

doi:10.19920/j.cnki.jmsc.2024.09.001

# 大数据热潮中被冷落的“第三数据”：以制造业为例<sup>①</sup>

党延忠\*, 胡德强\*, 徐照光\*  
(大连理工大学系统工程研究所, 大连 116024)

**摘要:** 在大数据研究风起云涌的当下, 有一类非常重要、非常有价值的数 据却被冷落了, 这就是本研究提出的“第三数据”。第三数据是在问题处理过程中, 由问题处理人员产生的具有主观特性的数据, 其中蕴含着密度极大的问题处理的经验性知识, 是明确问题、分析问题和解决问题不可或缺的重要知识资源。本研究界定了第三数据的概念, 讨论了第三数据的特点、来源。解析了从实践到经验再到数据的第三数据生成过程, 提出了第三数据的五星螺旋模型。实际案例分析表明, 第三数据的开发利用具有降低成本、提高效益的巨大价值。本文还提出了针对第三数据进行深入研究涉及的十个方面的课题。第三数据具有改造世界的内在属性, 还具有问题驱动的普遍性, 凡是存在问题的任何领域都以某种方式存在着第三数据。因此, 第三数据的理论成果具有更为一般的普遍意义, 不仅适用于制造业, 也适用于服务业, 即使在社会经济领域也具有普遍的适用性。

**关键词:** 大数据; 第三数据; 智能制造

**中图分类号:** F272 **文献标识码:** A **文章编号:** 1007-9807(2024)09-0001-10

## 0 引言

近年来大数据技术及人工智能的发展, 使得智能制造<sup>[1]</sup> 等人机混合智能驱动的管理系统逐渐涌现<sup>[2]</sup>。混合智能管理系统中人和机两大主体的驱动方式有着显著差异, 知识是人类智慧的基础, 而数据是机器智能的燃料。混合智能情境下, 人类智慧和机器智能应当共享、互补, 相互启发式增长<sup>[3]</sup>。个体知识如何汇聚成群体智慧, 并与机器智能进行协同决策是当今混合智能系统中需要研究的重要课题<sup>[4]</sup>。而人类知识如何获取、利用和管理则是其中的基本问题, 也依然是知识管理中的难题<sup>[5]</sup>。只有解决了该问题, 人类智慧才有可能融入智能系统, 才有可能与机器进行智能协同。已有研究已经开始尝试将人类知识引入到数

据驱动的智能机器中, 这种范式被称为“知信机器学习”(informed machine learning)<sup>[6, 7]</sup>。

当前的这些研究忽视了人机系统中一类非常重要的知识, 也即来源于人在问题解决过程中形成的经验知识。虽然已有学者利用大数据技术从“知识密集型”数据中挖掘产品开发<sup>[8]</sup>、故障诊断<sup>[9]</sup>、质量改进等<sup>[10]</sup>方面的知识, 但是这些研究并未对由人在问题处理过程中生成的“知识密集型”数据的产生机理与应用机制进行系统化的研究。

本研究基于汽车制造企业的应用成果, 提出了“第三数据”的新概念, 阐述了其来源和特点, 并利用真实案例讨论了“第三数据”的应用前景和意义, 试图引起学界和企业界对“第三数据”的思考和重视。

① 收稿日期: 2022-10-03; 修订日期: 2023-10-12。

基金项目: 国家自然科学基金资助项目(72231010; 72001034; 71871041)。

通讯作者: 党延忠(1954—), 男, 辽宁营口人, 博士, 教授, 博士生导师。Email: yzhdang@dlut.edu.cn;

胡德强(1988—), 男, 黑龙江齐齐哈尔人, 博士, 助理研究员。Email: deqianghu@dlut.edu.cn;

徐照光(1989—), 男, 江西景德镇人, 博士, 副教授。Email: zhgxu@dlut.edu.cn

# 1 第三数据的概念

## 1.1 第三数据的来源

产品制造过程是个复杂系统,一方面“产品制造系统”在生产中产生大量数据,另一方面,

“数据加工系统”把“产品制造系统”产生的数据与企业外部数据融合处理后,为“产品制造系统”再利用,从而形成图 1 所示的“产生 - 加工 - 利用 - 再产生”的螺旋式数据流,使数据不断地积累形成规模恢弘的“大数据”. 本文把这些数据分为三种类型:第一数据、第二数据和第三数据.

