

人机协同的数据智慧机制： 智慧教育的数据价值炼金术

彭红超¹ 祝智庭²

(1. 华东师范大学 教育学部教育信息技术学系, 上海 200062; 2. 华东师范大学 开放教育学院, 上海 200062)

[摘要] 同其他产业界一样,教育界日益重视数据的价值,但数据本身的价值密度很低,无法直接服务于教育,更别说智慧教育。数据智慧可以解决此问题,它包含数据、信息、知识、智慧四个层级。然而,目前对于数据如何经信息、知识跃升为智慧还没有较好的方法。对此,本研究研制了一种人机协同的数据智慧机制,通过人与机器两方面的理解力实现数据向智慧跃升。人具有“亲而知”“觉而知”“构而知”和“审而知”四层理解力;机器具有“感而知”“描而知”“掘而知”和“学而知”四层理解力。人机协同数据智慧机制以理解力为纽带,包括数据跃升为信息的数据关系组织机制,信息跃升为知识的信息模式识释机制,知识跃升为智慧的知识原理派生机制。人机协同的数据智慧机制能为智慧教育提供可行的数据价值提炼方案,助力智慧教育中教与学行为的优化。

[关键词] 数据智慧;智慧教育;数据价值;人机协同;精准决策

[中图分类号] G431 **[文献标识码]** A **[文章编号]** 1007-2179(2018)02-0041-10

万物互联的迅猛发展,使得数字世界变得可以和物理世界相比拟。国际数据资讯公司(International Data Corporation,简称IDC)分析显示,全球数据正以每两年翻倍的惊人速度增长,预计到2020年,数据量将达到44ZB(1ZB=1012GB)(IDC,2014)。在如此庞大的数字化资产面前,教育界也日益注重对其潜在价值的挖掘与应用。我国《教育信息化“十三五”规划》指出,到“十三五”末,要“制订出台教育数据管理办法,规范数据的采集、存储、处理、使用、共享……保证数据的真实、完整、准确、安全及可用”(中华人民共和国教育部,2016)。然而不幸的是,很多企业投入大量资金后,发现无法从大数据中析出高价值的见解(Ross et al., 2013),数

据本身没有任何价值倾向(祝智庭,彭红超等,2017),且价值密度很低(Demchenko & Gruengard, 2014)。这无疑增加了数据资产变为数据垃圾的风险。因此,如何从数据中提炼出高密度的价值已成为急需解决的问题(祝智庭,彭红超等,2017a)。

在智慧教育(Smarter Education,简称SerE)祝氏定义中,数据智慧(Data Wisdom)成为技术增能的三大智慧之一(另两个为文化智慧和教学智慧),以用于使数据信息从“知几无”(Know-Nothing)的状态变为“知最佳”(Know-Best)智慧状态(祝智庭,彭红超,2017b),从而诱发行动智慧。在智慧教育这种“弱规则”领域,数据智慧体现的是一种人机协同的新型认知范式,特别是需要专家经验领悟或洞悉数

[收稿日期] 2018-01-24

[修回日期] 2018-02-09

[DOI编码] 10.13966/j.cnki.kfjyyj.2018.02.005

[基金项目] 2017年度华东师范大学教育学部大学生科研基金项目“智慧教育境域中数据智慧机制设计研究”(ECNUFOE2017KY074)。

[作者简介] 彭红超,博士研究生,华东师范大学教育学部教育信息技术学系,研究方向:智慧学习生态、精准教学(hongchao5d@qq.com);祝智庭(通讯作者),博士,教授,博士生导师,华东师范大学开放教育学院,研究方向:教育信息化系统架构与技术标准、信息化促进教学变革与创新、技术使能的智慧教育、面向信息化的教师能力发展、技术文化等(ztzu@dec.ecnu.edu.cn)。

据模式中蕴含的教育因素,从而寻求更优质的学习服务(祝智庭,孙妍妍等,2017)。目前,学界对数据智慧模型中的数据、信息、知识、智慧四个层级已经取得一定共识,但数据如何经信息、知识跃升为智慧还没有较好的方法。对此,本研究将介绍本团队研制的人机协同的数据智慧机制,为智慧教育提供可行的数据价值提炼方案,助力智慧教育中教与学行为的优化。

一、概述

教育数据使教师的决策不再仅仅依赖主观经验,更可以依据数据证据,特别是在学习数据的记录与收集成为常态的今天。当然,数据本身没有任何价值倾向,它需要被提炼为信息、知识甚至智慧才能更好地为智慧教育服务,这就是数据智慧的价值所在。

(一) 数据智慧的起源与发展

数据智慧通常被称为 DIKW (Data-Information-Knowledge-Wisdom) 金字塔、DIKW 层级结构、知识层级结构、信息层级结构、智慧层级结构或知识金字塔等(Rowley, 2007),它是表示数据、信息、知识和智慧之间结构和/或功能关系的模型(Zins, 2007)。关于数据智慧的起源,目前还没有得到确认,较多业内人士认为,是由前捷克斯洛伐克教育家米兰·泽莱尼(Milan Zeleny)提出的(并没有给出图形表征)(Wallace, 2007)。泽莱尼于1987年将数据智慧层级结构的四个要素映射为知识的形成:知几无、知是何(Know-What)、知如何(Know-How)、知为何(Know-Why)(Zeleny, 1987)。1988年,该层级结构再次出现在国际通用系统研究学会(International Society for the Systems Sciences)的演讲中,并被美国组织理论家罗素·阿科夫(Russell Ackoff)1989年发表的论文中提及(Ackoff, 1989),因此很多资料将阿科夫作为数据智慧层级结构的“原始阐述者”(Rowley, 2007)(也没有给出图形表征)。

