

基于多源数据融合的共享教育数据模型研究

武法提¹, 黄石华²

(1.数字学习与教育公共服务教育部工程研究中心,北京 100875;

2.北京师范大学教育技术学院,北京 100875)

[摘要] 人工智能教育时代,传统的教育数据共享方法无法满足海量教育数据共享的时效性,进而影响智能教育系统响应的即时性与智能性,文章提出了一种基于多源数据融合的共享教育数据模型的建模方法。该建模方法首先对多源数据融合的概念、融合方法等内容进行分析,并对多种异构数据源的数据共享特性进行剖析,提取出“学习者、时间、空间、设备、事件”五维数据共享特性来对多源异构的教育数据进行数据融合分析;然后再结合国际通用的 xAPI (Experience API)数据规范,对融合后的数据进行规范化分析,生成通用的教育数据交换格式;最后,基于该数据交换格式,探讨了共享教育数据模型的总体架构及实现路径,并构建一个可重用、可共享的教育数据模型,以期为今后开展基于大数据的数据共享的研究提供一套切实可行的实践指导框架。

[关键词] 数据特性;多源数据融合;xAPI规范;数据共享模型

[中图分类号] G434

[文献标志码] A

[作者简介] 武法提(1971—),男,山东郓城人。教授,博士,主要从事智能学习系统设计研究。E-mail:wft@bnu.edu.cn。

一、问题的提出

人工智能教育时代,数据收集与共享是智能教育过程中非常重要的一个环节^[1],数据的共享程度直接影响着智能教育系统响应的即时性和智能性。而随着云计算、移动互联网、大数据、人工智能等“新技术”的日渐成熟,这些“新技术”赋予了智能学习终端设备全方位感知和采集数据的能力,能够捕捉学习者全范围、全过程的学习行为数据,形成海量的教育大数据,呈现出多源性、多模态、多样性的特点。这些“新技术”虽然给人们的数据获取带来极大便利,但也出现海量教育数据资源与各异构数据源难以获取所需数据之间的矛盾,以及异构数据之间数据共享的时效性较差等问题,而如何打通各异构数据源的“数据壁垒”,构建一个高度共享互通的教育数据模型,在提升海量教育数据共享的时效性的同时,也为智能教育系统提供更客观、全面、完整的数据支撑,则

是当前教育发展迫切需要解决的问题之一。基于此,文章提出一种基于多源数据融合的共享教育数据模型的建模方法,通过采用多源数据融合方法对海量异构教育数据进行特征级数据融合,再结合国际通用的 xAPI(Experience API)学习数据规范,构建一个可重用、可共享的教育数据模型,以提升海量教育数据共享的时效性,实现各异构数据源之间数据的高度共享与互通。

二、多源异构教育数据的数据融合分析

(一)多源数据融合的方法分析

数据融合最早应用于军事领域,它是一种多层次、多方面的数据处理过程,用于处理多源数据,对信息进行自动检测、联合、相关、估计和合成^[2],主要是为了实现较为准确的位置推断和身份估计,进而对战场状况、威胁程度和重要水平作出及时完整的评价^[3]。后来在传感器、地理空间、情报分析等多个领域得到了

应用与发展,尤其是互联网时代,多源数据融合逐渐成为大数据领域的重要研究方向^[4]。通过多源数据融合,可以实现多源信息的交叉印证,可以达到数据信息的相互补偿,并可以有效地减少数据量,以获取确定数据和深层次的语义知识^[4-5]。人工智能教育时代,教育大数据的生态逐渐形成,海量的教育数据呈现多源异构特性,而多源数据融合方法为解决新时代教育大数据的共享互通提供了一种新的解决思路,为构建一个重用、共享的教育数据模型提供了可行的实践视角。

而目前常见的多源数据融合方法主要有数据级融合、特征级融合和决策级融合等三种^[6]。数据级融合是属于最底层的数据融合,它是对原始数据经过简单预处理之后直接进行关联和融合,融合之后才作数据特征提取;特征级融合是先对数据进行特征提取,再对数据进行关联融合;决策级融合是先对各数据源进行决策,然后再将这些决策进行关联融合,最终获得整体一致性的决策结果^[7-9]。这三种数据融合方法的融合过程如图1所示。通过比较这三种数据融合方法,可以看出,数据级融合虽然能够最大程度上保留原始数据的特征,但融合的代价较高,时效性也较差,无法满足人工智能教育时代对数据的即时性等要求;决策级融合虽然具有很高的容错性和时效性,但它是以具体决策需求为出发点进行的数据融合,面对人工智能教育时代复杂多变的教育大数据环境,难以制定出具体适配的决策以进行数据融合;而特征级融合在保证即时性的同时,也能够最大程度上给出决策所需的特质信息,其融合结果也具有较高的精度,由于教育大数据不像图像数据具有融合的高精度,特征级融合方法很好地契合了人工智能教育时代教育大数据分析的要求。基于此,文章将采用特征级数据融合方法来构建共享教育数据模型。

