

基于数据挖掘的兽药质量风险预测

田兴国¹, 陈江涛², 吕建秋¹

(1. 华南农业大学新农村发展研究院, 广东广州 510642) (2. 华南农业大学数学与信息学院, 广东广州 510642)

摘要: 当前, 兽药残留已成为食品安全的源头问题之一, 养殖户缺乏鉴别假兽药的能力, 兽药质量风险较大, 为了提高养殖户的辨别能力, 减少不合格兽药的使用, 降低兽药使用风险, 通过整理中国兽医药品监察所的抽检数据, 运用 SPSS Modeler 软件, 以 C5.0、Logistic、神经网络构建数据挖掘的分类预测模型, 对兽药质量进行分类预测。发现三种分类模型的整体分类精度偏低, 对此, 选用组合分类器对模型进行了优化, 并对神经网络、二元逻辑回归-神经网络及决策树-神经网络进行了比较, 发现从分类精度以及泛化性能上来讲, 决策树-神经网络的整体表现最好, 最后, 本文构建了决策树-神经网络的兽药质量风险预测模型, 并对之进行了进一步的优化, 预测准确率能达到 74.34%, 可为养殖户的购买决策提供参考。

关键词: 兽药; 数据挖掘; 组合分类; 食品安全

文章编号: 1673-9078(2017)11-212-218

DOI: 10.13982/j.mfst.1673-9078.2017.11.030

Quality-Risk Prediction of Veterinary Drugs by Data Mining

TIAN Xing-guo¹, CHEN Jiang-tao², LV Jian-qiu¹

(1. Institute of Technology Management and Planning, South China Agricultural University, Guangzhou 510642, China)

(2. College of Mathematics and Information, South China Agricultural University, Guangzhou 510642, China)

Abstract: Veterinary drug residues had become one of the source problems for food security at present. It was difficult for farmers to identify the fake veterinary drugs, which resulted in the risks of veterinary drugs quality. To improve the identification ability of the farmers and reduce the utilization of unqualified veterinary drugs, the data mining classification prediction model, established by C5.0, Logistic, neural network, was used to classify and predict the quality of veterinary drugs by sorting the sampling data of Chinese Veterinary Drug Administration based on SPSS Modeler software. Results showed that the classification accuracy of the three models was low, which resulted in optimizing the model by combination of classifier, and the neural network, binary logic regression - neural network, decision tree-neural network were compared. The overall performance of decision tree - neural network was the best in classification accuracy and generalization performance. Finally, the model for predicting the veterinary drugs quality in decision tree-neural network was established and further optimized, and the prediction accuracy reached 74.34%.

Key words: veterinary drug; data mining; combination classification; food security

兽药的使用对农业以及畜牧业的发展发挥了重要作用, 通常合理使用兽药能有效增强动物免疫力, 治疗动物疾病, 促进生长^[1]。但由于市场上兽药的质量参差不齐, 兽药生产厂商存在着由于经济利益驱动生产不合格兽药的动机, 导致兽药质量给人类带来的风险越来越大, 可以说当前大多数食品安全问题的源头在于兽药或者添加剂的不正规使用^[2], 这种不正规使用导致的兽药残留正成为食品安全事件的高发地带之一, 例如双汇“瘦肉精”事件、“红心鸭蛋”事件和多宝

收稿日期: 2017-06-14

基金项目: 国家重大农技推广专项 (2015GJZDNJTG); 科技部创新方法专项 (2015IM010400A4)

作者简介: 田兴国 (1976-), 男, 博士, 助理研究员, 研究方向: 为食品安全

通讯作者: 吕建秋 (1964-), 男, 研究员, 研究方向: 科技管理, 农技推广

鱼事件等^[3-5]。无一不是利益驱使下违规使用兽药造成的药物残留, 从而引发的严重食品安全事件。当然, 随着市场对食品安全的重视程度越来越高, 越来越多的研究也开始关注兽药质量风险的管控与检测, 但多是基于事后检测的, 例如 Kaufmann^[6]提出的基于高分辨率质谱以及串联质谱方法对兽药残留的检测, Schneider^[7]提出的一种高通量、多类、多类型的分析方法检测牛肉中的兽药残留等。事后控制虽然对于食品安全保障必不可少, 但对于生产出不合格产品的养殖户来讲, 他们可能也是不合格兽药的受害者, 不合格兽药也给他们造成了精神、身体以及经济上的损失。因而从养殖户的角度对兽药使用进行事前控制也是值得认真思考的方向。

