

基于大数据的牛顿平台自适应学习机制分析*



——“教育大数据研究与实践专栏”之关键技术篇

万海鹏 汪丹

(北京师范大学 教育技术学院, 北京 100875)

摘要: 个性化和自适应学习系统是教育大数据应用服务的主要阵地, 自适应学习系统能够采集学习过程中的行为数据, 并对学生的学习兴趣、知识水平、学习风格、学习进度等做出分析和预测, 以提供个性化的学习服务。近年来被学术界和企业界所广泛认可的典型自适应学习平台——牛顿平台(Knewton Platform)正逐步兴起, 文章从自适应原理、核心技术、自适应服务三个方面对牛顿平台进行了剖析, 以期能够为教育大数据分析研究人员和自适应学习平台设计者提供理论参考和技术借鉴。

关键词: 教育大数据; 牛顿平台; 自适应学习; 知识图谱; 项目反应理论

【中图分类号】G40-057 【文献标识码】A 【论文编号】1009—8097(2016)05—0005—07 【DOI】10.3969/j.issn.1009-8097.2016.05.001

引言

随着大数据理念的深入人心, 教育大数据呈“爆炸式”增长之势。教育大数据产生于各种教育活动, 与传统教育数据相比, 教育大数据的采集具有更强的实时性、连贯性、全面性和自然性^[1], 其分析处理和应用服务更加多元化、智能化、个性化。学生和教师是教育领域的两大核心主体, 教育大数据通过用数据说话使个体真正认识自己, 将真正推动教学和学习的个性化^[2]。在教育大数据的支持下, 教师可以关注到每个个体的学习行为和学习轨迹、预测学习结果、诊断学习需求和问题, 从而基于数据开展因材施教; 学生可以了解到自己的偏好、知识缺陷、能力缺陷、发展目标等, 并能够获取适合自己的学习资源和学习指导服务。

自适应学习系统是教育大数据采集、分析和提供个性化自适应服务的重要载体。Brusilovsky^[3]于1996年对自适应学习系统进行了分类和总结, 提出了自适应学习系统(Adaptive Educational Hypermedia Systems, AEHS)的通用模型, 包括领域模型(Domain Model)、用户模型(User Model)、适应模型(Adaptive Model)、自适应引擎(Adaptive Engine)四个模块——领域模型由领域内的概念及其关系组成; 用户模型是对用户特征信息的抽象描述, 包括知识结构状态、学习目标、背景经历、认知风格、学习偏好和学习成绩; 适应模型定义了进行内容呈现自适应、链接导航自适应以及用户模型更新的规则; 自适应引擎主要用于响应用户请求, 根据用户模型来选择、组装和呈现页面, 根据用户与系统的交互记录来更新和维护用户模型等^[4]。

近期以来, 一个名为“Knewton”的自适应学习平台逐渐被人们所熟知, 它集数据科学、统计学、心理测量、内容绘图和机器学习于一身, 旨在最大限度地实现个性化。作为“教育大数据研究与实践专栏”的关键技术篇^[5], 本文将牛顿平台(Knewton Platform)为例, 从自适应原理、核心技术、自适应服务三个方面对其进行阐述和分析。

一 牛顿平台的自适应原理

自适应学习强调学习环境的适应性, 要求创设的个性化学习环境能够最大限度地适应学习

者的不同特征,并以此来开展个别化学习和针对不同能力的学习者进行“因材施教”^[6]。在牛顿平台看来,自适应学习系统应保持适应的持续性,可以对学生的学习表现和活动完成质量给予及时反馈,以在正确的时间基于正确的内容提供合适的学习指导,来最大化学生获得学习内容的可能性;同时基于给定活动的完成情况,自适应学习系统应能持续性地逐步引导学生进入下一个活动。