(二)异构多源教育数据的共享数据特性的提取

为了提高数据融合之后的数据重用度,文章的多源教育数据融合方法是通过提取各异构数据源的数据共享特性来进行特征级数据融合。而数据共享特性的提取过程,其实质就是从各异构数据源学习者所产

生的学习行为数据中提取学习情境特性的过程。纵观以往的学习情境的信息特性描述,不同学者从不同的视角将情境信息特性划分为不同的类型。以下是相关学者对情境信息特性划分的典型观点:如 DEY^[10]认为,情境信息包括位置、时间和周围环境等显式感知的情境信息,同时也包括社会关系、习惯、消费水平和喜好等蕴含感知的情境信息。Lieberman^[11]等人将情境分为用户情境、环境情境和应用情境三大方面,其中,用户情境包括活动、位置和描述等情境;环境情境包括时间、亮度、温度、天气、资源等情境;应用情境包括功能、维护、能源等情境。岳玮宁^[12]等人将情境信息分为自然环境、设备环境、用户环境三大类。顾君忠^[13]等人从用户为中心的视角,将情境信息分为计算情景、用户情境、物理情境、时间情境和社会情境等情境。而人工智能教育时代不同数据源的学习数据具有很明显的时空特性,且学习者的学习交互离不开设备的支持。基于上述的情境信息特性分析,文章将各异构数据源的共享数据特性提取为学习者情境、时间情境、位置情境、设备情境和事件情境5个维度情境信息特性(见表1)。

表1 学习者的学习情境分类

情境分类	情境要素举例
学习者情境	姓名、年龄、班级、学段、学校等信息
时间情境	课堂学习时间、课间休息时间、娱乐休闲时间、体育锻炼时间、晚自习时间等
位置情境	校内(宿舍楼、图书馆、实验楼、阅览室等);校外(家、休闲吧、科技馆、动物园等)
事件情境	看初一数学不等式的微课、做初一英语 Unit2 的作业、玩海岛奇兵射击类的游戏等
设备情境	硬件(平板电脑、智能手机、可穿戴设备等),软件(软件类型、程序功能、版本等)

(三)基于共享数据特性的特征级数据融合

以上5个维度的学习情境信息特性很好地表征了各异构数据源的共享数据特性,通过这5个共享数据特性,可以准确地描述各异构数据源中学习者真实的学习生活场景,进而可以很好地实现异构数据之间

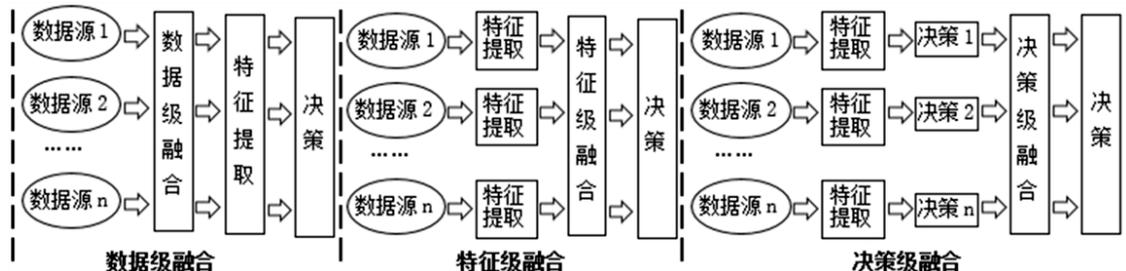


图1 三种不同数据融合方法的融合过程

的无缝对接。这5个共享数据特性代表5个数据维度,组合后可以构成学习者真实的学习场景:“学习者情境+时间情境+位置情境+设备情境+事件情境 \cong 学习场景”,它描述了“学习者、什么时间、什么地点、基于什么设备、做了什么事情”。基于这5个数据维度,采用特征级数据融合方法对教育数据进行融合。其特征级数据融合主要经过各数据维度的语义特征的分层提取、分层语义的特征级数据融合、跨维度跨分层的关联语义的特征级数据融合等融合过程,如图2所示。

