

决策分类器在空气质量数据分析中的实证对比

付悦¹, 夏小娜^{1,2,3}

(1. 曲阜师范大学 统计学院, 山东 曲阜 273165;

2. 曲阜师范大学 信息科学与工程学院, 山东 日照 276826;

3. 曲阜师范大学 中国教育大数据研究院, 山东 曲阜 273165)

摘要:分别使用传统统计学方法与机器学习方法对北京市空气质量状况展开实证分析,研究“判别分析”“决策树”“支持向量机”和“随机森林”四种分类方法对于同一数据集的分类结果。为了研究该问题,首先对传统分类方法构建的判别分析分类器与机器学习分类方法构建的三类分类器进行了理论介绍,随后以4:1的比例将北京市空气质量状况数据分为训练集与测试集,使用训练集构建分类器,对构建的分类器进行优化处理后,使用测试集进行分析及预测,根据预测结果对传统分类方法与机器学习分类方法进行详细实证对比和评价。采用传统分类方法与机器学习分类方法分别构建数据分类器,并应用于空气质量数据的统计和分析中,可为其他周期以及其他地区的空气质量研究提供方法支持,也可为类似数据的有效分类提供技术指导。

关键词:判别分析;决策树;随机森林;支持向量机;决策分类器;空气质量

中图分类号:TP39

文献标识码:A

文章编号:1673-629X(2020)07-0174-06

doi:10.3969/j.issn.1673-629X.2020.07.037

Empirical Comparison of Decision Classifiers in Air Quality Data Analysis

FU Yue¹, XIA Xiao-na^{1,2,3}

(1. Institute of Statistics, Qufu Normal University, Qufu 273165, China;

2. School of Information Science and Engineering, Qufu Normal University, Rizhao 276826, China;

3. China Institute of Big Data for Education, Qufu Normal University, Qufu 273165, China)

Abstract: The traditional statistical method and machine learning method are used to conduct empirical analysis on the air quality of Beijing respectively, and the classification results of discriminant analysis, decision tree, support vector machine and random forest for the same data set are studied. In order to study the problem, first of all, the discriminant analysis of traditional classification method to build a classifier with machine learning classification methods of three kinds of classifiers for the construction theory is introduced, and then in a ratio of 4:1 Beijing's air quality data can be divided into training set and test set, with the training set to build a classifier, for the construction of the classifier after optimized, using test set for analysis and prediction. According to the predicted results, the traditional classification method and machine learning classification methods are carried out empirical comparison and evaluation in detail. The traditional classification method and machine learning classification method are respectively used to construct data classifiers and apply them to the statistics and analysis of air quality data, which can provide methodological support for air quality research in other periods and other regions, and also provide technical guidance for effective classification of similar data.

Key words: discriminant analysis; decision tree; random forest; support vector machine; decision classifier; air quality

0 引言

近年来,使用机器学习进行数据挖掘的研究越来越丰富^[1-2]。在使用机器学习进行数据挖掘的各种算

法中,数据分类算法的研究具有重要意义。分类算法能够提取并刻画数据中的各个类别,基于这些不同的类别分析研究数据可以更加全面地理解数据与解决

收稿日期:2019-09-02

修回日期:2020-01-07

基金项目:山东省自然科学基金(ZR2016FM45);山东省社会科学规划研究项目(19CZKJ08);山东省研究生教育优质课程(SDYKC18079);山东省研究生教育创新计划项目(SDYY16092);曲阜师范大学交叉学科研究项目(XKJJC201901)

作者简介:付悦(1996-),女,硕士生,研究方向为数据分析、应用统计;夏小娜,博士,副教授,CCF会员(E3490G),研究方向为教育大数据、数据分析、数据挖掘。

