

大数据



大数据驱动创新方法 工作平台

崔立真 鹿旭东 郭 伟 刘士军 编著



山东大学出版社

大数据驱动创新方法工作平台

崔立真 鹿旭东 郭 伟 刘士军 编著

山东大学出版社

图书在版编目(CIP)数据

大数据驱动创新方法工作平台/崔立真等编著. --
济南: 山东大学出版社, 2017.10
ISBN 978-7-5607-5871-8

I. ①大… II. ①崔… III. ①数据处理软件—研究
IV. ①TP274

中国版本图书馆 CIP 数据核字(2017)第 261828 号

责任策划: 张晓林

责任编辑: 李 港

封面设计: 张 荔

出版发行: 山东大学出版社

社 址 山东省济南市山大南路 20 号

邮 编 250100

电 话 市场部(0531)88364466

经 销: 山东省新华书店

印 刷: 山东和平商务有限公司

规 格: 720 毫米×1000 毫米 1/16

13.25 印张 242 千字

版 次: 2017 年 10 月第 1 版

印 次: 2017 年 10 月第 1 次印刷

定 价: 32.80 元

版权所有, 盗印必究

凡购本书, 如有缺页、倒页、脱页, 由本社营销部负责调换

前 言

国务院印发的《关于加快构建大众创业万众创新支撑平台的指导意见》对大力推进大众创业万众创新平台建设和推动实施“互联网+”行动作出了具体部署,并特别提到了推进创新创业平台建设的“四众”新模式。

指导意见提出了一个“四众”概念,即众创、众包、众扶、众筹,要加快推动这“四众”新模式、新业态的发展。众创,是汇集大家的智慧搞创新,通过创业创新服务平台,聚集全社会各类创新资源,大幅降低创业创新成本,使每一个具有科学思维和创新能力的人都可参与创新,形成大众创造、释放众智的新局面。众包,是汇集众人的力量增加就业,借助互联网等手段,将传统由特定企业和机构完成任务向自愿参与的所有企业和个人进行分工,最大限度地利用大众力量,以更高的效率、更低的成本开拓创新、便捷创业。众扶,是汇集大家的力量来助推创业,通过政府和公益机构支持、企业帮扶援助、个人互助互扶等多种方式,共助小微企业和创业者成长,构建创业创新发展良好生态。众筹,是通过互联网平台向社会募集资金,更灵活高效地满足创业创新的融资需求,拓展创业创新的投融资新渠道。

同时,随着大数据时代的到来,大数据技术正在从产业结构、传统制造业升级、商业组织、“互联网+”和“大众创业,万众创新”等众多方面影响经济增长方式,助推创新,驱动发展。大数据分析形成的“相关关系”为我们认识世界提供了除“因果关系”外的另一种方式,引起科研方式的深刻变革,成为创新的新动力。大数据技术的一项重大意义在于其能够影响科学研究本身的发展,使科学从过去的假设驱动型转化为数据驱动型。应用大数据技术,人们开展科学研究不再是从提出自己的假设出发,而是先进行数据分析,然后再提出科学假设。大数据技术的兴起与发展,正在推动传统创新模式的变革。

本书共分为三篇:创新方法篇,对研发背景、传统创新方法研究、BDDI 方法论、BDDI 系统平台、项目关键技术和示范应用案例等进行了系统概述。平台功能需求篇。平台使用说明篇,包含需求挖掘和行为模式分析系统、模式创新智能

推荐系统、社会化众智众创互动反馈评价系统、大数据跨界融合公共支撑系统四个子系统使用手册和整个平台的演示文档。

本书在科技部创新方法工作专项课题(课题号:2015IM010200)的支持下完成,由崔立真、鹿旭东、郭伟、刘士军共同编著。创新方法篇由崔立真负责编撰完成,平台功能需求篇由鹿旭东和郭伟负责编撰完成,平台使用说明篇由郭伟和刘士军负责编撰完成。