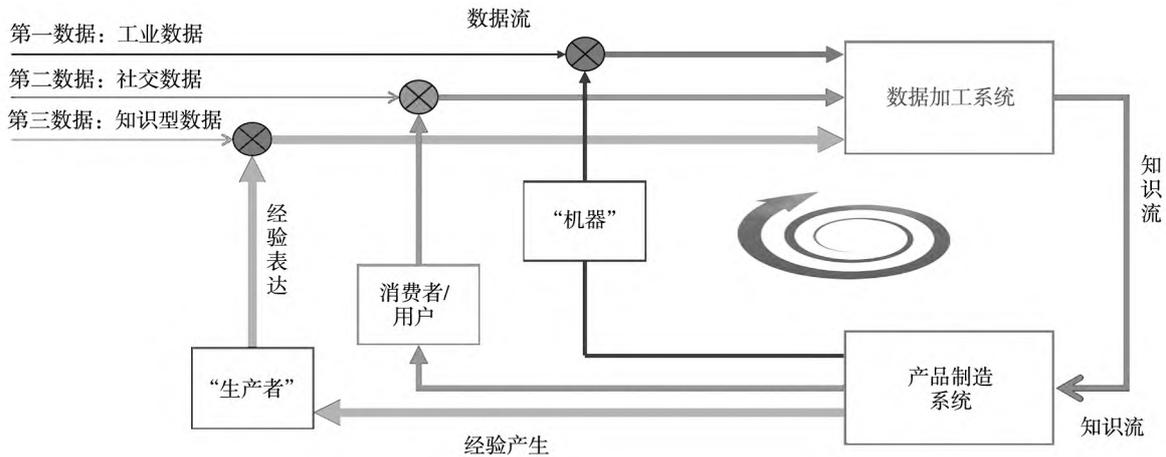


图 1 生产制造中的三类数据流

Fig. 1 Three types of data flows in manufacturing

“第一数据”(First Data)是图 1 中“产品制造系统”在产品全生命周期各个环节中产生的生产与运营数据,是对生产状态与运营情况的记录和描述. 一般由工业传感器、工业自动控制系统、工业物联网等技术自动采集或人工记录,并存储在 ERP、CRM、SCM 等各种信息系统中. 方便起见,本研究把“产品制造系统”中能够产生第一数据的产品全生命周期中的设备、系统和流程等统称为“机器”,因此第一数据就是机器生成数据,记为 MGD(machine generated data).

员统称为“生产者”,因此第三数据就是由生产者即人生成的记载问题处理经验的数据,称为生产者生成数据,记为 PGD(producer generated data). 尽管不同的“生产者”的工作岗位不同,但是其共同特点是都会遇到问题、处理问题和解决问题,都会在问题处理过程中产生问题处理经验,也都会把大脑中的经验外化并由人主动记录为第三数据.

“第二数据”(Second Data)是产品在消费和使用过程中,由消费者/用户为主、社会大众参与而生成的社交数据,简称用户生成数据,记为 UGD(user generated data),可参考 UGC<sup>[11]</sup>的概念予以理解. 第二数据是对产品消费/使用的体验、期望、需求甚至情感的表达,也是对产品的评价和情绪的宣泄,还包括对产品使用、保管、维护的经验交流. 比如汽车论坛、电子商务平台上的数据.

表 1 是“第三数据”的真实表格,是由“生产者”对设备工装故障的处理记录,“故障”就是一种“问题”.

“第三数据”(Third Data)是“产品制造系统”在产品全生命周期各个环节各个岗位上的工作人员,把问题处理的经验表达出来而生成的一种数据. 本研究把所有岗位参与工作的人

这张表完整地记录了一个问题的处理过程,其中蕴含着问题处理的经验性知识. 其中,“故障发生的状况”是对故障状态的描述;“故障发生要因”记录了故障的所有可能原因;“故障发生真因”是分析后得到的真实原因;“2、处理方法”、“存在问题的改善对策”和“再发防止”是故障处理的临时措施和永久解决方案,还有处理结果的记录. 一张表格记录一个故障问题的处理情况. 在真实世界中,不同企业、不同岗位、不同生产者所采用的表格不尽相同,不同企业存储在不同的信息系统中.

表1 “第三数据”的实例——一个真实问题处理过程的记录  
Table 1 Example of “Third Data”: A record of a real problem handling process

## 设备工装故障分析报告

---

**1、故障发生的状况 <状况/现象>**

2011年10月25日冲压A线生产J56后门内板，共生产1133件，模具停台25分钟。主要问题：  
1. OP10制件皱，裂  
2. 模具OP30 OP40堵废料

**2、处理方法**

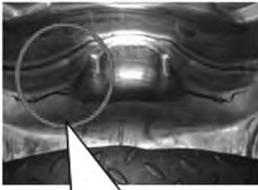
| 时间 | 问题 | 处理内容      |
|----|----|-----------|
| 14 | 1  | OP10制件裂调整 |
| 5  | 2  | OP30堵废料   |
| 6  | 3  | OP40堵废料   |

---

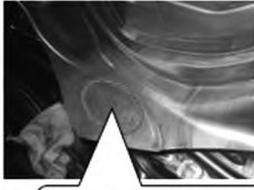
发生故障的详细情况 <系统/结构/组成>



OP10制件裂



OP10制件皱位置



OP20模具硬点

---

<故障发生要因(考虑所有要因)>

1. 模具压料面及拉伸走料慢，制件拉裂。
2. 模具压料面及拉伸筋拉毛造成，制件走料部均匀。
3. 涂油参数不合理。
4. 模具平衡块没配平制件局部走料控制不住。
5. 模具局部有硬点。

<故障发生真因> 1. 模具局部有硬点。2. 涂油参数不合理。3. 模具平衡块没配平

**3、存在问题的改善对策**

| NO | 存在的问题     | 对策实施内容        | 负责人 | 计划日期  | 完成日期 |
|----|-----------|---------------|-----|-------|------|
| 1  | 模具局部有硬点   | OP10模具进站修研压料面 |     | 10.27 |      |
| 2  | 模具顶料气缸动作慢 | 检查气缸          |     | 10.28 |      |

