

• 大数据驱动的中药智能制造专题 •

中药工业大数据关键技术与应用

徐冰^{1,2,3}, 史新元^{1,2,3}, 罗赣^{2,3}, 林兆洲⁴, 孙飞⁵, 戴胜云⁶, 张志强^{2,7,8},
肖伟^{9,10,11}, 乔延江^{1,2,3*}

(1. 北京中医药大学 中药信息学系, 北京 102400; 2. 北京市科委 中药生产过程控制与质量评价北京市重点实验室, 北京 102400; 3. 教育部 中药制药与新药开发关键技术工程研究中心, 北京 102400; 4. 首都医科大学 附属北京中医医院, 北京 100010; 5. 广东药科大学 中药学院, 广东 广州 510006; 6. 中国食品药品检定研究院, 北京 100050; 7. 中药配方颗粒关键技术国家地方联合工程研究中心, 天津 301700; 8. 北京康仁堂药业有限公司, 北京 101301; 9. 江苏康缘药业有限公司, 江苏 连云港 222001; 10. 中成药智能制造国家地方联合工程研究中心, 江苏 连云港 222001; 11. 中药制药过程新技术国家重点实验室, 江苏 连云港 222001)

[摘要] 在中药制造由工业 2.0 向更高水平迈进的过程中, 部分信息化和数字化基础较好的企业已积累了轻量级工业大数据, 成为企业资产的一部分。为促进中药工业大数据的应用, 该研究针对当前中药制造过程中存在的 Sigma 差距和知识匮乏等问题, 提出了以价值创造为导向的中药工业大数据三层架构设计原理, 即数据集成层、数据分析层和应用场景层。在数据集成层, 总结了以传感器为基础的中药关键质量属性感知关键技术。在数据分析层, 提出了由模型构建、验证、配置和维护组成的模型生命周期, 介绍了智慧中药系统(iTCM) 算法库和模型库。针对中药制造过程质量传递结构特点, 开发了“分块-集成”建模、递进建模和路径建模等系统建模关键技术; 针对中药制造过程高度专业性, 提出“数据+机制”双重驱动的高阶智能建模关键技术。最后结合中药注射剂、中药口服固体制剂和中药配方颗粒生产应用场景和需求, 介绍了基于工业大数据的中药生产工艺诊断、质量传递规律解析、实时放行检验和制剂处方智能设计典型案例。对中药工业大数据这一可再生资源的利用, 将有效促进中药制造知识积累和质量效益提升, 为实现中药制造智能化奠定基础。

[关键词] 中药; 智能制造; 工业大数据; 架构设计; 质量传递结构; 系统建模; 传感器

Key technologies and applications of industrial big data in manufacturing of Chinese medicine

XU Bing^{1,2,3}, SHI Xin-yuan^{1,2,3}, LUO Gan^{2,3}, LIN Zhao-zhou⁴, SUN Fei⁵, DAI Sheng-yun⁶,
ZHANG Zhi-qiang^{2,7,8}, XIAO Wei^{9,10,11}, QIAO Yan-jiang^{1,2,3*}

(1. Department of Chinese Medicine Information Science, Beijing University of Chinese Medicine, Beijing 102400, China; 2. Beijing Key Laboratory for Production Process Control and Quality Evaluation of Traditional Chinese Medicine, Beijing Municipal Science & Technology Commission, Beijing 102400, China; 3. Engineering Research Center of Key Technologies for Chinese Medicine Production and New Drug Development, Ministry of Education of People's Republic of China, Beijing 102400, China; 4. Beijing Hospital of Traditional Chinese Medicine, Capital Medical University, Beijing 100010, China; 5. School of Traditional Chinese Medicine, Guangdong Pharmaceutical University, Guangzhou 510006, China;

[收稿日期] 2019-12-19

[基金项目] 国家工信部智能制造综合标准化与新模式应用项目(KYYY20170820); 国家“重大新药创制”科技重大专项(2018ZX09201010); 国家自然科学基金项目(81403112)

[通信作者] * 乔延江, 教授, 博士生导师, 研究方向为中药信息学研究, Tel: (010) 53912117, E-mail: yjqiao@263.net

[作者简介] 徐冰, 副教授, 研究方向为中药质量与先进工艺控制, Tel: (010) 53912118, E-mail: xubing@bucm.edu.cn

6. National Institutes for Food and Drug Control, Beijing 100050, China; 7. National and Regional Joint Engineering Research Center for Key Technologies of Chinese Medicine Formula Granules, Tianjin 301700, China; 8. Beijing Temages Pharmaceutical Co. Ltd., Beijing 101301, China; 9. Jiangsu Kanion Pharmaceutical Co., Ltd., Lianyungang 222001, China; 10. National and Regional Joint Engineering Research Center for Key Technologies of Chinese Patent Medicine, Lianyungang, 222001, China; 11. State Key Laboratory of New-tech for Chinese Medicine Pharmaceutical Process, Lianyungang, 222001, China)

[Abstract] Along with the striding of the Chinese medicine (CM) manufacturing toward the Industry 4.0, some digital factories have accumulated lightweight industrial big data, which become part of the enterprise assets. These digital assets possess the possibility of solving the problems within the CM production system, like the Sigma gap and the poverty of manufacturing knowledge. From the holistic perspective, a three-tiered architecture of CM industrial big data is put forward, and it consists of the data integration layer, the data analysis layer and the application scenarios layer. In data integration layer, sensing of CM critical quality attributes is the key technology for big data collection. In data analysis and mining layer, the self-developed iTCM algorithm library and model library are introduced to facilitate the implementation of the model lifecycle methodologies, including process model development, model validation, model configuration and model maintenance. The CM quality transfer structure is closely related with the connection mode of multiple production units. The system modeling technologies, such as the partition-integration modeling method, the expanding modeling method and path modeling method, are key to mapping the structure of real manufacturing system. It is pointed out that advance modeling approaches that combine the first-principles driven and data driven technologies are promising in the future. At last, real-world applications of CM industrial big data in manufacturing of injections, oral solid dosages, and formula particles are presented. It is shown that the industrial big data can help process diagnosis, quality forming mechanism interpretations, real time release testing method development and intelligent product formulation design. As renewable resources, the CM industrial big data enable the manufacturing knowledge accumulation and product quality improvement, laying the foundation of intelligent manufacturing.

[Key words] Chinese medicine; intelligent manufacturing; industrial big data; architecture design; quality transfer structure; system modeling; sensor

doi: 10.19540/j.cnki.cjmm.20191220.301

中药工业是中医药大健康产业的主体。在《中国制造2025》引领和推动下^[1],智能制造日益成为中药制造业转型升级的重大趋势,是促进中药工业高质量发展并服务“健康中国”战略的重要举措,也是增强企业核心竞争力的必然选择。参照德国“工业4.0”^[2],当前我国中药制造业总体处于工业2.0水平^[3],全产业链发展不均衡,1.0、2.0和3.0并存。工业2.0阶段主要完成生产过程的自动化改造,工业3.0阶段主要完成制造业信息化和数字化,工业4.0阶段目标是使制造系统像人一样思考和工作,具有自主感知、学习、分析、决策、通讯与协调控制、以及协同工作能力,即实现真正意义上的制造智能化。

工业大数据是智能制造的核心驱动力,是制造系统产生智能行为的基础和原材料^[4]。在中药工业由2.0向更高水平迈进的过程中,传感器、智能装备、工业软件和工业IT设施的应用,为车间级和工厂级“大数据”的形成创造了条件。围绕典型中药智能制造模式,从海量数据中挖掘质量传递规律,扩充知识边界和工艺解析力,有利于实现产品质量的追溯与控制,促进中药制造过程精益操作、优化和持续改进。

当前药品生产企业对制造过程产生和积累的数据并没有进行充分利用,存在数据丰富/知识匮乏(data rich/knowledge poor)问题。如Manzano T等^[5]认为在药品制造过

程中,采集到的70%的数据没有进行利用。杨明^[6]认为中药制造业存在生产记录、检验报告等海量数据,但没有从数据中发现规律,数据有效利用率低。本研究针对当前中药制造过程中存在的问题,提出中药工业大数据架构设计原理,分析了中药工业大数据关键技术,介绍了大数据驱动的中药智能生产和产品设计案例,以期对中药工业大数据的应用提供参考。

1 中药制造系统内在矛盾

1.1 Sigma差距

Sigma差距指产品性能与消费者期望之间的差距。在六西格玛管理法中,6 Sigma是一个目标,该水平对应制造出100万件产品,只有3.4件有缺陷。Sigma水平越高,过程波动越小,制造成本和损失越低,满足消费者需求的能力也越强。Migliaccio提出在药品制造过程中,存在6 Sigma的市场期望和3 Sigma的制造过程的问题,即Migliaccio猜想^[7]。美国辉瑞公司Blackburn T等^[8-9]通过数据明确了药品制造系统约处于2.5~3.5 Sigma水平,产生差距的主要原因在于难以追溯导致缺陷的根本原因、对生产过程缺乏理解、以及过分依赖产品质量源于检验(quality by testing QbT)。

