

1-8-2019

Optimization of Material Release for Printed Circuit Board Template Based on Data Mining

Shengping Lü

1. School of Engineering, South China Agricultural University, Guangzhou 510642, China;;

Qiangsheng Yue

1. School of Engineering, South China Agricultural University, Guangzhou 510642, China;;

Liu Tao

2. School of Mechanical and Electrical Engineering, Anhui Jianzhu University, Hefei 230601, China;

Follow this and additional works at: <https://dc-china-simulation.researchcommons.org/journal>



Part of the Artificial Intelligence and Robotics Commons, Computer Engineering Commons, Numerical Analysis and Scientific Computing Commons, Operations Research, Systems Engineering and Industrial Engineering Commons, and the Systems Science Commons

This Paper is brought to you for free and open access by Journal of System Simulation. It has been accepted for inclusion in Journal of System Simulation by an authorized editor of Journal of System Simulation.

Optimization of Material Release for Printed Circuit Board Template Based on Data Mining

Abstract

Abstract: Data mining were employed for the optimization of material release of PCB (Printed Circuit Board) template. *PCB scrap ratio related parameters were specified and prediction model variables were chosen according to hypothesis test.* Multiple linear regression (MLR), Chi-squared automatic interaction detector, artificial neural network and support vector machine approaches for the prediction of scrap ratio were employed. *Evaluation indictors called as superfluous ratio, supplement release ratio and weighted sum of the two were presented;* the material release simulation was conducted and then the four approaches were compared and MLR was taken as the preferred one. *Adjust coefficient was introduced and optimized according to factory's cost model.* Finally, material release tool were developed and verified. Comparison results shown that superfluous and supplement release ratio has significant reduction which indicates that the approach can systematically reduce the cost of material release, waste of inventory, tardiness and so on.

Keywords

PCB, scrap ratio, data mining, prediction, optimization of material release

Recommended Citation

Lü Shengping, Yue Qiangsheng, Liu Tao. Optimization of Material Release for Printed Circuit Board Template Based on Data Mining[J]. Journal of System Simulation, 2018, 30(7): 2656-2665.

基于数据挖掘的印制电路样板投料优化

吕盛坪¹, 乐强生¹, 刘涛²

(1. 华南农业大学 工程学院, 广州 510642; 2. 安徽建筑大学 机械与电气工程学院, 合肥 230601)

摘要: 为了更准确确定 PCB (Printed Circuit Board) 样板投料, 基于车间历史数据开展挖掘分析。梳理报废率关联参数, 利用假设检验优选报废率预测建模参数。构建多元线性回归、卡方自动相互作用检测器、人工神经网络和支持向量机预测模型; 定义余数入库率和补投率及两者加权和评价指标, 开展投料仿真, 对比优选多元线性回归预测机制。引入调节系数, 结合企业成本模型进行优选; 开发实施投料控件并进行验证; 结果证明余数入库率和补投率较实际值均有明显降低, 可减少样板生产物料投入、库存浪费、补投拖期等成本, 为 PCB 样板投料优化提供新的参考手段。

关键词: PCB; 报废率; 数据挖掘; 预测; 投料优化

中图分类号: TN41; TP391.7

文献标识码: A

文章编号: 1004-731X (2018) 07-2656-10

DOI: 10.16182/j.issn1004731x.joss.201807028

Optimization of Material Release for Printed Circuit Board Template Based on Data Mining

Lü Shengping¹, Yue Qiangsheng¹, Liu Tao²

(1. School of Engineering, South China Agricultural University, Guangzhou 510642, China;

2. School of Mechanical and Electrical Engineering, Anhui Jianzhu University, Hefei 230601, China)

Abstract: Data mining were employed for the optimization of material release of PCB (Printed Circuit Board) template. PCB scrap ratio related parameters were specified and prediction model variables were chosen according to hypothesis test. Multiple linear regression (MLR), Chi-squared automatic interaction detector, artificial neural network and support vector machine approaches for the prediction of scrap ratio were employed. Evaluation indicators called as superfluous ratio, supplement release ratio and weighted sum of the two were presented; the material release simulation was conducted and then the four approaches were compared and MLR was taken as the preferred one. Adjust coefficient was introduced and optimized according to factory's cost model. Finally, material release tool were developed and verified. Comparison results shown that superfluous and supplement release ratio has significant reduction which indicates that the approach can systematically reduce the cost of material release, waste of inventory, tardiness and so on.

Keywords: PCB; scrap ratio; data mining; prediction; optimization of material release

引言

PCB (Printed Circuit Board)是电气和电子设备



收稿日期: 2017-04-20 修回日期: 2017-09-11;
基金项目: 国家自然科学基金(51605169), 广东省自然科学基金(2014A030310345);
作者简介: 吕盛坪(1982-), 男, 湖南新邵, 博士, 副教授, 研究方向为生产计划、工艺规划与调度优化、工业数据挖掘等。

的基体^[1]。客户个性化的需求使得 PCB 样板订单大量增加, 相应的生产模式也从传统的大规模批量生产转化为以大规模个性化生产为主要特征的长尾生产^[2]。更精准地确定各订单投料是影响车间上下游综合成本的关键。合理确定每个订单的投料面积将消减车间物料、生产、库存、回收处理和拖期等成本。但各订单具有个性化的功能特征, 所需要

<http://www.china-simulation.com>

• 2656 •

经历的工艺流程也不同, 确切的投料面积在生产前难以提前确定。车间投料人员通常基于不同层板的历史报废率均值计算确定各订单的投料面积, 调研中发现人工投料易导致车间超投和补投均较高。

剩余个性化 PCB 样板只能置于库存或直接销毁, 这导致物料、生产、存储和销毁成本的浪费。通过频繁的补投可以降低面积剩余, 但补投将带来车间在制品种增加和调度的不稳定, 影响车间生产周期和订单准交率。

