

基于 BI 的工序质量分析

童瑞明 米智伟

(上海大学机电工程与自动化学院 上海 200072)

摘要 质量管理中工序质量是很重要的环节。现在工序质量管理中存在的问题主要有缺乏质量诊断和改进的方法,历史数据没得到有效的利用。将最新的计算机技术 BI(商务智能)应用到工序质量管理中,对工序质量的相关数据进行集成,使用数据挖掘方法找出质量隐患,辅助生产人员进行质量管理,做出正确的决策,提升质量合格率。

关键词 商务智能,数据挖掘,决策树,工序质量

中图分类号 TP315 **文献标识码** A

Process Quality Analyse Based on BI

TONG Rui-ming MI Zhi-wei

(School of Electro-Mechanical Engineering and Automation, Shanghai University, Shanghai 200072, China)

Abstract Process quality is the important section of the quality control. Now the major problem that exist in process quality control is lacking method of quality diagnose and improvement, history data is not effective utilized. BI(Business Intelligence) is the latest computer technology, this paper applied it in the process quality control, to integrate all the data of process quality, find quality hidden trouble by data mining, assist operating staff to control quality and make correct decision to improve rate of satisfactory quality.

Keywords Business intelligence, Data mining, Decision tree, Process quality

1 引言

1.1 工序质量的重要性及概况

产品质量是企业生存之本,发展之本,品牌之本,是企业其他一切的基础。随着科学技术的不断进步和经济的快速发展,市场竞争日益激烈,产品的质量水平已成为影响企业参与市场竞争的关键因素。

在产品质量形成的全过程中,生产制造过程是最终实现产品质量的重要过程,而其中构成生产制造过程的基础是工序。每一道工序质量的好坏,最终都直接或间接地影响产品的质量,所以工序的质量是形成产品质量最基本的环节。

作为质量竞争的关键,工序质量控制愈加受到企业的重视。自 1924 年美国实验室工程师休哈特创立控制图至今,工序质量控制的理论和技术都有了很大的发展。

初期的工序质量控制理论仅以 SPC 技术为基础。SPC 理论能科学地区分出制造过程中产品质量的偶然波动与异常波动,从而对过程的异常及时预测,以便人们采取措施,消除异常,恢复过程的稳定。在此期间,SPC 理论在休哈特控制图理论的基础上有了较大的扩展,如累积和控制图、区域控制图、预控图、通用控制图、多元均值控制图、多元离散控制图。

1.2 企业概况

上海 H 公司专门从事汽车零部件产品的研发、制造与销售,生产的产品系列涉足汽车电器与发动机部件的领域,现有产品基本为国内外主机厂定点配套。自主研发的产品批量配

套于一汽集团、东风集团、上海大众、北京奔驰、一汽奥迪等 10 余家国内汽车主机厂,公司产品在国内的市场覆盖率已达到了 80%。同时批量返销欧洲和北美的汽车主机配套市场的是德国奥迪、美国克莱斯勒、法国雪铁龙等 10 余家国际著名高端客户的全球供应商。

H 公司通过 10 多年的积极探索,建立有效的产品质量保证体系,其从产品设计、生产到交付,全过程导入 APQP, FMEA, PPAP, SPC 等质量管理工具,采用 SQA 供应商管理和 6S 生产现场管理,同时还全面运用 ERP, PDM, OA 等企业信息化管理平台。

1.3 信息技术与质量管理

信息化管理是企业适应未来发展的必经之路,在企业数目不断增加,产品数据飞速增长的现实面前,传统的数据管理手段显得力不从心。所以,将信息技术和质量管理的有机结合起来,具有重要意义。数据挖掘技术作为一种先进的数据分析工具,为质量管理提供了全新的手段。

由于复杂生产过程具有工艺参数众多、机理复杂、非线性显著和动态变化等特点,因此难以用常规方法建立其精确的数学模型。同时,随着工业自动化技术和计算机信息技术的快速发展及应用,在生产现场又积累了海量的数据,激发了人们对分析这些数据的浓厚兴趣,以识别和发掘其中隐藏的规律,更好地指导生产实践。

