

doi: 10.7690/bgzd.2023.11.006

大数据下离散制造业产品质量分析综述

李君妍¹, 胡欣², 刘治红¹, 石义官^{1,3}, 张瀚铭¹

(1. 中国兵器装备集团自动化研究所有限公司长沙分公司, 四川 绵阳 621000;

2. 陆军驻广元地区军代室, 四川 广元 628000; 3. 北京理工大学机械与车辆学院, 北京 100081)

摘要: 为帮助企业识别产品问题和设计缺陷, 提高产品满意度, 需要对离散制造业产品质量进行大数据分析。阐述制造业产品质量大数据的特点, 从数据采集处理、质量预测、质量控制和质量追溯 4 个环节, 综述离散制造业产品质量分析的国内外研究进展和发展动态, 并指出各类理论方法在大数据背景下的挑战, 探讨应对这些挑战的解决途径与发展趋势。结果表明, 该分析可为进一步展开研究提供参考。

关键词: 智能制造; 质量分析; 质量预测; 质量控制; 质量追溯

中图分类号: TP3 **文献标志码:** A

Review on Product Quality Analysis of Discrete Manufacturing Industry Based on Big Data

Li Junyan¹, Hu Xin², Liu Zhihong¹, Shi Yiguan^{1,3}, Zhang Hanming¹

(1. Changsha Branch, Automation Research Institute Co., Ltd. of China South Industries Group Co., Ltd., Mianyang 621000, China; 2. Military Representative Office in Guangyuan District, Army Equipment Department, Guangyuan 628000, China; 3. School of Mechanical Engineering, Beijing Institute of Technology, Beijing 100081, China)

Abstract: In order to help enterprises identify product problems and design defects and improve product satisfaction, it is necessary to analyze the product quality of discrete manufacturing industry with big data. This paper describes the characteristics of product quality big data in manufacturing industry, summarizes the research progress and development trends of product quality analysis in discrete manufacturing industry at home and abroad from four aspects of data acquisition and processing, quality prediction, quality control and quality traceability, points out the challenges of various theoretical methods under the background of big data, and discusses the solutions and development trends to deal with these challenges. The results show that the analysis can provide a reference for further research.

Keywords: intelligent manufacturing; quality analysis; quality prediction; quality control; quality traceability

0 引言

随着物联网、云计算、大数据和人工智能等技术的发展, 离散制造业的生产数据量不断增加, 2015 年数据便超过了 100 EB, 预计未来十年将增加 20 倍^[1]。这些数据包括产品生产流程中的传感器数据、机器数据、工艺参数数据、设备状态数据和检测数据等。通过对这些生产数据进行质量分析, 可以发现生产中的瓶颈、优化生产流程、降低生产成本、提高生产效率和水平, 帮助企业提高竞争力, 促进企业的可持续发展。

对于离散制造业的质量大数据, 其特点主要有: 1) 大规模, 需特殊的工具和技术进行处理、分析和可视化; 2) 实时性, 需使用各类数据获取、传输和处理技术来进行分析; 3) 高多样性, 包含文本、音频、视频、图像和传感器数据等多种类型; 4) 高可靠性, 需保证其准确性、完整性和可信性, 以确

保它可以用于之后的生产决策以及其他关键任务; 5) 高价值性, 用于深入地了解客户行为, 优化业务运营, 做出更好的战略决策。

根据以上特点, 对于大数据下的离散制造业产品进行质量分析需要做到: 1) 数据采集以及处理。针对大量多源异构数据进行存储、处理和分析, 确保其准确性、完整性和可信性。2) 质量预测。属于事前控制, 通过对质量数据进行分析 and 建模, 以预测未来的产品质量, 减少不必要的生产环节。3) 质量控制。属于事中控制, 对实时生产制造过程进行实时监控, 针对突发故障进行处理, 以最大程度地减少产品的缺陷率。4) 质量追溯。属于事后控制, 通过建立有效的追溯体系和追溯系统, 挖掘质量数据中不同元素之间的相关性, 帮助决策者寻找质量问题产生的原因, 以改进生产制造过程。

收稿日期: 2023-07-06; 修回日期: 2023-08-17

基金项目: 国防基础科研项目 (JCKY2020209B002)

