

Predictive Classification of Cultivated Land at County Scale by Using Bayes Discriminant Analysis Algorithms

—A Case Study from Huixian City, Henan Province

Haiyang Wang¹, Jie Chen^{1*}, Xingxing Han¹, Daoquan Cheng²

¹School of Environment and Water Conservancy, Zhengzhou University, Zhengzhou Henan

²Soil and Fertilizer Station, Henan Province, Zhengzhou Henan

Email: *jchen@zzu.edu.cn

Received: Jul. 28th, 2017; accepted: Aug. 9th, 2017; published: Aug. 21st, 2017

Abstract

The purpose of this study is to attempt to simplify the evaluation of cultivated land fertility by applying the machine learning algorithm, which aims to explore a new approach to the application of machine learning method in the evaluation work of cultivated land fertility at regional scale. Based on Technical Specification for Investigation and Quality Evaluation of Cultivated Land Fertility (NY/T 1634-2008) and the local practices of cultivated land evaluation, the methods applied by this study generally are supposed to use the based data obtained by the Financial Subsidy Project for Soil Testing and Formulated Fertilization conducted in Huixian county, Henan Province, to establish linear discriminate functions. 10 soil and site condition factors including surface soil texture, soil profile characteristics, surface gravel degree, rapidly available potassium in soil, available phosphorous in soil, organic matter content in soil, irrigation guarantee rate, capacity for drainage, geomorphic types, and surface slope are selected as the discriminant variables of cultivated land fertility level. By constructing the model of Bayes discriminant functions, Bayes discriminant analysis is employed to determine, analyze and classify land fertility in 5922 sampled sites of the studied region using that Bayes discriminate functions. The results of the methods demonstrate a prediction accuracy reaching up 89.1% after mathematical statistics verification and back substitution verification which means the original data being returned back to the Bayes discriminant functions. Under the premise of identifying the standard of evaluation and classification of cultivated land, the discriminant analysis algorithm has a unique advantage in analyzing and classifying the fertility situation of cultivated land and predicting the grade of cultivated land.

Keywords

Evaluation of Cultivated Land Fertility, Bayes Discriminant Analysis, Discriminate Functions Test

*通讯作者。

基于Bayes判别分析算法的县域耕地地力等级预测

——以河南省辉县市为例

王海洋¹, 陈杰^{1*}, 韩杏杏¹, 程道全²

¹郑州大学水利与环境学院, 河南 郑州

²河南省土壤肥料站, 河南 郑州

Email: *jchen@zzu.edu.cn

收稿日期: 2017年7月28日; 录用日期: 2017年8月9日; 发布日期: 2017年8月21日

摘要

为了研究机器学习算法——判别分析算法在土壤质量和土地资源评价方面的应用, 以此来简化耕地地力评价工作, 探索区域尺度上机器学习方法在地力评价应用的新途径。以河南省辉县市为研究区, 基于辉县市测土配方施肥财政补贴项目耕地地力评价工作获取的基础数据, 依据我国农业部标准(NY/T 1634-2008)《耕地地力调查与质量评价技术规程》和该市耕地地力评价实践经验, 选取研究区表层土壤质地、土壤剖面特征、地表砾石度、速效钾、有效磷、有机质含量、灌溉保证率、排涝能力、地貌类型、坡度等10个土壤和立地条件因素作为耕地地力水平的判别变量, 采用Bayes判别分析算法建模, 对研究区评价单元的耕地地力状况进行判断分析和归类分级。通过对判别函数和判别结果进行统计验证和回代验证, 显示判别结果与原始资料相比一致率达89.1%。在耕地地力评价与分级标准确定的前提下, Bayes判别分析在区域尺度上对分析耕地地力状况、预测耕地地力等级方面具有独特优势。

关键词

耕地地力评价, Bayes判别分析, 判别函数检验

Copyright © 2017 by authors and Hans Publishers Inc.

This work is licensed under the Creative Commons Attribution International License (CC BY).

<http://creativecommons.org/licenses/by/4.0/>



Open Access

1. 引言

耕地地力也称耕地基础地力, 是指在特定气候区域由地形、地貌、成土母质、土壤理化性状、农田基本建设及肥力培育水平等要素综合构成的耕地生产能力[1] [2]。耕地地力水平是影响作物生长发育的核心因素, 直接决定着农产品产量和质量[3]。我国是一个耕地资源严重约束型国家, 人均耕地面积远低于世界平均水平, 且空间分布不均衡, 基础地力偏低, 中低产田比例较高。作为世界第一人口大国, 我国粮食安全战略将持续面临巨大压力与挑战。河南省是传统的农业大省, 是我国最重要的粮食主产区之一, 在国家粮食安全战略中的地位与作用突出。为实现《国家粮食核心区建设规划》、《河南省高标准粮田“百千万”工程建设规划》提出的战略目标, 河南省正坚持以耕地质量建设为核心, 依靠科技进步, 大力实施耕