显然, 同时减少超投和补投是消减车间成本的全局优化策略。但车间目前难以平衡这对 Pareto 目标, 主要以控制超投为主。寻求一种更优化的投料策略, 同时降低这两个目标具有重要的科学和工程实际意义。在此提出采用数据挖掘的手段对 PCB 样板进行报废率预测和投料优化。

数据挖掘是采用一系列的技术以抽取数据中蕴藏的有价值信息, 现已成功应用于工业、农业、商业等诸多领域^[3]。近年来, 随着大数据和云制造的兴起, 相关学者针对大数据的制造范式^[2]、大数据资源的云服务组合^[4]、基于大数据的智慧工厂框架构建^[4]等宏观角度开展了较深入研究。但与 PCB 的精益制造结合的数据挖掘研究相对较少。本研究从企业实际需求出发, 基于数据挖掘手段对其 PCB 样板投料进行优化, 主要内容包括 3 个方面: 1) 梳理 PCB 报废率相关参数, 利用假设检验和 MLR (Multiple linear regression) 参数检验机制优选 PCB 报废率预测建模参数; 2) 利用 MLR、CHAID (Chi-squared automatic interaction detector)、ANN (Artificial neural network) 和 SVM (Support vector machine) 等预测机制开展报废率预测; 结合 PCB 产品特点, 定义相应评价指标, 对比优选预测机制; 3) 引入调节系数, 结合企业成本模型进行优选; 开发实施投料控件并进行实施验证。

1 样本和方法

1.1 整体流程和抽样

本文整体流程如图 1 所示。其中数据来源于

广州某公司的 ERP (Enterprise Resource Planning) 系统。抽取 51 192 条有效记录, 其中下单时间在 2014-01-01~2015-12-31 之间的 40 526 条记录用作训练样本集(S1); 第一组测试样本(S2)为下单时间在 2016-01-1~2016-03-31 之间的 3 726 条记录, 第二组测试样本(S3)为订单下单时间在 2016-04-01~2016-07-31 的 6 940 条记录, 各组样本之间互不重叠。

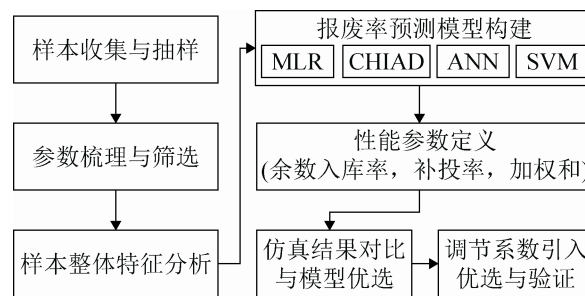


图 1 整体流程

Fig.1 Main flowchart

1.2 参数梳理与筛选

综合 ERP 中现有参数和数据预分析, 利用继承、派生、转换等方式, 共梳理与报废率具有较强关系的参数 53 个。为减少分析复杂度并提高建模效率, 以寻找对报废率具有积极贡献的重要参数^[6], 在此对离散和连续型并分别采用 F -检验和 t -检验进行筛选, 剔除连续型变异系数过小或离散型某一取值比例过大的参数。然后基于 MLR 的参数 t -检验剔除对报废率影响不显著参数, 最终保留 26 个参数(具体在 3.1 节讨论)。相应参数集合如表 1 所示。结合现有导出数据, 确定各数据范围, 并在“描述(取值范围)”列给出。

1.3 预测方法

统计分析和机器学习是常用测机制^[3]。MLR 作为常用的统计分析方法通过构建一个因变量与多个自变量的线性函数进行预测, 其中模型未知参数基于样本估计获得^[3]。

