

# 一种优化支持向量机回归算法的印刷工序损耗值预测方法

彭来湖<sup>1, 2</sup>, 孙海涛<sup>1</sup>, 李建强<sup>3</sup>, 胡旭东<sup>1</sup>

(1.浙江理工大学浙江省现代纺织装备技术重点实验室, 浙江 杭州 310000;

2.浙江理工大学龙港研究院, 浙江 温州 325000;

3.浙江大学生物医学工程与仪器科学学院, 浙江 杭州 310000)

✉laihup@zstu.edu.cn; 2365046290@qq.com; wzcnlj@126.com; xdhu@zstu.edu.cn



**摘要:** 针对印刷生产中物料需求计划的损耗值采用经验值的问题, 提出一种优化支持向量机回归算法的印刷工序损耗值预测方法。通过皮尔逊相关系数量化特征值选取; 采用布谷鸟搜索算法优化支持向量机回归算法的超参数选取, 建立损耗预测模型; 为验证模型的优越性, 分别与不同的特征值选取方案、优化算法、回归算法的模型进行对比。实验结果表明该损耗预测方法具有更高泛化性和预测精度, 决定系数、平均绝对百分误差、均方根误差分别为0.995、0.005、1.969, 为解决后续相关问题提供了技术支持。

**关键词:** 印刷工序; 损耗预测; 皮尔逊相关系数; 支持向量机回归算法; 布谷鸟搜索算法

**中图分类号:** TP274+2 **文献标识码:** A

## A Prediction Method of Printing Process Loss Value based on Optimized Support Vector Regression

PENG Laihu<sup>1,2</sup>, SUN Haitao<sup>1</sup>, LI Jianqiang<sup>3</sup>, HU Xudong<sup>1</sup>

(1.Zhejiang Key Laboratory of Modern Textile Equipment Technology, Zhejiang Sci-Tech University, Hangzhou 310000, China;

2.Zhejiang Sci-Tech University Longgang Research Institute, Wenzhou 325000, China;

3.College of Biomedical Engineering and Instrument Science, Zhejiang University, Hangzhou 310000, China)

✉laihup@zstu.edu.cn; 2365046290@qq.com; wzcnlj@126.com; xdhu@zstu.edu.cn

**Abstract:** Aiming at the empirical value problem of material demand planning loss value in printing production, this paper proposes a method to predict the loss value of printing process by optimizing Support Vector Regression. Firstly, the Pearson Correlation Coefficient is used to quantify the eigenvalue selection. Then, the Cuckoo Search algorithm is used to optimize the selection of super parameters of Support Vector Regression algorithm, and the loss prediction model is established. Finally, to verify the superiority of the model, it is compared with models of different eigenvalue selection schemes, optimization algorithms and regression algorithms. The experimental results show that the proposed loss prediction method has higher generalization and prediction accuracy, and the determination coefficient, average absolute percentage error, and root mean square error are 0.995, 0.005, and 1.969 respectively. It provides theoretical support for the follow-up related problems.

**Keywords:** printing process; loss prediction; Pearson Correlation Coefficient; Support Vector Regression algorithm; Cuckoo Search

## 1 引言(Introduction)

印刷生产作为典型的离散型制造行业<sup>[1]</sup>, 逐渐趋于订单式生产模式。物料需求计划(Material Requirement Planning, MRP)作为生产过程中重要的物资计划管理模式<sup>[2-3]</sup>, 在物料采

购中起着指导作用。但是, 生产工序的损耗值主要通过工程经验直接确定, 致使物资计划管理存在随意性等问题。

目前, 国内外学者针对MRP在印刷生产应用过程中的工序损耗值确定问题的研究较为匮乏, 但损耗值预测方法的

研究较多。何丹萍等<sup>[4]</sup>和吴丽娜等<sup>[5]</sup>采用神经网络进行损耗预测，但主要采用经验公式设置超参数，致使模型泛化性较差。袁义生等<sup>[6]</sup>提出一种优化支持向量机的预测方法，但易收敛到局部最优。同时，以上研究在进行特征选取时，均由执行者根据工程经验所得，忽略了特征选取对预测结果的影响。

本文以预测精确为判定依据，考虑MRP在印刷生产中的具体应用，进行印刷工序损耗值预测方法研究。首先采用皮尔逊相关系数(Pearson Correlation Coefficient, PCC)量化分析影响因素真实性，其次通过支持向量机回归算法(Support Vector Regression, SVR)进行模型训练，最后采用布谷鸟搜索算法(Cuckoo Search, CS)进行超参数智能寻优。

