

# 人工智能时代数据挖掘技术助力教育创新改革

## Data Mining Technology Assists Education Innovation and Reform in the Era of Artificial Intelligence

贾宁

Ning Jia

火箭军士官学校 中国·山东 潍坊 262609

Chinese People's Liberation Army Rocket Sergeant School, Weifang, Shandong, 262609, China

**摘要:** 伴随着教育信息化 2.0 时代的到来, 人工智能技术正加速融入高等教育教学活动, 院校教育也将迎来人机协同的智能时代, 必将会对传统的高等教育理念、教育方式产生重要的影响, 并加速教育供给侧结构性改革。

**Abstract:** With the advent of education informationization 2.0 era, artificial intelligence technology is accelerating into higher education teaching activities, and college education will also usher in an intelligent era of man-machine coordination, which will certainly have an important impact on traditional higher education concepts and education methods, and accelerate the structural reform of education supply side.

**关键词:** 人工智能时代; 数据挖掘; 教育创新改革

**Keywords:** era of artificial intelligence; data mining; educational innovation reform

**DOI:** 10.12346/csai.v2i1.9224

### 1 “大智移云物区”、数字经济与数字化转型

#### 1.1 “大智移云物区”与数字经济时代的到来

2020 年 11 月 20 日, 习近平总书记在亚太经合组织第二十七次领导人非正式会议上提出, 数字经济是全球未来的发展方向。根据国际数据公司 IDC 预测, 2025 年全球数据量将达 163ZB, 2019—2023 年大数据的市场收益复合年均增长率达到 13% 左右, 全球大数据产生的收益达到 3100 多亿美元。

数据已经成为创造和捕获价值的新经济资源。数据控制对于将数据转化为数字智能具有重要的战略意义。但事实上, 在全球“数据价值链”中, 价值和数据都集中在少数几个全球性平台和其他领先的跨国公司手中。很多国家只能成为这些数字平台的原始数据提供者, 同时不得不为平台所有者利用这些数据产生的数字智能付费。

联合国发布的《2019 年数字经济报告》中指出, 根据定义的不同, 数字经济的规模估计在世界国内生产总值的 4.5% 至 15.5% 之间, 其中中美两国合起来约占世界总量的

40%。也就是说, 在数字经济方面呈现出中美主导的“双寡头”格局。

《中国互联网发展报告 2020》中指出, 2019 年中国数字经济规模达 35.8 万亿元, 占 GDP 比重达 36.2%。数字经济结构持续优化升级, 产业数字化增加值占数字经济比重达 80.2%, 在数字经济发展中的主引擎地位进一步巩固, 向高质量发展迈出新步伐。

2020 年 9 月, 中国互联网络信息中心 (CNNIC) 发布的第 46 次《中国互联网络发展状况统计报告》中可以看到, 我国发展数字经济的基础非常雄厚, 网民规模达到 9.4 亿, 互联网普及率已经达到 67%, 城乡数字鸿沟显著缩小, 城乡地区互联网普及率差异为 24.1%, 成为精准扶贫和扶贫攻坚战的一个重要的抓手。此外, 我国形成了一系列基于互联网的商业模式, 比如在线教育用户规模达 3.81 亿, 在线医疗用户规模达 2.76 亿, 远程办公用户规模达 1.99 亿。这些网民慢慢形成了互联网的红利和互联网商业模式。

现在中国主要发展的是工业互联网, 5G 终端连接数已

【作者简介】贾宁 (1989-), 中国山东潍坊人, 硕士, 助教, 从事科学教育研究。

超过6600万，三家基础电信企业已开通5G基站超40万个，工业互联网领域已培育形成超过500个工业互联网平台。未来工业互联网会有进一步的发展，成为未来新的投资热点。

除此以外，网络视频的用户规模达到8.88亿，网络新闻的用户规模达到7.25亿，占网民整体的77.1%。可以看到，现在互联网大大改变了大家的生活。

之前大家认为发展比较滞后的在线政务公共服务，现在用户规模也达7.73亿，已经不再是短板。所以，中国产业链、价值链的完整性是世界上独一无二的，而且已经形成了O2O闭环和综合配套，为高质量发展和产业转型升级也打下了良好的基础。