(1)各数据维度的语义特征的分层提取,主要是对这5个不同数据维度进行语义特征的分层提取,确定各数据维度的语义属性,并确定各数据维度语义同级、上下级的多层语义逻辑关系。如时间维度的语义属性可以分为工作日和节假日大类语义信息,工作日又可以细分为课堂学习时间、自习时间等不同粒度的语义信息。

(2)分层语义的特征级数据融合,主要是将各异构数据源的教育数据,按照这些分层语义分类,采用相应的细粒度融合策略进行特征级数据融合,生成能准确描述学习者学习特征的场景数据,并消除数据结构和相同语义聚集在同一粒度上的不一致与冗余关系。

(3)跨维度、跨分层的关联语义的特征级数据融合,主要是为了更客观、精准地描述学习者的学习特征,根据不同维度、不同层面的相似语义,对这些具有关联语义的数据进一步进行特征级融合,生成具有深层次语义知识的场景数据(如学习者的学习习惯等)。

三、xAPI 规范对 5 维特征融合数据的规范化分析

(一) xAPI 规范与 5 维特征融合数据的融合分析
在得到 5 维特征融合的教育数据后,接下来就

需要对这 5 维特征融合数据进行规范化分析,而学习数据规范的选择是实现数据格式规范的关键。纵观以往学习数据规范标准的发展,主要经历了以下几个阶段^[14-16]:无标准阶段、AICC(The Aviation Industry CBT Committee)标准阶段、SCORM(Sharable Content Object Reference Model)标准阶段。AICC 标准虽然在一定程度上解决了课程资源的共享问题,但采用这种学习数据规范开发的课程资源迁移性不好,不一定能在不同的平台上运行。SCORM 标准是目前最为广泛应用的学习数据规范标准^[17],但这种标准只是针对课件等学习内容的数据规范,无法对课件学习以外的学习过程数据进行记录,也无法实现跨平台的数据共享与互通^[18],尤其是在面对人工智能教育时代动态多变的教育环境时,SCORM 标准难以解决新时代下多元异构数据的高度共享性问题。为了破除 SCORM 标准的局限,美国 ADL(Advanced Distributed Learning)组织推出了 xAPI(Experience API)数据规范,它不但兼容 SCORM 标准,而且可以记录任何一种学习或行为,并且可以跨平台进行数据共享与交换^[15],xAPI 规范的这种优势很好地契合了人工智能教育时代复杂多变的教育环境。

xAPI 规范的核心部件主要有两个^[15,19]:Statement 属性和 LRS(Learning Record Store)学习记录存储。其中,Statement 是定义了 xAPI 数据格式的语法,LRS 是定义了学习记录库(LRS)的数据的存储形式。根据 Statement 的“执行者(Actor)+动词(Verb)+对象(Object)”的声明结构,以及声明结构 Result、Context、Timestamp、Stored、Authority 等其他扩展属性。xAPI 规范的 Statement 声明结构很好地契合了 5 维特征融合数据的数据描述,可以生成 5 维特征融合数据的 Statement 声明结构,即“学习者(Actor)、时间情境(Time)、空间情境(Local_Context)、设备情境(Device_Context)、事件情境(Verb+ Object+Result)”,