问题。

在基于机器学习的分类算法兴起之前,研究人员多使用传统的分类方法如逻辑回归或者判别分析对数据进行分类^[3-4],这些分类方法有很大的局限性,比如逻辑回归多适用于二分类的情况,而判别分析多应用于数据的线性分类,正是由于传统分类方法的一些局限性,很多专家和学者在数据分类方面的研究不断向机器学习的分类算法靠近^[5]。

基于机器学习的分类算法有很多,比较常用的有决策树、随机森林和支持向量机。在数据的分类问题上,目前对于分类器的研究多集中在模型的构建以及优化方面,很少有研究将传统分类方法与机器学习分类方法进行对比。本研究以判别分析、决策树、随机森林和支持向量机为例,以北京市2018年1月到11月每小时的空气质量状况为样本,对以上几种算法进行实证分析并比较它们在处理空气质量状况问题中的应用效果及存在的差别。

1 研究基础

1.1 传统统计学分类方法-判别分析

传统的分类方法有很多,比如逻辑回归、判别分析等,基于本次研究是对空气质量状况进行分类,使用判别分析的分类方法进行实验最为合适,因此主要介绍和使用判别分析方法。使用判别分析分类的关键在于利用当前样本信息,选择合适方法建立相关判别规则,新样本便可以根据判别规则判断其所属类别^[6]。

根据判别规则的不同,判别分析主要分为距离判别、贝叶斯判别、Fisher判别和逐步判别,相关概念和思想可查阅相关文献^[7]。根据本研究所使用的数据集特点和对已有文献的分析论证,Fisher判别对空气质量状况的分类更加有效可行。

Fisher判别的基本思想是借助投影使数据降维。该方法首先选择一个适当的投影轴,使所有的样本点都投影到该轴上,使用Fisher判别投影的关键是使每一类的类内离差尽可能小,而不同类之间的类间离差尽可能大^[8]。

Fisher判别构造的线性判别函数表示为:

$$C(Y) = C_1 Y_1 + C_2 Y_2 + \cdots + C_p Y_p \quad (1)$$

其中, Y 为判别类别, C 为判别函数的系数。判别函数系数的估计表示为:

$$C = (\sum_1 + \sum_2)^{-1}(\mu_1 + \mu_2) \quad (2)$$

其中, μ, Σ 是变量 X 的均值与方差。

1.2 机器学习分类方法

1.2.1 支持向量机

支持向量机(SVM)是机器学习中常用的分类算法之一,该模型的关键是尽量扩大不同类别之间的分

类间隔,分类间隔越大得到的分类模型会越准确,详细解释可见参考文献[9]。

以一个二分类问题为例,使用支持向量机构建分类器就是在该数据空间中使用一个合适的超平面将原始数据分为两类。对于该超平面的寻找方法是通过寻找最大化边缘超平面进行。寻找最大化边缘超平面意味着将不同类别的数据以充分大的间隔进行分类即在使用支持向量机分类时,不仅考虑将正负样本分开,对于处在类别边缘的样本也要尽最大可能地正确分类,这便是使用支持向量机分类的关键所在^[10]。

核函数是使用支持向量机建模的重点所在,它可以对一些复杂非线性数据建模,核函数的关键原理是空间映射,维度的变化使复杂数据的分类变得简单。常用的核函数有高斯核函数、多项式核函数与字符串核函数,关于核函数的具体思想可查阅文献[11]。由于高斯核函数较为简单,适用范围较广,因此文中对空气质量状况的分类主要使用高斯核函数。

高斯核函数如式(3):

$$K(x, z) = \exp(-\|x - z\|^2 / 2\sigma^2) \quad (3)$$

其中,样本 $x, z \in \chi$ (χ 为欧氏空间的子集或者集合), σ^2 为数据方差。

1.2.2 决策树

决策树算法的关键是遍历所有节点,并在每一个节点处分类,直至遍历结束输出最终的数据分类结果^[12]。决策树的分类过程主要包括决策树的构建和优化。

首先需要从原始数据的所有特征中选取一个作为初始分类节点,然后遍历数据所有特征,直至数据不可再分;其次进行模型优化即剪枝操作。这样生成的初始决策树可能会存在节点过多的现象即出现过拟合,容易使决策树过于膨胀,为此需要对决策树进行剪枝,以此缩小树的规模与结构^[13]。