地质量培育措施,改善耕地土壤理化性状,提升耕地综合生产能力,强化抗御自然灾害的能力,保护农业生态环境。各项耕地质量培育、综合地力提升技术措施的有效实施,都离不开对耕地生产力现状的全面了解、系统评价以及对耕地地力等级的科学划分、对不同等级耕地资源空间分布格局的准确掌握。

在我国耕地地力评价与分级实践中,综合指数法是应用最为广泛的评价方法,也是我国农业行业标准(NY/T 1634-2008)中规定的耕地地力评价方法。这种方法是在确定评价因子及其权重和隶属度值的基础上,采用加法模型计算评价单元的综合地力指数,然后根据综合地力指数分布确定分级方案、划分地力等级[4]。尽管综合指数法在全国范围内、尤其是全国县域尺度耕地地力评价实践中具有基本一致的技术规程和实施方案,但步骤复杂,工作量大。在国内一些案例研究中,机器学习、决策分析等领域内的新技术、新方法被用于耕地地力评价,如粗糙集(Rough Set, RS)和支持向量机(Support Vector Machine, SVM)结合的组合法[5] [6]、ID3 (Iterative Dichotomizer 3)、C4.5、模糊支持决策算法[7] [8] [9] [10]以及分类与回归树算法[11]等。

判别分析(Discriminant Analysis)是一种依据已知类别的样本在若干指标上的观察值,建立关于指标的判别函数或判别准则,并据此对新样本进行分类的多元统计分析方法。根据建立的判别准则不同,具体的判别分析方法可分为距离判别法、Fisher 判别法、Bayes 判别法以及逐步判别法等。广义上,判别分析是一种机器学习(Machine Learning)技术,自 20 世纪 30 年代问世以来,在自然科学、人文社科以及经济管理领域的应用日益广泛[12] [13]。其中, Bayes 判别(Bayes Discriminant)是一种以 Bayes 法则(Bayes theorem)为核心的判别分析方法,它根植于 18 世纪英格兰著名数学家、长老会牧师 Thomas Bayes 的概率统计学思想,简述为:当无法准确知悉一个事物的本质时,可依靠与事物特定本质相关的事件出现的频率去判断其本质属性的概率;或换言之:支持某项属性的事件发生得愈多,则该属性成立的可能性就愈大。如果事件的结果不确定,那么量化它的唯一方法就是事件的发生概率;如果过去试验中事件的出现率已知,那么根据数学方法可以计算出未来试验中事件出现的概率,这就是 Bayes 判别分析的理论基础[14] [15] [16] [17]。Bayes 判别分析在国内矿藏评估[18] [19]、风险预警[20] [21] [22]、财务分析[23] [24]、交通规划[25] [26]、疾病诊断[27] [28]、智能识别[29] [30] [31]等领域应用广泛,但在土壤质量和土地资源评价方面的应用较为少见。本文尝试将 Bayes 判别分析引入耕地地力评价实践中,探索一条区域尺度上耕地地力分类预测及中低产田划分的新途径。

2. 数据与方法

2.1. 数据来源

研究区辉县市地处河南省西北部,地理坐标北纬 35°17'~35°50',东经 113°20'~113°57'之间(图 1),国土面积 2007 平方公里,境内西部为太行山脉,山地面积 1007 km²,平原面积 783 km²、丘陵 217 km²,全市耕地面积 5.34 × 10⁴ km²。本研究的主要数据源为辉县市测土配方施肥补贴项目及其耕地地力评价专项(2010 年~2012 年)获取的耕地表层土壤属性数据、立地环境数据、第二次土地调查土地利用数据库、最新修订的土壤图、地形图等相关图件资料等。2012 年辉县市依据农业部颁布的我国农业行业标准《耕地地力调查与质量评价技术规程(NY/T 1634-2008)》完成了全市耕地地力评价工作。实施过程中,以土壤图和土地利用现状图叠加产生的图斑作为耕地地力基本评价单元,选取表层土壤质地(x_1)、土壤剖面特征(x_2)、地表砾石度(x_3)、土壤速效钾(x_4)、有效磷(x_5)、有机质含量(x_6)、灌溉保证率(x_7)、排涝能力(x_8)、地貌类型(x_9)以及地表坡度(x_{10})等 10 个对辉县市耕地生产性能影响较大、区域内空间变异明显、且在时间序列上具有相对稳定性、与农业生产关系密切的因素作为耕地地力评价指标[32];利用层次分析法,建立评价指标与耕地潜在生产能力间的层次分析模型,计算单指标对耕地地力的权重;表层土壤质

