

结果表明:五福饮以生药 $10.5\text{ g}\cdot\text{kg}^{-1}$ 和 $21\text{ g}\cdot\text{kg}^{-1}$ (体重)灌胃荷瘤小鼠,可明显提高小鼠的脾淋巴细胞增殖反应($P<0.05$)、NK细胞活性($P<0.05$)、血清IFN- γ 、TNF- α 含量($P<0.05$),提示五福饮可增强荷瘤小鼠的免疫功能。五福饮各组对S₁₈₀移植瘤的抑制率在11%~16%,但与模型组比较均无明显差异($P>0.05$)。因此,笔者认为五福饮本身并不具有直接抑制肿瘤的作用,而是通过增强免疫功能起到一定的辅助抗肿瘤作用。这一结果将为中药五福饮的保健食品开发和肿瘤辅助治疗提供实验依据。

REFERENCES

- [1] ZHANG J B. Jingyue Quanshu(景岳全书) [M]. Shanghai Scientific and Technical Press, 1964: 981.
- [2] XU B L, SUN J. Experimental study on the effect of reinforcing drugs on cellular immune function [J]. Zhejiang J Tradit Chin Med(浙江中医杂志), 1996(5): 219-220.
- [3] WANG Z F, MIAO X L, CHEN D C, et al. Effects of ginsenosides on CD19 cell and NK cell in scalded rats with

- sepsis [J]. Chin J Integr Tradit West Med Intensive Crit Care(中国中西医结合急救杂志), 2007, 14(4): 219-221.
- [4] JI Y B. The pharmacology and application of anticancer(抗癌中药药理与应用) [M]. Harbin: Heilongjiang science and Technology Press, 1992: 28-30.
- [5] WANG J, YU Z, LI G S, et al. Experimental study on rehmanioside A in yin and regulating immune functions [J]. Chin Pharm J(中国药理学杂志), 2002, 37(1): 20-22.
- [6] LIU Z X, ZUO Z Y. Study on the immune pharmacological effect of Radix Angelicae Sinensis [J]. Chin J Integr Tradit West Med(中国中西医结合杂志), 1992, 12(6): 378-380.
- [7] MINISTRY OF HEALTH. Technical Standards for Testing and Assessment of Health Food(保健食品检验与评价技术规范) [M]. Tsinghua Tongfang Electronic Press, 2003: 22-23.
- [8] HUANG L, LI L M, TANG L Y, et al. Effects of Baishu decoction on immune function in mice [J]. Pharmacol Clin Chin Mater Med(中药药理与临床), 2012, 28(1): 114-115.
- [9] YAO S J, LIU B Y, SUN Y. Effects of Rhizoma Atractylodis Macrocephalae on NK and IL-2 activity of tumor-bearing mice [J]. Bull Med Res(医学研究通讯), 2005, 34(12): 52.
- [10] ZHANG Z S, SHI S, YANG C H, et al. Research on the effect of glycyrrhiza polysaccharide on immunity [J]. Prog Mod Biomed(现代生物医学进展), 2008, 8(10): 1835-1837.

收稿日期: 2013-01-21

数据挖掘技术在动物造模中的应用

余炜, 万悦(安徽朗诺医药有限公司, 合肥 230601)

摘要: 目的 将数据挖掘技术引入动物造模的过程中,使用数据挖掘技术建立动物造模的计算机仿真模型,展示数据挖掘技术在动物造模中的价值。方法 将30只SD大鼠,♂,随机分为对照组和实验组。实验组采用完全弗氏佐剂右后足趾腱膜下注射致炎造模,对照组同部位注射等量生理盐水。第0,14,21天测量大鼠足跖部的厚度;在不同时点对关节部位进行关节炎指数(AI)的评定;第0,7,14,21天观察大鼠的体重变化;第14天和第21天分别随机选取5只大鼠,断颈处死取血测IL-1 β 、IL-17、TNF- α 含量并观察病理切片。将上述采集的指标纳入数据库,利用SPSS MODELER进行数据挖掘分析。结果 利用数据挖掘技术建立了3个模型:综合模型、C5.0决策树模型和神经网络模型,三者的正确率依次为95%,90%,100%,累计增益曲线均有不同程度的明显提升,同AI评分及病理切片结果吻合。结论 本研究通过结合佐剂性关节炎大鼠造模的实例,详细对数据挖掘技术在动物造模中应用的流程、方法、优化和评价进行了说明和阐述,新的数据处理方法的引入可以针对多种评估指标构建出不同的预测评估模型,有效地对数据进行解析并提取有意义的信息辅助决策。

关键词: 数据挖掘; 动物造模; SPSS MODELER

中图分类号: R965.1

文献标志码: A

文章编号: 1007-7693(2013)09-0952-08

Application of Data Mining Techniques in the Process of Animal Modeling

YU Wei, WAN Kai(Anhui Renovo Pharmaceutical Co., Ltd., Hefei 230601, China)

ABSTRACT: OBJECTIVE To use data mining technology in the process of animal modeling to create a computer simulation model, and show the value of data mining technology. **METHODS** Thirty male SD rats were randomly divided into control