表1 参数规范
Tab. 1 Parameters specification

序号	参数名称	符号	描述(取值范围)	p-值
1	板厚	<i>Pt</i>	PCB 板厚(0.3~6.1)	<0.001 ^t
2	层数	<i>Ln</i>	PCB 铜箔层数(2~10)	<0.001 ^f
3	板镀次数	<i>Plfr</i>	板镀工艺频次(0~4)	<0.001 ^f
4	工序数	<i>Noo</i>	完成 PCB 板所需经历工序数(4~55)	<0.001 ^f
5	半固化片数	<i>NPP</i>	层压时所采用的半固化片总数(0~20)	<0.001 ^f
6	是否光电板	<i>Photb</i>	光电板为 1, 反之为 0.	<0.001 ^f
7	是否高频板	<i>Highfb</i>	高频板为 1, 反之为 0.	<0.001 ^f
8	是否 IPCIII 标准	<i>IPCIII</i>	采用 IPCIII 验收标准时为 1, 反之为 0	<0.001 ^f
9	内层最小线宽	<i>Mwil</i>	各层芯板中线宽最小值(0.9~137.8)	<0.001 ^t
10	内层最小线间距	<i>Msil</i>	各层芯板中线间距最小值(1.0~99)	<0.001 ^t
11	外层最小线宽	<i>Mwol</i>	上下层铜箔上线宽最小值(3~140)	<0.001 ^t
12	外层最小线间距	<i>Msol</i>	上下层铜箔上线间距最小值(3~140)	<0.001 ^t
13	是否有减薄铜	<i>Crd</i>	采用减薄铜工序时为 1, 反之为 0	<0.001 ^f
14	残铜率均值	<i>Arcr</i>	内层铜箔各层残铜率均值(0.15%~95%)	<0.001 ^f
15	是否内层干膜	<i>Dfil</i>	采用了内层干膜工艺为 1, 反之为 0	<0.001 ^f
16	是否有负片电镀	<i>Nflp</i>	有负片电镀为 1, 反之为 0	<0.001 ^f
17	总孔数	<i>Noh</i>	总钻孔孔数(0~116 406)	<0.001 ^f
18	沉铜通孔厚径比	<i>Hddr</i>	电镀前孔径尺寸比(0.816~12)	<0.001 ^f
19	是否需要二钻	<i>Secd</i>	需要二钻为1, 反之为0	<0.001 ^f
20	是否有树脂塞孔	<i>Phwr</i>	有树脂塞孔为 1, 反之为 0	<0.001 ^f
21	是否有背钻	<i>Bcdr</i>	有背钻为 1, 反之为 0	/
22	是否埋盲孔板	<i>Bbuv</i>	是埋盲孔板为 1, 反之为 0	/
23	有铅喷锡	<i>Haslwl</i>	采用有铅喷锡表面处理为 1, 反之为 0	<0.001 ^f
24	无铅喷锡	<i>Haslol</i>	采用无铅喷锡表面处理为 1, 反之为 0	<0.001 ^f
25	护铜剂	<i>Osp</i>	采用有机保焊膜表面处理为 1, 反之为 0	<0.001 ^f
26	图镀铜镍金	<i>Cnapp</i>	采用图镀铜镍金表面处理为 1, 反之为 0	<0.001 ^f
27	镀金手指	<i>Gfp</i>	采用镀金手指表面处理为 1, 反之为 0	<0.001 ^f
28	电镀硬金	<i>Godp</i>	采用电镀硬金表面处理为 1, 反之为 0	<0.001 ^f
29	电镀软金	<i>Snap</i>	采用电镀软金表面处理为 1, 反之为 0.	/
30	化学镍钯金	<i>Imnpa</i>	采用化学镍钯金表面处理为 1, 反之为 0	/
31	沉金/银/锡	<i>Iasa</i>	采用沉金、沉银、沉锡中的一种为 1, 反之为 0	0.001 ^f
32	是否有碳油	<i>Cboil</i>	有碳油为 1, 反之为 0	0.285
33	是否采用字符打印	<i>Chaprt</i>	采用字符打印为 1, 反之为 0	<0.001
34	是否采用阻焊塞孔	<i>Srph</i>	采用阻焊塞孔为 1, 反之为 0	<0.001
35	是否非绿油墨颜色		采用非绿油墨颜色为 1, 反之为 0	<0.001 ^f
36	Panel 包含交货单元数量	<i>Duap</i>	多个交货单元布局于一个 Panel 进行生产, 该厂只生产单拼板(1~1 008)	<0.001 ^t
37	补投次数	<i>Suprt</i>	同一订单号投料次数-1 (1~9)	
38	要求生产数量	<i>Reqq</i>	要求交货数量-引用库存数量(1~42 000)	<0.001 ^t
39	综合投入数量	<i>Relq</i>	生产中各订单所有投入交货单元数量(6~45 360)	
40	报废数量	<i>Scraq</i>	各订单报废数量(0~1 040)	
41	入库数量	<i>Qualq</i>	各订单合格入库数量(5~44 688)	
42	余数入库数量	<i>Surfq</i>	入库数量-要求生产数量(0~4 480)	
43	交货单元面积	<i>Punita</i>	各订单交货单元面积(0.001~0.267)	<0.001 ^t
44	要求生产面积	<i>Reqa</i>	$Reqq \times Punita$ (0.054~41.553)	<0.001 ^t
45	综合投入面积	<i>Rela</i>	$Relq \times Punita$ (0.175~76.532)	

续表

序号	参数名称	符号	描述(取值范围)	<i>p</i> -值
46	报废面积	<i>Scraa</i>	$Scraq \times Punita$ (0.000~20.976)	
47	入库面积	<i>Quala</i>	$Qualq \times Punita$ (0.14~75.842)	
48	余数入库面积	<i>Surfa</i>	$Surfq \times Punita$ (0~47.092)	
49	补投率	<i>Suprr</i>	一定周期内所有补投次数/总订单数×100%	
50	报废率	<i>Scrar</i>	$Scraa/Rela \times 100\%$ (0%~39.394%)	
51	综合合格率	<i>Qualr</i>	$Quala/Rela \times 100\%$ (60.606%~91.66%)	
52	余数入库率	<i>Surfr</i>	$Surfa/Reqa \times 100\%$ (0%~91.67%)	
53	历史良率	<i>Hquar</i>	2年内同一生产型号综合合格率的均值(2.903%~100%)	<0.001 ^t

注: 板厚、线宽/间距、铜厚、面积的单位分别为 mm、mil、OZ 和 m^2 。f 和 t 符号分别表示 *F*-检验和 *t*-检验小于 0.05, “/”表示连续型参数变异系数过小或离散型参数中某一取值比例过大; *p*-值为空表示该参数只做统计用。

机器学习主要目标是设计模型使得计算机能基于数据确定相应行为。决策树、ANN 和 SVM 是常见的机器学习算法^[7]。决策树采用递归划分训练样本以评估指定变量对因变量的影响,从而将样本分隔为具有相似特征的不同分组并预测结果^[3]。本文所采用的 CHAID 决策树通过反复地将样本划分为两个或两个以上的子节点,用于分类或连续型因变量的预测,任何预测变量(连续预测转化为序数预测)将被合并,直到对目标变量没有显著差异,划分过程中的合并与否将基于 Bonferroni-检测计算调整 *p*-值确定^[8]。

ANN 基于大量的神经元模拟大脑解决问题的方式进行训练并挖掘相应规律^[9]。本研究将采用具有一层隐含层的 BP(Back Propagation)人工神经网络进行预测。BP 人工神经网络隐含层和输出层中每一个节点都具有一个输入的加权求和函数;同时定义了一个激活函数用以确定一个节点在一个或多个输入节点条件下的输出。所有节点集可视为一个超平面,而隐含层神经元实质上是将非线性的样本转换为线性样本^[9],算法参数设置为默认值。