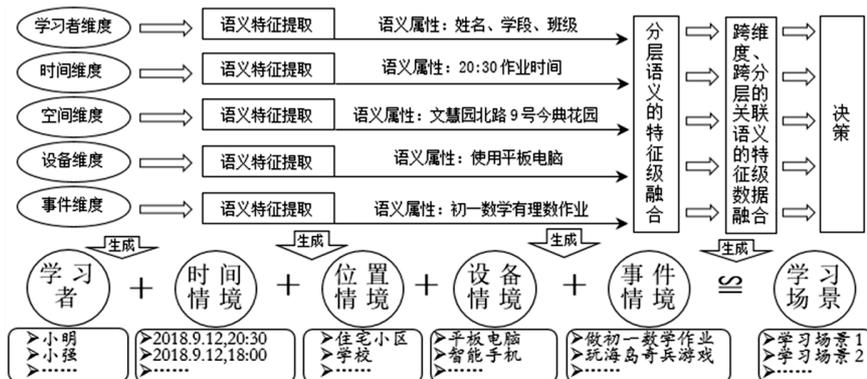


图2 基于5维数据共享特性的教育数据的特征级数据融合过程

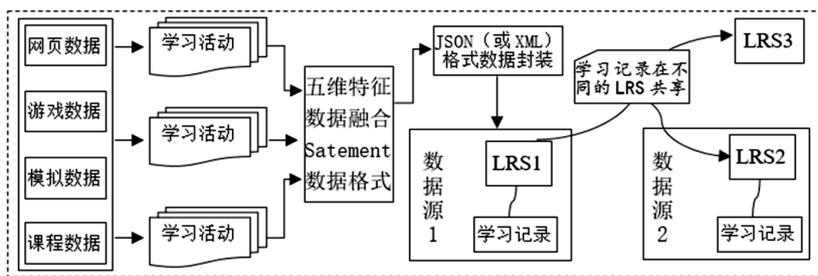


图3 xAPI 规范对5维特征融合数据进行规范化的运行机制

以 Statement 的属性来进行数据封装,生成通用的数据交换格式。然后再通过 xAPI 规范的另一个核心部件——LRS 学习记录库,将数据融合的规范化教育数据传送到 LRS 中进行记录并保存起来,实现 LRS 之间的数据共享和交换。xAPI 规范对 5 维特征融合数据进行规范化的运行机制如下(如图 3 所示):当不同数据源的学习活动或学习行为需要被跟踪记录时,xAPI 就会发出特征级数据融合的 Statement 表述格式,封装成 JSON 或 XML 等通用数据格式传递到 LRS 中,LRS 负责记录和存储,并与其他独立的 LRS 交换和共享这些学习经历记录,一个 LRS 可以与其他独立的 LRS 共享这些学习记录,LRS 也可以独立存在,也可以存在于不同的数据源中。

(二)5 维特征融合数据的通用数据规范格式分析

基于上述 xAPI 规范对 5 维特征融合数据进行规范化的运行机制,为了构建一个高度共享的教育数据模型,还需要对 5 维特征融合数据的数据规范格式进行分析,生成通用、标准的数据交换格式。根据前面分析得到的 5 维特征融合数据的数据结构:“学习者 (Actor)、时间情境(Time)、空间情境(Local_Context)、设备情境(Device_Context)、事件情境(Verb+ Object+ Result)”,再结合 xAPI 规范 Statement 声明结构的规范化描述,由此得到 5 维特征融合数据的通用数据规范格式:“Actor (学习者) + Time (时间情境) + Local_Context (位置情境) + Device_Context(设备情境) + (Verb + Object + Result + T-span)(事件情境)”。该通用数据格式映射出学习者的状态为:“{学习者} 附学习者的个人语义标签 | {某个时间点} 附时间分类语义标签 | {某个地点} 附地点分类语义标签, {使用什么设备} 附设备分类语义标签 | {做了某事, 结果如何,耗时多少} 附主题事件分类语义标签”。其实例化后为:“小明{分类语义标签:初一} | 10:00{分类语义标签:课堂学习} | 在北二附{分类语义标签:学校教 1 楼} | 使用 HUAWEIPad {分类语义标签:平板电脑} | 做作业:有理数{分类语义标签:初一数学第一单元}|得分:95{分类语义标签:优秀}”。由此得到 5 维

特征融合数据的数据格式规范是以“谁在什么时间、什么地点、使用什么设备、做了什么事情”对学习经历数据进行描述,其生成的通用数据格式 (JSON 格式)如图 4 所示:

```
{
  "Actor": [ { "小明": [ { "grade": "初一" } ] },
  "Times": [ { "10:00": [ { "time": "课堂学习" } ] },
  "Local_Context": [ { "北二附": [ { "local_context": "教 1 楼" } ] },
  "Device_Context": [ { "iPad": [ { "device_context": "平板电脑" } ] },
  "Verb": [ { "做": [ { "GBK": "created" } ] },
  "Object": [ { "作业": [ { "grade": "初一", "course": "数学", "knowledge_point": "不等式" } ] },
  "Result": [ { "95 分": [ { "level": "优秀" } ] },
  "T-span": [ { "0": [ { "task_duration": "作业时长" } ] }
}
```

