

基于机器学习的气流烘丝出口水分的预测模型的建立与研究

高 强, 邵春霖, 牛文巍

红云红河烟草(集团)有限责任公司昆明卷烟厂, 云南 昆明

收稿日期: 2023年11月16日; 录用日期: 2023年12月19日; 发布日期: 2023年12月28日

摘 要

在卷烟制丝加工过程, 气流丝、梗丝是卷烟配方中必不可少的填充物。气流烘丝设备作为制丝加工过程的重要工序点, 主要是利用高温气流对回潮后的叶丝进行高强度处理, 有效的降低香烟的焦油含量, 和使烟丝能够有较好的膨胀作用和青杂气去除效果。整个气流烘丝的特点就是: 工艺气体温度高, 热量快速传递到物料中, 干燥时间短, 瞬间脱水膨胀, 膨胀效果明显。所以说能够较好的控制烘丝出口水分对整个制丝过程有着举足轻重的作用。某卷烟工厂在制丝气流烘丝环节, 为了满足气流烘丝出口的冷床出口水分, 烘丝出口水分值的控制稳定性以及烘丝出口水分值的预测就显得尤为的重要, 本文通过基于机器学习的气流烘丝出口水分的预测模型的建立, 为生产过程提供准确的出口水分预测。

关键词

烘丝机, 水分控制, 机器学习, 模型预测

Establishment and Research of Prediction Model of Air Flow Drying Wire Outlet Moisture Based on Machine Learning

Qiang Gao, Chunlin Shao, Wenwei Niu

Hongyun Honghe Tobacco (Group) Co., Ltd. Kunming Cigarette Factory, Kunming Yunnan

Received: Nov. 16th, 2023; accepted: Dec. 19th, 2023; published: Dec. 28th, 2023

Abstract

In the process of cigarette silk making, air flow silk and stem silk are essential fillers in cigarette

文章引用: 高强, 邵春霖, 牛文巍. 基于机器学习的气流烘丝出口水分的预测模型的建立与研究[J]. 仪器与设备, 2023, 11(4): 416-423. DOI: 10.12677/iae.2023.114055

formulations. Air drying equipment, as an important process point in the process of silk processing, mainly uses high temperature air to carry out high strength treatment on the leaf silk after moisture return, effectively reduce the tar content of cigarettes, and make the tobacco can have a better expansion effect and green gas removal effect. The characteristics of the whole air drying wire are: High temperature of the process gas, rapid transfer of heat to the material, short drying time, instant dehydration and expansion, and obvious expansion effect. Therefore, the ability to better control the drying outlet moisture has a decisive role in the whole process of silk making. In the process of wire making and air drying in a cigarette factory, in order to satisfy the moisture at the cold bed outlet of air drying, the control stability of the moisture value at the wire drying outlet and the prediction of the moisture value at the wire drying outlet are particularly important. In this paper, the establishment of the prediction model of air drying outlet moisture value based on machine learning is used to provide accurate prediction of the moisture at the production process.

Keywords

Silk Dryers, Moisture Control, Machine Learning, Model Prediction

Copyright © 2023 by author(s) and Hans Publishers Inc.

This work is licensed under the Creative Commons Attribution International License (CC BY 4.0).

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



Open Access

1. 引言

昆明卷烟厂制丝线分为高档线(A线)、B\C线(薄板线), D线(气流烘丝线), 梗线(气流烘丝)一共5条生产线。其中梗线作为卷烟制丝必不可少的一条生产线, 同时作为生产加工中掺配的必要原料, 其中梗丝的生产过程控制对整个制丝的内在品质有着重要的影响。在梗丝的加工过程中, 其中一个关键指标就是梗加香出口水分。由于昆明卷烟厂的梗线设备工艺布局的特殊性, 梗烘丝出口通过吸风管道进行风力送丝连接梗加香设备, 中间环节存在着设备多, 环境影响大, 过程时间长等因素影响, 在烘丝出口与加香出口之间无水分控制设备, 他们之间存在着多变量关联因素, 相当于黑盒空间。对于加香出口水分的结果指标, 完全由烘丝出口水分而决定。对于整个烘丝出口水分值的输出结果的准确性就显得特别的重要。在整个烘丝出口水分值的控制过程主要由气流烘丝机的热风温度、工艺流量共同控制, 来实现烘丝出口水分稳定在设定中心值, 其中烘丝出口水份值通过PID反馈调节工艺流量进行闭环控制。所以烘丝出口水份值的准确性直接决定了加香出口水份。

