DOI:10.16781/j.0258-879x.2018.12.1390

·短篇论著·

基于决策树的住院烧伤患者医疗救治流程优化及规则挖掘

刘文宝，任东彦，陶 峰，陈国良*
海军军医大学（第二军医大学）海军医学系海军卫勤与装备教研室，上海 200433

[摘要] **目的** 探讨基于决策树的归纳分类算法在医疗救治流程优化中的应用。**方法** 以住院烧伤患者检测结果为基本资料，以医疗救治效率为决策目标，将基于决策树的归纳分类算法运用于医疗救治流程优化，构建其决策树模型，并挖掘出医疗救治流程优化的有用规则。**结果** 经过决策树流程优化，在 10 个病理属性中，有 4 个属性对确定患者的救治方案起到关键作用，即烧伤程度、血生物化学、血压、脉搏。当患者烧伤程度为轻度时，仅需通过考察血生物化学属性即可确定救治方案；当患者烧伤程度为中度时，首先通过考察血生物化学属性，进而再通过考察血压或脉搏属性即可确定救治方案；当患者烧伤程度为重度时直接采用紧急救治方案。**结论** 以决策树为代表的数据挖掘技术能够较好地辅助烧伤鉴别诊断，优化救治流程。

[关键词] 烧伤；临床方案；流程优化；决策树；算法
[中图分类号] R 826.54 **[文献标志码]** A **[文章编号]** 0258-879X(2018)12-1390-05

Process optimization and rule mining of medical treatment for burn inpatients based on decision tree

LIU Wen-bao, REN Dong-yan, TAO Feng, CHEN Guo-liang*
Department of Naval Health Service and Medical Equipment, Faculty of Naval Medicine, Navy Medical University (Second Military Medical University), Shanghai 200433, China

[Abstract] **Objective** To explore the application of inductive classification algorithm based on decision tree in optimization of medical treatment process. **Methods** Taking the test results of the burn inpatients as general data, we used inductive classification algorithm based on decision tree for medical treatment process optimization with medical treatment efficiency as the target. The model of decision tree was constructed and the rules for the optimization of medical treatment process were excavated. **Results** Among 10 pathological attributes, extent of burn, blood biochemistry, blood pressure and pulse played key roles in determining the patient treatment program after optimizing decision tree process. When the burn was mild, the treatment plan could be determined only by examining blood biochemistry indexes. When the burn was moderate, the treatment plan could be determined first by examining blood biochemistry indexes and then by examining blood pressure or pulse. When the burn was severe, emergency treatment plan should be adopted directly. **Conclusion** Data mining technology represented by decision tree can contributes to differential diagnosis of burn and optimization of the treatment process.

[Key words] burn patients; clinical protocols; process optimization; decision trees; algorithms
[Acad J Sec Mil Med Univ, 2018, 39(12): 1390-1394]

平时，医院的医疗救治分析手段主要集中在对病例病理特征的经验分析方面，且对各项检查（检验）指标过分依赖，往往医师要求检验（检查）越全面越好，因而对卫生资源消耗较多。在战时，时间紧迫且卫生资源有限的情况下，为提高救治效率，需要对伤病员进行快速诊断，传统的医疗救治分析手段显然不能满足战时“伤员多、时间紧、资源缺”的特点。随着大数据及医疗救治向智能化、现代化方向的发展，采用科学方法从大量的医疗数据中提取出有利于诊断决策的关键数据从而辅助战时伤病员鉴别诊断显得尤为重要[1-2]。本文充分利用数据挖掘技术，以住院烧伤患者检测结果为基本资料，结合烧伤病理特征，对烧伤医疗救治的流程进行分析和数据挖

[收稿日期] 2018-04-16 **[接受日期]** 2018-06-12
[基金项目] 军队后勤科研重点项目(BWS13C008, BWS17J020). Supported by Key Project of Military Logistics Research (BWS13C008, BWS17J020).
[作者简介] 刘文宝，博士，副教授．E-mail: pupfish@sina.com
*通信作者(Corresponding author). Tel: 021-81871109, E-mail: cgl307@126.com

