

质量分析

机器学习相关技术在以黄酮为特征的黄芪分类中的应用研究

石岩¹, 李宁², 魏锋^{1*}, 马双成¹

(1. 中国食品药品检定研究院, 北京 102629; 2. 北京市药品检验研究院, 北京 102206)

摘要 **目的:** 建立以黄酮类成分为特征的栽培黄芪、半野生黄芪和野生黄芪的三分类模型, 并且对自动机器学习技术和数据增强技术在药物分析领域中的应用进行探索和评价。**方法:** 首先, 对黄芪的黄酮类成分含量数据进行相关性分析、主成分分析, 建立决策树和逻辑回归模型, 根据模型分析黄酮类成分的重要性程度; 然后, 使用 TVAE 表格数据生成算法, 根据真实数据生成 600 批虚拟数据, 使用自动学习框架 AutoGluon, num_bag_folds 设为 5, 分别对 64 批真实数据和 600 批虚拟数据进行学习, 得到 2 组共 30 个模型, 依据准确率进行评估。**结果:** 对机器学习模型的分析可知, 芒柄花素、毛蕊异黄酮葡萄糖苷和刺芒柄花苷这 3 种黄酮类成分对于黄芪质量, 尤其是来源等级的控制具有重要意义; 2 组共 30 个模型预测准确率表明, 基于 NeuralNet 的模型和基于树模型的机器学习算法对于黄酮成分数据表征的黄芪而言分类效果最好; 数据增强技术生成的虚拟数据与真实数据在所训练得到的模型准确率趋势方面基本一致。**结论:** 机器学习相关技术在以黄酮为特征的黄芪分类中具有较好的应用价值。

关键词: 黄芪; 黄酮; 毛蕊异黄酮葡萄糖苷; 刺芒柄花苷; 毛蕊异黄酮; 山柰酚; 异鼠李素; 芒柄花素; 机器学习; 人工智能; 数据增强

中图分类号: R 917 文献标识码: A 文章编号: 0254 - 1793(2024)05 - 0866 - 08

doi: 10. 16155/j. 0254 - 1793. 2024. 05. 15

Research on the application of machine learning related techniques in the classification of Astragali Radix characterized by flavonoids

SHI Yan¹, LI Ning², WEI Feng^{1*}, MA Shuang - cheng¹

(1. National Institutes for Food and Drug Control, Beijing 102629, China;

2. Beijing Institute for Drug Control, Beijing 102206, China)

Abstract Objective: To establish a three classification model for cultivated, semi - wild, and wild Astragali Radix characterized by flavonoids, and explore and evaluate the application of techniques of automated machine learning and data augmentation in the field of drug analysis. **Methods:** Firstly, correlation analysis and principal component analysis were conducted on the flavonoid content data of Astragali Radix, and models of decision tree and logistic regression were established to analyze the importance of flavonoid components based on the models. Then, using the AutoGluon framework with 5 as num_bag_folds, 2 sets of 30 models respectively through 64 batches of real data and 600 batches of virtual data generated based on real data with the TVAE table data generation algorithm for training were obtained, and these models were evaluated by accuracy. **Results:** The analysis

* 通信作者 Tel: (010) 53852020; E - mail: weifeng@nifdc.org.cn

第一作者 石岩 Tel: (010) 53852081; E - mail: shiyan@nifdc.org.cn

李宁 Tel: 13811671528; E - mail: 642781540@qq.com

of machine learning models, indicated that formononetin, campanulin and onospin played the important roles in the quality control of Astragali Radix, especially for the source grade control. The accuracy of model prediction showed that the models based on Neural Net and tree - model always had the best classification effect for Astragali Radix. The virtual data generated by data augmentation technique is basically consistent with the actual data in terms of the accuracy trend of the model training process. **Conclusion:** Related techniques of machine learning have good application value in the classification of Astragali Radix characterized by flavonoids.

Keywords: Astragali Radix; flavonoids; campanulin; onospin; calycosin; kaempferol; isorhamnetin; formononetin; machine learning; artificial intelligence; data augmentation