数据挖掘技术在过去 20 年取得了飞速发展, 数据挖掘技术实现了将数据分析与计算机技术的有机结

合,增强了数据分析的可用性、可操作性以及可拓展性^[8]。目前,数据挖掘已经被广泛应用于各种行业之中,尤其在商业研究、决策支持、机器学习和人工智能等方面发挥了显著作用^[9~11]。数据挖掘在这些领域的巨大成功也促使了其向其他领域的拓展。例如 Bellazzi 等^[12]研究了数据挖掘技术在临床医疗上的应用原则以及方法,探讨了数据挖掘用于预测患者健康,辅助医疗决策的可能性。Algarni^[13]则将运用数据挖掘分析了在学生教育中,如何了利用学生数据,确定不同学生群体的优先学习需求,从而促进学生的发展。当然,随着物联网以及农业现代化的发展及需求,数据挖掘在农业领域的应用也越来越多,例如, Tripathy 等^[14]利用传感器获得的天气数据对病虫害进行预测。Ruß^[15]提出结合地理位置的空间数据对农业进行空间信息管理,来预测作物生长。也有研究开始将数据挖掘应用到食品安全领域,例如顾小林等人^[16]将关联规则算法应用到食品安全的信息预警之中。但总体上来讲,少有从事前控制的角度,将数据挖掘应用到兽药质量风险的评估之中。对此,本文从事前控制的角度入手,依据以往监测数据对兽药质量进行预测,以帮助养殖户进行购买决策。

1 数据来源

为满足全国兽药监督执法以及社会公众查询需要,2011年11月,中国兽医药品监察所正式启动国家兽药基础信息查询系统(以下简称查询系统)建设。截至2013年底系统包含11个数据库,约25.5万余条数据,数据来源于中国兽医药品监察所建立的国家兽药基础信息查询系统,登陆网址为:<http://sysjk.ivdc.gov.cn:8081/cx/>。

通过网页抓取工具共抓取2010~2013年化药监督抽检结果数据记录41883条,删除287条重复数据记录,有效数据记录41595条。鉴于论文研究目的是为养殖户购买兽药提供辅助决策,研究对象为猪用化学药物;因此选择化药监督抽检结果数据集中用药类别为“其他”数据记录40260条数据记录作为基础数据。

2 模型构建

建模使用的数据是事先在数据准备阶段筛选、生成好的“训练数据集”(“SJK”数据表)。目标变量为“训练数据集”(“SJK”数据表)中的“检测不合格”目标变量。输入变量为与兽药生产企业、被抽样单位、政府监管等相关影响因素的其他19个变量。

2.1 建模算法选择

兽药质量风险预测其本质就是预测养殖户购买兽药存在质量问题的倾向性。实际上就是预测兽药质量是否合格,应属于二元分类(响应)预测模型。对二元分类(响应)预测模型的而言,最常用的有神经网络、决策树、逻辑回归和多元线性回归4种算法。本文中兽药质量风险预测模型,输入和输出变量全部是分类变量。因此选择二元逻辑回归、决策树(C5.0)、神经网络3种模型建模,并对这些的预测结果进行评估,选择最优的模型进行部署。

2.2 建模数据流

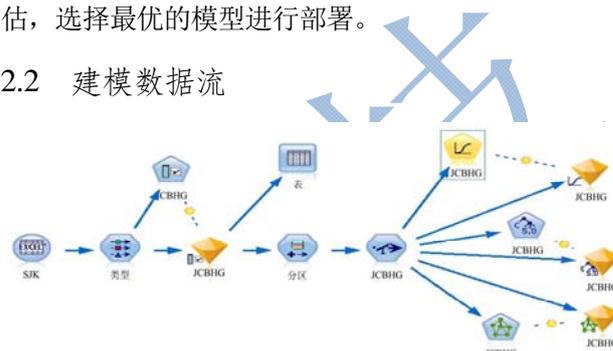


图1 兽药质量风险预测建模数据流图

Fig.1 Data flow diagram of quality-risk prediction modeling on veterinary drug

2.3 建模参数设置

(1)建模所用的变量中,除标称商品名、抽检单位与标称生产企业一致性、检测不合格字段为标志型字段,其他字段为分类型字段。检测不合格字段角色设为“目标”,其他字段角色设为“输入”。

