

基于PLS-SEM的烟草工艺参数关联关系分析研究

杨筱菡¹, 史益乐¹, 詹映²

¹同济大学数学科学学院, 上海

²上海创和亿电子科技有限公司, 上海

收稿日期: 2022年11月13日; 录用日期: 2022年12月3日; 发布日期: 2022年12月16日

摘要

本文研究如何发掘制丝流程中松散回潮、加料和切烘阶段中诸多工艺参数与各阶段以及最终输出产品的品质之间的关联性。通过对某烟厂制丝生产流水线上收集到的共121个工艺参数观察变量3064条记录进行清洗和变量筛选后, 选取了填充值作为最终输出指标, 建立形成性偏最小二乘结构方程模型展开分析探究。确定路径模式, 建立结构方程模型, 并给出模型评估结果, 以及提升填充值的参数调整优化建议。

关键词

制丝流程, 工艺参数, 关联发掘, 形成性偏最小二乘结构方程模型

Analysis on Association among Process Parameters of Tobacco Production Based on PLS-SEM

Xiaohan Yang¹, Yile Shi¹, Ying Zhan²

¹School of Mathematical Sciences, Tongji University, Shanghai

²Shanghai Micro Vision Technology Company Limited, Shanghai

Received: Nov. 13th, 2022; accepted: Dec. 3rd, 2022; published: Dec. 16th, 2022

Abstract

This research is aimed to explore the relationship between various process parameters in loose return process, casing process and drying process and the quality of each process and the final

output product in tobacco primary processing. After cleaning and variable screening of 3064 process parameter data of 121 indicators collected from a tobacco factory's tobacco primary processing line, we selected ratio of filled tobacco as the subject, determined a formative partial least squares structural equation model to carry out analysis and exploration. We designed path models and then established a structural equation model, and evaluation criteria and results as well as possible optimization suggestions are given.

Keywords

Tobacco Primary Processing, Process Parameters, Association Mining, Formative Partial Least Squares Structural Equation Modeling

Copyright © 2022 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

我国是世界上最大的烟草生产国，烟叶产量、卷烟产量与烟草行业从业人数均位列世界第一，烟草行业作为我国财政收入的一大来源，近年行业竞争愈发激烈。在卷烟生产中，工艺是关键，质量是核心，制丝流程的工艺水平成为诸多卷烟企业一较高下的决胜因素，对该流程的工艺参数分析将为工艺技术和生产品质的提升提供有力支持。

经过调研发现，烟厂坚持致力于引入新设备，新技术，改进过程质量管理体系，追求更高的产品质量，但实际应用效果却始终没有达到最优，海量的加工过程数据没有得到充分的处理和分析、统计过程控制与诊断(Statistical Process Control and Diagnosis, SPCD)理论没有落到实处、质量评价与分析过程仍主要依靠人为经验判断等，烟草的整个加工过程中质量损失较大、产品内在稳定性较低等一些有待改进。

在卷烟的三大加工工艺中，烟叶制丝流程是至关重要的一环，该流程产物的质量好坏决定了卷烟产品的品质高低。而制丝工艺主要包含松散回潮、加料、切烘等子过程，是一个多耦合、自相关、非线性、多干扰的复杂过程，对于该流程的参数分析和优化已经成为了一类具有挑战性的课题。因此，从制丝流水线参数数据出发的关联分析有助于探寻各阶段参数的关联关系，进而实现各阶段的参数优化，具有较高的研究价值与现实意义。

目前，有不少国内学者从不同角度切入该课题，结合不同的统计方法，尝试建立模型评估优化制丝过程的工艺参数。赵蓉蓉[1]和鲁宇童等人[2]采用正交试验设计方法，探究分析了松散回潮、加料和烘丝工序的工艺参数对叶丝结构、烟气、烟支的物理特性和感官质量的影响，并用直观分析、方差分析和综合平衡法进行参数的优化分析。赵佳成等人[3]从切丝工艺入手，对不同切丝宽度条件下的成品烟丝结构数据进行方差分析，之后利用灰色关联度分析法对其烟丝结构分布的稳定性进行综合评价。张刘渲楠[4]和雷振等人[5]利用烘丝机加工过程中的不同工艺参数的数据，通过 Pearson 相关性分析和主成分分析探究了烘丝机工艺参数的数据特征。何邦华等人[6]选取制丝工艺中 5 个关键参数进行均匀试验设计，之后根据各试验组检测的烟丝填充值、整丝率和碎丝率对工艺参数和质量指标进行 BP 神经网络建模，预测不同参数组合对应的烟丝物理质量。

