

一种基于制造大数据的产品工艺自适应设计方法

巍巍, 陈政, 袁君

(北京航空航天大学机械工程及自动化学院, 北京 100191)

摘要: 随着数字化与智能化生产方式在制造业中的广泛应用, 引导企业重视并发挥制造大数据的价值对革新产品工艺设计具有重要意义。本文旨在面向企业数据与业务深度融合的应用需求, 提出一种基于制造大数据挖掘的产品工艺自适应设计应用方法, 用于解决企业中制造数据利用率不足等问题。以企业制造数据为起点, 提出了“数据+知识+决策”的产品工艺自适应设计模式, 总结了该模式的制造数据挖掘与利用流程, 涵盖多源异构数据融合、数据清洗与预处理、数据变换与降维、数据挖掘、数据可视化和设计决策 6 个过程。最后以汽车产品焊接工艺为例, 建立焊接工艺参数与焊接缺陷的预测模型, 用于改善焊接工艺并提高焊接质量, 实现制造大数据驱动的焊接工艺自适应设计。研究表明, 制造大数据蕴含丰富的知识与模式, 可以指导产品设计决策, 支持实现不同制造环境下的产品工艺自适应设计; 建议进一步推动大数据与第五代移动通信技术等新兴技术的结合, 增加对大数据平台、算法设计平台研发的投入, 激发制造大数据与产品工艺设计的更大交融。

关键词: 产品工艺; 自适应设计; 制造大数据; 数据挖掘; 知识发现

中图分类号: TB472 **文献标识码:** A

A Product Process Adaptive Design Method Based on Manufacturing-Related Big Data

Wei Wei, Chen Zheng, Yuan Jun

(School of Mechanical Engineering and Automation, Beihang University, Beijing 100191, China)

Abstract: As digital and smart production methods being applied widely in manufacturing, enterprises should pay more attention to the values of manufacturing-related big data, which is important to the innovation of product process design. This study aims to propose a product process adaptive design method based on manufacturing-related big data. This method is proposed based on the requirement of enterprises for data-business deep integration and it is used for solving the insufficient utilization of manufacturing-related data among enterprises. To this end, a product process adaptive design model “data + knowledge + decision” is proposed and the data mining and utilization processes are summarized for the model, namely, multi-source heterogeneous data fusion; data cleaning and preprocessing; data conversion and dimensionality reduction; data mining; data visualization; and design decision. Subsequently, the automobile welding process is used as an example. A prediction model of the relationship between welding parameters and welding defects is established, aiming to improve welding quality and realize welding process adaptive design. This research reveals that manufacturing-related big data contains rich knowledge and patterns and thus can guide product design decisions and support the product process adaptive design under different manufacturing environments. In the future, to enhance the integration of manufacturing-related big data

收稿日期: 2020-05-26; 修回日期: 2020-06-28

通讯作者: 巍巍, 北京航空航天大学机械工程及自动化学院副教授, 研究方向为产品设计与智能制造; E-mail: weiwei@buaa.edu.cn

资助项目: 国家重点研发计划资助项目 (2018YFB1701703); 国家自然科学基金资助项目 (51675028); 中国工程院咨询项目“‘互联网+’行动计划战略研究 (2035)” (2018-ZD-02)

本刊网址: www.engineering.org.cn/ch/journal/sscae

with product process design, the integration of big data with 5G technology should be promoted and investment should be increased in the development of big data and algorithm design platforms.

Keywords: product process; adaptive design; manufacturing-related big data; data mining; knowledge discovery

一、前言

大数据、云计算等新兴信息技术快速兴起, 与现代工业技术迅速结合, 产生了众多新型生产制造模式, 如智能制造、数字孪生等 [1], 数据驱动成为解决复杂问题的有效途径与重要手段。随着数控机床、传感器、数据采集器和其他设备器件的大量应用, 制造企业在生产过程中积累了大量的制造数据 [2]。这些制造数据呈现出典型的大数据特性 [3], 数据背后隐含的关联规则与制造知识尚待发现, 需要进行数据挖掘以有效指导产品工艺的迭代和优化设计, 进而提高产品工艺设计对制造环境的适应性。

传统的产品工艺设计方法较多基于仿真与实验来进行 [4~6], 产品工艺设计与数据驱动的典型流程如图 1 所示。具体而言, 在新的设计任务出现后, 设计人员借助设计软件进行产品工艺的设计与规划; 在实际生产之前试制产品并进行试验验证, 确定最终设计方案后进行实际制造; 在实际制造过程中, 收集机器、刀具、工件等实时数据监控生产状态 [7]; 对生产出来的产品进行质量检测 [8], 产生产品检测数据。制造数据的挖掘过程就是通过分析处理制造数据, 探索出产品质量与工艺参数间的关系和规律; 新发现的知识可运用知识库技术进行规范化表达与结构化存储, 辅助设计优化决策 [9], 改进产品设计, 实现产品工艺的“后向设计”。目前基于实验与仿真的产品工艺设计方法在制造企业中应用相对成熟, 而制造大数据驱动的“后向设计”有所不足, 亟需解决产品设计对制造系统环境自适应差、产品迭代与版本更新滞后等问题。