为了保持持续自适应,即在任何时刻都能为学生做出个性化的学习进度安排,牛顿平台进行了概念层面的专业化数据(Proficiency Data,如知识概念掌握程度、学习投入程度、学习效率、优势劣势、活跃时间、预测分数等)采集处理,并建立了专业化数据与学习过程数据之间的关联映射。专业化的数据模型不仅能评估学生做了什么,还能在概念层面分析学生掌握了什么以及学生的学习就绪状态、认知投入、学习偏好、学习风格、学习策略等,并向学生呈现为下一步学习或评估所应该做的准备以及能力随时间变化的可视化图示。具体来说,牛顿平台的这种持续自适应主要体现在空间强化(Spaced Reinforcement)、记忆力和学习曲线(Retention and Learning Curves)、学生学习档案(Student Learning Profile)等方面^[7]。基于教育路径规划技术和学生能力模型,牛顿平台构建了自适应学习的基础框架,以最大程度地实现个性化。

1 基础结构

牛顿平台构建了一个基于规则、算法廉价的大规模规范化内容的基础设施(Heavy Duty Infrastructure)^[8],包括数据基础设施(Data Infrastructure)、推理基础设施(Inference Infrastructure)、个性化基础设施(Personalization Infrastructure)三部分。

数据基础设施部分主要负责收集、处理海量的专业化数据,涉及用于规划和管理各个概念之间关系的自适应本体(Adaptive Ontology),以及用于实时流和并行分布式流数据预处理的模型计算引擎(Model Computation Engine)。自适应本体是一组具有直观和可拓展性的概念对象及其关系的集合,这些概念和关系容易习得,且能很方便地用于表达学习内容之间的关系,为数据分析和自适应辅导提供基础支撑。模型计算引擎采用分布式的方式进行工作,能够将任务分解为细小的计算单元,以通过多台电脑实现高效的并行计算;而当某台电脑出现异常时,另外的电脑也能够及时取代并在任何状态下进行恢复。

推理基础设施部分的目标在于扩大数据集和从收集的数据中形成视图,包括心理测验引擎(Psychometrics Engine)、学习策略引擎(Learning Strategy Engine)和反馈引擎(Feedback Engine)。其中,心理测验引擎负责评估学生的概念掌握程度、内容参数、学习效率等,并通过推理的方式来扩充学生的数据集(包括挖掘学习偏好、认知风格、知识结构、能力水平、学习进度等),最终形成能综合表征学习者学习状态的信息档案面板;学习策略引擎主要用于评估学生对教学、评估、进度安排等方面变化的敏感程度,识别学生在学习过程中对学习资源、学习环境等改变做出的反应,并据此为学生选择合适的学习策略,如提供符合学习者认知风格的学习资源和导航、提供符合学习者学习水平的测评方案等;反馈引擎负责对数据和反馈结果进行归一化处理,并将它们返回到自适应本体库中,以丰富自适应本体的元数据信息,使知识概念与学生的学习过程信息之间建立更符合实际且可用的关联,进而提高推理和分析的精准性。

个性化基础设施部分主要利用所有合并数据所构成的整体网络为学生寻找最优的学习策略,包括推荐引擎(Recommendations Engine)、预测分析引擎(Predictive Analytics Engine)和归一化学习轨迹(Unified Learning History)。其中,推荐引擎负责从目标均衡性、学生的优劣势、投

入程度三个方面,为学生提供下一步操作的排序建议;预测分析引擎负责对学生的考量作预测,如达到教学目标的速率及完成程度、考试分数、概念的熟悉程度等;归一化学习轨迹的目的在于统一学生的个人账户,建立学生在不同学习应用、学科领域和学习时段与先前学习经验之间的关联,避免个性化推荐应用中常遇到的“冷启动”问题。

2 数据模型

数据模型(Data Model)是对现实数据特征的抽象,用于描述一组数据的概念和定义。对牛顿平台而言,数据模型是数据在系统中的存储方式,包括四部分内容^[9]:

(1) 知识图谱(Knowledge Graph)