掘，利用决策树算法和诊断指标建立判别规则，对检查（检验）指标进行筛选，选出对快速诊断影响最关键的指标并发现隐藏在医疗数据中的有用规则，从而揭示救治流程数据背后蕴含的规律，优化救治流程，提高救治效率。

1 决策树（decision tree）基础理论

决策树是用于分类和预测的主要技术，它着眼于从一组无规则的样例推理出决策树表示形式的分类规则，采用自顶向下的递归方式，在决策树的内部节点进行属性值的比较，并根据不同属性判断从该节点向下分枝，在决策树的叶节点得到结论[3-4]。因此，从根节点到叶节点就对应着一条合理规则，整棵树就对应着一组表达式规则。一个典型的决策树由一个根节点、若干个内部节点和若干个叶节点组成。叶节点与决策结果相对应，其他每个内部节点表示一个属性测试，每个分枝代表一个测试输出，每个叶节点代表一种类别，图 1 简要描述了决策树的生成过程。

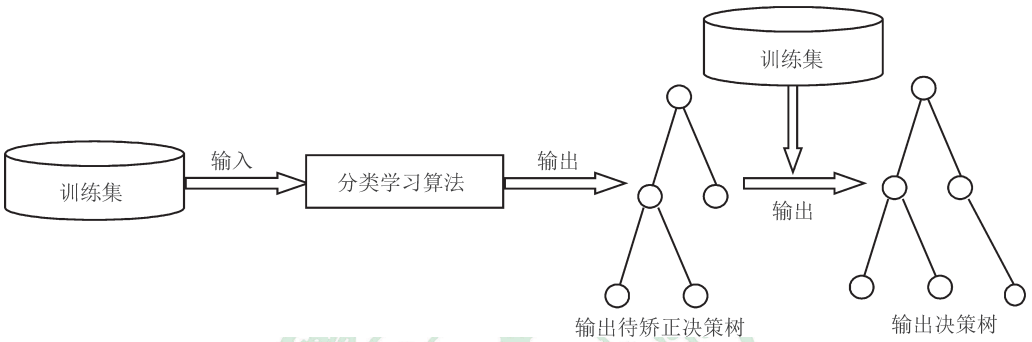

图 1 决策树生成过程

基于决策树算法的一个最大优点是它在学习过程中不需要使用者了解很多背景知识，只要训练样例能够用属性-值对的方式表示出来，就能使用该算法进行学习。常用的决策树算法有 ID3、C4.5、CART、CHIAD 和 PUBLIC 算法。决策树学习算法最为典型的是 ID3 算法，后来，C4.5 算法对 ID3 算法做出了较大改进，并且凭借其独特的特点和突出的优势在金融、医疗等行业得到成功应用[5]。

2 基于 C4.5 算法的医疗救治流程优化及规则挖掘算法设计

医疗救治过程是一个非线性、时变系统，涉及到众多参数，难以用准确的数学解析式描述参数的变化与救治效率之间的关系。本文采用 C4.5 算法构造决策树，并通过决策树获取不同精度的控制规则，发现医疗救治过程病理属性的变化与救治效率之间的关联，再利用这些规则实现医疗救治过程的优化与控制。基于 C4.5 算法的医疗救治流程优化及规则挖掘算法计算过程如下[6-7]：

（1）训练集的获取。不同于其他数据挖掘应用过程，医学上可以通过大量临床患者数据作为样本集，而无须通过随机生成一组属性数据进而通过重复该操作 n 次得到包含 n 个样本的训练集。

（2）属性选择和属性值量化。根据临床经验或要求从大量病理属性指标中选择若干属性作为决策属性集，然后对连续型属性值进行离散化。依据 C4.5 算法构造决策树，选取烧伤病理属性项“救治方案”为类别标识属性。属性项“烧伤程度”“血压”“脉搏”“呼吸”“尿量”“意识状态”“末梢循环”“血常规”“血生物化学”和“凝血酶原时间”作为决策属性集。