图4 5 维特征融合数据的通用数据格式

四、共享教育数据模型的构建

(一)共享教育数据模型的总体流程框架设计

数据模型是描述数据类型、数据联系、语义约束的集合,它的构成主要有 3 个部分:数据类型、数据联系、语义约束^[20]。其中,数据类型描述了数据的逻辑结构,数据联系定义了操作数据的方法,语义约束规定了数据的语义规则。基于此,文章将共享教育数据模型也划分为数据描述、数据语境、数据共享 3 个部分(如图 5 所示)。其中,数据描述是用于实现对信息资源的表示、发现、共享和重用;数据语境是数据信息所属主题的标注;数据共享是对数据交换格式进行规范化。

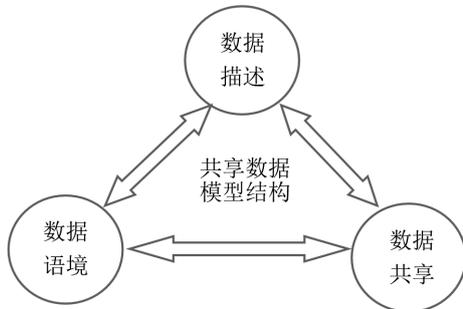


图5 共享数据模型的结构

基于图 5 所示的共享数据模型结构的逻辑关系,结合上述分析得到 5 维特征融合数据的通用数据规范格式,由此可以设计出共享教育数据模型的总体流程框架,如图 6 所示。该总体流程框架是通过对各异

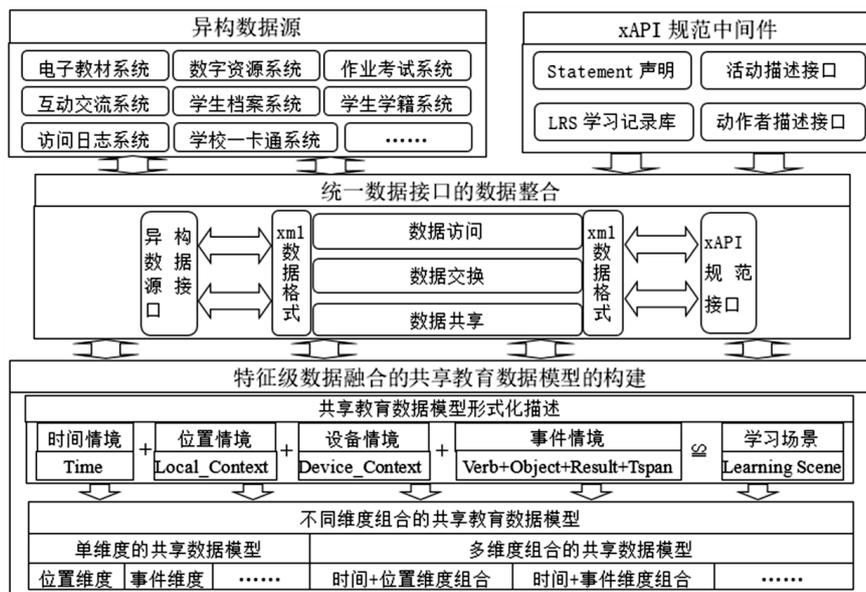


图6 基于多源数据融合的共享教育数据模型的总体流程框架

构数据源的数据特性进行归类分析,剖析出不同数据源的共享数据特征,提取为“学习者+时间情境+空间情境+设备情境+事件情境 \cong 学习场景”5维数据特征并进行特征级数据融合,结合xAPI通用数据规范,将融合后的教育数据转换为通用的教育数据格式规范,生成规范化的共享数据格式存储于LRS中,并对模型中5个独立数据维度进行组合分析,构建一个可重用、可共享的教育数据模型。该模型不但有利于各异构数据源的数据共享与分析,而且能完整地体现学习者真实学习活动全貌,从而为更深层次发掘学习者的学习需求和学习状态提供强有力的数据基础。

(二) 基于多源数据融合的共享教育数据模型的实现路径

基于上述共享教育模型的总体流程框架分析,共享教育数据模型的实现路径主要经过离散和异构数据的规范化运算、模型数据属性与语义唯一性运算、模型数据的形式化描述等3个关键运算环节。