地、土壤剖面特征、灌溉保证率、排涝能力、地表砾石度、地貌类型等概念型因子赋值后，其隶属度采用专家打分法确定，土壤有机质含量、土壤有效磷、速效钾以及地表坡度等定量因子则采用特尔斐法根据实测值评估对应的隶属度。

2.2. 研究方法

2.2.1. Bayes 判别函数

Bayes 判别分析的一个基本假设是对所研究对象(总体)在抽样前已有一定认识，通常用先验概率来描述这种认识。Bayes 判别就是根据总体的先验概率，使误判的平均损失达到最小而进行的判别。在本研究中，将研究区耕地评价单元分为 k 个不同耕地地力等级的总体 G_1, G_2, \dots, G_k ，每个总体含 p 个与土壤理化性质和立地条件相关的评价指标， $G_i = (X_1, X_2, \dots, X_p)^T$ ，对应 p 维样本空间 R_1, R_2, \dots, R_k ， p 元正态分布总体 $G_i \sim N_p(\mu^{(i)}, \Sigma_i)$ ， $i = 1, 2, \dots, k$ ，分布密度函数[33]为：

$$f_i(x) = (2\pi)^{-p/2} |\Sigma_i|^{-1/2} \cdot \exp\left\{-\frac{1}{2}(x - \mu^{(i)})' \Sigma_i^{-1} (x - \mu^{(i)})\right\} \quad (1)$$

不同地力等级的耕地评价单元的先验概率分别为 q_1, q_2, \dots, q_k ，其取值由研究区内各等级耕地的面积比确定(表 1)：

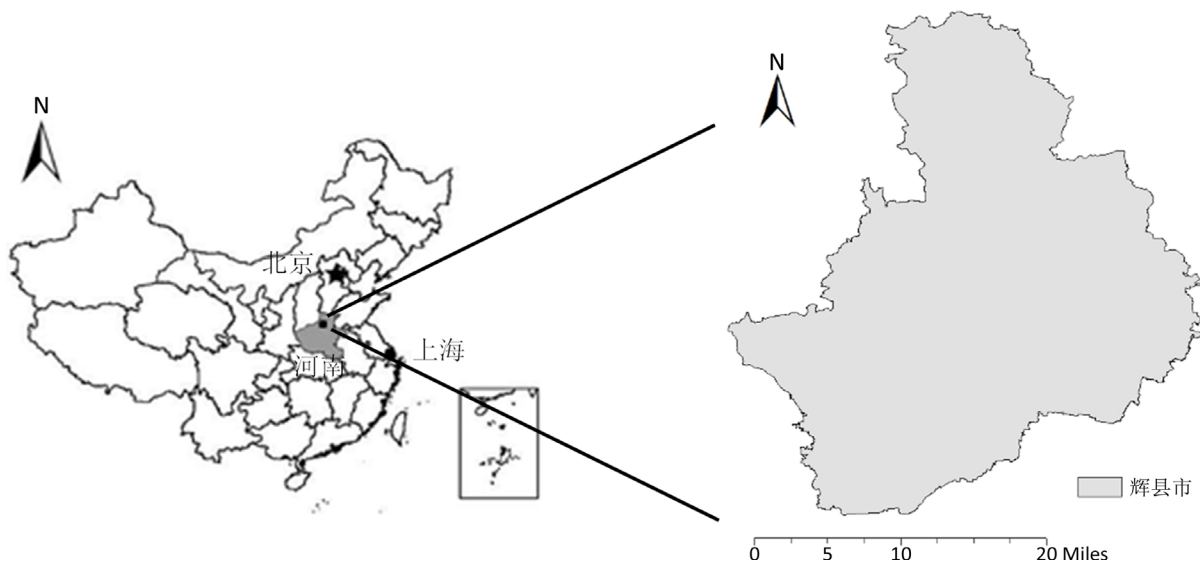


Figure 1. The map of the location of Huixian City

图 1. 辉县市地理位置示意图

Table 1. Prior probability of the land evaluation units with different fertility classes

表 1. 研究区不同地力等级耕地评价单元的先验概率

耕地等级 Cultivated land grade	一等地 First-class arable land	二等地 Second-class arable land	三等地 Third-class arable land	四等地 Fourth-class arable land
面积 Area/hm ²	21230	20500	15110	3160
比例 Proportion/%	35.38	34.17	25.19	5.27
先验概率 prior probability	0.35	0.34	0.25	0.06