黄芪为豆科植物蒙古黄芪 *Astragalus membranaceus* (Fisch.) Bge. var. *mongholicus* (Bge.) Hsiao 和膜荚黄芪 *Astragalus membranaceus* (Fisch.) Bge. 的干燥根,是我国传统的常用中药材和饮片,具有补气升阳、固表止汗、利水消肿等功效,在气虚乏力、食少便溏等症的临床治疗中多有应用^[1]。黄芪中的主要活性成分可以分为黄酮类、皂苷类及多糖类等^[2-4],黄酮类成分主要包括毛蕊异黄酮、山柰酚、芒柄花苷、芒柄花素等。前期对黄酮类成分的研究中^[5],对所测得的野生、半野生以及栽培品的黄芪中的主要黄酮类成分的含量分布,总结出一些规律,但是尚无法据此对主要黄酮类成分做到准确分类。目前市场上的黄芪多以栽培品为主,野生或半野生黄芪资源较少,价格相对也较昂贵,尤其是野生黄芪的资源几近枯竭,更是价高而难得^[6-8]。为了规范黄芪这一常用中药材的市场秩序,保护野生黄芪资源,有必要对不同来源的黄芪进行等级化研究,建立能够明确区分不同来源黄芪的数据模型,客观、准确地评判黄芪的质量。前期研究^[5]表明,黄芪中黄酮类成分的含量分布具有较多的重叠情况,仅靠简单的 1 个或几个成分含量的关系难以做到准确区分。为了探究野生和栽培对黄芪药材质量的影响,本研究根据黄芪所含特征黄酮类成分与含量,采用自动机器学习 (automated machine learning, AutoML) 与数据增强 (data augmentation) 技术的联合建模方法,对不同来源的黄芪药材的区分展开可行性研究。AutoML 是将机器学习与现实问题结合应用的端到端自动化的过程,可以帮助数据科学家、分析师和开发者高效、高质量地构建机器学习模型。数据增强技术是通过原始数据进行分布模拟合成,生成更多样本来扩充数据集的方法,可以提高模型的泛化能力和稳定性。本研究采用的 AutoML 框架为亚马逊公司的开源 Au-

toGluon, 而数据增强技术则采用了 4 种主流的生成表格数据的技术:高斯 Copula (Gaussian Copula)、条件表格生成对抗网络 (conditional tabular generative adversarial network, CTGAN)、Copula 生成对抗网络 (Copula generative adversarial network, CopulaGAN)、基于三元组的变分自动编码器 (triplet - based variational autoencoder, TVAE)。目前,生成表格数据的技术多是基于对抗神经网络 (GAN) 和变分自编码器 (VAE) 而实现的,区别主要在于数据转换及生成策略。本研究基于前期研究获得的真实实验数据,根据数据特点 (样本量少且数据分布缺乏均衡),采用数据增强获得足够数量虚拟数据,通过 4 种生成数据模型的评估得分,选定其中的 TVAE 作为生成虚拟表格数据模型,通过虚拟表格数据投喂训练 AutoML 模型,根据真实数据预测验证结果,优选 AutoGluon 框架的模型 bag 及堆叠 (stack) 形式。

对于药物分析领域内的机器学习相关应用而言,可将其流程总结如图 1 所示。本着“大胆假设,小心求证”准则,本研究在使用基本机器学习算法对黄芪进行基于黄酮类成分的分类预测的同时,初步揭示了不同来源黄芪样品中具有特征性的黄酮类成分。此外,还引入了目前机器学习领域中的相关热点技术,即 AutoML 技术和数据增强技术,对黄芪的数据进行了系列的应用研究,评估了 AutoML 各模型、二级融合模型及数据增强技术的应用效果。

1 数据与语言

本研究是前期黄芪中黄酮类成分测定研究工作的后续^[5],所使用的不同来源的黄芪样品以及测定数据均来自前期研究。黄芪样品即 1~2 年生人工栽培黄芪 30 批 (A1~A30), 5 年生半野生黄芪 24 批 (YX1~YX24), 5~6 年生野生黄芪 10 批 (HY1~HY10); 3 个类别黄芪 (人工栽培黄芪、半野生黄芪