其实,早在泽莱尼之前,亨利(Nicholas L. Henry)已经在知识管理境域中对数据、信息、知识作了区分,指出数据是一种原始事实,不像信息、知识那样会改变我们(Henry, 1974)。华裔地理学家段义孚(1982年)也在《未来学家》杂志中指出,除了智慧、知识、信息,还需增加一层数据(叶继元等,

2017;陈书悦,2013),哈伦·克利夫兰(Harlan Cleveland)则更加注重事实的重要性,提出了事实与观念(Facts & Ideas)-信息-知识-智慧的结构(Cleveland, 1982)。泽莱尼重点论述了数据、信息、知识、智慧四个要素及其映射,提到智慧之上应是开悟(Enlightenment),即用真理维度丰富依旧价值中立的智慧(Zeleny, 1987)。然而,开悟这层似乎并未得到学者们的认可。阿科夫即是其中一位,他依然关注 DIKW 四要素结构,并以此为基础,在知识与智慧之间增加了领悟(Understanding)层,形成 DIKUW 结构。与泽莱尼不同的是,阿科夫认为领悟是回答“为何”之问,而智慧用于处理价值观问题,它涉及价值判断(Ackoff, 1989)。吉恩·贝林格(Gene Bellinger)等人认为领悟不应作为单独的一层,应该是 DIKW 的纽带,正是领悟支撑了上一层到下一层的转变,领悟主要是在联通(Connectedness)基础上的认知(Bellinger et al., 2004)。唐纳德·克拉克(Donald Clark)将这种联通分为“事物部分属性之收集”“部分属性之联结”“联结成为整体”“诸多整体之联结”四个部分(Clark, 2010),并认为可以通过调查、吸收、行动、互动、内省五种学习策略实现领悟(Clark, 2012)。

对于数据智慧的要素/层级,其实一直没有统一的结果。比如,德邦斯·安东尼(Debons Anthony)等在数据层前增加“事件”“符号”“规则和公式”等层(Debons et al., 1988)。斯科特·卡彭特(Scott Carpenter)等在 DIKW 结构前增加环境层,在智慧后增加展望(Vision)层(Carpenter & Cannady, 2004),形成 EDIKWV 结构。惠勒也在智慧层后添加了转化(Transformation)层,形成 DIKWT 结构,并将参与(Engagement)作为各层要素转变的纽带(Wheeler, 2013)。我国学者顾基发认为需要引入道德(Moral)判定是非,因此提出了 DIKWM 结构。数据智慧的要素/层级五花八门,更有甚者提出了包含十个要素/层级的结构体系(Psuf10, 2014)。究其原因,主要有以下两点:1)专业知识的差异致使对数据智慧各要素的理解有异;2)应用目的的差异致使关注的侧重点不同。

(二) 智慧教育中数据智慧核心理念

教育技术旨在借助技术促发教育信息化,它是关于教育与技术的交叉学科。因此,作为教育信息

化创新发展新趋势的智慧教育(祝智庭, 2014), 它的数据智慧应该着重考虑教育中的“人”与技术中的“机器”的协同问题。此外, 智慧教育祝氏定义中的三大智慧各有分工: 文化智慧定“导向”, 数据智慧定“决策”, 教学智慧定“行动”(祝智庭, 彭红超, 2017c), 因此本研究介绍的数据智慧机制主要从人机协同角度出发, 解决“从数据中提炼新见解”和“基于新见解制定优质的决策方案”的问题。

从“数据智慧的起源与发展”不难看出, 数据智慧的要素/层级虽然不统一, 但基本上都包含数据、信息、知识、智慧这四个要素/层级, 本研究将采用这一共识, 即 DIKW 结构。原因有两点: 1) 相比原始事实、符号、标志、信号等术语, 数据更全面(数据含纳它们), 更符合智慧教育的理念(如数据驱动的精准确策理念); 2) 由于智慧教育中的数据智慧已有文化智慧的导向, 不需要在“智慧”层增加“道德”“展望”(Wiebe, 2009)等层。

智慧教育中人机协同的数据智慧如图 1 所示。其中, 数据是关于事物、事件属性的一组离散的、客观的符号、数字、信号等(Ackoff, 1989; Zins, 2007), 它是原始的, 本身没有任何意义(Bellinger et al., 2004)。信息是通过关系连接而赋予某种意义的数据(Bellinger et al., 2004; Gamble & Blackwell, 2001; Wallace, 2007), 这种关系(如相关关系)常表现为组织形式(如排列), 因此有学者认为信息是组织化或结构化的数据(Rowley, 2007)。知识是模式化或模式关联化信息(Bellinger et al., 2004)。数据、信息、知识都与过去有关, 它们处理现有或已有事物、事件(Ackoff, 1989)。智慧是原理框架下的最佳决策行为, 这种原理是已验证的有价值的知识, 它一部分由知识派生, 一部分由领域规定。智慧关注未来, 它能够明确地做出正确判断和决策, 从而正确地解决问题(祝智庭, 彭红超, 2017b), 甚至创造未来(Bellinger et al., 2004; Clark, 2010)。数据向智慧跃升的过程是数据的关系组织、信息的模式识释、知识的原理派生的过程。这是一个价值提炼的过程(随之失去的是可传播性), 它通过人与机器两方面的理解力(理解力作为数据智慧的纽带)来实现。人的方面分为“亲而知”“觉而知”“构而知”“审而知”四个层级; 机器方面分为“感而知”“描而知”“掘而知”“学而知”四个层级。人与机器各层理解

力的协同分别形成数据的“知几无”、信息的“知是何”、知识的“知如何”和智慧的“知何为”的认知状态。处于不同认知状态的各层在智慧教育中各有用途: 处于“知几无”状态的数据, 可用于教与学印记或学习资源的传播与存储; 处于“知是何”的信息可用于意义的可视化呈现(Ritholtz, 2010); 处于“知如何”的知识可用于见解的洞悉; 处于“知何为”的智慧可用于服务(含干预)的决策。