1. 离散、异构数据的规范化运算环节

离散、异构数据的规范化运算主要是将各异构数据源的学习行为轨迹数据,转换为结构化的教育数据规范的过程。该运算环节首先基于上述分析得到的5维特征数据融合的数据规范格式,将采集到离散、异构的学习者学习活动轨迹数据,再调用xAPI规范中间件的Statement声明和动作者描述接口,转换为结构化的学习行为数据,封装成“Actor + Time + Local_Context + Device_Context + (Verb + Object + Result + T-span)”的规范化描述的通用、标准数据规范格式,传递到LRS中进行记录与保存,生成规范化的教育数据集。其规范化过程如图7所示。

2. 模型数据属性的语义处理与唯一性运算

为了便于各异构数据源理解数据位置之间的关联关系,以发掘学习者潜在的学习行为模式和学习规律,需要对模型的数据属性作进一步的语义化处理与唯一性运算。该运算环节主要是对模型的数据属性打上语义标签,并对模型数据的语义进行唯一性计算的过程。首先对模型中5维融合数据的数据属性进行分



图7 异构、离散的数据转化为规范化数据的过程

表 2 具有唯一性语义的规范化数据集

操作者	分类语义	时间情境	时间分类语义	经纬度(位置情境)	位置分类语义	设备情境	设备分类语义	事件情境	事件分类语义
小明	初一	2019.9.12, 9:30	课堂学习时间	23.058098, 113.38562702088893	北二附	IPad2	平板电脑	做初一有理数作业	做作业:有理数{分类语义标签:初一数学第一单元}95分 {分类语义标签:优秀}30分钟
小明	初一	2019.9.12, 10:00	课堂学习时间	23.058098,113.38562702088893	北二附	IPad2	平板电脑	做初一有理数作业	做作业:有理数{分类语义标签:初一数学第一单元}95分{分类语义标签:优秀}30分钟
小明	初一	2019.9.13, 17:30	放学时间	23.057984, 113.38562702093856	枫蓝国际	Iphone5	手机	玩海岛奇兵游戏	玩游戏:海岛奇兵 {分类语义标签:战略类游戏} - -160分钟
小明	初一	2019.9.13, 17:31	放学时间	23.057984,113.38562702093856	枫蓝国际	Iphone5	手机	玩海岛奇兵游戏	玩游戏:海岛奇兵 {分类语义标签:战略类游戏} - -160分钟
小明	初一	2019.10.2, 10:05	晚自习时间	23.047608,113.37279899999999	今典花园	联想E450	台式电脑	查看 Unit 5 课件	查看 课件:Unit5 {分类语义标签:英语课件} - -125分钟
小明	初一	2019.10.2, 11:10	晚自习时间	23.047608,113.37279899999999	今典花园	联想E450	台式电脑	查看 Unit 5 课件	查看 课件:Unit5 {分类语义标签:英语课件} - -125分钟
.....

类语义化处理,即对 5 个相对独立的各数据维度,根据各数据维度的语义分类,给每一个数据维度的属性打上分类语义的标签,如时间数据维度以“天”为粒度的时间语义分类:早读时间、课堂时间等语义分类,生成具有分类语义的规范化数据集;然后对模型数据的语义进行唯一性计算,即去除具有分类语义的规范化数据中的重复数据,以保证模型中的数据没有相同语义的数据记录,进而提升学习行为分析结果的准确性。其模型数据属性的语义规范化过程见表 2。

3. 模型数据的形式描述

为了构建一个灵活性强、重用度高的共享教育数据模型,以减少 LRS 之间的数据交换量,并满足不同数据源对不同数据视图的需求,还需要对模型的数据进行形式化描述。该环节是基于模型数据属性与语义唯一性运算环节计算的结果,对模型中的学习者情境数据进行形式化描述,并进一步对这些独立的数据维度进行多维度组合,以构建一个多维度的共享教育数据模型。针对构成数据模型的 5 个独立数据维度,由