基金项目: 安徽省教育厅 2011 年重点项目(KJ2011Z230)

作者简介: 余炜, 男, 硕士 Tel: (0551)68778262 E-mail: YUWEI52213@163.com

group and model group, administered with intradermal injection of 0.25 mL normal saline and 0.25 mL complete Freund's adjuvant, respectively, and adjust the dose to 0.1 mL four days later. Measure the joint swelling degree on 0, 14 and 21d. Assess the arthritic index (AI) in different times and record rat weight on 0, 7, 14 and 21d. After 14 d and 21 d, five rats were randomly selected from each group, the ankle joints were prepared for histopathologic study and the level of IL-1 β , IL-17 and TNF- α in serum were determined by enzyme-linked immunosorbent assay, respectively. Then the collected data was entered into the database, and SPSS MODELER for data mining analysis was used. **RESULTS** Three simulation models were established with data mining techniques: the integrated model, C5.0 decision tree model and neural network model, the correct rate for these models were 95%, 90%, 100%, respectively. The cumulative gain curve of these models improved significantly in varying degrees, consistent with AI score and pathological results. **CONCLUSION** Combined with a practical example (adjuvant arthritis rats model), we describe and elaborate the application process, method, optimization and evaluation of data mining technology in animal modeling. With the introduction of new data processing methods, we can build a series of forecast evaluation models with different evaluation indicators, and make the right decision by using the information more effectively.

KEY WORDS: data mining; animal modeling; SPSS MODELER

实验研究是现代科学发展的基础, 动物实验是生命科学实验研究中重要的组成部分。学术界在 20 世纪 60 年代正式提出了“动物模型”的概念^[1]。动物模型是医学科学实验中, 采取特定的设计, 用动物模拟人体而制作的一种模型。目前使用动物模型已成为现代生物医学研究中一种非常重要的实验方法和手段, 大大促进了疾病的发生、发展规律和防治措施等方面的研究。

但是有关造模的因素非常复杂, 究竟哪一种造模因素最符合临床实际目前尚缺乏一致性标准。同时, 对于造出的较为典型的动物模型, 具有科学的评估标准和定量的评价手段不多。本实验拟利用数据挖掘技术针对当前实验研究中动物造模的不足进行弥补优化, 并通过一个佐剂性关节炎大鼠造模的实例进行技术路线的展示和说明。

1 数据挖掘简述

1.1 基本概念

数据挖掘, 又称为数据库中的知识发现。美国人工智能协会(AAAI)在 1989 年的会议上首次提出了数据库知识发现(KDD)的概念, 指出知识发现是一个过程, 在这个过程中, 计算机从数据库中识别出新的、有效的和具有潜在应用价值的数据和信息, 利用不同的算法将这些数据和信息加工成可以被我们理解的概念、规则和模式^[2]。虽然数据挖掘只是知识发现的一个步骤, 但在实际中往往等同看待。其发展出的多种数据处理方法, 可以有效弥补传统统计学方法的不足, 更有效地对多类型数据进行分析 and 处理。

数据挖掘^[3]是对存储于数据库中的大量数据, 通过查询和抽取的方式获得以前未知的有用信息、模式和规则的过程。其可以针对海量数据或者小样本高维数据进行有效地处理和分析。常用

模式及方法有: 分类、聚类、关联和预测。

数据挖掘的基本过程有 7 个步骤, 分别为:

①数据的准备阶段; ②数据的选择阶段: 对目的相关的数据进行检索和分析; ③数据的处理阶段: 对数据进行清洗; ④数据的变换阶段: 使数据更适合进一步的挖掘; ⑤数据的挖掘阶段: 使用人工智能的算法提取有用的信息; ⑥模型的解释与评估: 评价模型的计算结果对目的变量的解释程度, 同时发掘隐藏在数据背后的新价值; ⑦知识的展示: 用可视化的手段向客户形象地展示提取到的知识^[4]。这几个步骤构成一个循环, 整个过程是螺旋式的上升过程。

目前数据挖掘行业通行的准则为跨行业数据挖掘标准过程(Cross-Industry Standard Process for Data Mining, CRISP-DM)^[5], 其定位是面向行业、工具导向、面向应用。主要包括 6 个阶段: 商业理解、数据理解、数据准备、模型建立、模型评估和结果发布。

1.2 数据挖掘软件

行业内有 Statsoft Statistics、SAS Enterprise Miner、Oracle DM、Matlab、Angoss、SPSS MODELER(原 Clementine)等众多软件。SPSS MODELER 因拥有丰富的数据挖掘算法, 支持与数据库之间的数据和模型交换, 具有可视化操作界面、简单易用、分析结果直观易懂、图形功能强大等特点, 从众多的软件中脱颖而出, 当前最新版本为 14.2, 也是本研究中所采用的软件。

2 案例

2.1 动物造模材料与方法

2.1.1 试剂与仪器 完全弗氏佐剂(CFA, Sigma, 批号: 051M8725)。IL-1 β 、IL-17、TNF- α ELISA 检测试剂盒均购于合肥欣乐生物有限公司(批号:

20120302A)。蒸馏水,生理盐水为实验室自制。甲醛、乙醚、乙醇(无水)、硝酸、二甲苯、硫酸铝钾、苏木精、丙酮、伊红等均为国产分析纯。

550 酶标仪(日本 Bio-Rad 公司);DK-S22 电热恒温水浴锅(上海精宏);RM 2135 石蜡切片机(德国莱卡)。

2.1.2 动物分组及处理 SD 大鼠,30 只,♂,清洁级,7~9 周龄,体质量 200~250 g,购于安徽医科大学实验动物中心,皖医实动物第 01 号。进行适应喂养 7 d,检查健康后纳入正式实验,随机分成对照组和实验组,每组 15 只。

实验组大鼠采用右后足跖部皮下注射法,一次注入 CFA 0.25 mL,4 d 后同部位再次注入 CFA 0.1 mL,复制佐剂型关节炎(adjutant arthritis, AA)大鼠模型。对照组大鼠同部位注射等量生理盐水。其他不做任何处理。

2.1.3 观测指标 每天观察上述 2 组大鼠的生活状态(饮食及活动变化,毛色的改变)和关节部位红肿的情况;于实验第 0,14,21 天用游标卡尺测量大鼠足跖部的厚度,每次均测量 3 次,取算术平均值;在实验第 0,7,14,21 天观察大鼠的体质量变化。第 14 天和 21 天分别随机选取 5 只大鼠,断颈处死取血,用 ELISA 试剂盒作血清中细

胞因子的检测,同时制作踝关节病理组织切片。

造模的判断标准:采用关节炎指数(arthritis index, AI)和病理切片综合评定。关节炎评分法(0~4 级)。0 分:无关节炎;1 分:关节有红色斑点或轻度肿胀;2 分:关节中度红肿;3 分:关节重度红肿;4 分:关节严重红肿且不能负重。4 个关节炎的评分之和用 AI 表示,每个关节炎最高分为 4 分,4 个关节炎的总和最高为 16 分。通常,若实验组大鼠关节炎评分 AI>4 分,则认定为关节炎诱导成功,若 AI<4 分则认为造模失败。病理切片采用 HE 染色,光镜下观察大鼠关节滑膜增生、炎性细胞浸润及软骨骨质破坏及关节周围组织水肿等病理学变化。开始造模后 14 d,根据上述判定标准,造模成功。

2.2 基于数据挖掘的数据处理

2.2.1 主要思路 在动物造模中引入数据挖掘分析的主要思路为:先利用少量实验动物数据建模,将采集的指标纳入数据库,利用 SPSS MODELER 进行数据挖掘分析,进而利用这些样本数据训练出合适的数学模型并结合医药学专业知识对模型进行优化。在大范围造模开始前,依据计算机数据挖掘模型判定入选实验的动物。

完整的数据挖掘分析过程见图 1。

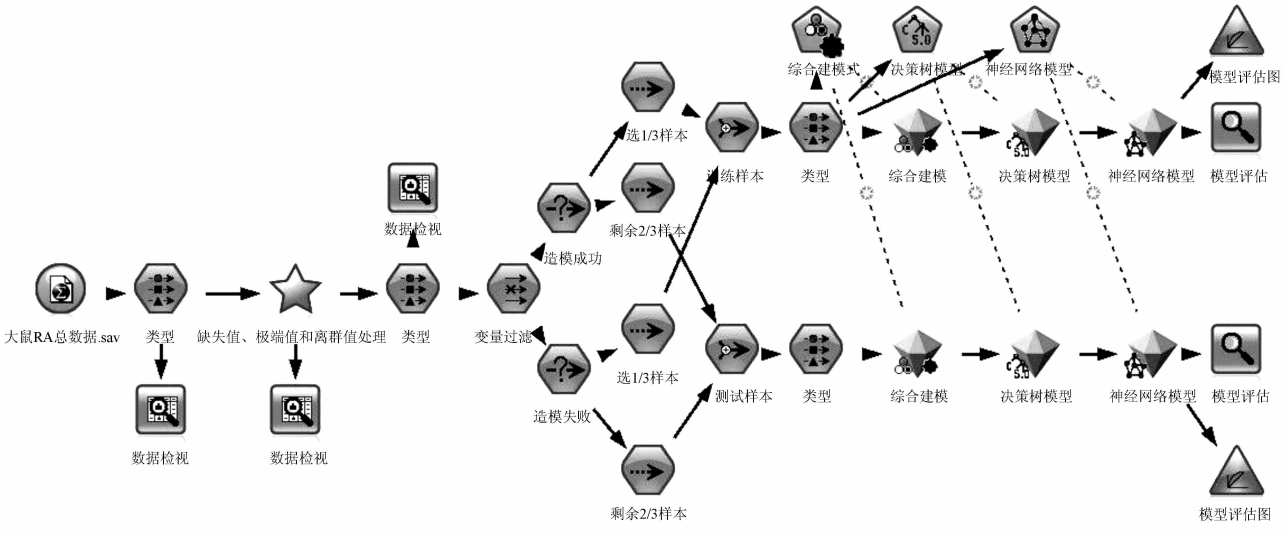


图 1 数据挖掘总模型图
Fig 1 The general plan of data mining

2.2.2 观测指标的采集 将上述观测结果按下列字段予以记录:SD 大鼠编号(ID);SD 大鼠组别(GROUP);SD 大鼠第 0 天体质量(WEIGHT1);SD