第一作者: 李君妍 (1997—), 女, 广东人。

1 国内外研究现状

离散制造业产品质量分析过程如图 1 所示。

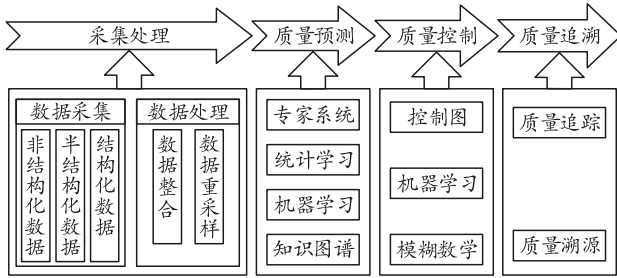


图 1 离散制造业产品质量分析过程

1.1 数据采集及处理

1.1.1 数据采集

产品质量数据主要从企业内部上线的数据采集设备和信息系统中采集，如各类传感设备、MES、ERP、SCM、CAN、CAPP 等。采集数据包括采集人员、机器、物料、方法、环境、测试 6 大类影响质量的数据，采用手工、半自动以及自动采集方式，可以分为非结构化数据、半结构化数据和结构化数据^[2]。结构化数据可分为定性数据和定量数据，定量数据包含几何参数、位置参数、公差参数、加工参数^[3]、设备参数、环境参数、质量缺陷和传感器采集参数；定性数据包含机器故障^[4]、半结构化数据包括表格^[5]、图像^[6]和非结构化数据^[5]。

1.1.2 数据处理

在对质量数据进行分析时，需要根据所使用的研究方法进行恰当的预处理。数据处理方法主要包括数据整合和数据重采样。数据整合，是通过多源异构数据合并到一个满足一致性和易访问性的数据集的过程，使用统一的协议和标准解决信息模型的异构性，常采用 MTConnect 和 OPC UA 协议。MTConnect 可将制造设备中的数据转换为结构化的格式，使其易于获取和分析，而 OPC UA 独立于硬件和操作系统，与前者不同之处在于其开放性和跨平台性。数据重采样，对于类别分布不平衡的数据集进行重新采样以使类别均衡，常用于需要对数据集的少数类重点关注的情况，会改变原始数据的数据分布，比起代价敏感学习具有普遍适用性，如 Feng Y 等^[7]便采用 SMOTE 方法和 Adaboost 方法相结合，以平衡数据用于装配质量预测。

1.2 质量预测

质量预测，即根据历史生产数据，建立预测模型，实时采集当前的产品数据，对未来时刻的产品

质量进行预测。其旨可以在产品的生产制造过程中，针对历史采集数据，评估质量趋势，提前预警，避免不必要且生产成本高昂的生产步骤，为工艺设计和参数优化提供参考，使企业以最低成本获得合格产品。现有质量预测方法主要使用统计学习、机器学习、知识图谱方法，如表 1 所示。

表 1 质量预测方法

方法	优点	缺点	适用范围
专家系统	可解释性强	难以建模复杂关系	机理清晰场景
统计学习	模型构建工作量小；预测速度快	难以适用于高维非线性数据	低维线性数据
机器学习	可拟合非线性关系	需大量数据	高维非线性数据

1.2.1 专家系统

专家系统方法从系统内在机理出发(如动力学、热力学等理论)，通过研究产品生产制造过程中的机理模型和专家知识，构建专家系统，并使用不确定性推理进行质量预测。孔玲爽等^[8]建立加工过程机理模型；林涛^[9]基于粗糙集理论进行知识提取建模，基于属性重要性进行不确定推理；O.A.Nada 等^[10]基于模糊推理进行层次模糊推理。

1.2.2 统计学习方法

统计学习方法针对在产品生产过程中的某个工序进行关键质量指标预测或进行偏差预测。Zhang S 等^[11]使用 PCA 和独立成分分析方法，识别构建回归模型以预测关键质量指标(零件表面平整度)，并使用动态矩阵控制方法提前 k 步预测；由于 PCA 方法无法保证在解释预测变量的方向预测因变量，S. Lynn 等^[12]使用加权 PLS 算法作为预测模型，采用滑动窗口模型更新数据集，并加入前一个时刻的预测值进行改进，结合多元回归分析进行预测。针对偏差预测，刘银华等^[13]通过构建偏最小二乘回归法进行装配过程中偏差源矩阵与关键测点数据的潜结构建模，再采用蒙特卡洛仿真法进行合格率预测。