传统质量控制方法关注于生产工序的质量波动性,通过统计推断预测工序的变异性,以便及时展开工艺调整。但这

童瑞明(1983—),男,硕士生,主要研究方向为 ERP、质量管理、BI;米智伟(1967—),男,博士,副教授,主要研究方向为计算机集成制造、供应链管理。

些传统方法一般针对生产的某个环节,很难从全局出发进行分析管理。在生产数据爆炸性增长的情况下,可以从大量质量数据中确定影响质量的重要因素。本文将 BI(Business Intelligence)技术应用到工序质量分析中,对生产过程中的所有数据进行集成,提高质量合格率,辅助生产人员进行质量管理及决策。

模糊分类法、人工神经网络、贝叶斯分类法和聚类分析等数据挖掘方法可以对工序质量进行分析和预测,取得了较好的应用效果。关联规则挖掘技术也较多地用于质量管理,但算法效率偏低,应用领域较小。基于聚类算法的质量管理中,很难兼顾各质量权重系数。相比之下,决策树(Decision Tree)方法具有易于处理变化的数据、对噪声数据具有健壮性、生产规则容易理解、识别效率高等优点。

2 公司使用传统的质量管理方法存在的问题

2.1 缺乏质量诊断的方法

目前,制造车间常采用统计过程控制来进行过程质量管理。比如运用 SPC 中质量控制图来实时监测产品在生产过程中的质量特性,确保生产过程始终处于统计控制状态,以保证产品质量,提高经济效益。但 SPC 只能对生产过程中当时的质量异常波动发生警报,并不能对异常进行诊断,且 SPC 主要为事中控制,很难对质量损失进行事前预防,必须寻找产生工序质量损失的深层次原因。

2.2 缺乏质量改进的方法

传统的质量改进是在业务系统上进行的,而且难以发现深层次的、不明显的质量问题。它只能判断当前生产过程中明显的质量问题,而对于如何应用历史数据,全面、客观、准确地分析生产中各因素,以及如何调整这些因素,则因为数据处理的成本、效率诸方面的原因而无法有效地进行。其在分析方法和分析系统上都存在着缺陷,因而不能根据历史数据,对产品质量合格率做进一步的提升。

3 工序质量分析系统总体框架

这里提出工序质量分析系统的总体框架,如图 1 所示。对工序质量分析用到的数据进行集成,将其存放在数据仓库中,然后利用数据挖掘方法对工序质量的相关数据进行处理,包括对历史数据进行挖掘,对班组人员、物料批次、工艺版次、设备、质量损失特性值做出评估。ERP 系统的数据包括 QM(质量管理),PP(生产计划),MM(物料管理)和 SD(销售与分销),通过 BAPI(业务应用程序接口)送到 BW(业务信息仓库)中;对于非 ERP 系统的数据,通过 DB Connect 传输到 BW 中。在 BW 中,数据先在 PSA(持续数据加载区)中进行整理,然后在 BW 中建立 Infocube(信息立方体),再通过 DM(数据挖掘)方法:决策树(Decision Tree)进行挖掘。

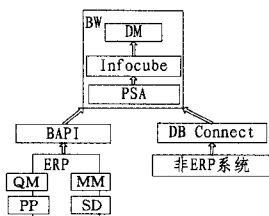


图 1 工序质量分析系统总体框架图

在 ERP 系统里存有质量管理的基本信息,包括物料供应商、使用的工艺方案、人员班组、生产线情况、废品率等。由于企业信息化的局限,还有些相关数据存在于非 ERP 系统中。对于质检时和生产中出现的质量问题,大部分数据可以从 QM(质量管理系统)中得到。

4 数据仓库的逻辑模型设计

4.1 传统的星型模型

星形结构是进行 OLAP 系统设计时常见的一种数据结构,它为多维数据分析提供了坚实的模型基础,可以为决策支持系统的处理优化数据;通过数据预连接和有选择的数据冗余,为访问和分析过程大大简化了数据。星形结构的中间是事实表,数据被大量地载入,周围属于维度表,被载入的数据较少。传统星型模型的缺点:在维度表中存在冗余记录;与事实表的历史化相比,维度表的历史化(即变化缓慢的维度)在模型中不容易实现。