（3）根据 C4.5 算法构造决策树。设 T 为数据集，类别集合为 $\{C_1,C_2,\cdots,C_k\}$，选择一个属性 V 把 T 分为多个子集。设 V 有互不重合的 n 个取值 $\{v_1,v_2,\cdots,v_n\}$，则 T 被分为 n 个子集 T_1，T_2，…，T_n，T_i 中所有实例的取值均为 v_i，令：$|T|$ 为数据集 T 的例子数；$|T_i|$为 $V=v_i$ 的例子数；$|C_j|=freq\ (C_j, T)$ 为 C_j 类的例子数；$|C_{jv}|$是 $V=v_i$ 例子中具有 C_j 类别的例子数。参照文献 [5] 的方法构建决策树，具体如下：

① 按公式计算类别信息熵：

$$H(C)=-\sum_{j=1}^{k}\frac{freq(C_j,T)}{|T|}\times\log_2\frac{freq(C_j,T)}{|T|}$$

② 按公式计算条件属性的熵：

$$H(C\backslash V)=-\sum_{i=1}^{n}\frac{|T_i|}{|T|}\times H(C)$$

③ 按公式计算信息增益：

$I(C, V)=H(C)-H(C\backslash V)$

④ 按公式计算属性 V 的信息熵：

$$H(V)=-\sum_{i=1}^{n}\frac{|T_i|}{|T|}\times\log_2(\frac{|T_i|}{|T|})$$

⑤ 按公式计算信息增益率：

$Gain\_ratio=I(C, V)/H(V)$

⑥ 根据信息增益率构造决策树。

（4）对决策树进行剪枝并提取规则加入知识库。当决策树创建时，由于数据中的噪声和孤立点，许多分枝反映的是训练集中的异常，即训练过度。为了使得到的决策树所蕴含的规则具有普遍意义，须对决策树进行剪枝。剪枝的技术包括预剪枝、后剪枝及其他方法。

（5）利用规则实现医疗救治流程的优化与控制。利用 C4.5 算法构造决策树并通过决策树获取不同精度的控制规则，发现医疗救治过程病理属性的变化与救治效率之间的关联，再利用这些规则实现医疗救治过程的优化与控制。在实际应用中，我们把每次根据决策规则改变救治过程而得到的相关结果加入训练集，作为下次挖掘的样本，根据新的训练集进行挖掘后的规则可信度和覆盖率更高。

3　基于决策树的住院烧伤患者医疗救治流程优化及规则挖掘实例分析

实现基于决策树的住院烧伤患者医疗救治流程优化，就是通过发现烧伤救治方案与烧伤病理属性的关系，发现有哪几类病理属性对患者的救治方案起到关键作用，有哪几类病理属性对患者的救治方案影响相对较小。

3.1　*烧伤患者病理属性的预处理*[8-9]　为了便于决策树的使用，结合烧伤病理特征，把涉及的烧伤病理数据分为 10 类：烧伤程度、血压、脉搏、呼吸、尿量、意识状态、末梢循环、血常规、血生物化学、凝血酶原时间。其中，血常规包括红细胞比容、血红蛋白、血小板；血生物化学包括丙氨酸转氨酶、天冬氨酸转氨酶、肌酐、尿素、钾、钠。对每一大类的病理数据进行离散化处理，用 $V(C)$ 函数表示，结合救治标准和专家经验分别给出轻度值范围或病情描述 A、中度值范围或病情描述 B、重度值范围或病情描述 C；当 $V(C)\in A$ 时记为“轻”，当 $V(C)\in B$ 时记为“中”，当 $V(C)\in C$ 时记为“重”。个别属性值结合具体情况也可分为两类“正常”和“异常”；以患者“呼吸”为例，$V(C)\in$(16～20 次/min) 为“轻”，$V(C)\in$(>20 次/min) 为“中”，$V(C)\in$(呼吸浅快) 为“重”。

3.2　*救治方案信息表的处理*　不同烧伤患者的救治方案不同，有时是病理属性相似的患者采用同一治疗方案，而有时因为某一项病理数据的不同又会采用不同的治疗方案，为了克服这一问题，临床医学专家人为地对烧伤患者救治方案进行分类，根据患者烧伤救治方案与患者烧伤病理属性的相近程度把烧伤救治方案分为 3 类，分别记为 F_1（一般处理）、F_2（常规治疗）、F_3（紧急救治）。通过以上对患者烧伤病理属性和患者救治方案信息表的预处理，以 20 例住院烧伤患者检测结果为基本资料，得到新的烧伤病理属性-救治方案信息表，如表 1 所示。