1.2.3 机器学习

离散制造业产品生产过程中存在产品品种多、批次多、批量小的特点，传统统计学习方法对于此类生产过程的适应性不强；因此，研究者提出使用支持机器学习方法来处理。

支持向量机(supported vector machine, SVM)基于统计学习理论，通过寻找一个满足分类要求的超平面以进行分类，该超平面两侧的空白区域越大则表明其分类性能越好。孙林等^[14]采用 SVM 方法，首次提出使用 SVM 模型的优势；考虑到装配工序

操作的影响,陈勤和^[15]从产品质量与装配工艺相关性的角度出发进行预测,解决了噪声、振动与声振粗糙度波动不稳定,难分析的问题。为得到更快的学习能力和更好的泛化性,避免陷入局部最优解,研究者们采用智能算法(如粒子群优化算法、遗传算法)对 SVM 进行改进。裘稼荣等^[16]使用灰熵关联分析筛选关键装配工艺参数,建立基于粒子群优化的 LSSVM 方法,对关键装配工序进行预测;陈资等^[17]使用遗传算法优化;刘明周等^[18]使用 PSO 优化方法,针对多工序装配过程进行预测,缺点在于虽然从多工序角度进行模型预测,但是没有考虑输入质量特性之间存在的耦合关系。

神经网络(artificial neural network, ANN)在分类、聚类、模式识别和预测中均有应用,主要用于数值范式中通用函数逼近,其具有自学习、适应性、容错性、非线性以及输入到输出映射的先进性等优异的特性。研究者们多使用自组织神经网络^[19]、反向神经网络^[20],使用 PSO 和 AGNES 算法^[21]进行优化。

深度学习提供端对端的质量预测模型,从收集的数据中自动学习质量特征,减少特征提取时间。运用于质量预测的深度学习模型主要有用于图像数据的卷积神经网络(convolutional neural networks, CNN)和针对时间序列数据的长短期记忆神经网络(long short-term memory, LSTM)。张爱华等^[22]使用 CNN 代替 ANN 来提高准确性;K.Imoto 等^[23]使用基于 CNN 的迁移学习方法解决模型使用早期数据较少问题,Liu Z 等^[24]发现产品加工过程与自然语言处理过程类似,提出使用 LSTM 方法搭建端对端框架,Bai Y 等^[25]提出使用粗糙知识和 AdaBoost 方法增强性能。

深度学习还可同知识图谱进行结合。王宵等^[26]从半结构化数据和非结构化数据提取工艺参数和质量指标,构造知识图谱,将知识图进行词向量嵌入,输入双门控循环单元和注意力机制结合的深度神经网络,预测制丝工艺中的松散回潮质量。

集成学习,即将多个弱学习器进行聚合,并选择高性能的学习器进行预测。Hung Y H^[27]集成了神经网络和提升决策树,提出自适应增强决策树算法;A.S.Struchtrup 等^[28]提出全面质量预测框架,自动执行数据预处理、特征选择、训练和验证,包括超参数优化,使用 7 种监督学习算法进行集成学习;针对离散型制造生产数据,蒋金瑜等^[5]采用随机森林、XGBoost、Light GBM 和 Cat Boost 4 种集成学

习算法进行质量分类预测,融合人工神经网络,进一步提高了分类性能。

1.3 质量控制和质量追溯

质量控制,针对过程质量实时管理和分析,对于某一个时序变量进行监控,若该变量变化趋势不符合预期则进行实时报警,主要采用统计过程控制方法(statistical process control, SPC)。该方法最早由 W.A.Shewhart^[29]提出,称为休哈特控制图。SPC 可对关键质量特性进行实时监控,建立分析用控制图和控制用控制图,对人员操作、加工精度、机器磨损、工艺变化、环境变化等影响造成的系统误差进行监控,对产品质量进行动态监测,实时图形化分析,形成质量趋势图。Page 等^[30]、Roberts 等^[31]、Lowry 等^[32]提出了累积和控制图、指数加权移动平均控制图及其改进,改进了休哈特控制图对于数据分布的敏感性,其侦测流程偏移及预警的效力与数据服从何种分布无关。SPC 方法只能进行过程异常波动的监控,无法对异常进行分析诊断以确定异常类型、发生原因和位置,故常与机器学习、模糊数学等方法进行结合以高效处理存在模糊性和不确定性特征的过程质量。如陈志强等^[33-34]提出了基于模糊信息的模糊控制图;刘艳永等^[35]提出了多元协方差阵未知时的样本多元协方差控制图;R.S.Guh 等^[36]通过将机器学习方法中的决策树方法检测多源控制图中的均值偏移,有效检测均值偏移,准确识别偏离原均值的变量以及其偏移方向;S.M.Acosta 等^[37]将差分进化算法优化的贝叶斯稀疏核的相关向量机融合控制图,监控产品故障。

