

基于场景感知的学习者建模研究

武法提, 黄石华, 殷宝媛

(北京师范大学 教育学部 教育技术学院, 北京 100875)

[摘要] “互联网+教育”时代,以往传统的学习场景正不断被重构,新型的学习场景逐渐向动态化、多元化、碎片化的模式发展,使得基于场景挖掘的新型个性化学习服务模式的重要性越发凸显。而学习者模型作为个性化学习服务系统的核心部件,目前还无法满足新时代下高精度的个性化服务要求。基于此,文章提出一种基于场景感知的学习者建模方法,该方法是基于近些年有关学习者特征分析模型研究的分析结果,融合学习者的场景特性,设计了一个6维度学习者特征分析模型,在此基础上,融合场景感知建模方法和频繁序列挖掘算法,计算不同学习场景下各维度学习者特征值,由此构建一个具有场景特性的个性化学习者模型,并基于该模型探讨了个性化学习服务推荐演化框架。

[关键词] 互联网+教育; 场景感知; 学习者特征; 学习者建模

[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,主要从事数字化学习环境与学习资源设计研究。E-mail: wft@bnu.edu.cn。

一、引言

“互联网+”时代的教育,是一个灵活、多样、开放、终身的个性化教育,是重视学习者的个性化和多样性发展的教育^[1],其教育形态是一个多元异构、动态多变、开放共享的教育形态。这种新型的教育形态,促使学习场景发生了巨大的变化,从学校传统的课堂学习场景逐渐扩张到校外碎片化的学习场景,这就要求个性化学习推荐服务需要在合适的学习场景中进行,从而引发研究者重点关注围绕“场景”进行个性化学习服务设计研究,以提高学习服务的精准度^[2]。而学习者模型作为个性化学习服务系统的核心部件,其模型的准确性直接决定着个性化学习服务系统的精准度。因此,如何基于动态多变的学习场景,构建一个具有场景特性的个性化学习者模型,便成为当前互联网学习环境亟须解决的重要问题之一。

二、学习者模型的相关研究

近年来,关于学习者模型的研究,逐渐从学术研究走向实际应用。目前,对于学习者模型的研究,国内外的权威机构出台了相关的学习者模型规范,其中比

较有代表性的规范有:PAPI (Public and Private Information) 规范^[3]、IMS-LIP (IMS Learner Information Package) 规范^[4]和 CELTS-11 学习者模型规范^[5]。PAPI 规范规定了学习者的公开与隐私信息,把学习者的相关信息分为6大类:个人信息、关系信息、安全信息、偏好信息、绩效信息和作品集信息。LIP 规范则通过将学习者信息封装成通用的一组信息包,以实现不同学习系统之间的数据交换与共享,它将学习者信息分为11类:身份验证,目标,质量、认证和许可,可访问性,活动,能力,兴趣,成绩报告单,联系,安全钥匙,关系等信息。而 CELTS-11 学习者模型规范主要借鉴了 PAPI 和 IMS-LIP 等规范,是结合中国的实际情况制定的学习者模型规范,它将学习者信息划分为8类:个人信息、学业信息、管理信息、关系信息、安全信息、偏好信息、绩效信息和作品集信息。这三个学习者模型规范,虽然从理论层面上能比较全面地囊括学习者各方面的特征信息,使得学习者的行为表现数据可以在不同学习系统之间进行交换、共享和重用,但这三个学习者模型规范缺乏对学习者的知识、情感等方面特征的描述,难以表征出学习者真实的学习状态,无法满足互联网学习环境下个性化的学习需求。为