3.3　*构造决策树*　以临床患者数据作为样本集并依据 C4.5 算法构造决策树。选取烧伤病理属性-救治方案信息表的属性项“救治方案”为类别标识属性。属性项“烧伤程度”“血压”“脉搏”“呼吸”“尿量”“意识状态”“末梢循环”“血常规”“血生物化学”“凝血酶原时间”作为决策属性集。

（1）计算信息增益率

① 计算类别信息熵：类别（决策）属性为“救治方案”，该属性分为 3 类：F_1、F_2、F_3。F_1（一般处理）=11，F_2（常规治疗）=7，F_3（紧急救治）=2，$F=F_1+F_2+F_3=20$。计算公式为：

$$I(F_1, F_2, F_3)=-\frac{11}{20}\log_2(11/20)-\frac{7}{20}\log_2(7/20)-\frac{2}{20}\log_2(2/20)=1.336\ 7$$

② 计算条件属性的熵：分别计算烧伤程度、血压、脉搏、呼吸、尿量、意识状态、末梢循环、血常规、血生化、凝血酶原时间的信息增益。按下式计算烧伤程度的信息增益：

$$E(\text{烧伤程度})=\frac{9}{20}I_{e1}(F_1, F_2, F_3)+\frac{9}{20}I_{e2}(F_1, F_2, F_3)+\frac{2}{20}I_{e3}(F_1, F_2, F_3)=0.639\ 7$$

其中，$I_{e1}(F_1, F_2, F_3)=-\frac{8}{9}\log_2(8/9)-\frac{1}{9}\log_2(1/9)=0.503\ 3$

$I_{e2}(F_1, F_2, F_3)=-\frac{6}{9}\log_2(6/9)-\frac{3}{9}\log_2(3/9)=0.918\ 3$

$I_{e3}(F_1, F_2, F_3)=-\frac{2}{2}\log_2(2/2)=0$

表 1　20 例烧伤患者病理属性－救治方案信息表

No.	烧伤程度	血压	脉搏	呼吸	尿量	意识状态	末梢循环	血常规	血生物化学	凝血酶原时间	救治方案
1	中度	异常	正常	正常	正常	清醒	正常	异常	异常	正常	F_2
2	轻度	正常	正常	正常	正常	清醒	正常	异常	正常	正常	F_1
3	轻度	正常	正常	正常	正常	清醒	正常	正常	正常	正常	F_1
4	中度	正常	正常	正常	正常	清醒	正常	正常	异常	正常	F_2
5	重度	异常	正常	正常	正常	不清	正常	异常	异常	正常	F_3
6	重度	异常	正常	异常	正常	清醒	正常	异常	异常	正常	F_3
7	中度	正常	正常	正常	正常	清醒	正常	异常	异常	正常	F_2
8	轻度	正常	正常	异常	正常	清醒	正常	正常	正常	正常	F_1
9	中度	正常	正常	正常	正常	清醒	正常	正常	异常	正常	F_2
10	中度	正常	正常	正常	正常	清醒	正常	正常	正常	正常	F_1
11	中度	正常	正常	正常	正常	清醒	正常	正常	正常	正常	F_1
12	轻度	异常	正常	异常	正常	清醒	正常	正常	正常	正常	F_1
13	中度	异常	异常	正常	正常	清醒	正常	正常	正常	正常	F_2
14	轻度	正常	正常	正常	正常	清醒	正常	正常	正常	正常	F_1
15	中度	正常	正常	正常	正常	清醒	正常	异常	异常	正常	F_2
16	中度	正常	正常	正常	正常	清醒	正常	异常	正常	正常	F_1
17	轻度	正常	异常	正常	正常	清醒	正常	正常	正常	正常	F_1
18	轻度	异常	正常	正常	正常	清醒	正常	异常	异常	正常	F_2
19	轻度	正常	正常	异常	正常	清醒	正常	正常	正常	正常	F_1
20	轻度	异常	正常	正常	正常	清醒	正常	正常	正常	正常	F_1

