

基于随机森林的建筑物场景微蜂窝模型的构建

陈明发^{1 2} 严天峰^{1 2 3}

¹(兰州交通大学电子与信息工程学院 甘肃 兰州 730070)

²(甘肃省无线电监测及定位行业技术中心 甘肃 兰州 730070)

³(甘肃省高精度北斗定位技术工程实验室 甘肃 兰州 730070)

摘 要 在市区微蜂窝中,信号的传播由于建筑物阻挡衰减很大。传统的微蜂窝模型在测试地区实际应用时存在很大误差的问题,对此提出改进市区微蜂窝传播模型。通过测试计算路径损耗斜率,提升了自由空间损耗模型精度,以改进算法分析建筑物损耗为基础进行建模。实验结果表明,本地区的路径损耗斜率为 23.51 dB/dec,改进路径损耗斜率的自由空间损耗模型与传统模型相比,精度提升了 1.5 dB;基于随机森林构建建筑物损耗模型明显优于其他算法构建的建筑物损耗模型。

关键词 微蜂窝 随机森林 路径损耗 误差

中图分类号 TN011 TP3 文献标志码 A DOI: 10.3969/j.issn.1000-386x.2020.10.019

CONSTRUCTION OF BUILDING SCENE MICROCELLULAR MODEL BASED ON RANDOM FOREST

Chen Mingfa^{1 2} Yan Tianfeng^{1 2 3}

¹(School of Electronic and Information Engineering, Lanzhou Jiaotong University, Lanzhou 730070, Gansu, China)

²(Radio Monitoring and Technology Center of Positioning Industry of Gansu Province, Lanzhou 730070, Gansu, China)

³(High-precision Positioning Technology Compass Engineering Laboratory of Gansu Province, Lanzhou 730070, Gansu, China)

Abstract In urban microcells, signal propagation is greatly attenuated due to building block. Because the traditional microcellular model has a lot of errors in the practical application of the test area, the paper proposes to improve the urban microcellular propagation model. By testing and calculating the slope of path loss, the accuracy of free space loss model is improved, and the model is based on the analysis of building loss by the improved algorithm. The experimental results show that the path loss slope in this region is 23.51 dB/dec, and the accuracy of the improved free space loss model is 1.5 dB higher than that of the traditional free space loss model. The building loss model based on random forest is obviously superior to building loss model based on other algorithms.

Keywords Microcell Random forest Path loss Error

0 引 言

微蜂窝预测模型是基于容量节俭进行微蜂窝系统设计的工具,对于移动通信的应用研究有着巨大的指导意义。微蜂窝预测半径在 1.6 km 以内,微蜂窝预测模型大体可以划分为两类:理论模型和经验模型。理论模型是应用电磁波理论计算得到的确定性模型;经

验模型是统计性模型,是在大量测量数据基础上形成的。经验模型中比较经典的就是 Lee 微蜂窝预测模型,它根据信号的传播距离,在自由空间损耗模型上添加不同场景下的损耗参数,与其他模型相比,复杂度不高且适用各类场景,经实验验证,其准确度较高^[1-3]。Lee 微蜂窝建筑物损耗分析方法研究的是建筑物损耗与射频路径上穿过建筑物宽度之和的关系,然后应用最小二乘法构建成模型,但是在市区场景下,建筑物形状

收稿日期:2019-06-19。甘肃省高等学校创新团队项目(2017C-09);中国铁路总公司科技研究开发计划重大课题(2017X013-A)。陈明发 硕士生,主研领域:无线电传播模型。严天峰 教授。

的多样性和布局的复杂性等对分析建筑物损耗造成较大影响,使用 Lee 建筑物简化算法来计算建筑物损耗存在较大误差,而精准地得到射频路径上的信号穿过每个建筑物的宽度又很困难。

本文基于实地测试数据对 Lee 模型进行了建模以及仿真,建立路径损耗参数,改进自由空间损耗模型;针对 Lee 对与建筑物损耗分析误差较大的问题,应用机器学习中随机森林和 GBRT 两种集成回归算法对建筑物损耗数据进行分析,建立建筑物损耗模型;综合自由空间损耗模型与建筑物损耗模型使得市区微蜂窝模型更为精准。