现阶段针对制丝工艺参数的研究大多从松散回潮、加料、切烘中的某一个阶段出发，探讨该阶段内

参数调整对制丝流程产品质量的影响,即单一阶段的关联发掘。然而,对于三阶段串联后各阶段间参数关联关系的探究则并不多见。而神经网络等机器学习方法建模虽然可以有很高的预测精度,但是缺乏对工艺过程的机理分析,缺乏实际指导意义。

简言之,在这一课题上,诸多学者从不同方向进行了探索性研究,提出了许多可行的制丝流程优化思路。目前,仍有许多方法和模型还未加以尝试,这一领域的研究依然面临着不小的挑战。

结构方程模型(Structural equation modeling, SEM)是一种融合了因素分析和路径分析的多元统计技术。自 Joreskog 和 Wiley (1973)提出以来,众多学者不断深入挖掘这一模型的可能性,SEM 分析逐渐成为当今以社会科学和心理学为代表的诸多领域中极为重要的一种统计方法并广泛用于实证研究中。黄秋杰[7]通过结构方程模型分析了中国财险公司的盈利能力,发现具有直接影响的是企业规模,并提出了提高盈利能力的相关举措。李新杰等[8]利用 SEM 模型对中部六省的数据展开研究,评估了新生代农民工的心能、智能、体能和技能四个人力资源能力。申珅[9]从形成性偏最小二乘结构方程模型(简称 PLS-SEM)出发,探究了烟草种植、税收及消费等因素对烟草生产的影响,并给出有针对性的政策建议。

总体来说,SEM 分析经过数十年的发展,已具备较为坚实完善的理论体系,得到了广大研究者的认可和广泛应用。与其它研究相比,结构方程模型除了能够计算评估各指标间的关联程度,还能将各阶段参数进行串联,刻画完整的工艺流程以及其中的指标关系,并进行相关统计计算,是目前比较理想的处理研究数据的模型。

因此,本研究基于某烟厂制丝流程流水线上的工艺参数数据展开,从数据分析的角度出发,数据进行清洗后,将研究重点主要集中在松散回潮工序、加料工序和烘丝工序的工艺参数关联关系发掘,通过建立形成性 PLS-SEM 探索各工艺参数对制丝流程的主要品质衡量指标填充值的影响,尝试在现有工艺经验的基础上发掘制丝流程中各阶段工艺参数与输出指标的多重关联性。对这些指标影响显著的阶段和显著的工艺参数给出可能的调整策略,为实际生产过程中工艺参数的调整优化给出可行建议。

2. 数据

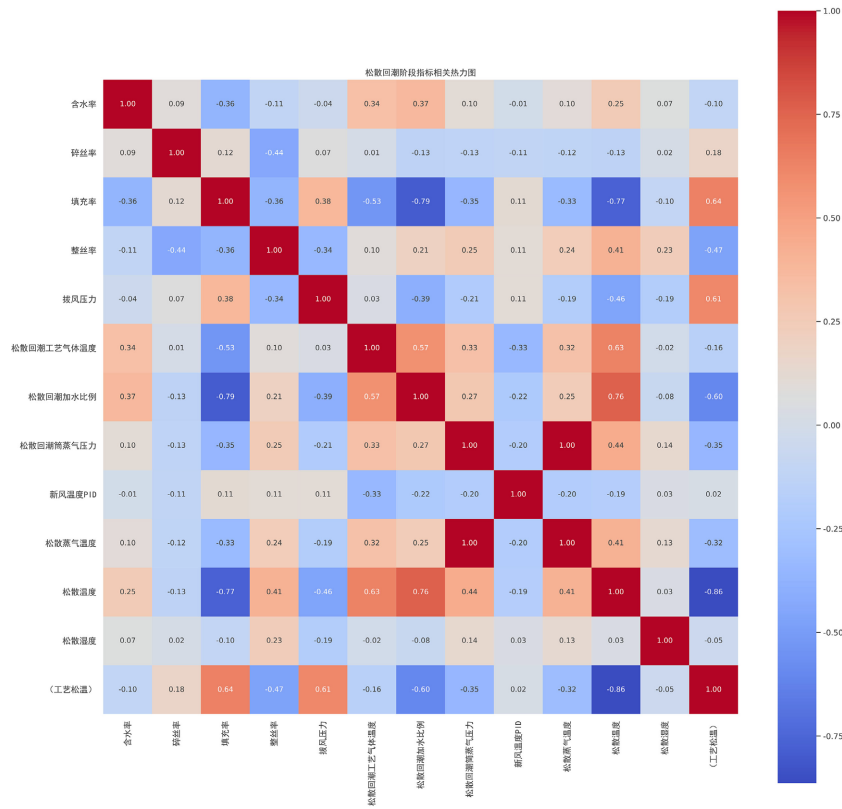
选取上海卷烟厂中华牌号制丝工艺线为研究对象,收集了 2019 年 1 月 1 日至 2020 年 9 月 27 日期间包括批次号、日期以及制丝工艺过程中松散回潮、加料和切烘三个阶段各类工艺参数共 121 个变量在内无缺失值的 3064 条生产记录。