根据梗线的生产安排, 每月的生产量约为60批左右, 2023年前9月的生产量约550批。其中梗加香出口含水率的CPK合格率在87%。其中13%的不合格率中, 烘丝出口水分的预测值不准确占比达到70%以上。所以针对烘丝出口水分的预测值的精准度提出了更高的要求。

通过选择合适的机器学习算法(如最邻近算法、随机森林等), 我们可以训练一个预测模型, 将历史数据与实际出口水分进行比较和分析, 以优化模型的准确性[1][2]。同时, 我们还可以利用交叉验证和模型调参等技术, 进一步提高模型的性能。一旦我们得到了一个准确可靠的预测模型, 我们将对其进行评估, 并与传统方法进行对比。通过比较模型的精确度, 我们可以评估其性能和优势。最终, 我们将该模型应用于实际烟草生产中, 为烟草生产过程提供准确的出口水分预测, 以优化烟草质量和市场竞争力。

2. 设备控制原理

气流烘丝机的工艺气体温度达到设定的工作温度时, 由振动输送设备将回潮后的松散烟丝均匀的输

送到进料气锁，到达进料管口，气锁对工艺气循环系统起密封作用，避免室内空气进入循环系统，影响系统温度。在设备内部，高温热风在内部收到风速的变化而形成负压，叶丝在气体的推动下载膨胀管道内向上运动，在这个过程中，每根烟丝被热风包围，烟丝可以快速脱水和定型。

某卷烟工厂的 D 线气流丝现有的加工模式条件下(默认工艺气温度相对恒定不变)，主要通过调整物料流量来进行出口水分的控制，设备在通过烘丝出口水分实际值与参数设定值的偏差，进行 PID 反馈调节，一直到烘丝出后水分的实际值与烘丝出后水分参数设定值相吻合，方可判断烘丝出后水分的稳定。根据工艺相关要求，烘丝后的冷床出口水分作为考核标准，例如：烘丝出口冷却水分是 $13\% \pm 3\%$ ，烘丝出口水分可能对应值是 14.5%，设备的反馈控制是由烘丝出口水分反馈调节。所以烘丝出口水分的预测值的设定的精准度就决定了最终的烘丝出口冷却水分值。所以如何构建一套烘丝出口水分预测模型就显得非常重要。

3. 原因分析

通过生产管理系统和 MES 的历史数据，挑选 5、6、7、8、9 月份的部分工作日，调查比对每天烘丝第一批的烘丝出口水分预设中心值与加香出口水分实际需求的烘丝出口水分中心值的偏离程度，统计相应的加香出口水分 CPK 是否合格(表 1)。

Table 1. Statistics on CPK qualification of flavored export moisture

表 1. 统计加香出口水分 CPK 合格情况

日期	当日第一批烘丝出口水分预设中心值	烘丝出口水分实际需求中心值	烘丝出口水分预设中心值与实际需求中心值的偏离程度	加香出口水分 CPK 是否合格
5.10	13.50%	13.90%	-0.40%	不合格
5.18	13.80%	13.60%	0.20%	合格
5.23	14.30%	14.20%	0.10%	合格
6.2	13.90%	13.50%	0.40%	不合格
6.6	13.60%	14.00%	-0.40%	不合格
7.10	13.80%	13.70%	0.10%	合格
7.12	13.60%	13.90%	-0.30%	不合格
7.13	13.50%	13.90%	-0.40%	不合格
7.19	13.80%	13.50%	0.30%	合格
7.20	14.20%	13.60%	0.60%	不合格
8.1	13.80%	14.00%	-0.20%	合格
8.8	13.70%	13.40%	0.30%	不合格
8.14	13.90%	13.50%	0.40%	不合格
8.16	13.80%	14.00%	-0.20%	合格
8.22	13.80%	13.70%	0.10%	合格
9.4	13.80%	14.20%	-0.40%	不合格
9.8	14.00%	13.70%	0.30%	不合格
9.12	13.90%	13.50%	0.40%	不合格
9.14	13.80%	13.60%	0.20%	合格