数据挖掘兴起于 20 世纪 90 年代, 究其本质在于从海量数据中提取、分析和发现知识 [10,11]。近年来, 研究者非常重视制造大数据的应用。Köksal 等 [8] 综述了 1997—2007 年数据挖掘在产品质量改进方面的应用情况, Ferrerio 等 [7] 将数据挖掘应用于飞机零部件钻孔工艺的优化决策, Zhang

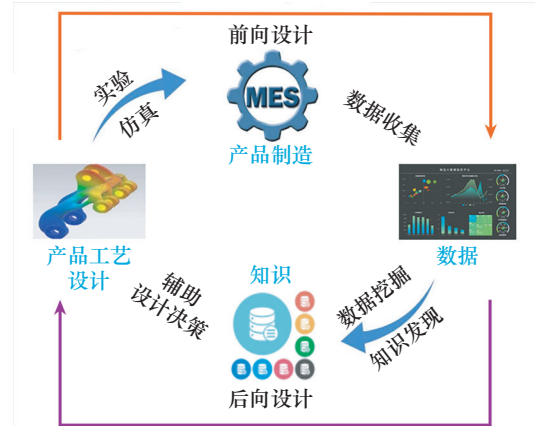


图 1 产品工艺设计与数据驱动流程图
注: MES 表示制造企业生产过程执行管理系统。

等 [12] 通过挖掘分析用户评论数据以优化产品创新、更新产品市场, Kretschmer 等 [13] 研究了数据挖掘在装配工艺设计决策中的应用, Tao 等 [14,15] 提出了基于大数据的数字孪生驱动的产品设计框架, Ma 等 [16] 提出了由数据驱动的产品族设计方法并在产品族设计实例中应用, Fu 等 [17] 举证分享了数据驱动产品创新的指导意见, Yu 等 [18] 讨论了大数据世代产品设计的新模式。

传统的数据挖掘技术多是针对结构化数据展开的, 而在海量、多源、异构的制造大数据场景下进行应用略显不足, 需要运用新兴技术进行数据融合、处理和分析。为此, 本文提出一种新型的“数据+知识+决策”产品工艺自适应设计模式, 以发挥制造大数据在产品设计中的驱动作用。

二、产品工艺自适应设计方法分析

随着数控机床、传感器在生产车间中的广泛应用以及物联网、人工智能等技术的快速发展, 制造企业收集到的制造数据越来越多, 数据增加速度也越来越快。如何充分发挥制造数据的价值成为企业发展面临的难题和机遇。“数据+知识+决策”的产品工艺自适应设计模式是解决这一问题的有效手段, 可以充分挖掘制造数据的潜在价值, 驱动

产品工艺的自适应设计，促进产品服务和产品质量的提升。

数据驱动的产品工艺自适应设计模式的数据挖掘过程如图 2 所示，具体步骤为：①从制造系统产生的数据出发形成制造大数据，分为结构化数据和非结构化数据，其中非结构化数据通过特征提取与信息提取等方式进行结构化处理来实现异构数据融合；②对海量数据进行选择与抽样，使抽样样本既能体现整体的分布情况，又不增加数据分析计算的复杂度；③对抽样后的数据进行清洗与预处理；④通过数据变换使数据降维，成为低维数据；⑤进行数据挖掘计算，挖掘数据中隐含的规律、模式和模型；⑥以数据可视化方式辅助设计决策，推动产品的自适应设计。

(一) 多源异构数据融合

在先进制造系统中，随着传感器种类的增多，产生的信息量不断增大，数据的表现形式也更加复杂，如力学传感器与速度传感器工作时产生时序数据、视觉传感器捕捉图像产生图像数据、产品数据管理系统 (PDM) 运行中产生数据等。制造大数据的显著特征是多样性、复杂性和不确定性，对制造大数据的统一表达是解决数据融合问题的关键和难点 [19]。

(1) 时序数据的融合。制造过程通常需要对机床和工件的状态进行监控，采集速度、应力、温度等时序数据，而这类数据的采集帧率较高

(40~80 fps)。为实现时序数据与其他类型数据的匹配和融合，需对时序数据进行下采样，运用平均值、方差等方法反映某一阶段的加工状态。

(2) 图像数据的融合。对于制造系统中的图像数据，需要提取图像中的信息以进行结构化表达。传统图像信息的提取是由手动完成的，效率较低。随着卷积神经网络模型在计算机视觉领域的应用，机器的图像感知能力有了跨越式发展，可运用机器进行图像数据的信息提取，将非结构化的图像数据转换为结构化的数据信息。