1 Lee 微蜂窝原理

基于市区建筑物场景下衰减预测的基础模型为 Lee 微蜂窝模型。接收端接收的强信号是来自于多条经过建筑多次反射的反射波,而不是穿透建筑的无线电波, Lee 微蜂窝传播模型的预测公式为:

$$P_r = P_t - L_{LOS}(d_A, h_1) - L_B + G_A + G_a \quad (1)$$

式中: P_t 为发射功率; L_B 为建筑损耗, B 表示射频信号路径上信号穿过建筑物的总厚度; G_A 是射频路径上信号到达第一个建筑物的损耗,由发射端到建筑物的距离决定。本次实验中 $G_A = 0$, $L_{LOS}(d_A, h_1)$ 为视距 d_A 处的路径损耗,即自由空间损耗。距离发射端的自由空间损耗的公式为:

$$L_{LOS} = \begin{cases} 20 \log \frac{4\pi d_A}{\lambda} & d_A < d_f \\ 20 \log \frac{4\pi d_A}{\lambda} + \gamma \times \log \frac{d_A}{d_f} & d_A > d_f \end{cases} \quad (2)$$

式中: d_f 为近场距离。微蜂窝的近场距离公式为:

$$d_f = \frac{4h_1 h_2}{\lambda} \quad (3)$$

式中: h_1 为发射天线高度; h_2 接收天线高度; λ 为波长^[4]。

Lee 微蜂窝建筑损耗模型为:

$$L_B = \begin{cases} 0 & B < 0.3 \\ -(1 + 0.5 \times \lg(B/3.04)) & 0.3 \leq B < 7.62 \\ -(1.2 + 12.5 \times \lg(B/7.61)) & 7.62 \leq B < 182 \\ -(17.95 + 3 \times \lg(B/182)) & 182 \leq B < 900 \\ -20 & B > 900 \end{cases} \quad (4)$$

式中: L_B 表示建筑物造成的信号传播损失; B (单位为 m) 为 Lee 建筑物损耗模型传输路径上所有建物宽度之和。

2 集成回归算法原理

集成算法是指由多个弱分类器模型组成的整体模型,主要有两大类: Boosting 和 Bagging^[5-7]。本文中选择 Boosting 中最具代表性的算法 GBRT 和 Bagging 中最具代表性的算法 Random Forest 进行数据预测分析。

2.1 随机森林回归算法

随机森林^[8]是复合决策树的集成机器学习算法。随机森林使用 bootstrap 方法构建 n 个训练集,每个训练集对应生成一个决策树,总体就有 n 个决策树,因为每个决策树的数据集都不相同,所以每棵树又有少量区别。最后对所有的决策树的预测结果取平均减少预测的方差,提高在测试集上的性能表现。相比较单棵树训练过程,随机性主要体现在两个方面:每次迭代是在原始数据集中重新抽样获得不同的训练集;对于每一个树节点,考虑不同的随机特征子集来进行分裂。

随机森林的数学模型公式如下:

$$f_{RF}(x) = \frac{1}{N} \sum_{i=0}^T h(x; a_i) \quad (5)$$

式中: N 为回归树模型回归树的数量。

2.2 GBRT 回归算法

GBRT^[9]算法是一种迭代决策树的集成机器学习算法,该算法由多棵决策树组成,所有树的结论加起来就是最终答案,GBRT 的数学模型 f_{GBRT} 如下:

$$f_{GBRT} = \sum_{i=0}^T h(x; a_i) \quad (6)$$

式中: T 为建立模型的回归树的数量; $h(x; a_i)$ 为一个回归树模型; a_i 为每个回归树模型的参数。算法的核心思想是第 T 树学习的是前 $T-1$ 棵树结论和的残差,通过叠加所有决策树的结论不断逼近真实值。为了防止过拟合,GBRT 算法加入了 Boosting 这一项。