F_1: 一般处理; F_2: 常规治疗; F_3: 紧急救治

③ 计算烧伤程度的信息增益：

$I_{\text{烧伤程度}}(F, E)=I(F_1, F_2, F_3)-E(\text{烧伤程度})=0.697\ 0$

④ 计算烧伤程度的信息熵：

$$H_{\text{烧伤程度}}(V)=-\frac{9}{20}\log_2(11/20)-\frac{9}{20}\log_2(7/20)-\frac{2}{20}\log_2(2/20)=1.369\ 0$$

⑤ 计算烧伤程度信息增益率：

$Gain\_ratio(\text{烧伤程度})=I_{\text{烧伤程度}}(F, E)/H_{\text{烧伤程度}}(V)=0.509\ 1$

采用同样的方法可计算出血压、脉搏、呼吸、尿量、意识状态（CS）、末梢循环（PC）、血常规（RBT）、血生物化学（BE）、凝血酶原时间（PT）的信息增益率：

$Gain\_ratio(\text{血压})=I_{\text{血压}}(F, E)/H_{\text{血压}}(V)=0.228\ 0$

$Gain\_ratio(\text{脉搏})=I_{\text{脉搏}}(F, E)/H_{\text{脉搏}}(V)=0.043\ 3$

$Gain\_ratio(\text{呼吸})=I_{\text{呼吸}}(F, E)/H_{\text{呼吸}}(V)=0.217\ 5$

$Gain\_ratio(\text{尿量})=I_{\text{尿量}}(F, E)/H_{\text{尿量}}(V)=\text{无穷大}$

$Gain\_ratio(\text{CS})=I_{\text{CS}}(F, E)/H_{\text{CS}}(V)=0.650\ 8$

$Gain\_ratio(\text{PC})=I_{\text{PC}}(F, E)/H_{\text{PC}}(V)=\text{无穷大}$

$Gain\_ratio(\text{RBT})=I_{\text{RBT}}(F, E)/H_{\text{RBT}}(V)=0.166\ 0$

$Gain\_ratio(\text{BE})=I_{\text{BE}}(F, E)/H_{\text{BE}}(V)=0.440\ 4$

$Gain\_ratio(\text{PT})=I_{\text{PT}}(F, E)/H_{\text{PT}}(V)=\text{无穷大}$

（2）确定选择结点

以信息增益率最大的属性为根节点，通过以上计算可知尿量、末梢循环和凝血酶原时间的信息增益率为无穷大，进一步通过对烧伤病理属性-救治方案信息表和原始病理数据分析可知，烧伤患者在不同烧伤情况下尿量、末梢循环和凝血酶原时间均正常，这是由于住院烧伤患者经过急诊救治后个别属性指标趋于稳定，收集到的数据为正常值，在构建决策树时此 3 项属性可忽略。因此，选择意识状态属性进行分枝进一步分析，意识状态为“不清醒”与烧伤程度为“重度”，对应的归类都为 F_3，该处形成叶节点。结合临床经验，此处选择烧伤程度属性进行分枝最为恰当，烧伤程度取“轻度”“中度”时对应的归类均不唯一，为 F_1 或 F_2。

在烧伤程度分别为轻度或中度时，计算血

压、脉搏、呼吸、尿量、血常规、血生物化学的信息增益率。选择增益率最大的一个属性作为第一层分类控制节点（即根节点），然后对每一个分枝再次计算决策属性的信息增益率，以确定下一步分类的属性项。如此，即可最终得到烧伤患者医疗救治流程优化决策树，如图 2。