(2)将数据集分成训练集和测试集,以提高模型的稳定性和延续性。本文兽药质量风险预测各模型建模分区(Distinct)节点设置训练分区的大小为70%,测试集分区大小为30%。

(3)数据表中兽药质量风险预测样本中不合格样本的比例为7%左右,相对于合格样本而言,兽药质量抽检不合格属于稀有事件。若采用一般的随机抽样,可能会造成对稀少事件的不易掌握。Linoff^[18]建议采用过抽样(Oversampling)或欠抽样(Undersampling)技术,增加样本中稀有事件的比率^[18],认为维持在10%~50%的比率中,通常可以得到不错的效果。SPSS Modeler Blance节点设置:合格:不合格=13:1。

3 模型评估

3.1 不同模型的误差分析

利用SPSS Modeler中输出(Output)选项卡中的分析(Analysis)节点可以对各种模型的误差进行比较。分别运行二元逻辑回归、决策树C5.0、神经网络模型

算法，并依次连接前面兽药质量检测不合格建模过程中3个模型，添加分析(Analysis)节点到质量检测不合格预测建模数据流编辑器，并连接到最后的神经网络模型。模型参数设置分析选项勾选重合矩阵和置信度图。运行分析节点，可得到如下结果。

表1 三种分类方法的一致性检测

Table 1 Consistency detection of three classification methods

分区	训练样本		测试样本	
一致	2534	74.75%	7904	75.35%
不一致	856	25.25%	2586	24.65%
总计	3390		10490	

表2 分类一致样本中的分类结果

Table 2 Classification results of the classification in consistent samples

分区	训练样本		测试样本	
正确	2064	81.45%	6374	80.64%
错误	470	18.55%	1530	19.36%
总计	2534		7904	

表3 一致性重合矩阵

Table 3 Coincidence matrix of consistency

数据样本	预测合格	预测不合格
训练样本实际合格	1040	238
训练样本实际不合格	232	1024
测试样本实际合格	5945	1404
测试样本实际不合格	126	429

表4 模型的一致性置信度报告

Table 4 Report of model consistency confidence

指标	训练样本	测试样本
范围	0.545~0.963	0.535~0.965
平均正确性	0.784	0.784
平均不正确性	0.729	0.744
正确性始终高于	0.944 (观测值的 1.22%)	0.953 (观测值的 1.06%)
不正确性始终低于	0.545 (观测值的 0%)	0.546 (观测值的 0%)
90.06%以上的准确性	0.775	
90.02%以上的准确性		0.86
2.0 以上的折叠正确性	0.907 (观测值的 80.89%)	0.904 (观测值的 86.96%)

综上所述，对兽药质量检测不合格预测而言，3个模型在测试集中整体预测精度不够非常理想，仅为80.64%；其原因可能是所选样本集中不合格样本所在的比例较少，模型训练时对训练样本进行平衡处理所致。同时，预测置信度均值在正确预测和错误预测的样本上差异较小，即高置信度并非能够得到高的预测正确率，低的置信度并非预测的错误率就高。测试集中，置信度大于0.86的样本，有90.02%的样本预测正确。

由表1可知，在测试集10490个样本中，测试集中3个模型预测结果相同的比例较高；3个模型对7904个样本有相同的预测值，占75.35%；2586个样本预测结果不同，占24.65%。

从表2中可以看出在3个模型给出相同的预测结果的7904个样本中，有6374个样本预测正确，占80.4%；1530个样本预测错误，占19.36%，模型总体预测精度一般。

另外，表3可以看出3个模型给出的相同预测结果7904个样本中，实际合格与不合格且模型预测正确的分别是5945和429个，3个模型在预测合格的正确率高于不合格，其原因可能是所选样本集中不合格样本所在的比例较少，模型训练时第训练样本进行平衡处理所致。

