# 工业大数据分析指南

Series on the Industrial Internet

工业互联网产业联盟 大数据系统软件国家工程实验室 编著

非外借

中国工信出版集团

電子工業出版社
PUBLISHING HOUSE OF ELECTRONICS INDUSTRY
http://www.phei.com.cn

“十三五”国家重点出版物出版规划项目
工业和信息产业科技与教育专著出版资金项目

国之重器出版工程
网络强国建设

工业互联网丛书

# 工业大数据分析指南

工业互联网产业联盟　大数据系统软件国家工程实验室　编著

電子工業出版社
Publishing House of Electronics Industry
北京 · BEIJING

## 内 容 简 介

如今，全球掀起了以制造业转型升级为首要任务的新一轮工业变革，工业领域通过与云计算、大数据、物联网、人工智能等技术结合引领新一轮科技革命，拉动工业经济的创新发展。工业大数据分析技术作为工业大数据的核心技术之一，可使工业大数据产品具备海量数据的挖掘能力、多源数据的集成能力、多类型知识的建模能力、多业务场景的分析能力、多领域知识的发掘能力等，对驱动企业业务创新和转型升级具有重大的作用。

本书围绕着“工业大数据分析”这一重要议题，对通用的工业大数据分析方法和分析流程进行归纳总结，对其关键共性进行辨识、抽象和提升，而非针对某一特定行业、企业或产品进行阐述。本书从工业大数据分析的概念、特殊性及常见的问题入手，提出了工业大数据分析框架，并详细阐述了业务理解、数据理解、数据准备、数据建模、模型的验证与评估、模型的部署这 6 个工业大数据分析的基本步骤，最后对工业大数据分析的未来进行了展望，为工业大数据分析相关技术研发、设计建模和应用落地提供了理论依据和标准化方法。

本书更加关注方法论而非某些具体的技术，因此具有更加广泛的通用性和相对普遍的指导意义，可供从事工业大数据分析等相关政策制定、技术研发、产业应用、产品推广的政府机关、科研机构、企业单位及个人参考借鉴。

未经许可，不得以任何方式复制或抄袭本书之部分或全部内容。
版权所有，侵权必究。

**图书在版编目（CIP）数据**

工业大数据分析指南 / 工业互联网产业联盟等编著. —北京：电子工业出版社，2019.10
（工业互联网丛书）
ISBN 978-7-121-37332-9

Ⅰ. ①工… Ⅱ. ①工… Ⅲ. ①制造工业—数据管理—指南 Ⅳ. ①F407.4-62

中国版本图书馆 CIP 数据核字（2019）第 189951 号

责任编辑：郭穗娟
印　　刷：固安县铭成印刷有限公司
装　　订：固安县铭成印刷有限公司
出　　版：电子工业出版社
　　北京市海淀区万寿路 173 信箱　邮编　100036
开　　本：720×1000　1/16　印张：8　字数：126 千字
版　　次：2019 年 10 月第 1 版
印　　次：2019 年 10 月第 1 次印刷
定　　价：78.00 元

凡所购买电子工业出版社图书有缺损问题，请向购买书店调换。若书店售缺，请与本社发行部联系，联系及邮购电话：（010）88254888，88258888。

质量投诉请发邮件至 zlts@phei.com.cn，盗版侵权举报请发邮件至 dbqq@phei.com.cn。

本书咨询联系方式：（010）88254502，guosj@phei.com.cn。

此为试读，需要完整PDF请访问：www.ertongbook.com

# “国之重器出版工程”
## 编辑委员会


编辑委员会主任：苗　圩

编辑委员会副主任：刘利华　辛国斌

编辑委员会委员：

| | | | | |
|---|---|---|---|---|
| 冯长辉 | 梁志峰 | 高东升 | 姜子琨 | 许科敏 |
| 陈　因 | 郑立新 | 马向晖 | 高云虎 | 金　鑫 |
| 李　巍 | 高延敏 | 何　琼 | 刁石京 | 谢少锋 |
| 闻　库 | 韩　夏 | 赵志国 | 谢远生 | 赵永红 |
| 韩占武 | 刘　多 | 尹丽波 | 赵　波 | 卢　山 |
| 徐惠彬 | 赵长禄 | 周　玉 | 姚　郁 | 张　炜 |
| 聂　宏 | 付梦印 | 季仲华 | | |

