# 第 3 章

# 大数据思维——革故鼎新

思维作为人类认识世界的理性形式，具有一种排斥非理性因素的内在禀赋。本书作者于 2017 年出版的《互联网+　思维与创新》一书中对“数据思维”已进行了思考与解读。在此基础之上，本章从哲学思维、赋能思维、模型思维和计算思维四个方面对大数据思维进行探究，以期指导大家理解和运用大数据技术，有助于拓展对数据价值的认识，从而发现大数据、关注大数据、管好大数据。

大数据思维导图如图 3-1 所示。

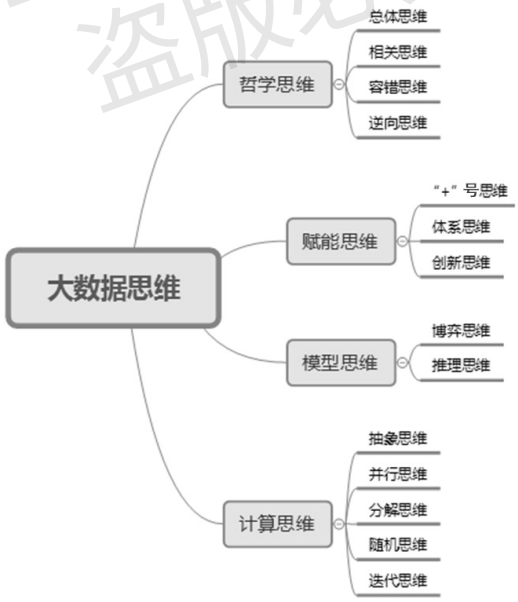


图 3-1　大数据思维导图

# 3.1 哲学思维

大数据是相对小数据而言的。人们接触到的信息或数据发生了变化，其思维方式也要进行相应的调整。在大数据时代，人们看待数据的思维方式，最关键的转变在于从自然科学思维转向哲学思维，由以往使用单一数据向使用全体数据转变，由重视数据的精确性向挖掘数据的内在价值、侧重数据的混杂性转变，由注重数据的因果性向利用数据的相关性转变。大数据时代的思维方式变革如图 3-2 所示。

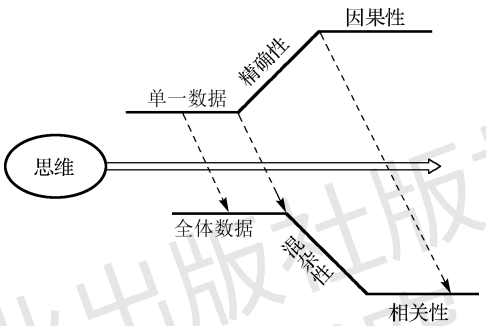


图 3-2　大数据时代的思维方式变革

## 3.1.1 总体思维

个体是总体的一个实例，人们观察到的个体是个别显现出来的事物的内在本质，但不包含事物的全部本质，要想全面认识事物，就必须采用总体思维。

在小数据时代，由于受技术发展水平的制约和数据采集、处理能力的限制，人们无法获得全部数据信息。在大数据时代，数据采集、存储、分析等技术获得了突破性进展，数据总量迅速增长，人们获取全部数据信息变为现实，不再受数据不足的限制，可以更加便捷地获取与研究对象有关的一切数据。此外，大数据强调数据的复杂性和完整性，可以进一步揭示事情的真相。因此，人们看待数据的思维方式逐渐转向总体思维，从而全面客观地认识事物的特征和机理。

例如，早期商场记录顾客消费数据只能通过销售额、会员卡等，数据相对滞后也不够准确，不便随时调整自己的销售策略，尤其是无法实时监测客流量。而如今，商场借助客流计数器、手机定位等技术，可以对客流量进行实时追踪，结