4.2 改进的星型模型

本文设计的星型模型对传统的星型模型进行了改进,其主要变化在于维度表不包含主数据的信息,而只存储了主数据的 SID 键。SID 键替代了作为维度表组成部分的特性。SID 表示代理标识。在图 2 中,这些键带前缀 SID。SID 表把这些维度表链接到它们各自的主数据表。主数据包含属性、文本和层级 3 个可选的组成部分。它们是分开存储的,属性存储在属性表中,文本存储在文本表中,层级存储在层级表中。维度表中 SID 是这些主数据表的主键。改进后的星形结构拥有这些优点:能更快地访问数据;拥有历史化的维度;主数据跨信息立方体的使用(共享维度);查询性能得到提高。

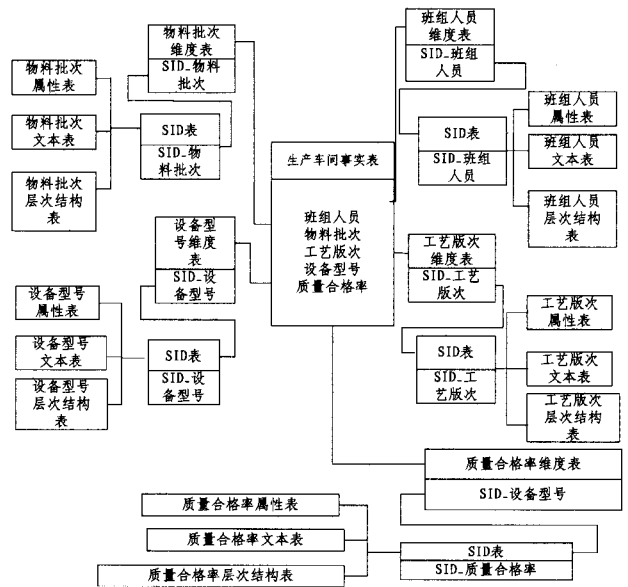


图 2 改进的工序质量数据集星型模型图

5 基于工序质量分析实例的决策树挖掘分析

5.1 决策树简介

决策树是以实例为基础的归纳学习算法。它从一组无次序、无规则的元组中推理出决策树表示形式的分类规则。它采用自顶向下的递归方式,对决策树的内部结点进行属性值的比较,并根据不同的属性值从该结点向下分支,叶结点是要

学习划分的类。从根到叶结点的一条路径就对应着一条合取规则,整个决策树就对应着一组析取表达式规则。1986年Quinlan提出了著名的ID3算法。ID3算法的优点是:算法的理论清晰,方法简单,学习能力较强。虽然ID3只对比较小的数据集有效,但在此处不会产生负面影响。

5.2 实例计算

此处采用以上建立的改进决策树模型对H公司质量损失数据集进行数据挖掘。车间生产的质量损失主要制约因素为生产的各大要素:班组人员,物料批次,工艺版次,设备型号,也即决策树数据挖掘中的条件属性。表1为经过数据聚集、数据转换、数据导入处理后的质量损失数据集的样本数据集。

表1 质量损失原因挖掘训练数据集

编号	属性				工序质量合格率
	物料批次	班组人员	设备型号	工艺版次	
1	1	甲	1	2	低
2	1	甲	1	1	低
3	2	甲	1	2	高
4	3	乙	1	2	高
5	3	丙	2	2	高
6	3	丙	2	1	低
7	2	丙	2	1	高
8	1	乙	1	2	低
9	1	丙	2	2	高
10	3	乙	2	2	高
11	1	乙	2	1	高
12	2	乙	1	1	高
13	2	甲	2	2	高
14	3	乙	1	1	低

信息熵的计算: $H(U) = -\sum_i P(u_i) \log_2 P(u_i)$ 。

类别 u_i 出现的概率: $P(u_i) = \frac{u_i}{S}$ 。

S表示例子集的总数, u_i 表示此类的例子数。对9个合格率高和5个合格率低有: $H(U) = (\frac{9}{14}) \log_2 \frac{14}{9} + (\frac{5}{14}) \log_2 \frac{14}{5} = 0.94\text{bit}$ 。