由于每天第一批的烘丝出口水分预设的中心值都是靠中控操作工的经验来决定的，所以每天第一批的烘丝出口水分预设中心值与实际需求中心值都会存在一定的偏离。而加香出口水分又是由烘丝出口水分所决定的，所以当预设值与实际值两者出现偏差时，就会对加香出口水分造成很大的影响，CPK 相应

也将会受到影响。从统计结果可以看出来，当烘丝出口水分预设中心值与实际需求中心值两者的偏离程度超过 0.3 时，加香出口水分的 CPK 也无法合格。所以烘丝出口水分的预设中心值与实际需求值的偏离程度和加香出口水分的 CPK 是否合格，有着极强的相关性。所以为了保证加香出口水分的 CPK 合格，就必须保证烘丝出口水分的预设中心值较为准确，偏差不能超过 $\pm 0.2\%$ 。最后确定造成梗加香出口水分偏离标准中心值过大的要因是因为烘丝出口水分预设中心值不准确，偏离实际需求中心值过大。

4. 机器学习训练与模型的建立

针对以上主要原因，可以提出具体的实施策略：主要针对烘丝出口水分的预测不够精准，导致加香出口水分的 CPK 不达标。通过选择合适的机器学习算法[3] (如最邻近算法、随机森林、catboost、LightGBM、卡尔曼滤波等预测算法)，我们可以训练一个较为理想的预测模型，将历史数据与实际出口水分进行比较和分析，以优化模型的准确性。同时，我们还可以利用交叉验证和模型调参等技术，进一步提高模型的性能。一旦我们得到了一个准确可靠的预测模型，我们将对其进行评估，并与传统方法进行对比。通过比较模型的精确度，我们可以评估其性能和优势。最终，我们将将该模型应用于实际生产中，为生产过程提供准确的出口水分预测，从而保障后续梗加香出口水分 CPK 的合格率。

4.1. KNN 算法分析

KNN (k-Nearest Neighbor), k 最近邻算法是一种常用的监督学习方法，其核心思想是，如果一个样本在特征空间中的 K 个最相邻的样本中的大多数属于某一个类别，则该样本也属于这个类别，并具有这个类别上样本的特性。

KNN 分类算法包括以下 4 个步骤：

- ① 准备数据，对数据进行预处理；
- ② 计算测试样本点(也就是待分类点)到其他每个样本点的距离；
- ③ 对每个距离进行排序，然后选择出距离最小的 K 个点；
- ④ 对 K 个点所属的类别进行比较，根据少数服从多数的原则，将测试样本点归入在 K 个点中占比最高的那一类。

如图 1，对目标水分进行预测，本实验中采用的数据级为处理过后的部分数据集，去除了数据波动较大的部分，只留下了稳定值。

预测结果如下：

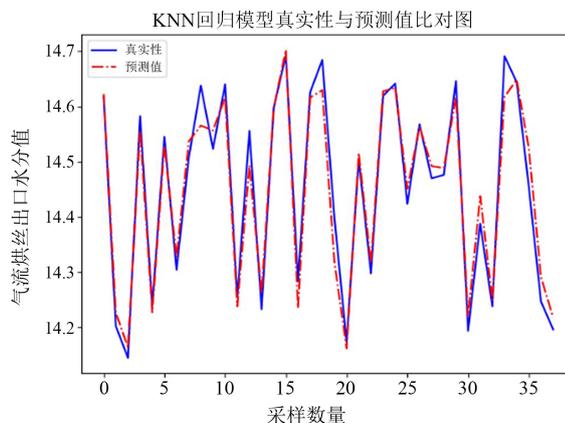


Figure 1. KNN algorithm analysis results
图 1. KNN 算法分析结果

模型的评估指标结果：均方误差为 0.0012，可解释方差为 0.961。

4.2. ACO 优化 LightGBM 回归方法

蚁群算法是一种用来寻找优化路径的概率型算法。用该方法对回归方法参数进行优化，再进行回归预测。

在算法初始化时，问题空间中所有的边上的信息素都被初始化为 T_0 。算法迭代每一轮，问题空间中的所有路径上的信息素都会发生蒸发，我们为所有边上的信息素乘上一个小于 1 的常数[4]。信息素蒸发是自然界本身固有的特征，在算法中能够帮助避免信息素的无限积累，使得算法可以快速丢弃之前构建过的较差的路径。蚂蚁根据自己构建的路径长度在它们本轮经过的边上释放信息素。蚂蚁构建的路径越短、释放的信息素就越多。一条边被蚂蚁爬过的次数越多、它所获得的信息素也越多。迭代②，直至算法终止。