SVM 是一种用于分类和回归的监督式机器学习模型。类似于 MLR, SVM 回归预测的目的是通过确定一个因变量与多个自变量之间的超平面;目标是减少实际值和预测值之间的偏差平方和;但 SVM 的超平面参数估计遵循的是 ϵ -不敏感损失函数参数估计最小化原则^[10]。实践证明 SVM 具有较好的预测精度和预测效果^[11]。本研究通过 SPSS

Modeler 14.1 实现。

1.4 性能度量

常用的均方误差在此无法完全反应预测效果。为了更好地评价预测性能,结合 PCB 生产特点,引入余数入库率和补投率相应预测值以及两者加权和作为预测性能评价指标。相应评价指标计算流程如图 2 所示,具体参数如表 2 所示。

投料数量预测定义为:

$$Relq_Pd = \frac{Reqq_Pd \times (1 + adj_cof)}{(100 - Scrar_Pd) / 100} \quad (1)$$

式中: *adj_cof* 为平衡余数入库率和补投率并优化综合成本引入的调节系数,将设置不同水平,并通过企业成本模型确定优选值。

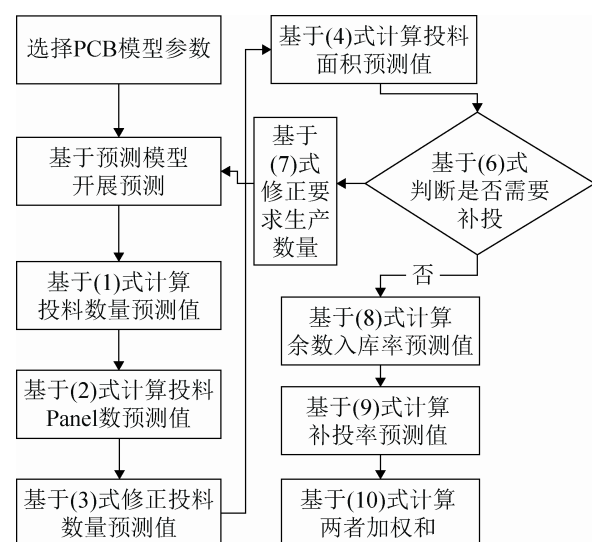


图2 性能指标计算流程

Fig.2 Flowchart for the calculation of indicators

表 2 评价指标相关参数
Tab. 2 Evaluation indicators related parameters

参数	符号
报废率预测值	<i>Scrar_Pd</i>
投料数量预测值	<i>Relq_Pd</i>
投料 Panel 数预测值	<i>Relp_Pd</i>
投料面积预测值	<i>Rela_Pd</i>
要求生产数量预测值	<i>Reqq_Pd</i>
要求生产面积预测值	<i>Reqa_Pd</i>
报废面积预测值	<i>Scraa_Pd</i>
余数入库率预测值	<i>Surfr_Pd</i>
订单补投次数预测值	<i>Suprt_Pd</i>
补投率预测值	<i>Suprr_Pd</i>

相应投料 Panel 数预测值通过下式计算:

$$Relp_Pd = \left\lceil \frac{Relq_Pd}{Duap} \right\rceil \quad (2)$$

式中: *Duap* 为表 1 中所定义的一个 Panel 中包含交货单元数量。

在此基础上, 投料数量将更新为:

$$Relq_Pd = Relp_Pd \times Duap \quad (3)$$

投料面积预测值将直接基于修正后的投料数量计算:

$$Rela_Pd = Relq_Pd \times Punita \quad (4)$$

相应报废面积预测值通过下式获得:

$$Scraa_Pd = Rela_Pd \times Scrar \quad (5)$$

定义某一次预测投料中订单补投次数:

$$Suprt_Pd = \begin{cases} 1, & Rela_Pd - Reqa_Pd - Scraa_Pd < 0 \\ 0, & \text{反之} \end{cases} \quad (6)$$

更新要求生产数量预测值:

$$Reqq_Pd = \lceil Relq_Pd - Scraa_Pd / Duap \rceil \quad (7)$$

相应的余数入库率预测值定义为:

$$Surfr_Pd = \frac{\sum Relq_Pd - Reqa - \sum Scraa_Pd}{Reqa} \times 100\% \quad (8)$$

$\sum Relq_Pd$ 为所有投料面积预测值之和,
 $\sum Scraa_Pd$ 为所有报废面积预测值之和。

n 个订单的补投率预测值为:

$$Suprr_Pd = \left(\sum_{i=1}^n \sum_j Surfr_Pd_i \right) / n \times 100\% \quad (9)$$

$\sum_j Surfr_Pd_i$ 为订单 *i* 的补投次数。

在此基础上, 余数入库率和补投率的预测值的加权和定义为:

$$WSum = w_1 \times Surfr_Pd + w_2 \times Suprr_Pd, \\ w_1 + w_2 = 1, 0 \leq w_1 \leq 1, 0 \leq w_2 \leq 1 \quad (10)$$

加权和是对余数入库率和补投率这对 Pareto 目标的综合; 当对比不同预测模型时, 如果无法保证某一模型在这两个目标上均优于其它模型, 难以评定算法的优劣。而在确定的权值下, 加权和越大相应预测机制所导致的超投和补投成本越大; 反之则越小, 对应预测模型越好。

2 结果与讨论

2.1 *t*-检验和参数优选

基于 *F*-检验和 *t*-检验优选的 35 个参数, 进一步利用 MLR *t*-检验筛选对报废率影响不显著参数, 作为各预测模型构建参数输入。表 3 中给出 *t*-检验结果, 基于显著性判断筛选: 层数、板镀次数、工序数、半固化片数、是否光电板、是否高频板、是否 IPCIII 标准、是否有减薄铜、残铜率均值、是否内层干膜、是否有负片电镀、孔数、沉铜通孔厚径比、是否有树脂塞孔、有铅喷锡、无铅喷锡、护铜剂、图镀铜镍金、镀金手指、电镀硬金、沉金/银/锡、是否采用阻焊塞孔、是否非绿油墨颜色、交货单元面积、要求生产面积、历史良率 26 个报废率预测建模参数。