2.3 最小二乘法

最小二乘法^[10]是线性拟合中最常见的一种,是通过最小化误差的平方和寻找最为逼近原始数据的函数表达式,形如 $y = f(x, b)$,求出函数关系。

3 测试方案描述

3.1 测试地点

本次实验选取的测试地点为以兰州交通大学电信楼为中心,以半径为微蜂窝极限距离 1 600 m 的圆为测试范围进行测试计算。测试场景和测试点分布使用

ArcGIS^[11] 绘制如图 1 所示。



图 1 测试场景图

图中黑色圆点为发射基站位置,黑色三角形为移动接收点。本次实验中发射天线高度为 45 m,接收天线高度为 2.5 m,测试频率为 400 MHz,频率为 400 MHz 时的近场距离为 600 m,微蜂窝的极限距离为 1 600 m。

3.2 测试设备信息

表 1 为本文测试的设备信息。

表 1 测试设备信息

设备名称	型号
射频合成信号发生器	AV1485
发射天线	AH-8000
接收天线	AH-8000
频谱仪	FSH8
功率放大器	AV80244
馈线	D10-01-07-3M D10-07-07-10M
可编程直流电源	IT6332B

经过测试以及分析得出,功率放大器和天线增益为 33 dB,设备损耗为 8 dB,馈线损耗为 5.3 dB。

为了验证市区微蜂窝场景中建筑对信号传播的影响,测试得到的数据要尽可能详细,包括接收点的功率值、经纬度值、海拔值。

3.3 射频路径上建筑物宽度计算

如图 1 所示,研究建筑物对信号传播的损耗就是研究射频路径上信号穿过建筑物的宽度与建筑物损耗之间的关系。Lee 建筑物损耗模型提出了简便计算射频路径上信号穿过建筑物宽度之和的方法:将测试地区的建筑物按照建筑物布局和街道分布划分为如图 1 所示的 5 个建筑物区块,分别计算 5 个建筑物区块的建筑物占比,由穿过建筑物区块的距离和建筑物占比得到等价建筑物区块总宽度 B ,若射频信号穿过图 1

中的 2 和 4 区域,则:

$$B = a_2 \cdot L_2 + a_4 \cdot L_4 \tag{7}$$

式中: a_i 为各个区域的建筑物面积占比; L_i 为射频信号穿过各个区域的距离。

在同一个建筑物区块内的两个测试点,建筑物占比是相同的,测试点与发射点的距离是相近的,但是由 Lee 建筑物损耗模型得到的建筑物损耗与实际测得建筑物损耗差值却很大。针对这一缺点,本文提出如下椭圆区域计算方法。

由惠更斯-菲涅尔原理可以得到:对电波传播造成较大影响主要是位于第一菲涅尔区域的障碍物。因此,以收发两点间距离的一半为长半轴,以第一菲涅尔半径 F 为短半轴构建收发两点间信号的椭圆传播区域,计算该椭圆区域的建筑物面积占比 a ,如图 2 所示。

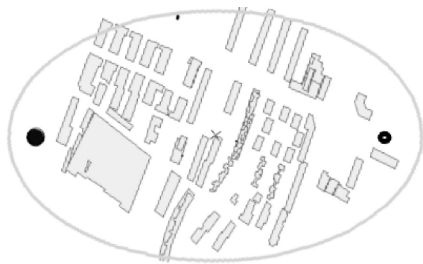


图 2 基于菲涅尔区域的椭圆区域示意图

得到等价建筑物区块总宽度 B :

$$B = L \times a \tag{8}$$

式中: L 为收发两点间的距离。

4 结果分析

4.1 改进自由空间损耗模型

由式(2)可以得到,自由空间损耗主要有两大参数决定,一是距离,二是路径损耗斜率 γ 。本文中使用的发射点 (E_T, N_T) 、接收点 (E_R, N_R) 的经纬度计算两点之间的距离,公式如下:

$$S = 2r \cdot \arcsin \left(\sin^2 \left(\frac{N_T - N_R}{2} \right) + \cos(N_T) \cdot \cos(N_R) \cdot \sin^2 \left(\frac{E_T - E_R}{2} \right) \right)^{1/2} \tag{9}$$