等人对“学习场景要素”的划分方法^[2],将学习场景的数据维度划分为:时间维度、空间维度、设备维度和事件维度4个维度。其中,设备是互联网学习环境下不可或缺的要害,它是互联网学习环境下人与物沟通的桥梁,是量化学习行为轨迹数据的技术支撑。由此得到融合场景特性的4维度学习者特征分析框架(如图2所示)。该分析框架中4个数据维度的组合,可抽象成语义化程度更高且易于分析的学习场景,其形式化描述为“时间维度+空间维度+设备维度+事件维度 $\cong$ 学习场景”,很好地还原了学习者真实的学习场景,即:“学习者在什么时间,什么地点,使用什么设备,做了什么事情”。这样的个性化学习者特征描述,更有利于深层次挖掘学习者潜在的学习需求。

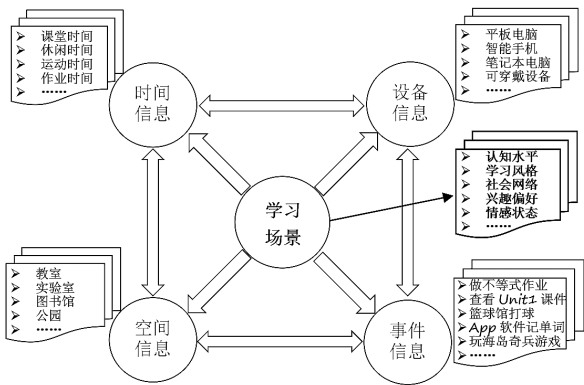


图2 融合场景特性的学习特征分析框架

四、基于场景感知的个性化学习者模型构建

(一)基于场景感知的个性化学习者模型的总体架构设计

为了构建一个闭环、自适应的个性化学习者模型总体架构,本文将个性化学习者模型划分为场景感知模块、场景数据计算模块、模型应用3个模块。其中,

场景感知模块是架构的基础模块,主要是通过智能感知技术主动感知并采集场景化的学习行为数据(包括时间信息、空间信息、设备信息、行为信息等),再融合通用的数据规范中间件,对采集到的场景化学习行为数据进行规范化描述,并对规范化描述后的场景数据进行清洗、处理,生成可供学习者特征值计算的场景数据集;场景数据计算模块是架构的核心模块,主要是基于前面生成的规范化的场景数据集,融合场景感知建模方法^[19]和频繁场景挖掘算法^[20]计算各维度的学习者特征值,生成涵盖基本信息维度、认知水平维度、学习风格维度、兴趣偏好维度、社会网络维度和情感状态维度6个子模型的学习者模型;模型应用模块是架构的关键模块,它是各个模块之间数据交互共享的枢纽,使得各个模块共同作用为学习者提供精准的个性化学习服务。其主要任务是根据学习者当前学习场景的需求,再融合相关的个性化学习推荐算法,筛选合适的个性化学习推送策略,找准个性化学习推送时机,提供给学习者适配的个性化学习信息资源。基于这三个模块的分析,可以生成基于场景感知的个性化学习者模型的总体架构,如图3所示。

(二)基于场景感知的个性化学习者模型的构建

从图3的模型总体架构可以看出,学习者模型主要包括学习者场景化行为数据的获取、处理以及模型输出的表示等流程。基于此,本文将个性化学习者模型的构建流程主要分为学习场景数据的采集、模型的特征值计算和模型的形式化描述3个关键步骤。

1. 基于场景感知的数据采集

互联网学习环境下,场景化学习行为数据的采集是借助互联网的智能感知能力,主动感知并采集学习者学习过程的学习情境数据,包括静态的学习情境数据和动态的学习情境数据。而为了便于系统的分析与