**4、再发防止/水平展开计划**

| 再发防止计划 |  |  |  |  |
|--------|--|--|--|--|
|        |  |  |  |  |

### 1.2 第三数据的特性与定义

与第一数据和第二数据相比,第三数据的主要特点:1)由“生产者”即“人”而非“机器”生成;2)对问题处理过程的记录;3)蕴含经验性知识。

由于上述特点衍生出一些特性,比如主观性、个体性、个别性、分散性、情境依赖性和路径依赖性等。主观性:第一、来源于经验的主观性,经验是人的认知结果,认知本身的主观性必然地被携带到第三数据中。第二、来源于对隐含于大脑中经验表达过程的主观性,表达的能力、方式方法因人而

异。第三、来源于个人意愿、自我保护和价值观等心理因素。每个“生产者”都是一个个体,每个个体都有自己的独特经验,因此记录经验的第三数据都是个体的、个别的,而且是分散的。

情境是指问题发生和处理的时间、地点、人物、条件甚至生产者当时的心境等主客观因素综合交叉所构成的“场景”。情境依赖性是指即使同样问题,由于情境不同所产生的经验和第三数据也不同。路径是指问题处理的过程“轨迹”。路径依赖性则是指即使是同一个问题由于处理过程的路径不同,则形成

的问题处理经验和第三数据也不同。

此外,第三数据还具有不确定性、不完备性、缺乏普适性等特性. 上述第三数据的特性使得对

第三数据的分析、挖掘和使用都有别于第一数据和第二数据,需要研究开发新的方法. 第三数据的特性归纳于表 2 所示.

表 2 “第三数据”与“第一数据”、“第二数据”的比较

Table 2 Comparison of “Third Data” with “First Data” and “Second Data”

| 名称          | 来源                  | 产生过程                            | 特性  | 内含                    | 表现形式     |
|-------------|---------------------|---------------------------------|---|-----------------------|----------|
| 第一数据<br>MGD | “机器”本质上是设备、系统和流程    | 1、产品全生命周期中产生; 2、自动/人工记录         | 客观性、纪实性、不允许掺杂个人主观意愿                         | 信息为主,极少量的规则性、关系性知识    | 结构化、半结构化 |
| 第二数据<br>UGD | 用户/消费者              | 1、使用/打算使用产品过程中; 2、在线发表          | 主、客观性混合掺杂、随意性、情感性、期望性、评价性                   | 信息、意愿、情感、情绪、价值观/偏好的混合 | 非结构化     |
| 第三数据<br>PGD | “生产者”本质上是工作在不同岗位上的人 | 1、在问题处理过程中产生经验; 2、经验拥有者对经验外化并记录 | 主观性、逻辑性、经验性、个体性、个别性、不确定性、不完备性、情境依赖性、路径依赖性等等 | 蕴含关于问题处理的经验性知识,密度最大   | 混杂、多样    |

根据第三数据的三个基本特点:问题处理、人、经验,定义如下.

**定义 1** 第三数据是在问题处理过程中,由“生产者”即人产生的蕴含问题处理经验的数据.

由于第三数据蕴含人的经验而且知识密度极大,所以属于知识型数据的一种类型.

## 2 第三数据的价值与意义

### 2.1 第三数据的价值——一个真实案例

冲压是汽车四大制造工艺的第一道工序. 一般希望生产线平稳且连续运行,但实际上既不平

稳也不连续,会产生正常和非正常两种停机现象. 前者是按计划安排的停机,如交接班、检修、改造或新车调试、节假日等. 后者是计划外停机,是生产中不希望出现的停机,是需要解决的“问题”,称为“停机问题”. 停机问题所浪费的时间称为“停机时间”. 图 2 是对国内多家汽车冲压生产线停机时间的平均统计结果. 其中,“工作时间”包括“正常停机”和“期望有效生产时间”两部分. 后者是指正常停机之外都应该生产合格产品的时间,其实不然,“实际有效生产时间”远小于“期望有效生产时间”,其差值就是“非正常停机时间”.

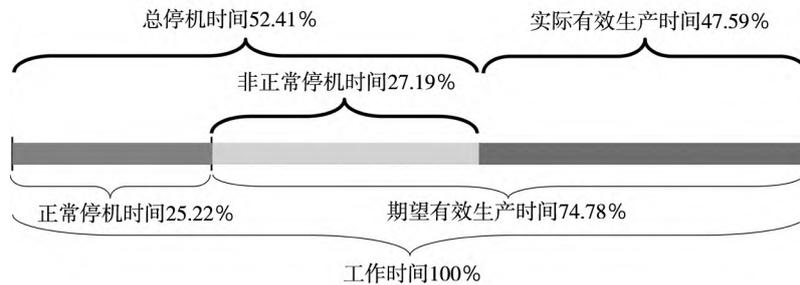


图 2 冲压线停机时间统计结果

Fig. 2 Statistical results of press line downtime

由图 2 可见,“非正常停机时间”占了工作时间的 27.19%,“实际有效生产时间”只有 47.59%,表明生产合格产品的有效时间还不到整个工作时间的二分之一. 停机时间是由停机问题造成

的,可通过解决“停机问题”来消除停机时间. “第三数据”正是为解决停机问题、消除停机时间而存在的数据,如果能够通过停机问题的处理,把其中占工作时间 23.795% 的停机时间消除掉,就相

当于无形中增加了一条生产线的生产能力. 这是因为现有的47.59%的“实际有效生产时间”是由两条生产线实现的,其中一条的平均有效生产时

间是  $47.59\% / 2 = 23.795\%$ . 如图3所示,因此,利用第三数据进行停机问题处理的效益是十分显著的.