本书在撰写过程中,山东大学的宋伟凤、李耀坤、张盘龙、杨哲、杨卫刚、傅静华等同学也参与了部分章节的材料组织、系统开发、文档整理和书稿校对等工作。在此,向所有对本书提供过帮助的单位和个人表示真诚的谢意。

本书在编写过程中,参考了大量文献,向所有作者表示诚挚的敬意与谢意!限于自身水平,书中难免存在不足之处,恳请广大读者批评指正。

作者

2017年7月31日

目 录

创新方法篇	(1)
1 大数据驱动创新方法	(3)
1.1 大数据驱动创新的背景与意义	(3)
1.1.1 创新——引领发展的第一动力	(3)
1.1.2 “互联网+”与大数据时代	(4)
1.1.3 大数据驱动创新的提出	(16)
1.2 创新方法研究	(17)
1.2.1 创新方法	(17)
1.2.2 创新相关方法	(19)
1.2.3 创新方法相关工具现状	(57)
1.3 大数据驱动创新方法论	(61)
1.3.1 BDDI 理论基础	(62)
1.3.2 BDDI 方法论	(69)
1.4 大数据驱动创新平台	(75)
1.4.1 BDDI 平台架构	(76)
1.4.2 BDDI 平台实现	(81)
1.4.3 BDDI 平台创新过程	(83)
1.5 平台采用的关键技术	(85)
1.6 BDDI 示范应用案例	(88)
2 TRIZ 方法介绍	(94)
2.1 创新方法论	(94)
2.1.1 TRIZ 的理论前提和基本认识	(94)
2.1.2 TRIZ 的理论体系	(94)
2.1.3 TRIZ 的解题流程及模式	(95)
2.2 TRIZ 基本概念(术语)	(96)

2.2.1	TRIZ 问题分析	(96)
2.2.2	TRIZ 系统分析	(96)
2.2.3	TRIZ 功能分析	(97)
2.2.4	建立功能结构的方法	(98)
2.2.5	TRIZ 系统裁剪	(99)
2.2.6	TRIZ 问题模型	(99)
2.2.7	TRIZ 解决矛盾的方法与流程	(101)
2.2.8	TRIZ 解决物场问题标准解法	(102)
2.2.9	TRIZ 技术进化规律(八大进化法则)	(103)
2.2.10	ARIZ 算法	(107)
2.2.11	计算机辅助创新技术(CAI)	(108)
2.3	TRIZ 示范应用报告案例	(109)
平台功能需求篇		(115)
3	大数据驱动创新方法工作服务平台需求说明	(117)
3.1	平台简介	(117)
3.2	平台功能需求	(117)
3.2.1	概述	(117)
3.2.2	主要功能需求	(118)
3.3	平台非功能性需求	(118)
3.3.1	性能需求	(118)
3.3.2	安全需求	(119)
3.3.3	质量需求	(119)
3.4	需求挖掘和行为模式分析系统需求说明书	(120)
3.4.1	总体需求概述及相关术语	(120)
3.4.2	功能性需求	(120)
3.4.3	主要业务	(120)
3.5	模式创新智能推荐系统需求说明书	(122)
3.5.1	总体需求概述及相关术语	(122)
3.5.2	功能性需求	(123)
3.5.3	解决方案推荐主要业务	(124)
3.5.4	方案管理主要业务	(125)
3.6	社会化众智众创互动反馈评价系统需求说明书	(126)