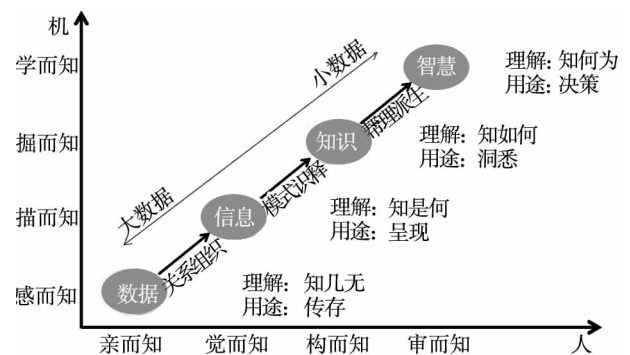


图 1 智慧教育中人机协同的数据智慧图谱

智慧教育中的数据分为大数据(Big Data)和小数据(Small Data)两部分, 数据具有巨量、迅变、高速、多样、稀值等特性, 常规的数据工具与统计方法无法获得与处理。小数据是小体量和特定属性的数据, 它具有情感温度, 机器算法无法处理此类问题。对此, 祝智庭团队提出, 在智慧教育中, 大数据部分可借助机器智能的优势处理, 小数据部分可借助专家智能的优势处理(祝智庭等, 2017)。这种人机协同的过程包含数据的关系组织、信息的模式识释、知识的原理派生三个阶段, 即数据智慧机制的三个阶段。

二、数据的关系组织机制

在智慧教育中, 数据由教育相关人员亲身践行(亲而知)与机器的感知记录(感而知)得来。这两方面数据的收集往往是在不知其用途的情况下发生的(特别是大数据部分), 这使得数据蕴含的价值更加稀疏, 很难直接看出事实的真实情况。因此, 通过特定的关系组织数据, 使其显露意义(即跃升为信息)有重要意义。图 2 展示了数据的关系组织机制, 共分为四个环节: 目的切入、关系确立、数据塑形和意义表征。前两者偏向于人工处理, 后两者偏向

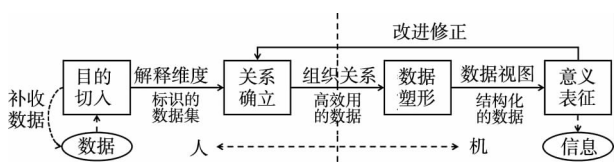


图2 数据的关系组织机制

于机器处理。

(一) 目的切入

离散的数据因缺少情境和解释而没有意义 (Rowley & Hartley, 2008), 因此, “目的切入”环节主要是在特定的学习情境下, 以智慧教育的监测目的为切入点, 制定解释维度。这里的目的不仅仅是期望实现的学习目标, 还包括希望得到的教与学过程中学生的行为反馈和能力变化等, 后者主要从采用的具体学习方略中析出。目的越具体越好, 最好具有高度的可操作性和可测量性, 可以采用格式 1 所示的结构表征, 其中 “[]” 表示可缺省, “{ } [; … n]” 表示 “{ }” 中的内容可重复 n 次, 并用 “;” 分割, “-” 是连接符, 将各部分连接起来。它表示在 “时间 (或期间)”, “学生” 在 “情境” 中用 “工具” 进行 “活动”, 学习 “内容”, 并显露出某些 “表现”, 这些表现达到了指定的 “标准”。例如, “7”, 个体-观看视频-杠杆原理/完成测试, 5/6” 表示在 7 分钟内, 每位学生观看视频中的杠杆原理, 完成测试中的 6 道题并至少答对 5 道题。格式 1 中的 “活动” 可以借助活动理论的层级结构 (活动层、行为层、操作层) 实现可操作性 (彭红超, 祝智庭, 2016), “内容” 在智慧教育中主要有学材、习材、创材 (祝智庭, 2016), “标准” 最好提供具体数值或特征以实现可测量。

[时间,] 学生-[情境-][工具-]活动[-内容]/{表现[, 标准]} [; … n]

每个目的的实时情况均需相应的数据来说明, 这为数据的意义解释提供了方向。由此, 可以以目的为切入点, 制定数据的解释维度。维度与目的的属性对应, 因为一个目的可能会有多个属性 (如格式 1 有 “时间” “学生” 等八个属性), 因此, 一个目的可以有多个解释维度。将需要的维度按照 DOM 树或其他结构组织起来, 便可形成解释维度体系, 用于 “关系确立” 环节。

另外, 智慧教育中的数据多是无目的收集得到

的 (大数据特性使然), 这些数据并非对切入的目的都有效用, 因此, 需要将有效用的数据集或资源作标识。切入的目的越不明确, 越要尽可能多的标识数据集或资源, 从而提高发现隐藏模式或关系的可能。对于解释维度需要但未被收集的数据, 需要补收, 并标识这是有效用的数据集。

(二) 关系确立

“关系确立” 环节需要考虑数据属性和解释维度两方面因素。数据属性表征数据的特性, 解释维度表征监测目的。从数据的可计算属性看, 数据可分为标称数据 (可进行 =、≠ 计算)、序数数据 (可进行 =、≠、>、< 计算)、区间数据 (可进行 =、≠、>、<、+、- 计算)、比率数据 (可进行 =、≠、>、<、+、-、×、÷ 计算) 四类 (Pang-Ning et al., 2011)。依据其属性, 数据的组织关系可分为类型、序列、间距、共基 (即共同基准: 零点) 四类, 这四类数据组织关系依据解释维度可进一步细分 (见表一)。

表一 数据组织关系类目

按数据属性分类	按解释维度分类
类型	学习环境、学习组、班级、学科、学习工具、学习活动、性别等。
序列	年级、成绩、学习流程、进步名次、学习动力、语法等。
间距	学习时间、摄氏温度等。
共基	及格率、年龄、视频暂停次数、学习速度、学习结果流畅度等。