大鼠第 7 天体质量(WEIGHT2);SD 大鼠第 14 天体质量(WEIGHT3);SD 大鼠第 21 天体质量(WEIGHT4);SD 大鼠第 0 天左关节肿胀度

(JOINT_SCORE1L); SD 大鼠第 0 天右关节肿胀度(JOINT_SCORE1R); SD 大鼠第 14 天左关节肿胀度(JOINT_SCORE2L); SD 大鼠第 14 天右关节肿胀度(JOINT_SCORE2R); SD 大鼠第 21 天左关节肿胀度(JOINT_SCORE3L); SD 大鼠第 21 天右关节肿胀度(JOINT_SCORE3R); 第 14 天血清中 IL-1 β 含量(IL_1A); 第 21 天血清中 IL-1 β 含量(IL_1B); 第 14 天血清中 IL-17 含量(IL_17A); 第 21 天血清中 IL-17 含量(IL_17B); 第 14 天血清中 TNF- α 含量(TNF_a1); 第 21 天血清中 TNF- α 含量(TNF_a2)。

2.2.3 数据录入和导入 将上述数据录入 SPSS STATISTICS 后, 将生成的数据文件导入 SPSS MODELER 中构建总的数据库。利用 Type 节点(类型)对数据类型进行定义, 本例中将 SD 大鼠编号(ID)字段角色设为 None; SD 大鼠组别(GROUP)字段角色设为: Target; 其余字段设为: In。

2.2.4 数据清洗 选择 Data Audit(数据检视)节点对导入的数据进行分析处理, 主要处理缺失值、离群值和极端值。并采用合适的方法弥补缺失值(通常选用 Algorithm 方法处理)。根据本实验, 将离群值和极端值的诊断标准定为: Interquartile ranges from upper/lower quartiles(四分位数法)较好。具体的修正方法选择 Coerce(用距离其最近的正常数据予以替代)较好。处理完毕后选择生成缺失值、离群值和极端值的处理超节点。

2.2.5 数据预分析 通过 Data Audit(数据检视)节点了解各个字段的统计学描述, 探索合适的处理方法, 利用 Feature Selection(特征选择)节点检视各字段的重要性。

2.2.6 数据建模 数据准备过程结束后, 通过选择自动综合建模节点和不同的独立建模节点, 利用录入的样本数据训练仿真模型。本例中采取从总体数据库中分层随机抽取 1/3 样本构建训练样本库, 用于数学仿真模型的生成(图 1 中的上半支路径)。

根据不同建模节点的不同特点和数据要求, 对不同的模型参数进行详细设定。本例中主要选择自动综合建模节点(系统整合了 5 个模型进行综合评分, 包括决策树、Logistic 回归、贝叶斯网络、最邻近结点 KNN 和支持向量机 SVM)、C5.0 决策

树建模节点和神经网络建模 3 个节点。

2.2.6.1 综合建模 综合建模节点主要是利用 MODELER 软件中的“自动分类器”节点, 使用多种不同方法估算和比较名义(集合)或二元(是/否)目标的模型, 使得在一次建模运行中即可尝试各种方法。通过综合选择可以选用算法(C&R 树、QUEST、CHAID、C5.0、Logistic 回归、决策列表、贝叶斯网络、判别式、KNN 和 SVM), 实验选项的多个组合^[6]。类似于组成一个由多个领域专家组成的专家团, 各个专家都对数据给出自己的理解和评估结论, 最终按照置信度加权投票生成一个综合模型用于评分和评价。

2.2.6.2 C5.0 决策树模型 C5.0 是众多决策树分析法中的一种, 目前常见的其他算法还有 C&RT、QUEST、CHAID 等。其中最具影响的是 Quinlan 于 1986 年提出的 ID3(Iterative Dichotomic Version3)算法。C5.0 是 ID3 的扩展, 它将分类领域从类别属性扩展到数值型属性, 主要在执行效率和内存使用方面进行了改进^[7]。本实验在建立动物模型预测训练时采用的是 ID3 的改进算法 C5.0。

2.2.6.3 神经网络模型 人工神经网络^[8], 就是基于模仿生物大脑的结构和功能而构成的一种信息处理系统或计算机, 简称神经网络, 简称为 ANN(Artificial Neural Network)。其本质上是人脑处理信息方式的简化模型, 通过模拟大量相互连接的简单处理单元工作(类似神经元的抽象化版本)。神经网络是功能强大的一般函数估计器, 只需要最少的统计或数学知识就可以对其进行训练或应用: 神经网络是神经系统运转方式的简单模型。其基本单元是神经元, 通常将其组织到层中, 这些处理单元都位于层中。模型通常有三部分: 一个输入层, 其单元代表输入字段; 一个或者更多的隐藏层; 以及一个输出层, 其单元代表输出字段。这些单元通过不断变化的连接强度或权值连接。输入数据显示在第一层, 其值从每个神经元传播到下一层的每个神经元。最终从输出层中输出结果。