合商场内部的管理系统和银联实时反馈的商品销售额，可以实时获取顾客的全部消费数据，并依据这些数据调整商场的经营管理策略，提升商场的运营效率。例如，如果发现当天有爆款商品存在缺货现象，可以及时向上游公司订货来补充货源，确保当天的顾客可以买到自己喜欢的商品，提升顾客体验。

### 3.1.2　相关思维

事物之间存在关联。发现其相互作用的规律是人们探究世界的重要途径，亦是更高层次的认知需求。例如，在医学领域，在发现病人颈椎膨出可能压迫神经引发头疼后，医生应综合考虑头部和其他间接引发头疼的部位的病因关联机理，对症下药。

传统的因果思维主要关注事物之间的逻辑推理，通常采用的研究方法分为三步：一是假定事物之间存在某种关系，分析产生某个结果的可能原因；二是建立原因和结果的共变关系，根据原因与结果的时序关系来消除非确定性的因素；三是最终通过推断验证产生该结果的原因。

因果关系只是事物之间关系的一种，事物之间的相关关系普遍存在。因果关系 A→B 表示产生 B 的原因是 A，但有可能是 C 导致了 A 和 B 的发生，此时 A 和 B 之间不再是因果关系，而是相关关系。例如，血压升高可能引发头疼，但颈椎膨出会同时引发血压升高和头疼，因此血压升高和头疼之间是相关关系。

相关关系指当某个或一批变量变化时，与之对应的一些变量按照某一规律在一定范围内变化，这一规律就是这些变量之间的相关关系，也称相关系数。根据相关程度，相关关系可以分为完全相关、不相关和部分相关。传统相关系数计算可采用皮尔森相关系数、斯皮尔曼等级相关系数等方法。大数据的非线性和高维性催生了基于互信息的相关系数、基于矩阵计算的相关系数、基于距离的相关系数等新方法。

在大数据时代，数据分析在一定条件下不再刻意寻找因果关系，而是更注重数据之间的相关关系。从大量的数据中挖掘复杂多样的相关关系，从而揭示事物之间深层次的内在机理。

### 3.1.3　容错思维

在小数据时代，由于样本数量相对较少，所以必须确保样本数据的精确性，

否则由通用模型分析得出的结果就会出现偏差。例如，在显微镜下观察物体时，为了确保测量结果的精确性，对采样的要求十分苛刻，如果采样结果不够精确，就会影响测量结果的精确性。

当可用数据愈加丰富时，单一数据是否绝对精确不再是一个大问题。人们的思维方式从精确思维转变为容错思维，不再苛求个别数据，容许适量误差，关注整体数据质量，全面认识事物。例如，顾客会存在代购商品的行为，致使商场很难精准了解每名顾客的消费习惯，但可通过对全部顾客消费习惯的分析，优化柜台布局、人员配置和营销策略。

### 3.1.4 逆向思维

事物有两面性，人们容易看到其熟悉的一面，忽视其陌生的一面。逆向思维“反其道而思之”，可以取得出乎意料的结果。例如，某自助餐厅生意火爆，但顾客浪费严重，为了防止顾客浪费食物，该餐厅规定，凡是浪费食物者罚款 10 元，结果生意一落千丈；后来餐厅经营者转变思维方式，规定凡是没有浪费食物者奖励 10 元，结果生意重新火爆起来，并且杜绝了浪费行为。这就是采用了逆向思维，从处罚转变为奖励，让顾客认为自己占到了便宜。

逆向思维是一种对常见或结论性事物反过来思考的方式。它是对传统思维的挑战，能够突破传统思维、习俗思维和常识思维的刻板印象，改变由经验和习俗造成的僵化的认知模式。

若想在竞争日益激烈的大数据时代有所建树，就不要一味盲目跟随，必须运用逆向思维，发现其他人发现不了的闪光点，将冷门变成热点，将不可能变成可能。

## 3.2 赋能思维

赋能被广泛应用于商业管理中，强调注重信息共享，突破组织管理“深井”，在沟通上透明，在决策上去中心化，赋予员工自主工作的权利，激发整个系统内在的活力。在“互联网+”高速发展的基础上，大数据继承了“互联网+”的倍增、加速、提升等特性，因此“+”号思维是赋能思维的基础；大数据关注宏观，与体