**专家委员会委员**（按姓氏笔画排列）：

**于　全**　中国工程院院士

**王少萍**　“长江学者奖励计划”特聘教授

**王建民**　清华大学软件学院院长

**王哲荣**　中国工程院院士

**王　越**　中国科学院院士、中国工程院院士

**尤肖虎**　“长江学者奖励计划”特聘教授

**邓宗全**　中国工程院院士

**甘晓华**　中国工程院院士

**叶培建**　中国科学院院士

**朵英贤**　中国工程院院士

**刘大响**　中国工程院院士

**刘怡昕**　中国工程院院士

**刘韵洁**　中国工程院院士

**苏彦庆**　“长江学者奖励计划”特聘教授

**孙逢春**　中国工程院院士

**邬贺铨**　中国工程院院士

**朱英富**　中国工程院院士

苏哲子　中国工程院院士

李伯虎　中国工程院院士

李应红　中国科学院院士

李新亚　国家制造强国建设战略咨询委员会委员、
中国机械工业联合会副会长

杨德森　中国工程院院士

张宏科　北京交通大学下一代互联网互联设备国家
工程实验室主任

陆建勋　中国工程院院士

陆燕荪　国家制造强国建设战略咨询委员会委员、原
机械工业部副部长

陈一坚　中国工程院院士

陈懋章　中国工程院院士

金东寒　中国工程院院士

周立伟　中国工程院院士

郑纬民　中国计算机学会原理事长

郑建华　中国科学院院士

**屈贤明** 国家制造强国建设战略咨询委员会委员、工业和信息化部智能制造专家咨询委员会副主任

**项昌乐** “长江学者奖励计划”特聘教授，中国科协书记处书记，北京理工大学党委副书记、副校长

**柳百成** 中国工程院院士

**闻雪友** 中国工程院院士

**徐德民** 中国工程院院士

**唐长红** 中国工程院院士

**黄卫东** “长江学者奖励计划”特聘教授

**黄先祥** 中国工程院院士

**黄先祥** 中国工程院院士

**黄维中** 国科学院院士、西北工业大学常务副校长

**董景辰** 工业和信息化部智能制造专家咨询委员会委员

**焦宗夏** “长江学者奖励计划”特聘教授

# “工业互联网丛书”
# 编辑委员会


编辑委员会主任：陈肇雄

编辑委员会副主任：张　峰　王新哲

专家委员会委员：

王钦敏　周　济　邬贺铨　陈左宁　李伯虎

孙家广　刘韵洁　方滨兴　张　军　房建成

主　　编：王建民

编辑委员会委员：

韩　夏　刘　多　谢少锋　赵志国　闻　库

谢远生　李　颖　鲁春丛　杨宇燕　陈家春

祁　锋　尹丽波　徐晓兰　余晓晖　林　啸

冯　伟　付景广　徐　波　赵　征

本书编写组成员：王建民　郭朝晖　王　晨

# “工业互联网丛书”
# 总序

2019 年正值互联网发明 50 周年，也是中国全功能接入互联网 25 年。中国互联网用户普及率接近 60%，以国内电商为代表的消费互联网业务全球领先，互联网已经深入到社会生活的方方面面，数字经济对经济增长的贡献日益显著。现在我国消费互联网在教育、医疗、养老、文化、旅游和海外电商等应用领域还有待拓展和深化，但更大的发展空间在工业互联网，工业互联网可看做互联网的“下半场”，既是互联网发展的新动能，也是拉动经济增长的新引擎。2019 年的《政府工作报告》提出，打造工业互联网平台，拓展“智能+”，为制造业转型升级赋能。工业互联网的提出不仅是与消费互联网新旧动能的接续，而且正好与中国从中高速发展向高质量发展的转型时间对应，是实现质量变革、效率变革和动力变革的关键。

“工业互联网丛书”站在新一代科技革命到来和国际竞争面临前所未有不确定性的时代格局下，透析了工业互联网提出的背景，解读了工业互联网体系的构成，以丰富的实践案例佐证了工业互联网的成功应用；对照国外工业互联网的发展战略与布局，分析了中国发展工业互联网的有利条件与不利因素，展望数字经济时代中国工业互联网发展的机遇与挑战，提出了深入促进工业互联网发展的建议。

互联网走过 50 年，而工业互联网现在才刚开始。互联网从面向人到面向企业，在技术要求、实施主体、产业生态、商业模式等方面都有很大差异。工业互联网在全球也处在发展过程中，对于工业现代化任务很重的中国来说，全面实现工业互联

网的路还很长，但中国工业互联网的实践一定会对全球工业互联网的发展做出自己的贡献。

“工业互联网丛书”的写作团队对国内外工业互联网的情况有较全面的了解，深入企业获得第一手的案例与诉求，已编写出数十本有关工业互联网的白皮书和研究报告，在此基础上用简洁和通俗的语言介绍工业互联网。在中国工业互联网起步阶段“工业互联网丛书”的推出正当其时，但对工业互联网的理解和应用肯定随时间而深化，有待更多的实践来补充和完善。