## 2 印刷生产中的MRP(MRP in printing production)

### 2.1 MRP应用分析

在印刷工序的物料采购中，主要基于物料清单、物料基本信息、订单与工单情况、申购单与采购单情况等信息，通过分析MRP的实际应用过程，得到净采购和生产需求，实现物料动态平衡，并生成采购计划和生产计划，而在整个MRP的运算过程中，主要分为物料需求和损耗两个部分。

假设印刷生产过程中，在产出物的产需求数为 $oun$ ，根据物料清单关系，单一产出物的投入物B物料理论需求为 $md$ ，总物料需求为 $B_{need}$ ，投入物预计总损耗为 $B_{loss}$ 且B物料来自采购，则B物料需采购数 $B_{all}$ 可表示如下：

$$B_{all} = B_{need} + B_{loss} = oun \times md + B_{loss} \quad (1)$$

考虑到当前库存、物料申购情况，总物料需求不包括已有的物料数 $B_{have}$ ，而已有物料主要由B物料已领用数量 $claim$ 、可用库存数 $inven$ 、申购数 $prod$ 三个部分构成，其中可用库存数包括现有库存数 $inven_{now}$ 、安全库存数 $inven_{safe}$ 、非同一订单的锁库数 $inven_{lock}$ 、非同一订单的工单已分配数 $inven_{give}$ ，则B物料需采购数 $B_{all}$ 可表示如下：

$$\begin{cases} inven = inven_{now} - inven_{safe} - inven_{lock} - inven_{give} \\ B_{have} = claim + inven + prod \\ B_{all} = B_{need} + B_{loss} = oun \times md - B_{have} + B_{loss} \end{cases} \quad (2)$$

当 $B_{all} > 0$ 时，表示需要继续采购B物料，反之则表示当前的订单暂时不需要额外采购；而在目前的印刷生产过程中，物料损耗 $B_{loss}$ 的确定主要由采购人员在物料需求计算完成后，凭借工程实践经验以固定值或者按一定比例的形式直接确定，致使 $B_{all}$ 存在误差，影响了采购的精准性和印刷生产的正常进行。

### 2.2 印刷过程损耗分析

印刷生产包括构思设计、电雕制版、印前处理、调色对

版、印制小样及签样印刷6个过程，查看文献[7]—文献[9]及进行实际的现场调研发现，企业在印刷生产过程中，较大的损耗点主要出现在调色对版环节时对调色和套印的调机损耗，而不同工单的色数要求对应了不同的承印物损耗值。目前，一般企业主要通过设置上机损耗值和色数损耗控制该损耗值，但在实际生产过程中，由于采用废料去套色、打样等人为控制方法，导致设置的理论值无法满足实际需求。

印刷工序中所涉及的物料众多，本文基于文献分析，以及后续搭建模型在印刷工序中的普适性，主要选取承印物损耗作为印刷生产损耗的研究对象；而承印物的损耗与印刷上色过程中承印物正面上色数、反面上色数、正反面上色总数等具有一定联系。夏自由<sup>[8]</sup>提出，批量生产时根据不同的印刷数量自定义调整合理的放数，以减少换墨和换纸操作，可以减少物料损耗。故工单预产数量也可以影响印刷过程的损耗数量。

## 3 PCC-SVR损耗预测方法(PCC-SVR loss prediction method)

### 3.1 PCC-SVR损耗预测方法总体结构

根据印刷工序中损耗影响因素与损耗值之间的相关性，可将损耗预测问题转化为求解特征值和标签 $B_{loss}$ 之间的回归问题。为提高损耗预测结果的实用性， $B_{loss}$ 可换算成对应工单的同规格印刷件数值 $ls$ ，使得预测的损耗结果可以直接累加到工单预产印刷件数量上。通过对累加后的印刷件数量的物料理论需求计算，代替原来的物料理论需求和损耗，得到新的 $B_{all}$ ：

$$B_{all} = oun \times (md + ls) \quad (3)$$

损耗预测整体方案流程图如图1所示，利用PCC量化特征值和标签之间的关联性，采用SVR进行回归问题求解，采用CS对模型中的超参数进行最优解选取。然后通过优化后的SVR模型对数据进行训练，将训练后的损耗模型对印刷工序损耗进行预测，输入特征值，输出损耗预测值 $ls$ 。