将地力等级为 i 的样本错判为地力等级 j 造成的损失为 $L(j/i)$, 发生这种错判的概率为:

$$W(j/i, R) = \int_{R_j} f_i(X) dX, \quad i \neq j \quad (2)$$

错判造成的总平均损失[34]为:

$$g(R) = \sum_{i=1}^k q_i \sum_{j=1}^k L(j/i) W(j/i, R) = \sum_{j=1}^k \int_{R_j} \left(\sum_{i=1}^k q_i L(j/i) f_i(X) \right) dX \quad (3)$$

样品落入总体 G_i 对应的空间 R_i 的概率即后验概率为:

$$P(G_i/X) = \frac{q_i f_i(X)}{\sum_{j=1}^k q_j f_j(X)} \quad (4)$$

当 $P(G_h/X) = \max_{1 \leq i \leq k} P(G_i/X)$ 时, 样本 X 被判入地力等级为 h 的总体 G_h 。Bayes 判别分析的本质其实就是构造一个对 p 维空间的最优划分, 使得后验概率最大, 同时使错判造成的平均损失最小[35]。错判地力等级的损失可以给出定性的分析, 但很难定量描述, 故假定各错判损失均相同。如此, 平均错判损失最小与后验概率最大相互等价, 即使得 $y_i(X) = q_i f_i(X)$ 最大[34], p 元正态分布总体的概率函数见式(1)。

令 $Z_i(X) = \ln \left[(2\pi)^{0.5p} y_i(X) \right]$, 则判别函数[33] [34]为:

$$Z_i(X) = \ln q_i - \frac{1}{2} \ln |\Sigma_i| - \frac{1}{2} x' \Sigma_i^{-1} x - \frac{1}{2} \mu^{(i)'} \Sigma_i^{-1} \mu^{(i)} + x' \Sigma_i \mu^{(i)} \quad (5)$$

后验概率记为:

$$P(G_i/X) = \frac{\exp\{Z_i(X)\}}{\sum_{j=1}^k \exp\{Z_j(X)\}} \quad (6)$$

把未知地力等级的样本 X 代入判别式, 分别计算 $Z_i(X)$, 使 $Z_i(X)$ 为最大的总体 G_i , 如后验概率 $P(G_i/X)$ 也最大, 则把样品 X 归为总体 G_i , 耕地地力等级为 i 。

2.2.2. Bayes 判别检验与修正

① 检验原假设 H_{01} : $\mu^{(1)} = \mu^{(2)} = \dots = \mu^{(i)}$

一般耕地地力共有多个等级, 对于总体数量 $k > 2$ 的情况通常分两步来检验, 首先检验 k 个等级的耕地总体的评价指标均值向量是否全相等, 构建维尔克斯统计量[34]:

$$\Lambda(p, n-k, k-1) = \frac{|L_{xx}|}{|L_{xx} + B|} \quad (7)$$

L_{xx} 为各总体组内离差阵之和: $L_{xx} = \sum_{i=1}^k L_{xx}^{(i)}$,

B 为组间离差阵: $B = \sum_{i=1}^k n_i (\bar{X}^{(i)} - \bar{X})(\bar{X}^{(i)} - \bar{X})'$,

构造统计量 $V = -\left(n-1-\frac{p+k}{2}\right) \ln \Lambda(p, n-k, k-1)$ 近似服从自由度为 $p(k-1)$ 的 χ^2 分布, 根据 p 值大小判断是否拒绝原假设。若接受原假设, 即各等级耕地的指标均值全部相等, 说明这些耕地的土壤理化性状、立地条件、障碍因素等条件没有显著差异, 耕地等级需进行合并; 若不全相等, 则进行下一步两两逐对检验。

② 检验原假设 H_{02} : $\mu^{(i)} = \mu^{(j)}$ ($i, j = 1, 2, \dots, k; i \neq j$)

检验每两个总体之间差异的显著性, 构造统计量[34]为:

$$F_{ij} = \frac{(n-p-k+1)n_i n_j}{p(n-k)(n_i+n_j)} d_{ij}^2 \sim F(p, n-p-k+1) \quad (8)$$

服从分子自由度为 p , 分母自由度为 $n-p-k+1$ 的 F 分布, d_{ij}^2 为两个总体间的马氏距离:

$$d_{ij}^2 = (\bar{X}^{(i)} - \bar{X}^{(j)})' S^{-1} (\bar{X}^{(i)} - \bar{X}^{(j)})$$

$$S^{-1} \text{ 为组间协方差阵的逆矩阵: } S = \frac{L_{xx}}{n-k} = \frac{\sum_{i=1}^k L_{xx}^{(i)}}{n-k}$$