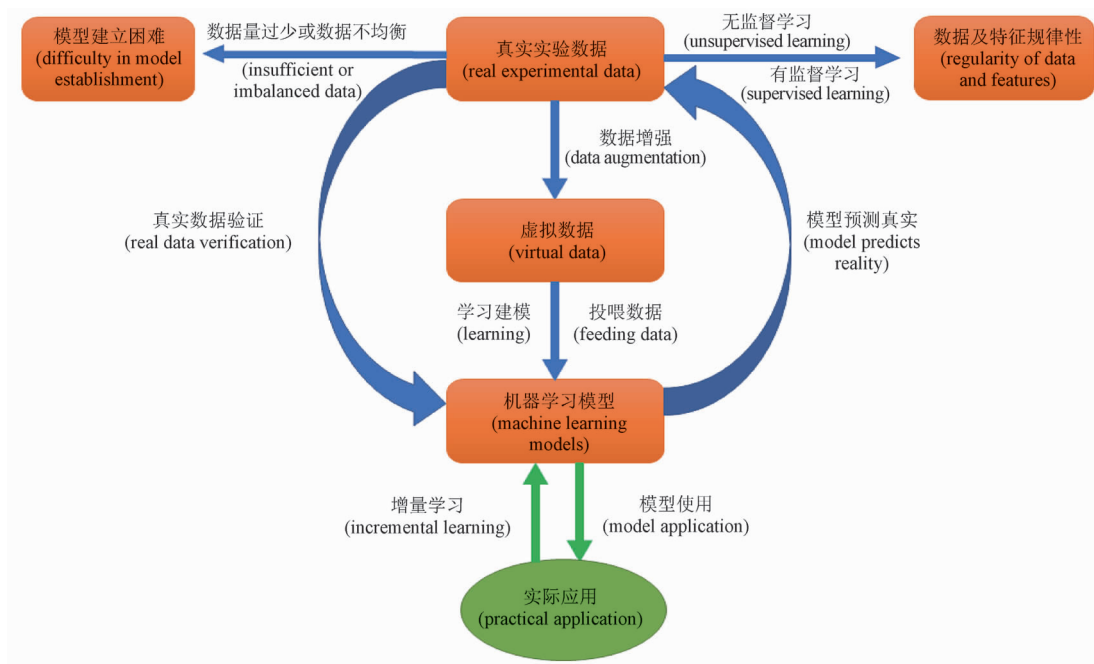


图 1 研究流程示意图

Fig. 1 Research flow diagram

和野生黄芪)分别以 ZP、BY 和 YS 表示。测定数据即以上 64 批黄芪样品中毛蕊异黄酮葡萄糖苷、刺芒柄花苷、毛蕊异黄酮、山柰酚、异鼠李素和芒柄花素的含量,连同六者的含量总和,共 7 组数据作为输入特征。