知识图谱是概念与概念之间关系的集合,是牛顿平台用于精准定位学生学习状态的重要方式,其结构如图1所示。其中,圆圈代表概念;连线代表各概念之间的关系;箭头指向表示前一个概念是后一个概念的先修概念,二者之间是先修关系(Prerequisite Relationships)。牛顿平台的知识图谱是通过自适应本体来建立的,具有可扩展、可伸缩、可测量的特性。自适应本体由模块(内容片段)、概念、内容与概念之间的关系三种元素构成,其关系类型主要有:包含(Containment),表示该内容或概念从属于更大的群组;评估(Assessment),表示该内容提供了学生掌握状态的信息;教学(Instruction),表示该内容在教授某个具体的概念;先修(Prerequisiteness),表示学习该概念之前需要先掌握另外一个层级更低的概念^[10]。基于自适应本体,研究者和教师就能对典型的课本内容进行概念映射和标注。利用这种标注好的课本内容数据,结合学习交互数据、心理测验数据,牛顿平台就能自动为学生生成个性化的知识图谱。



图1 知识图谱(源于《牛顿平台技术白皮书》)

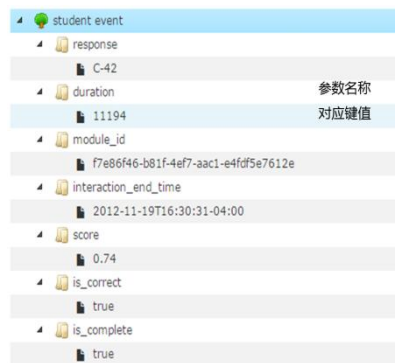


图2 学生响应事件数据结构示意图

(2) 学生事件(Student Events)

学生事件是学生与学习内容交互时产生的系列数据流,主要用于对学生的能力进行实时推断。牛顿平台收集来自不同合作伙伴产品中所生成的交互数据流信息,用于为学生的个性化分析与推荐作支撑。学生响应事件数据的存储与交换格式,包括试题编号、作答持续时间、试题所属模块、交互结束时间、得分、正误状态以及完成状态,其数据结构示意图如图2所示。

(3) 目标管理(Goal Management)

目标管理是对学生学习结果数据的分析和处理。牛顿平台能够为学生提供可持续更新的学习目标档案,档案内容包括学生未学习的内容、已学习的内容、知识概念掌握的状态水平、成绩排名以及如何学得更好的推荐信息。随着学生使用平台的时间变长,档案将会变得越来越智

能。目标档案数据的存储与交换格式，包括知识概念名称、所属模块、目标分数、开始时间、目标时间、推荐模块候选集、模块推荐数量等。

(4) 推荐与分析 API (Recommendations and Analytics API)

推荐与分析 API 作为学习者个性内容推荐与分析服务的接口，能为学习者持续提供内容推荐，并在学习进度、概念熟练程度、学习投入等方面进行精准推断。个性推荐与分析诊断数据的存储与交互格式，包括推荐模块、学习案例、目标模块、预期分数、置信区间、评估时间等。

二 牛顿平台的核心技术

1 项目反应理论 (Item Response Theory, IRT)

项目反应理论将学习者对测试项目的反应（应答）通过表示测试项目特性的参数和被测试学习者能力的能力参数及其组合的统计概率模型来表示，其中表示项目特性的参数主要有难度系数和区分度^[11]。传统的项目反应理论一般针对问题、项目来设计相关参数，且运用过程中通常存在两大误区：一是认为学生的能力是个常量，二是倾向于用一个参数来表示学生的能力。

考虑到能力的发展变化以及多种能力之间的相互连接，牛顿平台对传统的项目反应理论进行了扩展，并从问题层级的表现来对学生的能力建模——认为学生的能力参数会随时间而变化；同时，对学生能力的表征不再局限于某个唯一的参数，而是通过利用聚焦于概念层面的知识图谱来对学生能力进行评估和表征。