决策树分类的准确性与每个节点特征的划分规则相关,规则通常为信息增益准则、信息增益比准则与Gini指数准则这三种,分别对应三种决策树算法^[14]。在三种决策树算法中,使用Gini指数作为划分准则的CART算法是前两种算法的改进,这里主要介绍CART算法^[15]。

由于CART算法使用Gini指数作为划分准则,首先需要定义样本 X 的Gini值,计算公式表示为:

$$\text{Gini}(X) = 1 - \sum_{k=1}^{|k|} (c^k / |X|)^2 \quad (4)$$

其中, c 为特征总数, c^k 即属于第 k 类的样本数。

设 M 为样本特征, m 是它可能的取值, X_1 和 X_2 是样本集 X 分割的两部分,具体表示为:

$$\begin{cases} X_1 = \{(X, Y) \in X \mid M(x) = m\} \\ X_2 = X - X_1 \end{cases} \quad (5)$$

此时, X 的 Gini 指数可表示为:

$$\text{Gini}(H, M) = \frac{X_1}{X} \text{Gini}(X_1) + \frac{X_2}{X} \text{Gini}(X_2) \quad (6)$$

1.2.3 随机森林

随机森林是决策树的拓展算法,它的主要思想是集成学习,基本单位是决策树。与决策树算法不同的是,随机森林是由数个决策树构成,处在森林中的每个决策树会生成一个分类结果,随机森林利用集成学习方法分析所有决策树的分类结果,投票数最多的分类作为最终结果,即 Bagging 算法^[16]。

(1) Bagging 算法。

Bagging 算法需要在原始数据集中进行多次随机采样,将每次采样结果作为一个训练样本分类器直至所有样本训练结束,在多次训练中选择出现次数最多的分类作为最终分类器,再将测试样本输入最终分类器中预测分类结果,随机森林是 Bagging 算法的典型应用^[17]。

(2) 随机森林构建流程

首先从原始数据集中进行 n 次随机采样,根据采样得到 n 个样本集,利用 n 个样本集训练决策树模型,每个决策树分类器生成的分类结果是不同的,所有的决策树组成随机森林。在随机森林中,选择出现次数最多的决策树模型作为最终的分类模型。具体构建流

程可见参考文献[18]。

2 实证分析

基于北京市 2018 年 1 月至 11 月每小时的空气质量监测数据进行实证研究,分别使用传统分类方法与三种机器学习分类方法对北京市空气质量进行分类,根据结果分析比较几种方法的有效性与准确性。

空气质量指数(air quality index, AQI)可用来衡量当前地区的空气质量状况。AQI 的高低反映了空气污染程度,空气质量状况通常可分为优、良、轻度污染、中度污染和重度污染五级^[19]。AQI 主要受空气中污染物的影响,这些污染物主要包括 PM2.5、PM10、CO、NO₂、SO₂ 与 CO。根据空气污染物浓度的高低,对当前空气质量进行分类,借助分类结果研究主要污染物对空气质量的影响。

在构造分类器之前,将北京市 2018 年 1-11 月每小时的空气质量数据以 4:1 的比例分为训练集与测试集,训练集用来对数据进行学习建模,测试集用来检验所建模型的优良性。

2.1 线性判别分析分类

由于北京市空气质量数据为线性分类数据,同时该数据为多分类情况,因此,在传统的分类方法中,选择线性判别方法。

使用训练集构造线性判别分类器后,带入测试数据进行判断,测试结果如表 1 所示。

表 1 判别分析方法对空气质量状况的预测分类结果

空气质量状况	优	良	轻度污染	中度污染	重度污染	成功分类比例/%
优	647	96	0	0	0	87.53
良	115	632	65	0	0	77.83
轻度污染	0	127	287	12	0	69.32
中度污染	0	5	54	185	7	73.70
重度污染	0	0	3	9	12	50