本研究中涉及的数据处理、AutoML 及数据增强等技术,均在 Python 计算机编程语言(美国 Python

Software Foundation, version:3.8.8)中编码和运行。

2 方法与结果

2.1 真实数据相关性分析及主成分分析

将前期研究^[5]得到的 64 批黄芪样品的特征变量进行 Pearson 相关性分析,绘制各成分数据之间相关性热图,见图 2。

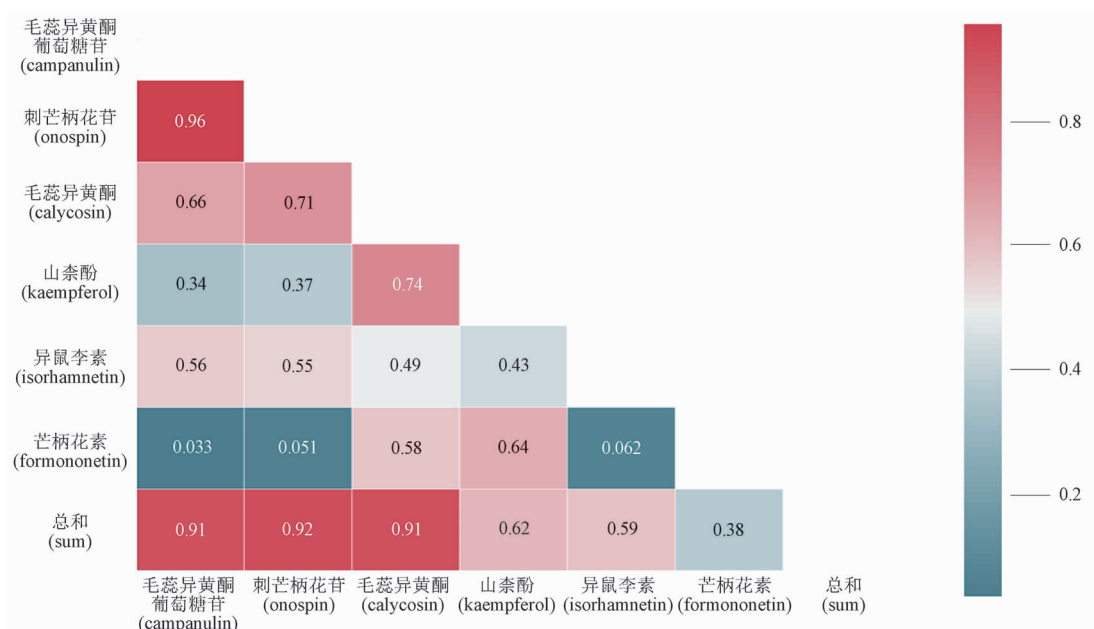


图 2 Pearson 相关性分析热度图

Fig. 2 Pearson correlation analysis heat map

对数据进行标准化预处理,然后进行主成分分析,以累计方差达 95% 以上为标准选取主成分。实际选

取了 4 个主成分,累计方差为 97.6%。64 批样品在这 4 个主成分之间形成的二维空间的得分分布情况见图 3。

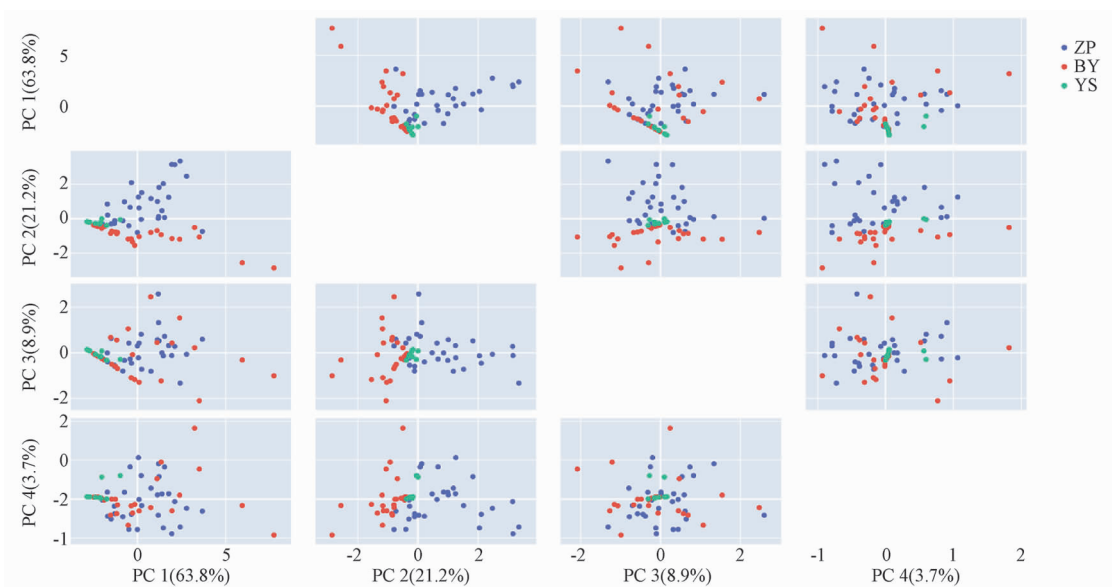


图 3 主成分分析得分分布图

Fig. 3 Score plot of principle component analysis

2.2 基于决策树算法的黄酮类成分特征性研究

决策树以信息熵表征不纯度和计算信息增益,选择最大深度为 3,绘制决策树树状图,见图 4。其

中,芒柄花素、毛蕊异黄酮葡萄糖苷和刺芒柄花苷这 3 个成分分别分布在根节点及近根节点,表明其在分类中的重要性。

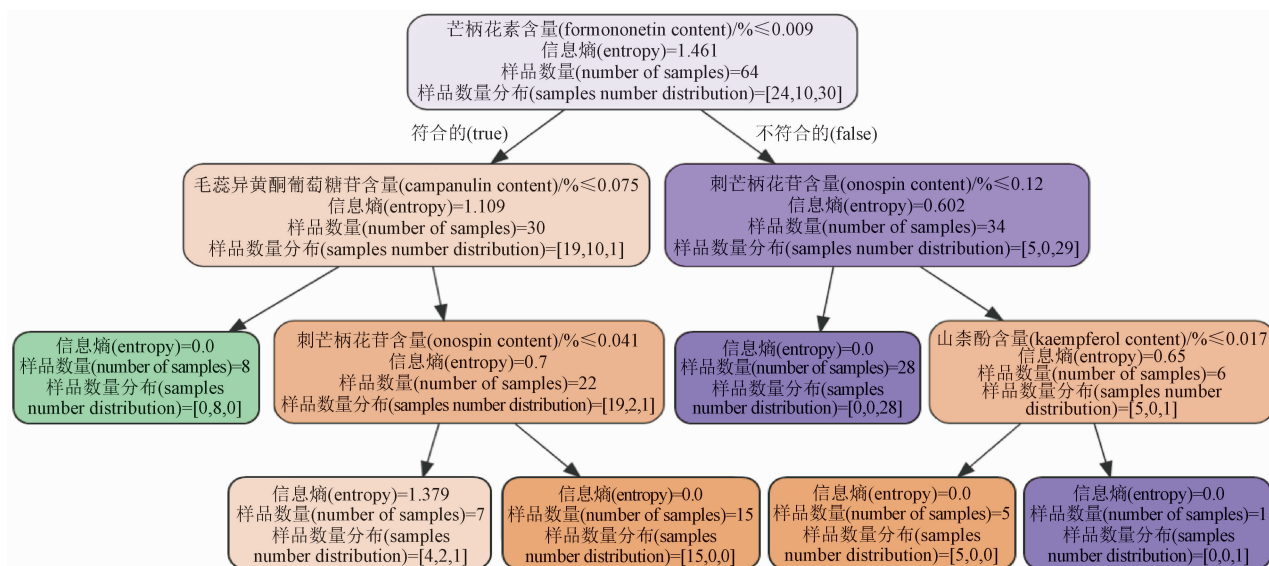


图 4 决策树示意图

Fig. 4 Decision tree diagram

2.3 基于真实数据的逻辑回归模型

对数据进行标准化处理,然后将变量数据及分类信息输入逻辑回归分类模型,根据多分类的实际

情况选择 Softmax 回归,惩罚项参数选择岭回归 (l2),优化算法选择 lbfgs。由于 YS 类别中样品数量较少,且该模型并非最终目标模型,故选择 5 折交叉

验证法实现模型的粗略验证评价。

2.4 基于真实数据的数据增强和虚拟数据生成

使用 TVAE, 分别对不同来源黄芪样品的黄酮类成分含量及其含量总和进行数据的深度学习, epochs 为 400, batch_size 为 10, 每个来源的黄芪样品模拟生成 200 批数据, 共模拟生成 600 批数据。