2.2 预测性能对比

基于上述预测机制和优选参数, 开展虚拟仿真投料, 测试报废率预测模型的训练和预测效果。针对训练样本, 表 4 给出了不同预测机制均方误差; 可以看出 ANN 的均方误差最小, 随后依次为 MLR、CHAID 和 SVM。

进一步对比 3 组样本的余数入库率(*Surfr*)和补投率(*Suprr*)预测值, 具体结果如表 5 所示。其中 S2、S3 的 *Suprr* 较 S1 分别增加了 13.17%、30.84%,

而相应的 *Surfr* 分别降低了 28.38%、52.07%。其主要原因来自于车间管理改革, 2016 年 1 月和 4 月开始(分别对应于样本组 S2、S3)车间管理层强制车

间投料人员一次投料不能超过要求生产数量的 9% 和 6%。调整后的结果显示目前车间的投料策略和管理技术难以同时优化余数入库率和补投率。

表 3 MLR 参数检验
Tab. 3 MLR parameter testing

序号	变量	<i>B</i>	<i>SE</i>	<i>Beta</i>	<i>t</i>	<i>Sig.</i>
1	板厚	0.075	0.079	0.017	0.949	0.343
2	层数	0.632	0.024	0.421	26.582	<0.001
3	板镀次数	1.012	0.12	0.108	8.405	<0.001
4	工序数	0.156	0.01	0.564	15.184	<0.001
5	半固化片数	-0.042	0.017	-0.028	-2.455	0.014
6	是否光电板	0.367	0.142	0.013	2.583	<0.001
7	是否高频板	0.550	0.13	0.013	4.226	<0.001
8	是否 IPCIII 标准	0.976	0.087	0.046	11.220	<0.001
9	内层最小线宽	-0.005	0.004	-0.010	-1.188	0.235
10	内层最小线间距	-0.012	0.01	-0.016	-1.238	0.216
11	外层最小线宽	-0.005	0.005	-0.010	-1.155	0.248
12	外层最小线间距	0.008	0.01	0.011	0.815	0.415
13	是否有减薄铜	0.587	0.089	0.021	6.564	<0.001
14	残铜率均值	0.021	0.001	0.211	14.891	<0.001
15	是否内层干膜、	0.378	0.083	0.013	4.565	<0.001
16	是否有负片电镀	1.127	0.141	0.115	8.020	<0.001
17	总孔数	7.49E-005	0.0	0.153	26.154	<0.001
18	沉铜通孔厚径比	0.180	0.02	0.124	9.155	<0.001
19	是否需要二钻	-0.058	0.062	-0.003	-0.936	0.35
20	是否有树脂塞孔	-1.104	0.209	-0.025	-5.275	<0.001
21	有铅喷锡	2.631	0.268	0.172	9.809	<0.001
22	无铅喷锡	2.697	0.271	0.136	9.943	<0.001
23	护铜剂	2.652	0.282	0.076	9.397	<0.001
24	图镀铜镍金	3.915	0.274	0.159	14.268	<0.001
25	镀金手指	0.417	0.122	0.010	3.411	<0.001
26	电镀硬金	0.748	0.155	0.027	4.813	<0.001
27	沉金/银/锡	2.664	0.266	0.246	10.005	<0.05
28	是否采用字符打印	0.125	0.135	0.003	0.927	0.354
29	是否采用阻焊塞孔	-0.389	0.062	-0.045	-6.265	<0.001
30	是否非绿油墨颜色	0.139	0.065	0.007	2.134	0.033
31	Panel 中包含交货单元数量	-0.004	0.003	-0.011	-1.729	0.084
32	要求生产数量	4.08E-005	0.0	0.003	0.542	0.588
33	交货单元面积	37.202	0.903	0.233	41.213	<0.001
34	要求生产面积	-0.121	0.006	-0.119	-19.540	<0.001
35	历史良率	-0.099	0.003	-1.227	-33.801	<0.001

注: *B* 为各参数系数估计值; *SE* 为标准偏差; *Beta* 为标准系数; *t* 为 *t*-检验值; *Sig.* 为显著性值的临界水平。

表4 均方误差对比

Tab. 4 Comparison of mean squared error (MSE)

	MLR	CHAID	ANN	SVM
SE	3.08	3.18	3.05	3.22

表5 不同预测算法余数入库率和补投率对比

Tab. 5 Comparison of *Surfr* and *Suprr* for the four prediction approaches

样本	实际值	MLR	CHAID	ANN	SVM
S1	17.54%	20.64%	21.33%	20.83%	27.97%
	11.31%	3.77%	3.98%	3.66%	3.05%
S2	19.85%	21.84%	21.38%	21.65%	26.36%
	8.1%	3.89%	4.01%	3.81%	3.24%
S3	22.95%	21.38%	22.68%	21.63%	26.78%
	5.42%	3.87%	3.91%	3.85%	3.14%

注：各组样本上下行分别为补投和余数入库率。

对比结果同时显示，S2、S3 预测余数入库率明显低于实际余数入库率，SVM 获得最低的余数入库率；对于 S1，其中 MLR、CHAID、ANN 以及 SVM 所得余数入库率预测值较实际值分别下降了 66.67%，64.81%，67.63%，73.03%；对于 S2，分别下降了 51.98%，50.49%，52.96%，60%；对于 S3，分别下降了 28.59%，27.85%，28.97%，42.07%。4 种预测机制均可减少余数入库率，从而降低物料、生产、库存和销毁处理成本。

