

基于教育大数据的大规模私有在线课程中 交互活动模式的研究

程 罡¹, 孙 迪², 尚伟伟³

(1.国家开放大学 学习资源部, 北京 100039;

2.大连理工大学 高等教育研究院, 辽宁 大连 116024;

3.华东师范大学 马克思主义学院, 上海 200062)

[摘 要] 作为我国开放高等教育的主要形式,大规模私有在线课程(MPOCs)的学习管理系统积累了大量学习者行为数据。对于这些数据的分析,能够有效地探究教学与学习的规律,提高开放高等教育的实践和评估。文章运用隐马尔科夫模型识别了 MPOCs 中不同课程阶段的学习者交互活动的序列模式。研究表明,MPOCs 中的学习者在周学习和考试周的行为模式存在明显差异,他们会根据不同的课程阶段调整学习策略和侧重点,但小组学习和讨论对于学习成效的影响不大,这一研究结果与传统主流研究中关于协作学习及讨论的正向研究结果有比较明显的差异。这些研究结果有助于学习者和教师动态地梳理教育教学过程,促使他们采用不同学习策略来促进教学与学习,以及从更加细致的角度来评估开放高等教育的教学效果。

[关键词] 大规模私有在线课程;交互活动;模式分析与评估;隐马尔科夫模型

[中图分类号] G434

[文献标志码] A

[作者简介] 程罡 (1982—),男,湖南娄底人。副教授,博士,主要从事在线教育与学习分析研究。E-mail:chenggang@ouchn.edu.cn。孙迪为通讯作者,E-mail:sundi@dlut.edu.cn。

一、引言

随着信息技术的发展,支撑在线学习的学习管理系统(LMS)在高等教育中得到广泛的应用。这些系统收集和记录了大量的学习者交互数据、个人数据、系统信息和学术信息^[1],这些数据是课程层面的重要的教育大数据^[2]。在此基础上,针对教育大数据的研究方法——学习分析技术(Learning Analytics, LA)也应运而生。学习分析突破了传统研究方法在数据和分析方法上的局限,通过数据挖掘的算法,对教育大数据中的交互模式进行深度挖掘和分析,可以呈现出传统研究方法无法探究的潜在学习模式,加深了我们对教学与学习的认知、理解和评估^[3],有助于我们更好地对学与教的过程进行反思和预测、干预和优化^[4]。

研究表明,不同的学习群体在不同的学习阶段,其活动模式也有着不同的特征^[5]。然而,这些研究仅限于小部分学习者在传统协作学习情境下的行为,并没有针对大规模在线环境的学习者,特别是在大规模私有在线课程(Massive Private Online Courses, MPOCs)中的学习者^[6]。MPOCs 是我国开放高等教育(即开放大学)的主要授课形式,对于 MPOCs 学习者的关注并开展深入研究是非常适时和有必要的。

二、研究现状

(一) 基于教育大数据和学习分析的学习者交互研究

现阶段,基于教育大数据,一些研究人员采用了多种学习分析技术,如聚类、序列模式挖掘和隐马尔

基金项目:2021 年大连理工大学基本科研项目“基于学习分析和 AIOT 技术的教学研究”[项目编号:DUT21RC(3)046]

可夫模型(hidden Markov model,HMM)等,对学习过程中的交互模式进行了挖掘和探究。

在探究在线学习者协作学习行为的特征上,Perera等人采用了聚类 and 顺序学习模式挖掘的方法追踪分析在线协作学习的交互数据^[7]。该研究比较了不同组的结果,确定了某些最常见和最不常见的模式,从而使学习者和教师能够看到群体的学习行为,发现问题,并及时提供反馈。另外,Martinez等人利用序列挖掘的方法分析了学习者的频繁连续合作模式^[8],并探索了高、低不同成就群体行为模式的差异。为了探究学习者的行为模式与学习绩效的关系,Jeong等人专注于研究不同成就的学习者之间的行为差异^[9],他们使用HMM方法重点分析了学习者在活动模块与过渡行为之间的操作模式。结果表明,高成就学习者的学习行为更具线性,并且在不同的课程过程中也表现得更为一致。

在探究在线学习者的行为模式和交互轨迹上,Chiu等人利用时序分析的方法分析了187名大学生在学习过程中生成的概念图数据,从中挖掘出了典型的学习行为模式^[10]。Kinnebrew等人采用情境化和差分序列挖掘的方法评估并比较了40名8年级学生在网络环境下的交互行为^[11]。研究人员运用序列模式挖掘得到了不同小组的交互指标,并采用差分序列挖掘的频率算法和分级技巧,极大地提高了研究效率。