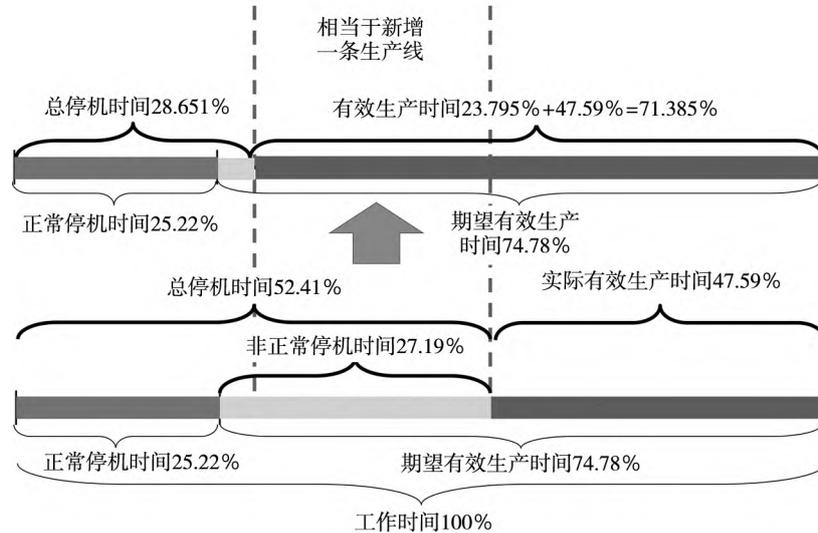


图3 停机问题管理的效益

Fig. 3 Benefits of downtime management

### 2.2 第三数据的研究意义

第三数据蕴含着问题处理经验,经验是知识的原始状态,与第一数据和第二数据相比,第三数据是一种知识密度最大而且知识类型最完备的数据. 因此:

第三数据是企业获取竞争优势的知识资源. 从资源秉性上看,第三数据是由人把经验外化出来而得到的蕴含知识的数据,是制造业企业创新发展、获取竞争优势的优质知识资源 (knowledge resource).

第三数据是人工智能的知识来源. 来源于第三数据的经验性知识不仅是人类智慧的底色,也是人工智能 (AI) 的根本,可以说没有知识就没有 AI.

第三数据是普遍存在的知识资源. “问题”的本质是矛盾,而矛盾具有普遍存在性<sup>[12]</sup>,因此“问题”也具有普遍存在性、“问题处理经验”也是普遍存在的. 事实上,在企业调研中发现,制造业企业中已广泛存在着不同形式的第三数据,已经是一种普遍存在的知识资源.

第三数据是改造世界和支持创新发展的知识资源. “问题”是需要人类主动处理的矛盾,矛盾

是一切事物发展的根本动力. 从这一点来讲,第三数据在本质上具有改造世界的作用,是可以支持创新发展的知识资源.

第三数据应用成果的可迁移性. 由于第三数据的普遍性,问题处理逻辑的通用性,可以认为在一个领域提出的方法具备迁移到相似领域的特质,对相似问题的处理具有参考作用,可以启发人们对相似领域的创新思考. 实践也证明了制造业中某一领域的第三数据研究成果可以迁移到相似领域.

## 3 第三数据的产生过程

### 3.1 从实践到经验

本研究所说的实践是指对问题处理的完整行为过程,经验则是指在问题处理过程中所产生的认知结果.

问题处理过程的一般逻辑如图4所示,包括五个逻辑步骤. 每一个逻辑步骤及整体过程都是多次反复的试错过程,即“猜测-试验-再猜测-再试验”直到结果满意为止,其中既有成功也有失败.

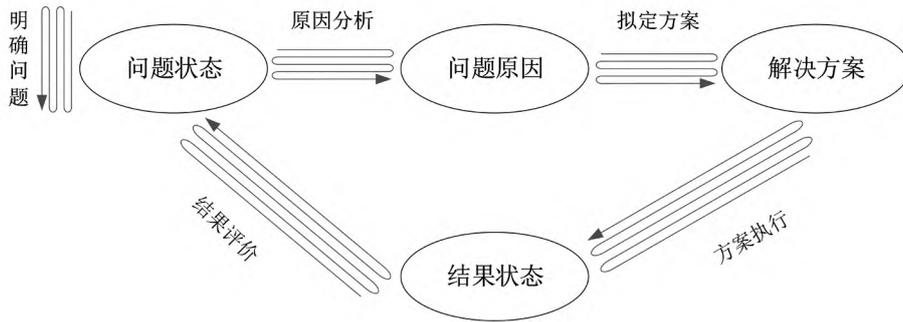


图 4 问题处理过程

Fig. 4 Problem handling process

为了对问题处理行为进行描述,定义如下.

