

自适应学习平台的关键技术与典型案例

□文 / 万海鹏



万海鹏

理学博士，现为首都师范大学教育学院讲师，本硕博均就读于北京师范大学教育学部教育技术学专业，学习期间受国家留学基金委资助前往多伦多大学教育研究院学习一年，主要从事计算机教育应用、适应性学习方面的研究，已发表学术期刊和国际会议论文数十篇，已授权国家发明专利两项，主持或参与省部级项目多项。

任何科学技术的发展进步都是为人类服务的，随着人工智能技术的迅猛发展，如何用人工智能技术来变革传统的教育模式成为业内人士研究的热门问题。其中，作为“AI+教育”中的一个重要应用领域，自适应学习由于能够针对不同的学习者提供不同的解决方案，为“因材施教”的个性化学习提供了新的可能。本文将以此Knewton的自适应学习平台为例，阐述其中的关键技术与典型案例。

自适应学习长期以来一直是教育研究领域所关注的话题，也是“人工智能+教育”的一个重要应用领域。自适应学习平台最早可追溯到20世纪90年代，Brusilovsky[1]对其进行了系统化的分类和总结，提出了自适应学习系统（也称AEHS，Adaptive Educational Hypermedia Systems）的通用模型，包括领域模型、学生（用户）模型、教学模型以及自适应引擎四个模块。同时，Brusilovsky认为基于用户模型（User Model）来实现自适应是AEHS的核心，而用户模型的构建与更新是关键。其中用户模型主要涉及学习目标、知识结构、用户背景、学习经历、用户偏好五个方面，而自适应则主要表现在导航和呈现两个方面。

美国教育技术办公室从学习分析和数据挖掘的视角描绘了自适应学习系统的通用组件和数据流模型，如图1所示，主要涉及学习内容、自适应引擎、干预引擎、预测模型、仪表盘以及学生学习数据和学生人口信息数据，并包含五个数据流过程。其中，学习内容部分主要用来进行内容的管理、维护以及传送，通过与学生的相互作用来支持学生学习的个性化科目内容和评价方案；学生学习数据（或其他大数据存储库）主要用于存储做了时间标记的学生输入数据，以及学生在操作系统时所捕捉到的行为数据；预测模型部分是将学生人口数据（从一个外部的学生信息系统中获得的）和学生学习数据中的学习/行为数据进行整合，跟踪学生的学习进展并对其未来的行为和表现进行预测；仪表盘是利用预测模型的输出，并结构化地生成学习结果“仪表盘”，并为各类用户提供可视化反馈；自适应引擎则主要基于预测模型的输出调整内容传送组件，并根据学生表现的优劣及兴趣大小来传送内容，以保证持续的学习质量的提升；干预引擎则允许教师、管理者或者系统开发者干预甚至推翻整个自适应学习系统，从而为学生的学习提供更好的服务。

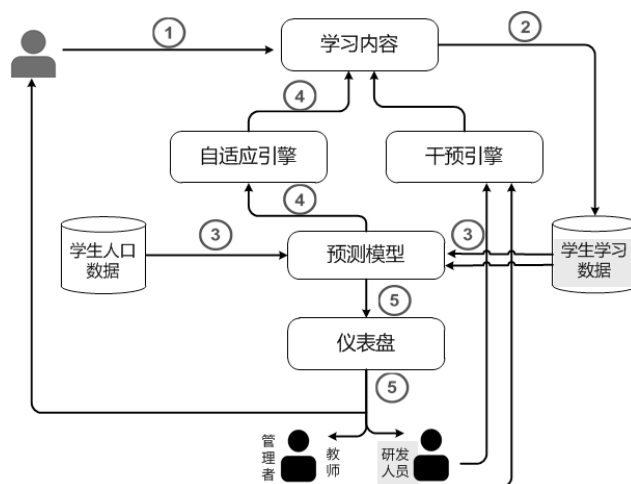


图 1：典型自适应学习平台组件及数据流

在上述自适应学习框架的基础上，随着信息技术的发展，自适应学习的理论及其核心技术的实现获得了长足的发展和进步。下面将以美国的 Knewton 作为自适应学习平台的典型案例，从数据驱动的学习服务要素和核心技术两个方面对其实施自适应学习的关键技术和方法进行详细介绍 [3]。

[1]Brusilovsky, P. Methods and techniques of adaptive hypermedia[J]. User modeling and user-adapted interaction, 1996, 6(2-3), 87-129.