3.6.1 总体需求概述	(126)
3.6.2 功能性需求	(126)
3.6.3 系统角色	(127)
3.7 大数据跨界融合公共支撑系统需求说明书	(128)
3.7.1 总体需求概述	(128)
3.7.2 功能性需求	(129)
平台使用说明篇	(131)
4 大数据驱动创新方法工作服务平台使用手册	(133)
4.1 引言	(133)
4.1.1 手册编写目的	(133)
4.1.2 平台简介	(133)
4.1.3 系统关系图	(134)
4.2 系统门户	(134)
4.2.1 首页	(134)
4.2.2 登录和注册	(134)
4.2.3 后台管理	(135)
4.3 需求挖掘和行为模式分析系统	(136)
4.3.1 创意管理	(136)
4.3.2 创意引导	(137)
4.4 模式创新智能推荐系统	(140)
4.4.1 方案管理	(140)
4.4.2 解决方案推荐	(140)
4.5 社会化众智众创互动反馈评价系统	(142)
4.5.1 首页	(142)
4.5.2 信息详情界面	(142)
4.6 大数据跨界融合公共支撑系统	(144)
4.6.1 架构设计	(144)
4.6.2 节点任务	(144)
4.6.3 树干节点	(144)
4.6.4 树枝节点	(145)
4.6.5 树叶节点	(145)

5	需求挖掘和行为模式分析系统用户手册	(147)
5.1	需求挖掘和行为模式分析系统简介	(147)
5.2	使用说明	(147)
5.2.1	需求挖掘和行为模式分析系统概述	(147)
5.2.2	业务流程图	(147)
5.2.3	需求挖掘和行为模式分析	(148)
5.2.4	创意管理	(154)
5.3	案例	(155)
5.3.1	黏胶长丝调湿工艺改进	(155)
5.3.2	可弹琴的高跟鞋	(156)
6	模式创新智能推荐系统用户手册	(159)
6.1	模式创新智能推荐系统简介	(159)
6.2	使用说明	(159)
6.2.1	模式创新智能推荐系统概述	(159)
6.2.2	业务流程图	(159)
6.2.3	方案管理	(159)
6.2.4	解决方案推荐	(160)
6.2.5	方案评论	(165)
6.3	案例	(165)
6.3.1	纸硬币集成终端电源问题解决	(165)
6.3.2	无变速器机床	(167)
7	社会化众智众创互动反馈评价系统用户手册	(170)
7.1	软件简介	(170)
7.2	使用说明	(170)
7.2.1	社会化众智众创互动反馈评价系统概述	(170)
7.2.2	使用指南	(171)
8	大数据跨界融合公共支撑系统用户手册	(175)
8.1	适用网页类型	(175)
8.2	可爬取数据类型	(177)
8.3	基本知识储备	(177)
8.4	爬虫整体架构	(177)
8.4.1	架构设计	(177)
8.4.2	节点任务	(178)

8.5 爬虫详细参数.....	(178)
8.5.1 树干节点.....	(178)
8.5.2 树枝节点.....	(182)
8.5.3 树叶节点.....	(184)
8.6 注意事项.....	(191)
9 大数据驱动创新方法工作服务平台使用实例	(192)

创新方法篇

创新是指以现有的知识和物质,在特定的环境中,改进或创造新的事物。党的十八届五中全会提出的“五大发展理念”中,排在首位的就是“创新发展”。创新是引领发展的第一动力。

大数据驱动创新以大数据为基础,通过数据融合、存储、计算、分析和可视化,形成大数据跨界融合公共支撑平台。其中,大数据来源于互联网的网络世界和行业的信息系统。基于大数据跨界融合公共支撑平台,实现需求的实时洞察与智能捕获,以及模式创新的智能推荐。

本篇介绍大数据驱动创新方法和 TRIZ 方法等。

1 大数据驱动创新方法

党的十八大以来,国家把创新发展提高到了事关国家和民族前途命运的高度,摆到了国家发展全局的核心位置。为了顺应大数据创新时代的潮流,我们充分利用大数据并结合行业特点研发了基于众智的大数据驱动创新平台。