考虑到原数据集内共 121 个工艺参数变量,直接使用原数据集进行建模分析,变量过多可能导致模型运算量过大,很多变量间的相关关系也会导致建模效果不稳健。基于此,在不损失过多信息的前提下必须精简建模的变量和数据集。删除描述批次号和日期时间的变量;计算数据的样本方差值,删除部分在生产过程中取值基本固定不变(变量标准化后的样本方差值 ≤ 0.05)的变量。结合异常值检测,在删除变量的异常值后,绘制变量的分布图,通过分布的卡方拟合优度删除分布情况较差的变量。对 3064 条工艺过程记录进行清洗和变量初步筛选后,依照制丝工艺过程中松散回潮、加料和切烘的工艺流程阶段,依次讨论各阶段变量间的(多重)共线性情况,通过计算变量 Pearson 相关系数和方差膨胀因子(VIF)值对变量间的(多重)共线性进行评估,最终确定各阶段的观察变量(具体见表 1)。

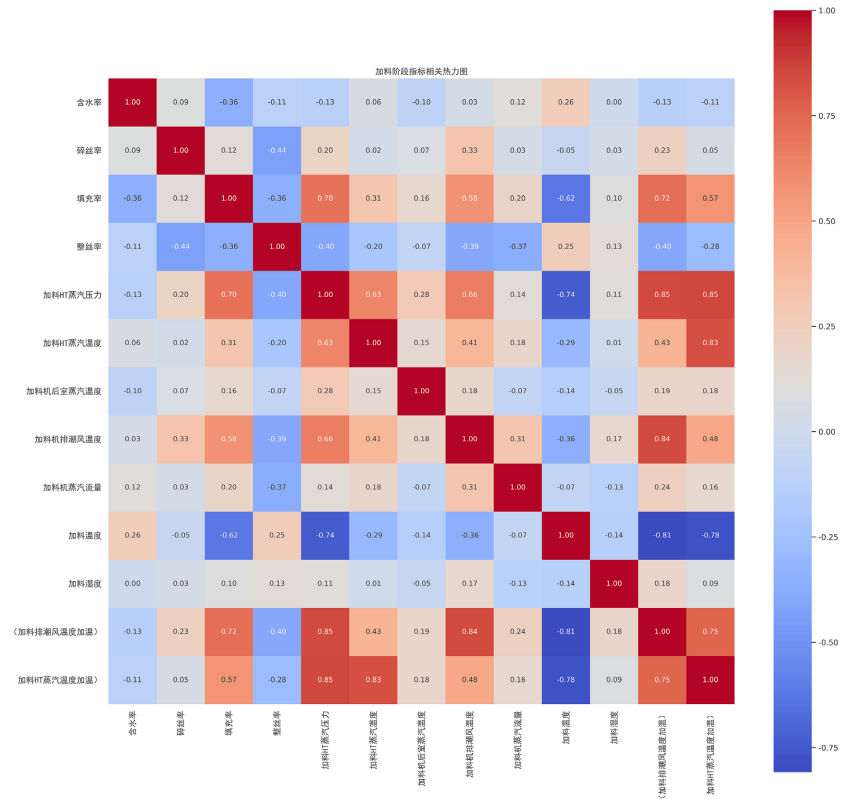
Diamantopoulos 和 Sigauw 等人[10] 2006 年指出,为保证建模效果保持在较高水平,最好将 SEM 模型中的观察变量间的 VIF 控制在 3.3 以下。经验证可得数据集中所有自变量指标的 VIF 值都小于 3.3。三个阶段的变量的相关系数热力图如下图 1,相关系数最大值不超过 0.41。