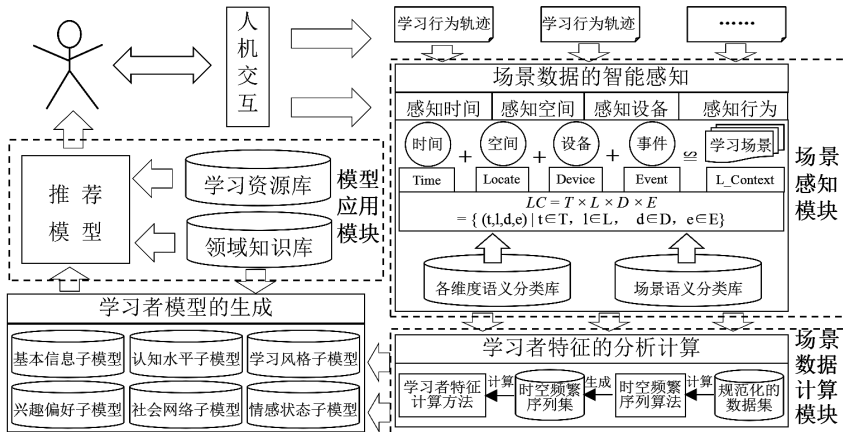


图3 基于场景感知的个性化学习者模型的总体流程架构

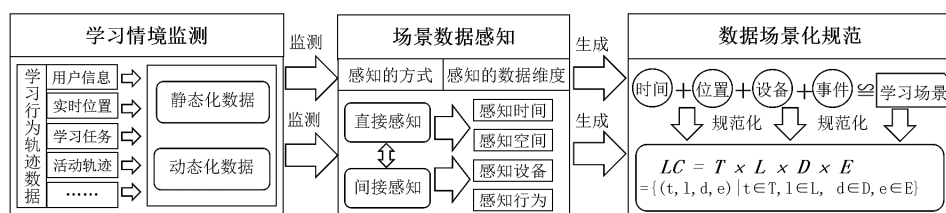


图4 学习情境数据采集流程

计算,还需要将采集到的学习行为数据作进一步的规范化处理,提取成系统易于识别和处理的场景化数据。因此,基于场景感知的数据采集过程主要分为3个阶段:学习情境监测阶段、场景数据感知阶段和数据场景化规范阶段3个阶段。其采集的流程如图4所示。

(1)学习情境监测阶段。该阶段是对学习者产生的学习行为数据进行归类分析,生成感知设备易于感知的学习情境数据,在此基础上,借助互联网的智能感知设备,主动监测并采集这些学习情境数据变化信息。

(2)场景数据感知阶段。该阶段主要是将学习情境数据转化为场景化的学习行为数据。而为了更全面、更完整地获取学习者的学习场景数据,本文采用直接感知和间接感知的方式。直接感知主要是基于感知设备本身来进行数据采集,如用于获取温/湿度、心跳、血压等传感器数据以及识别学习者地理位置的GPS、无线网络定位等数据的获取;间接感知是感知设备无法直接获取,则由外设系统或环境等间接获取,如学习者的学习任务、社交网络等情境。基于这些数据感知方式,一旦感知到变化的学习情境数据,系统就将这些原始学习情境数据转换切割为4维度的场景化学习行为数据。

(3)数据场景化规范阶段。该阶段主要采用规范化的数据格式,将场景化学习行为数据转换为直接被机器识别的结构化、标准化的学习场景数据,生成的规范化学习场景数据可以表示为: $LC = T \times L \times D \times E = \{(t, l, d, e) | t \in T, l \in L, d \in D, e \in E\}$,其中, t 为时间集合 T 中的元素, l 为位置集合 L 中的元素, d 为设备集合 D 中的元素, e 为事件集合 E 中的元素。

2. 基于频繁场景挖掘的模型特征值计算

基于前面采集到的具有高度语义、规范化的学习场景数据集,见表2。而基于频繁场景挖掘的模型特征值计算就是从这些学习场景数据集中,挖掘出与学习者特征关联的频繁学习场景序列模式,结合各维度特征值的计算方法,计算学习者模型的特征值。其计算过程主要经历原始学习场景数据集的语义唯一性

编码、基于时空聚类学习场景发现计算、频繁学习场景挖掘计算、基于频繁场景挖掘的学习者特征值计算这四个步骤,其具体的算法设计如下:

表2 原始的学习场景数据集

时间	空间	设备	事件
2018.9.12, 10:02	教室	平板电脑	课堂练习
2018.9.25, 10:50	教室	平板电脑	课堂讨论
2018.9.16, 18:00	教室	智能手机	玩游戏
2018.9.20, 18:30	地铁站	智能手机	玩游戏
2018.9.18, 18:50	超市	智能手机	买东西
2018.9.26, 21:30	家	笔记本电脑	课后作业
.....			
2018.9.22, 21:05	家	网络电视	看电视

(1)学习场景数据集的语义唯一性编码

表3 编码后的学习场景数据集

时间	空间代码	空间语义	设备代码	设备语义	事件代码	事件语义
2018.9.12, 10:02	1 01	教室	d 01	平板电脑	e 01	课堂练习
2018.9.25, 10:50	1 01	教室	d 01	平板电脑	e 05	课堂讨论
2018.9.16, 18:00	1 01	教室	d 04	智能手机	e 02	玩游戏
2018.9.20, 18:30	1 02	地铁站	d 04	智能手机	e 02	玩游戏
2018.9.18, 18:50	1 04	超市	d 04	智能手机	e 06	买东西
2018.9.26, 21:30	1 03	家	d 02	笔记本电脑	e 04	课后作业
.....						
2018.9.22, 21:05	1 03	家	d 03	网络电视	e 03	看电视

该步骤是将采集到的规范化学习场景数据集进行语义唯一性编码,生成具有语义标注的学习场景数据集。根据学习场景的数据维度语义分类,对原始的学习场景数据集(见表3)的空间维度、设备维度、事件维度进行语义的唯一性编码,由此生成编码后的学

习场景数据集 R。其中,表中 101 的编码是代表空间维度的语义为“教室”,d01 的编码是代表设备维度的语义为“平板电脑”,e01 的编码是代表事件维度的语义为“做课堂练习”。

(2)基于时空聚类学习场景发现计算

该步骤是对编码后的学习场景集 R 进行时空聚类分析,识别出学习者不同类别学习场景的过程。将编码后的学习场景数据集 R 以时间为关键字,按照时间的先后顺序进行排序,得到排序后的学习场景数据结果集 R',遍历计算结果集 R'中相邻数据记录的时空相似度,将相似度高的学习场景数据记录合并为一个学习场景,生成相对独立的学习场景序列,并对合并后的学习场景中的事件维度进行频繁事件序列计算,由此生成的频繁学习场景数据集 FR 见表 4。表中的{(st01, <e01, e05>), (st01, e02)}的序列集表示学习者在课堂学习场景下的事件项集,项集中 st 表示学习场景的编码标识,如 st01 代表“课堂学习”场景,e 表示事件维度的编码标识,如:e01 代表“课堂练习”事件。

表 4 频繁场景数据集

时空唯一标识	某一时空发生的事件序列集
st01	{ (st01, <e01, e05>), (st01, e02) }
st04	{ (st04, e06) }
st02	{ (st02, e02) }
.....	
st03	{ (st03,<e03, e04>) }

(3)频繁学习场景挖掘计算

该步骤是对前面生成的频繁学习场景数据集 FR 进行事件频繁项的支持度计算的过程,便于挖掘不同学习场景的行为模式。扫描频繁学习场景数据集 FR,根据时空关联的事件项,统计不同学习场景下产生的事件频繁项的支持度,由此生成不同学习场景下的事件频繁项支持度的结果集 RS,见表 5,其结果集的序列模式为:学习者在教室进行课堂练习和课堂讨论的频繁度达到 50%,在教室里玩游戏的频繁度达到 40%,在地铁站里玩游戏的频繁度达到 65%等。

表 5 不同学习场景的事件频繁项支持度统计结果集

事件项集	支持度
(st01, <e01, e05>)	50%
(st01, e02)	40%
(st02, e02)	65%
.....	
(st03,e03)	10%