1.1 大数据驱动创新的背景与意义

1.1.1 创新——引领发展的第一动力

1.1.1.1 创新的定义

创新是指以现有的知识和物质,在特定的环境中,改进或创造新的事物(包括但不限于各种方法、元素、路径、环境等),并能获得一定有益效果的行为。

创新方法是为了科学、高效地开展创新活动,又多又好地获得创新成果,加快增强自主创新能力所采取的思想、行动、程序、手段、工具和技巧的总和。

创新方法论是研究创新过程中有没有逻辑顺序、规则、方法,以及有什么样的顺序、规则、方法的哲学学科。

1.1.1.2 创新的重要性

(1) 国家与创新

党的十八大以来,习近平总书记针对创新发展提出了一系列重要思想和论断。党的十八届五中全会提出的“五大发展理念”中,排在首位的就是“创新发展”。创新是引领发展的第一动力。

实施创新驱动发展战略,要推动以科技创新为核心的全面创新,坚持需求导向和产业化方向,坚持企业在创新中的主体地位,发挥市场在资源配置中的决定性作用和社会主义制度优势,增强科技进步对经济增长的贡献度,形成新的增长动力源泉,推动经济持续健康发展。

(2) 企业与创新

创新是企业生存与发展的根本。美国著名经济学家、管理大师熊彼特曾说

过;创新是企业对生产要素的重新组合。创新可促进企业组织形式的改善和管理效率的提高,从而使企业不断提高效率,不断适应经济发展的要求。对一个企业而言,创新可以包括很多方面,如管理创新、文化创新、技术创新、制度创新、营销创新等。纵观当代企业,唯有不断创新,才能在竞争中处于主动,立于不败之地。

1.1.2 “互联网+”与大数据时代

1.1.2.1 “互联网+”的应用

(1)“互联网+”的内涵与意义

“互联网+”是创新 2.0 下的互联网发展新业态,是知识社会创新 2.0 推动下的互联网形态演进及其催生的经济社会发展新形态。通俗地说,“互联网+”就是“互联网+各个传统行业”,但并不是两者简单相加。

《国务院关于积极推进“互联网+”行动的指导意见》指出,“互联网+”是把互联网的创新成果与经济社会的各领域深度融合,推动技术进步、效率提升和组织变革,提升全社会的创新力和生产力,形成更广泛的、以互联网为基础设施和实现工具的经济发展新形态。

威客理论和互联网进化论的创始人刘锋分析了互联网和脑科学的关系(见图 1-1)。

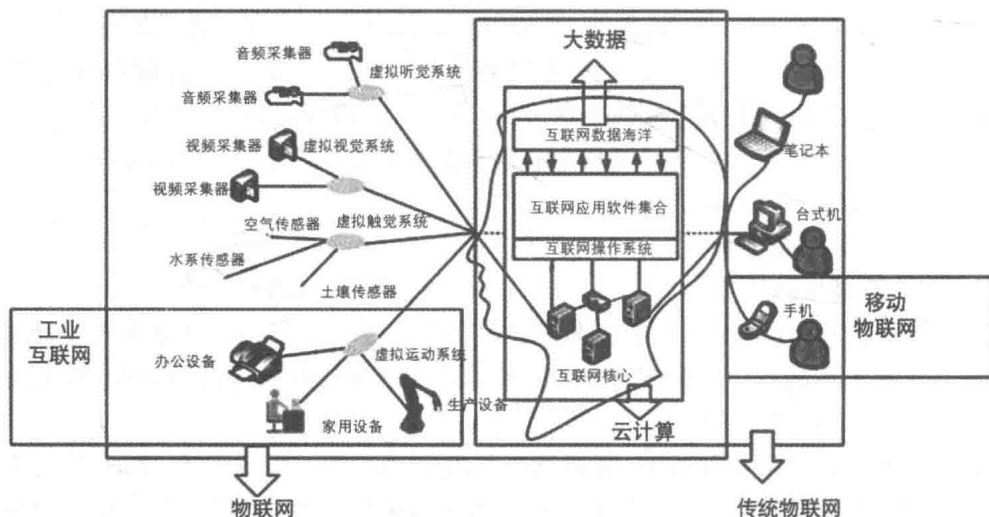


图 1-1 互联网和脑科学的关系

A.物联网是互联网大脑的感觉神经系统。因为物联网重点突出了传感器感知的概念,也具备了网络线路传输、信息存储和处理、行业应用接口等功能,而

且也往往与互联网共用服务器、网络线路和应用接口,使人与人(Human to Human,H2H)、人与物(Human to Thing,H2T)、物与物(Thing to Thing,T2T)之间的交流变成可能,最终将使人类社会、信息空间和物理世界(人—机—物)融为一体。