2.5 基于 AutoGluon 框架的多模型训练及基于真实数据的模型评价

使用 python 语言环境下的 AutoGluon 自动机器

学习框架, 对 64 批真实的黄芪数据以及模拟生成的 600 批数据分别进行学习, 得到 2 组模型; 交换数据, 分别使用模拟生成的 600 批数据和 64 批真实的黄芪数据作为测试集对模型进行验证, 即真实数据和模拟生成的数据各自进行对方数据训练得到的模型的外部验证。自动学习框架 AutoGluon 选择不使用多层堆叠模式, num_bag_folds 选择 5。框架内各学习算法模型的测试集验证准确率和训练集交叉验证准确率结果见表 1。

表 1 模型预测准确率

Tab. 1 Model prediction accuracy

序号 (code)	模型 (model)	真实数据建模准确率 (accuracy based on true data)		模拟数据建模准确率 (accuracy based on analog data)	
		测试集 (test set)	交叉验证 (cross validation)	测试集 (test set)	交叉验证 (cross validation)
1	NeuralNetTorch_BAG_L1	0.968	0.844	0.875	0.993
2	ExtraTreesEntr_BAG_L1	0.967	0.891	0.891	0.993
3	ExtraTreesGini_BAG_L1	0.965	0.875	0.891	0.992
4	NeuralNetFastAI_BAG_L1	0.963	0.953	0.813	0.992
5	WeightedEnsemble_L2	0.963	0.953	0.891	0.997
6	LightGBM_BAG_L1	0.945	0.891	0.828	0.992
7	CatBoost_BAG_L1	0.943	0.875	0.844	0.992
8	RandomForestGini_BAG_L1	0.942	0.906	0.844	0.997
9	RandomForestEntr_BAG_L1	0.938	0.891	0.844	0.995
10	XGBoost_BAG_L1	0.913	0.859	0.828	0.988
11	LightGBMXT_BAG_L1	0.910	0.859	0.891	0.997
12	LightGBMLarge_BAG_L1	0.873	0.859	0.813	0.992
13	KNeighborsDist_BAG_L1	0.845	0.781	0.828	0.973
14	KNeighborsUnif_BAG_L1	0.840	0.688	0.828	0.973

3 讨论

3.1 真实数据相关性分析解析

图 2 中方格代表横纵坐标变量之间的 Pearson 相关性, 方格颜色越红, 代表 2 个变量之间相关性数值越大, 即相关性越强。在排除含量总和情况下, 毛蕊异黄酮葡萄糖苷与刺芒柄花苷属于极强相关, 二者相关系数为 0.96; 而芒柄花素与毛蕊异黄酮葡萄糖苷、刺芒柄花苷及异鼠李素三者之间属于极弱相关, 相关系数分别为 0.033、0.051 和 0.062。该项研究结果初步显示芒柄花素属于比较特殊的成分。

3.2 真实数据主成分分析解析

为避免数据各变量差异对算法产生不必要的影响, 在进行主成分分析及后续的逻辑回归之前, 首先对数据进行标准化预处理。图 3 为数据经过标准化

处理后各批样品主成分得分分布图。该图显示, 总体来说 YS 样品分布相对集中, 而 BY 与 ZP 样品分布较分散; YS 样品较多分布在 BY 和 ZP 样品之间位置; ZP、BY 和 YS 样品的分布并无明显的区域界限差异。可见, 简单的常规分类算法可能较难取得良好的效果。

3.3 真实数据的决策树分类模型解析

决策树算法是机器学习的一种基础分类算法, 可以采用树形作为形象化表示和理解, 所有样品由根节点出发, 依据一定指标规则对特征变量进行选择而分枝, 最终到达代表类别的叶节点。其中节点位置特征变量的选择直接决定了决策树模型分类效果, 该处通常以信息增益 (information gain, IG) 为具体选择指标, 以信息熵 (information entropy) 或基尼

(Gini) 系数作为衡量。本研究以信息熵为模型的度量,绘制决策树形象化树形示意图,如图 4 所示。由图 4 可见,选择在根节点以芒柄花素为特征,表明芒柄花素可以给样品分析带来最大的信息增益,即最大化信息熵。此结果与图 2 的相关性分析结果相契合,表明芒柄花素在黄芪的黄酮类成分中具有较为重要的指标性意义。依据该成分含量是否 $> 0.009\%$,可将野生黄芪、栽培黄芪与大部分的半野生黄芪区分开来。此外,决策树分析结果还显示结合上述芒柄花素成分含量情况,毛蕊异黄酮葡萄糖苷成分含量以 0.075% 为界,可将 10 批野生黄芪中

的 8 批从 64 批总样本中区分出来;刺芒柄花苷成分含量以 0.12% 为界,可将 30 批栽培黄芪中的 28 批从 64 批总样本中区分出来。为了避免样本过拟合情况的发生,决策树分析进行了剪枝设定,决策树进行到根节点下 3 层即止。