(二) 数据清洗与预处理

在数据驱动的产品工艺自适应设计模式中，数据的体量和质量都发挥着至关重要的作用。数据质量包括数据的准确性、完整性、一致性和有效性。其中，准确性指数据与物理世界相符合的程度，完整性指数据中有效值所占的比例，一致性指数据对指定约束的满足程度，有效性则表征数据的价值密度。

数据清洗指将“脏数据”清除以提高数据质量，包括数据异常值与缺失值的处理、去噪等。对于数据异常值，可以采用距离度量或聚类的方法检测数据集中的离群点，删除与数据集中心距离过大的数据点。对于数据集中的缺失值，插值是数据处理的有效手段，通过数据填充使数据集趋于完整。针对数据中的噪声，可以使用平滑滤波等算法进行去噪。制造大数据中通常包含大量重复的数据点，对于这

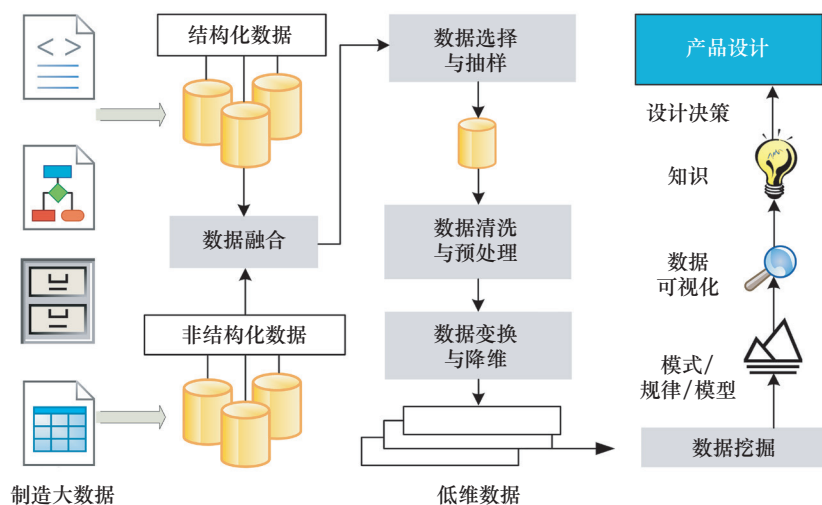


图 2 产品工艺自适应设计模式的数据挖掘过程

种数据集要进行数据降重,减少数据冗余。

(三) 数据变换与降维

制造大数据是制造系统与制造过程的数字化表达,制造系统中收集到的数据越多,对制造系统的完整描述就越有帮助,而这为数据挖掘工作带来维数灾难问题。维数灾难是处理高维数据时遇到的最大问题之一,不仅影响数据分析算法的时间和空间复杂度,还会导致数据分析算法的不收敛问题。

制造过程中采集的各种数据通常具有一定的相关性,如焊接过程中的电压与电流、机床主轴的转速与切削速度等。这种相关性会造成维度的冗余,增加不必要的计算,因而数据降维就显得尤为重要。数据降维指从高维的数据空间中保留合适的特征数据并剔除冗余数据,降低数据维度。降维后的数据既能保留原有信息量,又能避免维数灾难。

(四) 数据挖掘

制造大数据挖掘的主要目的是进行预测与规则提取。数据预测是用已知变量预测其他变量的未来值,数据规则提取则是找到数据中可被理解并可用于指导产品设计的隐藏规则。常用的数据挖掘方法有分类、回归、聚类与关联性分析 [20]。

(1) 分类分析与回归分析方法是进行数据预测的主要方法,都是对特征空间进行映射。分类分析方法将特征空间映射到离散的变量,回归分析方法将特征空间映射到连续的变量。分类分析和回归分析方法的预测过程如图 3 所示。

首先给定一个训练数据集 T :

$$T = (x_1, y_1), (x_2, y_2), \dots, (x_n, y_n) \quad (1)$$

式 (1) 中, $x_i \in R^n$ 是输入, $y_i \in R$ 是输出, 分类

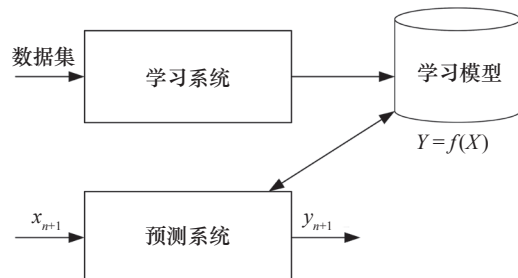


图 3 分类分析与回归分析方法的预测过程

问题 y_i 是离散型变量,回归问题 y_i 是连续型变量。目的是构建一个预测模型 $Y=f(X)$,对于新的输入 x_{n+1} 预测得到输出 y_{n+1} 。