B.云计算是互联网大脑的中枢神经系统。在互联网虚拟大脑的架构中,互联网虚拟大脑的中枢神经系统是将互联网的核心硬件层、核心软件层和互联网信息层统一起来,为互联网各虚拟神经系统提供支持和服务。从定义上看,云计算与互联网虚拟大脑中枢神经系统的特征非常吻合。在理想状态下,物联网的传感器和互联网的使用者通过网络线路和计算机终端与云计算进行交互,向云计算提供数据,接受云计算提供的服务。

C.大数据是互联网智慧和意识产生的基础。随着博客、社交网络、云计算、物联网的兴起,互联网上的数据信息正以前所未有的速度增长和累积。互联网用户的互动,企业和政府的信息发布,物联网传感器感应的实时信息,每时每刻都在产生大量的结构化和非结构化数据。这些数据分散在整个互联网网络体系内,体量极其巨大,并且蕴含了对经济、科技、教育等领域非常宝贵的信息。这就是互联网大数据兴起的根源和背景。

与此同时,以深度学习为代表的机器学习算法在互联网领域的广泛使用,使得互联网大数据开始与人工智能进行更为深入的结合。这其中就包括在大数据和人工智能领域领先的世界级公司,如谷歌、微软、百度等。2011年,谷歌开始将深度学习运用在自己的大数据处理上。互联网大数据与人工智能的结合为互联网大脑的智慧和意识产生奠定了基础。

D.工业4.0和工业互联网本质上是互联网运动神经系统的萌芽。互联网中枢神经系统也就是云计算中的软件系统控制工业企业的生产设备、家庭的家用设备、办公室的办公设备,通过智能化、3D打印、无线传感等技术使机械设备成为互联网大脑改造世界的工具。同时,这些智能制造和智能设备也源源不断地向互联网大脑反馈大数据流,供互联网中枢神经系统决策使用。

E.“互联网+”的核心是互联网的进化和扩张,反映了互联网从广度、深度方面融合和介入现实世界的动态过程。加快推进“互联网+”发展,有利于重塑创新体系、激发创新活力、培育新业态和创新公共服务模式,对打造大众创业、万众创新,启动公共产品、公共服务“双引擎”,主动适应和引领经济发展新常态,形成经济发展新动能,实现中国经济体制增效升级具有重要意义。

(2)“互联网+”时代的创业创新——某鸡尾酒的互联网思维

如果问鸡尾酒跟互联网有什么关系,可能大部分人认为就是利用互联网进行宣传和销售,而在安兴华创业过程中,从市场调研、产品设计到推广和销

售,各个环节都贯彻着互联网思维。

首先进行市场需求调研,确定鸡尾酒是否有市场需求,目标客户群体有什么样的特征,产品该如何定价等问题。传统的市场调研方式通常是发放问卷或对客户进行访谈,而安兴华很好地利用了互联网这一工具。安兴华制作了一个无法真正完成购买行为的测试网站,让百度联盟在其网站上发布广告,通过比较不同广告的点击率,发现鸡尾酒相关广告点击率和购买点击率远高于均值,女性购买者占了七成,购买者主要集中在北上广一线城市,不同价位的购买点击率呈均匀分布。这些数据说明鸡尾酒的市场需求是极大的,尤其是在一线城市,女性也确实是消费的主力人群,而且客户并不是一味地追求廉价。这让安兴华确定了鸡尾酒的市场需求足够大,尤其是推出一些针对女性消费者的产品,将会非常受欢迎。