表 1 为线性判别分析方法对北京市空气质量状况的分类结果。结果显示,空气质量状况为“优”时有 87.53% 的数据分类正确,数据判错类别“良”;空气质量状况为“良”时有 77.83% 的数据分类正确,判错类别大多为“优”;空气质量状况为“轻度污染”时有 69.32% 的数据分类正确,判错类别大多为“良”;空气质量状况为“中度污染”时有 73.7% 的数据分类正确,判错类别大多为“轻度污染”;空气质量状况为“重度污染”时仅有 50% 的数据分类正确。线性判别分类器的成功率在 78.25% 左右,结果并非很理想。同时,成功分类比例显示每一类的判错概率或高或低,而处于两个类别之间的数据使用线性判别方法并不容易判断,这会导致分类不准确,错误率升高;对于数据量较

小的类别如“重度污染”,使用该方法出错率很高。

2.2 支持向量机分类

一般的支持向量机只能将数据分成两类,由于空气质量分为五级,需要使用多分类的分类器。在用 R 语言构造分类器之前,需要为分类设定合适的 gamma 和 cost 参数,它们分别表示支持向量的个数和惩罚系数。其中,gamma 函数在 0.001 到 100 之间, cost 在 10 到 100 之间。通过多次测试,最终得到最优参数为 gamma=0.1, cost=100,以此为基础所构建支持向量机分类器并完成测试数据的分类,结果如表 2 所示。

表 2 显示,空气质量状况为“优”时有 96.50% 的数据分类正确,数据判错类别大多为“良”;空气质量状况为“良”时有 90.63% 的数据分类正确,判错类别

为“优”和“轻度污染”;空气质量状况为“轻度污染”时有 83.7% 的数据分类正确,判错类别大多为“良”;空气质量状况为“中度污染”时有 96.5% 的数据分类正确,判错类别大多为“轻度污染”;空气质量状况为“重度污染”时有 90.47% 的数据分类正确,有两个判

为“中度污染”。支持向量机分类器的成功率在 91.51% 左右,结果较好。同时,成功分类比例显示每一类的判错概率或高或低,但大多在 90% 以上,没有过低情况,这使得整体精度达到 91.51%。

表 2 支持向量机对空气质量状况的预测分类结果

空气质量状况	优	良	轻度污染	中度污染	重度污染	成功分类比例/%
优	716	25	1	0	0	96.50
良	31	841	54	2	0	90.63
轻度污染	0	40	339	26	0	83.70
中度污染	1	2	13	218	1	96.50
重度污染	0	0	0	2	19	90.47

2.3 决策树分类

最初构建的决策树可能存在过拟合现象从而使决策树较大,需要对决策树进行剪枝。表 3 反映了交叉验证误差 $xerror$ 与复杂度参数 cp 的关系,通过分析它们之间的关系确定最终决策树的大小。表 3 显示损失函数最小的节点为第 10 节点,此时最小交叉验证误差为 0.129,由于标准误差为 0.006 09,因此最优树为 $0.129 \pm 0.006 09$,以此为基准进行剪枝。

表 3 交叉验证误差与复杂度关系表

序号	cp	error	xstd
1	0.407 9	1.000	0.010 92
2	0.160 1	0.592	0.010 72
3	0.118 0	0.431	0.009 82
4	0.059 6	0.313	0.008 76
5	0.049 3	0.254	0.008 06
6	0.030 9	0.204	0.007 36
7	0.018 8	0.173	0.006 85
8	0.013 3	0.155	0.006 53
9	0.012 1	0.141	0.006 41
10	0.010 0	0.129	0.006 09

