

基于大数据的智慧工厂制造优化技术研究

潘 伟

(安徽海螺信息技术工程有限责任公司,安徽 芜湖 241000)

摘 要:在分析智慧工厂国内外研究现状的基础上,对制造业如何利用物联网、云计算、大数据等新兴技术实现智慧工厂展开研究。探讨了集成制造物联、智能制造执行与制造协同的智慧工厂内涵,以及智慧工厂具备的大数据特征。基于工业大数据平台的应用,提出不依赖精确的数学模型、复杂的算法实现工厂运行分析与决策的新模式。在研究大数据环境下生产制造动态优化技术的基础上,探讨了其中涉及的大数据集成、大数据存储、相关性分析和相关性描述等大数据分析方法。改进频繁序列模式挖掘技术,提出一种能够支持结构化数据、非结构化数据,又无需建立复杂数学模型的设备故障知识推送方法,实现设备故障智能诊断功能。并以提高设备运转率为例,阐述该模式在生产工厂应用的可行性。

关键词:智慧工厂;大数据;决策分析;相关性分析;序列模式挖掘

中图分类号:TP18

文献标识码:A

文章编号:1673-629X(2019)01-0202-04

doi:10.3969/j.issn.1673-629X.2019.01.042

Research on Manufacturing Optimization of Intelligent Factory Based on Big Data

PAN Wei

(Anhui Conch Information Technology Engineering Company, Wuhu 241000, China)

Abstract: On the basis of the analysis of the current research situation at home and abroad in the intelligent factory, research how the manufacturing industry uses Internet of things (IoT), cloud computing, big data and other emerging technologies to achieve smart factories. The smart factory's nature composed of internet of thing, intelligent manufacturing execution system and manufacturing collaboration is discussed, and its big data characteristics is also analyzed. Based on the application of big data platform in industry, we propose a new mode without relying on accurate mathematical models and complex algorithm to realize plant operation analysis and decision. Aiming at the key technology of big data driven manufacturing process dynamic optimization, the big data analysis methods, including big data integration, big data storage, correlation analysis and correlation description is discussed, and a device fault knowledge push method based on frequent sequential pattern mining technology is proposed to realize the intelligent diagnosis function of equipment fault. Taking the equipment operation rate as an example, the feasibility of applying this model to plant is expounded.

Key words: smart factory; big data; decision analysis; correlation analysis; sequential pattern mining

0 引 言

随着生产要素成本的不断提高,中国制造业正经历着从低附加值、劳动密集型向高附加值、高技术含量转变。生产制造和装备制造业作为供给侧改革的重点领域,正处于升级换代过程中,需引入智能化和信息化技术,在融合工艺知识的基础上,推动生产、管理的转型,实现“降成本、补短板”和跨越式发展。

智慧工厂不同于传统的基于 MES 或智能设备管理的智能工厂模式,其应对工厂的制造执行进行智慧升华,在制造物联的基础上,发现规律并提供智能服务

的协同^[1-2]。随着智慧工厂业务需求的扩展与细分,传统的“数学建模+算法优化”模式的应用已显得捉襟见肘^[3-5],如何从工业大数据中挖掘知识来指导生产的优化,便成为学术界研究的重点。第一个大数据分析平台于 2013 年投入运行^[6],该平台通过流处理模式实时分析海量数据,并提供分析结果,让工业生产由被动执行转为主动预测模式。

文中将结合国内外研究工作,深入探讨生产制造和装备制造行业智慧工厂的内涵及特征,分析智慧工厂的技术体系及大数据关键技术,提出通过构建工业

收稿日期:2018-01-25

修回日期:2018-05-30

网络出版时间:2018-09-21

基金项目:国家重点研发计划资助项目(2016YFB0303405)

作者简介:潘 伟(1981-),男,硕士,高级工程师,硕导,研究方向为智能制造系统及数据挖掘分析等。

网络出版地址: <http://kns.cnki.net/kcms/detail/61.1450.TP.20180920.1536.034.html>

和综合性能预测模型;最后基于预测模型优化控制参数,以保证设备运转率等指标达到要求。

3 大数据驱动的智慧工厂制造优化技术

针对提高设备故障诊断的知识重用效率问题,文中提出一种基于序列模式挖掘技术的既能支持非文本知识数据又无需建立复杂数据模型的设备故障诊断知识推送方法^[16-17]。该方法通过对大量故障诊断知识所使用的检测数据频率进行分析,提取故障模型中的频繁知识序列,然后结合设备工程师的故障诊断知识使用行为和频繁知识序列对故障诊断的支持度,实现设备故障知识的推送,从而达到设备故障智能诊断功能。