表4的一致性的置信度报告显示，测试集中样本的预测置信度最小值为0.53，最大值为0.963。对正确预测的样本，其预测置信度的平均值为0.784 (Mean Correct)，对错误预测的样本，其预测置信度的平均值为0.744 (Mean Incorrect)。预测置信度在0.953之上的样本其预测通常是正确的，因而只有观测值的1.06%的样本符合此条件。预测置信度在0.546之下的样本其预测通常是错误的，因而没有样本符合此条件。置信度大于0.86的样本中，有90.02%的样本预测正确。置信度大于0.904的样本(包含86.96%的样本)，预测的正确率相比全部样本改善20%。

3.2 不同模型的收益对比

利用SPSS Modeler中图形(Graphs)选项卡中的评估(Evaluation)节点可以对各种模型进行统一评估。添加评估(Evaluation)节点到兽药质量检测不合格预测建模数据流编辑器，并连接到最后神经网络模型。在图选项图标类型分别选择增益和提升两个选项，勾选累积散点图、包含基线和包含最佳线。运行评估节点，输出评估结果如图2所示。

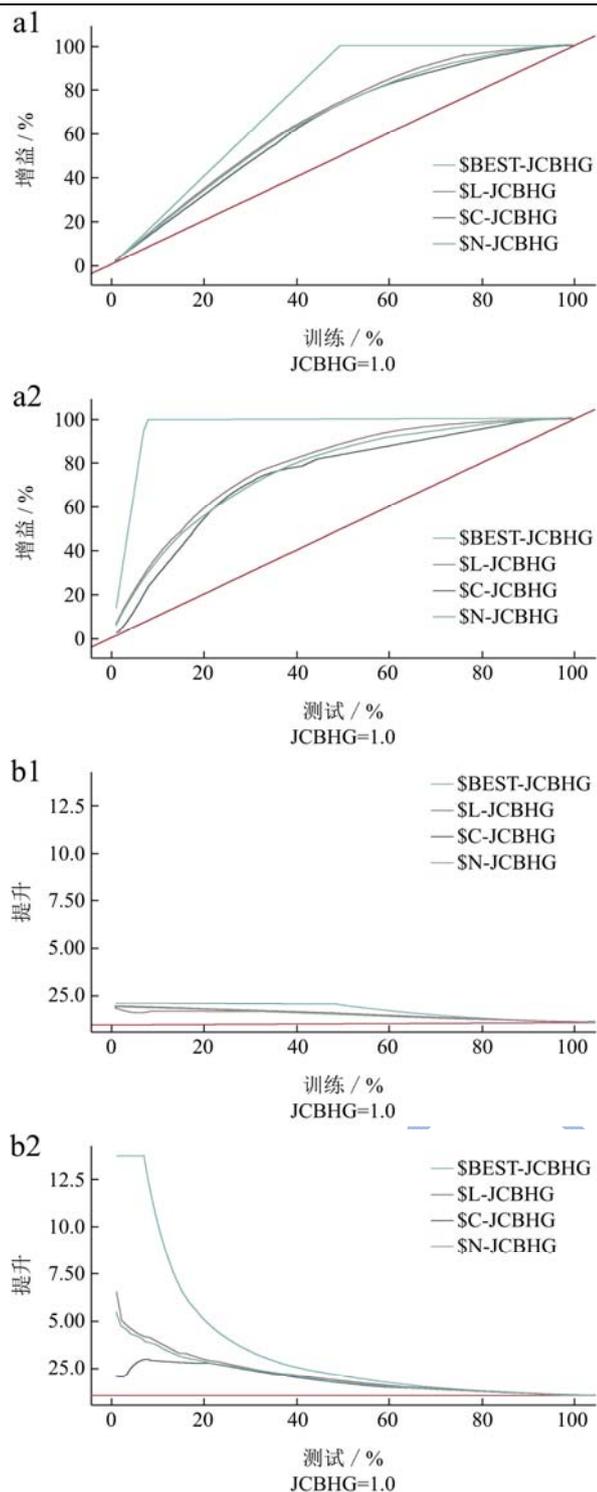


图2 模型的增益和提升图

Fig.2 Graph of gain and ascension of the model

由图 2 可知, 对兽药质量检测不合格预测中, 3 个模型的总体性能有所差异, 以二元逻辑回归和神经网络模型表现较好, 二元逻辑回归模型表现稍好, 决策树 C5.0 模型表现最差。测试样本中, 二元逻辑回归模型预测前 20 个百分位预测出不合格样本占总的不合格样本的 58.56%; 前 1 个百分位不合格预测的性能是随机抽样预测的 6.4 倍。3 个模型与最优模型都有一