(2) 聚类分析指用非监督学习的方法将实例分为自然组,基于实例数据背后的隐藏模式,将相似的实例聚集于对应的簇中。聚类分析方法在制造数据挖掘中广泛应用,例如,在故障预测中收集大量的故障实例,根据故障特征进行聚类,发现故障中可能存在的子类。常用的聚类算法有 K-means 法 [21]、最大期望值算法 [22]、贝叶斯聚类 [23] 等。

(3) 关联性分析可以表达为“规则形式”。例如,“if X then Y”的关联规则可以解释为满足条件 X 的某产品很可能也会满足条件 Y。制造大数据中的关联性分析是确定设计约束的重要因素之一,通过挖掘关联规则可以发现影响产品质量的因素,从而在设计决策时调整设计变量,达到自适应设计的目的。

(五) 数据可视化

数据可视化指数据在视觉上的直观表现形式,利用图像处理、计算机图形学 and 用户界面等对数据进行表达、建模和动画演示。数据可视化通常采用散点图、矩阵图、曲面等直观表现形式,在庞大的数据集中呈现有价值的信息 [24],加深用户对数据的理解,加快信息的获取速度。根据数据类型可将数据可视化分为高维数据可视化、时序数据可视化、层次数据可视化和网络数据可视化 [25]。

(六) 设计决策

通过数据挖掘得出的知识与关联模式可用于辅助设计决策,如决定设计参数、产品架构、工艺流程等(见图 4)。当确立新的执行方案并下达执行后,数字孪生系统将从制造执行系统中收集产生的数据。这些数据在设计决策中有如下应用:一是在数据库中存储用于数据挖掘;二是及时向设计者提供反馈并辅助设计决策,验证产品工艺设计决策的可行性与合理性。

三、制造大数据在产品工艺自适应设计中的应用——以数据挖掘为例

为更好地了解制造大数据在产品工艺设计中发挥的驱动作用,以数据挖掘在某汽车产品焊接工艺

优化中的应用为案例进行阐述。汽车产品在焊接制造过程收集到的工艺数据主要有用于焊接过程控制的工艺参数、用于质量检测的焊缝射线图像等。焊缝射线图像数据不能直接用于数据分析，需要先进行焊缝射线图像缺陷检测，然后再分析焊缝缺陷与焊接工艺参数之间的关系。

(一) 焊接射线图像缺陷检测

射线检测是焊接质量检测中广泛采用的方法，可以精确检测出缺陷的位置、形状、大小及分布情

况。对于射线检测过程中产生的焊缝射线图像（见图5），传统的图像处理方法多是由专业人员利用经验对缺陷进行检测评价来进行的，而这种方法存在检测效率低、对检验人员的技术能力要求高、无法适用于大量图像分析工作等问题。为此，运用语义分割算法进行焊缝缺陷的自动检测，将射线图像的每一个像素进行分类，如背景、气孔、夹渣、裂纹等。

基于卷积神经网络，采用“编码器-解码器”的网络结构进行焊缝射线图形语义分割（见图6）。对于输入的射线图像，首先进行“卷积+批归一化+激活函数”计算；然后使用 2×2 的最大池化层进行下采样，在进行5次相同的操作后，输入图像的尺寸变为原来的 $1/32$ ；然后反池化操作进行上采样，同样也进行“卷积+批归一化+激活函数”计