中药制造亦存在Sigma差距问题。根据中国食品药品检定研究院发布的国家药品抽检年报^[10-12]2016至2018年,

我国中成药抽检不合格百万分率(defects per million opportunities ,DPMO) 分别为 9 977(61/6 114 不合格批次/总抽检批次) 、16 143(158/9 787) 和 19 586(177/9 037) ,约相当于 3. 83 3. 64 和 3. 56 Sigma 水平; 中药饮片抽检不合格百万分率分别为 182 580(501/2 744) ,102 229(211/2 064) 和 121 716(278/2 284) ,约相当于 2. 41 2. 78 和 2. 67 Sigma 水平。国家药品抽检中针对生产环节的抽样量较少 根据抽检数据计算的 Sigma 水平未考虑生产一次成型率(right first time ,RFT) 等因素的影响,因此实际中药制造的平均 Sigma 水平应低于上述值。

1.2 研发知识与制造知识不匹配

2009 年 ICH Q8(R2) 指南中鼓励在药物开发过程中采用“质量源于设计(quality by design ,QbD) ”理念^[13],充分理解关键物料属性(critical material attributes ,CMAs) 、关键工艺参数(critical process parameters ,CPPs) 和产品关键质量属性(critical quality attributes ,CQAs) 之间的多维组合和交互作用。达成 QbD 的关键在于获取代表性的物料模拟生产过程中物料的变化,研究物料变化和工艺参数变化对过程输出的影响,进而制定相应的工艺控制策略,以确保输出始终符合目标产品质量。

中药制剂研发过程中用到的原料批次和数量是通常是有限的,加之中药原料的天然属性,研发过程无法涵盖所有的原料质量波动。此外,中药开发过程采用的随机化控制试验设计(DOE) 方案中,多重考察工艺参数的变化,较少考虑 CMAs 与 CPPs ,CQAs 之间的交互,见图 1。DOE 除随机误差外,平衡了处理之外的全部有关因素,获取的知识范围有限。由于缺乏对原料质量波动、以及原料波动如何传递至产品的理解,制造过程往往采取保守的工艺控制策略,即将工艺参数控制在有限的范围或固定点,并强调终产品质量检验。在中药制造过程中还存在以下特点,使得研发 DOE 小样本数据难以应对生产质量控制的复杂度: ①设备的规模和控制参数维度发生变化,除 CMAs ,CPPs 和 CQAs 外,尚需考虑和控制更多的条件和参数; ②生产物料的批次数目和质量不确定性增加,中药材和中药饮片质量波动对产品质量一致性的影响可高于生产工艺^[14]; ③新的物料特征与高维过程参数交互可生成新的工艺知识,突破工艺设计空间的范围。解决中药研发知识与制造过程所需知识不匹配的矛盾,迫切需要引入新的技术和方法。

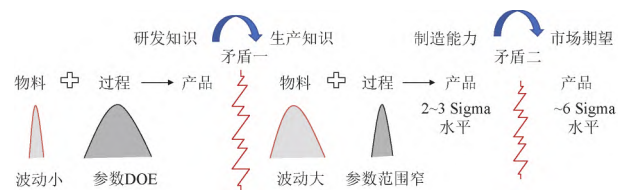


图 1 中药制造系统的内在矛盾

Fig. 1 Chinese medicine production system contradictions

1.3 药品监管科学的进步

质量源于设计的提出者 Juran 博士曾指出,QbD 的内核是质量三部曲(quality trilogy)^[15],即质量计划、质量控制和质量改进。质量三部曲已贯穿在 ICH Q8 ,Q9(质量风险管理)^[16] ,Q10(制药质量体系)^[17] 和 Q12(产品生命周期管理的技术和法规考虑)^[18] 指南体系中。工艺设计空间、知识管理、既定条件等工具,为制造工艺变更、改进和创新提供了灵活性和可操作性。以美国食品药品监督管理局(FDA) 为代表的国外药品监管机构,在以 QbD 为基础的政策环境下,积极推动制药新技术的应用,如 2004 年 9 月颁布《过程分析技术(PAT) 工业指南》^[19]、2017 年 9 月颁布《促进制药创新和现代化的新兴技术应用工业指南》^[20]、2019 年 2 月颁布《连续制造的质量考量工业指南(草案) 》^[21]。美国 FDA 的 Yu L 等^[22] 提出未来药品制造的 6 Sigma 目标,达成这一目标的路径要素包括经济驱动和监管科学、QbD、新兴技术、持续改进和卓越运营。

一些具有国际影响力的行业协会亦积极推动新兴技术与监管政策的协调与融合,如国际制药工程协会(ISPE) 欧洲工作组在全球工业 4.0 革命的背景下,创新提出制药 4.0 (Pharma 4.0)^[23]。认为制药 4.0 = ICH Q10 + 数字化,制药 4.0 包含数字化成熟度和数据完整性源于设计(data integrity by design ,DibD) 2 个使能要素。数字化成熟度促使制药企业向数据驱动型、精益型组织转变。DibD 是监管机构关注重点,要求数据系统清晰透明,数据流和工艺流互相关联。在制药 4.0 环境下,GXP 相关信息系统的集成和多学科知识管理为基于大数据的制药过程智能决策和质量持续改进奠定了基础。

国家药品监督管理局(NMPA) 在 2018 年 6 月成为 ICH 管理委员会成员后,正逐步转化和实施国际最高技术标准和指南。如 2019 年 11 月,NMPA 药品审评中心发布《已上市中药药学变更研究技术指导原则(征求意见稿) 》^[24],提出: 鼓励持有人借鉴 ICH 相关指导原则中的“质量源于设计”“设计空间”“既定条件”等理念和方法,在加强对产品工艺、质量研究的基础上,开展变更管理相关工作……研究工作越系统、深入,生产过程中积累的数据越充分,对上市后的变更研究越有帮助。

2 中药工业大数据系统架构

高效管理制造数据流,收集大数据,有助于应对中药制造系统不确定性挑战,为破解中药研发和制造系统的内在矛盾提供可行方案。开展工业大数据应用,第一步是构建系统架构^[25]。系统架构是一个系统共性关键和基本组成方式的抽象与归纳,是解决不同场景业务诉求的共同基础。中药工业大数据系统架构既要立足中药制造的特点和需求,解决中药行业问题,又要保持与跨产业智能制造、工业互联网和工业大数据相关架构和标准的兼容^[26-28],以利于技术共享、创新和降低实施成本。本研究提出了中药工业大数据三层系

统架构,见图2,包括数据集成层、数据分析层和应用场景层,即以大数据集成为基础(储存价值)、以中药工业大数据挖掘

分析和数据建模为关键(发现价值)、以应用和服务于中药产品生命周期业务需求为导向(实现价值)。

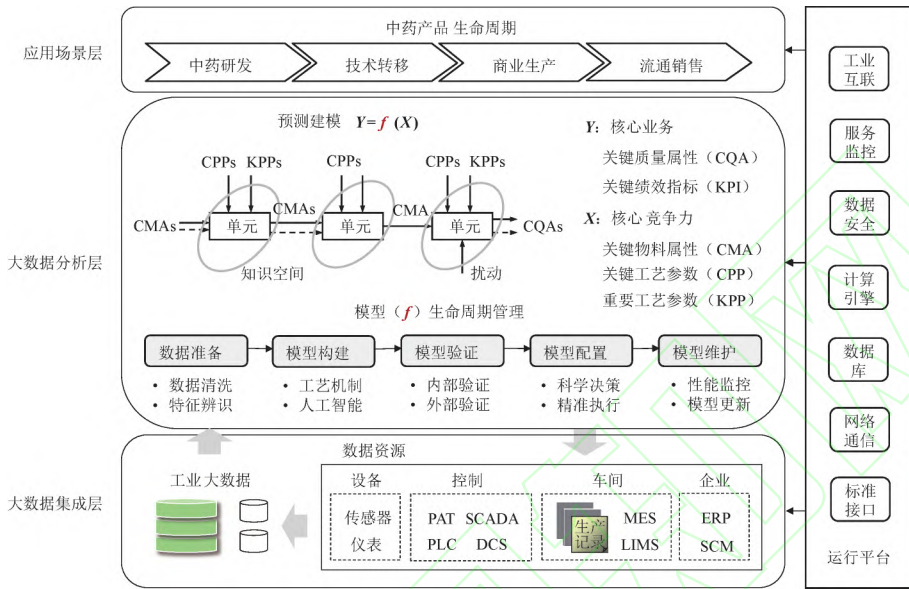


图2 中药工业大数据系统架构

Fig. 2 System architecture of Chinese medicine industrial big data

2.1 数据集成层

数据集成层采集不同来源的数据并集中存储,为中药工业大数据的分析和应用提供准确的基础数据。狭义的中药工业大数据集中于工业设备和工业软件所产生、采集和处理的数据、中间过程控制和质量检验数据等,广义中药工业大数据还包括生产经营相关数据。按照企业纵向层级方向,可能的数据资源包括设备层仪表和传感器采集的数据, PAT、集散控制系统(DCS)、数据采集和监视控制系统(SCADA)等过程控制系统产生的数据,车间层制造执行系统(MES)和高级计划与排程系统(APS)数据,质量部门实验室管理系统(LIMS)数据,以及企业层的企业资源管理系统(ERP)和供应链管理系统(SCM)产生的数据等。数据的多样性、互补性和全面性有助于提高中药工业大数据的质量。