定的差距, 还有进一步改善的空间。

综合来看, 兽药质量风险预测模型的预测精度不够理想, 预测置信度均值在正确预测和错误预测的样本上差异不大, 模型还有提升的空间, 需进一步优化。

4 模型优化

4.1 优化思路

鉴于兽药质量风险预测 3 个模型中神经网络和逻辑回归模型性能表现较好, 且两者性能表现相差不多; 决策树性能表现较差。因此采用两种变量筛选方式都去尝试比较, 构建决策树-神经网络和逻辑回归-神经网络兽药质量风险预测组合模型。具体方法是先将决策树或二元逻辑回归模型用于筛选兽药质量风险预测 3 个模型的输入变量, 减少非核心变量对兽药质量风险分类预测结果的影响, 然后用决策树或逻辑回归选择出重要性排序靠前的相关 (关联系数大于或等于 0.05) 变量作为神经网络 (BP) 模型的输入层, 最终输出组合模型的分类型预测结果。这样的两阶段组合模型将决策树或逻辑回归模型与神经网络模型相结合, 旨在筛选出对分类结果有重要影响的要素指标, 增强神经网络模型的可解释性, 同时通过去除部分非重要指标对神经网络模型的影响。

4.2 筛选有效输入变量

打开前面构建的兽药含量不合格预测决策树和二元逻辑回归模型, 可得到决策树 C5.0 和二元逻辑回归模型选择预测变量重要性对比如表 5 所示。

表 5 决策树 C5.0 选择变量与二元逻辑回归选择变量对比

Table 5 Comparison of selected variables of C5.0 decision tree and binary logistic regression

逻辑回归选择变量		决策树选择变量	
变量名称	重要性	变量名称	重要性
QYDQYZX	0.2072	QYCJPL	0.6195
QYCJPL	0.1761	QYDQYZX	0.2482
QYBHGL	0.1516	QYBHGL	0.0645
QYSZDQ	0.1065	BCSPM	0.0462
BCYDWXZ	0.1058	CJDWSF	0.0079
BCSPM	0.0952	QYSZDQ	0.0064
YT	0.0789	JX	0.0058
JX	0.0489	CJJD	0.0014
SJJG	0.0214		
CJJD	0.0042		
CFY	0.0034		
GMPSJG	0.0009		

由表 5 可知, 二元逻辑回归模型预测变量重要性大于或等于 0.01 的共有 9 个变量, 其预测变量重要性总值为 0.9916; 决策树 C5.0 模型预测变量重要性大于或等于 0.01 的共有 4 个变量, 其预测变量重要性总值为 0.9784。在此分别选取二元逻辑回归模型筛选的 9 个变量和决策树模型筛选的 4 个变量作为第二部分神经网络模型输入层的属性指标。

4.3 优化后分类精度比较

表 6 单一与组合模型的预测精度比较

Table 6 Comparison of predict accuracy of single model and combined model

	单一神经网络/%	二元逻辑回归-神经网络/%	决策树-神经网络/%
训练集 假正率	29.83	30.06	30.92
训练集 假负率	22.81	25.79	27.08
训练集 总正确率	73.58	72.02	70.97
测试集 假正率	29.59	30.40	30.71
测试集 假负率	26.75	28.50	29.69
测试集 总正确率	70.56	69.72	69.36

由表 6 可知, 单一的神经网络模型、二元逻辑回归-神经网络组合模型、决策树-神经网络模型组合模型总体的对测试样本的预测精确度和假负率都相差不多, 单一神经网络模型稍好, 二元逻辑回归-神经网络组合模型次之。

