。接下来,为了更好地挖掘数据信息与标签间的关联,将Classifier Chains算法的链式思想与ML-KNN算法进行了融合,提出一种层链多标签学习算法LCML-KNN。LCML-KNN根据专家小组指导将不同标签进行了层链划分处理,在第一层链的预测标签信息输出后经独热编码转化为新特征扩充了数据特征维度,并将扩充后的数据集代入第二层链分类器进行进一步的学习与预测。通过在两个数据集上的各类多标签学习算法的实验对比,全面而客观地验证了LCML-KNN算法的鲁棒性和优越性。

在未来研究中,为了进一步提升决策制定的准确性与合理性,一方面可以通过进一步挖掘现有标签的潜在关联来提升LCML-KNN算法性能,另一方面可以对救治决策进行更为细致科学的子决策划分并提取对应的判定因素。

参考文献:

- [1] 郑泽源,谢芬,许硕贵,等.院前创伤救治体系发展与“移动急诊科”方案[J].创伤外科杂志,2021,23(7): 318-321.
ZHENG Zeyuan, XIE Fen, XU Shuogui, et al. Development of pre-hospital trauma care system and plan of “mobile emergency department”[J]. *Journal of Traumatic Surgery*, 2021, 23(7): 318-321.
- [2] 李博,李莉,白冰,等.基于流程重组建立一体化创伤急救流程[J].中华急诊医学杂志,2015,24(1): 105-107.
LI Bo, LI Li, BAI Bing, et al. Establish an integration trauma first aid process based on process reorganization[J]. *Chinese Journal of Emergency Medicine*, 2015, 24(1): 105-107.
- [3] 李喆,吕立文,沈印.德国创伤中心见闻及启示[J].重庆医学,2021,50(5): 876-879.
LI Zhe, LYU Liwen, SHEN Yin. Experience and inspiration from the German Trauma Center[J]. *Chongqing Medicine*, 2021, 50(5): 876-879.
- [4] 张鹏,杜哲,刘中砥.我国医生对严重创伤评估能力的多中心调查[J].中华急诊医学杂志,2021,30(5): 533-536.
ZHANG Peng, DU Zhe, LIU Zhongdi, et al. A multi-center survey of Chinese doctors' ability to assess severe trauma[J]. *Chinese Journal of Emergency Medicine*, 2021, 30(5): 533-536.

- [5] FOLORUNSO S O, FASHOTO S G, OLAOMI J, et al. A multi-label learning model for psychotic diseases in Nigeria[J]. *Informatics in Medicine Unlocked*, 2020, 19: 100326.
- [6] WANG Boyan, HU Xuegang, LI Peipei, et al. Cognitive structure learning model for hierarchical multi-label text classification[J]. *Knowledge-Based Systems*, 2021, 218: 106876.
- [7] HOU Jinhui, ZENG Huanqiang, CAI Lei, et al. Multi-label learning with multi-label smoothing regularization for vehicle re-identification[J]. *Neurocomputing*, 2019, 345: 15-22.
- [8] HE Junjun, LI Cheng, YE Jin, et al. Multi-label ocular disease classification with a dense correlation deep neural network[J]. *Biomedical Signal Processing and Control*, 2021, 63: 102167.
- [9] FENG Shou, FU Ping, ZHENG Wenbin. A hierarchical multi-label classification algorithm for gene function prediction[J]. *Algorithms*, 2017, 10(4): 138.
- [10] YU Chongchong, NING Yaqian, QIN Yong, et al. Multi-label fault diagnosis of rolling bearing based on meta-learning[J]. *Neural Computing and Applications*, 2021, 33(10): 5393-5407.
- [11] READ J, PFAHRINGER B, HOLMES G, et al. Classifier chains for multi-label classification[J]. *Machine Learning*, 2011, 85(3): 333-359.
- [12] ZHANG Mingling, ZHOU Zhihua. ML-KNN: A lazy learning approach to multi-label learning[J]. *Pattern Recognition*, 2007, 40(7): 2038-2048.
- [13] JIANG Baoguo. The challenges faced in the field of trauma care in China[J]. *Chinese Journal of Surgery*, 2015, 53(6): 401-404.
- [14] 程晓斌, 赵先柱, 张连阳, 等. 多发伤院内紧急救治规范探讨[J]. *创伤外科杂志*, 2010, 12(1): 4-7.
CHENG Xiaobin, ZHAO Xianzhu, ZHANG Lianyang, et al. Study of standard procedure for in-hospital emergency management of multiple injury patients[J]. *Journal of Traumatic Surgery*, 2010, 12(1): 4-7.
- [15] BROWN J B, ROSENGART M R, FORSYTHE R M, et al. Not all prehospital time is equal: Influence of scene time on mortality[J]. *Journal of Trauma and Acute Care Surgery*, 2016, 81(1): 93-100.
- [16] LIER H, BERNHARD M, KNAPP J, et al. Approaches to pre-hospital bleeding management: Current overview on civilian emergency medicine[J]. *Der Anaesthesist*, 2017, 66(11): 867-878.
- [17] MOTA M A L, SANTOS M R, SANTOS E J F, et al. Trauma prehospital hypothermia prevention and treatment: An observational study[J]. *Journal of Trauma Nursing*, 2021, 28(3): 194-202.
- [18] CLARE A, KING R D. Knowledge discovery in multi-label phenotype data[C]//*Proceedings of European Conference on Principles of Data Mining and Knowledge Discovery*. [S.l.]: Springer, 2001: 42-53.
- [19] FÜRNKRANZ J, HÜLLERMEIER E, MENCÍA E L, et al. Multilabel classification via calibrated label ranking[J]. *Machine Learning*, 2008, 73(2): 133-153.
- [20] READ J, PFAHRINGER B, HOLMES G. Multi-label classification using ensemble of pruned sets[C]//*Proceedings of the 8th IEEE International Conference on Data Mining*. [S.l.]: IEEE, 2008: 995-1000.
- [21] TSOUMAKAS G, KATAKIS I, VLAHAVAS I. Random k-labelsets for multilabel classification[J]. *IEEE Transactions on Knowledge and Data Engineering*, 2010, 23(7): 1079-1089.
- [22] TSOUMAKAS G, KATAKIS I. Multi-label classification: An overview[J]. *International Journal of Data Warehousing and Mining (IJDWM)*, 2007, 3(3): 1-13.
- [23] TSOUMAKAS G, KATAKIS I, VLAHAVAS I. Mining multi-label data[M]//*Data Mining and Knowledge Discovery Handbook*. Boston, MA: Springer, 2009: 667-685.

作者简介:



赵鹏飞(1993-),男,硕士研究生,研究方向:数据挖掘、人工智能等,E-mail: 343539633@qq.com。



刘华(1981-),通信作者,女,讲师,硕士,研究方向:人机交互等,E-mail: 56162097@qq.com。