3. 方法与模型

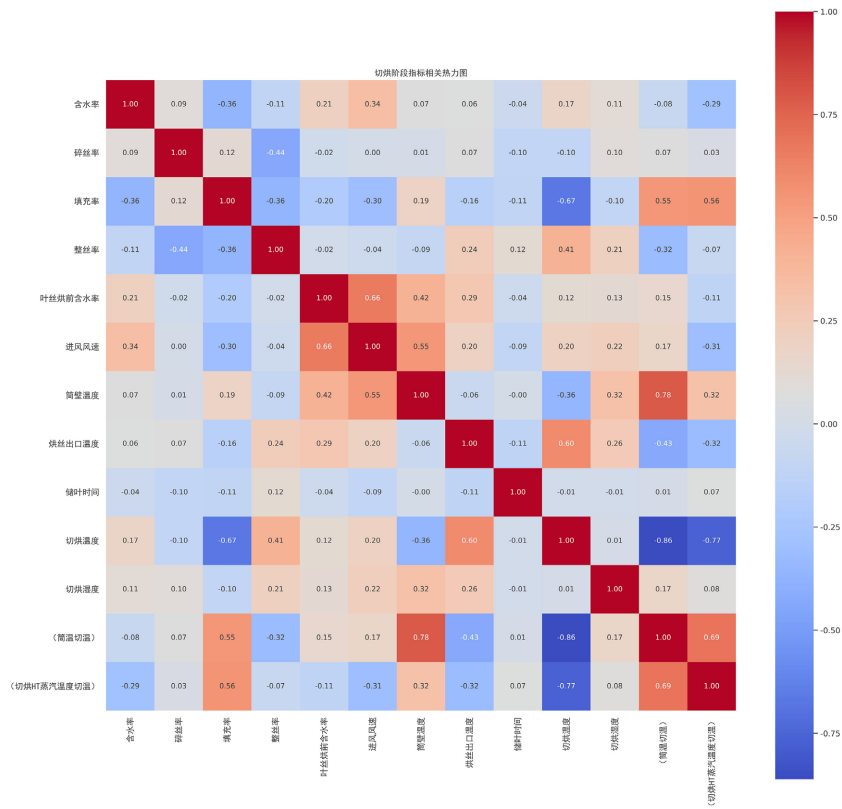
一个完整的 SEM 包含测量模型和结构模型两部分,前者用于刻画观察变量与潜在变量之间的关联情况,后者用于描述潜在变量之间的关联情况,其中测量模型又可以分为反映性模型(reflective model)与



(a)



(b)



(c)

Figure 1. Diagram of variable correlation coefficient
图 1. 变量相关系数图

形成性模型(formative model)两种不同情形。两种测量模型的主要差异在于观察变量和潜在变量之间的因果关系不同。对于反映性测量模型，潜在变量是“因”，观察变量是“果”，观察变量可以视为对应潜在变量含义的一种表现，同一潜在变量的观察变量之间具有高相关性且可互相替代；形成性测量模型则正相反，观察变量是“因”，潜在变量是“果”，观察变量解释对应潜在变量的一部分信息，进而同一潜在变量的观察变量之间不相关也无法互相替代。

另一方面，根据估计关联性时使用的算法不同，SEM 可分为基于变量间协方差结构方程模型(Covariance-Based SEM, 简称 CB-SEM)与 PLS-SEM 两类模型。CB-SEM 是基于变量间的协方差矩阵展开、使用极大似然估计法进行关联性估计的模型，更偏向于对理论的检验性分析，对于数据量和数据分布也有较高要求；PLS-SEM 则从变量的方差出发，使用以最小二乘回归为基础的统计方法对变量之间的关联性进行分析，在发掘数据间关联结构和预测方面有较好表现，对于数据量和分布的要求也相对较低。一般情况下，对于模型复杂度较高、涉及形成性构念、数据分布非正态或小样本的情形下，PLS-SEM 具有更好表现。

本研究中，考虑到潜在变量所包含的信息为各观察变量的结合，即观察变量是“因”，潜在变量是“果”，另一方面，本研究所使用的部分数据分布非正态，且研究属于探索式研究，所以选择形成性 PLS-SEM 作为研究方法。