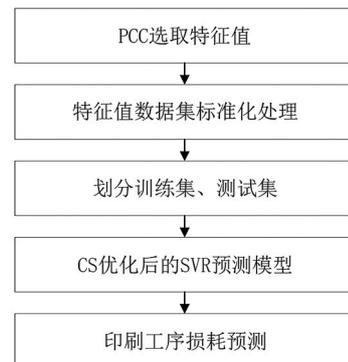


图1 损耗预测方案流程图

Fig.1 Flow chart of loss prediction scheme

### 3.2 SVR损耗预测模型建立

#### 3.2.1 特征值选取

在损耗预测模型搭建中，特征值的选取直接影响预测结果的准确性，为确定相关损耗影响因素的真实性，将该问题转化为求解每一个损耗影响因素与损耗值之间相关性的问题；而PCC作为衡量两个变量之间关系密切程度的方法，可用于该问题的求解。通过计算特征值中每一项与 $ls$ 之间的相关系数，当结果大于等于0.7时，表示该影响因素与损耗值高度相关。

#### 3.2.2 特征值标准化

不同的损耗因素之间数值大小相差较大，承印物正面上色数、反面上色数、正反面上色总数较工单预产数量在数值上会存在上百甚至上千倍的差距，易导致在进行模型训练时，影响损耗值目标结果，使模型无法学习其他的损耗影响特征。为了消除不同特征值的量纲对最终结果的影响，对数据集中特征值数据进行标准化处理，使特征值之间具有可比性。

#### 3.2.3 构建SVR损耗预测模型

支持向量机回归算法可用于解决回归问题，是一种常用于数据趋势预测的数学模型<sup>[10]</sup>，相比传统的算法，如正规方程算法、随机梯度下降算法、岭回归算法等，它具有更好的预测与拟合的效果，基本的函数表达式如下：

$$\begin{cases} f(x) = \sum_{i=1}^n (\alpha_i - \alpha_i^*) K(x_i, x_j) + b \\ 0 \leq \alpha_i \leq C \\ 0 \leq \alpha_i^* \leq C \end{cases} \quad (4)$$

式(4)中， $\alpha_i, \alpha_i^*$ 为拉格朗日乘子； $K(x_i, x_j)$ 表示满足Mercer定理的核函数； $C$ 表示松弛变量的惩罚系数。选取最常用的高斯核函数作为 $K(x_i, x_j)$ ，通过调整惩罚系数 $C$ 和高斯核系数 $\gamma$ ，降低损耗预测值 $f(x)$ 和 $ls$ 之间的误差，实现印刷工序损耗预测模型的搭建。

### 3.3 模型超参数最优解选取

布谷鸟搜索算法是一种通过模拟某些种属布谷鸟的寄生育雏，有效求解最优化问题的算法。在进行最优解搜索过程中，主要采用局部随机游走和全局探索性随机游走的平衡组合，由宿主鸟发现外来鸟蛋的概率 $P_a$ 控制，在解决局部最优问题和提高泛化性方面具有较好的效果。局部随机游走函数表示如下：

$$x_i^{t+1} = x_i^t + \alpha s \otimes H(p_a - \varepsilon) \otimes (x_j^t - x_k^t) \quad (5)$$

式(5)中， $x_i^t$ 是任意 $t$ 时刻第 $i$ 个点的位置； $\alpha$ 是步长比例因子； $s$ 是步长； $x_j^t$ 和 $x_k^t$ 是通过随机置换随机选择的两个解； $H(p_a - \varepsilon)$ 是阶跃函数； $p_a$ 是发现概率的固定值； $\varepsilon$ 是从