**定义 2** 问题处理事实 (problem handling facts, PHF),是在问题处理过程中真实发生的、完整的行为路径.

PHF 是对问题处理“实践”这种行为过程的

概念化,可用一条“行为轨迹”来描述.试错的行为因人而异、因情境而异,其“轨迹”即 PHF 也不尽相同.图 5 表示处理同一个问题的三个不同轨迹,即三个 PHF,纵向虚线自上而下对应时间从小到大,表示问题的处理进程.

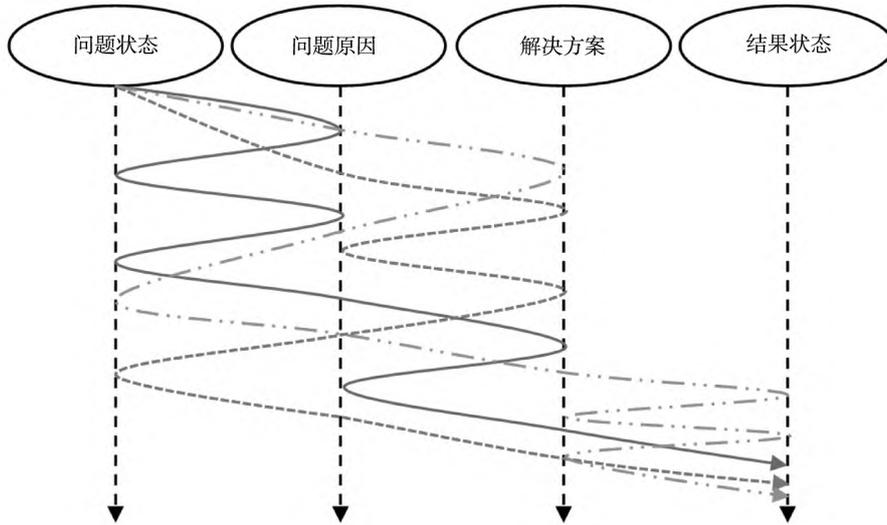


图 5 同一个问题的三个 PHF——三条问题处理行为轨迹

Fig. 5 Three PHFs for the same problem: Three trajectories of problem handling

每一个行为过程即 PHF,一定伴随着生产者的认知过程,图 5 表示三个认知过程.行为是显性活动,认知是隐性活动,两者不可分,只要生产者亲身参与问题解决的实践,就一定自觉不自觉地在大脑中留下认知结果.

**定义 3** 问题处理经验 (problem handling experience, PHE),是指伴随着问题处理事实 PHF,在生产者大脑中形成的隐性的、不可直接观测的认知结果.

每一个问题处理的 PHF 都对应一条 PHE,每一条 PHE 都包括三种经验:状态及其关系经验 (experience with states and their relationships, E-

SR);情境依赖经验 (experience with situation dependence, E-SD);分析思路经验 (experience in analyzing processes, E-AP). 每一条经验可以表示为

$$PHE = \{E-SR, E-SD, E-AP\} \quad (1)$$

### 3.1.1 状态及关系经验 E-SR

状态是指问题、原因和方案是“什么样”的表现形式,状态之间的关系就是表现形式之间的关系,E-SR 包括

$$E-SR = \{know-States, know-Relationships\} \quad (2)$$

其中 know-States = { know-what, know-why, know-how }.

且 Know-Relationships = { know-cause and problem relationship, know-solutions and causes relationship, know- solutions and results relationship, know-who }.

即“问题与原因的关系”、“原因与方案的关系”、“方案与结果的关系”以及对这些“关系”、“谁知道”的知识。

3.1.2 情境依赖经验 E-SD

图4所示的问题处理过程总是在特定情境中发生的,因此经验具有“情境依赖性”(situation dependence)<sup>[13]</sup>,把依赖于问题情境(situation)的经验记为

$$E-SD = \{ \text{know-when, know-where, know-conditions} \} \quad (3)$$

每个问题情境都是具体的,所以情境依赖的经验不具有一般性和普适性。

3.1.3 分析思路经验 E-AP

问题处理事实 PHF 反映问题处理的真实试错过程,其中即有成功也有失败,且反复次数、步长以及每次反复结果和投入资源均有不同,而且试错过程具有“路径依赖性”(path dependence)<sup>[14]</sup>,因此

即使相同问题也会产生不同的 PHF. 由于 PHF 中的成败步骤反映了问题处理的“过程经验”. 经过加工就可以在大量的 PHF 基础上提炼出问题处理的最佳思路,记为

$$E-AP = \{ \text{know-thinking process} \} \quad (4)$$

3.2 从经验到数据

从知识管理的角度可以把 PHE 称为隐性知识<sup>[15]</sup>,为了便于知识共享,需要把 PHE 从大脑中“拿出来”,变为别人可见的问题处理数据。

定义4 问题处理数据(problem handling data, PHD),是指对问题处理事实 PHF 进行符号化描述的数据,其中蕴含着问题处理经验 PHE.