中国工程院院士　邬贺铨

2019 年 8 月 12 日



此为试读，需要完整PDF请访问：www.ertongbook.com

# 序　言

如今，全球掀起了以制造业转型升级为首要任务的新一轮工业变革，工业大数据作为引领这场变革的主要驱动力，已经成为当今工业领域的热点之一。

新一代信息技术与制造业的深度融合，将促进工业领域的服务转型和产品升级，重塑全球制造业的产业格局。为紧紧抓住这一重大历史机遇，抢占制造业新一轮竞争制高点，党中央高度重视并做出长期性、战略性部署。党的十九大报告指出，要“加快建设制造强国，加快发展先进制造业，推动互联网、大数据、人工智能和实体经济深度融合”。

工业大数据是智能制造的核心，以“大数据+工业互联网”为基础，用云计算、大数据、物联网、人工智能等技术引领工业生产方式的变革，拉动工业经济的创新发展。工业大数据分析技术作为工业大数据的核心技术之一，可使工业大数据产品具备海量数据的挖掘能力、多源数据的集成能力、多类型知识的建模能力、多业务场景的分析能力、多领域知识的发掘能力等，对驱动企业业务创新和转型升级具有重大的作用。可以从以下 3 个方面来理解。

首先，资源优化是分析的目标。企业之间竞争的本质是资源配置效率的竞争，优化资源配置效率是企业技术创新应用的主要动力，也是工业大数据分析的核心目标。工业大数据分析是实现新一代信息技术与制造业融合的重要技术支撑，其目的是不断优化资源的配置效率，实现生产全过程的可视化、高端定制化生产、产品生产节能增效、供应链配置优化、企业智能化管理等，达到提升质量、降低成本、灵活生产、提高满意度等目的，促进制造业全要素生产率的提高。

其次，数据建模是分析的关键。来自产品生命周期各个环节中的海量数据，为工业大数据分析提供了前提和基础，而海量的工业数据如果不经过清洗、加工和建模等处理是无法直接应用于实际的业务场景的。工业大数据分析通过模型来描述对象，构建复杂工业过程与知识之间的映射，实现知识清晰化、准确化的表达。

最后，知识转化是分析的核心。确定性和稳定性是工业应用的两个基本特点，这就决定了工业大数据分析技术就是感知信息和提炼知识，其核心在于如何把海量数据转化为信息，把信息转化为知识，把知识转化为决策，以解决制造过程的复杂性和不确定性等问题。

本书是在新形势下对工业大数据分析的关键共性问题进行辨识、抽象和提升，适应当前工业大数据的应用需求和技术变革，具有较为广泛的通用性和相对普遍的指导意义，适于工业领域的企业、机构研究和参考。希望通过与业界的分享，共同推动工业大数据的开发利用和应用推广，为制造强国和网络强国建设添薪助力！

谢少锋

2019 年 5 月

# 编写说明

工业大数据是工业领域相关数据集的总称，是工业互联网的核心，是智能制造的关键。工业大数据分析技术作为工业大数据的核心技术之一，是工业智能化发展的重要基础和关键支撑。为此，在工业互联网产业联盟的指导下，大数据特设组主持编写了这本《工业大数据分析指南》。

本书旨在对通用的工业大数据分析方法和分析流程进行归纳和总结，对其关键共性进行辨识、抽象和提升，而非针对某一特定行业、企业或产品进行阐述。本书更加关注方法论而非某些具体的技术，因此，具有更加广泛的通用性和相对普遍的指导意义。

本书共 9 章，第 1 章论述了工业大数据分析的概念、特殊性及常见的问题。第 2 章提出了工业大数据分析框架，简要介绍了 CRISP-DM 模型，并针对该模型落地的难点及其使用的指导思想展开讨论。第 3～8 章依次对业务理解、数据理解、数据准备、数据建模、模型的验证与评估、模型的部署这 6 个 CRISP-DM 模型的基本步骤进行了详细的阐述，从需求分析到目标评估，从数据来源到数据分类，从数据预处理到建模过程，从模型验证到部署问题处理，对每个步骤中的原理方法、分析过程、处理方式、问题排除等都进行了讲解和说明。第 9 章对工业大数据分析的未来进行了展望。

本书由工业互联网产业联盟大数据特设组组长单位清华大学（大数据系统软件国家工程实验室）牵头编写，在编写过程中得到了工信部领导的悉心指导和相关单位的有力支持。特别感谢清华大学孙家广院士、工信部信软司谢少锋司长、中国信

息通信研究院余晓晖副院长等给予的全面指导。同时，北京工业大数据创新中心的李三华、田春华，清华大学的任艮全、徐哲、强逍等在本书的编写过程中也给予了无私的帮助，在此表示诚挚的谢意！

工业大数据作为新兴概念，其数据分析的原则、手段、方法和流程还很模糊，对海量数据的挖掘、分析和处理等技术仍在不断地发展和进步，由于作者自身的能力和水平有限，本书不可避免地存在诸多的缺点和不足，期待各位读者能够积极发现问题，并予以批评指正。

工业互联网产业联盟　大数据系统软件国家工程实验室

2019 年 5 月

# 目录 / Contents





此为试读，需要完整PDF请访问：www.ertongbook.com