补投率对比结果显示，对于 S1，MLR、CHAID、ANN、SVM 所得预测补投率较实际值分别上升了 17.67%，21.61%，18.76% 和 59.46%；对于 S2，4 种预测模型所得预测补投率较实际值分别上升了 10.02%，7.71% 和 9.07%，32.76%；对于 S3，MLR、CHAID 以及 ANN 所得预测补投率较实际值分别下降了 6.84%，1.17% 和 5.75%，而 SVM 所得值增加了 16.69%。

上述对比仍无法直接确定预测机制优劣。进一步基于(10)式加权和进行对比。因为 w_1 和 w_2 无法明确，所以设置 *Suprr_Pd* 权值 0.1~0.9 九个水平(间隔为 0.1，且满足 $w_1+w_2=1$)，对于样本 S1、S2 和 S3，分别计算 *WSum*，并拟合 *WSum* 随 *Suprr_Pd*

权值变化趋势，具体结果如图 3 所示。

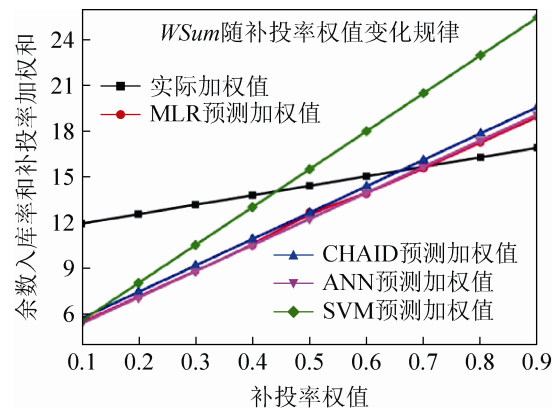
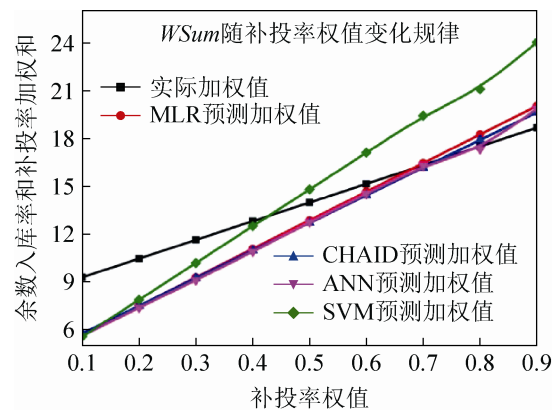
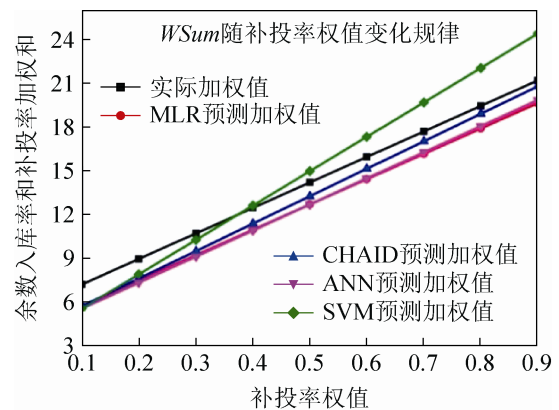
(a) S1 的 *WSum* 随 *Suprr_Pd* 变化趋势(b) S2 的 *WSum* 随 *Suprr_Pd* 变化趋势(c) S3 的 *WSum* 随 *Suprr_Pd* 变化趋势

图3 加权和随补投率预测值权值变化趋势

Fig.3 *Wsum* changes along with the weight of *Suprr_Pd*

从图 3(a)~(c)拟合曲线可以看出，不同权值条件下，SVM 所得 *WSum* 最大，CHAID 次之，MLR 和 ANN 最小且拟合曲线几乎重合；对于 3 组样本，在补投率权值小于 0.4 时，4 种预测机制所得 *WSum*

均小于 W_{Sum} 实际值; 对于 S1、S2, 在 $Suprr_Pd$ 权重小于 0.7 时, 4 种预测机制所得 W_{Sum} 均小于 W_{Sum} 实际值; 对于 S3, 除 SVM 外, 其它预测机制所得 W_{Sum} 均小于 W_{Sum} 实际值。按照企业经验 W_{Sum} 中 w_2 一般小于 0.7, 整体上可以选择 MLR 和 ANN 作为预测模型。因 ANN 对外是一个黑箱且参数配置较为困难, 而 MLR 能通过权值系数直观反映各参数对报废率的影响, 在此优选 MLR。

2.3 调节系数优选

基于报废率多元回归预测模型, 设置调节系数 0%, 0.5%, 1%, 1.5%, 2.0%, 2.5%, 3.0% 七个水平, 开展投料仿真, 计算仿真投料后相应余数入库率和补投率。同时结合车间实际成本模型计算对应的余数入库成本、拖期成本和补投成本, 基于上述 3 种成本优选综合损失最小所对应调节系数。表 6 给出了计算相应成本的参数描述; 表 7 是实施车间针对第三组样本(S3)计算成本所涉及的核心参数及其结果, 对比综合损失发现当预设余数入库率为 1% 时, 综合成本最低。

2.4 实施与验证

为了进一步验证上述方法, 结合企业应用需求, 基于 MLR 所优选参数和预测机制, 设置其调节系数 1%, 开发实施投料优化控件, 该控件与车间 ERP 和制造执行系统具有数据集成双向接口, 根据新订单相应参数, 可手工选择订单或后台自动触发预测引擎并向 ERP 和制造执行系统反馈预测报废率、下料 Panel 数、综合投入数量等关键结果。