式中: $r = 6\,378.137$ km 为地球赤道半径。

测试地区的路径损耗斜率 γ 以前往往是参考类似地区的路径损耗斜率得到的,而每个地区的地形、建筑物分布、建筑物形状、街道走向又各不相同,因此引用其他地区的 γ 计算本地区的自由空间损耗往往会对近场距离以外的点造成自由空间损耗误差。本文将测试计算本地区的路径损耗斜率 γ ,从而提高自由空间损耗模型精度。将损耗数据进行线性分析,得到路径损耗斜率图如图 3 所示。

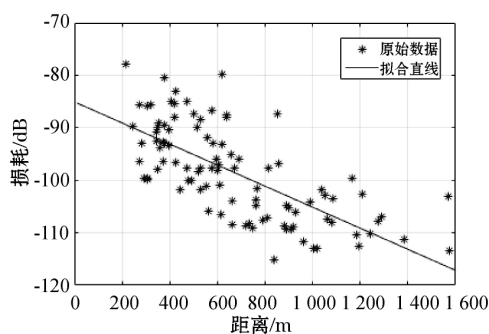


图3 路径损耗斜率图

取图3中的7组数据分别计算路径损耗斜率^[12],再取平均值:

$$\gamma = 23.51 \text{ dB/dec}$$

研究地区的微蜂窝路径损耗斜率为23.51 dB/dec。已知路径损耗斜率与距离,得到改进自由空间损耗模型与传统的自由空间损耗模型对比如图4所示。

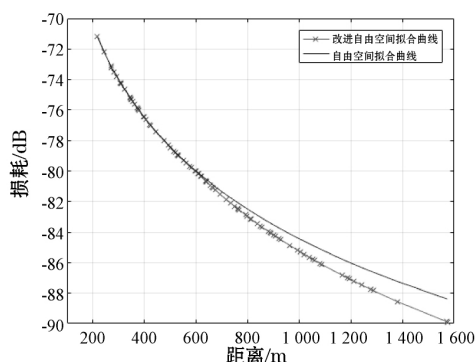


图4 自由空间损耗模型

可以看出,自由空间损耗误差主要在近场距离以外,误差在1.5 dB左右。

4.2 基于集成机器学习算法的建筑物损耗模型

将数据集按照3:7的比例划分为测试集和训练集,使用随机森林、GBRT两种集成算法构建市区建筑物损耗的预测模型,模型的构建流程如图5所示。

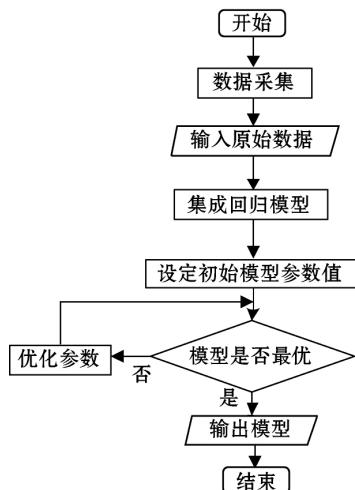
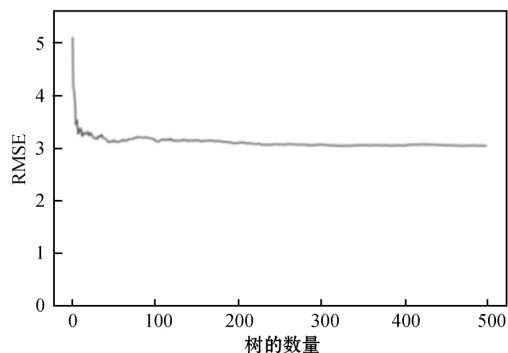


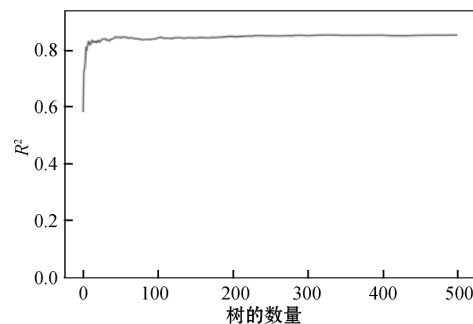
图5 模型构建流程图

4.2.1 参数设置

集成机器学习算法中回归树的数量是影响算法的预测性能的重要参数。本文依据模型在数据集上的评价参数(RMSE、 R^2)与回归树数量之间的关系曲线选择最佳的回归树数量进行模型训练,两种算法的关系曲线分别如图6和图7所示,其构建的市区建筑物场景的微蜂窝预测模型达到训练的最优。