频繁知识序列表示众多设备工程师对设备故障分析时参考的共同知识项,是进行知识需求判断的重要依据。基于序列模式挖掘技术的频繁知识序列挖掘,能够从大量的序列中找到满足最低支持度的频繁序列。支持度表示为:

$$\text{st}(s) = \frac{\text{count}(s)}{\text{size}(D)} \quad (1)$$

其中, $\text{count}(s)$ 为以 s 为子序列的序列数量; $\text{size}(D)$ 为数据库中序列的总数量。

将 GSP(generalized sequential pattern) 算法应用于频繁知识序列的挖掘^[18-19],可识别出故障模型中所有叶节点的频繁知识序列。设备故障模型节点 t_i 的频繁知识序列集为 $\text{FMS}_i = \{\text{fms}_{i1}, \text{fms}_{i2}, \dots, \text{fms}_{in}\}$, 其中 $\text{fms}_{ij} = \langle e_1, e_2, \dots, e_n \rangle$ 是第 j 个频繁知识序列。频繁知识序列的支持度记作 $\text{st}(\text{fms}_{ij})$

在设备故障诊断知识项推送时,要考虑当前设备故障所产生的振动频率、工艺参数、电流电压、质量检测等知识与频繁知识序列之间的相似度。在生产实际过程中,故障知识序列通常较长,与频繁知识序列完全匹配的可能性较小。文中提出按照最近、最高频的原则,抽取故障诊断知识项集合组成故障活动窗口^[20],并假设活动窗口的长度为 3 进行分析和知识推送。

假设设备工程师正在诊断的是设备 t_i 的故障;序列 $\text{cms} = \langle b_1, b_2, \dots, b_v \rangle$ 为当前设备运行所产生的知识序列,其中 b_v 是最近故障或供电异常的知识项所组成的元素; $w = \langle b_{v-2}, b_{v-1}, b_v \rangle$ 为该知识序列的故障活动窗口。从设备 t_i 的频繁故障知识序列集合 FMS_i 中选出与 w 具有相同 3 元素的频繁知识序列,组成集合 $\text{FMS}_i = \{\text{fms}_{v1}, \text{fms}_{v2}, \dots, \text{fms}_{vn}\}$ 。截取 fms_{vj} 中与 w 具有相同 3 元素序列的 3 元素子序列及其下一个元素,组成 4 元素的序列 $v_j = \langle c_{j1}, c_{j2}, c_{j3}, c_{j4} \rangle$, $V = \{v_1, v_2, \dots, v_n\}$ 为上述 4 元素序列的集合。

设 $\text{rm} = \{m_1, m_2, \dots, m_n\}$ 为当前知识使用行为是

“最近故障或供电异常”时,应被推送的知识项所组成的集合,其中 $m_i \in \{c_{13} \cup c_{23} \cup \dots \cup c_{n3}\}$, 且 $k_i \notin b_v$ 。集合 rm 中的知识项将按照推送系数 $I(m_i)$ 的降序进行推送,推送系统如下:

$$I(m_i) = \sum_{j=1}^n \text{st}(\text{fms}_{vj}) \left(\frac{|c_{j1} \cap b_{v-2}|}{|c_{j1} \cup b_{v-2}|} + \frac{|c_{j2} \cap b_{v-1}|}{|c_{j2} \cup b_{v-1}|} + \frac{|c_{j3} \cap b_v|}{|c_{j3} \cup b_v|} \right) \times y$$

$$y = \begin{cases} 1, m_i \in c_{j3} \\ 0, m_i \notin c_{j3} \end{cases} \quad (2)$$

设 $\text{rm} = \{m_1, m_2, \dots, m_n\}$ 为当前知识使用行为是“负荷异常”时,应被推送的知识项所组成的集合,其中 $m_i \in \{c_{13} \cup c_{23} \cup \dots \cup c_{n4}\}$ 。集合 rm 中的知识项将按照推送系数 $I(m_i)$ 的降序进行推送,推送系统如下:

$$I(m_i) = \sum_{j=1}^n \text{st}(\text{fms}_{vj}) \left(\frac{|c_{j1} \cap b_{v-2}|}{|c_{j1} \cup b_{v-2}|} + \frac{|c_{j2} \cap b_{v-1}|}{|c_{j2} \cup b_{v-1}|} + \frac{|c_{j3} \cap b_v|}{|c_{j3} \cup b_v|} \right) \times y$$

$$y = \begin{cases} 1, m_i \in c_{j4} \\ 0, m_i \notin c_{j4} \end{cases} \quad (3)$$