(4)基于频繁场景挖掘的学习者特征值计算

该步骤是基于上述生成的不同学习场景的事件频繁项的支持度的结果集 RS,提取与各维度的学习者特征量化指标集,统计频繁度大于 1 的事件项,再融合各维度学习者特征值的计算方法,计算各维度的学习者特征值,构建了 6 维度的个性化学习者模型。其具体的挖掘算法如下:

输入:

√不同学习场景的事件频繁项的支持度的结果集 RS={RS₁,RS₂,RS₃,RS₄,RS₅,RS₆},其中 RS_i表示第 i 个维度的学习者特征的量化指标集;

√各维度量化指标的最小支持度 min-sup。

输出:各维度学习者特征值的结果集

Step1:初始化 RS 数据集;

While(RS 不为空)

{

Step2:遍历一次结果集 RS,依次计算 RS_i的量化指标的每个事件项的支持度;

Step3:删除支持小于最小支持度的频繁项,将剩下频繁项按照支持度的大小排序得到频繁项集 RS_i;

Step4:基于生成的各维度学习者特征的频繁 1-项集 RS_i,结合各维度的学习者特征计算方法,计算各维度的学习者特征值;

}

Step5:输出最终各维度的学习者特征值的结果集,生成 6 维个性化学习者模型。

以上基于频繁学习场景挖掘的学习者特征值计算方法,不但从不同学习场景描述了学习者特征状态的变化过程,而且还有利于发现学习者的频繁事件序列模式,进而挖掘出学习者的学习行为模式,便于构建更客观、更准确的个性化学习者模型。

3. 个性化学习者模型的形式化描述

基于前面融合场景特性的 6 维学习者特征模型分析,本文的个性化学习者模型的形式化将采用 6 元组 LM 来表示,每一个元组对应每一个学习者特征维度。该 6 元组表示为:LM=(LM_B,LM_C,LM_S,LM_M,LM_R,LM_P),其中,LM_B代表基本信息子模型,LM_C代表认知水平子模型,LM_S代表学习风格子模型,LM_M代表情感状态子模型,LM_R代表社会网络子模型,LM_P代表兴趣偏好子模型,并且每一个子模型的元组是由序偶对<δ,φ,LM>来组成,其中,δ表示时间维度的信息,φ表示空间维度信息,LM表示某一学习场景下的学习特征状态值。基于这样的形式化描述,由此构成的学习者模型的形式化描述如图 5 所示。

$$LM = \begin{bmatrix} \langle \delta_B, \Phi_B, LM_B \rangle \\ \langle \delta_C, \Phi_C, LM_C \rangle \\ \langle \delta_S, \Phi_S, LM_S \rangle \\ \langle \delta_M, \Phi_M, LM_M \rangle \\ \langle \delta_R, \Phi_R, LM_R \rangle \\ \langle \delta_P, \Phi_P, LM_P \rangle \end{bmatrix} = \begin{bmatrix} \langle \delta_{B_1}, \Phi_{B_1}, LM_{B_1} \rangle, \langle \delta_{B_2}, \Phi_{B_2}, LM_{B_2} \rangle, \dots, \langle \delta_{B_n}, \Phi_{B_n}, LM_{B_n} \rangle \\ \langle \delta_{C_1}, \Phi_{C_1}, LM_{C_1} \rangle, \langle \delta_{C_2}, \Phi_{C_2}, LM_{C_2} \rangle, \dots, \langle \delta_{C_n}, \Phi_{C_n}, LM_{C_n} \rangle \\ \langle \delta_{S_1}, \Phi_{S_1}, LM_{S_1} \rangle, \langle \delta_{S_2}, \Phi_{S_2}, LM_{S_2} \rangle, \dots, \langle \delta_{S_n}, \Phi_{S_n}, LM_{S_n} \rangle \\ \langle \delta_{M_1}, \Phi_{M_1}, LM_{M_1} \rangle, \langle \delta_{M_2}, \Phi_{M_2}, LM_{M_2} \rangle, \dots, \langle \delta_{M_n}, \Phi_{M_n}, LM_{M_n} \rangle \\ \langle \delta_{R_1}, \Phi_{R_1}, LM_{R_1} \rangle, \langle \delta_{R_2}, \Phi_{R_2}, LM_{R_2} \rangle, \dots, \langle \delta_{R_n}, \Phi_{R_n}, LM_{R_n} \rangle \\ \langle \delta_{P_1}, \Phi_{P_1}, LM_{P_1} \rangle, \langle \delta_{P_2}, \Phi_{P_2}, LM_{P_2} \rangle, \dots, \langle \delta_{P_n}, \Phi_{P_n}, LM_{P_n} \rangle \end{bmatrix}$$