3.4 真实数据逻辑回归模型解析及指标性成分规律总结

逻辑回归模型预测的平均准确率为 82.2% ,表明模型在部分程度上具有一定的预测能力,参与建模的各黄酮类成分在 BY、YS 和 ZP 3 个等级分类模型方程中的系数见表 2。

表 2 逻辑回归模型系数

Tab. 2 Coefficients of logistic regression model

类别 (category)	毛蕊异黄酮葡萄糖苷 (campanulin)	刺芒柄花苷 (onospin)	毛蕊异黄酮 (calycosin)	山柰酚 (kaempferol)	异鼠李素 (isorhamnetin)	芒柄花素 (formononetin)	总和 (sum)
BY	1.405	0.978	-1.057	-0.422	0.240	-1.177	0.268
YS	-0.776	-0.898	0.218	0.233	-0.580	-0.792	-0.532
ZP	-0.630	-0.080	0.840	0.190	0.340	1.969	0.263

从原理上来看,逻辑回归模型是根据计算样本二分类的概率大小,来预测样本的二分类问题,但也能解决样本多分类的预测问题^[9-11],后者主要通过 Softmax 回归实现^[9]。Softmax 回归,即多项式逻辑回归,首先计算各类别得分,然后使用 Softmax 函数(归一化指数)对各得分求得各类别概率值高低^[8],计算过程中,各变量的系数大小一定程度上可以表征该变量对各类别得分影响的大小,从而影响最终分类概率。因此,表 2 列出的逻辑回归各变量系数,在一定程度上是这几个黄酮类成分在黄芪等级分类中重要性的表征。逻辑回归模型对 BY、YS、ZP 等级分类的系数绝对值排前 2 位的分别为毛蕊异黄酮葡萄糖苷和芒柄花素、刺芒柄花苷和芒柄花素,以及芒柄花素和毛蕊异黄酮,其中以芒柄花素系数出现的次数最多,可见芒柄花素在黄酮类成分中对于黄芪来源等级的表征的重要性。

真实数据分布及异常值等情况如箱式图(图 5)所示,6 个黄酮类成分中,以毛蕊异黄酮葡萄糖苷与刺芒柄花苷的含量分布最为集中,其次为芒柄花素。这 3 个成分的含量数据均有异常值出现,且均为数值较大的异常值。

从以上相关性分析、决策树模型和逻辑回归分类模型来看,芒柄花素是区分不同来源等级黄芪的

重要指标性成分,其次为毛蕊异黄酮葡萄糖苷和刺芒柄花苷;从箱式图也可看出,这 3 个成分数据分布较集中,适用于表征黄芪药材质量信息,对黄芪质量,尤其是来源等级的控制具有重要意义。若仅以这 3 个成分的数据区分黄芪来源等级,分类策略参考决策树分析,基于这 3 个成分的决策树分析结果可到树的第 2 层,64 批数据有 60 批分类正确,准确率已可达到 93.8% 。

3.5 数据增强技术模拟生成的数据与真实数据关系

机器学习是一种基于样本数据的科学,样本数据的大小和质量能够直接影响其应用的效果。本研究真实数据仅 64 批,尤其是 YS 类别仅有 10 批,对于机器学习来说,样本数量较少。此类情况在药物分析领域内较为普遍,样本收集的困难和数据的实验成本在一定程度上制约了实验数据的样本规模。因此,本研究尝试以真实数据为基础,利用深度学习中的数据增强技术生成虚拟数据,以考察此途径的可行性。

本研究分别使用了 4 种主要的表格数据生成技术,即 GaussianCopula、CopulaGAN、CTGAN 和 TVAE,最终选择采用 TVAE 技术。在处理高维、稀疏和不平衡的数据时,以及具有复杂关系和约束的数据时,TVAE 常常具有较好的效果。见表 3。