在这种情况下，企业的发展越来越倚重于数字化的转型，根据IDC在2018年所做的调查，中国TOP1000大企业有70%以上把数字化转型作为公司战略核心，也就是说数字化转型变成了公司战略层面的东西，其重要性不言而喻。在2019年至2022年期间，IDC预测企业数字化转型支出将会超过1万亿美元，至2022年中国GDP的65%将与数据有关。数字化转型使资源要素重新配置，生产制造更加智能，专业分工更加精细，经济效益更加显著。企业通过数字化转型对价值链进行重新定位，并力求提升价值链地位。2017年10月，世界经济论坛发布了《第四次工业革命对供应链的影响》白皮书显示，数字化转型将使制造业企业成本降低17.6%、营收增加22.6%，使物流服务业成本降低34.2%、营收增加33.6%，使零售业成本降低7.8%、营收增加33.3%。也就是说，数字化转型起到降本增收的作用。

## 1.2 “大智移云物区”的发展态势

### 1.2.1 “大”和“移”

大数据和移动互联方面，2019年全球移动数据流量为190EB，全球大数据市场收入490亿美元，已经形成产业化。

### 1.2.2 “智”

人工智能方面，呈IBM、Google、Honeywell“三强争霸”，同时这三方在争夺“量子霸权”，截至2020年底，IBM有18台量子计算机，Google有5台，Honeywell有6台。中国在量子计算方面发展也很快。

### 1.2.3 “云”

2019年云存储数据量首次超过设备存储，这意味着人类越来越依赖云存储。

### 1.2.4 “物”

根据《2020年物联网模组行业研究报告》，预计2020年物联网连接数为281亿，收入规模高达7.1万亿美元。根据梅特卡夫定律，网络节点增长带来网络价值平方增长，同时单节点价值增长，进一步带来节点数增长。

### 1.2.5 “区”

根据Gartner发布的2020年区块链报告，区块链在促进信息互联网向价值互联网变革过程中的重要价值已受到普遍认同，核心技术框架已基本形成。2020年4月，区块链

作为基础设施被纳入“新基建”，5月，央行数字货币内测及试点，10月，央行数字货币正式在深圳落地。

## 1.3 中国对“大智移云物区”与数字经济的政策推动

2017年7月20日，国务院印发《新一代人工智能发展规划的通知》（国发〔2017〕35号）。该文对新时代下数字经济发展的环境作了系统的阐述，核心体现在“五新”上，即人工智能会催生新技术、新产品、新产业、新业态、新模式。2018年3月5日，《政府工作报告》提出，“实施大数据发展行动，加强新一代人工智能研发应用”。2018年12月，中央经济工作会议把5G、人工智能、工业互联网、物联网定义为“新型基础设施建设”。2020年4月20日，国家发改委进一步明确，“新型基础设施建设”包括信息基础设施、融合基础设施和创新基础设施三方面。具体包括5G基建、特高压、城际高速铁路和城市轨道交通、新能源汽车充电桩、大数据中心、人工智能、工业互联网。

数字经济的核心要素就是数据要素。因此，2020年5月11日发布的《中共中央国务院关于新时代加快完善社会主义市场经济体制的意见》提出，加快培育发展数据要素市场，建立数据资源清单管理机制，完善数据权属界定、开放共享、交易流通等标准和措施，发挥社会数据资源价值<sup>[1]</sup>。

2020年10月29日，党的十九届五中全会通过的《中共中央关于制定国民经济和社会发展第十四个五年规划和二〇三五年远景目标的建议》中提出，“发展数字经济，推进数字产业化和产业数字化，推动数字经济和实体经济深度融合，打造具有国际竞争力的数字产业集群。加强数字社会、数字政府建设，提升公共服务、社会治理等数字化智能化水平。建立数据资源产权、交易流通、跨境传输和安全保护等基础制度和标准规范，推动数据资源开发利用。扩大基础公共信息数据有序开放，建设国家数据统一共享开放平台。保障国家数据安全，加强个人信息保护。提升全民数字技能，实现信息服务全覆盖。积极参与数字领域国际规则和标准制定”。