LightGBM 是一个用于梯度提升决策树的框架。LightGBM 的方法流程可以分为数据准备，特征工程，参数设置，数据集加载，训练模型，提前停止策略，模型评估，参数调优，预测，结果分析和部署，几个主要步骤。

如果模型表现良好，可以将其部署到生产环境中，用于实际应用。

如图 2，对目标水分进行预测，本实验中采用的数据级为处理过后的部分数据集，去除了数据波动较大的部分，只留下了稳定值。

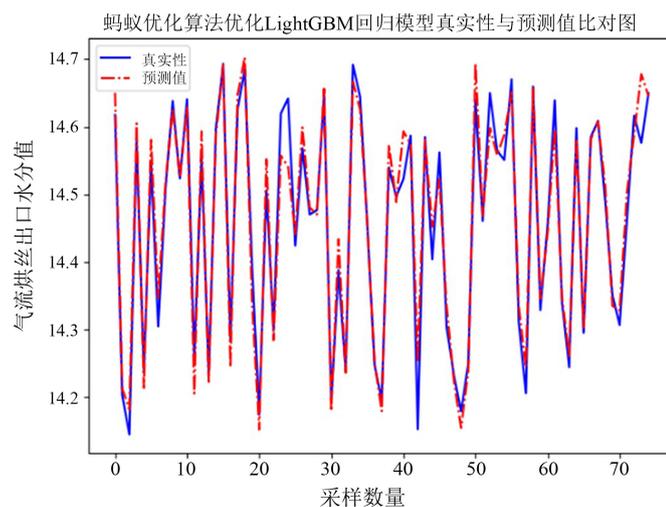


Figure 2. ACO optimized LightGBM regression analysis results
图 2. ACO 优化 LightGBM 回归分析结果

模型的评估指标结果：均方误差为 0.0012，可解释方差为 0.9593。

4.3. ACO 优化 Catboost 回归方法

CatBoost 是一种用于回归问题的梯度提升树算法，它特别适用于处理分类特征和高基数特征。可以用于预测数值型目标变量的值。CatBoost 可以帮助识别和理解数据中的模式、趋势和关联性。通过训练回归模型，可以了解特征之间的重要性，以及它们与目标变量之间的关系。

如图 3，对目标水分进行预测，本实验中采用的数据级为处理过后的部分数据集，去除了数据波动较大的部分，只留下了稳定值。

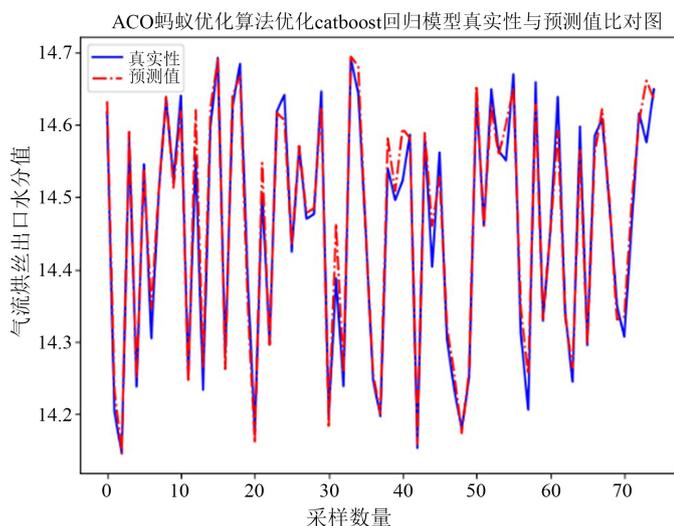


Figure 3. ACO optimization Catboost regression analysis results
图 3. ACO 优化 Catboost 回归分析结果

模型的评估指标结果：均方误差为 0.0008，可解释方差为 0.974。

4.4. ACO 优化随机森林回归方法

随机森林是一个包含多个决策树的分类器，并且其输出的类别是由个别树输出的类别的众数而定。随机森林对于回归问题的解决非常有效，它可以用于预测连续数值型目标变量，该方法是一种非参数化的算法，因此可以很好地适应非线性关系。它不对数据做出线性假设，因此可以捕捉到复杂的非线性模式。