均匀分布中提取的随机数。全局探索性随机游走函数表达式如下：

$$x_i^{t+1} = x_i^t + \alpha L(s, \lambda) \quad (6)$$

因此，可将求解模型超参数的最优解问题转换为求解CS中布谷鸟最优鸟巢位置的问题，实现对模型的优化，优化流程图如图2所示。

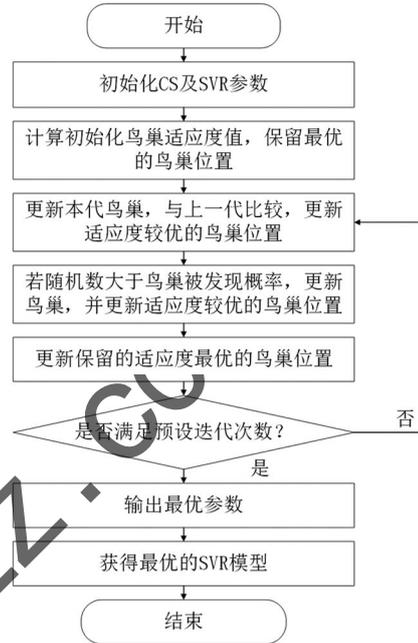


图2 CS优化SVR流程图

Fig.2 SVR flow chart of CS optimization

首先，初始化CS和SVR模型，设置鸟巢数量、最大迭代数量、寻优维度、宿主鸟发现外来鸟蛋的概率 $P_a$ 和SVR超参数等参数，采用预测值与真实值之间的决定系数作为CS的适应度函数，分别计算CS中每个鸟巢对应的适应度值，并选取其中适应度值最接近1的作为最优鸟巢位置 $x_{best}$ 。

其次，采用公式(6)更新鸟巢位置，得到新的鸟巢位置 $x_i^{t+1}$ ，并重新计算它们的适应度值，将本代适应度值与上一代比较，用适应度较优的鸟巢取代较差的鸟巢位置，得到新的鸟巢位置 $x_i^{t+1}$ 。同时，用随机数与 $P_a$ 进行对比，若大于 $P_a$ ，则通过公式(5)对 $x_i^{t+1}$ 位置进行改变，保留一组较优的鸟巢位置，更新 $x_i^{t+1}$ ，否则维持不变。

最后，从 $x_i^{t+1}$ 中选取最优鸟巢位置与 $x_{best}$ 进行比较，更新 $x_{best}$ ，并重复多次更新的 $x_{best}$ 操作，直至迭代次数达到预设值，则输出 $x_{best}$ 对应的超参数值作为损耗预测模型中的超参数最优解。

## 4 实验验证(Experimental verification)

### 4.1 特征值选取

选取温州某印刷企业的560条历史工单数据进行实验验证，将数据集中承印物正面上色数、反面上色数、正反面上

色总数、工单预产数量与损耗值通过PCC方法，分别得到对应的相关系数为0.965、0.714、0.984、0.996。根据皮尔逊相关系数的性质，可以得出选取的四个影响因素与损耗值存在高度的相关性，可作为损耗值预测模型的特征值进行模型的搭建。

### 4.2 印刷工序损耗预测结果分析

对特征值数据进行标准化处理后，设定CS中的鸟巢数为20个，最大迭代数为200次， $P_a$ 为0.25；SVR的惩罚系数 $C$ 为[0,200]，高斯核系数 $\gamma$ 为[0,50]。输入训练集，输出如图3所示的迭代次数与对应适应度值的曲线图。