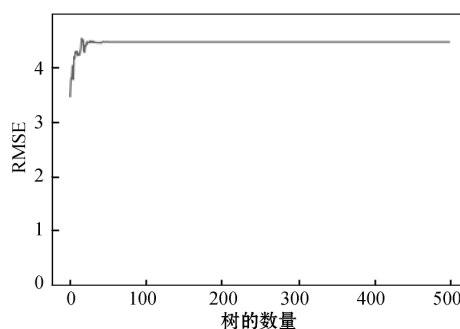


(a)

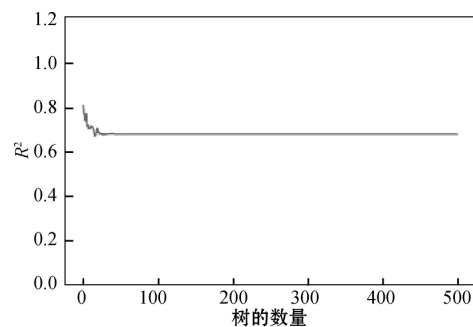


(b)

图6 随机森林回归树的选定



(a)



(b)

图7 GBRT 回归树的选定

由图 6 和图 7 得到构建集成损耗模型时的最佳回归树的数量如表 2 所示。

表 2 回归模型回归树的选定

回归模型	树的数量	R^2	RMSE
RF 模型	200	0.844 1	3.143
GBRT 模型	65	0.720 0	4.964

4.2.2 建筑物损耗建模对比分析

本文选取了 Lee 建筑物损耗模型与随机森林、GBRT 两种集成算法构建的模型进行对比,结果如图 8-图 10 所示。

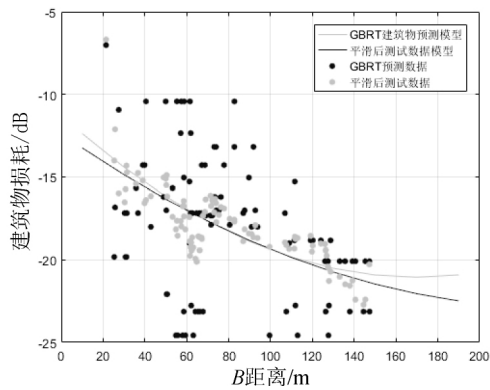


图 8 GBRT 建筑物损耗模型与实测模型对比

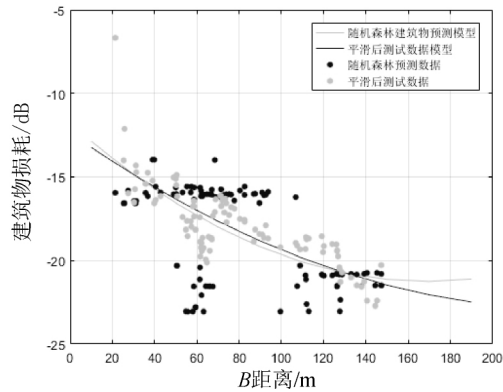


图 9 随机森林建筑物损耗模型与实测模型对比

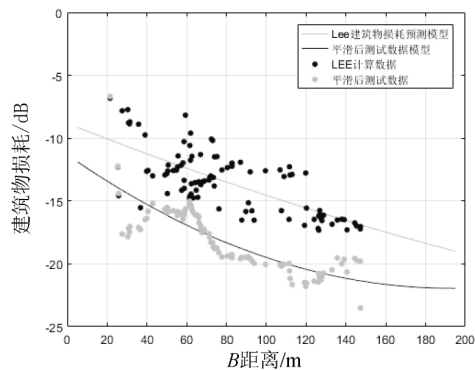


图 10 Lee 建筑物损耗模型与实测模型对比

图 8、图 9 中横轴上的距离是由式 (8) 计算得到的信号传播区域中信号穿过的总宽度; 图 10 中横轴上的

距离是按照 LEE 建筑物模型简化算法得到信号穿过建筑物的总宽度。由图 8-图 10 得到三种建筑损耗模型在测试集上的精度如表 3 所示。

表 3 回归模型精度对比

回归模型	R^2	RMSE
RF 模型	0.970 8	0.594 7
GBRT 模型	0.911 7	1.025 1
Lee 建筑物损耗模型	0.777 0	1.783 8