质量追溯是指在产品出现质量问题时,追溯产品生产过程中所涉及各类影响因素,寻找质量问题产生的原因,改进生产制造过程。质量追溯可以分为质量追踪和质量溯源。质量追踪多使用 RFID 技术串联产品生命周期信息,建立完成质量追溯信息链条,如 Wang J 等^[38]开发了针对柔性制造装配线的信息追溯系统,陈晓明^[39]从关键节点出发建立产品追溯模型。质量溯源则多使用关联知识挖掘方法,建立影响因素和质量之间的影响关系进行质量溯源,如 Bernard 等^[40-41]挖掘制造零件大小功能障碍和启动延迟之间的关联关系,I.Kunttu 等^[42]将缺陷、缺陷原因以及其对应的缺陷图像关联起来,He Y 等^[43]建立功能、物理和过程产生的节点和故障之间的关联关系。通过识别产品生命周期不同阶段可能的成因因素,可以进行质量追溯。

2 挑战

质量分析存在以下挑战: 1) 多源异构且存在噪声的数据融合问题。数据多源异构, 表现差异大, 采样策略多, 数据价值密度低, 没有标准化的规则来指示应将哪些数据预处理技术应用于特定类型的制造问题, 难以建立标准规则进行解析。2) 数据丢失问题。机器传感器收集时间序列数据时, 可能无法在特定时间段内获取正确的测量值, 这往往被记录为缺失, 或者具有超出其属性的最小或最大限制的不相关数值, 丢失蕴含信息。3) 数据不平衡问题和维度灾难问题。产品制造时多为正常制造生产, 产品出现质量问题的概率较低, 因此, 正常数据远多于非正常数据, 这会导致数据不平衡问题, 同时, 质量数据高维会导致维度灾难, 导致使用机器学习方法存在挑战。4) 知识抽取问题。各类故障机理与演化规律隐含在以振动、声场、图像、文本等多源异构大数据中, 如何构建自动化方法, 将这些数据进行融合并分析萃取大数据中的本质和演化机理信息存在一定挑战。

3 发展趋势

质量分析的发展趋势如下: 1) 数据中心发展。从专业领域下的小样本数据库向标准大数据库和知识图谱转变, 以共享数据, 复用模型, 优化方法, 如通过知识添加数据集避免采集数据不充分问题, 调整数据分布以解决数据集偏移问题, 设计神经网络结构以提高精度, 约束神经网络的预测值于控制方程投影的空间内。2) 验证手段发展。从数字可视化向数字孪生转变。以往多使用设备参数可视化、质量指标可视化、预测结果可视化等手段来挖掘验证因果关系、相关关系和互斥关系, 现通过未加工部件的虚拟加工与验证可进行事前分析质量, 并进行实时改进。3) 预测模型发展。从浅层智能模型向深度学习模型转变以实现自动特征提取、复杂关系拟合以及提高模型精度, 可从领域泛化、特征推理和知识嵌入进行模型优化, 如改进应用于 NLP 和 CV 方向的深度学习方法、使用知识推理以提高深度学习模型在不完整数据集上的精确度、运用知识嵌入优化深度学习模型网络拓扑。4) 计算手段发展。从云计算向边缘计算转变。以往由于硬件限制, 需要将数据汇总到在云计算中心来高效使用其计算资源, 现可在边缘设备进行分层分析, 再汇总到云计算中心来进行分析计算。