条件熵的计算: $H(U|V) = -\sum_j P(v_j) \sum_i P(u_i | v_j) \log_2 P(u_i | v_j)$ 。

属性 A_1 取值 v_j 时,类别 u_i 的条件概率: $P(u_i | v_j) = \frac{u_i}{v_j}$ 。

属性 A_1 为物料批次,取值 v_1 为1号批次, v_2 为2号批次, v_3 为3号批次。在物料批次处,1号批次有5个,2号批次有4个,3号批次有5个。所以有: $P(u_1 | v_1) = \frac{2}{5}, P(u_2 | v_1) = \frac{3}{5}, P(u_1 | v_2) = \frac{4}{4}, P(u_2 | v_2) = 0, P(u_1 | v_3) = \frac{2}{5}, P(u_2 | v_3) = \frac{3}{5}$ 。

所以可得,物料批次属性的条件熵为: $H(U|V) = \frac{5}{14} (\frac{2}{5} \log_2 \frac{5}{2} + \frac{3}{5} \log_2 \frac{5}{3}) + \frac{4}{14} (\frac{4}{4} \log_2 \frac{4}{4} + 0) + \frac{5}{14} (\frac{2}{5} \log_2 \frac{5}{2} + \frac{3}{5} \log_2 \frac{5}{3}) = 0.694\text{bit}$; 互信息计算: 对物料批次属性,有

$I(\text{物料批次}) = H(U) - H(U|V) = 0.94 - 0.694 = 0.246\text{bit}$ 。

类似可得: $I(\text{班组人员}) = 0.029\text{bit}$; $I(\text{设备型号}) = 0.151\text{bit}$; $I(\text{工艺版次}) = 0.048\text{bit}$ 。

由以上计算可以看出,互信息最大的特征物料批次应该作为决策树树根,这样就可以得到图3所示的决策树。

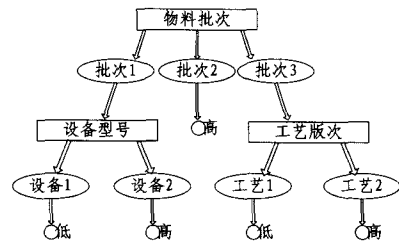


图3 工序质量决策树

5.3 结果分析

从最终决策树和决策规则可以看出,在班组、设备、物料、工艺因素中,物料批次是影响工序质量损失的第一因素,设备和工艺是影响工序质量损失的第二因素。因此加强物料质量检验,选择优秀的物料供应商是减少工序质量损失和控制质量成本的关键因素。此外还要对生产设备进行定期的维护,对工艺不断深入、优化。

用此决策树获得的知识规则对训练集数据实例进行工序质量损失原因挖掘分析验证说明,其在实际应用中准确率高。说明利用基于BI的工序质量挖掘模型对工序质量损失原因进行挖掘分析是可行的。

结束语 传统的质量记录和控制方法已经力不从心,为此笔者运用决策树算法建立了工序质量损失原因挖掘的决策树知识挖掘模型。抽取数据集训练集样本进行挖掘分析,车间生产实际应用证明该模型可行且效率较高。根据决策树分析结果得出了导致工序质量损失的主要原因,为生产管理者进行质量改进提供了依据和参考,为加强生产过程的质量控制奠定了基础。基于数据挖掘的方法有效地利用了工厂的历史数据,使看起来杂乱无章的数据变成企业的宝贵财富。

参考文献

- [1] 余腊生,李强.数据挖掘在质量管理系统中的应用研究[J].计算机工程与设计,2010,31(10):2327-2334
- [2] 王惠君,方明.数据挖掘技术ID3算法在学生评教中的应用研究[J].中国科技信息,2010,19:258-260
- [3] Tomassini M, Luthi L, Giacobini M. The structure of the genetic programming collaboration network[J]. Genetic Programming and Evolvable Machines, 2007, 8(1): 97-103
- [4] 刘建平,孙树栋,王军强.基于改进决策树的工序质量损失原因挖掘研究[J].机械科学与技术,2009,28(10):1353-1358
- [5] 王晓华,苏宏业,渠瑜,等.面向电信欠费挖掘的数据质量评估策略研究[J].计算机工程与应用,2011,47(12):220-233
- [6] 宋彦坡.数据挖掘技术及其在工业生产中的应用[J].计算机测量与控制,2004,12(10):944-947
- [7] 周纪梦.质量管理统计方法[M].北京:中国统计出版社,1999:3-11