模型的解释能力由 R^2 、外部权重、外部因子载荷来反映，通过路径系数及 Bootstrapping 检验结果反映模型变量间的因果关系。

4. 结果与分析

本研究在数据清洗的基础上,使用 Smart PLS 3 软件(3.2.9 版本软件)以填充值为输出指标,建立了形成性 PLS-SEM。采用结构方程模型探究了制丝过程中松散回潮、加料和切烘三个阶段的输出指标与最终输出指标的关联性;用测量模型发掘了各阶段的观察变量与潜在变量之间的关系。考虑到各阶段的观察变量数量较多,因此本研究使用重复指标方法,将各阶段输出指标设置为二阶外生潜在变量,在其下分别定义“温度”和“湿度”两个一阶外生潜在变量,用以具体探究各观察变量的作用情况。图 2 示意了本研究中使用的结构方程模型,其中潜在变量用圆圈表示,观察变量用矩形框表示。

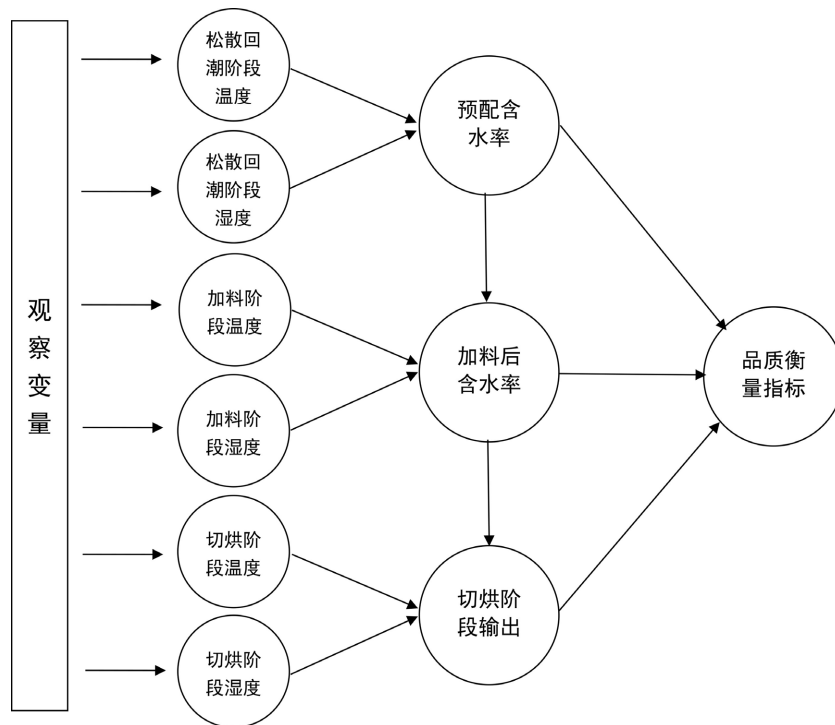


Figure 2. Diagram of structural equation model path
图 2. 研究用结构方程模型路径示意图

各阶段相关联的观察变量与潜在变量由表 1 给出。并给出了各阶段观察变量在一阶外生潜在变量和二阶外生潜在变量下的 VIF 值。从表中可知所有观察变量的 VIF 均小于 3.3, 可以认为观察变量间基本不存在(多重)共线性带来的影响, 这一结论与数据清洗阶段的讨论与处理结果是一致的。

Table 1. VIF value of variables
表 1. 观察变量的 VIF 值

二阶构念	一阶构念	观察变量	一阶构念下 VIF	二阶构念下 VIF
预配含水率	松散回潮阶段温度	新风温度 PID	1.14	1.163
		松散回潮工艺气体温度	1.257	2.197
		松散蒸汽温度	1.13	1.187
		松散温度等级	1.05	1.167

Continued

		拔风压力	1.255	1.59
预配含水率	松散回潮阶段湿度	松散回潮加水比例	1.216	2.197
		松散湿度	1.067	1.113
		加料 HT 蒸汽温度	1.242	1.28
	加料后含	加料机后室蒸汽温度	1.045	1.077
		加料机排潮风温度	1.311	1.515
加料后含水率		加料温度	1.193	1.22
		加料 HT 蒸汽压力等级	1.014	1.116
	加料阶段湿度	加料机蒸汽流量	1.03	1.185
		加料湿度	1.017	1.098
		切烘温度等级	1.141	1.158
	切烘阶段输出	烘丝出口温度	1.219	1.489
		切烘 HT 蒸汽温度切温	1.245	1.907
切烘阶段输出		筒壁温度	1.141	2.415
		进风风速	1.852	3.155
	切烘阶段湿度	储叶时间	1.01	1.027
		切烘湿度	1.054	1.262
		叶丝烘前含水率	1.775	1.975