2.2 数据分析层

数据分析层针对特定的业务目标,灵活组织中药工业大数据集,柔性选择组合使用各类大数据分析挖掘算法和技术,从大数据中获得有价值的信息和知识,为应用场景提供决策服务。知识的集中体现是模型,即以函数关系 $Y=f(X)$ 的形式表征过程“输入-输出”量对的集合^[29],是制造过程物理实体及其相互关系在虚拟空间的映像。在预测建模前,把抽象的业务问题转变为数据可分析的核心业务指标(Y),如关键质量属性、关键绩效指标(KPI)等。为获得高质量模型,应结合专业知识,寻找相关性高的特征变量(X),是企业核心竞争力的体现。在工业大数据建模过程中,实施模型生

命周期管理,以促进模型质量持续改进。

2.3 应用场景层

应用场景分布于中药产品设计、开发、生产制造和供应销售的整个生命周期,其决定大数据价值的上限。应针对不同的场景,组织相应的数据,构建各类应用。如在中药生产过程中,设备控制具有实时性和精确性特点,所需数据种类不多,但数据量庞大,且要求数据感知、分析、决策和控制能形成实时闭环的信息流。企业决策层关注制造活动方向的正确性和前瞻性,要求从全局数据中高度提炼有价值信息。除解决已知问题外,工业大数据也可用于未知问题的预测,发现未知的变异和潜在的问题,如设备性能衰退导致的质量漂移等。

2.4 运行平台

为实现上述3个层次的协同运行,企业应部署相应的IT设施,如高速网络、高性能数据库服务器、操作系统、计算引擎、工业APP等。各层次通过标准接口交互,模块之间松耦合,保持可扩展性。工业大数据包括结构化数据、半结构化数据和非结构化数据,数据的类型、性质决定存储结构。目前中药生产过程数据多来自仪表和传感器、DCS、SCADA等,属于结构化数据,可通过数据接入工具部署至关系数据库,如MySQL、Oracle等。随着数据量和数据类型(如系统日志、文本等)的增加,系统面临大规模多源异构数据集,也可选择非结构化查询语言数据库(NoSQL)、非关系型数据库(如HBase)或基于云的基础架构。云计算服务包括基础设施即服务(IaaS)、平台即服务(PaaS)和软件即服务(SaaS),部署

类型包括私有云、公有云和混合云。目前中药企业不同层级、不同业务部门之间的数字化基础设施专业性强,但系统之间往往缺乏统筹,存在重复性建设、可拓展性差的问题。IT 基础设施的布局应与应用需求挂钩,注重系统建设从工具向平台、整体架构从局部向整体的转变。

3 中药工业大数据感知关键技术

仪表和传感器是中药制造过程的眼睛、耳朵和鼻子等“感官”,是中药工业大数据的源头。根据自动化仪表在信息的获取、处理和传递过程中的不同作用,可分为检测仪表、显示仪表、控制仪表和执行器等。检测仪表用来检测生产过程中压力、温度、流量、物位、物性等参数。通过仪表通信接口芯片,将欲发送的数据转换成标准通信格式的数字信号,经发送电路送至通信网络,实现数据存储或与其他系统进行通信。传感器可定义为接收和响应某种信号的装置。感知对象涉及物理信号(如光、电、磁、热、声等)、机械信号(如方位、密度、黏度、速度、力矩等),以及化学信号(如成分浓度)和生物学信号(如生物活性)这些信号与中药关键质量属性密切相关。智能传感器可在传感器内部完成原始数据处理,并可通过标准接口实现数据交换。

随着 PAT 技术和方法在化学药和生物制品生产中的应用普及,各种在线(on-line)、原位(in-line)、非接触式(noninvasive)传感器在中药生产在线质量控制中的应用逐渐增多。在不同中药制造单元和关键质量属性检测中,可选择的传感器见表 1^[30-49]。目前近红外(NIR)传感器在中药生产过程应用最为广泛,一些研究机构和生产企业也在积极引入新型传感器如机器视觉、衰减全反射中红外(ATR IR)、太赫兹(THz)光谱、聚焦反射光束测量(FBRM)和颗粒录影显微镜(PVM)、近红外成像(NIR-CI)等高光谱成像、激光诱导激发光谱(LIBS)、声发射(AE)、以及针对中药性味感知的新型传感器芯片等。

表 1 针对不同关键质量属性的传感器

Table 1 Sensors for analyzing different critical quality attributes

| 单元/对象 | CQAs | PAT 传感器 |
|-------|----------------|--|
| 药材饮片 | 鉴别、等级、Q-marker | 机器视觉 ^[30] 、NIR ^[31] 、THz ^[32] |
| 提取物 | 含量、水分 | NIR ^[33-35] |
| 提取浓缩 | 浓度、动力学 | NIR ^[36-37] 、UV-Vis |
| 醇沉 | 浓度、沉淀粒子 | NIR ^[38-39] 、IR、FBRM+PVM ^[40] |
| 柱色谱 | 浓度、终点 | NIR ^[41] |
| 结晶 | 含量、粒径 | ATR IR、FBRM+PVM |
| 流化床干燥 | 水分 | NIR ^[42] |
| 混合 | 均匀度 | NIR ^[43-44] 、NIR-CI、LIBS |
| 湿法制粒 | 粒径、终点 | NIR、CCD、力学传感器、AE |
| 干法制粒 | 薄片密度、均匀度 | NIR、IR、热成像 |
| 压片 | 质量差异、均匀度 | NIR、NIR-CI ^[45] 、LIBS、THz |
| 包衣 | 衣膜厚度 | NIR ^[46-47] 、CCD、THz |
| 制剂产品 | 浓度、含量、可见异物 | NIR ^[48-49] 、激光散射 |

中药制造过程中物料种类丰富(药材/饮片、提取液、浸膏、粉末、颗粒、不同剂型)、物料状态多样(固体、液体、气体、固液/气液混合)、理化性质差异显著、物质基础复杂,要求传感器具备快速、灵敏、易于清洗、稳定、可靠等能力。在选择传感器时,除针对有效物质基础的响应,还应注意将中药制造过程中的模糊的、不确定性的经验知识“翻译”成能客观化和定量表达的信息。如作者采用针刺式压力传感器模拟“触觉”感官^[50],对人参润制“药透水尽”进行客观化表征,发现穿刺力为 10 N 时,人参药材可达到传统经验判断的完全软化状态,可以作为“药透”的标准,而且人参润制过程中穿刺力随时间变化呈一级动力学,可以根据人参初始状态实现润制终点预测。又如作者为克服药典颗粒剂溶化性目视评价的主观性,设计采用在线浊度传感器,对溶液浑浊程度进行量化表征,基于上市中成药颗粒剂溶化过程浊度分析大数据,建立了可量化的中药颗粒剂溶化性分级评价标准^[51],可为相关产品质量改进提供明确目标。

现代药品质量传感器正在向智能化、集成化、网络化、新材料化方向发展,随着微电子机械系统(MEMS)技术的应用^[52],尺寸越来越小、灵敏度更高、选择性更好,价格也越来越低。采集哪些数据,选择何种传感器,如何在生产过程中对仪表和传感器进行布局,将在一定程度上决定中药工业大数据的质量。大量数据感知工具的使用,将扩展中药产品质量评价的维度,能更精确、全面和透明的还原中药制造过程本来面貌。采用物联网(IoT)技术汇聚多种感知技术采集的信息,并对多源信息进行融合处理^[53-54],有助于提高系统核心业务指标预测的精度和可靠性。