4.4 优化后模型的稳健性比较

模型的稳健性也可以理解为模型的泛化性或推广性, 即训练完毕的模型用于其他样本数据, 模型的准确度也不会发生过大的波动。泛化性即经训练样本训练后的模型用其它检验样本进行检验时, 精确度的波动性。如果精确度的波动较小, 说明这个模型经训练样本训练后对其它样本适用性比较强, 选用样本的标准差作为稳健性的衡量指标:

$$\sigma = \sqrt{\sum_{i=1}^2 (X_i - \bar{X})^2} \quad (1)$$

其中, $X_i(i=1, 2)$ 为训练样本和测试样本的精度。

表 7 单一模型与组合模型稳健性对比

Table 7 Comparison of robustness of single model and combined model

模型	假正率	假负率	分类精度下降
神经网络	0.1646	2.7804	2.1362
二元逻辑回归-神经网络	0.2338	1.9162	1.6256
决策树-神经网络	0.1427	1.8457	1.1391

兽药质量检测不合格单一模型与组合模型对样本稳健性对比如表 7 所示。

由表 7 可知, 决策树-神经网络组合模型的总分类精度标准差最低, 仅为 1.1391, 仅为单一神经网络模型的 53.33%, 二元逻辑回归-神经网络组合模型的 70.07%; 决策树-神经网络组合模型的假负率标准差最低, 仅为 1.8457, 为单一神经网络模型的 66.38%, 二元逻辑回归-神经网络组合模型的 96.32%; 决策树-神经网络组合模型总分类精度和假负率标准差远低于单一神经网络模型和二元逻辑回归-神经网络组合模型。说明决策树-神经网络组合模型稳健性远远优于单一神经网络模型和二元逻辑回归-神经网络组合模型。

4.5 决策树-BP 神经网络组合模型参数优化

本文所选最优模型是决策树-BP 神经网络组合模型, 其在训练集上分类预测精确度稍微差一点。可能的原因之一前面构建的神经网络模型参数设置采用的是默认方式, 网络结构过于简单, 分类预测精度不高;

由于任意一个三层 BP 神经网络只要具备一定的中间层神经元数, 就一定可以模拟出任何非线性结果, 中间层神经元数的确定对于模型的映射效果影响较大, 对于中间层神经元数的设计, 虽然目前还不存在一个最权威、最标准的确定方法, 但是每一个 BP 神经网络都存在一个最优的中间神经元数, 它的设计往往需要根据设计者的经验和多次实验来确定。本文数据挖掘选择的采取是经验计算公式来确定神经网络隐含层数量一个合理的范围。

$$n_1 = \sqrt{n + m} + a \quad (2)$$

其中, n_1 为隐含层节点数, n 为输入节点数, m 为输出节点数, a 为介于 1~10 的常数。

表 8 最优决策树-神经网络模型参数设置

Table 8 Optimal parameters settings of the decision tree-neural network combination model

参数名	参数值
模型	神经网络多层感知器
结构	4-7-1
激活函数	sigmoid
训练方法	增强模型准确性(Boosting)
使用最大训练时间 (每个组件模型)	15 分
用于 Boosting 组件模型数量	10
过度拟合防止集合/%	30

利用 SPSS Modeler 软件选择“增强模型准确性(boosting)”方法产生一系列“成分模型”, 其中各个模型

都是在整个数据集上构建。在构建下一个成分模型之前，将根据前一个成分模型的残差对记录进行加权。具有较大残差的个案将被给予较高的分析权重，因此下一个成分模型将较好地侧重于这些记录。这些成分模型共同构成一个整体模型。该整体模型采用组合规则对新记录进行评分，可用的规则取决于目标的测量级别。此方法采用推进方式构建整体模型，这将生成一系列模型以获得更精确的预测结果；但与标准模型相比，此整体模型需要更长的构建与评分时间，模型的参数设置如表 8 所示。

表 9 参数优化前后决策树-神经网络组合模型预测精度比较

Table 9 Comparison of predict accuracy of the decision tree-neural network combination model before and after the optimal parameters