2020年12月，国家发改委、中央网信办、工信部、国家能源局联合印发《关于加快构建全国一体化大数据中心协同创新体系的指导意见》。这是中央政府职能部门发布的第一个关于加快构建全国一体化大数据中心协同创新体系的正式文件，标志着中国进入数字经济时代。

## 1.4 “大智移云物区”重新定义了贸易

比如数字服务贸易、电子商务等，也就是说，数字贸易逐渐纳入全球的多边贸易体系。同时，“大智移云物区”也重新定义了货币。比如2020年10月，中国央行数字货币正式落地。数字货币从理论上讲可以提高交易的便利程度，降低交易成本，减少信息不对称。

## 2 人工智能与数据挖掘

### 2.1 人工智能与大数据挖掘效果

人工智能是大数据挖掘的利器，它能解决大数据、大容

量、大速度的问题。因为人类智能的内存有限，运行的速度有限。而人工智能能大大弥补人类智能在这两个方面的短板。所以从理论上讲，它是进行数据挖掘很好的助力。但实际上，利用人工智能炒股效果却非常不好。

2017年10月18日，全球第一只应用人工智能进行投资的ETF基金——AI Powered Equity ETF（代码：AIEQ）在纽交所上市。其使用IBM的Watson系统的认知和大数据处理能力，来分析美股的投资机会。AIEQ身上有很多让人类羡慕的特性，比如365天每天24小时始终在工作；同时分析6000多只美国挂牌股票；每天分析上百万条相关的公告文件、财报、新闻及社群文章；利用量化择时、量化选股、因子分析、事件驱动等N种量化模型选股；不断地深度学习，数据处理得越多，选股能力就会更强大等。

但实际效果怎么样呢？截至2020年11月，AIEQ的股价仅上升了42%左右，而美股纳斯达克指数大盘上升了接近79%，也就是说，它跑输了30多个点，所以它的股价上涨远远低于大盘的上涨。

那么为什么人工智能下象棋、围棋很好，炒股却一塌糊涂？因为机器人没有办法像人类一样，对一些定性的主观的事物进行判断。

## 2.2 人类智能与人工智能的边界

人类智能最后的领地，第一是感情，但这点现在已经被突破了，机器可以通过相应算法训练感情；第二就是职业判断。人工智能和人类智能的边界在不同的数据挖掘的路径中是不一样的。比如，基于信息本身的平面挖掘，肯定人工智能是强势的。但如果是基于传导机制的链式挖掘，即前后关系变成因果关系的判断，人工智能的优势就会弱化很多。

## 3 人工智能对教育供给侧结构性改革

教育供给侧结构性改革对人工智能方面影响最大的是会计与金融的融合，这个融合体现在培养高位人才的一些要求，新复合型人才既要懂人工智能、懂科技，又要懂金融、懂战略，归纳为十六个字，即“会计为表、金融为里，战略为本，科技为气”。也就是说，现在企业会计编报表越来越次要，分析报表越来越重要。并且除了对数据进行挖掘，对公司战略的制定和风险管理也越来越重要。在此背景下，2018年教育部开始对人工智能学科和专业建设加快了推进。现在中国的人工智能整体实力已经领跑全球了，人工智能论文发表数量全球第一，企业数量、融资规模全球第二，但是美中不足的是论文的影响力指标稍微落后一点。

2020年2月21日，教育部公布2019年度普通高等学校本科专业备案和审批结果，新增备案专业数量最多的人工智能，共有180所高校新增，其他新增的部分专业与人工智能相关度较高。说明人工智能专业以及交叉学科成为我们教育供给侧改革的主流方向和主要实施路径。

随着产业数字化带来的数据基础的日益成熟，人工智能

在营销、金融、数字政府、零售、医疗等行业落地持续推进。如何产业落地成为人工智能的主要命题。人工智能技术无法直接解决业务需求，需根据具体业务场景和目标，形成可规模化落地的产品与服务。人工智能在数据、算法业务场景理解、服务方式、产出投入等方面面临挑战。

针对中国人工智能产业发展瓶颈，《新一代人工智能发展年度报（2019—2020）》建议：充分利用新基建机遇，协调推进各类数据中心、5G网络部署，全面提升端侧的数据计算、采集及传输能力，为传统产业全面向数字化转型打造坚实的泛在计算基础。