图5 学习者模型的形式化描述

(三)模型的应用框架设计

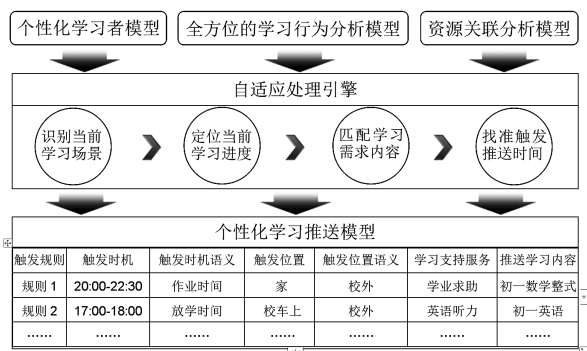


图6 场景化的个性化学习服务推送的演化框架图

为了形成一个闭环、自适应的个性化学习者模型架构,本文基于上述构建的个性化学习者模型,进一步分析了互联网学习环境下个性化学习服务推送的演化过程,提出了一个场景化个性化学习服务推送的演化框架(如图6所示)。该演化框架结合个性化学习者模型、学习行为分析模型和资源关联分析模型等,利用相关的数据挖掘算法,设计了自适应的学习服务推送引擎,以识别出学习者的当前学习场景,定位到当前学习者的学习进度,匹配学习者需求的学习内容,找准恰当的推送时机,选择恰当的推送方式,主动地为学习者推送适配的学习资源。通过分析互联网学

习环境下的学习服务推送的演化框架,及时地了解学习者的真实学习需求,从而高效、高质量地为学习者推送精确的学习服务信息,促使学习者采取自我改进的学习策略,主导控制自己的学习,从而有效地提高学习者的学习效率。

五、结 语

随着互联网+教育的深度融合,如何构建一个精确性高、实时性强的个性化学习者模型以应对互联网学习环境下高精度的个性化学习服务要求,是加快推进教育现代化进程的重要阶段。本文提出的基于场景感知的个性化学习者模型,是通过深入分析学习者特征维度的基础上,将场景特性融合到学习者建模的设计中,构建一个能表征不同学习场景下学习者特征状态变化的个性化学习者模型,并基于该模型设计了场景化的个性化学习服务推送的演化框架,试图从场景的视角探索新型的个性化学习服务模式,以满足互联网学习环境下高精度的个性化学习服务需求。目前,模型的研究尚处于初始探索阶段,下一步的工作是对模型进行迭代优化与完善,基于该模型设计开发原型系统并应用到实践中进行实验,以验证该模型的有效性与准确性。

[参考文献]