图 1 为对决策树进行剪枝操作后的结果。该图显示,对空气质量状况进行分类时决策树所构造的根节点为“良”,意为空气质量为“良”时所占比重相对较大,在根节点之下,根据不同的空气污染物再次对数据分类,直至所有分类结束,共构建了五类空气质量的叶节点。

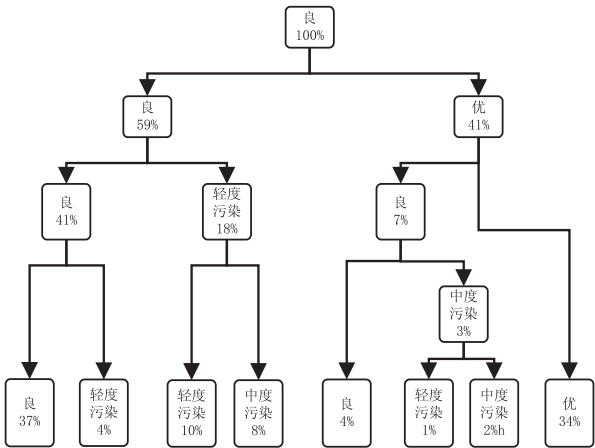


图 1 决策树分类器示意图

使用图 1 得到的决策树数据分类器,带入测试数据得到表 4。

表 4 决策树对空气质量状况的预测分类结果

空气质量状况	优	良	轻度污染	中度污染	重度污染	成功分类比例/%
优	743	0	0	0	0	100
良	41	771	0	0	0	95
轻度污染	10	47	369	0	0	86.6
中度污染	2	8	35	206	0	82.1
重度污染	0	0	6	15	0	0

表 4 为决策树分类器对测试数据的分类结果。空气质量状况为“优”时全部数据分类正确;空气质量状况为“良”时有 95% 的数据分类正确,判错类别为“优”;空气质量状况为“轻度污染”时有 86.6% 的数据分类正确,判错类别大多为“良”;空气质量状况为“中度污染”时有 82.1% 的数据分类正确,判错类别大多

为“轻度污染”;空气质量状况为“重度污染”时没有数据分类正确。决策树分类器的成功率在 92.7% 左右,结果比较好。但当空气质量状况为“重度污染”时数据全部分类错误,通过对数据分析可以得到在北京市的空气质量状况为“重度污染”时的数量非常少,仅占 0.009 7%,同时由于“重度污染”的 AQI 与“中度污

染”的 AQI 临界值比较接近,导致“重度污染”的类别区分度很小,使空气质量状况为“重度污染”的类别全部分类错误。

2.4 随机森林分类

使用随机森林分类器对数据集分类,首先默认生成 500 棵决策树即 500 个分类器,接下来借助分类器指标 MeanDecreaseGini^[20] 分析分类变量的重要性。

图 2 是使用训练数据构建分类器得到的 MeanDecreaseGini 图,PM2.5 与 PM10 在六个变量中居于首要

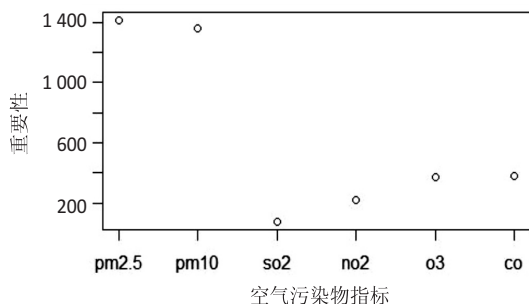


图 2 MeanDecreaseGini

表 5 随机森林对空气质量状况的预测分类结果

空气质量状况	优	良	轻度污染	中度污染	重度污染	成功分类比例/%
优	735	8	0	0	0	98.92
良	19	782	11	0	0	96.31
轻度污染	1	32	393	1	0	92.04
中度污染	0	3	20	228	0	90.84
重度污染	0	0	0	1	20	95.28