下文举例说明该推送方法的具体实现过程。假设当前设备故障所产生的知识序列为 $\text{cms} = \langle (E, F)(B, F)(C, D, G)(D) \rangle$, 当前的设备故障诊断任务是任务节点 t_i 的实例,任务节点 t_i 的频繁知识序列集合为 $\text{FMS}_i = \langle \text{fms}_{i1}, \text{fms}_{i2}, \text{fms}_{i3} \rangle$, 其中 $\text{fms}_{i1} = \langle (B)(C, D, F)(B, D, E)(B, G)(C, D) \rangle$, 其支持度 $\text{st}(\text{fms}_{i1}) = 30\%$; $\text{fms}_{i2} = \langle (B, D)(C, E)(D, E)(D, G)(B, F)(H, I) \rangle$, 其支持度 $\text{st}(\text{fms}_{i2}) = 40\%$; $\text{fms}_{i3} = \langle (D, G)(B, C)(B)(B, E) \rangle$, 其支持度 $\text{st}(\text{fms}_{i3}) = 35\%$ 。

(1) 捕捉当前故障知识序列的活动窗口 $w = \langle (B, F)(C, D, G)(D) \rangle$ 。

(2) 从集合 FMS_i 中选出与 w 具有相同 3 元素连续子序列的序列,组成集合 $\text{FMS}_v = \langle \text{fms}_{v1}, \text{fms}_{v2} \rangle$ 。其中, $\text{fms}_{v1} = \text{fms}_{i1} = \langle (B)(C, D, F)(B, D, E)(B, G)(C, D) \rangle$, 其支持度 $\text{st}(\text{fms}_{v1}) = \text{st}(\text{fms}_{i1}) = 30\%$; $\text{fms}_{v2} = \text{fms}_{i2} = \langle (B, D)(C, E)(D, E)(D, G)(B, D)(H, I) \rangle$, 其支持度 $\text{st}(\text{fms}_{v2}) = \text{st}(\text{fms}_{i2}) = 40\%$ 。

(3) 截取序列 fms_{vj} 中与 w 有 3 元连续子序列的子序列及其下一个元素,构成集合 $v_1 = \langle (B)(C, D, F)(B, D, E)(B, G) \rangle$, $v_2 = \langle (B, D)(C, E)(D, E)(D, G) \rangle$, $V = \{v_1, v_2\}$ 。

(4) 若设备工程师当前的知识使用行为是“最近故障或供电异常”,设 rm 为应被推送的知识项所组成

的集合,则 $rm = \{B, E\}$, 知识项 B 与 E 的推送系数分别为:

$$I(B) = 0.3 \times (\frac{1}{2} + \frac{2}{4} + \frac{1}{3}) = 0.4$$
$$I(E) = 0.3 \times (\frac{1}{2} + \frac{2}{4} + \frac{1}{3}) + 0.4 \times (\frac{1}{3} + \frac{1}{4} + \frac{1}{2}) \approx 0.83$$

因此,向设备工程师推送的知识项应符合 $\langle (E) (B) \rangle$ 的顺序。

(5)若设备工程师当前的知识使用行为是“负荷异常”,设 rm 为应被推送的知识项所组成的集合,则 $rm = \{B, D, G\}$, 知识项 B, D 和 G 的推送系数分别为:

$$I(B) = 0.3 \times (\frac{1}{2} + \frac{2}{4} + \frac{1}{3}) = 0.4$$
$$I(D) = 0.4 \times (\frac{1}{3} + \frac{1}{4} + \frac{1}{2}) \approx 0.43$$
$$I(G) = 0.3 \times (\frac{1}{2} + \frac{2}{4} + \frac{1}{3}) + 0.4 \times (\frac{1}{3} + \frac{1}{4} + \frac{1}{2}) \approx 0.83$$