基于该控件开展实施, 截取实施以来 2 372 个订单作为验证样本(S4), 实施后的余数入库率和补投率分别为 4.89% 和 16.61%, 较 S2、S3 的余数入库率分别下降 39.62%、9.78%, 补投率分别降低 16.32%、38.17%。验证结果显示本研究所确定的预测机制能更精准的自动确定各订单的投料面积和数量, 从而进一步消减车间物料、生产、库存、回收处理、补投和拖期等成本。同时可减少车间投料经验影响和投料人员。

表 6 成本计算相应参数说明
Tab. 6 Specification of parameters related to cost calculation

No.	参数	符号	说明
1	需要补投次数	$TSuprt_Pred$	该样本预测补投累计次数
2	需要补投面积	$RSupra_Pred$	按照预测各订单需要补投面积和
3	余数入库面积	$Surfa_Pred$	同表 1 种定义
4	引用库存面积比例	$Ivuger_Pred$	引用库存中的余数入库面积和预测新产生的余数入库面积之比
5	单位面积平均售价	$Apua_Pred$	平均售价 2000/ m^2 左右
6	余数入库成本	$Surfpc_Pred$	$Surfa_Pred \times (1 - Ivuger_Pred) \times Apua_Pred$
7	拖期 > 3 天的订单数	$Anom3d_Pred$	约有 1/4 的补投订单其拖期大于 3 天(一般 3 天即以上需要赔偿)
8	拖期订单平均面积	$Aato_Pred$	平均值约为 5.5 m^2
9	单个订单拖期订单折款	$Disctpo_Pred$	拖期订单平均折款损失 2 200 元左右
10	拖期成本	Tdc_Pred	$Anom3d_Pred \times Disctpo_Pred$
12	补投面积	$Supra_Pred$	基于需要补投面积, 考虑 Panel 圆整后实际需要重新投料面积
13	补投成本	$Prodinca_Pred$	$Prodinca_Pred \times Apua_Pred$
14	综合损失	$Compl_Pred$	$Surfpc_Pred + Tdc_Pred + Prodinca_Pred$

注: 其中统计结果基于车间 2016 年 1-8 月累计历史数据。

表7 成本统计计算表
Tab. 7 Statistic table for the computation of cost

<i>Adjst_coef</i>	<i>Surfr</i>	<i>Suprr</i>	<i>TSuprt</i>	<i>RSupra</i>	<i>Surfa</i>	<i>Ivuger</i>	<i>Apua</i>	<i>Surfpc</i>
0%	3.89%	21.84%	1 516	565.389 4	1 194.289 9	48.43%	2 000	1 231 790.6
0.5%	4.18%	19.97%	1 386	535.215 7	1 283.324 4	48.43%	2 000	1 323 620.8
1%	4.56%	17.74%	1 231	503.405 6	1 399.990 2	48.43%	2 000	1 443 949.9
1.5%	4.93%	16.24%	1 127	481.217 2	1 513.585 9	48.43%	2 000	1 561 112.5
2%	5.32%	14.89%	1 033	458.340 1	1 633.321 9	48.43%	2 000	1 684 608.2
2.5%	5.63%	14.19%	985	441.202 5	1 728.496 7	48.43%	2 000	1 782 771.5
3%	5.92%	13.39%	929	425.996 3	1 817.531 2	48.43%	2 000	1 874 601.6

<i>Adjst_coef</i>	<i>Anom3d</i>	<i>Aato</i>	<i>Prodinc</i>	<i>Tdc</i>	<i>Supra</i>	<i>Prodinca</i>	<i>Compl</i>
0%	379	5.5	2 200	833 800	565.389 4	1 130 778.8	3 196 369.4
0.5%	346.5	5.5	2 200	762 300	535.215 7	1 070 431.4	3 156 352.2
1%	307.75	5.5	2 200	677 050	503.405 6	1 006 811.3	3 127 811.2
1.5%	281.75	5.5	2 200	619 850	481.217 3	962 434.56	3 143 397.1
2%	258.25	5.5	2 200	568 150	458.340 1	916 680.25	3 169 438.5
2.5%	246.25	5.5	2 200	541 750	441.202 6	882 405.11	3 206 926.6
3%	232.25	5.5	2 200	510 950	425.996 3	851 992.65	3 237 544.3

注：表头_Pred 均被省略；要求生产面积(Reqa)为 30 701.54 m²。

3 结论

本文提出了基于数据挖掘优化 PCB 样板投料的新机制。首先梳理和规范定义了与 PCB 报废样本直接相关的核心参数，并基于假设检验优选了与报废率预测建模相关的 26 个参数：主要包括层数、板镀次数、工序数、半固化片数、是否光电板、是否高频板、是否 IPCIII 标准、是否有减薄铜、残铜率均值、是否内层干膜、是否有负片电镀、孔数、沉铜通孔厚径比、是否有树脂塞孔、有铅喷锡、无铅喷锡、护铜剂、图镀铜镍金、镀金手指、电镀硬金、沉金/银/锡、是否采用阻焊塞孔、是否非绿油墨颜色、交货单元面积、要求生产面积、历史良率等，为类似挖掘分析和参数优选提供了重要参考。

构建了 MLR、CHAID、ANN 和 SVM 四种预测模型；引入了余数入库率、补投率及两者加权和评价指标，开展虚拟投料仿真，对比优选 MLR 预测机制。引入调节系数，结合企业实际成本模型进行优选，设置 0%~3.5% 7 个不同调节水平，基于 MLR 预测模型和优选调节系数，通过仿真投料发现 1%时能最小化车间投料综合损失；在此基础上，基于 MLR 和 1%调节系数开发投料优化控件，以

实施车间实际数据进行验证，结果显示余数入库率较两组测试样本分别下降 39.62%，9.78%，补投率降低 16.32%，38.17%，具有明显的经济效益。

本研究主要基于一个车间数据，鉴于该车间生产特点，其数据具有一定的局限性，比如极少的埋盲孔板(包括高密度板)、10 层以上板和多拼板；同时 PCB 产品和工艺参数非常之多，其它未列举参数缺乏一定量的高质量的数据积累，所筛选的参数需要结合不同车间甚至同行业其它企业数据进行适当增删，以提高预测模型参数的通用性。