系思维的出发点是一致的，因此体系思维是赋能思维的支撑；创新是赋能的有效手段，因此创新思维是赋能思维的方式。

### 3.2.1 “+”号思维

“+”号的本义是用来表示整数中的正数或加法运算，后来这个符号被赋予了丰富的抽象意义。“+”号思维更多是从倍增、加速、提升的角度阐释大数据的赋能思维。

“倍增”：顾名思义就是成倍地增加或增长。运用大数据赋能，可对传统行业起到倍增器的作用，达到 1+1>2 的效果。例如，健康医疗大数据、金融大数据已展现出了强大的发展前景。运用大数据技术，通过嫁接其他行业的价值对企业进行创新，提升产业转型升级能力，加快构建新型产业体系，引领行业新发展。

“加速”：大数据赋能后，效益或效能得到了指数级增长。数据量剧增，间接刺激了运用数据能力的提升，极大地提高了决策的精度和速度。例如，“让数据多跑路，让群众少跑腿”，通过构建创新型政务服务体系，实现扁平化的管理模式，缩短办事流程，提升政务服务效率。

“提升”：大数据赋能就是要让数据的价值最大化。对收集到的数据进行严密且富有逻辑的整理、分析、关联，发掘出具有价值和意义的信息。例如，基于交通大数据，导航软件可以规划出最快到达目的地的路线，从而节省时间。

### 3.2.2　体系思维

大数据思维可以从诸多事物并存运行、相互协同的条件中探究其本质。这种从宏观视角全面系统地认识事物的思维方式称为体系思维，也可理解为为了实现某个目标而采取的一系列方法的组合。

体系思维是整体的。既要全面认识事物，还要分步骤解决问题。在复杂系统中，很多事情难以找到原因，但是通过相关性来发现问题，把握宏观规律，就可以找到解决问题的方法和策略。因此，重视事物的普遍联系与相互作用，强调系统论的视角，综合全面地思考、处理问题，是体系思维的重要体现。

体系思维是成长的。所谓的成长体现在演化和进步的过程中，在这个过程中，一些功能会被抛弃或退化，一些功能会变得越来越强大，进而形成新的体系结构。

体系思维是全样本的。全体数据将取代样本数据，表现出部分数据所不能表现出的细节。例如，尽可能将物价指数、营商指数和其他经济指标进行全样本收集与分析，构建经济模型，预测未来经济走向。

### 3.2.3 创新思维

在运用大数据时，需要突破常规思维的障碍，提出新颖的解决方案，从而产生意想不到的结果。大数据的创新思维可以从跨界、智慧等方面去认识。

跨界思维就是要敢于打破常规，从表面上没有关系的其他领域，寻找内在的相关性突破口。例如，在社会安全领域，警察可以使用日常生活的多种数据来识别各种犯罪活动。根据情报，某地区有部分人员私自加工管制刀具，如果采取逐户检查的常规方法，任务量大、耗时长，很难完成任务。由于加工管制刀具使用的机械设备具有独特的用电特征，当地警察通过分析该地区用户的用电量波形来及时发现犯罪团伙。

为什么人类的大脑具有智慧？究其原因是它可以进行智能研判和科学决策。大数据思维使大数据像人脑一样具有生命力。2016 年 3 月，围棋人工智能程序 AlphaGo 以总比分 4∶1 轻松取胜棋手李世石；2017 年 10 月，AlphaGo Zero 在与 AlphaGo 的对阵中取得了 100∶0 的战绩；2019 年 1 月，在《星际争霸 2》项目中，AlphaStar 5∶0 战胜职业选手 TLO，5∶0 战胜 2018 年 WSC 奥斯汀站亚军 MaNa 。这些人工智能程序的“智慧”来源于大量对抗数据和自我对弈训练数据。运用深度学习方法，经过多次训练，这些人工智能程序不仅掌握了对弈知识，而且在对抗过程中，还可以做到创造新的招式。当然，这还只是一个游戏程序，离人类的智能还有很大的距离，但是它让我们看到了运用数据和工具，机器可以达到甚至超越人类的某方面智能。云计算、机器学习等技术的发展，正在大力推动大数据向经济社会各领域、各行业渗透，一个巨大的智慧网络将改变大众的生活和人类的未来。