	决策树-神经网络 (前) /%	决策树-神经网络 (后) /%
训练集	假正率	30.92
	假负率	27.08
	总正确率	70.97
测试集	假正率	30.71
	假负率	29.69
	总正确率	69.36
		72

经过增加模型隐含层数量，选择“增强模型准确性(推进)”方法进行训练，所得决策树-神经网络组合模型预测精度与优化前模型预测精度比较如表 9 所示。

由表 9 可知，决策树-神经网络组合模型经过参数优化后，模型在训练集上的总正确率为 71.33%，比优化前提高；在测试集上的总正确率为 72%，比优化前提高 2.64%；且在测试集上的模型正确率比训练集高。说明经过参数优化模型分类预测的精度提高，模型性能表现有所提高。

5 模型验证

以 2014 年一季度的国家兽药基础信息查询系统公布的最新数据验证模型的稳定性。从国家兽药基础信息查询系统中抽出 4081 条有效数据；经过数据清洗、数据整理和数据变换等，最后用于验证的数据为 3975 条记录，存入“SJK2014”数据表。

表 10 验证集上分类预测结果表

Table 10 Classification predict results in the validation set

观测	预测		预测正确率 /%
	0	1	
0	2772	943	74.62
1	77	183	70.38
预测命中率/%	97.30	16.25	74.34

由表 10 可知，模型在验证集上预测正确率为 74.34%，模型预测兽药检测不合格样本数为 183 个，占全部兽药不合格样本的 70.38%；占全部预测为兽药不合格样本的 16.25%，比随机抽样预测 (Random Rate) 的 6.54% 高近 10 个百分点。因此模型分类预测精确性比较理想。

响应率是指经过模型打分，按照概率分数排序后的某区间或累计区间的观察对象中，属于正 (1) 的观察对象占该区间或该累计区间总体观察对象数量的百分比。响应率是指经概率分数排序后，某区间或累计区间实际兽药不合格样本占全部兽药样本的百分比。响应率越大，说明在该区间或该累计区间模型的预测准确度越高，模型分类效果越好。模型在验证集上响应率曲线如图 3 所示。

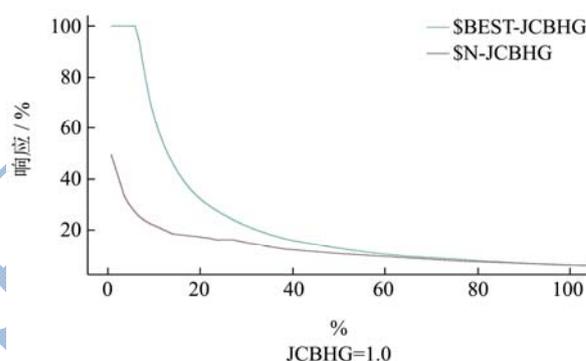


图 3 模型在验证集上响应率曲线图

Fig.3 Response rate Curve of model in the validation set

由图 3 可知，经决策树-神经网络模型得到的分数最高的前 5% 的兽药样本中，检测不合格兽药样本高达 30 左右，而兽药样本的整体不合格率在 6.54%；得分最高的前 20% 的兽药样本中，检测不合格兽药样本在 18 左右；得分最高的前 30% 的兽药样本中，检测不合格兽药样本在 16 左右；远远高于兽药样本的整体不合格率 (6.54%)。这表明模型的稳定性较好可以用于模型后续的开发应用。

综上所述，从模型分类预测准确率和响应率来看，模型在验证集上的表现还是比较理想。最终确定决策树-神经网络组合模型为最优预测模型，可以应用次模型到后续的兽药质量风险预测辅助决策系统开发中。

6 结论

以数据挖掘理论为基础，选用 C5.0、Logistic 和神经网络算法作为初始算法构建了兽药质量风险的预测模型，发现单一分类模型有改进的可能。最后，从组合分类的角度入手，构建了决策树-神经网络的兽药质量风险预测模型，并与单一神经网络、Logistic-神

经网络的组合模型进行比较发现,决策树-神经网络的分类精度与其他分类算法相差不大,但泛化性能即模型的稳健性方面要明显好于其他模型,最后本文以决策树-神经网络模型为基础,经过参数优化,构建了决策树-神经网络模型的兽药质量风险预测模型,总体精度达到了74.34%,可为养殖户的兽药选择提供参考。