于学习者维度是静态化的信息数据,其形式化描述在模型中不作标注,由此,模型的学习情境数据描述可以用 4 元组 Preference 来表示,各元组既相互独立又相互联系,以构成不同组合的多维度的共享教育数据模型,这 4 元组分别是:时间情境维度、位置情境维度、设备情境维度和事件情境维度,依次用字符表示为 PT、PL、PD、PO,每一个元组是由序偶对<P,δ>来组成,其中,δ 表示各属性的分类语义,由此构成的共享教育数据模型 Preference 的形式化描述公式为:

$$Preference = \langle P, \delta \rangle = \left\{ \begin{array}{l} \langle P_{T1}, \delta_{T1} \rangle, \langle P_{T2}, \delta_{T2} \rangle, \dots, \langle P_{Tn}, \delta_{Tn} \rangle \\ \langle P_{L1}, \delta_{L1} \rangle, \langle P_{L2}, \delta_{L2} \rangle, \dots, \langle P_{Ln}, \delta_{Ln} \rangle \\ \langle P_{D1}, \delta_{D1} \rangle, \langle P_{D2}, \delta_{D2} \rangle, \dots, \langle P_{Dn}, \delta_{Dn} \rangle \\ \langle P_{O1}, \delta_{O1} \rangle, \langle P_{O2}, \delta_{O2} \rangle, \dots, \langle P_{On}, \delta_{On} \rangle \end{array} \right.$$

上述的形式化描述公式,以时间情境维度为例,<P_T, δ_T>可以是由<18:30, δ_{T1}=晚饭时间><20:30, δ_{T2}=晚自习时间><22:30, δ_{T3}=晚睡时间>等组成的集合。模型中既可以构成位置情境、事件情境等单维度的数据模

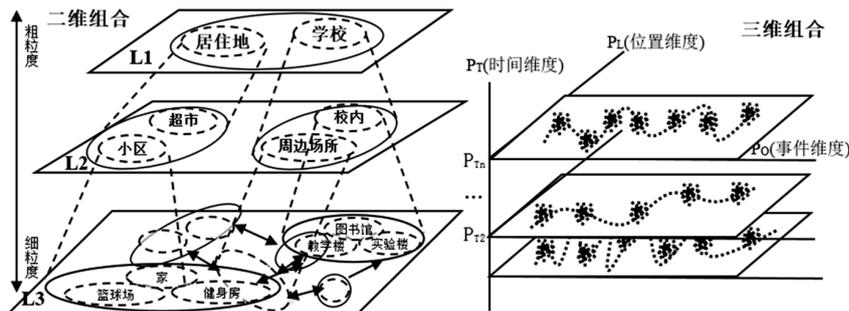


图 8 不同维度组合的数据模型视图

型视图,也可以组合成多维度的数据视图,如图8所示,以便于各数据系统根据自身数据分析的需求,动态地获取个性化的数据模型视图,这样构建出的共享教育数据模型不但可以减少各异构数据系统之间数据交换的数据量和传输量,同时也降低了数据分析的复杂度,有效减少了数据信息的缺失,有利于深层次挖掘学习者真实的学习需求。

五、结 语

人工智能教育时代,构建一个可重用、可共享的教育数据模型,以规范人工智能教育环境下多源异构

的教育数据,实现各异构数据的高度共享,是当今时代教育发展亟需解决的问题之一。文章提出的基于多源数据融合的共享教育数据模型,通过对多源异构教育数据进行数据融合,结合 xAPI 学习数据规范,对融合后的教育数据进行规范化描述,生成通用、标准的数据交换格式,在此基础上,构建了一个可共享、可重用的教育数据模型,实现了各异构数据源之间的数据共享与交换,便于智能教育系统获取更全面、完整的学习记录数据,以提升数据共享的时效性,进而使得学习行为分析的结果更客观、及时、准确,有利于提升智能教育系统响应的即时性与智能性。

[参考文献]