PHD 具有“第三数据”的三个基本特点,即,问题处理、人和经验,如定义1所示,因此 PHD 就是本文所说的“第三数据”,在制造业中就是“生产者生成数据”PGD,是由人产生的数据. 对表1这类实际数据进行抽象概括可得到表3的逻辑结构,表头的结构顺序反映了问题处理的逻辑顺序. 表3的每一行都记录试错过程每一步骤的状态,无论对和错. 从 step -1 到 step -n 的 n 行反映了问题处理的整个过程。

表3 PHD 即第三数据的逻辑结构表  
Table 3 PHD that is logical structure of “Third Data”

| 试错步骤    | 问题状态  | 问题原因  | 解决方案  | 结果状态  | 情境条件 |
|---------|-------|-------|-------|-------|------|
| Step -1 |       |       |       |       |      |
| Step -2 |       |       |       |       |      |
| .....   | ..... | ..... | ..... | ..... |      |
| Step -n |       |       |       |       |      |

类似表3的每一张表只反映一个问题处理事实 PHF,多个问题则需要多个数据表. 依此,可以建立第三数据的数据库。

表1中的“故障发生的状况 < 状况/现象 >”和“发生故障的详细情况(系统/结构/组成)”对应表3中的“问题状态”,其描述方式有文字、字符和图像;“故障发生要因(考虑所有要因)”、“故障发生真因”对应表3的“问题原因”;“处理方法”、“存在问题的改善对策”和“再发防止/水平展开计划”对应表3中的解决方案;“2011年10月25日”等是时间因素、“冲压A线生产后门内板”等是生产地点和问题产品上的位置,还有相关的负责人和其他条件因素等情境因素. 显然,表1并不像表3那样完备,特别是表1中缺乏对试错步骤的记录。

4 第三数据的研究问题

第三数据是大数据研究中的一个新视角、新对象和新领域,若使第三数据产生价值,除上述从实践到经验再到数据的两个过程外,还需要三个过程才能构成完整、可行的知识生产和价值生成机制,即从数据到模型,从模型到知识,从知识再到实践,完整过程如图6所示,本文称为“五星螺旋模型”. 其中“问题处理实践”是存在于真实世界中的客观事实,“问题处理经验”是存在于精神世界中的主观认知,“问题处理数据”是存在于符号世界中的第三数据,“问题处理模型”是存在于符号世界中的知识表示模式,“问题处理知识”是内化于精神世界中的模型内涵和数据内涵,只有

构建这样的螺旋通道,才能使第三数据中蕴含的经验性知识为新的实践提供支持.否则只能在实践与经验的短链条中低效循环,如图6实践与经验之间的双向箭头所示.根据图6的五星螺旋模型列举十个相关研究课题.以下省略“问题处理”四个字,简称实践、经验、数据、模型和知识.

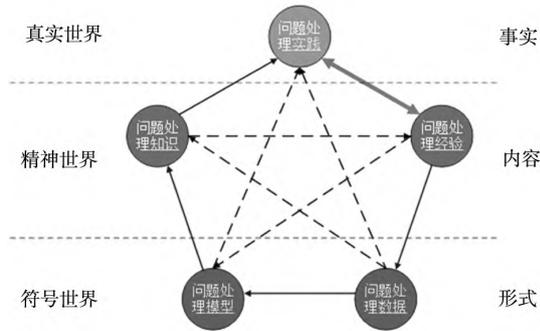


图6 第三数据的五星螺旋模型

Fig.6 Five-star spiral model for “Third Data”

第一, 第三数据价值的产生机制研究.

图6中的双向箭头表示“实践-经验”的互动,是短链条的“自然过程”,其缺点是难以知识交流、存储和共享且价值效率低下.然而,借助符号世界中的数据 and 模型可以人为构造出图6中的全部有向边组合成的多个不同的知识螺旋<sup>[16, 17]</sup>,比如“实践-经验-数据-实践”,“实践-经验-数据-模型-实践”等等,每一个螺旋都是实际应用场景背后的一种逻辑,虽然价值效率各不相同但都超越了上述的“自然过程”.需要结合实际场景深入研究每一个知识螺旋的过程机制及实现路径与方法.显然,第三数据是每一个知识螺旋的“关键结点”,不可或缺,否则每个知识螺旋的价值都不能形成.

第二, 问题处理效益的提高方法研究.

从知识螺旋的源头来看,问题处理是一个试错过程,如何利用已有的经验和知识减少无用功,提高试错过程的有效性,同时产生高质量的新经验,需要对经验产生机制、方法进行研究.

第三, 经验变为第三数据的外化机制研究.

经验外化是指把隐藏于大脑中的经验变为可以交流和共享的数据,外化过程受到个人心理、价值因素的直接影 响.如何使每一位“生产者”都能自愿地把个人的宝贵经验外化出来?这就需要对外化的激励机制、过程、方法和工具以及能够蕴含

经验的最佳数据结构进行研究.

第四, 模型设计与第三数据的建模研究.

模型的内涵是问题处理的知识,是对大量经验的汇集和升华.因此需要根据第三数据的特点和特性研究有利于知识表示、推理计算和更新进化的模型;同时还要研究第三数据的知识挖掘方法、建模方法以及软件工具和技术.

第五, 知识及经验内化的学习机制研究.

模型中蕴含的知识、数据中蕴含的经验需要通过学习内化为知识,才能为价值产生提供基础,因此需要学习理论、学习方法和学习机制进行研究.

第六, 第三数据的数据治理研究.