[2]U.S. Department of Education, Office of Educational Technology, Enhancing Teaching and Learning Through Educational Data Mining and Learning Analytics: An Issue Brief, Washington, D.C., 2012.

[3] 万海鹏, 汪丹. 基于大数据的牛顿平台自适应学习机制分析 [J]. 现代教育技术, 2016,(05):5-11.

一、数据驱动的学习服务要素

Knewton 是一家自适应学习基础服务提供商，其所研发的自适应学习平台可以帮助学校、出版商和技术开发者为学习者提供个性化的教育内容。Knewton 提供的服务主要基于对学习者的分析，其基本要素主要包括数据的收集和处理、基于数据的推理以及个性化推荐。其中数据的收集和处理主要实现学习者在线学习过程中测评结果、学习行为、个体档案等信息的个性化采集，是对学习者在线学习需求进行归纳和分析的基础，该过程中分布式的数据处理引擎至关重要；基于数据的推理主要实现了以学习者兴趣、偏好、知识结构、知识掌握、学习绩效、心理特征等方面为推理特征，以学习者需要的学习服务（包括学习情境、策略、内容等）为输出特征的推理过程，该过程的实现依赖于大规模学科知识图谱和推理引擎；个性化推荐服务不仅基于个体的特征对学习者进行内容的单方面推荐，还可以构建更加均衡的推荐服务引擎，整合全平台数据，通过分析个体的均衡、优劣势、学习投入度等方面，挖掘学习者最佳的学习目标、学习内容、推进策略、学习路径等，形成对学习者全方位、立体化的学习辅助。

上述服务实现过程中，建立了如图 2 所示的数据模型，包括知识图谱、学生事件、目标管理和推荐分析 API。

1. 知识图谱：知识图谱是 Knewton 平台进行个性化学习服务的基础。Knewton 的知识图谱涵盖了不同的学科和知识领域，主要通过专家构建的方法进行设置，学科的核心概念为图谱中的节点，而节点的连边主要以包含、评



图 2: Knewton 数据模型

估、先修、教学关系为主。在知识图谱构建完成后，需要将知识图谱与平台中相应的素材进行关联，当学习者学习过程中遇到了问题，即可通过素材先定位到知识图谱，然后通过知识图谱的推理找到导致学习者问题的原因。

2. 学生事件：学生事件主要是学习者在线学习过程中发生的行为的记录。学习者的学习行为是影响其学习效果的重要因素，同时也是预测其学习偏好、为其进行动态学习规划的重要依据。Knewton 平台中根据学习者与内容交互的类型，定义了试题作答、试题作答起止时间、交互时间以及交互结果等多方面的数据。通过这些数据与学习者个

人档案中数据的整合实现对学习者的有效估计。

3. 目标管理：学习者在线学习过程是以目标为导向的，而由于学习者个体目标的模糊性，平台需要根据学习者的学习历史、即时状态以及评价等方面信息动态地更新学习者的学习目标。而目标的计算主要以知识的掌握状态为标准进行衡量，当学习者开展了一段时间的学习后，其学习状态发生改变，其学习目标也相应产生变化，从而实现学习者学习状态的智能诊断和学习的智能规划。

4. 推荐分析 API：学习者学习过程中所需的内容推荐和分析结果报表，均可通过 API 接口的形式进行提供。当学习者随时随地产生学习需求时，系统即可调用分析推荐 API，为学习者调取学习档案、发现薄弱点，并为其提供练习测试，辅助其学习的不断改进。