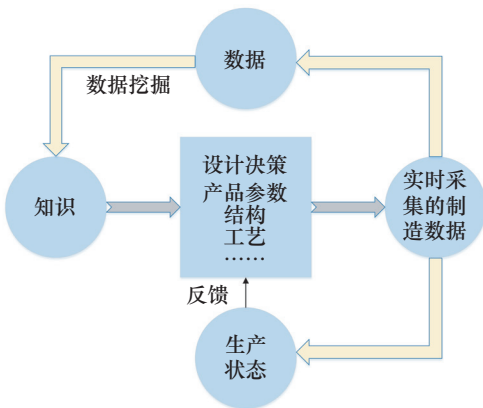


图4 数据挖掘辅助设计决策流程

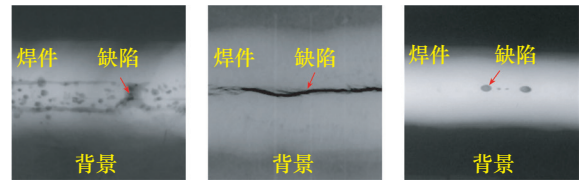


图5 焊缝射线图像示意图

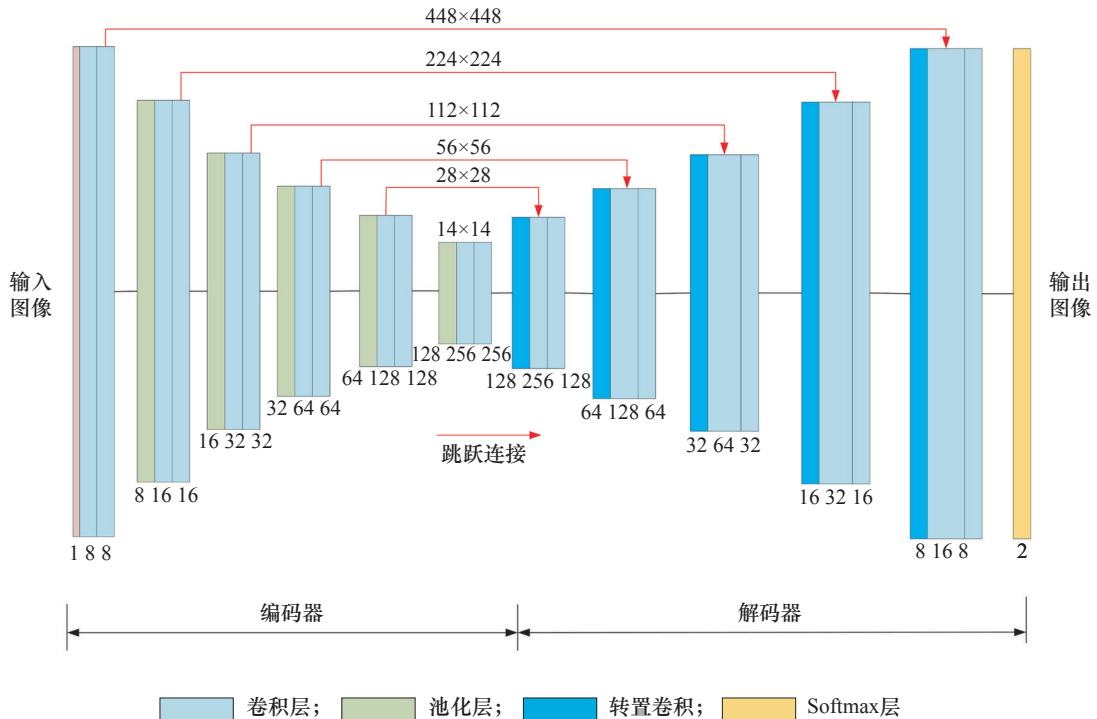


图6 基于“编码器-解码器”结构的焊缝射线图像语义分割

算,直至图像恢复到输入图像尺寸;最终使用归一化指数函数(Softmax)对输出像素值进行分类并统计各缺陷部分占总像素个数的比例。

(二) 工艺参数和焊接质量关联性模型

某品牌汽车在生产过程中采用了钢板预涂底漆后组对焊接的生产工艺,但产品焊接质量问题突出,亟需改进焊接工艺。已有数据包括 200 幅焊缝的射线图像,生产中用于过程监控采集的焊接电流、电压、保护气体流量,预涂漆膜的厚度等。针对上述数据,研究影响焊接质量的工艺因素以优化工艺设计。

(1) 用事先训练好的语义分割缺陷检测模型对射线图像进行检测,统计各类型缺陷的数量和比例。如图 7 所示,气孔缺陷问题在这批产品中较为突出,因此需要分析气孔缺陷与工艺参数间的关系。

(2) 对各图像对应的工艺参数时序数据进行下采样,得到每个焊缝所对应的工艺参数。利用皮尔逊相关系数法分析各工艺参数与气孔缺陷的相关性并进行可视化展示。可以看出,焊接电流、电压、保护气体流量、预涂漆膜的厚度对气孔缺陷均有影响(见图 8)。在此基础上,建立工艺参数与气孔数量间的回归模型,即采用全连接神经网络构建回归模型(见图 9)。模型输入层包括漆膜厚度、焊接速度、焊接电流、气体流量 4 个节点输入,输出层输出的是在该工艺参数条件下焊接产品的气孔率。

(3) 按照 6:4 的比例将 200 个实例数据分成训练集和验证集,选择均方差损失作为损失函数。如图 10 所示,经过 200 次迭代后均方差损失降到 0.02