加快中国开源开放生态体系研发与应用推广，建设可跨硬件和算法平台、兼容国内外主流开源平台的开源开放体系。

鼓励人工智能产业人才引入、知识产权质押融资、贸易便利化等领域政策先行先试，加快融入国际产业链。

加强对人工智能专业学科指导建设，按照“厚基础”“重交叉”“宽口径”思路，培育一批专业技能扎实、科学素养高、动手实践能力强、具备开阔产业应用视角和国际前瞻视野的人才。鼓励高校内部整合或者多个高校联合形成人工智能学科群，培育更多符合产业高质量发展所需的复合型人才。

## 4 以数据驱动助力教育改革创新改革框架

### 4.1 自我修正的决策能力

人工智能教育大脑遵循统一的数据规范以及互通融合的交互机制，通过挖掘分析多源头、多类型、多情境的行为数据，努力克服“主观经验”和“零散片乱”等弊端，为教育治理决策与教学创新实践提供更加系统科学的证据支持。同时，人工智能教育大脑自我修正的决策能力可以保障教育数据全域互联、智能感知、数据开放以及融合应用等，有利于洞察、解构和关联学习者的学习状态以及教师的教学行为，实现动态化、精准化、个性化的教学诊断和有效干预目标。这也与已有实证研究结论“基于科学决策的智能决策能够提升教师的数据素养”相一致。需要强调的是，人工智能教育大脑“用证据说话、让数据动起来”等特征，并非唯数据论，而是通过多种渠道发挥数据驱动教育的最大化价值，推动智能技术与教育生态系统健康有效地融合。

### 4.2 融合统一的协同机制

教育治理或教学实践过程动态变化且复杂多样，需要依赖多元融合的证据支持。人工智能教育大脑其融合统一的协同机制，是破解教育系统功能割裂或数据碎片化等不良现象的重要利器。

一方面，数据标准由“分散多样”转向“融合统一”，打通行为表征、数据采集、诊断评估等立体性、动态化过程，挖掘教育数据的真正价值，为教育决策和数据治理提供切实可行的应对措施。比如，美国预测分析报告（Performance Analysis Reporting，简称PAR）参照学习者人口学数据、教

学管理数据、学习过程数据、成绩数据以及财务信息数据等综合视角,定制化地构建辍学风险因素模型。迄今为止,PAR已经帮助351所学校有效降低了辍学率。

另一方面,数据中枢由“碎片式整合”转向“网状化融合”,通过汇聚多元化的数据类型,打通数据之间的流通机制,有利于把控数据的内在机理和规律,提高“学生画像”和“教师画像”的精确度。与之类似,已有研究从横向分工网状化以及纵向结构扁平化等方面精准把控数据的传输过程,加强多模态数据流的融合互通,推进大规模数据协同的开放共享<sup>[2]</sup>。

### 4.3 因材施教的教育赋能

人工智能教育大脑为因材施教的教育赋能提供了有力支持。首先,汇聚智能环境下多模态教育数据,建立标准化、统一化的数据交互基准,保障物理空间与数字空间的全时全域全要素映射,实时采集互通融合的多元化数据。其次,通过表征关联学习者的外显行为和内隐特征,能够建构有效行

为发现模型。再次,运用分类算法、聚类算法、时间序列等方法深度挖掘学习者的行为特征或能力水平,诊断学习者的过程性行为 and 异常状态等。最后,通过对学习者实时动态画像以及因素挖掘,发现单体和群体素养指标薄弱项,并采用多重匹配算法,提取适合个体素养发展的干预方案或优化策略。显然,人工智能教育大脑贯穿教育实践过程的始终,不仅强调多元化的数据来源以及统一规范的数据交互规则,也非常重视行为数据过程性诊断、最佳干预方案选择、实践效果分析等过程的科学性与合理性<sup>[3]</sup>。

### 参考文献

- [1] 崔娟.计算机网络技术中人工智能的应用探讨[J].电脑编程技巧与维护,2020(8).
- [2] 徐阳.大数据时代下人工智能在计算机网络技术中的应用[J].电视技术,2023(4).
- [3] 黄建.人工智能产业发展对教育行业的影响与挑战——以ChatGPT为例[J].经济界,2023(2).