二、核心技术

持续自适应 (Continuous Adaptivity)

Knewton 的持续自适应主要体现在空间强化、记忆力和学习曲线、学生学习档案三个方面 [4]。

强化理论是美国心理学家和行为科学家斯金纳等人提出的一种理论，也叫做操作条件反射理论、行为修正理论。传统强化的标准训练方法一般要求学生在较短的时间内学习应用新的概念或技巧，直至他们掌握，甚至达到精通。与传统的强化不同，空间强化，也可理解为分布式强化，是指新的概念或技巧被学习吸收时，以前学过的概念或技巧也被强化。因为新的学习材料往往采用逐步渐进的方式，以学习者熟悉的形式进行设计或者基于学习者熟悉的内容进行组织和编排，从而实现了新旧知识之间的贯通，保证了获得新知识的同时，旧知识也得到很大程度的温习。空间强化经常发生在一个较长的时期内，注重知识在时间上的延伸分布和筛选抽取。空间强化使得 Knewton 能够让学生以不断累积的方式来培养自身的能力，并且一旦学会就能长久记住。

受到艾宾浩斯记忆保持力和学习曲线的启发，Knewton 利用指数增长和衰退曲线来描述学生学习和忘记的能力曲线，其建立的假设前提是对于接触过的主题内容，学生对于这个主题的知识或能力的记忆就会深入一层；而对于没有接触的主题内容，学生很有可能会在一段时间后忘记这个主题的内容。控制记忆力比率的遗忘曲线表示为： $R = e^{-\frac{t}{S}}$ ， R 代表记忆保留值， S 代表记忆力相对强度， t 表示时间，随着时间的延长，记忆保留值逐渐变小。通过将这个遗忘曲线整合进推荐与分析引擎中，Knewton 可以根据学生在何时以及如何接触学习内容等信息来捕获学生知识增长和减少的方式，最终监控学生在课程中的发展轨迹。

学生学习档案是保持持续更新的，这种持续更新的学习档案有助于深入剖析学生理

[4]Knewton 适应性学习白皮书 [EB/OL].http://www.sjsjw.com/kf_www/article/106_3104_3856.asp.

解材料的内部过程，比如哪些是学生真正掌握的，哪些还没有掌握，哪些地方存在误解等，同时还有助于学生逐步形成对自己学习风格、学习优劣势的全面深刻理解，进而最大限度地发挥他们的学习潜能。例如，学生学习完某门课程后，当他进入新的一门课程时，这门新课程将会利用学生最近生成的相关学习数据、已形成的学习轨迹，开始为这位学生构建学习档案，并力求学生从课程学习的一开始就能够达到最大程度的适应性。一旦有了足够的学习记录数据，Knewton 便能逐步揭示并更新学生的学习模式、可能存在的学习盲点、对媒介的喜好、对学习内容粒度和强度的偏好等。

项目反应理论 (Item Response Theory)

传统的项目反应理论一般是针对问题、项目来设计相关参数，且运用过程中通常会存在两大误区：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，Knewton 对传统的项目反应理论进行了扩展，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是随时间变化的，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进行对学生能力的评估和表征。

假设你正在教一群四年级学生的数学复习课，且安排了一个包含十道题目的考试。这十道题目中有两道非常简单，两道非常难，其余的都是中等难度。现在假设有两名学生参与了测试，并且都答对了九道题目。第一名学生错了一道简单的试题，而第二名学生则错了一道比较难的试题，那么哪一名学生对材料更了解呢？在传统的评价机制下，他们可能都得 90 分，且评级为 A，然后进入下一个测试。但是这种利用测量工具来评估学生能力的方法会带来一个问题，那就是测试问题并不具备统一的特质。在考虑不同问题之间差异的同时，Knewton 基于学生问题层面的表现而不是整体测试成绩来对其能力进行建模。对于理解不同问题所带来的贡献，Knewton 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及答题者个人能力信息的贡献来计算解释。