通过计算可得,对于最终输出指标填充值,模型 R^2 与 R^2_{adjusted} 均为 0.688,参照 Chin [11]提出的结构方程模型 R^2 评估标准,这一数值大于 0.67,因此可以认为在现有模型中,“预配含水率”、“加料后含水率”与“切烘阶段输出”三个外生潜在变量对于最终输出指标填充值的解释度与预测精度都达到了较高水平。此外,作为对 R^2 的补充说明,本研究计算了填充值这一指标的预测相关度,计算结果 $Q^2 = 0.682 > 0$,也反映了现阶段建立的是一个能够较好预测填充值的模型。

基于良好的预测精度与预测相关性,本研究在 PLS-SEM 的基础上,利用原数据对填充值进行预测,从 10 折交叉验证的预测结果可以看到,对于填充值的预测结果中,均方根误差 $RMSE = 0.058$ 与平均绝对误差 $MAE = 0.046$ 都较低,可以认为本研究构建的形成性 PLS-SEM 对于填充值表现出高模型解释度以及高预测精度。图 3 给出了填充值模型路径图,图中包含各观察变量的外部权重、潜在变量的判定系数 R^2 以及潜在变量间的路径系数。

观察变量在一、二阶外生潜在变量下的外部权重以及显著性检验的结果如表 2 所示。

可以看到,一、二阶外生潜在变量下外部权重均不小于 0.1 且显著,这一类变量包括“松散回潮工艺气体温度”、“松散蒸汽温度”、“松散回潮加水比例”、“加料机排潮风温度”、“加料温度”、“切烘温度等级”、“切烘 HT 蒸汽温度切温”、“进风风速”等。这些指标反映了对应外生潜在变量较多信息,同时也具有较高的显著性。

在松散回潮阶段,“松散回潮工艺气体温度”与“松散回潮加水比例”分别是该阶段中对一阶外生潜在变量“温度”和“湿度”影响最显著的两个变量,其中后者对于二阶外生潜在变量“预配含水率”的影响最显著;在加料阶段中,对于“温度”和“湿度”影响最显著的分别是“加料温度”和“加料 HT 蒸汽压力等级”两个变量,“加料温度”对二阶外生潜在变量“加料后含水率”的影响更显著;切烘阶