4 结束语

笔者首先总结了质量分析中质量预测、质量控制和质量追溯的研究进展, 其次探索了之后质量分析任务面临的挑战, 包括数据融合问题、数据丢失问题、数据不平衡问题和维度灾难问题、知识抽取问题, 最后提出了之后产品质量分析的发展趋势, 包括数据中心发展、验证手段发展、预测模型发展和计算手段发展, 希望可以为进一步展开大数据下离散制造业产品质量分析的研究提供启发。

参考文献:

- [1] YIN S, KAYNAK O. Big Data for modern industry: challenges and trends [point of view][J]. Proceedings of the IEEE, 2015, 103(2): 143-146.
- [2] HASHEM I A T, YAQOOB I, ANUAR N B, et al. The rise of "big data" on cloud computing: Review and open research issues[J]. Information Systems, 2015, 47: 98-115.
- [3] LI J, TAO F, CHENG Y, et al. Big data in product lifecycle management[J/OL]. The International Journal of Advanced Manufacturing Technology, 2015, 81(1): 667-684.
- [4] 方喜峰, 赵良才, 吴洪涛. 基于数据挖掘的产品质量控制建模方法[J]. 机械工程学报, 2005(11): 24-29.
- [5] 蒋金瑜. 面向离散型制造的生产数据采集与产品质量预测方法研究[D]. 重庆: 重庆交通大学, 2021.
- [6] WANG T, CHEN Y, QIAO M, et al. A fast and robust convolutional neural network-based defect detection model in product quality control[J]. The International Journal of Advanced Manufacturing Technology, 2018, 94(9): 3465-3471.
- [7] FENG Y, WANG T, HU B, et al. An integrated method for high-dimensional imbalanced assembly quality prediction supported by edge computing[J]. IEEE Access, 2020, 8: 71279-71290.
- [8] 孔玲爽, 阳春华, 王雅琳. 一种基于灰色模型和机理模型集成的质量预测模型及其应用[J]. 湖南文理学报(自然科学版), 2007, 19(2): 68-70.
- [9] 林涛. 基于粗糙集知识建模的焊缝成形质量专家系统预测研究[D]. 上海: 上海交通大学, 2008.
- [10] NADA O A, ELMARAGHY H A, ELMARAGHY W H. Quality prediction in manufacturing system design[J]. Journal of Manufacturing Systems, 2006, 25(3): 153-171.
- [11] ZHANG S, DUBAY R, CHAREST M. A principal component analysis model-based predictive controller for controlling part warpage in plastic injection molding[J/OL]. Expert Systems with Applications, 2015, 42(6): 2919-2927.
- [12] LYNN S, RINGWOOD J V, MACGEARAILT N. Weighted windowed PLS models for virtual metrology of an industrial plasma etch process[C]//2010 IEEE