概率图形化模型 (Probabilistic Graphical Models)

概率图形化模型涉及贝叶斯网络、马尔科夫随机域等数据统计方法，且允许数据科学家在多维空间中通过成百上千的参数来分析和操作概率分布。换言之，概率图形化模型允许数据分析师对一个应用效果建立复杂的分析评估模型，将观测到的学习活动与评估联系起来。

概率图形化模型的核心是通过一个概念的掌握状态，推断其他概念的掌握程度。Knewton 通过利用学生已经掌握的知识来分析和推断学生可能有能力学习的知识内容。例如，Knewton 发现分数掌握到某种程度可以帮助学生掌握小数，小数掌握到某种程度可以帮助学生掌握求幂运算。基于这种逻辑分析关系，Knewton 就能够确定对于分数的掌握和幂运算的

掌握之间的联系，进而利用这种发现的类型关系来促进推荐引擎性能的优化和提升。

凝聚的层次聚类 (Hierarchical Agglomerative Clustering)

层次聚类就是通过对数据集按照某种方法进行层次分解，直到满足某种条件为止，根据层次分解的不同策略可以分为凝聚和分裂两种形式。凝聚的层次聚类是一种自底向上的策略，首先将每个对象作为一个簇，然后通过迭代的方式将这些原子簇合并为越来越大的簇，直到所有的对象都在一个簇中，或者满足某个终结条件。

Knewton 的凝聚的层次聚类主要用于设计学生分组算法、选择作为分组依据的学生特征以及发现大规模分组内部的可能潜在结构，为满足不同教学需求的科学分组提供便利。例如，Knewton Math Readiness 包含了层次聚类的样例，其提供了一个分组面板，对使用相同教材的学生，教师可以根据学生对概念理解的不同级别进行相应分组。

三、展望

知识图谱是实现自适应学习的基础部件，目前很多自适应学习平台（包括 Knewton）只是将其作为平台运行背后的支撑性功能模块，而较少向学习者进行直接展示。应该说，在教育领域，知识图谱作为描述特定领域内的知识集合，一般可用于进行教学知识结构化原始表征的初始形式。而学习是一种过程性的个性化体验，需要个体投入认知努力，如果能够在知识图谱上叠加学习者个体的认知过程信息，并在学习过程中为学习者提供及时的学习认知状态反馈，必将有益于学习的真正发生，也将极大地提升学习者开展自适应学习过程的体验。

学习认知地图 (learning cognitive map) 不仅能够表征已映射过程数据的知识学习状态，推理挖掘未掌握知识的学习路径，还应该能够预测缺乏过程数据的知识学习状态，形成符合认知特性的资源和人际网络推荐。基于学习认知地图功能特征的考虑，设计了学习认知地图的概念模型，如图 3 所示，具体包括知识点内容、知识

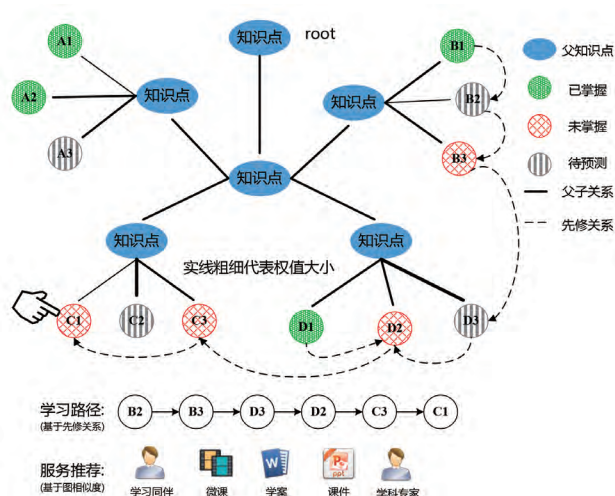


图 3: 学习认知地图概念模型

[5] 万海鹏, 余胜泉. 基于学习元平台的学习认知地图构建 [J]. 电化教育研究, 2017, 38(09): 83-88+107.