2.3 结果

2.3.1 数据清洗的结果 数据清洗过程主要包括填充缺失值, 修正离群值和极端值, 详细的处理过程见图 2。

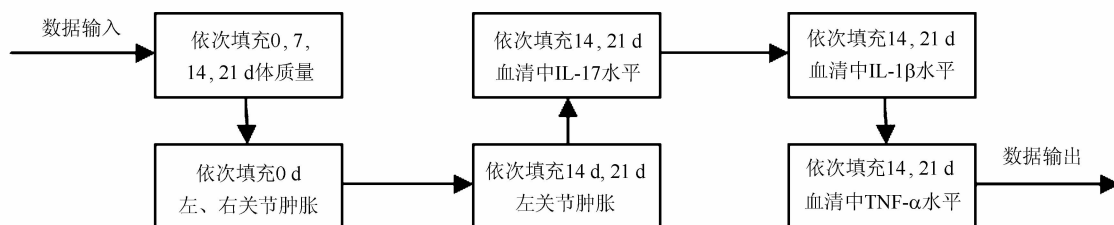


图2 缺失值、离群值和极端值的处理

Fig 2 The process of deal with missing values, outliers and extreme values

2.3.2 数据预分析重要性筛选的结果 通过分析可知, 本次采集的所有变量中重要的如表 1 所示。

表 1 检视各字段的重要性

Tab 1 Check the importance of the fields

等级	字段	测量类型	重要性	值
1	IL_1A	连续型数据	重要	1.0
2	IL_1B	连续型数据	重要	1.0
3	IL_17A	连续型数据	重要	1.0
4	JOINT_SOCRE3R	连续型数据	重要	1.0
5	IL_17B	连续型数据	重要	1.0
6	JOINT_SOCRE2R	连续型数据	重要	1.0
7	WEIGHT3	连续型数据	一般重要	0.922
8	WEIGHT2	连续型数据	不重要	0.786
9	WEIGHT4	连续型数据	不重要	0.541
10	WEIGHT1	连续型数据	不重要	0.424

注: 重要: 值>0.95; 一般重要: 值 0.9~0.95; 不重要: 值<0.9

Note: Important: value>0.95; Marginal: value 0.9~0.95; Unimportant: value<0.9

2.3.3 模型评价的结果 数据建模完成后, 可利用模型评价节点(主要是 Analysis 节点和 Evaluation 节点)对模型进行评价。本例中采取建模的 1/3 样本和剩余的 2/3 样本对计算机生成的数学仿真模型进行评估。1/3 样本训练样本评估结果见表 2、图 3。可见数据挖掘生成的模型正确性较高, 各模型的结果一致性较好。累计增益图中增益曲线越往左上方凸出, 效果越好, 从图中可以看出使用仿真模型后的增益曲线获得了比较明显提升。

表 2 训练模型评估 A

Tab 2 Assessment of the training model-A

模型类别	样本数	判断正确	判断错误	正确率/%	错误率/%
综合模型	10	10	0	100	0
C5.0 决策树模型	10	10	0	100	0
神经网络模型	10	10	0	100	0

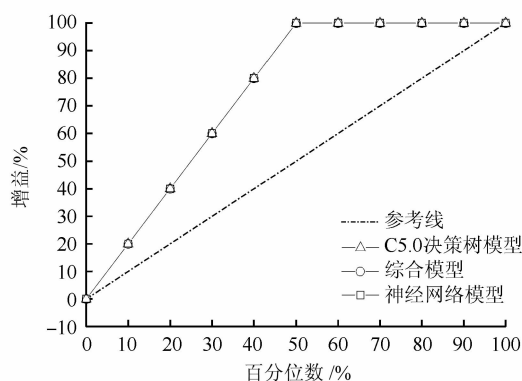


图 3 训练模型评估 B

Fig 3 Assessment of the training model-B

再利用检测样本评估模型(图 1 中的下半支路径), 结果见表 3、图 4。可见综合建模节点的正确率为 95%, C5.0 决策树的正确率为 90%, 而神经网络的正确率为 100%。结果显示, 利用神经网络的数据挖掘模型能获得把握度相当高的结果。累计增益曲线的不同程度提升也都比较明显。

表 3 测试模型评估 A

Tab 3 Assessment of the test model-A

模型类别	样本数	判断正确	判断错误	正确率/%	错误率/%
综合模型	20	19	1	95	5
C5.0 决策树模型	20	18	2	90	10
神经网络模型	20	20	0	100	0