参考文献：

- [1] Marques A C, Cabrera C J, Malfatti C F. Printed circuit boards: A review on the perspective of sustainability [J]. Journal of Environmental Management (S0301-4797), 2013, 131: 298-306.
- [2] 姚锡凡, 周佳军, 张存吉, 等. 主动制造—大数据驱动的新兴制造范式[J]. 计算机集成制造系统, 2017, 23(1): 172-185.
Yao X F, Zhou J J, Zhang C J, et al. Proactive manufacturing—a big-data driving emerging manufacturing paradigm [J]. Computer Integrated Manufacturing Systems, 2017, 23(1): 172-185.
- [3] Han J W, Kamber M, Pei J. Data mining: concepts and techniques [M]. 3rd Edition. San Francisco: Morgan

- Kaufmann, 2011.
- [4] 张影, 翟丽丽, 王京. 大数据背景下的云联盟数据资源服务组合模型[J]. 计算机集成制造系统, 2016, 22(12): 2920-2929.
Zhang Y, Zhai L L, Wang J. Cloud computing federation data resource sever composition in big-data background [J]. Computer Integrated Manufacturing Systems, 2016, 22(12): 2920-2929.
- [5] 吕佑龙, 张洁. 基于大数据的智慧工厂技术框架[J]. 计算机集成制造系统, 2016, 22(11): 2691-2697.
Lyu Y L, Zhang J. Big-data-based technical framework of smart factor [J]. Computer Integrated Manufacturing Systems, 2016, 22(11): 2691-2697.
- [6] Bolón-Canedo V, Sánchez-Marroño N, Alonso-Betanzos A. A review of feature selection methods on synthetic data [J]. Knowledge and Information Systems (S0219-1377), 2013, 34(3): 483-519.
- [7] Philip C C L, Zhang C. Data-intensive applications, challenges, techniques and technologies: a survey on big data [J]. Information Science (S0020-0255), 2014, 275: 314-347.
- [8] Ture M, Tokatli F, Kurt I. Using Kaplan–Meier analysis together with decision tree methods (C&RT, CHAID, QUEST, C4.5 and ID3) in determining recurrence-free survival of breast cancer patients [J]. Expert Systems with Applications (S0957-4174), 2009, 36 (2, Part 1): 2017-2026.
- [9] Wikipedia. Artificial neural network [EB/OL]. (2016-10) [2017-09-05]. https://en.wikipedia.org/wiki/Artificial_neural_network.
- [10] Smola A J, Schölkopf B. A tutorial on support vector regression [J]. Statistics and Computing (S0960-3174), 2004, 14(3): 199-222.
- [11] 任艳, 周小敏, 关威, 等. 支持向量回归机在颜色测温中的仿真应用[J]. 系统仿真学报, 2016, 28(11): 2736-2741.
Ren Y, Zhou X M, Guan W, et al. Simulation approach to temperature measuring using image color based on support vector regression [J]. Journal of System Simulation, 2016, 28(11): 2736-2741.

(上接第 2655 页)

- [9] 张继华, 邓研, 郭凤仪, 等. 永磁操作机构真空断路器的智能控制器的设计[J]. 辽宁工程技术大学学报(自然科学版), 2011, 30(12): 751-756.
Zhang Jihua, Deng Yan, Guo Fengyi, et al. Design of controller for permanent magnetic actuator vacuum circuit breaker[J]. Journal of Liaoning Technical University (Natural Science), 2011, 30(12): 751-756.
- [10] 吕锦柏, 王毅, 谢将剑, 等. 基于线圈电流的永磁真空断路器控制方法[J]. 高电压技术, 2013, 39(4): 860-868.
Lv Jinbo, Wang Yi, Xie Jiangjian, et al. Control Method for Permanent Magnetic Vacuum Circuit Breaker Based on Coil Current[J]. High Voltage Engineering, 2013, 39(4): 860-868.
- [11] 刘博, 张建峡. 电容器恒流充电方法的分析与研究[J]. 工业控制计算机, 2013, 26(1): 116-118.
Liu Bo, Zhang Jianxia. Analysis and Research of Constant Current Charging Method for Capacitor[J]. Industrial Control Computer, 2013, 26(1): 116-118.
- [12] 阳峰. 基于双管 Buck-Boost 变换器的电容器充电电源研究[D]. 武汉: 华中科技大学, 2012.
Yang Feng. A Thesis Submitted in Partial Fulfillment of the Requirements for the Degree of Master of Engineering[D]. Wuhan: Huazhong University of Science and Technology, 2012.
- [13] 赵红强, 徐建源, 秦祖荫. 单线圈永磁驱动机构中驱动器的研制[J]. 华通技术, 2005 (1): 11-15.
Zhao Hongqiang, Xu Jianyuan, Qin Zuyin. The Permanent Magnetic Machine Driver of Vacuum Circuit Breaker[J]. Huatong Technology, 2005 (1): 11-15.
- [14] 危立辉. 储能电容开环 PWM 控制恒流充电装置[P]. 中国, 201010555730.X.
- [15] 汪先兵, 费树岷, 徐清杨, 等. BP 神经网络 PID 控制的永磁真空开关储能电容恒流充电特性分析[J]. 电工技术学报, 2015, 30(10): 212-218.
Wang Xianbing, Fei Shumin, Xu Qingyang, et al. Constant Current Charging Characteristic Analysis of Storage Capacitor Based on BP Neural Network PID Control for Permanent Magnet Vacuum Switch[J]. Transactions of China Electrotechnical Society, 2015, 30(10): 212-218.