点之间的关系、学习认知状态、知识点簇下知识点间关系权重、学习路径和服务推荐六方面的内容 [5]。其中知识点内容是指根据教学大纲，在满足教师日常教学需求的基础上整理形成的知识单元，由高到低分为不同层级，层级越低，知识单元越具体；知识点之间的关系是指知识对象之间的连结，或是表示相关的对象之间连结的意义，包括父子关系和先修关系；学习认知状态是指学习者对知识学习目标掌握程度的描述，分为已掌握（绿色密网格）、未掌握（红色疏网格）和待评估（灰色竖线）；知识点簇下知识点间关系权重是指在同一知识点簇的集合中，不同兄弟知识节点对于父亲知识节点的重要程度，以连线的粗细表示，越粗表示所占的权重越大；学习路径是指为了掌握目标知识点（当前处于未掌握状态）而提供的先修知识点序列，是指知识点学习顺序层面上的学习路径，且对于具有多个学习目标要求的知识点来说，具有针对不同学习目标要求的个性化学习路径；学习服务推荐是指在综合考虑学习者所处的学习目标层级、学习认知状态的基础上为学习者推送的用于辅助其达成学习目标的学习活动、数字资源（如微课、课件）和人际网络（如学习同伴、知识专家）。

在上述概念模型的指导下，北京师范大学未来教育高精尖创新中心团队基于学习元平台（Learning Cell Knowledge Community）[6] 研发了可用于表征学习者在线学习过程中知识结构和认知状态的学习认知地图，如图4所示，以为自适应学习的未来发展提供参考和借鉴。学习认知地图，一方面可以帮助学习者规划学习路径、推荐与知识点建立关联的学习资源和学习活动，另一方面可以帮助学习者寻找知识点背后的学科专家以及与学习者学习步调最一致的学习同伴。

学习路径及内容推荐

最适合的学习同伴

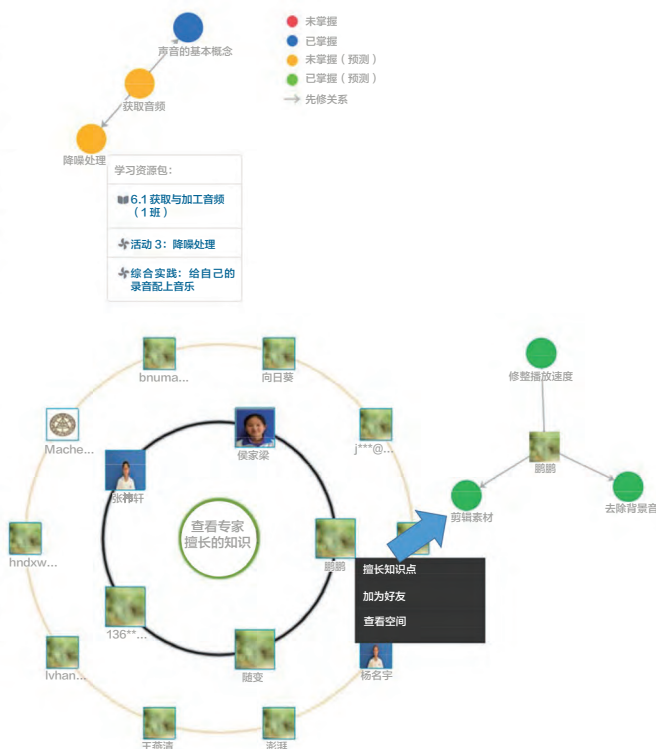


图 4：学习元平台学习认知地图



查看内容精选

[6] 余胜泉, 杨现民, 程罡. 泛在学习环境中的学习资源设计与共享——“学习元”的理念与结构[J]. 开放教育研究, 2009,15(01):47-53.