参考文献

- [1] Beyene T, Tesega B. Rational veterinary drug use: its significance in public health [J]. *Journal of Veterinary Medicine & Animal Health*, 2014, 6(7)
- [2] Stolker A A M, Zuidema T. Residue analysis of veterinary drugs and growth-promoting agents [J]. *Trac Trends in Analytical Chemistry*, 2007, 26(10): 967-979
- [3] 何计国.从“瘦肉精”事件看国内食品安全问题[J].*中国猪业*,2011,4:4-7
HE Ji-guo. Discussion on domestic food safety from "Clenbuterol" event [J]. *China Swine Industry*, 2011, 4: 4-7
- [4] 刘震.苏丹红借红心鸭蛋“还魂”餐桌[N].*大众科技报*, 2006-11-19(A02)
LIU Zhen. Sudan red 'resurrection' to table by red duck [N]. *Popular Technology News*, 2006-11-19(A02)
- [5] 叶佳林.“多宝鱼事件”的思考[J].*中国水产*,2006,373(12): 14-16
YE Jia-lin. Reflections on the event of 'Treasure fish' [J]. *Chinese Aquatic Products*, 2006, 373(12): 14-16
- [6] Kaufmann A, Butcher P, Maden K, et al. Reliability of veterinary drug residue confirmation: high resolution mass spectrometry versus tandem mass spectrometry [J]. *Analytica Chimica Acta*, 2015, 856(856): 54
- [7] Schneider M J, Lehotay S J, Lightfield A R. Validation of a streamlined multiclass, multiresidue method for determination of veterinary drug residues in bovine muscle by liquid chromatography-tandem mass spectrometry [J]. *Analytical and Bioanalytical Chemistry*, 2015, 407(15): 4423-4435
- [8] Aggarwal C C. *Data Mining* [M]. Springer International Publishing, 2015
- [9] Hartl K, Jacob O. Determining the business value of business intelligence with data mining methods [J]. *Data Analytics*, 2015
- [10] Bawden D. Data mining and decision support: integration and collaboration [J]. *Journal of Documentation*, 2005, 61(3): 443-445
- [11] Dean J. *Big data, data mining, and machine learning: value creation for business leaders and practitioners* [M]. John Wiley & Sons, 2014
- [12] Bellazzi R, Ferrazzi F, Sacchi L. Predictive data mining in clinical medicine: a focus on selected methods and applications [J]. *Wiley Interdisciplinary Reviews Data Mining & Knowledge Discovery*, 2011, 1(5): 416-430
- [13] Algarni A. Data mining in education [J]. *International Journal of Advanced Computer Science & Applications*, 2016, 7(6)
- [14] Tripathy A K, Adinarayana J, Sudharsan D, et al. Data mining and wireless sensor network for agriculture pest/disease predictions [C]// *Information and Communication Technologies. IEEE*, 2011: 1229-1234
- [15] Ruß G, Brenning A. Data Mining in Precision Agriculture: Management of Spatial Information [C]// *Computational Intelligence for Knowledge-Based Systems Design, International Conference on Information Processing and Management of Uncertainty, Ipmu 2010, Dortmund, Germany, June 28 - July 2, 2010. Proceedings. DBLP*, 2010: 350-359
- [16] 顾小林,张大为,张可,等.基于关联规则挖掘的食品安全信息预警模型[J].*软科学*,2011,25(11):136-141
GU Xiao-lin, ZHANG Da-wei, ZHANG Ke, et al. Food safety information early warning model based on association rules mining [J]. *Soft Science* 2011, 25(11): 136-141
- [17] 张洁,崔上元.基层兽药市场存在的问题及对策[J].*养殖与饲料*,2011,2:62-63
ZHANG Jie, CUI Shang-yuan. Problems and counter-measures of traditional veterinary drug market [J]. *Breeding and Feed*, 2011, 2: 62-63
- [18] Gordon S Linoff, Michael J A Berry. *数据挖掘技术*[M].北京:清华大学出版社,2013
Gordon S Linoff, Michael J A Berry. *Data mining technology* [M]. Beijing: Tsinghua University Press, 2013