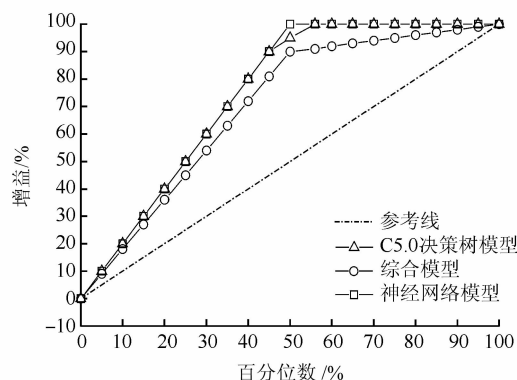


图 4 测试模型评估 B

Fig 4 Assessment of the test model-B

进一步检视神经网络模型的细节，结果见图5。可知该模型为多层感知器型神经网络，输入层有10个解释变量，分别是通过变量重要性评估排序后的前10个变量，共10个神经结点；1个隐藏层，该隐藏层中有3个神经结点；输出层有1个神经节点。

输入层的激活函数采用 Hyperbolic Tangent 形式，即：

$$\gamma(c) = \tanh(c) = \frac{e^c - e^{-c}}{e^c + e^{-c}}$$

输出层的激活函数采用 Softmax 形式，即：

$$\gamma(c_k) = e^{c_k} / \sum_{j \in \Gamma} e^{c_j}$$

式中 c 为输入函数； $\gamma(\cdot)$ 为激活函数。

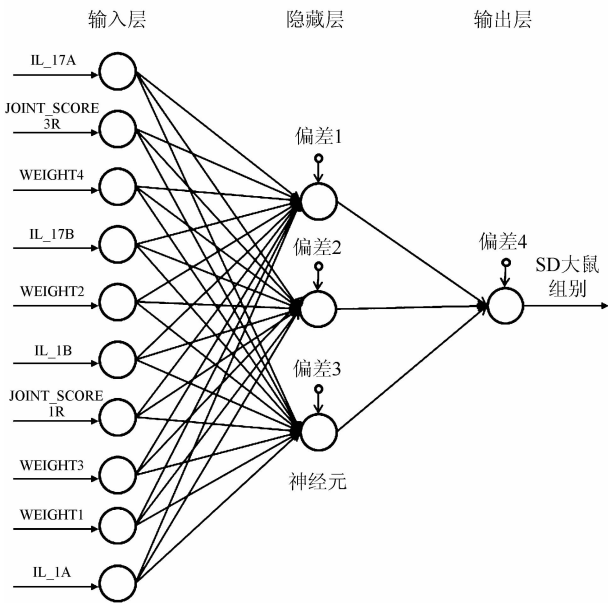


图5 神经网络模型
Fig 5 A brief outline of the Artificial Neural Network model

2.3.4 关节炎评分结果 仿真模型判定造模成功的关节炎评分为 7.6 ± 2.10 ，造模失败的评分为 2.3 ± 0.73 。

2.3.5 病理切片对比结果 仿真模型判定造模成功的病理切片显示：滑膜有不同程度的充血、水肿及大量中性粒细胞、单核细胞等炎性细胞浸润。滑膜增生明显，出现关节软骨破坏、纤维化。见图6。

仿真模型判定造模失败的病理切片显示：双侧踝关节组织学结构正常，滑膜衬里细胞1~2层，少量纤维细胞及血管。见图7。

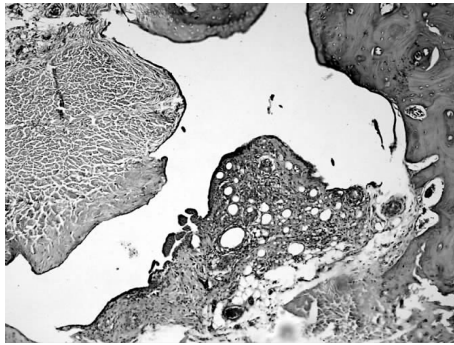


图6 大鼠踝关节病理切片(100×)
Fig 6 Histological section of ankle(100×)

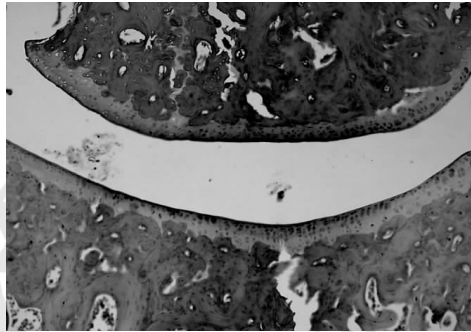


图7 大鼠踝关节组织切片(100×)
Fig 7 Histological section of ankle(100×)

2.4 讨论

2.4.1 在动物造模中引入数据挖掘分析的原因传统的动物造模有以下几点不足：①大多参照文献报道选取独立指标进行判定，缺乏多指标综合评估的方法和手段；②很多造模成功与否的确定性指标(如病理组织切片)必须处死模型后才能取得，而很多时候客观条件限制了这种验证手段的实施；③判断成功造模的动物模型需要多年造模经验，需要将这些专家意见纳入到判定系统中提高造模动物成功与否的正确率；④传统造模路径中的一些检测耗时耗力且不经济，需要一种方式将一些不必要的检测略去，在保证准确度的基础上加快筛选速度，降低筛选成本；⑤不同造模方法的评估存在定性为主，缺少定量比较的不足；⑥一些造模方法获得的数据是广义小样本(样本容量小于变量个数，表现为高维数据小样本)，传统的统计学方法在多维数据降维分析方面存在着不足；⑦传统复制模型着重于造模成功后，对成功造模路径的模仿，限于客观条件，往往不能针对扩大规模后的动物造模实施精确的多指标检测，特别是对于一些复杂动物模型，哪些是关键的指