为了去掉对说明监测目的不相关的数据集, 在关系确立前, 标识的数据集或资源需要加以归类 and 过滤, 之后从中提取的数据将被转换成可数据分析的格式 (形成高效用的数据)。进行关系确立时, 依据解释维度表征的监测目的, 高效用的数据需要被判定是变量还是常量, 在数据意义的可视化表征时, 常量多在标题中描述, 变量多作为图表的坐标轴。无论是变量还是常量, 均需判定其数据属性的类型, 并依据相应的属性类型结合监测目的, 确立对应的数据组织关系。例如, 在 “7”, 个体-观看视频-杠杆原理/完成测试, 5/6” 中, “7” “个体” “5” 等数据是变量, 其组织关系可以定为 “时间-共基” “学生-类型” 和 “答对题数-共基”。“观看视频” “杠杆原理” “完成测试” “6” 等数据是常量, 其组织关系可以定

为“活动-类型”“内容-类型”和“答对题数-共基”。在可视化表征时,常量部分可在标题中表述为“学生观看杠杆原理视频完成6道测试题的情况”。

(三) 数据塑形

高效用的数据来源于不同数据集(数据库、文本、XML等),会出现数据冗余、重复、不一致、遗漏值、无效数据等问题。这一环节只需解决与数据组织关系相关的数据存在的这些问题即可。对于遗漏值,可通过数据验证,借助冗余信息尽可能将它们补填好。无法填充的,需要删除遗漏值所在的记录或属性,同时一并删除无效数据。接下来,将这些不同来源的数据进行数据聚合(依据相同的属性或数据域),形成统一的数据视图。数据视图还需要进行数据清洗,删除冗余、重复的数据,修正不一致数据。对于不一致数据,如果是单位不统一造成的,需要进行单位换算,如果是属性名称不一致造成的,需要修正为统一的名称,如果是内容本身有冲突,则需要保留来源于最高可信用数据集的数据。最后,清除所有与数据组织关系无关的数据,便得到意义表征需要的结构化数据。

(四) 意义表征

意义表征所用到的所有数据均来源于上一环节最终得到的数据视图中的数据,可以使用类似 SELECT 的语句对这些原始数据进行可视化表征,也可以先数据分析再表征。这一环节使用的数据分析类型主要为描述型分析,解析“发生了什么”(描而知)。描述型分析主要有集中趋势描述和离散趋势描述两类,前者主要包括平均数(算数、加权、几何、调和)、中位数、众数等方法,后者主要包括极差、平均差、方差、标准差等方法。

数据意义的可视化表征首先要选定坐标系的类型,教学中常用的有直角坐标系和极坐标系(如饼图),偶尔也会用到地理坐标系;之后选取坐标轴的尺度类型,基本原则为:对于区间数据和比率数据,表征学生相关变化的幅度时,使用等距尺度,如散点图等;表征学生相关变化的速度时,使用比例尺度,如标准变速图(彭红超,祝智庭,2017)。另外,标称数据一般使用分类标尺,序数数据一般使用顺序标尺。最后的可视化呈现常用即席报表和仪表盘的形式,前者是静态数据表征,后者是动态数据表征(Camm et al., 2017),当然,也会用到网络图、极区

图等特殊的可视化形式。

通过描述分析和可视化表征,数据便跃升为富有意义的信息,这些信息很容易被师生感受辨别(觉而知),形成对数据意义的初步理解。

三、信息的模式识释机制

通过关系组织机制,数据跃升为信息,但信息通常只表征“何人、何时、何处、何事”等浅显事实。对于智慧教育的精准决策来讲,事实背后隐含的模式更有价值。识别出信息中内隐的模式并对其进行解释(即模式识释),可使信息跃升为知识,表征事情是如何发生的,即“知如何”。图3描绘了信息的模式识释机制,共分为四个环节:机器方面为特征提取和信息模化,人工方面为意义理解和模式解释。

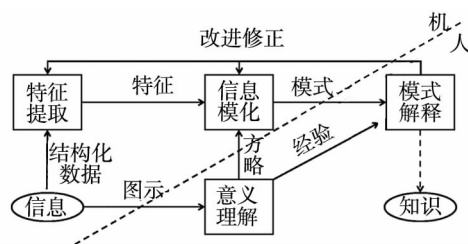


图3 信息的模式识释机制

(一) 特征提取

通过数据的关系组织机制得到的是结构化数据和即席报表、仪表盘等图示。结构化数据组成的测量空间的维度通常过于庞大且并非所有维度都很重要,还会降低算法的性能,因此需要通过映射或变换降低维度,获取一组数量较少的更能反映本质的特征。从形式上看,这些特征属于数学特征(另两类为物理特征、结构特征)(盛立东,2010),非常抽象,不易被人感知,但适合计算机处理。

特征提取可选取两种途径:1)直接选取法,即直接选取失真最小的低维子空间,具体可用最优搜索法、次优搜索法、遗传算法等;2)变换降维法,即对测量空间进行变换映射来降维,具体可用基于距离(如类内距、类间距)可分性测度的特征提取法和基于离散 K-L 变换的特征提取法(即主成分分析法)等。

每种特征提取方法都有(或构造)自己的可分性判据,通过识别可分性判据满足的条件,可以衡量哪种提取方法最有效。一般讲,可分性判据需满足

四个条件(周丽芳等,2013):1)与误分概率(或误分概率的上、下界)有单调关系,即判据取最大(小)值时,误分概率最小(大);2)特征相互独立;3)判据有非负性(即当 $i \neq j$ 时 $J_{ij} > 0$,当 $i = j$ 时 $J_{ij} = 0$, J_{ij} 为第 i 类和第 j 类的判据)和对称性(即 $J_{ij} = J_{ji}$);4)加入新特征后,判据值不减。当然,可分性判据很难同时满足这四个条件,一般满足一个或几个即可。