- [1] 吴文峻. 面向智慧教育的学习大数据分析技术[J]. 电化教育研究, 2017(6): 90-96.
- [2] U.S. Department of Defense. Data fusion subpanel of the joint directors of laboratories [R]. Washington: Technical Panel for C3, Data fusion lexicon, 1991.
- [3] STEINBERG A N, BOWMAN C L, WHITE F E. Revisions to the JDL data fusion model[C]// Proceedings of SPIE - The International Society for Optical Engineering, 1999: 430-442.
- [4] 化柏林, 李广建. 大数据环境下多源信息融合的理论与应用探讨[J]. 图书情报工作, 2015, 59(16): 5-10.
- [5] 胡永利, 朴星霖, 孙艳丰, 尹宝才. 多源异构感知数据融合方法及其在目标定位跟踪中的应用[J]. 中国科学: 信息科学, 2013, 43(10): 1288-1306.
- [6] 姜延吉. 多传感器数据融合关键技术研究[D]. 哈尔滨: 哈尔滨工程大学, 2010.
- [7] 王海颖. 多源数据关联与融合算法研究[D]. 无锡: 江南大学, 2016.
- [8] 刘桐欢. 无线传感器网络多源数据融合技术研究[D]. 北京: 北京交通大学, 2014.
- [9] 贺雅琪. 多源异构数据融合关键技术研究及其应用[D]. 成都: 电子科技大学, 2019.
- [10] DEY A K. Providing architectural support for building context-aware application[D]. Atlanta: Georgia Institute of Technology, 2000.
- [11] LIEBERMAN H, SELKER T. Out of context: computer systems that adapt to, and learn from context [J]. IBM systems journal, 2000, 39(3/4): 617-632.
- [12] 岳玮宁, 董士海, 王悦, 等. 普适计算的人机交互框架研究[J]. 计算机学报, 2004, 27(12): 1657-1664.
- [13] 顾君忠. 情景感知计算[J]. 华东师范大学学报(自然科学版), 2009, 20(5): 1-20.
- [14] BAKHOUYI A, DEHBI R, LTI M T, et al. Evolution of standardization and interoperability on e-learning systems: an overview[C]// Ohrid: 2017 16th International Conference on Information Technology Based Higher Education and Training (ITHET), IEEE, 2017: 1-8.
- [15] Technical introduction to xAPI[EB/OL]. [2014-02-13]. <https://xapi.com/tech-overview/>.
- [16] 顾小清, 郑隆威, 简菁. 获取教育大数据: 基于 xAPI 规范对学习经历数据的获取与共享[J]. 现代远程教育研究, 2014(5): 13-23.
- [17] SHARIAT Z, HASHEMI S M, MOHAMMADI A. Research and compare standards of e-learning management system: a survey[J]. Information technology and computer science, 2014, 6(2): 52-57.
- [18] GULZAR Z, LEEMA D A A, AYZAZ A S. Educational data mining using SCORM specifications on learning management system[J]. International journal of applied engineering research, 2015, 10(82): 592-600.
- [19] 方海光. 基于 xAPI 学习记录的 LMS 网络系统架构研究[J]. 中国电化教育, 2015(2): 65-69.
- [20] Office of Management and Budget. The data reference model v2.0 [R/OL]. [2006-12-21]. http://www.whitehouse.gov/OMB/egov/documents/DRM_2_0_Final.pdf.

(下转第 103 页)

and intrinsic mechanism are not clear. Therefore, this study investigates three pairs of relationships: immersion and learning motivation, presence and embodied concern, schema and concept generation. An experiment is designed to detect the learning differences between IVR and the video, and the results of the experiment are analyzed from two aspects of learning experience and cognition. It is found that IVR has advantages in the process of spatial presence, depth perception and schematization. On this basis, the contradiction between the fidelity of IVR materials to the objective and the artificial selective construction is discussed. It is considered that the schematization is a basic principle for selective construction. From the perspective of geospatial thinking, this paper interprets the meaning of embodiment, and then gives new connotation to situated learning with the thought of embodied cognition.

[Keywords] Geographic Space; Sense of Presence; Schema; Embodied Cognition; Immersive Virtual Reality

(上接第 65 页)

Research on Shared Education Data Model Based on Multi-source Data Fusion

WU Fati¹, HUANG Shihua²

(1.Engineering Research Center of Digital Learning and Educational Public Service, Beijing 100875;

2. School of Educational Technology, Beijing Normal University, Beijing 100875)

[Abstract] In the era of artificial intelligence education, the traditional education data sharing method cannot meet the timeliness of the sharing of massive education data, which in turn affects the immediacy and intelligence of the response of intelligent education systems. This paper presents a modeling approach for a shared education data model based on multi-source data fusion. Firstly, this paper analyzes the concept and fusion method of multi-source data fusion, as well as the data sharing characteristics of multiple heterogeneous data sources, and extracts the five-dimensional data sharing characteristics of "learner, time, space, device, event" to perform data fusion analysis of multi-source and heterogeneous education data. Secondly, combined with internationally accepted xAPI (Experience API) data specification, the standardized analysis of the fused data is conducted to generate a common exchange format of education data. Finally, based on the data exchange format, this paper discusses the overall architecture and implementation path of the shared education data model, so as to provide a practical guidance framework for future research on big data-based data sharing.

[Keywords] Data Characteristics; Multi-source Data Fusion; xAPI Specification; Data Sharing Model