如结果接受原假设, 则均值相等的两个地力等级的土壤理化性状、立地条件等因素差异不明显, 应合并为一个总体, 与其他均值向量不同的耕地总体重新构建判别函数; 若拒绝原假设, 说明两类地力等级已经得以最优划分, 判别函数有效。

3. 结果与讨论

3.1. 判别函数的构建

按照《耕地地力调查与质量评价技术规程(NY/T 1634-2008)》, 将研究区耕地评价单元分为 4 个地力等级的总体, 每个总体由表层土壤质地(x_1)、土壤剖面特征(x_2)、地表砾石度(x_3)、速效钾(x_4)、有效磷(x_5)、有机质含量(x_6)、灌溉保证率(x_7)、排涝能力(x_8)、地貌类型(x_9)、坡度(x_{10})等 10 个评价因子描述。其中, 1、2、3、4 地力等级总体的先验概率分别为 0.35、0.34、0.25、0.06, 基于研究区测土配方施肥项目耕地地力评价中使用的评价因子对耕地地力影响的权重及其对应隶属度数据, 采用耕地自然要素评价指数来反映耕地潜在生产能力的高低, 关系式为:

$$IFI = f_1 x_1 + f_2 x_2 + \dots + f_n x_n \quad (9)$$

其中, IFI 为耕地地力指数, f_i 为耕地自然属性分, x_i 为该评价因子的权重。假设各总体的协方差矩阵相等, 且均为多元正态总体, 根据各因子权重数据、原始数据及其隶属度求得耕地地力指数, 利用统计分析工具 SPSS 判别分析模块, 计算 4 个总体 Bayes 判别函数各项系数, 结果列入表 2。

将表 2 中的自变量系数和常数项代入 4 个耕地地力等级的 Bayes 判别函数, 自变量为评价单元的耕地地力指数, 获得其线性函数形式:

$$Z_1 = 138.120x_1 + 65.255x_2 + 53.563x_3 + 226.449x_4 + 259.720x_5 + 341.849x_6 \\ + 34.822x_7 + 160.030x_8 + 89.433x_9 + 138.956x_{10} - 454.234$$

$$Z_2 = 126.895x_1 + 58.882x_2 + 56.784x_3 + 211.825x_4 + 232.774x_5 + 313.147x_6 \\ + 31.463x_7 + 155.024x_8 + 82.857x_9 + 134.640x_{10} - 392.033$$

$$Z_3 = 112.016x_1 + 47.279x_2 + 55.777x_3 + 194.910x_4 + 211.605x_5 + 286.183x_6 \\ + 26.852x_7 + 145.853x_8 + 70.403x_9 + 128.033x_{10} - 314.129$$

$$Z_4 = 99.097x_1 + 40.423x_2 + 38.759x_3 + 177.568x_4 + 191.291x_5 + 267.767x_6 \\ + 22.129x_7 + 131.919x_8 + 62.706x_9 + 105.360x_{10} - 245.494$$

将未知地力等级评价单元的耕地地力指数分别代入 4 个地力等级的判别函数, 当 Z_i 最大时, 样本落入对应等级的概率最大, 从而将样本划入这一地力等级。基于研究区所有耕地评价单元的相关数据进行 Bayes 判别函数的回代检验, 结果列入表 3。

结果显示存在误判情况, 原始资料中耕地地力等级为二等的的评价单元在 Bayes 判别规则下有 283 个评价单元判为一等地, 148 个评价单元判为三等地; 三等地评价单元有 54 个误判为二等地, 23 个误判为

Table 2. Coefficients of Bayes discriminant functions
表 2. Bayes 判别函数系数

自变量 Independent variable	系数 Coefficients	耕地地力等级 Cultivated land fertility grade			
		1	2	3	4
表层土壤质地 Surface soil texture		138.120	126.895	112.016	99.097
土壤剖面特征 Soil profile characteristics		65.255	58.882	47.279	40.423
地表砾石度 Surface gravel degree		53.563	56.784	55.777	38.759
土壤速效钾 Rapidly available potassium in soil		226.449	211.825	194.910	177.568
土壤有效磷 Available phosphorous in soil		259.720	232.774	211.605	191.291
有机质含量 Organic matter content		341.849	313.147	286.183	267.767
灌溉保证率 Irrigation guarantee rate		34.822	31.463	26.852	22.129
排涝能力 Drainage capacity		160.030	155.024	145.853	131.919
地貌类型 Geomorphical types		89.433	82.857	70.403	62.706
地面坡度 Ground slope		138.956	134.640	128.033	105.360
常量 constant		-454.234	-392.033	-314.129	-245.494