因此,向设备工程师推送的知识项应符合 $\langle (G) (D) (B) \rangle$ 的顺序。

4 结束语

随着以计算机和互联网为代表的信息技术在生产制造和装备制造行业的不断深入应用,制造业在两化融合浪潮的推动下正在经历一场产业革命,向生产透明化、制造智能化、产业协同全球化的智慧工厂方向发展。文中在对智慧工厂内涵和工业大数据应用技术分析的基础上,讨论了基于大数据驱动的制造过程动态优化技术方法体系。在制造优化技术方面,提出一种基于序列模式挖掘的设备故障知识推送方法,重点讨论了在设备故障诊断过程中,针对种类多异的故障信息,如何提高故障诊断知识推送的支持度,提高了设备故障诊断知识推送的可实现性。后续工作将以生产专家控制、能源优化控制为应用背景,从提高生产效率、提升产品质量、降低产品能耗、提高设备利用率等角度出发,利用上述技术实现传统生产制造和装备制造工厂从数字化、信息化、智能化到智慧化发展的转型。

参考文献:

[1] CHAND S, DAVIS J F. What is smart manufacturing[J]. Time Magazine Wrapper, 2010, 16(1): 28-33.

[2] OCTAVIAN M, THEODOR B, SILVIU R. vMES: virtualization aware manufacturing execution system[J]. Computers in Industry, 2015, 67: 27-37.

[3] 刘晓冰,焦璇,宁涛,等. 基于双链量子遗传算法的柔性作业车间调度[J]. 计算机集成制造系统, 2015, 21(2): 495-502.

[4] 宋代立,张洁. 蚁群算法求解混合流水车间分批调度问题[J]. 计算机集成制造系统, 2013, 19(7): 1640-1647.

[5] 张其亮,陈永生. 求解双向无等待混合流水车间调度问题的粒子群优化算法[J]. 计算机集成制造系统, 2013, 19(10): 2503-2509.

[6] DUTTA R, MORSHED A, ARYAL J, et al. Development of an intelligent environmental knowledge system for sustainable agricultural decision support[J]. Environmental Modelling & Software, 2014, 52: 264-272.

[7] SONG Bo, JIANG Zuhua. Proactive search enabled context-sensitive knowledge supply situated in computer-aided engineering[J]. Advanced Engineering Informatics, 2013, 27(1): 66-75.

[8] 谢展鹏,张超勇,邵新宇,等. 基于 Memetic 算法的有限缓冲流水线车间调度问题[J]. 计算机集成制造系统, 2015, 21(5): 1253-1261.

[9] 崔维伟,陆志强,潘尔顺. 基于多目标优化的生产调度与设备维护集成研究[J]. 计算机集成制造系统, 2014, 20(6): 1398-1404.

[10] 刘明周,王小巧,马靖,等. 基于互信息和博弈论的复杂机械产品装配质量控制阈在线优化方法及应用[J]. 计算机集成制造系统, 2014, 20(11): 2798-2807.

[11] 姚锡凡,于森,陈勇,等. 制造物联的内涵、体系结构和关键技术[J]. 计算机集成制造系统, 2014, 20(1): 1-10.

[12] 中国科学院先进制造领域战略研究组. 中国至 2050 年先进制造科技发展路线图[M]. 北京: 科学出版社, 2009.

[13] 国家自然科学基金委员会工程与材料科学部. 机械工程学科技发展报告[M]. 北京: 科学出版社, 2010.

[14] 李振,贾瑞玉. 一种改进的 K-means 蚁群聚类算法[J]. 计算机技术与发展, 2015, 25(12): 28-31.

[15] 唐丹,张正军,王俐莉. 基于改进的近邻传播聚类算法的 Gap 统计研究[J]. 计算机技术与发展, 2017, 27(1): 182-185.

[16] 戚后林,顾磊. 基于密度与最小距离的 K-means 算法初始中心方法[J]. 计算机技术与发展, 2017, 27(9): 60-63.

[17] 谢志明,王鹏,黄焱. 多维数据 K-means 谱聚类算法改进研究[J]. 计算机技术与发展, 2017, 27(10): 60-64.

[18] 吕锋,张玮玮. 4 种序列模式挖掘算法的特性研究[J]. 武汉理工大学学报, 2006, 28(2): 57-60.

[19] SRIKANT R, AGRAWAL R. Mining sequential patterns: generalizations and performance improvements[M]. Berlin, Germany: Springer-Verlag, 1996.

[20] MOBASHER B, DAI H, LUO T, et al. Effective personalization based on association rule discovery from web usage data[C]//Proceedings of the 3rd international workshop on web information and data management. Atlanta, Georgia, USA: ACM, 2001: 9-15.