- [1] 陈丽,郭玉娟,王怀波,郑勤华.新时代信息化进程中教育研究问题域框架[J].现代远程教育研究,2018(1):40-46,87.
- [2] 武法提,黄石华,殷宝媛.场景化:学习服务设计的新思路[J].电化教育研究,2018(12):63-69.
- [3] IEEE Learning Technology Standards Committee(LTSC)[EB/OL].[2018-12-01]. <http://ltsc.ieee.org/>.
- [4] IMS learner information package summary of changes version 1.0.1 final specification[EB/OL].[2005-01-04].https://www.imsglobal.org/profiles/lipv1p0p1/imsip_sumcv1p0p1.html.
- [5] 祝智庭.网络教育技术标准研究[J].电化教育研究,2001(8):72-78.
- [6] BRUSILOVSKY P, MILLÁN E. User models for adaptive hypermedia and adaptive educational systems [J]. The adaptive web, lecture notes in computer science, 2007(4321):3-53.
- [7] 王晓琳,高丹丹,张际平.智能授导系统中的学习者本体构建[J].电化教育研究,2009(4):56-58,77.
- [8] 陈仕品,张剑平.适应性学习支持系统的学生模型研究[J].中国电化教育,2010(5):112-117.
- [9] 王怡涵.能力导向终身学习个性化推荐系统学习者模型构建[D].长春:东北师范大学,2011.
- [10] CHRYSAFIADI K, VIRVOU M. Student modelling approaches: a literature review for last decade[J]. Expert system with applications, 2013(40):4715-4729.

- [11] 黄焕. 面向 e-Learning 的学习者情感建模及应用研究[D]. 武汉: 华中师范大学, 2014.
- [12] 牟智佳, 武法提. 电子书包中基于学习者模型的个性化学习资源推荐研究[J]. 电化教育研究, 2015, 36(1): 69-76.
- [13] JELENA N, ANDRINA G, VLADO G. Anatomy of student models in adaptive learning systems: a systematic literature review of individual differences from 2001 to 2013[J]. Educational computing research, 2015, 51(4): 459-489.
- [14] TADLAOUI M A, AAMOU S, KHALDI M, CARVALHO R N. Learner modeling in adaptive educational systems: a comparative study[J]. Modern education and computer science, 2016, 8(3): 1-10.
- [15] 岳俊芳, 陈逸. 基于大数据分析的远程学习者建模与个性化学习应用[J]. 中国远程教育, 2017(7): 34-39.
- [16] 孙力, 张婷. 网络教育中个性化学习者模型的设计与分析[J]. 远程教育杂志, 2017, 35(3): 93-101.
- [17] 孙海民. 个性特征对网络学习行为影响研究的关键问题探究[J]. 电化教育研究, 2012, 33(10): 50-55, 63.
- [18] 哈佛·席尔瓦. 多元智能新视点丛书: 多元智能与学习风格[M]. 张玲, 译. 北京: 教育科学出版社, 2003: 146.
- [19] ACHAR A, LAXMAN S, SASTRY P S. A unified view of the apriori-based algorithms for frequent episode discovery[J]. Knowledge & information systems, 2012, 31(2): 223-250.
- [20] BIRANT D, KUT A. ST-DBSCAN: an algorithm for clustering spatial-temporal data [J]. Data & knowledge engineering, 2007, 60(1): 208-221.

Research on Learner Modeling Based on Context-awareness

WU Fati, HUANG Shihua, YIN Baoyuan

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

[Abstract] In the era of "Internet + education", as traditional learning scenes are constantly being reconstructed, the new learning scenes are gradually developing into dynamic, diversified and fragmented models, which makes new personalized learning service models based on scene mining more and more important. However, the learner model, as the core component of the personalized learning service system, cannot meet the high-precision personalized service requirements in the new era. Therefore, this paper proposes a learner modeling method based on context awareness. On the basis of the analysis results of recent studies on learner feature analysis models, together with the scene features of learners, a 6-dimensional learner feature analysis model is designed. Then, the context-awareness modeling method and frequent sequence mining algorithm are used to calculate the value of learner feature in each dimension under different learning scenes, and a personalized learner model with scene features is constructed. Finally, based on that model, the evolution framework of personalized learning service recommendation is discussed.

[Keywords] Internet + Education; Context-awareness; Learner Characteristics; Learner Modeling