2 Knewton API

Knewton API^[12]是连接应用场景与合作平台的桥梁，以云服务的方式被第三方企业调用，如图 3 所示。

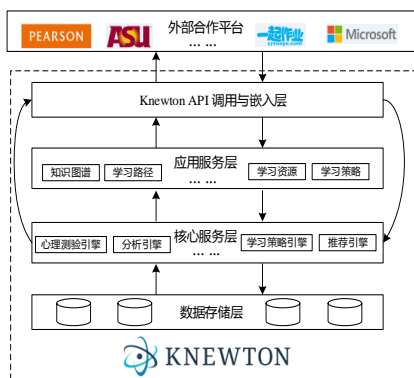


图 3 牛顿平台与第三方应用集成框架

其中，核心服务层负责与牛顿平台的数据库打交道，并以表单的方式向应用服务层提供预处理后的数据信息，其中典型的服务就是知识的图谱化工作。基于本体库，图服务能时时进行图式化内容的信息更新，并结合实际需求对图谱进行基于历史版本的改造。应用服务层负责与推理引擎（包括心理测验引擎、推荐引擎等）进行对接，而这些引擎的正常工作都有赖于核心服务层所提供的可直接利用的数据。API 调用与嵌入层则负责收集来自合作企业平台中产生的系列信息，并根据需要以消息的方式通知系统中的其它服务层。例如，当有内容需要加入知识图谱或学生注册了一门课程，API 调用与嵌入层接收这些信息后，便立即通知相应的核心服务

层进行响应，并在数据存储层进行存储。具体说来，Knewton API 能为合作企业提供下述三个层面的服务：

①学习历史记录层面，Knewton API 采集了学习者历史学习过程中所表现出的一系列学习偏好和差距，可帮助学生保持在新课程中的积极性。学习历史记录档案包括学生所知道的内容、掌握的水平、学得最好的课程、如何学得最好的推荐信息，并能持续性地进行更新。

②学习交互数据分析层面，Knewton API 能将海量数据转变为认知交互模型、估计向量、数据框架和可人为操控的视图，并向教师、家长、管理者和学生提供深层次的教学和内容分析报告——教学分析指标包括熟练程度、就绪状态预测、分数预测、活跃时间，基于该指标，教师可以在更加准确知晓学生缺点的前提下指导学生，年复一年地比较课程数据，按学期、按年度进行课程的改进和完善；内容分析可以帮助教师、出版者和管理者确定教学材料中最丰富和最薄弱的部分、需精细讲授和评估的部分，保证内容的持续更新，确保学生不会使用过时的教材。

③个性化推荐层面，Knewton API 通过综合考虑内容要素、学习者要素和目标要素来决定对下一步所应学习内容的推荐。其中，内容要素包括模块关系、教育意义、评估价值、问题难度、持续时间和学习投入程度，学习者要素包括概念熟悉程度、评估需求、复习需求、学习步调和材料重复接受度，目标要素则包括目标模块、目标分数、达标日期、可推荐模块。

三 牛顿平台的自适应服务

1 差异化引导的自适应学习过程

牛顿平台提供差异化的学习辅导服务，即利用项目反应理论对学生的学习状态进行测试评估，基于学生问题层面的表现而不是整体测试成绩来对学生的能力进行建模。对于理解不同问题所带来的贡献，项目反应理论并没有同等看待，而是针对每一个问题提供了包含问题信息和答题者个人能力信息的贡献计算解释。下文将以一个差异化引导任务为例，来阐述牛顿平台的实时自适应学习过程^[13]。

如图 4 所示，Amy、Bill 和 Chad 三位学生有同样的学习目标——理解乘法公式、一位数乘法、两位数乘法、解决乘法应用题；这四个概念的先修知识分别是乘法符号、理解乘法、100 以内的整数乘法、用乘法解决问题。比如，要理解“两位数乘法”必须先理解“一位数乘法”；下方排列的小图形代表每位同学答的题目及正误信息，每道题对应的图形与它们所属的知识点框中的图形类型一致，图形的填充与否代表正误信息，实心表示正确，空心表示错误。