由上述对比结果可得: (1) 由集成回归算法构建的建筑物损耗模型普遍优于 Lee 建筑物损耗模型。(2) 当 RMSE 和 R^2 趋于稳定时,基于随机森林算法构建的模型的 RMSE 最小且 R^2 最大,但是基于 GBRT 算法构建的模型恰恰相反,并且 GBRT 模型预测值与实测值误差明显大于随机森林模型。综上,基于随机森林算法构建的建筑物损耗模型是稳定而且最优越的。

5 微蜂窝模型评估

结合三种建筑物损耗模型和改进自由空间损耗模型得到微蜂窝模型预测曲线在测试集上的表现以及误差累计分布曲线如图 11、图 12 所示。可以看出,基于随机森林构建的市区建筑物场景损耗模型最贴近测试集数据曲线,因此最终选择随机森林回归模型进行微蜂窝损耗分析。比较最终模型与实测损耗测试集数据发现,基于随机森林实现的市区建筑物场景微蜂窝模型的 R^2 为 0.918 5, RMSE 为 2.321 4,最大误差不超过 4 dB。造成本次实验误差的原因分析如下: (1) 由于接受天线架设的高度为 2.5 m,测试时天线的抖动可能形成测试值与实际值存在偏差; (2) 测试时来往行人和车辆的干扰也会造成实际值与测试值的偏差; (3) 发射点与接收点的地势高度差产生的有效天线高度增益或损耗也会使实验数据产生误差。

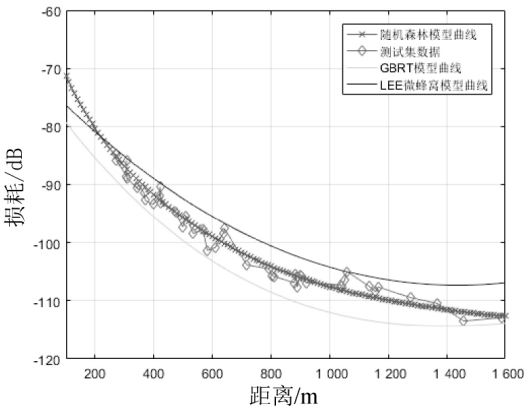


图 11 模型曲线对比

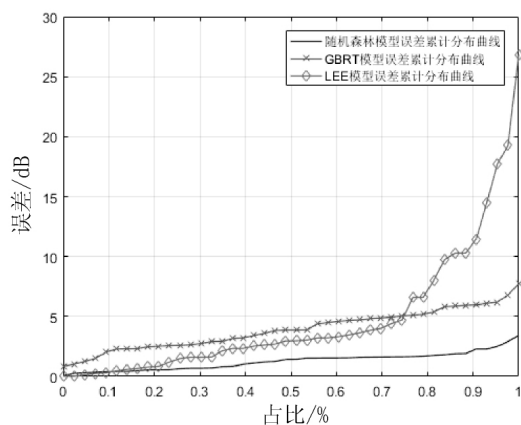


图12 误差累计分布曲线

6 结 语

在信号穿过市区建筑物的场景下,建筑物的阻挡使得电磁波在到达接收端时存在多路信号且多路信号之间存在相位差,多路信号的合成过程中形成了多径衰落,造成信号的衰减,同时也使得市区建筑物场景下微蜂窝的电波损耗预测难度增大^[13-15]。随机森林、GBRT与最小二乘法相比较具有很大的优势,可以充分应用数据含有的特征而又不担心过拟合。使用基于随机森林算法构建的微蜂窝模型在预测损耗时,预测误差不超过4 dB,预测精度较高,可以有效精准地应用于城市环境下损耗的预测。由于市区建筑物的多样性以及测试环境周围的复杂性,造成信测试点信号衰减的因素也不尽相同,因此使用随机森林算法对市区建筑物的形状和测试环境进行分类判别,再结合实测数据分析。未来将对市区建筑物微蜂窝预测模型进行更进一步优化,并构建更贴近实际的微蜂窝模型。