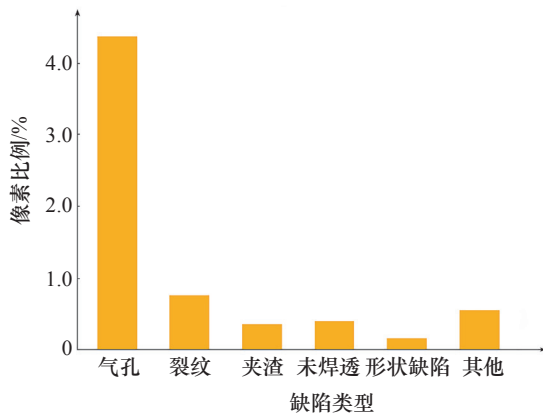


图 7 6 种缺陷类型的检测结果

以下,这表明缺陷检测回归模型具有较优的预测性能,在帮助设计人员选择工艺参数、减少焊接缺陷并提高产品质量方面具有较好的应用效果。

(三) 缺陷检测模型在产品工艺设计中的应用

缺陷检测模型可用于焊接产品的检验过程。传统手工测试方法效率低、检查周期长,焊接产品的设计人员无法及时获得焊接工艺的反馈结果。而缺陷检测模型可以根据焊接工艺与焊接缺陷之间的回归关系预测产品工艺设计方案是否合理,辅助设计者作出正确的设计决策,提高产品焊接项量。

四、结语

本文提出了一种制造大数据驱动的产品工艺自适应设计方法,阐述了制造数据如何在产品工艺设计中发挥价值、制造数据挖掘存在的问题以及数据挖掘应用步骤;以数据挖掘在焊接工艺中的应用为例进行焊接缺陷预测,推动焊接工艺过程改进。研究表明,制造大数据挖掘衍生出的知识与模式为产品设计人员提供了有效的辅助指导,推动了产品工艺的迭代优化,实现了大数据驱动产品工艺自适应设计的目标。

为更好地促进制造大数据驱动的产品工艺设计自适应方法的应用与推广,针对相关产业提出如下发展建议。

(1) 深入推动第五代移动通信(5G)技术与制造大数据的应用发展。制造大数据是进行数据挖掘与分析的基础,也是驱动企业向创新设计转型的动力。依托 5G 技术,制造企业可以收集到更全面的数据,从而实现生产全要素的分析。

(2) 加强大数据平台与算法设计平台的研发。大数据平台是对制造数据进行高效存储、读取、导入、导出的基础,研发稳定可靠的大数据平台可以提高数据挖掘的效率,保证知识利用的时效性。大数据平台也是确保企业信息安全的重要支撑。算法设计平台是进行数据挖掘算法快速设计的基础,研发集成度高的算法设计平台可以降低数据挖掘应用的门槛,让更多的企业设计人员运用制造大数据挖掘来创造价值。