4 中药工业大数据分析关键技术

中药工业大数据中隐含的模式和规则往往无法靠经验或直觉发现,需要借鉴大数据分析的思维和将低质、碎片化的数据转变成高质、高价值密度信息,然后借助数据挖掘、机器学习、人工智能方法,并结合中药专业知识和判断,建立与应用相关联的数学模型,将实体关系透明化,为解决应用需求提供洞察力和决策力。中药工业大数据分析包括数据准备、特征提取、模型构建、验证、配置和维护,各分析模块不断交互和循环,构成模型生命周期。

4.1 数据准备

数据准备包含从处理原始数据到构建最终数据集的过程,其目的是为解决业务问题提供正确的、预定格式的数据。数据准备包括数据选择、清洗、变换等内容,在处理过程中往往需要多种方法的组合尝试,是工业大数据分析中耗时最多($\geq 50\%$)的步骤。

4.1.1 数据选择 根据具体业务目标确定数据分析的范围,选择性地将不同来源、格式、性质的原始数据子集在物理上或逻辑上有机汇集,覆盖尽可能全面的样本和各类变化,提高数据的代表性、多样性和与分析目标的关联性。

4.1.2 数据清洗 原始数据中可能存在缺失/空值、输入或

传输错误、不完整、不同步、异常、编码不一致等情况。这些数据与建模目标关联性差,或不符合模型对数据格式的要求,需要删除、修改或变换问题数据,提高数据质量。如辨识离群点并进行删除处理、采用平滑法处理噪音数据、采用动态时间弯曲(DTW)对数据进行同步^[55]、采用插补法填充空值等。

4.1.3 数据变换 是将数据转化为可以进行统计分析和数据建模的形式。如当前大多数建模技术,均针对表格数据(tabular data)进行操作,若存在异构性数据子集,可通过数据切块、切片或旋转,将原始数据集转换成行列表,如行为批次、列即描述该批次的变量信息。当数据量纲不同时,应当进行数据标准化或尺度化。当数据量大时,可采用抽样或降维方法将大量数据降低至可处理的规模,提高分析效率。

4.2 特征提取

特征提取是中药工业大数据分析中最具创造性和创新性的部分。中药工业大数据多存在具体的物理意义,采集的数据多针对具体需求,具有流程性、时序性、关联性、可解析性特点。特征变量从原始变量中提取,不仅能反映中药制造过程的本质表现,而且有助于建模变量的选择和阐释,提高模型的预测性能,成为企业在工业大数据应用中的核心竞争力。中药制造过程高层次特征的归纳依赖对工业过程的深刻理解,与领域知识密切相关。

除依靠工艺专家进行特征提取外,对于拥有较少先验知识的应用场景,可采用探索性数据分析的方法探究数据的结构和规律、可能的交互等。如采用散点图、直方图、时间序列、饼图、柱状图、热图等对数据进行可视化和对比,了解数据分布模式、变化趋势。通过计算 Pearson 相关系数、灰色关

联度、方差膨胀因子等了解变量相关性,从而抽取有用的变量。通过描述性统计,提取均值、标准差、中位数、分位数、最大值、最小值等特征参数。通过属性构造添加新的变量,填补丢失信息。或通过主成分分析、因子分析等对高维数据进行降维处理,以少数隐变量为特征表达多个变量之间相关性和原始数据的主要变化信息。一些新型人工智能(AI)算法,如深度学习,可自动提取简单而抽象的特征,并组合成更加复杂的特征^[56]。在探索性分析过程中,可尝试不同方法转换数据,特征提取的质量可以用数据建模进行评价。

4.3 数据建模

建立模型是中药工业大数据分析的核心。与互联网大数据和商业大数据重视相关性不同,中药工业大数据重视因果关系。企业在中药工厂设计时,对变量逻辑关系有一定了解,但一些常规经验无法定义的重要关系在工厂设计时不完全所知或不确定。数据建模可看做归纳过程,即根据数据特征学习数据集的固有结构、模式和关系,将工业大数据由感知空间提升到认知空间。

4.3.1 iTCM 算法库和模型库 作者建立了 iTCMTM 算法库,即一套标准数据处理、建模和优化方法和程序,在应用时将准备好的数据插入每个模型,可自动匹配最佳模型,或根据数据性质和背景知识选择合适的建模算法,见图3。iTCM 算法库中的基本建模方法分为描述性建模法和预测性建模法。描述性建模没有响应变量引导,即无监督学习。主要用于探索数据的基础结构或模式、提取有意义的特征,常用方法有聚类、潜结构投影等。在中药制造过程中,将新批次数据与历史数据比较,判断数据结构性差异,可辅助工艺诊断。

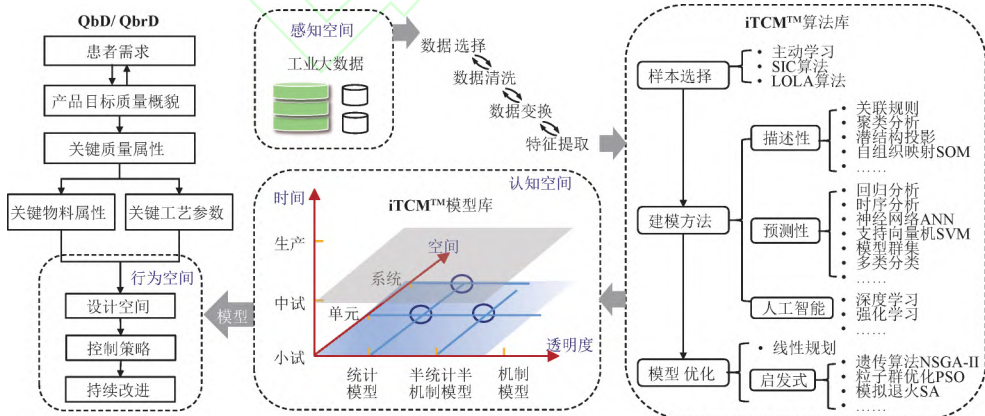


图3 iTCM 算法库和模型库

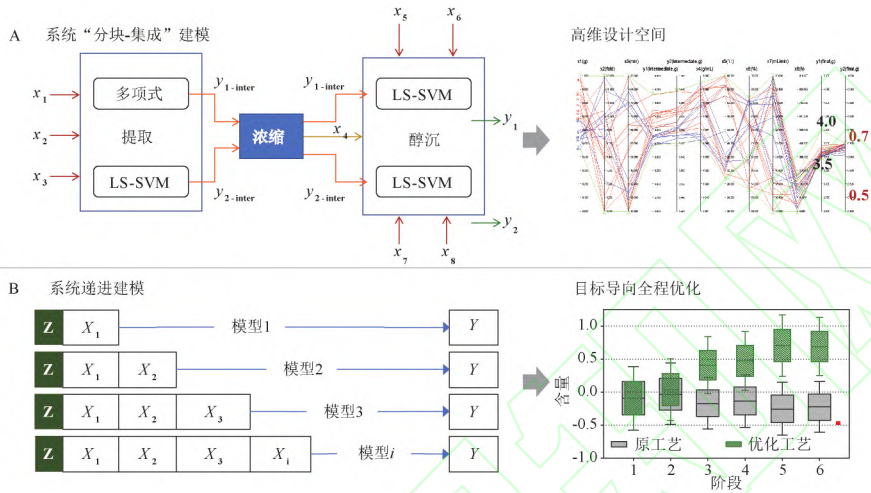
Fig. 3 The iTCM algorithm base and model base

预测性建模有响应变量引导,即监督学习。主要用于寻找响应变量和自变量之间的函数关系,对新的输入做出预测。常用方法有回归分析,时序分析,神经网络(ANN)^[57],支持向量机(SVM)^[58],Bagging、Boosting 和随机森林等模型

群集算法,多类分类^[59]等。根据响应变量是否为连续型变量,可分为定量预测建模和定性预测建模。选定模型后,应用优化算法(线性规划或进化算法)调整模型超参数到最佳值,避免过拟合或欠拟合。

作者针对中药工业大数据的特点,建立了 iTCMTM 模型库,以归纳不同场景数据建模实践的共性经验,是业务需求和算法的整合,见图 4。模型库中的模型类型以 3 个维度分类,即空间、时间和透明度。从空间维度,可分为工艺单元模型和工艺系统

模型;从时间维度,可分为小试模型、中试模型和生产模型,以及规模传递模型^[60];从透明度维度,可分为统计模型、半统计半机制模型和机制模型。针对 QbD 和质量源于重设计(quality by redesign, Qbrd) 中不同应用场景^[61-62] 方便快速匹配模型。