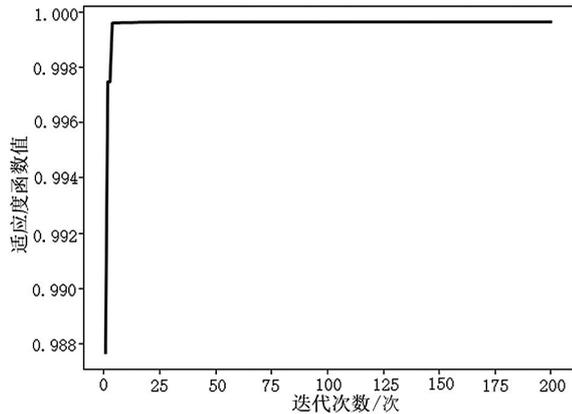


图3 迭代次数与对应适应度值曲线图

Fig.3 Curve of iteration times and corresponding fitness values

通过图3中的曲线输出，选取最高点对应的惩罚系数和高斯核系数作为SVR模型中超参数的设置，输入测试集得到图4中的预测结果。

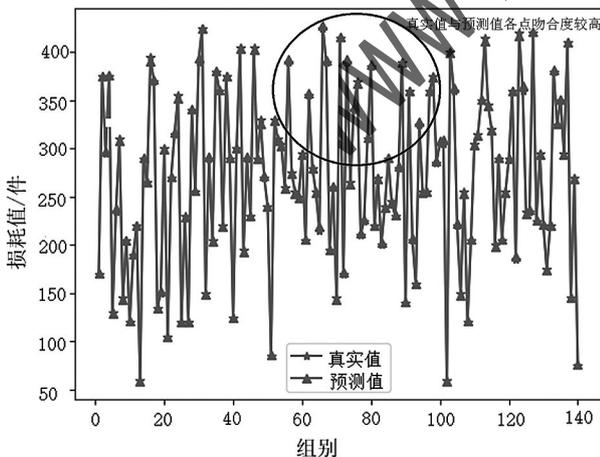


图4 优化后SVR模型损耗预测结果

Fig.4 Loss prediction results of optimized SVR model

从图4中可以看出，印刷工序的损耗值预测结果与真实值较为接近。为进一步衡量印刷损耗模型预测结果和真实结果之间的吻合度及模型预测偏差，选择决定系数、平均绝对百分误差和均方根误差作为模型的评价目标，通过实验得到相

应的结果分别为0.995、0.005、1.969。

### 4.3 对比实验

#### 4.3.1 与其他优化算法进行对比

为验证在解决印刷工序损耗值预测问题中，采用CS方法优化SVR的优越性，本文选取其他主流的优化算法，如遗传算法(Genetic Algorithm, GA)、粒子群优化算法(Particle Swarm Optimization, PSO)进行对比分析，通过设置相同的迭代数、种群数量和适应度函数等，得到迭代次数与对应适应度值的曲线图如图5所示。

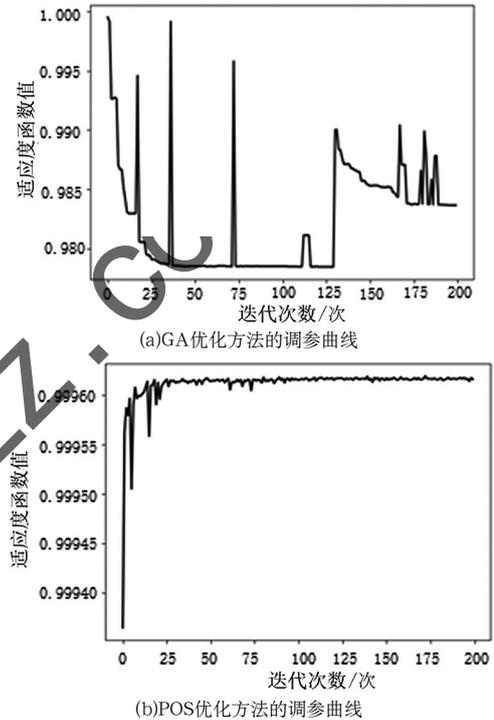


图5 其他优化方法的调参曲线图

Fig.5 Parameter adjustment curve of other optimization methods

从图5中可以看出，GA和PSO在迭代调参过程中，随着迭代次数的增多，整体曲线呈不断波动变化，针对多个局部范围内的超参数设定，分别存在多个不同的最优解，易使优化过程陷入局部最优的情况，从而导致模型超参数最优解选取有误。与图3对比，选用CS具有更好的泛化性和通用性，不易陷入局部收敛，更适用于解决本文所提的损耗预测问题。

#### 4.3.2 与其他回归算法对比

为进一步验证本文采用SVR解决印刷损耗回归问题的优越性，本文与其他主流的机器学习回归算法，如正规方程算法、随机梯度下降算法和岭回归算法进行对比分析，得到图6中的预测结果。从图6可以看出，三种回归模型相对SVR模型存在较多吻合度低的异常点，部分工单数据的损耗值预测与真实值相差较大，不适用于当下的应用环境。