标也缺乏足够的认知，对于造模路径背后关键变量的把握不足。

针对上述问题，本研究尝试采用数据挖掘辅助建模这一新的方式对动物模型进行评估和推广。此种方式在动物造模领域的应用国内未见文献报道，利用此技术，一方面可以更加综合地反映整个建模过程，将不同观察类别下的多个评估指标整体纳入评价体系，提高造模成功与否的判断正确性；另一方面也可利用少数实验样本构建的仿真模型用于未来同类实验推广及大范围的动物造模判定。

2.4.2 模型优化 首先，通过数据检视，利用预先分析筛选重要变量，基于这些字段的评估，我们就可以对整个检测指标的构成进行优化，在未来的动物建模过程中删除一些不必要的检测(例如本例中的 TNF- α 和体质量)，节约人力和资金，在保证造模有效性的基础上进一步优化造模路径。

其次，针对生成的仿真模型，可以尝试修改相关模型参数进行优化。例如，如果误认为造模失败的动物是成功的，将其入选后续药理实验是不合理的，我们可以通过决策树模型中误分类损失矩阵参数调整来优化模型。将风险成本由原来的 1 提高为 2，见表 4。

表 4 误分类损失矩阵

Tab 4 Misclassification costs Matrix

		预测情况	
		造模失败	造模成功
实际情况	造模失败	0	2(原始设定为 1)
	造模成功	1	0

最后，针对不同的实验和造模路径，对于连续性数据，也可以通过结合 EXCEL 进行连续数据离散化处理，进而基于离散化后的数据进行分析 and 仿真建模。

2.4.3 模型应用和推广 不同的计算机算法构建的仿真模型具有不同的特点，各自具有不同的优缺点，例如决策树模型具有较好的辨识规则，模型简单易懂，但是稳健性不足；神经网络模型具有广泛地适应性，能有效处理各种类型的数据，但是在应用规则上不如决策树清晰，所以针对不同的采集指标，选择不同的算法构建不同的模型十分必要。

有了上述数据挖掘生成的仿真动物模型，我

们未来便可将新造模的动物模型指标数据录入数据文件，链接上述模型节点进行评定，或者根据仿真模型的规则集进行判断评分。另外还可以根据发现的数据关系构建新的关联指标进行分析。也可以尝试纳入更多的评价字段构建更加系统的仿真模型，通过逐步完善来构建不同动物模型的计算机仿真模型仓库，用于后续的相关研究。针对造模中纳入的更多评估指标，可以进一步通过关联数据挖掘技术来发现潜在的指标关系，结合路径分析技术和因果分析技术对各指标背后的作用机制进行探索。

3 结语

动物造模是否成功，必须以科学的判定标准为基础，依靠现代的手段和方法对隐藏在定量数据背后的价值进行发掘。数据挖掘技术可以有效地对数据进行深度解析并提取有意义的信息辅助决策。

在互联网、银行业、电信业等领域，数据挖掘技术得到了非常有效地开发和利用。但是其在医药卫生领域的应用才刚刚起步，特别是动物造模仿真评估方面的应用还不够普遍。未来，数字化变革将渗透到医学研究的各个领域，医药不同领域的人士都要积极地提高自身信息化水平，数据挖掘技术的掌握和应用能够迅速成为一种优势，通过医药应用的拓展和融合，数据挖掘技术将在以科学证据支持为中心的医学发展过程中，显现出巨大的应用前景。

致谢：感谢安徽省教育厅 2011 科技重点项目资助了本研究(项目编号：KJ2011Z230)。同时，感谢 IBM SPSS CHINA 为本研究提供了正版 SPSS STATISTICS 20 和 SPSS MODELER 14.2 软件。

REFERENCES

[1] HUANG W H, QU Y S. Replication principle of animal models in traditional Chinese medicine research [J]. Mod J Tradit Chin Med Pharm(现代中医药), 2003(2): 15-16.

[2] LU Q C, ZOU P. Research on data mining and its application [J]. J Kunming Univ Sci Technol(Sci Technol)(昆明理工大学学报: 理工版), 2002, 27(5): 62-70.

[3] THURASINGHAM, BHAVANI M. Data Mining: Technologies, Techniques, Tools and Trends [M]. Boca Raton: CRC Press, 1999.

[4] FAYYAD U M, PIATETSKY-SHAPIO G, SMYTH P. From data mining to knowledge discovery in databases [J]. AI Magazine, 1996, 17(3): 37-54.