A. “分块-集成”建模; B. 递进建模。

图 4 中药工艺系统建模

Fig. 4 System modeling of Chinese medicine multi-unit production process

4.3.2 系统建模 中药制药过程具有原料复杂多样、流程长、优化控制对象多等特点。组成中药制药过程系统的各单元之间在结构上密不可分,在功能上相互协调影响,将中药质量从原料经由一系列单元操作逐步传递到最终产品。从质量传递结构上看,中药生产工艺系统可分为串联、并联和串并混联 3 种。串联型中药制药过程是中药质量传递的基本载体,串并混联更具一般性。如何从系统角度应对原料和过程波动,是目前中药制造过程面临的重要挑战。

在中药制药过程系统建模方面,作者开发了“分块-集成”建模法^[63],递进建模法^[64]和路径建模(path modeling)法^[65-67]等。“分块-集成”建模法是将中药制造系统按功能划分为若干子系统,分别建立各子系统模型,然后建将子系统模型关联集成,构成全局模型;中药提取、浓缩和醇沉三工序“分块-集成”系统模型,以及基于该模型开发的工艺参数高维设计空间,见图 4A。递进建模法针对串联型结构工艺系统,模拟中药制造系统“质量累积”效应,即将中药生产过程视为多阶段间歇过程,模拟生产进行到不同阶段时对终产品质量可能产生的影响,见图 4B。基于递进模型计算制药过程多阶段概率轨迹(probability trajectory, PT),以可视化形式量化生产进行到每一工序时符合终产品质量目标的概率,是过程全局不确定性的概括。路径模型针对串并混联型结构中中药工艺系统,对原始高维显变量进行降维,通过测度模型表征数据降维过程中,中药制药工艺高维显变量空间和低维潜变量空间之间的多维组合和相互作用,通过结构模型表征潜

变量因子和产品关键质量属性之间的关系,明确各工艺单元对终产品质量的贡献,以及各工艺单元之间的耦合协同机制。

4.3.3 “数据+机制”双重驱动的知识发现 “数据和机制双重驱动”的知识发现,有助于建立高可靠性和高精度的模型。中药工业大数据的低质性、隐性与其工业应用中的低容错性之间存在矛盾^[68]。如数据中存在噪音或数据多样性不足,可导致预测精度不高。为提高数据模型质量,可借助机制分析理清数据模块之间的逻辑关联,辅助特征变量辨识。反之,通过工业大数据也可增进过程机制的理解,促进中药制药过程从黑箱模型到白箱模型的转变。在与药品相近的化工领域, Venkatasubramanian V^[69] 分析指出以深度学习和数据驱动为特征的 AI 发挥更大作用的条件已经出现,但需发展一种理论框架,即不仅能定性定量预测宏观特征,而且能阐释背后的机制。在中药口服固体剂剂共性关键环节的研究中,作者发展了部分半统计半机制模型,如采用物性分析结合非接触式在线光谱传感器,理解影响混合均匀度的粒子特征^[70];在高速剪切湿法制粒(HSWG)工艺中,引入无量纲分析模型,即模式地图(regime map)^[71-72],指导 HSWG 过程放大和参数优化;在中药干法制粒过程和压片过程,分别建立了物料辊压行为分类系统(RCBCS)^[73]和物料压缩行为分类系统(CBCS)^[74],指导物料物性改进和配方设计与优化。

4.4 模型验证、部署和维护

4.4.1 模型验证 在模型应用前,应对模型的预测精度和

可靠性进行充分验证。验证方法包括内部验证和外部验证。内部验证分简单验证和交叉验证。简单验证即随机从原始数据集中选择 1 个子集作为验证集,其余数据作为校正集,以校正集建立模型,以验证集评价校正集的预测性能。交叉验证将原始数据集分成 K 组,每组分别做一次验证集,其余 $K-1$ 组作为校正集,以 K 个模型的预测性能平均水平评价模型质量。外部验证即选择新的数据集验证模型性能。定量模型预测精度评价指标通常选择均方根误差 (root mean square error, RMSE)、决定系数 (R^2)、交叉验证有效性 (Q^2) 等^[75]。分类模型性能常采用分类准确率、错误率、灵敏度、特异度等。

4.4.2 模型部署 通过验证后的模型,应部署至相应的业务系统,匹配相应的工艺控制策略^[76],并以界面友好、易于理解和可视化、可重复使用的形式呈现。如将中药工业大数据分析流程和算法封装成程序和工具,配置于工业软件系

统,辅助制造过程科学决策,或向车间层、边缘层发布指令,或通过执行器精准控制物理系统,实现控制操作闭环。

4.4.3 模型维护 模型部署以后,应建立模型性能持续监控机制、定期维护或更新机制^[77-78]。中药原料质量波动大,高质量的过程模型需要更多历史数据。如模型预测精度下降时,可设法调试模型参数,或添加新的数据,进行重新训练,确保模型覆盖各种类型数据。随着生产智能化水平的提高,也可建立主动学习、遗忘等自适应机制^[70,79],实现模型可持续优化。

5 中药工业大数据应用

中药工业大数据本身并不重要,利用大数据创造价值才是根本目的。本节介绍常规工厂和数字化工厂中数据挖掘应用案例,见图 5。涉及数据均源自大规模批次生产。在中药制造业和工厂发展的不同时期,数据存储能力不同,衡量数据规模的尺度是相对的。

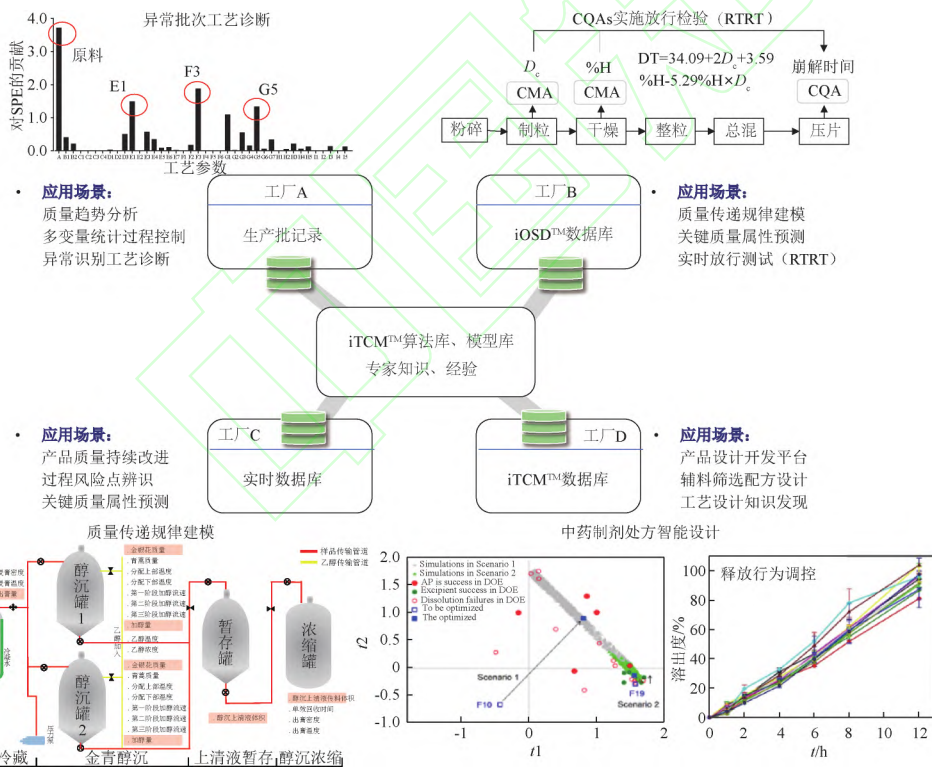


图 5 中药工业大数据应用案例

Fig. 5 Applications of Chinese medicine industrial big data

案例 1 至案例 3 分别介绍了工业大数据在 3 个中药生产场景中的应用。案例 4 介绍中药工业大数据可以反馈到设计端,指导新产品开发,或对现有产品进行质量重设计、质量优化和工艺改进,打通中药产品生命周期的数据闭环。

5.1 案例 1: 工艺诊断

数据描述: 某工厂 A (常规工厂) 3 年间 115 批清开灵注射液中金银花前处理生产历史批记录数据,包括 9 个工艺单

元,40 个工艺变量,总计 4 600 个数据点^[80-81]。分析前将数据进行均值标准化处理,以消除量纲差异。

工艺诊断: 对数据进行主成分分析,计算各主成分贡献率,选择累积和百分比大于 85% 的主成分个数。根据多变量统计过程控制 (MSPC) 中的 Hotelling T^2 、SPE 统计量是否超出其控制限,判断金银花前处理生产工艺是否出现了异常状况。当异常批次出现时,采用贡献图确定异常情况发生的原

因。如图 5 工厂 A 应用场景中的 SPE 贡献图,该异常批次主要与变量 A(代表金银花药材中绿原酸的含量)有关,并且与工艺参数 F3、E1 和 G5 有一定关联。