四等地；四等地评价单元有 139 个误判为三等地。与测土配方施肥项目开展的辉县市耕地地力评价资料相比，误判率为 10.9%。

3.2. 判别函数检验与分析

依照前文中的式(7)构造维尔克斯统计量，计算结果 $\Lambda(10, 5918, 3) = 0.556$ ，变换为近似服从自由度为 30 的 χ^2 分布的统计量， $\chi^2(30) = 3473.002 \gg \chi^2_{0.005}(30) = 53.672$ ，即 Bayes 判别函数耕地地力等级分类效果显著的概率为 99.5%。对每两个耕地地力等级之间差异的显著性进行检验，构造服从分子自由度为 10、分母自由度为 5909 的 F 分布的统计量，1 等到 4 等地力的每两个总体之间统计结果如表 4。

查 F 分布临界值表， $F_{0.01}(10, \infty) = 2.23$ ，远小于表 3 中每 2 个地力等级之间的计算值，故认为研究区耕地评价单元的被归入明显不同的地力等级的概率为 99%。综上所述，本研究构建的 Bayes 判别函数对总体做到最优划分。为进一步对判别结果进行分析，从而证明 Bayes 判别函数应用于耕地地力评价的有效性，现随机抽取 20 个样本的判别结果与原始数据对比见表 5。

**表示判别结果与原始资料不一致的样本。如表 5 所呈现，评价单元样本代入 Bayes 判别函数结果最大，落入该函数对应的总体类别的后验概率最大，这一类别即为样本单元的地力等级。随机抽取 20 个样本单元，有 2 个显示错判，原因分析如下：

1) 本研究采用的 Bayes 判别法考虑了各耕地评价单元的先验概率，根据抽取的待测评价单元属性特征来修正判别分析前已有的认识，从而得出后验概率；而测土配方施肥项目开展的辉县市耕地地力评价则不需要依赖任何先验描述，受先验概率的影响，二者的评价结果难免会有所不同。

2) 测土配方施肥项目开展的辉县市耕地地力评价划分级别的方法,是用评价单元数与耕地地力综合指数制作累积频率曲线图,根据单元综合指数的分布频率,在频率曲线图的突变处划分级别。本研究通过机器学习建立 Bayes 判别函数,将样本数据代入函数,根据函数值大小判断样本所属地力等级,避免前者方法在累积频率曲线突变临界点附近样本归属难以取舍,操作过程也更为简便。

故出现 Bayes 判别结果与原始资料不一致属正常现象,根据判别函数的检验和回代法的验证, Bayes 判别用于耕地地力等级预测是可行、有效的。

4. 主要结论

1) 河南省辉县市的案例研究表明, Bayes 判别分析算法的最大优势是可以用于多组判别问题,在多指标、多因素分类系统的样本归属判别领域具有巨大的应用潜力,非常适用于区域耕地评价单元的地力等级归属判别,具有相对简捷、高效、精准的特点。

2) Bayes 判别分析算法在研究对象的变量指标较多时,不能逐步进行筛选,剔除对总体影响不大、没有统计意义的变量,只能人为选择有用的评价指标,本文案例是基于特定范围内(完整的县域)、特定气候(一般来说,一个县域内的气候特征是基本相似的)条件下研究耕地生产力高低,若延伸到大地域范围,必然联系实际对评价指标进行增加或删减。可结合逐步判别法,利用其对评价指标有进有出的筛选,找出与研究区域关系最密切的评价指标。

Table 3. The results of Bayes discriminant analysis for cultivated land fertility in the study area

表 3. 研究区耕地地力等级判别结果

原始等级 Original level	预测等级 Coefficients	一等地 First-class arable land	二等地 Second-class arable land	三等地 Third-class arable land	四等地 Fourth-class arable land	总计 Total
一等地 First-class arable land		769	0	0	0	769
二等地 Second-class arable land		283	1396	148	0	1827
三等地 Third-class arable land		0	54	1602	23	1679
四等地 Fourth-class arable land		0	0	139	1508	1647
总计 Total		1052	1450	1889	1531	5922

Table 4. The calculation results of F distribution between levels of cultivated land

表 4. 各等级耕地间 F 分布计算结果

	一等地 First-class arable land	二等地 Second-class arable land	三等地 Third-class arable land	四等地 Fourth-class arable land
一等地 First-class arable land		335.964	2149.566	4712.255
二等地 Second-class arable land	335.964		1441.52	4550.76
三等地 Third-class arable land	2149.566	1441.52		1048.561
四等地 Fourth-class arable land	4712.255	4550.76	1048.561	

Table 5. The random sampling of Bayes discriminant analysis results
表 5. 随机抽取 Bayes 判别分析结果