参 考 文 献

- [1] Lee W C Y. Mobile cellular telecommunications analog and digital system[M]. McGraw-Hill Professional, 1995.
- [2] 李彤, 姜媚, 王豪行, 等. 利用射线跟踪方法进行微蜂窝电波传播预测[J]. 通信学报, 1997, 18(11): 1-7.
- [3] 李萌, 孙恩昌, 张延华. 无线信道模型研究与展望[J]. 中国电子科学研究院学报, 2012, 7(4): 362-364.
- [4] Lee W C Y. Integrated wireless propagation model[M]. McGraw-Hill Professional, 2015.
- [5] 周志华. 机器学习[M]. 北京: 清华大学出版社, 2016.
- [6] 李航. 统计学习方法[M]. 北京: 清华大学出版社, 2012.
- [7] Wang H, Zhou K, Wu R F. A dynamic integrated model in machine learning and its application[C]//2010 Second WRI Global Congress on Intelligent Systems, 2010: 75-78.

- [8] 李国, 江晓东. 基于提升回归树与随机森林的风电功率集成预测方法[J]. 电力系统及其自动化学报, 2018, 30(11): 70-74.
- [9] Manna S, Biswas S, Kundu R, et al. A statistical approach to predict flight delay using gradient boosted decision tree[C]//2017 International Conference on Computational Intelligence in Data Science(ICCIDS). IEEE, 2017: 1-5.
- [10] 陈明晶, 方源敏, 陈杰. 最小二乘法和迭代法圆曲线拟合[J]. 测绘科学, 2016, 41(1): 194-197, 202.
- [11] 张岱琼. 基于 ArcGIS 的地图标注方法的研究[J]. 经纬天地, 2017(5): 42-50.
- [12] 薛会军. 基于电波传播路径损耗斜率的固定监测站覆盖范围预测分析[J]. 中国无线电, 2006(7): 55-57.
- [13] 王绍民, 赵道木, 林强. 对光波及电磁波衍射本性的认识 and 判断[J]. 浙江大学学报(理学版), 2000, 27(1): 109-113.
- [14] 张财元, 刘冰炎, 刘武超. 探究多径衰落信道中的信道估计与均衡技术[J]. 中国新通信, 2018, 20(18): 6-7.
- [15] Liu J. Wireless multipath fading channels modeling and simulation based on Sum-of-Sinusoids[C]//2016 First IEEE International Conference on Computer Communication and the Internet. IEEE, 2016: 165-168.

(上接第98页)

- [9] Mohamed A E A, Ahmed A, Aboul E H. Whale optimization algorithm and moth-flame optimization for multilevel thresholding image segmentation[J]. Expert Systems with Applications, 2017, 83(4): 242-256.
- [10] Wang J, Weng T Y, Zhang Q F. A two-stage multiobjective evolutionary algorithm for multiobjective multidepot vehicle routing problem with time windows[J]. IEEE Transactions on Cybernetics, 2019, 49(7): 2467-2478.
- [11] 李小川, 刘媛华, 王影歌. 求解最小最大 VRP 的精英反向学习鱼群算法[J]. 传感器与微系统, 2020, 39(2): 140-143, 152.
- [12] 王敏, 陈峰, 张磊石. 具有反向学习能力的串车调度算法研究[J]. 交通运输系统工程与信息, 2019, 19(2): 102-107, 115.
- [13] Xia Y K, Zhuo F. An adaptive tabu search algorithm for the open vehicle routing problem with split deliveries by order[J]. Wireless Personal Communications, 2018(1): 1-15.
- [14] Rajeev G, Raman M. A hybrid of ant colony and firefly algorithms (HAFA) for solving vehicle routing problems[J]. Journal of Computational Science, 2018, 25: 28-37.
- [15] Pop P C, Fuksz L, Marc A H, et al. A novel two-level optimization approach for clustered vehicle routing problem[J]. Computers & Industrial Engineering, 2018, 115: 304-318.