(3) 推进企业的典型应用示范推广。重点提高

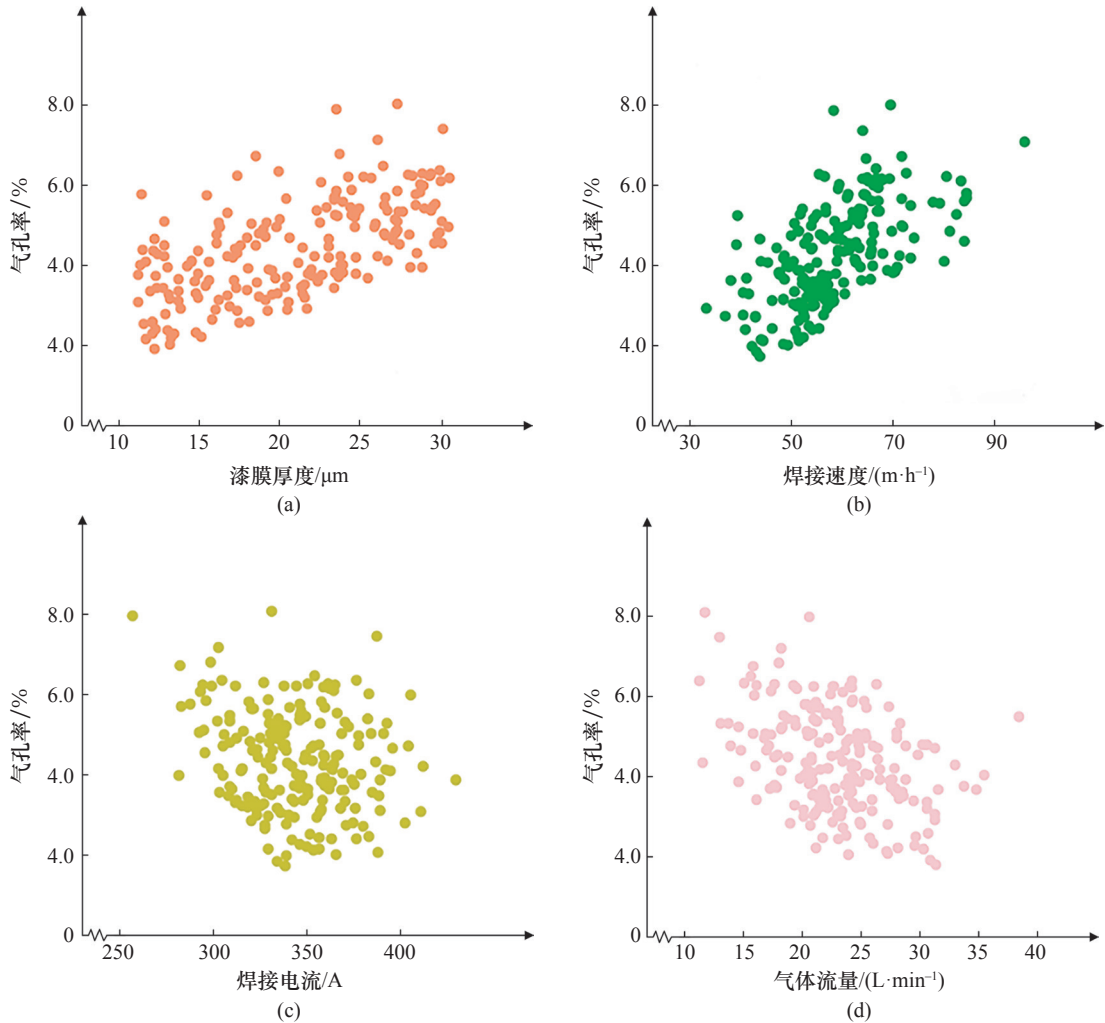


图8 焊接工艺参数与气孔缺陷的相关关系

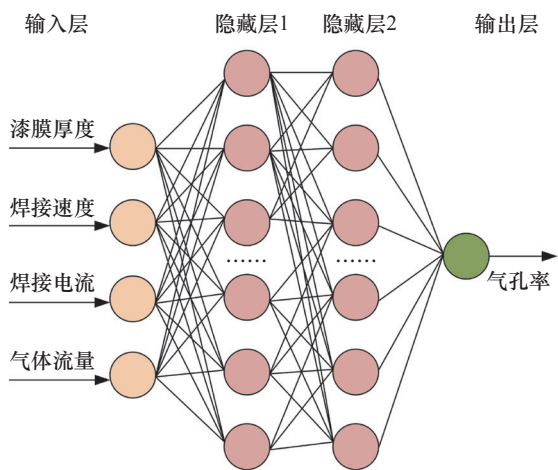


图9 全连接神经网络结构回归模型

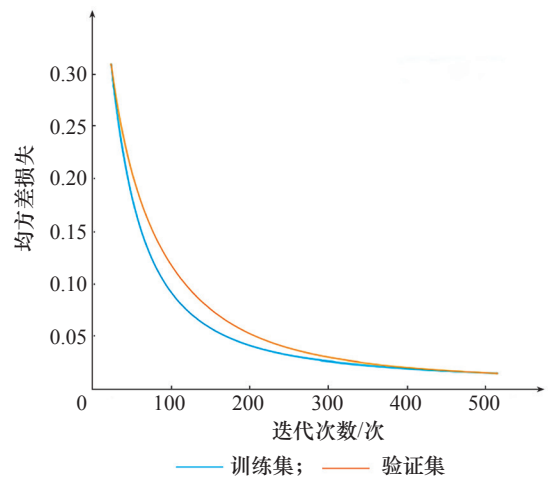


图10 训练集与验证集中的均方差损失

典型企业的应用示范推广水平，革新传统制造企业的生产观念，以先进、有效的应用案例来推广制造大数据在企业中的应用。

参考文献

- [1] Tao F, Qi Q, Liu A, et al. Data-driven smart manufacturing [J]. Journal of Manufacturing Systems, 2018, 48: 157-169.
- [2] Zhang Y, Ren S, Liu Y, et al. A framework for big data driven