样本序号 Sample number	原始等级 Original level	预测等级 Predictive level	后验概率 Posterior probability P(G = i D = d)				到等级质心的马氏距离 Mahalanobis distance to the centroid
			一等	二等	三等	四等	
			First class	Second class	Third class	Fourth class	
1	1	1	0.642	0.358	0.000	0.000	1.045
2	3	3	0.000	0.000	0.993	0.007	.190
3	2	3**	0.000	0.225	0.775	0.000	3.945
4	1	1	0.893	0.107	0.000	0.000	.204
5	3	3	0.000	0.000	0.994	0.006	2.365
6	1	1	0.975	0.025	0.000	0.000	.124
7	1	1	0.765	0.235	0.000	0.000	.615
8	2	2	0.136	0.864	0.000	0.000	.709
9	3	3	0.000	0.000	0.608	0.392	4.406
10	4	4	0.000	0.000	0.094	0.906	3.136
11	4	4	0.000	0.000	0.000	1.000	4.427
12	4	4	0.000	0.000	0.094	0.906	3.136
13	2	1**	0.790	0.210	0.000	0.000	.734
14	3	3	0.000	0.000	0.996	0.004	3.271
15	4	4	0.000	0.000	0.043	0.957	1.708
16	1	1	1.000	0.000	0.000	0.000	.615
17	1	1	0.922	0.078	0.000	0.000	.188
18	4	4	0.000	0.000	0.000	1.000	7.146
19	3	3	0.000	0.000	0.993	0.007	.190
20	4	4	0.000	0.000	0.146	0.854	12.732

3) Bayes 判别分析需要考虑研究对象的先验概率和错判损失, 这些因素对实验结果的影响大小仍有待考证, 有些试验中先验概率和错判损失直接可以忽略, 本研究则忽略了错判损失这一条件, 找到一种使二者最大限度发挥作用的计算方法很有必要, Bayes 判别算法仍有待改良。

基金项目

国家自然科学基金项目“基于分类距离-环境协变量回归模型的土壤数字化制图研究”(40971128)。

参考文献 (References)

- [1] 吴克宁, 郑义, 康鸳鸯, 等. 河南省耕地地力调查与评价[J]. 河南农业科学, 2004, 19(9): 49-52.
- [2] 陈艳春. 兴安县耕地地力等级划分与评价[J]. 南方农业学报, 2011, 42(7): 768-770.
- [3] 王瑞燕, 赵庚星, 李涛, 等. GIS 支持下的耕地地力等级评价[J]. 农业工程学报, 2004, 20(1): 308-310.
- [4] 田有国, 辛景树. 耕地地力评价[M]. 北京: 中国农业科学技术出版社, 2009: 1-326.
- [5] 赖红松, 吴次芳. 基于粗糙集和支持向量机的标准农田地力等级评价[J]. 自然资源学报, 2011, 26(12): 2141-2154.