(二) 意义理解

通过数据的关系组织机制得到的即席报表、仪表盘等图示会显示分散的物理特征和结构特征,这些特征对计算机而言比较复杂,却容易被人感知。本环节主要目的是理解图示中的物理特征和结构特征,为数学特征的信息模化与模式解释分别提供方略框架与经验依据。

格式塔心理学的知觉组织原则指出,人们按照接近律、相似律、连续律、闭合律、同命律将分散的特征组织成整体(邵志芳,2013)。对整体的理解可以从两方面入手(构而知):自我认知和与人互动。自我认知指通过观察、思考、推理等方式将整体与原有知识经验建立联系并革新原有的认知图式,必要时可以查阅相关材料,以便促使原有知识经验的提取。当个体无法理解信息的意义时,可从与人互动,采用小组研讨、头脑风暴、专家咨询等方式发挥群体智慧。

心理学研究表明,人类很容易并且高精度地从即席报表、仪表盘中的得到对整体的理解。在智慧教育中,有三类群体需要从上述可视化图表中得到对整体的理解(知是何):教育者、受教育者、数据分析师。前两者主要关注教与学方面的理解,如学习者当前的学习状态、能力水平等。数据分析师则偏重于关注数据挖掘与学习分析方面的理解,如应该采用何种信息模化的方略才能得到内隐在浅显事实中的模式,应该依据哪些已有的模式解释经验才能准确地估所得模式是否有用以及将有用的模式内化为人类的新知识。

(三) 信息模化

特征提取环节得到了低维的更能反映本质的数学特征,依据意义理解环节得到的方略框架,可以对这些数学特征进行信息模化,从而得到内隐在信息中的模式(即规律或趋势)。

信息模化主要采用数据挖掘技术,挖掘有趣的

模式的过程(Han et al., 2012),即有效性、新颖性、潜在有用性、最终可理解性(史忠植,2011)。意义理解环节得来的方略其实是某种数据挖掘模型框架,例如贝叶斯、决策树、人工神经网络等。利用选定的数据挖掘模型框架处理数学特征,最终会得到有趣的模式,如类、概念、联系、序列模式、结构模式等。目前较为流行的是基于人工神经网络的数据挖掘技术,通过构建适切的神经网络(BP神经网络、卷积神经网络等),之后通过学习大量的训练集(主要有监督学习和无监督学习)让机器具有类似人脑处理信号的能力,以对数据特征逐步抽象(祝智庭,彭红超,2017d),得到“有趣的”模式(Pattern)。在这里,抽象过程主要表现为“事物之部分属性、部分属性之联结、联结成为整体”。对于机器来讲,这些有趣的模式即知识(Han et al., 2012),它通过信息挖掘得到(掘而知)。

(四) 模式解释

信息模化得到的模式需要解释,以便解决可能存在的冗余或无法满足需求的问题以及将模式内化为人类知识的问题(构而知)。前一问题是对模式本身的解释(也称模式评估),后一问题是对模式理解的解释,解决这两方面的问题,均需依据意义理解环节得到的经验。前一问题可以借助模式兴趣度来判定。模式兴趣度的度量方式可依据经验选择,客观度量方面供选择的方式有支持度-置信度框架、提升度、覆盖率、余弦等,主观度量方面主要看模式是否出乎意料或提供了对采取行动至关重要的信息(Han et al., 2012)。度量后如果发现模式冗余或无关,则将其删除,如果发现模式无法满足需求,则变换特征提取的方法或信息模化的框架。这一问题的解决主要由数据分析师完成,特别是客观度量方面,需要专业的数据分析本领。对模式兴趣度的主观度量,数据分析师需要与教育管理者和教师交流、研讨,明确其需求,并以此判定模式是否对当前管理与教学有价值。后一问题的解决主要由教育管理者和教师完成,其借助自身已有的管理与教学经验来理解得到评估的有趣模式。这些得到理解的有趣模式与得到理解的整体建立意义联系后(按照同化、顺应方式),形成新的知识。得到新知识后,管理者与教师便可进行见解的洞悉,形成对“知如何”的认知,即洞悉事情发生的来龙去脉,明确事情是如

何一步步发展至此的。这其实是知识创新的过程。依据著名的 SECI 过程模型(Nonaka, 1991),经过社会化(Socialization)、外在化(Externalization)、组合化(Combination)和内在化(Internalization)四阶段的螺旋式发展可实现见解的洞悉,也可以实现知识在个人、群体、组织、跨组织层面的传递与创造。

四、知识的原理派生机制

发生了什么事情(信息),事情是如何发生的(知识)均是对已发生或正发生事情的理解。它有助于类似问题的处理,但这种理解还不足以应用于新的情境或问题(即迁移),因此,知识仍需进一步跃升为智慧,满足精准决策这一需求(智慧主要面向前所未遇的问题)。图4刻画了知识跃升为智慧的机制——知识的原理派生机制,共四个环节:专家端倪、价值判断、机器学习、决策生成。智慧的形成,主要由人完成,当然,通过机器学习技术,机器也已经显露出这方面的本领。

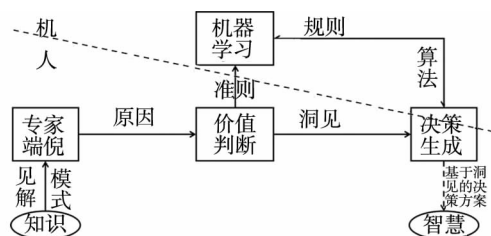


图4 知识的原理派生机制

(一) 专家端倪

信息的模式解释会得到两类知识:机器的知识——模式;人的知识——见解。模式表征教育数据的规律和趋势,见解是对模式理解后而有了主意。这种主意多是对学生的学习是否存在问题或潜在问题,以及是否需要变换学习服务或提供干预措施的断定。其实,无论断定结果如何,最好都能找出原因(Know-Why),这是迁移的基础。