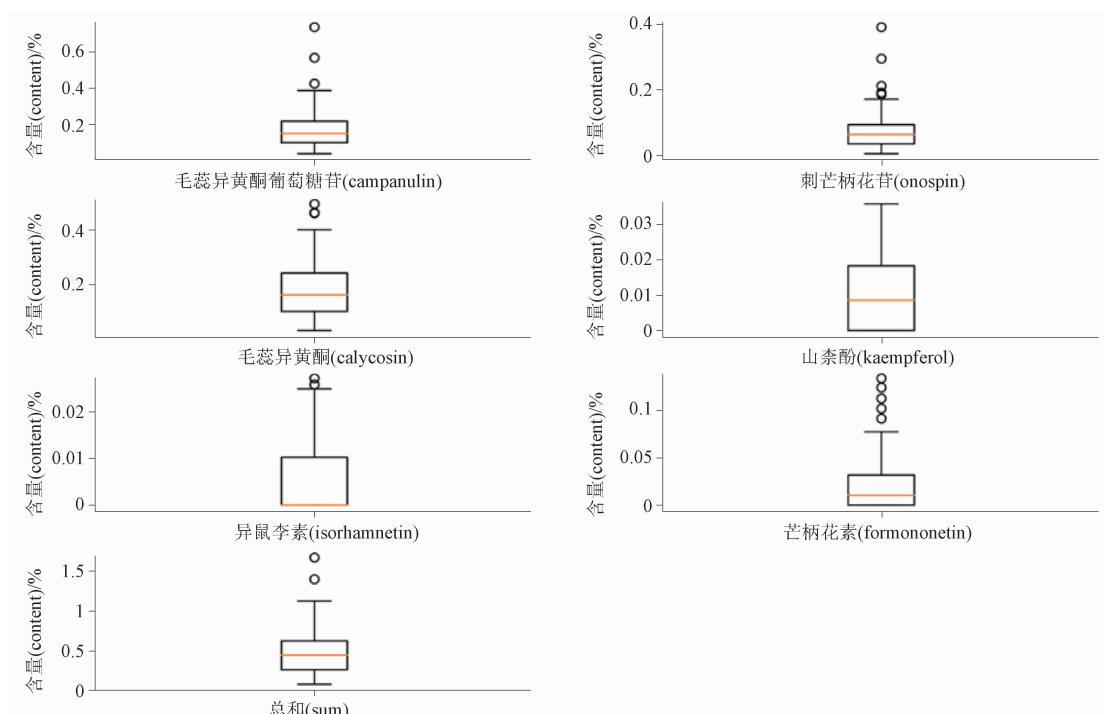


图5 数据箱式图

Fig. 5 Data box diagram

表3 数据增强结果评价

Tab. 3 Evaluation for results of data augmentation

类别 (category)	相似程度打分 (score on similarity)			
	GaussianCopula	CopulaGAN	CTGAN	TVAE
ZP	0.82	0.63	0.70	0.80
BY	0.69	0.68	0.68	0.74
YS	0.72	0.79	0.77	0.82

3.6 基于 AutoGluon 的模型建立及相关研究

AutoGluon 是亚马逊公司开发的 AutoML 框架,与通常的 AutoML 框架基于超参数搜索优化策略不同,AutoGluon 依赖于融合多个无需超参数搜索的模型,在相同数据上训练多个不同类型的模型,然后以 stacking 形式通过线性模型输出加权和结果,其中权重是通过训练获得。此外,AutoGluon 还采用了源于 K-则交叉验证的 K-则交叉 bagging 技术,对同类别的模型使用不同初始权重或数据块,对输出的结果做平均以减小模型输出方差。AutoGluon 还有多层 stacking 技术可供选择,但为了避免过拟合产生,多配合 K-则交叉 bagging 技术使用。AutoGluon 除了对表格数据进行处理外,还支持文本、图像等多模态数据学习。得益于这些先进的技术和思想,Auto-

Gluon 在机器学习竞赛中获得了较好的成绩,在预测房价比赛中获得了第 1 名,更在国际知名的机器学习 Kaggle 竞赛上,在对泰坦尼克号生还预测中轻松取得前 10% 的成绩。

表 1 展示了框架内各模型分别使用真实数据和虚拟数据情况下的学习预测准确率,内含交叉验证与测试集验证结果。首先,基于真实数据建立的 AutoML 内各模型以 NeuralNet 模型 (NeuralNetTorch 和 NeuralNetFastAI) 及树模型类 (ExtraTreesEntr、ExtraTreesGini、LightGBM、CatBoost、RandomForestGini 和 RandomForestEntr 等) 准确率最高,同时,WeightedEnsemble 作为模型融合的元模型 (Meta Model) 在真实数据交叉验证和模拟数据的测试集验证中都具有相当高的正确率;其次,基于模拟数据建立的 AutoML 内各模型中,主要以树模型类 (ExtraTreesEntr、ExtraTreesGini、LightGBMXT、CatBoost、RandomForestGini 和 RandomForestEntr 等) 准确率最高,同时,WeightedEnsemble 在真实数据交叉验证和真实数据的测试集验证中均具有最高的正确率。这部分结果表明:基于 NeuralNet 的模型和基于树模型的机器学习算法对于黄酮成分数据表征的黄芪而言分类效果最好,然而,从模型可解释性及复

杂程度考虑,基于树模型的各算法应该是更合适的选择;AutoML框架中各模型的各种验证显示,相较于真实数据而言,模型对于模拟生成的数据往往能够产生较好的分类预测结果,可见本研究使用的基于TVAE算法的数据增强技术模拟生成的数据分布与真实数据的确尚存在一定的差异。然而,从各类模型预测结果的准确率趋势来看,两类数据的确存在一定的相似性,对于药物分析数据而言,数据增强技术是可以用于数据不足之时的模型选择,甚至是用于模型训练。