- International Conference on Industrial Technology. IEEE, 2010: 309-314.
- [13] 刘银华, 孙芮, 吴欢. 基于车身尺寸数据流潜结构建模的装配质量预测控制[J]. 中国机械工程, 2019, 30(2): 237-243.
- [14] 孙林, 杨世元. 基于 SVM 的柔性生产模式下生产过程质量智能预测[J]. 系统工程理论与实践, 2009, 29(6): 139-146.
- [15] 陈勤和. 发动机装配拧紧工艺与冷试振动相关性及其质量预测研究[D]. 重庆: 重庆大学, 2018.
- [16] 裘稼荣, 曾鹏飞, 邵伟平, 等. 基于 PSO-LSSVM 弹药装配质量预测方法[J]. 兵工学报, 2022, 43(9): 2379-2387.
- [17] 陈资, 陈春雨, 张根保. 基于 GA-SVM 的数控机床装配质量预测模型[J/OL]. 制造技术与机床, 2021(9): 97-100, 106.
- [18] 刘明周, 吕旭泽, 王小巧. 发动机曲轴多工序装配的质量预测模型研究[J]. 汽车工程学报, 2016, 6(1): 22-28.
- [19] CHEN W C, TAI P H, WANG M W, et al. A neural network-based approach for dynamic quality prediction in a plastic injection molding process[J]. Expert Systems with Applications, 2008, 35(3): 843-849.
- [20] 张朝武. M 型变速箱装配总成密封质量控制研究[D]. 重庆: 重庆理工大学, 2021.
- [21] 冯尔磊. 基于 SMT 大数据的产品质量预测方法研究与实现[D]. 西安: 西安电子科技大学, 2018.
- [22] 张爱华, 魏浩, 马晶, 等. 基于时间序列深度学习的超窄间隙焊接质量预测方法[J]. 电焊机, 2020, 50(8): 43-47, 138.
- [23] IMOTO K, NAKAI T, IKE T, et al. A CNN-based transfer learning method for defect classification in semiconductor manufacturing[C]//2018 international symposium on semiconductor manufacturing (ISSM). IEEE, 2018: 1-3.
- [24] LIU Z, ZHANG D, JIA W, et al. An adversarial bidirectional serial-parallel LSTM-based QTD framework for product quality prediction[J]. Journal of Intelligent Manufacturing, 2020, 31(6): 1511-1529.
- [25] BAI Y, XIE J, WANG D, et al. A manufacturing quality prediction model based on AdaBoost-LSTM with rough knowledge[J]. Computers & Industrial Engineering, 2021, 155: 107227.
- [26] 王宵, 徐成现. 结合知识图谱与深度神经网络的烟丝松散回潮质量预测方法[J]. 轻工机械, 2022, 40(4): 100-104.
- [27] HUNG Y H. Improved Ensemble-Learning Algorithm for Predictive Maintenance in the Manufacturing Process[J/OL]. Applied Sciences, 2021, 11(15).
- [28] STRUCHTRUP A S, KVAKTUN D, SCHIFFERS R. Adaptive quality prediction in injection molding based on ensemble learning[J]. Procedia CIRP, 2021, 99: 301-306.
- [29] SHEWHART W A. Quality control charts[J/OL]. The Bell System Technical Journal, 1926, 5(4): 593-603.
- [30] PAGE E S. Continuous inspection schemes[J]. Biometrika, 1954, 41(1/2): 100-115.
- [31] ROBERTS S W. Control Chart Tests Based on Geometric Moving Averages[J]. Technometrics, 1959, 1(3): 239-250.
- [32] LOWRY C A, WOODALL W H, CHAMP C W, et al. A Multivariate Exponentially Weighted Moving Average Control Chart[J]. Technometrics, 1992, 34(1): 46-53.
- [33] 陈志强, 张公绪. 模糊技术在控制图中的应用[J]. 中国质量, 1996.
- [34] 陈志强, 卜祥民. 模糊控制图及其应用[J]. 北京科技大学学报, 1996(S1): 56-58.
- [35] 刘艳永, 孙静, 张公绪. 多元协方差控制图[J]. 北京航空航天大学学报, 1999(1): 89-91.
- [36] GUH R S, SHIUE Y R. An effective application of decision tree learning for on-line detection of mean shifts in multivariate control charts[J]. Computers & Industrial Engineering, 2008, 55(2): 475-493.
- [37] ACOSTA S M, OLIVEIRA SANT'ANNA A M. Machine learning-based control charts for monitoring fraction nonconforming product in smart manufacturing[J/OL]. International Journal of Quality & Reliability Management, 2022, ahead-of-print[2023-02-20]. <https://doi.org/10.1108/IJQRM-07-2021-0210>.
- [38] WANG J, WONG L E C. RFID-enabled tracking in flexible assembly line[J]. The International of Advanced Manufacturing Technology, 2010, 46(1-4): 351-360.
- [39] 陈晓明. 面向离散制造的订单跟踪与产品质量追溯系统研究[D]. 杭州: 浙江大学, 2010.
- [40] KAMSU-FOGUEM B, RIGAL F, MAUGET F. Mining association rules for the quality improvement of the production process[J]. Expert Systems with Applications, 2013, 40(4): 1034-1045.
- [41] FOURNIER-VIGER P, NKAMBOU R, TSENG V S M. Rule Growth: Mining Sequential Rules Common to Several Sequences by Pattern-Growth[C]//Proceedings of the 2011 ACM Symposium on Applied Computing. New York, NY, USA: Association for Computing Machinery, 2011: 956-961.
- [42] KUNTTU I, LEPISTÖ L, RAUHAMAA J, et al. Multiscale Fourier descriptors for defect image retrieval[J]. Pattern Recognition Letters, 2006, 27(2): 123-132.
- [43] HE Y, ZHU C, HE Z, et al. Big data oriented root cause identification approach based on Axiomatic domain mapping and weighted association rule mining for product infant failure[J]. Computers & Industrial Engineering, 2017, 109: 253-265.