如图 4，对目标水分进行预测，本实验中采用的数据级为处理过后的部分数据集，去除了数据波动较大的部分，只留下了稳定值。

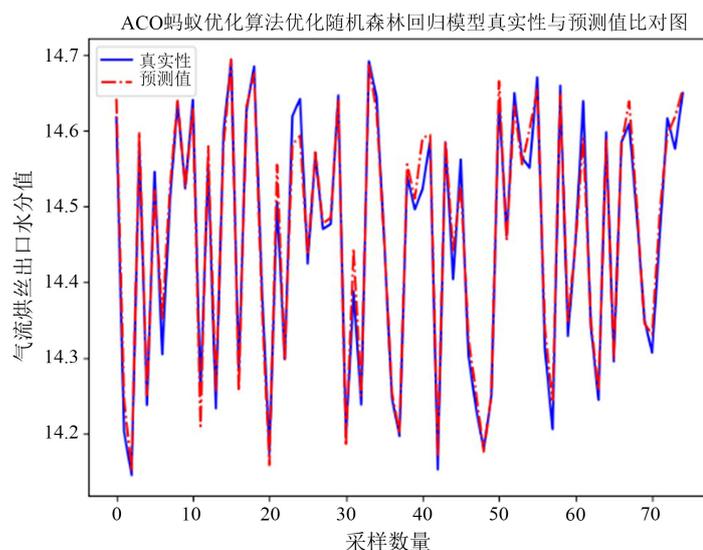


Figure 4. ACO optimization random forest regression analysis results
图 4. ACO 优化随机森林回归分析结果

模型的评估指标结果：均方误差为 0.0006，可解释方差为 0.9797。

4.5. 基于 RNN 的时序预测方法

RNN 通过在每个时间步骤上接收输入和前一个时间步骤的隐藏状态，来捕捉序列数据中的动态模式。在时序预测任务中，RNN 会根据已经观察到的序列数据，学习到序列中的模式和规律，并基于这些规律进行未来值的预测。

基于 RNN 的时序预测一般包含以下步骤，数据准备，网络建模，模型训练，模型预测。

在时序预测中，RNN 模型可以学习到序列中的长期依赖关系，并对未来的数值进行预测。

本次实验中的数据包含了数据集的全体数据，整合之后作为模型的输入(图 5)。

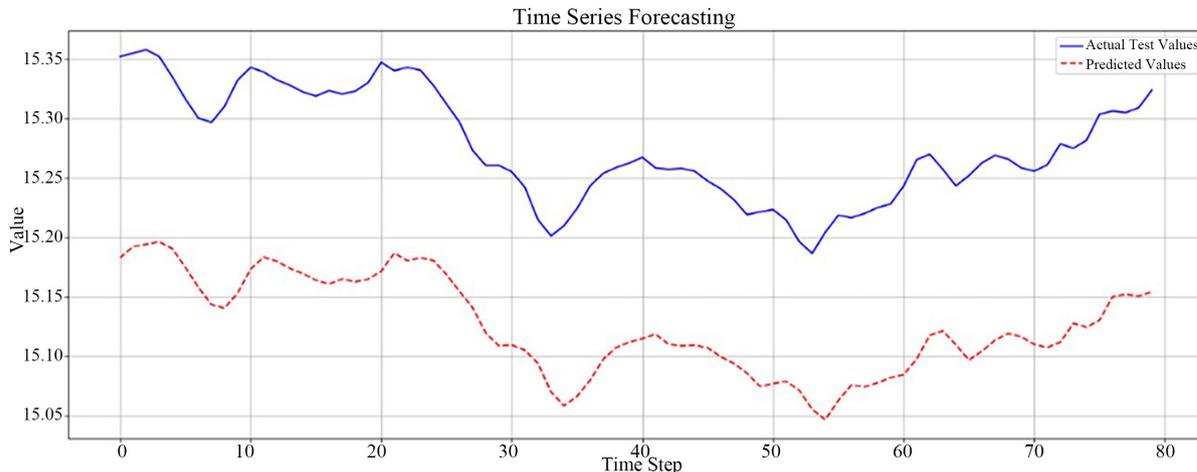


Figure 5. RNN based on time series prediction value and real value comparison chart