当前,制造企业中已经存在着大量的第三数据,但是处于一种混杂状态,为了提高第三数据的规范化、科学化、完备化和系统化程度以及提高形成和采集的效率,需要开展第三数据的数据治理 (data governance)<sup>[18]</sup>研究,更好地促使第三数据从自发产生向自觉产生过渡.

第七, 第三数据与知识管理的关系研究.

图6的第三数据的五星螺旋模型,与野中郁次郎的SECI<sup>[19]</sup>知识螺旋模型在内涵上有许多相似之处,但是又有不同.因此有必要对五星螺旋模型与SECI知识螺旋之间的关系开展研究,比如隐性知识与经验的关系、显性知识与第三数据的关系以及知识螺旋的动态机制等.

第八, 第三数据与人工智能的关系研究.

知识的“用户”除了人之外,还有智能机器,当智能机器加入到图1的“产品制造系统”时,智能机器也变成了“生产者”,因此第三数据既为“人”所用,也为“智能机器”所用.如此,智能机器是否也能产生“经验”,也能生成第三数据,也能通过“学习”内化为知识,这也是值得深入思考和需要研究的问题.

第九, 第三数据与第一数据和第二数据的关系研究.

虽然第三数据与第一数据、第二数据具有很多不同特性,但是都与图1中的“产品制造系统”有关,客观上三者之间必然存在一定的联系,从不同的侧面反映了“产品制造系统”的整体情况,第三数据中蕴含的经验性知识如何与第二数据中情感性知识、第一数据中的规则性知识相互融合,将会为企业产生更全面更完整的知识,一定会给企

业带来更大的价值<sup>[20]</sup>. 这是更全面、更深刻的“知识系统工程”<sup>[21]</sup>的研究课题,需要对第三数据与第一数据、第二数据融合的知识系统的体系架构、融合方法、模型、算法和工具展开广泛而深入的研究.

第十,第三数据的普适性研究.

由于“问题”的本质是矛盾,矛盾具有普遍性,因此“问题”也具有普遍性,即无论是制造业领域,还是服务业领域,甚至社会经济领域等各种领域都广泛地存在着矛盾和问题;再加之问题处理逻辑的通用性,使得第三数据在广泛领域中具有普遍性.但是,在推广到服务业和社会经济等更为广泛的领域时,需要结合各自领域的特点开展领域性的深入研究,比如人生病的“病”就是需要治疗的“问题”,医案就是第三数据.在社会经济系统乃至管理领域中的案例文本,其本质上也是关于问题处理过程经验的记录,也属于第三数据.因此,关于第三数据的概念、理论、方法和工具等研究成果具有一定的普适性.除了针对“产品制造系统”开展研究之外,在社会经济等广泛的领

域也存在着大量需要深入研究的课题<sup>[22]</sup>.

## 5 结束语

第三数据是当前大数据热潮中被冷落的一种新的数据类型,本研究分析了第三数据的特性及其与第一数据、第二数据的差别,阐述了第三数据独特的应用价值;给出了问题处理事实、问题处理经验、问题处理数据、问题处理模型、问题处理知识的概念,并结合真实世界、精神世界、符号世界的相互转换关系构造了基于第三数据的五星螺旋模型(图6);根据第三数据的特性和五星螺旋模型提出了十个方面值得深入思考的研究课题.

本文的研究成果已经在某汽车制造企业应用了十几年,实践证明第三数据既有理论研究价值也有实际应用价值.希望不同领域的广大学者和企业界人士对“第三数据”引起重视,在深入实际应用的同时,开展“第三数据”的理论方法研究,构建并完善理论架构,在大数据研究领域促进大数据和知识管理理论的创新发展.

## 参考文献:

- [1]周 济. 智能制造——“中国制造 2025”的主攻方向[J]. 中国机械工程, 2015, 26(17): 2273 - 2284.  
Zhou Ji. Intelligent manufacturing: Main direction of “Made in China 2025”[J]. China Mechanical Engineering, 2015, 26(17): 2273 - 2284. (in Chinese)
- [2]张 维, 曾大军, 李一军, 等. 混合智能管理系统理论与方法研究[J]. 管理科学学报, 2021, 24(8): 10 - 17.  
Zhang Wei, Zeng Dajun, Li Yijun, et al. Hybrid intelligence management system research: Theory and methods[J]. Journal of Management Sciences in China, 2021, 24(8): 10 - 17. (in Chinese)
- [3]Zhou J, Li P, Zhou Y, et al. Toward new-generation intelligent manufacturing[J]. Engineering, 2018, 4(1): 11 - 20.
- [4]曾大军, 张 柱, 梁嘉琦, 等. 机器行为与人机协同决策理论和方法[J]. 管理科学学报, 2021, 34(6): 55 - 59.  
Zeng Dajun, Zhang Zhu, Liang Jiaqi, et al. Machine behavior and human-machine collaboration decision: Theory and methods[J]. Journal of Management Sciences in China, 2021, 34(6): 55 - 59. (in Chinese)
- [5]李雨桐, 党延忠. 基于交互记忆系统的团队发展状况分析方法[J]. 管理科学学报, 2016, 19(4): 16 - 31.  
Li Yutong, Dang Yanzhong. Analytical method of team development status based on transactive memory system[J]. Journal of Management Sciences in China, 2016, 19(4): 16 - 31. (in Chinese)
- [6]Von Rueden L, Mayer S, Beckh K, et al. Informed machine learning: A taxonomy and survey of integrating prior knowledge into learning systems[J]. IEEE Transactions on Knowledge and Data Engineering, 2023, 35(1): 614 - 633.
- [7]Karniadakis G, Kevrekidis I, Lu L, et al. Physics-informed machine learning[J]. Nature Reviews Physics, 2021, 3(6): 422 - 440.
- [8]Rathore A, Ilavarasan P. Pre-and post-launch emotions in new product development: Insights from twitter analytics of three products[J]. International Journal of Information Management, 2020, 50: 111 - 127.
- [9]Rajpathak D. An ontology based text mining system for knowledge discovery from the diagnosis data in the automotive domain [J]. Computers in Industry, 2013, 64(5): 565 - 580.
- [10]Xu Z, Dang Y. Automated digital cause-and-effect diagrams to assist causal analysis in problem-solving: A data-driven approach[J]. International Journal of Production Research, 2020, 58(17): 5359 - 5379.