## 3.3 模型思维

人们基于自己的知识体系来解决工作、学习和生活中的各种问题。同样的信息，由于受众方的知识体系不同，采取的评价模型不同，可能得出大相径庭的结

果。模型是大数据时代的动力，其更易实现大数据时代的事件挖掘、用户行为分析。

### 3.3.1　博弈思维

企业在管理数据的时候，只存储而不分析和应用数据，意味着存储成本的不断增加。这时可以使用一些低价的存储方案来存储使用量较少的数据，以减少存储成本。但是，如果数据继续呈指数级增加，那么存储成本也会增加。以前，由于数据量小，这一问题造成的影响还不太明显。现在，存储和使用成为很多企业在决策中不得不面对的问题。如果不大量存储数据，企业就可能因为没有存储需要的数据而错失发展良机；如果大量存储数据而不能应用，也会给企业的运营成本带来压力。这就是企业数据管理面临的“博弈”。

要解决上述企业面临的问题，就需要用到博弈思维。在运用博弈思维时，首先要确定博弈的参与者及其行动，然后建立相关的博弈模型，如标准博弈模型、序贯博弈模型、合作博弈模型等。

在大数据时代，企业之间的合作和竞争更加频繁，企业之间既是合作关系又是竞争关系，此时采取博弈思维来指导企业战略决策是一个不错的选择，可参考其中的合作博弈模型。假设有 $N$ 个相关企业，它们之间可以是竞争关系，也可以是合作关系。在某次经济活动中，合作者要承担的合作成本为 $C$，其他企业可以获得的相关收益为 $B$。如果企业之间是竞争关系，则不会产生任何成本和收益。而企业合作的潜在收益可以通过合作优势比率 $B/C$ 来衡量。这就是简单的合作博弈模型。

### 3.3.2　推理思维

在大数据时代，全数据集中的数据元素是繁多的，难以建立所有数据元素之间的关系，因为数据元素之间的关系随着数据元素数目的变化呈指数级变化。这时可以采用推理思维，基于构建的某种基本知识，通过现存的数据元素之间的关系推理出未知的数据元素之间的关系。例如，如果数据元素 a 和 b 之间是夫妻关系，而数据元素 a 和 c 之间是父子关系，根据人类的家庭关系知识，可以推理数据元素 b 和 c 之间是母子关系。现有的推理模型有基于逻辑的推理模型、基于图

的推理模型和基于表示的推理模型。

在推理思维中，一般先将数据元素知识化，即用某种形式化的语言表示相应的数据元素，如逻辑元素、图节点等。这种表示方法就是符号化。符号化的缺点是难以进行计算，推理的复杂程度较高。因此，基于表示的推理模型采用向量来对数据元素进行知识化表示，将对未知的推理转化为对相关向量的操作。对向量的操作计算模型有距离模型、翻译模型和矩阵模型。距离模型用向量的距离表示数据元素之间的某种关系。而翻译模型则建立了向量之间的语义关系，将三元组 <*h*,*r*,*t*>分别用向量表示，如果该三元组成立，则 *h*、*r*、*t* 在向量空间上会有如图 3-3 所示的关系。矩阵模型则根据向量中的数据构建与其所对应的矩阵，通过对矩阵的操作进行学习推理。典型的矩阵方法是 Nickel 等提出的 RESCAL。