图 5. RNN 基于时间序列的预测值与真实值对比图

模型的评估指标结果：均方误差为 0.18，可解释方差为 0.9929。

4.6. 模型测试结果汇总

对于研究结果，采用平均绝对误差 MAE，均方误差 MSE，可解释方差， R^2 四个指标来衡量模型预测结果的准确度(表 2)。

Table 2. Summary of research results

表 2. 研究结果汇总

	MAE	MSE	可解释方差	R^2
ACO 优化 LightGBM 回归	0.0256	0.0012	0.9593	0.9587
ACO 优化 Catboost 回归	0.0205	0.0008	0.974	0.9721
ACO 优化随机森林回归	0.0188	0.0006	0.9797	0.9785
RNN 时序预测	0.180	0.05435	0.9929	0.9912

5. 效果验证

通过建模分析，除了工艺气温度，工艺流量，来料水分较大因素外，还有空气环境温湿度也对烘丝出口水分有较大影响，在建模分析的维度变量中，还缺少环境温湿度的影响分析模型。所以该模型目前的分析仅仅是对历史数据的分析建模，有一定的缺陷型。但是通过目前的模型预测，与实际生产拟合来看，目前的精准性对生产同样具有很好的指导作用。通过近一个月数据通过 9 月的生产记录来看，模型

预测烘丝出口水份值与实际的烘丝出口水分如下表 3、表 4 所示：

Table 3. Comparison between the moisture value predicted by different models and the actual moisture content at the dryer outlet
表 3. 不同模型预测烘丝出口水份值与实际的烘丝出口水分抽样对比

模型 批次	455		460		465		470		475		480	
	实际	预测										
RNN	13.157	14.184	14.230	15.032	14.484	15.069	14.338	14.970	15.653	15.792	14.155	15.133
ACO 优化												
LightGBM 回归	14.156	14.253	14.975	15.012	14.718	14.827	14.567	14.521	15.813	15.754	14.392	14.413
Catboost	13.484	13.763	14.463	14.573	14.816	14.826	14.567	14.521	15.812	15.754	14.392	14.413
随机森林	13.484	13.561	14.463	14.550	14.718	14.852	14.641	15.010	15.813	15.770	14.392	14.477

Table 4. Deviation of moisture content at airflow drying wire outlet of different model stem lines in September
表 4. 9 月不同模型梗线气流烘丝出口含水率偏差

模型 批次	450~454	455~459	460~464	465~469	470~474	475~479
RNN	0.080	0.077	0.144	0.138	0.091	0.082
ACO 优化						
LightGBM 回归	0.055	0.059	0.0751	0.065	0.063	0.051
Catboost	0.0553	0.057	0.0751	0.064	0.0633	0.051
随机森林	0.0708	0.071	0.110	0.110	0.089	0.078

根据对比模型预测与实际生产记录数据可以看出，两者的偏差均在 0.2% 以内，预测模型具有很好的指导意义。

6. 结论

通过选择合适的机器学习算法，如最邻近算法、随机森林、catboost、lightGBM、卡尔曼滤波等预测算法，将历史数据与实际出口水分进行比较和分析，以优化模型的准确性。同时，我们还可以利用交叉验证和模型调参等技术，进一步提高模型的性能。最终，我们将该模型应用于实际生产中，为生产过程提供准确的出口水分预测，从而保障后续梗加香出口水分 CPK 的合格率。

参考文献

- [1] 张楚安, 朱俊召, 王乐军, 等. 烘丝机入口水分控制模型的建立及应用[J]. 安徽农业科学, 2023, 51(9): 178-183.
- [2] 刘穗君, 刘颖, 赵万莹, 等. 薄板烘丝机出口含水率稳定性控制方法研究[J]. 中国烟草学报, 2022, 28(5): 41-46.
- [3] 江婷, 罗先喜. 基于数据驱动的松散回潮出口水分预测模型分析[J]. 机电工程技术, 2023, 52(4): 119-123.
- [4] 王丽芳, 丁姗, 李丽萍. 烘丝出口水分预测模型的建立[J]. 中国科技纵横, 2020(5): 58-59.