专家端倪是数据模式启发的专家探究法(祝智庭等,2017),可分为三步:联想、推理和归因。联想用来探寻相关的线索情报(诸多整体之联结),可以按照相似律(如形状、结构、功能、性质等的相似)、接近律(如时间、空间的接近)、对比律(如性质、特点等的不同与相反)展开联想。推理用来推导未知的线索关联,可以采用演绎推理(一般到特殊)、归

纳推理(特殊到一般)和类比推理(特殊到特殊、一般到一般)策略实现。归因用来进一步探求线索的因果联系,可以采用著名的穆勒五法,包括求同法、求异法、求同求异并用法、共变法、剩余法(刘江,2004)。专家端倪主要有教育教学专家完成,如教育学者、管理者、一线教师等。

(二) 价值判断

智慧是价值驱动的(Başkarada & Koronios, 2013),因此,在原因得到领悟后,专家还需要对学习印记理解(如发生了什麼,如何发生的,为什么会发生等)进行价值判断。价值判断有助于形成透彻见解(即洞见),例如,“如何确定学习服务的时间、条件、对象、模式等才能有高效能”(Bellinger et al., 2004)。由此,数据智慧将由关注“正确地做事”转为关注“做正确的事”。

价值判断同样由教育教学专家完成,此环节需要考虑教育价值和成效价值两个指标。智慧教育的价值集中体现在“精准、个性、协同、优化、思维、创造”六大原则(祝智庭 & 余平,2017)中。因此,教育价值指标应以这六方面的价值等级为依据。具体讲,精准方面可分为目标精准、问题精准、辅助精准三层;个性方面可分为个体特征、适性服务、需求愿景三层;协同方面可分为协作、合作、群智三层;优化方面可分为适性最优、优势互补、迭代进化三层;思维方面可分为思得准、思得全、思得巧三层;创造方面可分为做得了、做得好,做得妙三层。另外,成本、效益也是影响服务决策的关键因素,因此,成效价值可以采用“效益/成本”来衡量。通过审辨思维,教育者需要决断何时、何种服务的综合价值是最高的(审而知)。此外,除了洞见,价值判断也会形成一系列学习服务准则。

(三) 机器学习

价值判断得到的服务准则和决策生成的服务决策规则可以作为训练集供机器学习。通过学习,机器自身的算法将得到优化(学而知)。这样,机器制定的个性化学习服务方案也会更適切。

为提高机器学习的性能,对于简单的学习服务决策的学习可采用浅层学习(Shallow Learning)技术,复杂的学习服务决策的学习可采用深度学习(Deep Learning)。浅层/深度学习得到的相关特征可用于强化学习(Reinforcement Learning)技术(焦

李成等, 2017), 利用它蕴含的奖惩制度(即回报函数)生成最优的算法。最后, 借助迁移学习(Transfer Learning)(史忠植, 2011), 学到的知识(包括优化的算法)可在不同情境中应用。其中, 浅层/深度学习可以解决强化学习时特征状态需要人工设定的难题; 迁移学习可以解决新问题中数据采集, 甚至无法获取的难题。

(四) 决策生成

人得到对学习服务的见解(价值判断的结果)、机器得到优化的算法之后(机器学习的结果), 便可各自或协同开展学习服务决策的生成工作, 这是一种基于洞见的行为(insight-based action), 形成了“智慧行为”, 可简称为“智行”。

在智慧教育中, 精准的个性化学习服务是智慧人才培育的基础。这方面, 比尔与梅琳达·盖茨基金会(K-12 Education Team, 2014)提出的个性化学习服务四要素——学习者画像(Learner Profiles)、能基发展(Competency-Based Progression)、个人学习路径(Personal Learning)、柔性学习环境(Flexible Learning Environments)具有一定的影响力。依据这四个要素, 人机在制定学习服务决策时, 可以从个体特征、个人表现、个性需求、学习生态环境(祝智庭等, 2017b; 祝智庭等, 2017c)四方面着手。另外, 借用个性化学习适配处方模型(祝智庭等, 2017), 还可以实现从班级、小组、个人三个层面生成低成本、高精度的基于洞见的决策方案。

由上可知, 知识的原理派生机制最终形成的原理形式化为规则, 表现为基于洞见的决策(insight-based decision-making), 可简称为“智策”, 此过程伴随着是否提供服务、何时提供服务、提供何种服务、如何服务等决策。至此, 教育数据实现了向智慧的跃升, 形成了极具可资行为的“行数据”(Actionable Data), 这是从数据智慧转换为教学智慧的阀门。

五、结 语

智慧教育培育的人才有两方面智慧: 思的智慧和做的智慧(祝智庭, 彭超红等, 2017b)。思的智慧主要为思维模式, 做的智慧主要是创造潜能。对于人工智能来说, 思维模式体现在从数据中提炼新见解, 创造潜能体现在基于新见解的决策方案(智策)。因此, 数据智慧包含透彻的新见解(洞见)和智策(基

于洞见的决策方案)两方面, 二者即为数据价值。

由上述的数据智慧机制可知, 随着数据价值的提炼, 数据对人的依赖程度会增大, 特别是处理前所未遇的教学问题时。不过庆幸的是, 这类问题被解决后, 就变成经验和知识, 通过逐渐成熟的机器学习技术, 这种经验和知识可以被机器习得。这样, 后续再有类似的问题出现, 机器便能自动处理。