5.2 案例 2: 实时放行检验

数据描述:(口服固体制剂)工厂 B 围绕天舒片、桂枝茯苓胶囊等口服固体制剂共性关键环节,建立了粉体中间体物理性质综合评价平台,物性参数包括密度类参数:松密度和振实密度(D_c);流动性参数:休止角(α)、粉体流动时间(t'')和豪斯纳比(IH);粒径参数: D_{10} 、 D_{50} 、 D_{90} ,分布宽度(Span)和范围(Width)、粒径 $<50\ \mu\text{m}$ 所占百分比(%Pf)、相对均匀度指数(I0);可压性参数:颗粒间孔隙率(I_c)、卡尔指数(IC)和内聚力指数(Icd);稳定性参数:水分(%HR)和吸湿性(%H);质地参数:物料粘附性、内聚性、弹性和弹性系数。对目标产品混合、制粒、流化床干燥、整粒、压片和包衣等工序的物料进行取样分析后,将数据存储在自主设计开发的 iOSD™ 数据库。以天舒片崩解时间(DT)的实时放行检验(RTRT)方法建立为例,测定了 49 批次制剂生产过程原料和中间体的 72 个物性参数为输入,以素片崩解时间(DT)为输出,总计 3 577 数据点^[82-83]。

RTRT 方法:参照欧盟 GMP 附录 17 标准^[84]和日本药品与医疗器械管理局(PMDA)片剂 QbD 开发模板^[85],通过系统建模、特征筛选,最终确定影响天舒片崩解时间的关键物料属性为湿法制粒所得湿颗粒的振实密度 D_c 和沸腾干燥所得干燥颗粒的吸湿性%H,以此 2 个关键物料属性为自变量建立的天舒片 DT 实时放行测试模型为: $DT = 34.09 + 2 \times D_c + 3.59 \times \%H - 5.29 \times \%H \times D_c$,该模型的决定系数 $R^2 = 0.9017$,外部验证集平均相对误差为 3.69%。该 RTRT 模型可代替天舒片 DT 常规检验方法,并可用于指导湿法制粒和沸腾干燥的预测性调控,提高一次成品率。

5.3 案例 3: 质量传递规律建模

数据描述:工厂 C(数字化工厂)为热毒宁注射液生产线^[86],装备实现了自动化,采用 DCS 控制原料提取、浓缩、醇沉、萃取、干燥等过程,实现温度、流量、pH、真空度、密度等工艺参数数据的在线抓取和存储,数据生成从传统的人工生成的单通道模式转变为人工生成和自动生成的双通道模式。以醇沉工序为例,包括浓缩液分配、3 次加醇、静置、过滤、收醇等步骤,为解析醇沉质量传递规律,收集该工段参数,包括 2 个部分:第一部分为生产批记录参数,如浸膏密度、浸膏温度、乙醇温度、乙醇浓度、单效回收时间、出膏密度和出膏温度;第二部分由设备温度和流量仪表生成或计算,包括出膏量、金银花质量、青蒿质量、分配温度、加醇流速、加醇量、醇沉上清液体积、醇沉上清液传料体积。该案例收集了 2017—2018 年间热毒宁金青醇沉过程的 259 批生产数据,每批数据点 3 202 个,总计 829 318 个数据点^[87]。

质量传递规律建模:综合采用关键质量属性定义、数据地图、数据选择、数据清洗、特征提取、探索性建模分析等方

法,从生产大数据中提取 48 个物理意义明确的特征参数,并进一步辨识得到 15 个潜在工艺参数和 9 个关键工艺参数。发现当前采取醇沉前药液密度控制和加醇后药液含醇量控制的单变量模式,尚无法应对饮片原料波动的影响,如每批提取浓缩后所得浸膏的固形物量不同,导致相同密度的药液具有不同体积,因此加醇量随初始浸膏体积的变化而变化,这一波动直接影响醇沉上清液体积,进而影响上清液浓缩时间。

5.4 案例 4: 中药制剂配方智能设计

数据描述:作者联合(中药配方颗粒)工厂 D,以生产物料为载体,建立了中药提取物粉末物理性质综合评价平台和 iTCM™ 智慧中药系统^[88],该系统通过物性大数据、物理指纹图谱^[89]和物性周期表^[90]对不同来源中药粉体的质量概貌、波动规律和“药辅合一”科学内涵^[91]进行解析,结合工艺智能建模、仿真和优化技术,辅助深刻理解中药口服固体制剂成型过程中关键物料属性、关键工艺参数和产品关键质量属性之间的关系,预测工艺中试放大和生产过程中可能出现的问题,提高辅料筛选和工艺控制可靠性,节省研发投入,降低工艺批准后变更带来的成本和风险。

以中药片剂为例,建立了基于原辅料物理性质的处方智能设计方法^[92],包括:①从 iTCM 数据库中选择输入物料数据,建立输入物料的物性对最终处方性质影响的数学表达式,即定量预测模型;②建立期望获得的最终处方性质的定量表达式,即目标方程,包括约束条件的确立;③从可能性空间中搜索得到满足目标方程和限制条件的可行结果。如以三七总皂苷骨架缓释片为例^[93-94],在不改变处方中原辅料比例的约束下,通过选择合适物性的 HPMC 辅料,可实现 PNS 凝胶骨架片中人参皂苷成分溶出曲线的调控,降低原辅料质量波动对片剂质量一致性的影响,提高处方稳健性。

6 总结和展望

中药工业大数据是中药产品制造过程中信息系统和物理系统交互的桥梁。本研究提出的中药工业大数据三层架构致力于在数字化世界打通感知空间、认知空间、行为空间之间的信息回路。在中药产品制造过程中,工艺系统的不确定性导致输出产品质量的分散性,中药工业大数据中蕴藏的丰富信息可产生解决过程质量控制问题的新的可能性。基于工业大数据的系统建模是研究中药制造系统结构与功能、发现过程质量传递一般规律、探究如何消除或减小不确定性对输出性能分散性的影响,以及从全局角度实现中药生产过程质量可靠性设计和控制的关键手段。

当前,部分数字化技术设施较好的中药企业数据量已达到 TB 级/年的规模,工业大数据和智能装备、工业软件一样,逐渐成为企业基础设施和资产的一部分。这些轻量级中药工业大数据具备了传统大数据的部分特点,如大容量(volume)、多样性(variety)、价值性(value)和真实性(veracity)等。随着 5G、AI、工业互联网、物联网等“新型基础设施建设”的应用,

高速性(velocity)、闭环性(closed-loop)、安全性(safety)是中药工业大数据今后发展的方向。中药工业大数据属于可再生资源,通过多学科交叉协作,开展数据重组、扩展、再利用,不断突破原有分析框架,将有效促进中药制造知识积累,为早日实现中药制造智能化奠定基础。

[参考文献]