4 总结与展望

传统医药是优秀传统文化的重要载体,发展中医药事业应“传承创新发展”,应立足于“中医药特点”,发挥“中医药的独特优势”^[12]。整体思维是中医药的主要思维规律,该认识具有重悟性而不重论证的特点^[13]。从创新发展和中医药整体思维角度出发,对于中药领域以及机器学习的本质而言,机器学习的相关技术应是当前应当尝试和发展的研究方向。本研究正是基于此宗旨,以黄芪样品中黄酮类成分含量为分析数据,使用了目前机器学习领域的热点技术,在解决黄芪不同等级的三分类及特征性黄酮类成分问题的基础上,考察实践了不同类型机器学习模型效果以及数据增强技术模拟生成的数据在模型训练中的应用效果,确定了AutoML框架和数据增强技术在中药质量分析和评价领域的应用价值及应用前景。

一般来讲,机器学习所建立的预测模型属于“黑盒子”,外界只清楚模型的输入和输出结果,至于其内部对输入数据的计算过程,则模糊不清。若仅仅将其用于样本分类或定量结果的预测,则低估了该领域技术的实际应用价值,如果能在准确率基础之上,开展基于输入数据的模型解释能力的研究,可以更加丰富和加强其在中药质量分析和评价领域的应用场景和应用意义。鉴于此,急需展开对于机器学习模型的特征解释,这也是下一步研究的着力点。

参考文献

- [1] 中华人民共和国药典 2020 年版. 一部 [S]. 2020:315
ChP 2020. Vol I [S]. 2020:315
- [2] DU HW, ZHAO XL, ZHANG AH, *et al.* Rapid separation, identification and analysis of *Astragalus membranaceus* Fisch. using liquid chromatography – tandem mass spectrometry [J]. J Chroma-

- togr Sci, 2014, 52(3):226
- [3] LI K, GAO F, WANG G, *et al.* Identification of cultured and natural *Astragalus* root based on monosaccharide mapping [J]. Molecules, 2015, 20(9):16466
- [4] LEE SM, JEONG JS, KWON HJ, *et al.* Quantification of isoflavonoids and triterpene saponins in *Astragalus Radix*, the root of *Astragalus membranaceus*, via reverse – phase high – performance liquid chromatography coupled with integrated pulsed amperometric detection [J]. J Chromatogr B, 2017, 1070:76
- [5] 石岩,贾天颖,李向日,等. 黄芪中多种黄酮类成分的测定研究 [J]. 药物分析杂志,2022,42(7):1120
SHI Y, JIA TY, LI XR, *et al.* Quantification of flavonoid compounds in *Astragalus Radix* [J]. Chin J Pharm Anal, 2022, 42(7):1120
- [6] 张丽,钱大玮,卜凡淑,等. 基于UPLC – MS的黄芪药材质量评价研究 [J]. 药物分析杂志,2020,40(4):722
ZHANG L, QIAN DW, BU FS, *et al.* Study on quality evaluation of *Astragalus Radix* base on UPLC – MS [J]. Chin J Pharm Anal, 2020, 40(4):722
- [7] 赵晨光,李存玉,杨珊,等. 基于道地产区蒙古黄芪的质量差异性分析 [J]. 中国中药杂志,2020,45(13):3183
ZHAO CG, LI CY, YANG S, *et al.* Analysis of quality difference based on *Astragalus membranaceus* var. *mongolicus* in genuine region [J]. China J Chin Mater Med, 2020, 45(13):3183
- [8] 裴文菡,何凡,程青松,等. 中药黄芪质量评价方法的研究进展 [J]. 中国现代应用药学,2020,37(5):620
PEI WH, HE F, CHENG QS, *et al.* Research progress on the quality evaluation methods of traditional Chinese medicine *Astragalus Radix* [J]. Chin J Mod Appl Pharm, 2020, 37(5):620
- [9] 李航. 统计学习方法 [M]. 第2版. 北京:清华大学出版社, 2019:91
LI H. Statistical Learning Methods [M]. 2nd Ed. Beijing:Tsinghua University Press,2019:91
- [10] HASTIE T, TIBSHIRANI R, FRIEDMAN J. The Elements of Statistical Learning: Data Mining, Inference, and Prediction [M]. 2nd Ed. New York:Springer, 2009:119
- [11] GÉRON A. Hands – on Machine Learning with Scikit – learn, Keras & TensorFlow: Concepts, Tools, and Techniques to Build Intelligent Systems [M]. Sebastopol:O’REILLY, 2017:134
- [12] 习近平. 在教育文化卫生体育领域专家代表座谈会上的讲话 [J]. 中华人民共和国国务院公报, 2020 (28):7
XI JP. Speech at the symposium for representative of expert in the fields of education, culture, health, and sports [J]. Gazette of the State Council of the People’s Republic of China, 2020(28):7
- [13] 黄璐琦. 对中医药发展规律及特点的传承与创新认识 [J]. 中医杂志,2022,63(17):1601
HUANG LQ. Inheritance and innovation of the development rules and characteristics of traditional Chinese medicine [J]. J Tradit Chin Med, 2022, 63(17):1601

(本文于2023年7月19日收到)