- [6] 魏善沛, 章景, 王凯. 粗糙集与 SVM 的组合算法在人工林地力评价中的应用[J]. 中南林业科技大学学报, 2013, 33(5): 1-5.
- [7] 曹丽英, 孙学生, 赵月玲, 等. 一种基于决策树算法的耕地地力等级评价[J]. 东北林业大学学报, 2011, 39(2): 93-96.
- [8] 司巧梅, 宋丽. 决策树在耕地地力等级评价中的应用[J]. 安徽农业科学, 2011, 39(12): 7328-7329, 7417.
- [9] 孙微微, 胡月明, 刘才兴, 等. 基于决策树的土壤质量等级研究[J]. 华南农业大学学报, 2005, 26(3): 108-110.
- [10] 田剑, 胡月明, 王长委, 等. 聚类支持下决策树模型在耕地地力评价中的应用[J]. 农业工程学报, 2007, 23(12): 58-63.
- [11] 王莉莎, 李勇, 沈健林, 等. 应用回归树分析双季稻区水稻土地力特征[J]. 生态学杂志, 2013, 32(1): 232-240.
- [12] Dulam, J. (2005) Discriminate Analysis for Dust Storm Prediction in the Gobi and Steppe Regions in Mongolia. *Water, Air & Soil Pollution: Focus*, **5**, 37-49. <https://doi.org/10.1007/s11267-005-0725-0>
- [13] Pham-Gia, T., Turkkan, N. and Bekker, A. (2007) Bounds for the Bayes Error in Classification: A Bayesian Approach Using Discriminant Analysis. *Statistical Methods and Applications*, **2007**, 7-26. <https://doi.org/10.1007/s10260-006-0012-x>
- [14] Bruss, F.T. (2014) 250 Years of "An Essay Towards Solving a Problem in the Doctrine of Chances. By the Late Rev. Mr. Bayes, F.R.S. Communicated by Mr. Price, in a Letter to John Canton, A.M.F.R.S. *Jahresbericht der Deutschen Mathematiker-Vereinigung*, **115**, 129-133. <https://doi.org/10.1365/s13291-013-0069-z>
- [15] Puga, J.L., Krzywinski, M. and Altman, N. (2015) Bayes' Theorem. *Nature Methods*, **12**, 277-285. <https://doi.org/10.1038/nmeth.3335>
- [16] 曹玲玲, 潘建寿. 基于 Fisher 判别分析的贝叶斯分类器[J]. 计算机工程, 2011, 37(10): 162-164.
- [17] Wang, C., Wang, E.Y., Xu, J.K., et al. (2010) Bayesian Discriminant Analysis for Prediction of Coal and Gas Outbursts and Application. *Mining Science and Technology*, **20**, 520-523.
- [18] 丁帅伟, 姜汉桥, 陈民锋, 等. 基于模糊 C 均值聚类算法和贝叶斯判别函数研究深水油藏分类评价[J]. 西安石油大学学报(自然科学版), 2014, 29(2): 43-49.
- [19] 唐俊, 王琪, 马晓峰, 等. Q 型聚类分析和判别分析法在储层评价中的应用: 以鄂尔多斯盆地姬塬地区长₈储层为例[J]. 特种油气藏, 2012, 19(6): 28-31.
- [20] 刘在涛, 王栋梁, 张维佳, 等. 基于贝叶斯判别分析的地震应急响应等级初判方法[J]. 地震, 2011, 31(2): 114-120.
- [21] 赵钢铁钢, 杨大文, 李明亮. 超越概率贝叶斯判别分析方法及其在中长期径流预报中的应用[J]. 水利学报, 2011, 42(6): 693-698.
- [22] 吴凌峰, 兰乾玉, 吴凤山, 等. 基于贝叶斯判别法的工程岩爆危险评估与预测[J]. 施工技术, 2014(S1): 104-108.
- [23] 邹亚宝, 林佳佳. 基于贝叶斯判别的我国上市公司财务预警研究[J]. 广东外语外贸大学学报, 2012, 23(3): 31-36.
- [24] 黄振, 朱珺, 张为. 基于 Bayes 判别分析法的上市公司财务风险研究[J]. 洛阳理工学院学报(社会科学版), 2012, 27(3): 26-28.
- [25] 廖勇. 基于 Bayes 分析的区域综合运输通道出行方式分担率研究[J]. 物流技术, 2015, 34(8): 109-111.
- [26] 刘三豆. 公路软基处理方案技术决策的 Bayes 判别分析法[J]. 公路与汽运, 2016, 173(2): 116-118.
- [27] 宋淑钦, 张艳, 方丽珊, 等. 应用 Bayes 判别分析异位妊娠的早期诊断[J]. 医学综述, 2012, 18(23): 4095-4097.
- [28] 邵华飞, 刘剑芳, 姚丽波, 等. 探讨 Bayes 判别分析在食管癌淋巴结转移 CT 诊断中的价值[J]. 放射学实践, 2015, 30(8): 835-837.
- [29] 宫凤强, 李夕兵, 张伟. 基于 Bayes 判别分析方法的地下工程岩爆发生及烈度分级预测[J]. 岩土力学, 2010, 31(1): 370-377.
- [30] 刘树仁, 宋亚奇, 朱永利. 神经元形态分类和识别方法的研究[J]. 计算机工程与设计, 2013, 34(2): 670-674.
- [31] 张龙, 潘家荣, 朱诚. 基于近红外光谱的西湖龙井茶产地的精细判别[J]. 红外, 2015, 36(12): 41-46.
- [32] 聂长明, 袁长岭, 梅俊杰, 等. 河南省辉县市耕地地力评价工作报告[R]. 辉县: 辉县市土壤肥料管理站, 2012.
- [33] 袁志发, 周静芋, 等. 多元统计分析[M]. 北京: 科学出版社, 2002: 266-271.
- [34] 高惠璇. 应用多元统计分析[M]. 北京: 北京大学出版社, 2005: 175-211.
- [35] 戴佑才, 钟诚. 基于贝叶斯判别分析方法的膨胀土胀缩等级分类[J]. 公路工程, 2011, 36(5): 125-127.

期刊投稿者将享受如下服务：

1. 投稿前咨询服务 (QQ、微信、邮箱皆可)
2. 为您匹配最合适的期刊
3. 24 小时以内解答您的所有疑问
4. 友好的在线投稿界面
5. 专业的同行评审
6. 知网检索
7. 全网络覆盖式推广您的研究

投稿请点击：<http://www.hanspub.org/Submission.aspx>

期刊邮箱：hjas@hanspub.org