地位,而 SO₂ 的重要性相对来说是最低的,通过分析有关 AQI 影响因素的文献及数据资料得知^[21],最影响北京市空气质量状况的空气污染物确实是 PM2.5 与 PM10。

使用测试数据对随机森林分类器进行检验,得到表 5。空气质量状况为“优”时有 98.92% 的数据分类正确,判错类别为“良”;空气质量状况为“良”时有 96.31% 的数据分类正确,判错类别为“优”和“轻度污染”;空气质量状况为“轻度污染”时有 92.04% 的数据分类正确,判错类别大多为“良”;空气质量状况为“中度污染”时有 90.84% 的数据分类正确,判错类别大多为“轻度污染”;空气质量状况为“重度污染”时有 95.28% 的数据分类正确,判错为一个“中度污染”。随机森林分类器的成功率在 94.98% 左右,结果良好。表 5 的数据显示,对于每一个类别的空气质量状况,判断正确的概率比较均匀,均在 90% 以上,这使得整体判对概率达到 94.98%,精确度比较高,分类比较准确。

3 分类结果比较

本研究将传统的统计学分类方法和机器学习的三

种分类方法运用于北京市 2018 年度空气质量状况的数据分析中,通过结果论证,其分类特点如表 6 所示:

表 6 四种分类器优缺点对比分析

分类器	优点	缺点
判别分析	不需要调整参数无优化权重问题,计算效率较高,判别方法多样	结果误差较大,边界数据分类困难
支持向量机	使用核技巧建模,方便处理复杂的非线性数据,决策结果由支持向量决定,边界预测稳定	多分类情况运行缓慢,少数支持向量决定结果而非样本空间维数,非支持向量对模型影响较小
决策树	可以得到每个节点的分类情况,生成可理解的规则	样本量差距较大时,数据量小的类别易被删除,造成误差;类别多时错误累计较快
随机森林	每一棵树是一个决策树,多棵树投票在多分类时结果更加准确,算法易于理解	数据量较少时不便于抽样,造成结果误差较大

(1)传统的统计学分类方法虽然不需要调整参数的优化权重问题,但是预测精度是最低的,而且对于每一类的预测精度都低。与三种机器学习的分类方法比较均处于劣势位置。

(2)使用支持向量机对数据分类的结果较传统统计学分类方法略好。支持向量是支持向量机分类器分类的关键,对于处于两类别之间的数据的分类相对较

好。但支持向量机的算法运行较为缓慢,多分类时会造成较大的误差以及分类的低效率。

(3)使用决策树分类器的结果优于支持向量机。能够比较清晰地看到每个分类节点,利于分析空气中不同污染物对于空气质量的影响。缺点在于使用该方法构建分类器时一种类别的数据比重较小且在各节点处的分类不明显,会导致删除此类数据的分类,适合处

理不同类别的数据分布理想的样本。

(4)使用随机森林分类器对数据分类的结果是机器学习分类方法中最优的,缘于随机森林的评定结果建立在对所有决策树结果的综合分析基础上,而且分类错误的情况比较少,准确率也最高。

因此,机器学习的分类算法明显优于传统统计学分类方法,同时,具有全部决策树综合分析能力的随机森林分类器最稳定最有效。

4 结束语

通过对传统统计学分类方法和机器学习的三种分类算法进行比较,得到以下结论:

传统分类器有它独特的优点,但是缺点也比较明显,比如线性判别只能用于处理线性问题,而且在不同类别之间的数据边界不清晰时分类结果并没有那么准确。当数据类别之间的区分度不清晰时,机器学习算法更加具有优势。

机器学习的三种分类算法各有优劣,适用情景有所差别。针对本研究的北京市空气质量状况数据集,随机森林分类器的分类效率与分类精度明显优于决策树分类器和支持向量机分类器,分类结果更有统计学意义和参考价值。