- product lifecycle management [J]. *Journal of Cleaner Production*, 2017, 159: 229–240.
- [3] 王国成. 从3V到5V: 大数据助推经济行为的深化研究 [J]. *天津社会科学*, 2017 (2): 94–99.
Wang G C. From 3V to 5V: A further research of big data boosting economic behavior [J]. *Tianjin Social Sciences*, 2017 (2): 94–99.
- [4] Thomas W, Oenoki T, Altan T. Process simulation in stamping—Recent applications for product and process design [J]. *Journal of Materials Processing Technology*, 2000, 98 (2): 232–243.
- [5] 周方明, 刘威, 王贤, 等. 排气阀凹槽机器人自动化堆焊工艺设计 [J]. *江苏科技大学学报(自然科学版)*, 2014, 28(2): 135–139.
Zhou F M, Liu W, Wang X, et al. Exhaust valve groove robotic automation welding process design [J]. *Journal of Jiangsu University of Science and Technology (Natural Science Edition)*, 2014, 28(2): 135–139.
- [6] Qiao Y H, Han J, Zhang C Y. Jig of passenger vehicle chassis design and welding process simulation [C]. *Wuhan: 2010 International Conference on Mechanic Automation and Control Engineering*, 2010.
- [7] Ferreiro S, Sierra B, Irigoien I, et al. Data mining for quality control: Burr detection in the drilling process [J]. *Computers & Industrial Engineering*, 2011, 60(4): 801–810.
- [8] Köksal G, Batmaz I, Testik M C. A review of data mining applications for quality improvement in manufacturing industry [J]. *Expert Systems with Applications*, 2011, 38(10): 13448–13467.
- [9] Tucker C S, Kim H M. Data-driven decision tree classification for product portfolio design optimization [J]. *Journal of Computing and Information Science in Engineering*, 2011, 9(4): 1–14.
- [10] Fayyad U, Stolorz P. Data mining and KDD: Promise and challenges [J]. *Future Generation Computer Systems*, 1997, 13(2–3): 99–115.
- [11] Han J, Kamber M, Pei J. *Data mining: Concepts and techniques (third edition)* [M]. San Francisco: Morgan Kaufman Publishers, 2012.
- [12] Zhang H, Rao H G, Feng J Z. Product innovation based on online review data mining: A case study of Huawei phones [J]. *Electronic Commerce Research*, 2018, 18(1): 3–22.
- [13] Kretschmer R, Pfouga A, Rulhoff S, et al. Knowledge-based design for assembly in agile manufacturing by using data mining methods [J]. *Advanced Engineering Informatics*, 2017, 33: 285–299.
- [14] Tao F, Sui F Y, Liu A, et al. Digital twin-driven product design framework [J]. *International Journal of Production Research*, 2019, 57(12): 3935–3953.
- [15] Tao F, Cheng J F, Qi Q L, et al. Digital twin-driven product design, manufacturing and service with big data [J]. *The International Journal of Advanced Manufacturing Technology*, 2018, 94(9–12): 3563–3576.
- [16] Ma J & Kim H M. Product family architecture design with predictive, data-driven product family design method [J]. *Research in Engineering Design*, 2016, 27(1): 5–21.
- [17] Fu X, Asorey H. Data-driven product innovation [C]. *Sydney: The 21th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, 2015.
- [18] Yu C, Zhu L. Product design pattern based on big data-driven scenario [J]. *Advances in Mechanical Engineering*, 2016, 8(7): 1–9.
- [19] Bleiholder J, Naumann F. Data fusion [J]. *ACM Computing Surveys*, 2008, 41(1): 1–41.
- [20] 王光宏, 蒋平. 数据挖掘综述 [J]. *同济大学学报(自然科学版)*, 2004, 32(2): 246–252.
Wang G H, Jiang P. Survey of data mining [J]. *Journal of Tongji University(Natural Science Edition)*, 2004, 32 (2): 246–252.
- [21] 孙吉贵, 刘杰, 赵连宇. 聚类算法研究 [J]. *软件学报*, 2008, 19(1): 48–61.
Sun J G, Liu J, Zhao L Y. Clustering algorithms research [J]. *Journal of Software*, 2008, 19(1): 48–61.
- [22] 夏棒, Emilion R, 王惠文. Dirichlet混合样本的EM算法与动态聚类算法比较 [J]. *北京航空航天大学学报*, 2019, 45(9): 1805–1811.
Xia B, Emilion R, Wang H W. Comparison of EM algorithm and dynamic clustering algorithm of Dirichlet mixtum samples [J]. *Journal of Beijing University of Aeronautics and Astronautics*, 2019, 45(9): 1805–1811.
- [23] 朱杰, 陈黎飞. 类属数据的贝叶斯聚类算法 [J]. *计算机应用*, 2017, 37(4): 1026–1031.
Zhu J, Chen L F. Bayesian clustering algorithm for categorical data [J]. *Journal of Computer Application*, 2017, 37(4): 1026–1031.
- [24] 许文鹏. 数据可视化系统架构的设计与实现 [D]. 北京: 北京交通大学(硕士学位论文), 2015.
Xu W P. Design and implementation of data visualization system architecture [D]. Beijing: Beijing Jiaotong University(Master's thesis), 2015.
- [25] 汪玉薇. 面向中医处方的关联规则挖掘模型改进与应用研究 [D]. 武汉: 湖北中医药大学(硕士学位论文), 2017.
Wang Y W. A study on better models for association rule mining in Chinese medicine prescriptions and their applications [D]. Wuhan: Hubei University of Chinese Medicine(Master's thesis), 2017.