[5] XUE W, CHEN H G. Data Mining Methods and Applications

- in Clementine(Clementine 数据挖掘方法及应用) [M]. Beijing: Publishing House of Electronics Industry, 2010.
- [6] IBM Corporation. ModelingNodes. [DB/OL].IBM Corporation, 2011: 84-96[2012-08-19]. <ftp://ftp.software.ibm.com/software/analytics/spss/documentation/modeler/14.2/en/>.
- [7] ZHANG J W, HAN G S, ZHANG W. Application of C5. 0 Algorithm in passing ball training of RoboCup [J]. Computer Simulation(计算机仿真), 2006, 23(4): 132-134.
- [8] JIANG Z L. Introduction to Artificial Neural Networks (人工神经网络导论) [M]. Beijing: Higher Education Press, 2001: 15-54.
- [7] ZHANG J W, HAN G S, ZHANG W. Application of C5. 0

收稿日期: 2012-04-20

水蛭对血瘀证家兔血脂代谢及相关基因表达的影响

杨洪雁¹, 张香东², 刘可园¹, 杜智恒¹, 白秀娟^{1*} (1.东北农业大学动物科学技术学院, 哈尔滨 150030; 2.吉林农业大学中药材学院, 长春 130118)

摘要: 目的 观察水蛭对实验性血瘀证家兔血脂代谢及其相关基因表达的影响。方法 采用复合因素(饥饿+高脂饲料+注射肾上腺素)建立血瘀证模型, 造模结束后将实验动物随机分成5组, 高中低水蛭剂量组、模型对照组、空白对照组, 三个剂量组分别按 0.30, 0.15, 0.075 g·kg⁻¹ 给血瘀证家兔灌胃 30 d。分别于造模结束时、灌胃第 10, 20, 30 天心脏取血, 检测血清总胆固醇(TC)、甘油三酯(TG)、低密度脂蛋白胆固醇(LDL-C)含量; 采用荧光实时定量 PCR 技术检测家兔肝脏低密度脂蛋白受体(LDL-R)、载脂蛋白 E(ApoE) mRNA 的表达。结果 与模型组相比, 水蛭能显著降低血清中 TC、TG、LDL-C 水平($P<0.05$), 能显著上调血瘀证家兔肝脏中 LDL-R mRNA、ApoE mRNA 的表达($P<0.05$)。结论 水蛭具有调节血瘀证家兔血脂代谢的作用, 其机制可能与上调 LDL-R 基因和 ApoE 基因转录水平有关。

关键词: 水蛭; 血瘀证; 血脂代谢; LDL-R; ApoE

中图分类号: R972.6 文献标志码: A 文章编号: 1007-7693(2013)09-0959-05

Effect of Hirudo on Blood Metabolism and Its Related Gene Expression in Blood Stasis Syndrome Rabbits

YANG Hongyan¹, ZHANG Xiangdong², LIU Keyuan¹, DU Zhiheng¹, BAI Xiujuan^{1*} (1.College of Animal Science and Technology, Northeast Agricultural University, Harbin 150030, China; 2.College of Chinese Traditional Medicine, Jilin Agricultural University, Changchun 130118, China)

ABSTRACT: OBJECTIVE To observe the effects of Hirudo on blood metabolism and its related gene expression in blood stasis syndrome rabbits. **METHODS** The blood stasis syndrome rabbit's models induced by complex factors(starvation+high-fat feeding+adrenaline) were established. Model rabbits randomly divided into five groups, rabbits in Hirudo groups were daily administered with Hirudo(0.30, 0.15, 0.075 g·kg⁻¹) for 30 days, and rabbits in the normal and model groups were daily administered with 0.9% NaCl. The serum TC, TG, LDL-C in each different treated rabbit group were determined after models established, 10, 20, 30 d administration, respectively. The mRNA expressions of LDL-R and ApoE in liver tissue were quantified by Realtime-PCR. **RESULTS** Compared with normol group, Hirudo groups could remarkably decrease the levels of serum TC, TG, LDL-C($P<0.05$). The result of realtime-PCR demonstrated that the Hirudo could remarkably increase the mRNA expressions of LDL-R and ApoE in liver tissue of rabbit's model group. **CONCLUSION** Hirudo has the effect of regulating lipid metabolism of blood stasis syndrome rabbits, the mechanism is likely related to increasing LDL-R and ApoE mRNA expression.

KEY WORDS: Hirudo; blood stasis syndrome; lipid metabolism; LDL-R; ApoE

水蛭是一味传统中药。《神农本草经》曾记载:“水蛭味咸平,主逐恶血、瘀血、月闭,破血瘕积聚。无子,利水道。”传统多用于血滞经闭,微瘀积聚及跌打损伤等症^[1]。近年来,水蛭的医用

研究越来越广泛,从水蛭的化学成分、炮制方法到药理作用均日益深入^[2-4]。大量研究表明水蛭具有调血脂、抗血栓的作用^[5-7]。本实验观察水蛭对实验性血瘀证家兔血脂代谢的影响,并且采用实

作者简介: 杨洪雁, 女, 博士 Tel: 13796075409 E-mail: xiaopiqiu_yes@163.com *通信作者: 白秀娟, 女, 博士, 教授, 博导 Tel: 13766843278 E-mail: bxj630306@163.com