在明确需求之后,安兴华开始设计产品。产品究竟该采用何种瓶型设计呢?鸡尾酒这种产品的外观是极其重要的,优雅的外观设计能在第一时间吸引消费者的眼球。设计师提供了几种瓶身设计,如何选择呢?安兴华认为,重要的是消费者喜欢哪种瓶型,而不是自己喜欢哪种。有了互联网的帮助,了解消费者的喜好并不是一件难事。他设计了一个微信上的性格测试,将哪种瓶型更美观作为一道题放在了测试中,通过微信好友间的相互转发,仅仅1小时就获取了两万多份答案,而实际上半个小时的时候答案已经显现。在三种酒瓶设计中,最终中间的酒瓶因被更多消费者喜欢而胜出(见图1-2)。



图 1-2 被消费者选择的最终设计

在销售的过程中,互联网更是提供了无限大的平台。首先,安兴华在朋友圈推出了颜值抵现金的活动,客户在其页面自拍并转发朋友圈,即可根据客户的颜值给予一定的现金优惠。这一活动在8小时内就有47万的页面访问量、10万

的独立访客,5天后取得了195万的页面访问量、48万人品牌触达、转发率8.5%的良好成绩,很好地对品牌进行了宣传。为了更好地实现数据驱动营销,他们创办了自己品牌的公众号,经过某次对二次复购人群的回访发现,40%的客户是买给家人,其中75%的是买给妈妈,于是紧接着推出了感恩节的营销活动——“这个感恩节,我要买给妈妈”。这一策略让他们的鸡尾酒在感恩节期间取得了很棒的销量。

安兴华将鸡尾酒与互联网相结合,大大提高了产品的销量。

1.1.2.2 大数据的应用

(1) 大数据的基本概念

随着社交网络、电子商务和移动互联网的发展,人类社会的数据量迅速激增。据统计,目前人类一年产生的数据量就相当于人类进入现代化以前所有历史的总和。

在维克托·迈尔-舍恩伯格及肯尼斯·库克耶编写的《大数据时代》中,大数据不采用随机分析法(抽样调查)这样的捷径,而采用所有数据进行分析处理。大数据具有4V特点:Volume(大量)、Velocity(高速)、Variety(多样)、Value(价值)。

对于“大数据”(Big data),研究机构高德纳(Gartner)给出了这样的定义:“大数据”是需要新处理模式才能具有更强的决策力、洞察发现力和流程优化能力来适应海量、高增长率和多样化的信息资产。

根据维基百科的定义,大数据是指无法在可承受的时间范围内用常规软件工具进行捕捉、管理和处理的数据集合。

(2) 大数据的价值

大数据技术的战略意义不在于掌握庞大的数据信息,而在于对这些含有意义的数据进行专业化处理。换言之,如果把大数据比作一种产业,那么这种产业实现盈利的关键在于提高对数据的“加工能力”,通过“加工”实现数据的“增值”。

谷歌公司每天都会收到来自全球超过30亿条的搜索指令。经过多年数据的累计,谷歌公司建立了“咳嗽”“发热”等搜索关键字与流感地区的联系。于是在2009年,谷歌成功地在美国预测了冬季流感的传播,并且精确到地区和州。2013年奥巴马连任成功,成为第一位依靠大数据当选的美国总统。竞选过程中他的数据挖掘团队为其制定竞选策略,分析选民舆论,使其处处占得先机,最终获得胜利。再来看看大家熟悉的电子商务,如淘宝、亚马逊或京东的页面都经过了大数据技术的优化。你所看到的页面都是根据你的购物行为进行显示的。你所希望购买的商品会被显示在最显眼的地方,你喜欢的商品会被推荐。你在这些网站购物越多,这种分析就越准确,购物体验也就越流畅。

毫无疑问,大数据正在改变着我们的生活。过去几年,无论是医疗、健康、交