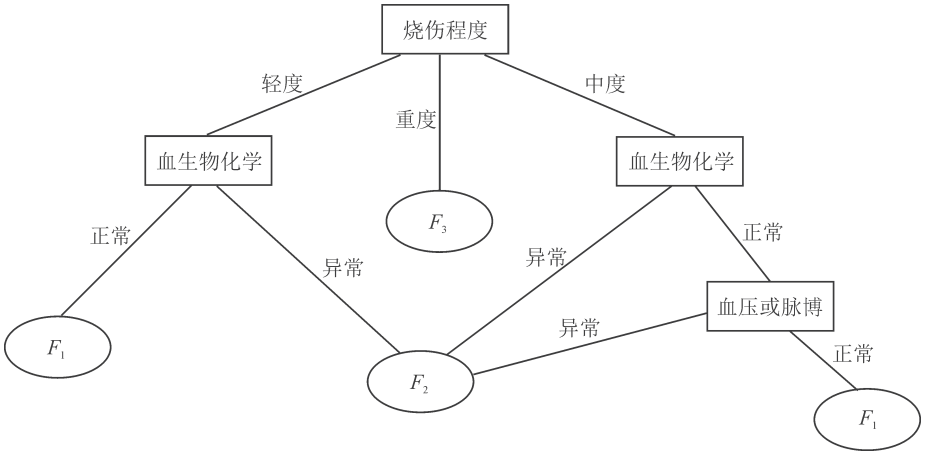

图 2　住院烧伤患者医疗救治流程优化决策树

F_1: 一般处理; F_2: 常规治疗; F_3: 紧急救治

3.4　*医疗救治流程优化规则挖掘*　结合本文构造出的住院烧伤患者医疗救治流程优化决策树，从根节点到叶节点就对应着一条合理规则，整棵树就对应着表达式规则，详述如下：（1）经过决策树流程优化，在 10 个病理属性中有 4 个属性对确定患者的救治方案起到关键作用，即烧伤程度、血生物化学、血压、脉搏；（2）当患者烧伤程度为轻度时，仅需通过考察血生物化学属性即可确定救治方案；（3）当患者烧伤程度为中度时，首先通过考察血生物化学属性，进而再通过考察血压或脉搏属性即可确定救治方案；（4）当患者烧伤程度为重度时直接采用治疗方案 F_3。

4　结　论

决策树的生成过程是一个不断通过样本集优化改进的过程，图 2 是根据表 1 烧伤病理属性-救治方案信息表构建的住院烧伤患者医疗救治流程优化决策树。本文以 20 例住院烧伤患者检测结果为基本资料进行分析，样本量偏少，随着医疗信息化的发展和住院烧伤患者医疗信息的丰富，决策树的构建会更加优化。总之，利用 C4.5 算法可构造出住院烧伤患者医疗救治流程优化决策树并通过决策树获取控制规则，从而发现医疗救治过程病理属性的变化与救治效率之间的关联，进而利用这些规则实现住院烧伤患者医疗救治过程的优化与控制。

[参 考 文 献]

[1] 罗堃,代冕. 数据挖掘技术在医疗大数据中的应用研究[J]. 信息与电脑(理论版),2016(6):45-47.

[2] 张世红,徐国桓,刘会霞,龚文涛. 数据挖掘在医学上的应用[J]. 医学情报工作,2004(6):408-410.

[3] 徐蕾,贺佳,孟虹,王忆勤,贺宪民,范思昌,等. 基于信息熵的决策树在慢性胃炎中医辨证中的应用[J]. 第二军医大学学报,2004,25:1009-1012.
XU L, HE J, MENG H, WANG Y Q, HE X M, FAN S C, et al. Application of decision tree based on entropy in traditional Chinese medicine symptom analysis of chronic gastritis[J]. Acad J Sec Mil Med Univ, 2004, 25: 1009-1012.

[4] 孟宜成,刘文奇,李月秋. 基于粗糙集的决策树优化算法[J]. 昆明理工大学学报(理工版),2009,34:95-98.

[5] 张婧,王书海. C4.5 算法在医疗保险数据挖掘中的应用研究[J]. 石家庄铁道学院学报(自然科学版),2008,21:37-40.

[6] 刘之家,蓝贞雄,廖伟志. 基于 C4.5 算法的混合生产过程优化与控制[J]. 广西教育学院学报,2010(4):171-174.

[7] 潘浩,蔺莉. 基于决策树的毕业生课程优化算法设计[J]. 信息技术,2010(8):80-83.

[8] 孙向东,董长征,陈晓妍,唐玲,王侃. 决策树在原发性肝癌鉴别诊断中的应用[J]. 医学信息学杂志,2015,36:56-59.

[9] 何爱香,张勇. 基于遗传算法和决策树的肿瘤分类规则挖掘[J]. 山东大学学报(理学版),2007,42:91-95.

[本文编辑]　尹　荼