- [11]尹 鹏, 丁栋虹, 豆国威. 视频平台用户生成内容投资和定价决策[J]. 管理科学学报, 2020, 23(10): 116 – 126.  
Yin Peng, Ding Donghong, Dou Guowei. User generated content investment and pricing decisions in online video markets [J]. Journal of Management Sciences in China, 2020, 23(10): 116 – 126. (in Chinese)
- [12]Zhao H K, Liu X P, Zhang X, et al. The effects of person-organization fit on lending behaviors: Empirical evidence from Kiva[J]. Journal of Management Science and Engineering, 2022, 7(1): 133 – 145.
- [13]王众托. 管理是科学还是艺术——从知识层面认识管理[J]. 中国科学院院刊, 2007, 22(1): 39 – 44.  
Wang Zhongtuo. Is management a science or an art: Understanding management from knowledge perspective[J]. Bulletin of Chinese Academy of Sciences, 2007, 22(1): 39 – 44. (in Chinese)
- [14]Carlile P, Reberntsch E. Into the black box: The knowledge transformation cycle[J]. Management Science, 2003, 49(9): 1180 – 1195.
- [15]Howells J. Tacit knowledge[J]. Technology Analysis & Strategic Management, 1996, 8(2): 91 – 106.
- [16]王志平, 王众托. 超网络理论及其应用[M]. 北京: 科学出版社, 2008.  
Wang Zhiping, Wang Zhongtuo. Hypernetwork Theory and its Applications[M]. Beijing: Science Press, 2008. (in Chinese)
- [17]Fang J, Liu Q, Tang M, et al. Network science faces the challenge and opportunity: Exploring“network of networks”and its unified theoretical framework[J]. Journal of Applied Analysis & Computation, 2016, 6(1): 12 – 29.
- [18]Abraham R, Schneider J, Vom Brocke J. Data governance: A conceptual framework, structured review, and research agenda[J]. International Journal of Information Management, 2019, 49: 424 – 438.
- [19]Nonaka I, Takeuchi H. The Knowledge-creating Company: How Japanese Companies Create the Dynamics of Innovation [M]. New York: Oxford University Press, 1995.
- [20]Escobar C, MCGovern M, Morales-Menendez R. Quality 4.0: A review of big data challenges in manufacturing[J]. Journal of Intelligent Manufacturing, 2021, 8(32): 2319 – 2334.
- [21]王众托. 知识系统工程(第二版)[M]. 北京: 科学出版社, 2015.  
Wang Zhongtuo. Knowledge Systems Engineering (Second Edition)[M]. Beijing: Science Press, 2015. (in Chinese)
- [22]Xu Y J, Li S P, Xiong X, et al. Intraday volatility spillover between the Shanghai and Hong Kong stock markets: Evidence from A + H shares after the launch of the Shanghai-Hong Kong stock connect[J]. Journal of Management Science and Engineering, 2017, 2(4): 290 – 317.

## The neglected “Third Data” in the big data boom: Taking the manufacturing industry as an example

DANG Yan-zhong<sup>\*</sup>, HU De-qiang<sup>\*</sup>, XU Zhao-guang<sup>\*</sup>

Institute of Systems Engineering, Dalian University of Technology, Dalian 116024, China

**Abstract:** At present, there is a very important and valuable data that has been neglected in the surging tide of big data research, which is the “Third Data” proposed in this paper. The Third Data is the data with subjective characteristics generated by problem handlers in the process of problem handling. The Third Data contains the empirical knowledge of problem handling with a great density, and is an indispensable knowledge resource for problem definition, problem analysis and problem solving. This paper defines the concept of the Third Data, and discusses the characteristics and sources of the Third Data. The generating process of the Third Data is analyzed through “practice-experience-data” and a five-star spiral model of Third Data is proposed. Actual case analysis shows that the exploitation and utilization of Third Data have significant value in reducing the cost and improving the benefit. This paper also puts forward ten topics related to the in-depth study of Third Data. The Third Data has the inherent property of transforming the world, as well as a problem-driven universality. Each aspect contains a problem related to Third Data in some way. Therefore, the theoretical results of Third Data have a more general and universal meaning, applicable not only to manufacturing industry but also to services, even in the social and economic fields.

**Key words:** big data; Third Data; intelligent manufacturing