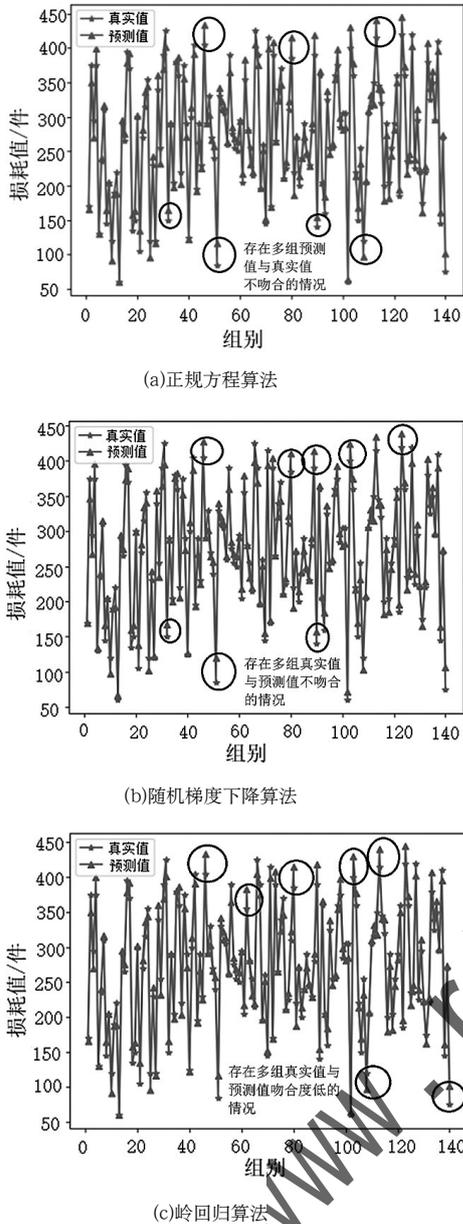


图6 其他回归算法的损耗预测结果

Fig.6 Loss prediction results of other regression algorithms

### 4.3.3 与其他特征值选取对比

为验证本文采用相关性较高的承印物正面上色数、反面上色数、正反面上色总数、工单预产数量四个特征值解决印刷工序损耗预测问题的优越性，根据四个特征值相关系数按照从大到小的顺序排序，以表1中选取特征值的分别搭建三个优化后的SVR回归模型。

表1 模型特征值选取

Tab.1 Model eigenvalue selection

模型	特征值
模型1	工单预产数量
模型2	工单预产数量、正反面上色总数
模型3	反面上色数、正反面上色总数、工单预产数量

模型输出如表2所示的评价指标对比结果，基于计算结果得到本文采用模型的决策系数大于模型1、模型2和模型3，以及平均绝对百分误差和均方根误差小于模型1、模型2和模型3的结果，因此具有更好的吻合度和精确度。实验验证了采用本文提出的特征值选取方案解决印刷工序中损耗值预测问题相比采用1—3个特征值更具优越性。

表2 评价指标对比

Tab.2 Comparison of evaluation indicators

模型	决定系数	平均绝对百分误差	均方根误差
模型1	0.551	0.238	59.647
模型2	0.986	0.013	5.176
模型3	0.989	0.013	5.038
本文采用模型	0.995	0.005	1.969

## 5 结论(Conclusion)

本文提出了一种基于PCC和CS优化SVR的印刷工序损耗值预测方法，用于解决MRP在印刷生产应用过程中印刷工序损耗值确定的问题，同时该方法量化了特征值选取的过程，兼顾了特征选取对预测结果的直接影响。

从实验数据可以看出，本文提出的损耗预测方法的决定系数、平均绝对百分误差和均方根误差分别为0.995、0.005、1.969，相较其他的优化算法、回归算法和特征值选取方案，具有更好的泛化性和预测精度。用该方法代替经验值设定，可以有效地提高物资计划管理的科学性、精准性和决策效率，为后续相关问题的研究提供了解决思路和技术支持。

## 参考文献(References)

- [1] 张亚洲.面向MES软包装加工质量管控关键策略与技术研究[D].北京:北京印刷学院,2022.
- [2] MILNE R, MAHAPATRA S, WANG C T. Optimizing planned lead times for enhancing performance of MRP systems[J]. International Journal of Production Economics, 2015, 167(6):220-231.
- [3] ZHU B, ZHANG Y Q, DING K, et al. Lot-sizing decisions for material requirements planning with hybrid uncertainties in a smart factory[J]. Advanced Engineering Informatics, 2022, 51(1):101527.
- [4] 何丹萍,徐卓成,曹惠云,等.基于机器学习和卫星图像的路径损耗预测[J].电波科学学报,2022,37(03):372-379.
- [5] 吴丽娜,何丹萍,艾渤,等.基于多层感知器神经网络的路径损耗预测研究[J].电波科学学报,2021,36(03):396-404.
- [6] 袁义生,钟青峰.基于支持向量机的光伏逆变器损耗预测方法[J].电力电子技术,2018,52(03):36-39.
- [7] 郑毅.A塑料软包装公司的作业成本法应用研究[D].武汉:中南财经政法大学,2020.

(下转第5页)