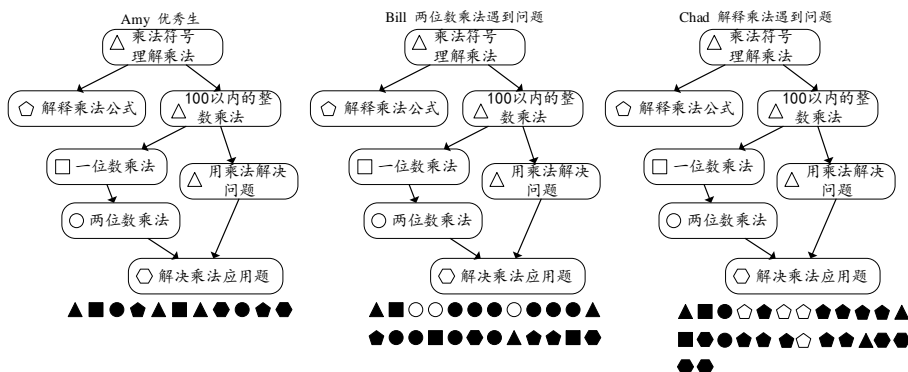


图 4 同一目标不同学生的自适应学习过程

从图4可以看出,这三位学生所答的前三道题目是一样的,由于第三道题Bill答错了,与其他两位同学出现了不同的学习状态,故三位同学开始呈现出不同的学习路径——Bill在理解“乘以两位数”时遇到了困难,故继续回答与这个主题相关的题目,而Amy和Chad进入到一个主题;从第四题的回答结果来看,Amy回答正确继续完成接下来的题目,而Chad回答错误继续回答与“理解乘法公式”这个主题相关的题目。图4展现了三位同学为达到同样的学习目标而进行的自适应学习过程,从中可以看出牛顿平台的差异化指导有助于学生更多地关注自己的薄弱环节,而不至于在已经掌握的环节上做无用功——平台引导那些学习困难的学生(如Bill和Chad)继续回答与问题主题相关的题目,直到他们理解、做对题目进而掌握概念;对于那些掌握程度较好的学生(如Amy),牛顿平台则向其提供按照自己步调学习的机会。

2 创建自适应课程

牛顿平台支持教师、家长及学生创建自适应课程,每门课都由创建者自由选择的若干个任务组件构成,每个任务组件内包含一定数量的题目,并以上述方式为学习者提供不同的学习路径。为自适应课程选择任务组件有两种方式:①接受推荐。牛顿平台基于知识图谱和用户教学行为数据分析,向用户推荐其可能感兴趣的组件,而用户可将推荐的组件加入到创建的个性化课程里。②根据目录树选择任务组件。牛顿平台已经内置大量的任务组件,以学科—一年级—主题—子主题等任务多维关键词表征,用户可以通过多维搜索的方式来主动选择组件以创设个性化课程。比如,数学学科在六年级水平上有代数I和代数II、数据分析和概率统计、几何和三角函数、数和计算四大主题;在大主题下又有子主题,如数和计算这一主题包含数的概念、复数、比率等子主题;每个子主题又对应一个任务组件。

课程创建完成后,用户可通过站内信和邮件的形式邀请学生加入、激活课程并参与学习。同时,用户还需要填写课程名称和课程详细描述、关联K-12课程大纲和任务完成时间,以方便其他用户查找和使用课程。牛顿平台将跟踪这门课程所有学习者的学习动态,向用户报告学生的基本信息和总体任务的完成情况,包括学习进度、学习困难、任务完成情况统计等信息,一方面方便教师和家长掌握学生动态,另一方面也利于用户进一步组建个性化的课程。