另外, 从机器角度看, 数据智慧中的数据、信息、知识、智慧本质上都是数据。随着数据价值的提炼, 层层无关数据会被“清洗”掉, 因此, 这个过程也是大数据走向小数据的过程。大数据主要考量相关关系, 小数据则转向考量因果关系, 因此, 数据智慧也是相关关系与因果关系融合的智慧。从事物联通性角度看, 数据是对“事物部分属性”的表征, 而数据到信息、知识再到智慧的跃升, 是“部分属性之联结、联结成为整体、诸多整体之联结”的过程。这是局部到整体、具体到抽象的过程。这与教育领域的深度学习理念相吻合。本质上, 深度学习是激发更多神经元参与学习, 参与的神经元越多, 感官得来的信息越能够抽象为更高级的意义, 当意义触及事物本质或原理时, 便到了迁移的条件。因此, 很多学者也认为深度学习是为迁移而学。这一点和人工智能领域的深度学习相通。

最后, 需要指出的是, 本研究虽然从线性关系的角度解读了数据跃升为智慧的过程机制, 但这并不说明数据、信息、知识、智慧之间是简单的线性关系, 已有研究表明, 它们之间均有重叠的部分(叶继元等, 2017)。后续, 本团队将把研究成果应用于深度学习, 打造以数据智慧为骨架的深度学习服务。

[参考文献]

- [1] Ackoff, R. L. (1989). From data to wisdom [J]. *Journal of applied systems analysis*, 16(1): 3-9.
- [2] Baškarada, S., & Koronios, A. (2013). Data, Information, Knowledge, Wisdom (DIKW): A semiotic theoretical and empirical exploration of the hierarchy and its quality dimension [J]. *Australasian Journal of Information Systems*, (18): 5-24.
- [3] Bellingier, G., Castro, D., & Mills, A. (2004). Data, information, knowledge, and wisdom [EB/OL]. [2017-05-22]. <http://www.systems-thinking.org/dikw/dikw.htm>.
- [4] Camm, J. D., Cochran, J. J., Fry, M. J., Ohlmann, J. W., Anderson, D. R., Sweeney, D. J., & Williams, T. A. (2017). *商业数据分析* [M]. 北京: 机械工业出版社: 5.

- [5] Carpenter, S. A., & Cannady, J. (2004). Tool for sharing and assessing models of fusion-based space transportation systems[C]. Virginia: American Institute of Aeronautics and Astronautics;11-14.
- [6] 陈书悦(2013). 说一说 DIKW 金字塔 [EB/OL]. [2017-11-11]. http://blog.sina.com.cn/s/blog_5be3027d0101mtq2.html.
- [7] Clark, D. (2010). Understanding and Performance[EB/OL]. [2017-05-22]. <http://www.nwlink.com/~donclark/performance/understanding.html>.
- [8] Clark, D. (2012). Context[EB/OL]. [2017-05-23]. <http://www.nwlink.com/~donclark/knowledge/context.html#connection>.
- [9] Cleveland, H. (1982). Information as a resource[J]. *Futurist*, 16(6): 34-39.
- [10] Debons, A., Home, E. E., & Cronenweth, S. (1988). *Information science: An integrated view*[M]. Boston: G. K. Hall;5.
- [11] Demchenko, Y., & Gruengard, E. (2014). Big data course and learning model for online education (LMO) [EB/OL]. [2017-07-07]. <http://www.uazone.org/demch/presentations/edison2014wsh-big-data-edu-v01.pdf>.
- [12] Gamble, P. R., & Blackwell, J. (2001). *Knowledge management: A state of the art guide*[M]. London: Kogan Page Publishers;43.
- [13] Han, J. W., Kamber, M., & Pei, J. (2012). *数据挖掘概念与技术*[M]. 北京:机械工业出版社;9-15.
- [14] Henry, N. L. (1974). Knowledge management: A new concern for public administration [J]. *Public Administration Review*, 34(3): 189-196.
- [15] IDC(2014). *The Digital Universe of Opportunities: Rich data and the increasing value of the Internet of Things*[EB/OL]. [2017-11-10]. <https://www.emc.com/leadership/digital-universe/2014iview/index.htm>.
- [16] 焦李成, 赵进, 杨淑媛, 刘芳(2017). 深度学习、优化和识别[M]. 北京:清华大学出版社;195-208.
- [17] K-12 Education Team (2014). *Early progress: Interim research on personalized learning* [R]. Bill & Melinda Gates Foundation;6.
- [18] 刘江(2004). *逻辑学:推理和论证*[M]. 广州:华南理工大学出版社;167-177.
- [19] Nonaka, I. (1991). The knowledge-creating company[J]. *Harvard Business Review*, Nov/Dec: 96-104.
- [20] Pang-Ning, T., Steinbach M., & Kumar V. (2011). *数据挖掘导论*[M]. 北京:人民邮电出版社;16-17.
- [21] 彭红超, 祝智庭(2016). 面向智慧学习的精准教学活动生成性设计[J]. *电化教育研究*, (8): 53-62.
- [22] 彭红超, 祝智庭 (2017). 以测辅学:智慧教育领域中精准教学的核心机制[J]. *电化教育研究*, (3): 94-103.
- [23] Psuf10(2014). A walking tour of the pathway to wisdom[EB/OL]. [2017-05-22]. <https://psuf10.wordpress.com/2014/02/16/a-walking-tour-of-the-pathway-to-wisdom/>.
- [24] Ritholtz, B. (2010). Intelligence hierarchy: Data, information, knowledge, wisdom [EB/OL]. [2017-11-11]. <http://ritholtz.com/2010/12/hierarchy-of-visual-knowledge/>.
- [25] Ross, J. W., Beath, C. M., & Quaadgras, A. (2013). You may not need big data after all[EB/OL]. [2017-07-13]. <https://hbr.org/2013/12/you-may-not-need-big-data-after-all>.
- [26] Rowley, J. (2007). The wisdom hierarchy: Representations of the DIKW hierarchy[J]. *Journal of information science*, 33(2): 163-180.
- [27] Rowley, J. E., & Hartley, R. J. (2008). *Organizing knowledge: An introduction to managing access to information*[M]. Aldershot: Ashgate Publishing, Ltd. ;5-6.
- [28] 邵志芳(2013). *认知心理学——理论、实验和应用(第二版)*[M]. 上海:上海教育出版社:9-11.
- [29] 盛立东(2010). *模式识别导论*[M]. 北京:北京邮电大学出版社;239.
- [30] 史忠植 (2011). *知识发现(第二版)*[M]. 北京:清华大学出版社.
- [31] Wallace, D. P. (2007). *Knowledge management: Historical and cross-disciplinary themes*[M]. Westport, Conn: Libraries unlimited;1-14.
- [32] Wheeler, S. (2013). *Learning theories for the digital age* [EB/OL]. [2017-11-11]. https://www.slideshare.net/timbuckteeth/learning-theories-for-the-digital-age?from_action=save.
- [33] Wiebe, J. (2009). Seven levels of interaction [EB/OL]. [2017-11-11]. <https://onemindblog.wordpress.com/tag/data-information-knowledge-and-wisdom/>.
- [34] 叶继元,陈铭,谢欢,华薇娜(2017). 数据与信息之间逻辑关系的探讨:兼及 DIKW 概念链模式[J]. *中国图书馆学报*, (3): 34-43.
- [35] Zeleny, M. (1987). Management support systems: Towards integrated knowledge management [J]. *Human systems management*, 7(1): 59-70.
- [36] 祝智庭(2014). 以智慧教育引领教育信息化创新发展[J]. *中国教育信息化*, (9): 4-8.
- [37] 祝智庭(2016). 智慧教育新发展:从翻转课堂到智慧课堂及智慧学习空间[J]. *开放教育研究*, 22(1): 18-26.
- [38] 祝智庭,彭红超(2017a). 数据智慧:人机协同的认知新范式[R]. 杭州:国家自然科学基金委员会.
- [39] 祝智庭,彭红超(2017b). 智慧学习生态:培育智慧人才的系统方法论[J]. *电化教育研究*, (4): 5-14.
- [40] 祝智庭,彭红超(2017c). 智慧学习生态系统研究之兴起[J]. *中国电化教育*, (6): 1-10.
- [41] 祝智庭,彭红超(2017d). 深度学习:智慧教育的核心支柱[J]. *中国教育学报*, (5): 36-45.
- [42] 祝智庭,彭红超,雷云鹤(2017). 解读教育数据智慧[J]. *开放教育研究*, (5): 21-29.
- [43] 祝智庭,孙妍妍,彭红超(2017). 解读教育大数据的文化意蕴[J]. *电化教育研究*, (1): 28-36.
- [44] 祝智庭,余平(2017). 智慧城市教育公共服务评价指标体系研制[J]. *开放教育研究*, (6): 49-59.
- [45] 中华人民共和国教育部(2016). 《教育信息化“十三五”规划》[EB/OL]. [2017-05-21]. <http://www.moe.edu.cn/srsite/A16/>