- [1] 国务院. 国务院关于印发《中国制造2025》的通知[EB/OL]. [2015-05-19]. http://www.gov.cn/zhengce/content/2015-05/19/content_9784.htm.
- [2] Germany Federal Ministry of Education and Research, Industrie 4.0 Working Group. Recommendations for implementing the strategic initiative Industrie 4.0 [EB/OL]. (2013-04-08) [2019-12-19]. https://en.acatech.de/wp-content/uploads/sites/6/2018/03/Final_report__Industrie_4.0_accessible.pdf.
- [3] 操秀英. 张伯礼代表: 中药生产线也要瞄准工业4.0[N]. 科技日报, 2018-3-11(9).
- [4] 张洁, 秦威, 鲍劲松, 等. 制造业大数据[M]. 上海: 上海科学技术出版社, 2016.
- [5] Manzano T, Langer G. Getting ready for Pharma 4.0-data integrity in cloud and big data applications[J]. Pharm Eng, 2018, 38(5): 72.
- [6] 杨明. 中药制剂质量过程控制研究[J]. 中国食品药品监管, 2019(7): 87.
- [7] Suresh P, Basu P. Improving pharmaceutical product development and manufacturing: impact on cost of drug development and cost of goods sold of pharmaceuticals[J]. J Pharm Innov, 2008, 3(3): 175.
- [8] Blackburn T, Mazzuchi T, Sarkani S. Overcoming inherent limits to pharmaceutical manufacturing quality performance with QbD (quality by design) [J]. J Pharm Innov, 2011, 6(2): 69.
- [9] Blackburn T. Quality by design (QbD) and production system fundamental sigma limits with a pharmaceutical industry example [D]. Washington: George Washington University, 2012.
- [10] 中国食品药品检定研究院. 2016年国家药品抽检年报[EB/OL]. (2017-10-09). [2019-12-19]. <https://www.nifdc.org.cn/nifdc/fwzn/ypchysxyj/9872.html>.
- [11] 中国食品药品检定研究院. 2017年国家药品抽检年报[EB/OL]. (2018-06-05). [2019-12-19]. <https://www.nifdc.org.cn/nifdc/fwzn/ypchysxyj/11119.html>.
- [12] 中国食品药品检定研究院. 国家药品抽检年报(2018)[EB/OL]. (2019-08-07). [2019-12-19]. <https://www.nifdc.org.cn/nifdc/fwzn/ypchysxyj/20190807101035.html>.
- [13] International Conference on Harmonization of Technical Requirements for Registration of Pharmaceuticals for Human Use (ICH). Pharmaceutical development. Q8(R2) [EB/OL]. [2019-12-19]. http://www.ich.org/fileadmin/Public_Web_Site/ICH_Products/Guidelines/Quality/Q8_R1/Step4/Q8_R2_Guideline.pdf.
- [14] 钟文, 陈莎, 章军, 等. 重点是原料, 还是工艺? ——以葛根芩连汤为例探讨中成药质量一致性控制方法[J]. 中国中药杂志, 2016, 41(6): 1027.
- [15] Juran J M. Juran on quality by design: the new steps for planning quality into goods and services[M]. New York: The Free Press, 1992.
- [16] International Conference on Harmonization of Technical Requirements for Registration of Pharmaceuticals for Human Use (ICH). Quality risk management. Q9 [EB/OL]. [2019-12-19]. http://www.ich.org/fileadmin/Public_Web_Site/ICH_Products/Guidelines/Quality/Q9/Step4/Q9_Guideline.pdf.
- [17] International Conference on Harmonization of Technical Requirements for Registration of Pharmaceuticals for Human Use (ICH). Pharmaceutical quality system. Q10 [EB/OL]. [2019-12-19]. http://www.ich.org/fileadmin/Public_Web_Site/ICH_Products/Guidelines/Quality/Q10/Step4/Q10_Guideline.pdf.
- [18] International Conference on Harmonization of Technical Requirements for Registration of Pharmaceuticals for Human Use (ICH). Technical and regulatory considerations for pharmaceutical product lifecycle management. Q12 [EB/OL]. [2019-12-19]. http://www.ich.org/fileadmin/Public_Web_Site/ICH_Products/Guidelines/Quality/Q12/Q12_Final_Concept_Paper_July_2014.pdf.
- [19] U. S. Food and Drug Administration. Guidance for industry PAT—a framework for innovative pharmaceutical development, manufacturing, and quality assurance [EB/OL]. [2019-12-19]. <https://www.fda.gov/media/71012/download>.
- [20] U. S. Food and Drug Administration. Advancement of emerging technology applications for pharmaceutical innovation and modernization—guidance for industry [EB/OL]. [2019-12-19]. <https://www.fda.gov/media/95444/download>.
- [21] U. S. Food and Drug Administration. Quality considerations for continuous manufacturing—guidance for industry (draft guidance) [EB/OL]. [2019-12-19]. <https://www.fda.gov/media/121314/download>.
- [22] Yu L, Kopcha M. The future of pharmaceutical quality and the path to get there[J]. Int J Pharm, 2017, 528(1/2): 354.
- [23] Binggeli L, Heesakkers H, Wölbelling C, et al. Special report: 2018 Europe Annual Conference. "Pharma 4.0: hype or reality?" [J]. Pharm Eng, 38(4): 40.
- [24] 国家药品监督管理局药品审评中心. 关于公开征求《已上市中药药性变更研究技术指导原则》意见的通知[EB/OL]. (2019-11-08) [2019-12-19]. <http://www.cde.org.cn/news.do?method=viewInfoCommon&id=314972>.
- [25] 郑树全, 宗宇伟, 董文生, 等. 工业大数据: 架构与应用[M]. 上海: 上海科学技术出版社, 2017.
- [26] 工业和信息化部, 国家标准化管理委员会. 国家智能制造标准体系建设指南(2018年版)[EB/OL]. (2018-08-14) [2019-12-19]. http://www.gov.cn/xinwen/2018-10/16/content_5331149.htm.
- [27] 工业和信息化部, 国家标准化管理委员会. 工业互联网综合标准化体系建设指南[EB/OL]. (2019-01-25) [2019-12-19].

- http://www.gov.cn/xinwen/2019-03/08/content_5371933.htm.
- [28] 中国电子技术标准化研究院, 全国信息技术标准化技术委员会大数据标准工作组. 工业大数据白皮书(2019版) [R/OL]. (2019-04-01) [2019-12-19]. <http://www.cesi.ac.cn/201904/4955.html>.
- [29] 石国琳, 徐冰, 林兆洲, 等. 中药质量源于设计方法和应用: 工艺建模[J]. 世界中医药, 2018, 13(3): 543.
- [30] 徐曼菲, 吴志生, 刘晓娜, 等. 从辨色论质谈中药质量评价方法[J]. 中国中药杂志, 2016, 41(2): 177.
- [31] 白钢, 侯媛媛, 丁国钰, 等. 基于中药质量标志物构建中药材品质的近红外智能评价体系[J]. 药学报, 2019, 54(2): 197.
- [32] 田珺宏, 周俊, 郑潇潇, 等. 基于太赫兹时域光谱仪的中药材鉴定技术研究[J]. 中国医疗设备, 2018, 33(7): 9.
- [33] Xue Z, Xu B, Shi X, et al. Overall uncertainty measurement for near infrared analysis of cryptotanshinone in tanshinone extract [J]. *Spectrochim Acta A*, 2017, 170: 39.
- [34] Luo G, Xu B, Shi X, et al. Rapid characterization of tanshinone extract powder by near infrared spectroscopy [J]. *Int J Anal Chem*, 2015, doi: 10.1155/2015/704940.
- [35] 刘倩, 徐冰, 罗赣, 等. 丹参提取物中辅料糊精的近红外快速定量分析[J]. 世界中医药, 2013, 8(11): 1287.
- [36] 隋丞琳, 林兆洲, 吴志生, 等. 药渣粒径对中药提取过程在线检测取样系统影响的模拟研究[J]. 分析化学, 2013, 41(12): 1899.
- [37] 刘爽悦, 李文龙, 瞿海斌, 等. 基于近红外光谱的丹红注射液提取过程质量在线检测方法研究[J]. 中国中药杂志, 2013, 38(11): 1657.
- [38] 徐冰, 罗赣, 林兆洲, 等. 基于过程分析技术和设计空间的金银花醇沉加醇过程终点检测[J]. 高等学校化学学报, 2013, 34(10): 2284.
- [39] 贾帅芸, 徐冰, 杨婵, 等. 基于SIC算法的丹参醇提过程近红外定量模型更新研究[J]. 中国中药杂志, 2016, 41(5): 823.
- [40] 袁佳. 丹参和红花提取液醇沉工艺及包裹损失现象的研究[D]. 杭州: 浙江大学, 2011.
- [41] 李建宇, 徐冰, 张毅, 等. 近红外光谱用于大孔树脂纯化栀子提取物放大过程的监测研究[J]. 中国中药杂志, 2016, 41(3): 421.
- [42] 王晴, 徐芳芳, 张欣, 等. 在线近红外光谱监测桂枝茯苓胶囊流化床干燥过程水分的方法研究[J]. 中草药, 2019, 50(22): 5429.
- [43] Xue Z, Xu B, Yang C, et al. Method validation for the analysis of licorice acid in the blending process by near infrared diffuse reflectance spectroscopy [J]. *Anal Methods-UK*, 2015, 7: 5830.
- [44] 杨婵, 徐冰, 张志强, 等. 基于移动窗F检验法的中药配方颗粒混合均匀度近红外分析研究[J]. 中国中药杂志, 2016, 41(19): 57.
- [45] 吴志生, 陶欧, 程伟, 等. 基于光谱成像技术的乳块消素片活性成分空间分布及均匀性研究[J]. 分析化学, 2011, 39(5): 628.
- [46] 夏春燕, 徐芳芳, 张欣, 等. 近红外光谱分析技术快速测定天舒片的包衣薄膜厚度[J]. 中国实验方剂学杂志, 2019, doi: 10.13422/j.cnki.syfx.20192251.
- [47] 夏春燕, 徐芳芳, 张欣, 等. 近红外光谱快速预测天舒片包衣终点研究[J]. 中草药, 2019, 50(21): 5223.
- [48] 徐冰, 薛忠, 罗赣, 等. 基于 β -容量容忍区间的血管通注射液近红外定量分析不确定度评估[J]. 中草药, 2015, 46(6): 832.
- [49] 孙飞, 徐冰, 戴胜云, 等. 近红外分析用于中药产品质量实时放行测试的可靠性研究[J]. 中华中医药杂志, 2017, 32(12): 5316.
- [50] 连传运, 徐冰, 赵亚楠, 等. 基于针刺压力传感器的人参润制“药透水尽”科学内涵研究[J]. 中国中药杂志, 2020, 45(1): 267.
- [51] 于佳琦, 徐冰, 黄雨妍, 等. 基于在线浊度传感器的中药颗粒剂溶化性评价和分类研究[J]. 中国中药杂志, 2020, 45(1): 259.
- [52] Yan H, Siesler H W. Quantitative analysis of a pharmaceutical formulation: performance comparison of different handheld near-infrared spectrometers [J]. *J Pharmaceut Biomed*, 2018, 160: 179.
- [53] 张娜, 徐冰, 贾帅芸, 等. 丹参提取过程多源信息融合建模方法研究[J]. 中草药, 2018, 49(6): 1304.
- [54] Dai S, Lin Z, Xu B, et al. Metabolomics data fusion between near infrared spectroscopy and high-resolution mass spectrometry: a synergetic approach to boost performance or induce confusion [J]. *Talanta*, 2018, 189: 641.
- [55] Salvador S, Chan P. Toward accurate dynamic time warping in linear time and space [J]. *Intel data anal*, 2007, 11(5): 561.
- [56] He K, Zhang X, Ren S, et al. Spatial pyramid pooling in deep convolutional networks for visual recognition [J]. *IEEE T Pattern Anal*, 2015, 37(9): 1904.
- [57] E. P. 9.0. 5.21. Chemometric methods applied to analytical data[S]. 2017.
- [58] 徐冰, 王星, Dhaene T, 等. 基于遗传算法的多目标最小二乘支持向量机在近红外多组分定量分析中的应用[J]. 光谱学与光谱分析, 2014, 34(3): 638.
- [59] Lin Z, Jia S, Luo G, et al. Dealing with heterogeneous classification problem in the framework of multi-instance learning [J]. *Talanta*, 2015, 132: 175.
- [60] Lin Z, Xu B, Li Y, et al. Application of orthogonal space regression to calibration transfer without standards [J]. *J Chemometr*, 2013, 27(11): 406.
- [61] 徐冰, 史新元, 吴志生, 等. 论中药质量源于设计[J]. 中国中药杂志, 2017, 42(6): 1015.
- [62] Nikolic S. Quality by redesign of a legacy product [J]. *Pharm Eng*, 2019, 39(2): 52.
- [63] 徐冰. 中药制剂生产过程全程优化方法学研究[D]. 北京: 北