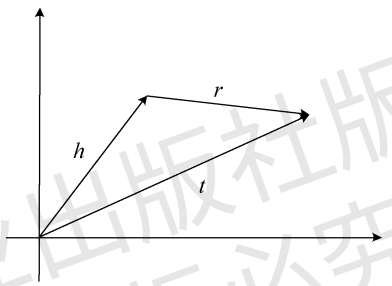


图 3-3　翻译模型示例



## 3.4　计算思维

计算思维是运用计算机科学领域的相关基础概念进行问题求解的过程，通过算法的思想来识别、分析问题，形成适合用计算机来处理问题的解决方案。

### 3.4.1　抽象思维

如果要将鸡、苹果、香蕉、鸭分成两类，大家都会将鸡和鸭分成一类，将苹果和香蕉分成一类，因为前者是“家禽”，后者是“水果”。用家禽概括鸡和鸭反映了人们的抽象思维方式。抽象思维是人们在认识、处理事物的过程中用概念的形式，对客观事物进行间接总结和概括的过程。抽象思维通过科学合理的概念抽象来获取事物的本质，其核心思想就是去除多余和细节，保留关键要素。

在运用抽象思维时，首先从要解决的问题中抽取关键要素，然后根据这些信息来筛选数据。抽取的关键要素就构成了解决该问题的一个概念模型。最终抽取的关键要素取决于解决的具体问题、应用的具体场景等。例如，都是导航服务，地铁导航服务和自驾导航服务所抽取的关键要素就不同。在地铁导航服务中，抽取的关键要素是地铁站点信息、站点换乘信息，而对站点之间的具体距离等物理环境信息不做过多要求；而在自驾导航服务中，则要关注道路的具体宽度、路口距离、红绿灯等物理环境信息，从而根据用户不同要求规划不同路线。

抽象思维正确与否一定要经过社会实践的检验，只有经过验证的抽象思维才是正确的。抽象思维指导人们认识并开展相关的社会实践，同时社会实践的结果又可以巩固并拓展抽象思维。空洞、不切实际的抽象思维是毫无意义的，而科学的抽象思维一定是建立在充分的社会实践基础上的。

## 3.4.2　并行思维

在现实中，我们面对的问题大多数是串行的，但许多问题其实可以并行处理。学者洪加威就并行处理曾讲述了“证比求易算法”童话。故事讲的是某国的国王向邻国的公主求婚，公主问了一个问题，如果国王在一天之内能够求出 48770428433377171 的一个因子，她就接受国王的求婚。可是国王试了 3 万多个数，还是没有得到结果。国王恳求公主告诉他答案，公主说，223092827 就是一个因子。国王请求公主再给他一次机会，公主答应了。为了把握住这次来之不易的机会，国王向宰相求教。宰相认真思考后认为这个数是 17 位，如果这个数存在因子，那么最小的一个因子不会超过 9 位。于是，宰相向国王建议：按照自然数的顺序发给全国的每个百姓一个编号，等公主给出数后，立即将这个数通报全国，每个百姓用这个数除自己对应的编号，如果没有余数，则立刻将自己的编号上报国王。果然，国王很快就求婚成功了。

在这个故事中，国王使用的是串行方法，宰相提出的是并行方法。显然，在处理这个问题上，并行方法比串行方法的效率要高很多。并行思维的智慧在于把一个人无法完成的任务分配给其他人，让其他人参与进来共同完成。例如，查询操作可能涉及海量数据集，为了高效地获取查询结果，可以先把海量数据集水平切分，放置到多个存储器数据库中，然后将查询请求分发到这些数据库引擎中，再将这些数据结果合并，就可以得到实际的结果。

互联网上的信息共享和对这些信息的运用也是典型的并行思维。大数据的价值在于将其置于“收集应用程序”的良性循环中，并将更多的数据输入其中，以在实际生产生活中产生价值。例如，在音乐、视频和商品领域，很多网站都有推荐功能，让用户自己来收藏喜欢的音乐、视频和商品。企业发展基于用户的选择，采用科学合理的算法为用户重新推荐，这就形成了一个循环：“分析—推荐—反馈—再分析—再推荐”。