段, “切烘 HT 蒸汽温度与切烘温度温差” 和 “进风风速” 在各自构成的一阶外生潜在变量下最显著, 前者对 “切烘阶段输出” 的贡献更显著。

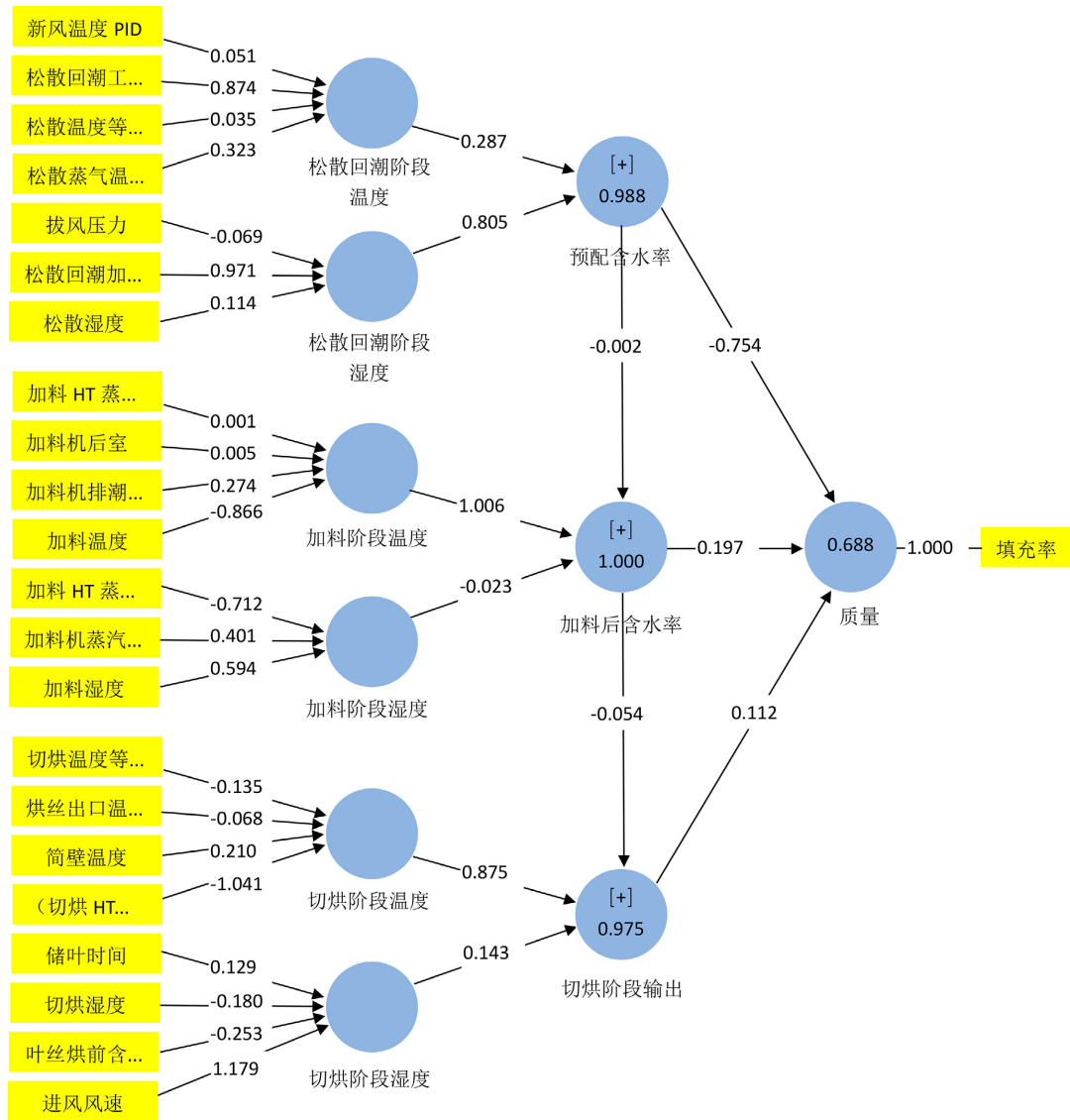


Figure 3. Path analysis model of variable filling value
图 3. 填充值路径分析模型

Table 2. Weighted value and significant test of variables
表 2. 观察变量的权重值及显著性检验

工艺阶段	构念层级	构念	观察变量	外部权重	T 统计量
松散回潮阶段	一阶	松散回潮阶段温度	松散回潮工艺气体温度	0.874	58.152**
			松散蒸汽温度	0.323	14.904**
	二阶	松散回潮阶段湿度	松散回潮加水比例	0.971	154.858**
			预配合水率	0.739	58.378**

Continued

加料阶段	一阶	加料阶段温度	加料机排潮风温度	0.274	16.524**
			加料温度	-0.866	79.215**
		加料阶段湿度	加料 HT 蒸汽压力等级	-0.712	13.122**
切烘阶段	二阶	叫料后含水率	加料温度	-0.867	73.583**
		切烘阶段温度	切烘 HT 蒸汽温度与切烘温度温差	-1.041	2.728**
	一阶	切烘阶段湿度	进风风速	1.179	2.639**
		二阶	切烘阶段输出	切烘 HT 蒸汽温度与切烘温度温差	-0.891

**表示高度显著。

Table 3. Path coefficients of significant variables and Bootstrapping analysis

表 3. 各显著变量的路径系数和 Bootstrapping 分析结果

路径	路径系数	VIF	T 统计量	p 值
松散回潮阶段温度→预配含水率	0.287	1.446	20.783	0.000
松散回潮阶段湿度→预配含水率	0.805	1.446	75.738	0.000
加料阶段温度→加料后含水率	1.006	2.573	258.923	0.000
切烘阶段温度→切烘阶段输出	0.875	2.603	33.791	0.000
切烘阶段湿度→切烘阶段输出	0.143	1.237	5.599	0.000
加料后含水率→切烘阶段输出	-0.054	2.313	2.568	0.010
预配含水率→填充值	-0.754	2.986	44.103	0.000
加料后含水率→填充值	0.197	2.99	11.495	0.000
切烘阶段输出→填充值	0.112	2.81	2.646	0.008

表 3 列出了潜在变量之间的路径系数以及路径系数的显著性检验结果，反映了各潜在变量之间的影响关系。在结构模型中，各路径的潜在变量之间的 VIF 均小于 3.3，因此认为潜在变量之间基本不存在(多重)共线性，结构模型后续各项分析与检验不会受到影响。