分类节点比较多时可以优先考虑决策树分类器,并且通过剪枝操作以减少分类模型的过拟合情况。但是在分类节点比较少或者每个节点均比较重要时慎用此类方法。

当数据集呈现二分类的情形时,支持向量机最为适用。但是在数据集的多分类情况下,应当减少支持向量机的使用。

本研究的方法优化和实证分析过程可为不定周期、存在多类别的数据集分析和统计提供支持,具有应用实践意义,并提供有效且通用的建模方法和技术指导。

参考文献:

- [1] 宋毅,徐志明. 个性化搜索用户兴趣更新学习及评价研究[J]. 计算机技术与发展,2018,28(6):64-66.
- [2] 郑顾平,闫勃勃,李刚. 基于机器学习的多车牌识别算法应用研究[J]. 计算机技术与发展,2018,28(6):129-132.
- [3] 李玉榕,项国波. 一种基于马氏距离的线性判别分析分类算法[J]. 计算机仿真,2006,23(8):86-88.
- [4] 涂艳,王翔宇. 基于机器学习的P2P网络借贷违约风险预警研究——来自“拍拍贷”的借贷交易证据[J]. 统计与信息论坛,2018,33(6):69-76.
- [5] 何清,李宁,罗文娟,等. 大数据下的机器学习算法综述[J]. 模式识别与人工智能,2014,27(4):327-336.
- [6] LI Wenbo. Comprehensive evaluation research on circular economic performance of eco-industrial parks[J]. Energy Procedia,2011,5:1682-1688.
- [7] LI F,FERGUS R,PERONA P. Learning generative visual models from few training examples:an incremental Bayesian approach tested on 101 object categories[J]. Computer Vision and Image Understanding,2005,106(1):59-70.
- [8] 蒲莉娟. 模式判别中的子空间分析方法研究[D]. 西安:西安电子科技大学,2014.
- [9] 王云鹏. 基于L₀-SVM特征选择算法的研究及在人脸识别中的应用[D]. 西安:西安科技大学,2017.
- [10] 郑宇晨. 基于支持向量机的英文字符识别研究[J]. 计算机技术与发展,2019,29(1):106-109.
- [11] YU H,YANG J,HAN J,et al. Making SVMs scalable to large data sets using hierarchical cluster indexing[J]. Data Mining and Knowledge Discovery,2005,11:295-321.
- [12] 史众,陈立平,陈天恩. 决策树在精准农业中的应用现状与发展趋势[J]. 北方园艺,2011(16):218-220.
- [13] KODIKARA G R L,WOLDAI T. Spectral indices derived, non-parametric decision tree classification approach to lithological mapping in the Lake Magadi area, Kenya[J]. International Journal of Digital Earth,2018,11(10):1020-1038.
- [14] 崔凤焦. 表情识别算法研究进展与性能比较[J]. 计算机技术与发展,2018,28(2):145-149.
- [15] 吴泽泰. 基于位置指纹的WiFi室内定位算法研究与系统设计[D]. 广州:华南理工大学,2017.
- [16] 王秋城. 基于OCTEON多核处理器的网络流量分类技术研究[实现][D]. 北京:北京邮电大学,2017.
- [17] KAMINSKA J A. A random forest partition model for predicting NO₂ concentrations from traffic flow and meteorological conditions[J]. Science of the Total Environment,2018,651:475-483.
- [18] 陈诚. 基于Hadoop的数据挖掘算法并行化研究[D]. 广州:广东工业大学,2015.
- [19] 李静. 考虑空气质量及供电能力的主动配电网优化调度研究[D]. 北京:华北电力大学,2016.
- [20] 周涛,汝小龙. 北京市雾霾天气成因及治理措施研究[J]. 华北电力大学学报:社会科学版,2012(2):12-16.
- [21] 王宇燕,王杜娟,王延章,等. 改进随机森林的集成分类方法预测结肠癌存活性[J]. 管理科学,2017,30(1):95-106.