四 结语

牛顿平台开创了教育大数据个性化服务设计和应用的先河,通过采集学生的在线学习数据,可精准分析和预测学生的优势、不足、学习兴趣、认知投入水平。牛顿平台正与合作企业在学习内容提供、学习过程管理、学习产品分发以及市场推广应用等方面开展深入合作,颇具规模的在线教育生态系统圈已现雏形。牛顿平台目前所提供的自适应功能在很大程度上满足了学生、教师、家长以及学校管理者的需求,为学习、评估和管理带来了极大便利。然而,本研究发​​现牛顿平台的自适应数据主要源于试题解答记录,与本研究所期待的全学习过程数据还存在一定差距,故后续的研究还需对数据予以进一步的优化和改进。此外,自适应学习中有关用户模型、位置模型、设备模型和情境模型等方面的内容^[14],也需引起研究者的重点和持续关注。

参考文献

[1]杨现民,王榴卉,唐斯斯.教育大数据的应用模式与政策建议[J].电化教育研究,2015,(9):54-61、69.

- [2]姜强,赵蔚,王朋娇,等.基于大数据的个性化自适应在线学习分析模型及实现[J].中国电化教育,2015,(1):85-92.
- [3]Brusilovsky P. Methods and techniques of adaptive hypermedia[J]. User Modeling and User-adapted Interaction. 1996,(2):87-129.
- [4]陈品德,李克东.适应性教育超媒体系统——模型、方法与技术[J].现代教育技术,2002,(1):11-17.
- [5]杨现民,唐斯斯,李冀红.教育大数据的技术体系框架与发展趋势——“教育大数据研究与实践专栏”之整体框架篇[J].现代教育技术,2016,(1):5-12.
- [6]陈仕品.适应性学习支持系统的学生模型研究[D].重庆:西南大学,2009:10.
- [7][8]Knewton Company. Heavy duty infrastructure for the adaptive world[OL].
<<https://www.knewton.com/assets-v2/downloads/knewton-intro-2014.pdf>>
- [9][12]Jaffe J. Here is the deck for presenting adaptive learning in the E-education world[OL].
<https://lists.w3.org/Archives/Public/www-archive/2015Jan/att-0004/Knewton_W3C_presentation_V2.pdf>
- [10]Wilson K, Nichols Z. Knewton technical white paper[OL]. <<http://learn.knewton.com/technical-white-paper>>
- [11]傅德荣,章慧敏,刘清堂.教育信息处理(第2版)[M].北京:北京师范大学出版社,2012:120.
- [13]Green-Lerman H. Visualizing personalized learning[OL].
<<https://www.knewton.com/resources/blog/adaptive-learning/visualizing-personalized-learning/>>
- [14]贾积有,马小强.适应性和个性化学习系统研究前沿——与国际著名教育技术专家金书轲教授对话[J].中国电化教育,2010,(6):1-5.

Investigation on Adaptive Learning Mechanism of Big Data based on Knewton Platform

——The Key Technology of “Research and Practice Column about Big Data in Education”

WAN Hai-peng WANG Dan

(School of Educational Technology, Beijing Normal University, Beijing, China 100875)

Abstract: Personalized and adaptive learning system is one of the vital applications of big data in education, which can collect learning behavior data and provide analysis and prediction on students' learning interest, knowledge level, learning style and learning progress. A typical adaptive learning system——Knewton Platform which has been widely recognized by academia and enterprise in recent years is rising gradually. A deep analysis on Knewton Platform, from the aspects of adaptive mechanism, core techniques and adaptive service, was made to provide some theory and technology references for the researchers of big data in education and designers of adaptive learning platform in the future.

Keywords: big data in education; Knewton platform; adaptive learning; knowledge graph; item response theory

*基金项目: 本文为北京师范大学教育学部学生科研基金资助项目“泛在环境下自适应在线学习模型的设计与实现——以学习元平台为例”(项目编号: 15秋-03-01)的阶段性研究成果。

作者简介: 万海鹏, 在读博士, 研究方向为学习知识地图、移动与泛在学习、在线学习评价、计算机教育应用, 邮箱为 dnhp@163.com。

收稿日期: 2016年2月20日

编辑: 小米