又，全部路径的路径系数 p 值 < 0.05，说明这些路径上的影响关系都是显著的。“加料阶段温度→加料后含水率”这一路径的路径系数高达 1.006。可以认为在以填充值作为最终输出指标的结构方程模型中，“加料阶段温度”对“加料后含水率”的影响是绝对显著的，即在加料流程中，温度对于这一阶段的输出质地起决定性作用。

此外，“松散回潮阶段湿度”和“切烘阶段温度”分别对“预配含水率”和“切烘阶段输出”的影响更为显著，即湿度与温度分别是决定松散回潮阶段和切烘阶段输出质地的关键因素。在松散回潮、加料和切烘三个阶段输出指标中，松散回潮阶段的输出质地对填充值造成的影响更大，体现在“预配含水率”对于填充值的影响更加显著，二者之间存在较强的负相关关系，即“预配含水率”每产生单位变化，填充值随之产生-0.754 单位变化。

综上测量模型和结构模型分析可得，填充值的显著影响路径依次为松散回潮加水比例→松散回潮阶段湿度→预配含水率→填充值。

5. 结论

本研究对某烟厂制丝流水线上获得的工艺参数数据进行筛选清洗后，以填充值为最终输出指标建立

了偏最小二乘法测度下的形成性结构方程模型, 发掘松散回潮、加料和切烘阶段对于最终输出指标具有显著影响的工艺阶段与具体参数。与其它研究相比, 本研究不仅仅关注某一阶段的工艺参数带来的影响, 而是着眼于制丝流程多阶段串联后显著关联关系的发掘。模型评估结果表明, 本研究建立的结构方程模型对于填充值具有较好的模型解释度/预测精度。松散回潮阶段的影响相较于其它阶段更显著, 表现为负向影响(-0.754)。具体而言, 该阶段显著工艺参数为“松散回潮加水比例”。因此, 生产过程中对于制丝流程产品的填充值的提升有必要重点关注松散回潮阶段的湿度参数的调整。总体而言, 湿度的影响要强于温度, 生产过程中需要更多关注湿度参数的调整。

基金项目

受上海市浦江人才计划资助项目(项目编号 2019PJD055)和高等学校大学数学教学研究与发展中心项目(项目编号 CMC20210411)资助。

参考文献

- [1] 赵蓉蓉. 卷烟制丝线关键工序工艺参数对烟丝质量的影响及工艺参数优化研究[D]: [硕士学位论文]. 郑州: 河南农业大学, 2017.
- [2] 鲁宇童, 张艺洁, 丁丹阳, 等. 基于正交设计的制丝工艺参数对叶丝质量的影响[J]. 湖南农业大学学报(自然科学版), 2018, 44(2): 27-31.
- [3] 赵佳成, 高辉, 王慧, 等. 切丝宽度对烟丝结构及其分布稳定性的影响[J]. 云南农业大学学报: 自然科学版, 2017(4): 668-677.
- [4] 张刘渲楠. 基于数据驱动的烘丝机工艺参数分析与优化[D]: [硕士学位论文]. 昆明: 昆明理工大学, 2018.
- [5] 雷振, 何金华, 黄传喜. 不同烘丝工艺参数对烟丝综合质量的影响[J]. 安徽农业科学, 2019, 47(5): 198-200+231.
- [6] 何邦华, 罗璟, 刘泽, 等. 烟丝物理质量的制丝工艺神经网络模型及其评价[J]. 云南农业大学学报: 自然科学版, 2016, 31(5): 874-879.
- [7] 黄秋杰. 基于结构方程模型的财险公司盈利能力影响因素分析[D]: [硕士学位论文]. 大连: 东北财经大学, 2012.
- [8] 李新杰, 孙泽厚. 新生代农民人力资源能力分析——基于 SEM 模型的中部六省实证研究[J]. 武汉理工大学学报(社会科学版), 2013, 26(6): 926-931.
- [9] 申琨. 基于 PLS-SEM 的烟草生产研究[J]. 中国市场, 2018(7): 26-28.
- [10] Diamantopoulos, A. and Siguaw, J.A. (2006) Formative versus Reflective Indicators in Organizational Measure Development: A Comparison and Empirical Illustration. *British Journal of Management*, **17**, 263-282. <https://doi.org/10.1111/j.1467-8551.2006.00500.x>
- [11] Chin, W.W. (1998) Issues and Opinion on Structural Equation Modeling. *MIS Quarterly*, **22**, 7-16.