s3342/201606/t20160622_269367.html.

[46]周丽芳,李伟生,黄颖(2013). 模式识别原理及工程应用

[M]. 北京:机械工业出版社:90-102.

[47]Zins, C. (2007). Conceptual approaches for defining data,

information, and knowledge[J]. Journal of the Association for Information Science and Technology, 58(4): 479-493.

(编辑:魏志慧)

Human-Machine Collaborated Data Wisdom Mechanisms: The Alchemy of Data Value for Smarter Education

PENG Hongchao¹ & ZHU Zhiting²

(1. Department of Education Information Technology, Faculty of Education, East China Normal University, Shanghai 200062, China; 2. School of Open Learning and Education, East China Normal University, Shanghai 200062, China)

Abstract: Data value is increasingly emphasized in education, like other industries. However, the density of this value is so low that it cannot serve education purpose directly, let alone serve the smarter education. Data wisdom refers to a class of models for representing structural and/or functional relationships between data, information, knowledge, and wisdom. The four levels of data wisdom provide a way for solving the above problems. But, the effective methods to make data evolve into wisdom through information and knowledge have not been found. Thus, we developed a kind of human-machine collaborated data wisdom mechanisms. It describes how to achieve the evolution of data into wisdom by using the understanding of both human and machine. Human understanding includes four layers: know by practicing, know by sensing, know by constructing and know by critiquing. Machine understanding also includes four layers: know by perceiving, know by describing, know by mining and know by learning. With the above-mentioned assumption, data wisdom mechanisms can be divided into three parts: the mechanism of the relational organization of data, the mechanism of pattern recognition and interpretation of information and the mechanism of principle derivation of knowledge. The first part is for data evolution into information which has four steps: a) introduce monitoring goals of teaching and students behavior, b) determine the relationship of monitored data, c) organize data based on determined relationship, d) and represent the meaning of organized data by using visual charts and dashboards. The second part is for information evolution into knowledge which also has four steps: a) extract the features of objects from organized data, b) inquiry meaning from visual charts and dashboards, c) mine information patterns from extracted objects' features via inquired meaning, d) and explain and evaluate the mined information patterns. The third part is for knowledge evolution into wisdom which is four steps: a) inquiry the reasons for the behavior and performance of students inspired by explained and evaluated information patterns, b) evaluate the value of student behavior and performance based on inquired reasons, c) learn the service criteria from b) and the service decisions from d) by machines, and d) generate learning service decisions based on the insight obtained from b). We hope that the proposed data wisdom mechanisms can provide a feasible scheme of extracting value from data for the smarter education, and then it can further help to optimize the teaching and learning behaviors in smarter education.

Key words: data wisdom; smarter education; data value; human-machine collaboration; precise decision making