### 3.4.3 分解思维

分解就是将一个复杂的问题分解成若干个较小的、更简单的子问题加以解决，然后对子问题的结果进行合并，得到最终的答案。我们都熟悉曹冲称象的故事。将无法分解的大象替换为可以分开称量的石头，同时借用木船和水的组合代替木秤，从而逐一称出石头的重量，最后求得大象的重量。这其实就是典型的分解思维。

分解思维的运用最著名的案例就是 Google 的 MapReduce 数据处理工具。MapReduce 的设计初衷是实现对大规模网页数据进行高效、快速的并行检索。这个过程可简单概括为：通过 Map（映射）操作获取海量网页的内容并建立索引，将大任务拆分成小的子任务，并且完成子任务的计算；通过 Reduce（规约）操作根据网页索引处理关键词，最后将中间结果合并成最终结果。

大数据技术提供了更强的数据分析与处理能力。当单个数据过于宏观时，通过对数据进行不同维度的分解可以获得更详细的数据。

### 3.4.4 随机思维

在计算机科学中，随机数是与概率相关的概念，随机数的显著特点是它的生成和前面的数没有任何关系。当我们只能通过推断求得一个问题的近似解，而无法用分析的方法来求得精确解时，可以运用随机思维。随机思维是求得近似解的有效方法。随机思维的流程是，首先建立一个与问题相关的概率模型，使所求问题的解正好是该模型的特征量；然后通过生成大量的随机数进行模拟统计实验，统计出某事件发生的百分比，只要实验次数足够多，该百分比就能接近事件发生的概率；最后利用前面建立的概率模型求得结果。

看似随机的问题，通过收集海量的数据，一切似乎又是可预见的。这也是我们所说的经验价值，或者说从不确定性中寻找确定性。例如，利用灾害预警系统来预测灾害并分析。通过将一定周期内灾害与气候、天气、土壤及自身发展因素等资料信息数据化，形成自然灾害风险级别与灾害影响因素的拟合函数，再建立回归方程，预测变量的依赖性。再如，在预防犯罪方面，过去警察只能沿街巡视，现在使用数学模型来分析大数据，警察可以有效地预测城市中何时何地可能发生犯罪行为，之后进行有针对性的巡逻。

### 3.4.5　迭代思维

“迭”是反复、屡次的意思，“代”是替换、替代的意思，“迭代”合在一起就是反复替换的意思。迭代方法是一个不断用变量的旧值递推新值的过程。迭代的核心思想基于这样一个事实：序列项由前一项的值可以得到后一项。按照这个法则，如果从一个已经知道的第一项开始，经过有限次的重复，最后会产生一个序列。著名的斐波那契数就是迭代方法的经典例子。1202 年，意大利数学家斐波那契在其《计算之书》中描述了一个“兔子问题”。假设一对兔子一个月即可发育成熟，两个月就可以产下一对小兔子，以后每个月都可以产下一对小兔子。小兔子也是一个月发育成熟，两个月开始产下一对小兔子。如果农户在年初有一对刚出生的小兔子，那么到了年末该农户有多少对兔子？每个月的兔子对数可以组成一个数列，这个数列就是 1,1,2,3,5,8,13,21,34,…。

迭代是为了接近目标而不断重复过程与反馈的工作，这是在工作中不断追求更好结果的一种典型做法。迭代思维有两个要点：一个是由小到大，从小处和细节入手，逐步构建全局；另一个是迭代的速度要快，要能快速构建出全局，即快速迭代。例如，各种互联网产品运用的就是典型的迭代思维。在产品发布以后，商家根据用户的使用反馈来评估产品和升级产品。基于用户反馈的大数据分析，为产品的升级提供了可靠的依据。