- 京中医药大学, 2013.
- [64] Xu B, Lin Z, Wu Z, et al. Target-oriented overall process optimization (TOPO) for reducing variability in the quality of herbal medicine products[J]. *Chemometr Intell Lab*, 2013, 128: 144.
- [65] Sun F, Xu B, Dai S, et al. A novel framework to aid the development of design space across multi-unit operation pharmaceutical processes—A case study of *Panax notoginseng* saponins immediate release tablet [J]. *Pharmaceutics*, 2019, 11(9): 474.
- [66] Sun F, Xu B, Zhang Y, et al. Latent variable modeling to analyze the effects of process parameters on the dissolution of paracetamol tablet[J]. *Bioengineered* 2017 8(1): 61.
- [67] Sun F, Xu B, Zhang Y, et al. Statistical modeling methods to analyze the impacts of multiunit process variability on critical quality attributes of Chinese herbal medicine tablets [J]. *Drug Des Dev Ther*, 2016, 10: 3909.
- [68] 李杰. 工业大数据: 工业 4.0 时代的工业转型与价值创造 [M]. 北京: 机械工业出版社, 2015.
- [69] Venkatasubramanian V. The promise of artificial intelligence in chemical engineering: is it here, finally? [J]. *AICHE J*, 2019, 65(2): 466.
- [70] Shi G, Xu B, Zhang Z, et al. Raw material variability and its impact on the online adaptive control of cohesive powder blend homogeneity using NIR spectroscopy [J]. *Processes*, 2019, 7(9): 568.
- [71] Luo G, Xu B, Zhang Y, et al. Scale-up of a high shear wet granulation process using a nucleation regime map approach [J]. *Particuology*, 2017, 31: 87.
- [72] Zhang Y, Cheng B, Zhou W, et al. Improved understanding of the high shear wet granulation process under the paradigm of quality by design using *Salvia miltiorrhiza* granules [J]. *Pharmaceutics*, 2019, 11(10): 519.
- [73] Yu J, Xu B, Zhang K, et al. Using a material library to understand the impacts of raw material properties on ribbon quality in roll compaction [J]. *Pharmaceutics*, 2019, 11(12): 662.
- [74] Dai S, Xu B, Zhang Z, et al. A compression behavior classification system of pharmaceutical powders for accelerating direct compression tablet formulation design [J]. *Int J Pharm*, 2019, 572: 118742.
- [75] Gorissen D, Couckuyt I, Laermans E, et al. Multiobjective global surrogate modeling, dealing with the 5-percent problem [J]. *Eng Comput*, 2010, 26(1): 81.
- [76] 连传运, 徐冰, 王秋平, 等. 中药质量源于设计应用: 工艺控制策略 [J]. *世界中医药*, 2018, 13(3): 561.
- [77] Xu B, Wu Z, Lin Z, et al. NIR analysis for batch process of ethanol precipitation coupled with a new calibration model updating strategy [J]. *Anal Chim Acta*, 2012, 720: 22.
- [78] Li J, Xu B, Zhang Y, et al. Determination of geniposide in gardenia jasminoides ellis. fruit by near infrared spectroscopy and chemometrics [J]. *Anal Lett*, 2016, 49(13): 2063.
- [79] Kucherenko S, Giamalakis D, Shah N, et al. Computationally efficient identification of probabilistic design spaces through application of metamodeling and adaptive sampling [J]. *Comput Chem Eng*, 2020, 132: 106608.
- [80] 周海燕, 徐冰, 史新元, 等. 清开灵注射液中金银花液生产工艺多元统计过程监控研究 [J]. *世界科学技术——中医药现代化*, 2012, 14(4): 1877.
- [81] 周海燕, 徐冰, 史新元, 等. 统计过程控制在栀子前处理生产工艺中的应用 [J]. *中国实验方剂学杂志*, 2012, 18(11): 16.
- [82] 夏春燕, 徐冰, 徐芳芳, 等. 天舒片素片崩解时间实时放行检验研究 [J]. *中国中药杂志*, 2020, 45(1): 250.
- [83] 王晴, 徐冰, 王芬, 等. 桂枝茯苓胶囊内容物吸湿性预测建模研究 [J]. *中国中药杂志*, 2020, 45(1): 242.
- [84] EU GMP. Annex 17: real time release testing and parametric release [S]. 2018.
- [85] MHLW sponsored QbD Drug Product Study Group. Sakura Bloom Tablets P2 Mock [EB/OL]. [2019-12-19]. http://www.nihs.go.jp/drug/section3/QbD_P2mock_SakuraBloom141114_E.pdf.
- [86] 辛国斌. 智能制造探索与实践: 46 项试点示范项目汇编 [M]. 北京: 北京电子工业出版社, 2016.
- [87] 杜慧, 徐冰, 徐芳芳, 等. 大数据驱动的热毒宁注射液金青醇沉关键工艺参数辨识研究 [J]. *中国中药杂志*, 2020, 45(1): 233.
- [88] Dai S, Xu B, Shi G, et al. SeDeM expert system for directly compressed tablet formulation: a review and new perspectives [J]. *Powder Technol*, 2019, 342: 517.
- [89] 张毅, 徐冰, 孙飞, 等. 中药提取物粉末物理指纹谱研究及应用 [J]. *中国中药杂志*, 2016, 41(12): 2221.
- [90] 戴胜云, 徐冰, 张志强, 等. 中药粉体物性周期表的建立及应用 [J]. *中华中医药杂志*, 2019, 34(5): 1902.
- [91] 王芬, 徐冰, 张坤峰, 等. 基于粉体压缩行为的中药半浸膏粉“药辅合一”科学内涵研究 [J]. *中国中药杂志*, 2020, 45(1): 274.
- [92] 戴胜云. 中药直接压片处方智能设计方法研究 [D]. 北京: 北京中医药大学, 2019.
- [93] Zhang Y, Xu B, Wang X, et al. Optimal selection of incoming materials from the inventory for achieving the target drug release profile of high drug load sustained-release matrix tablet [J]. *AAPS Pharm Sci Tech*, 2019, 20: 76.
- [94] Zhang Y, Xu B, Wang X, et al. Setting up multivariate specifications on critical raw material attributes to ensure consistent drug dissolution from high drug-load sustained-release matrix tablet [J]. *Drug Dev Ind Pharm*, 2018, 44(11): 1733.

[责任编辑 孔晶晶]