对于特定的研究主题,如预测评估、学习风格等也是研究者关注的重点。Venant等人使用序列模式挖掘的方法发现了学习者行为模式与其在最终评估测试中的表现之间存在一定的关系,其研究结果确认了某些学习策略与学习成绩之间的相关性^[12]。Fatahi等人利用数据挖掘的算法挖掘了一些高频的序列行为模式。结果表明,序列行为模式对于学习风格的预测具有较高的准确性^[13]。

通过文献分析,可以发现以往的研究大多是基于一定设计的实验研究,集中在协作学习、校园学习、专题研究性学习等情境中,研究对象的范围相对狭窄。大多数研究没有涉及MPOCs中学习者的非实验性的日常学习行为分析。其次,现有研究过于注重不同成就群体的行为差异研究,忽略了对不同学习阶段的行为模式的研究。此外,研究人员没有明确区分在线学习的交互行动和交互活动。事实上,学习者之间的交互至少可以分为两个层次:底层是微观的交互行为,指具有教育意义的典型行为,如查看、删除、写入、创建等;第二层是交互活动,一般指学习支持系统中设计的学习活动模块,每个模块包括一系列的底层交互

活动,如在Moodle中,对应于模块,交互活动可能包括介绍、内容、论坛、测验、作业等。一些研究人员侧重于挖掘交互行为,而其他人可能关注交互活动。在一个理论框架内明确地区分这两个层次的学习者交互,可以使我们的研究重点更具体和突出,从而连贯清晰地推进在线学习中的交互研究。

(二)典型的学习分析方法

由于研究学习交互的侧重点不同,研究者所采用的学习分析的方法也有很大差异。针对交互行为的研究方法集中于序列挖掘算法和聚类算法,而针对交互活动的研究方法通常采用隐马尔可夫模型(Hidden Markov Model,HMM)。

HMM以贝叶斯统计为基础,是一种基于显性数据的概率揭示隐性状态的方法。在以教育大数据为基础的学习分析研究中,HMM通过对一段时间内大规模显性学习行为数据的时序特征进行概率挖掘,进而分析出显性行为背后的潜在行为模式,以揭示学习的内在规律^[14-16]。HMM由3组参数来揭示隐性活动的状态:

- 初始概率向量 π :隐性活动状态的初始概率。
- 转移概率矩阵 A :每个隐性活动之间的转移概率,表示学习者在不同的隐性状态中转移流动的可能程度。
- 输出概率矩阵 B :用于检测隐性活动状态下特定显性活动的概率。因为每个隐性活动状态可能包括几个可观察的显性活动。显性活动概率表示在一个特定的隐性活动状态中,组成这个隐性状态的每一个显性活动的概率。

由一系列连贯的隐性活动状态组成的学习活动序列就是在线学习交互活动的潜在模式。通常,研究人员采用著名的Baum-Welch(BW)方法来计算HMM的参数 (π, A, B) ^[17]。给定每个数据集,HMM中输入的显性活动序列是可观察到的学习者从课程开始到结束的所有纳入研究范围之内的交互活动序列。针对一个显性活动序列的数据集,HMM算法可以挖掘出若干个参数集 (π, A, B) ,而我们需要的是一个最优参数集 (π, A, B) ,这个最优的参数集可以最大限度地拟合可观察的活动序列所组成的隐性活动状态序列^[14,18]。HMM输出的最大拟合的潜在活动模式是在确定最佳隐性活动状态基础上生成的最佳参数集 (π, A, B) 。

综上所述,通过对文献研究和学习分析方法的梳理,在本研究中,尝试突破现有研究有限的实验设计、学习环境范围不足、研究群体过窄、研究数据量级较小等局限,聚焦于大规模私有在线课程中学习者的非实验性的日常学习行为,采用HMM模型作为分析方

法,侧重于交互活动层面上的典型模式分析,探究不同学习阶段中学习者的潜在活动模式。因此,本文的研究问题如下:

- (1)在 MPOCs 的不同阶段,学习者的日常交互活动有哪些典型模式;
- (2)不同阶段的学习模式有哪些异同。

三、研究方法与过程

(一)数据与方法

本文的研究背景是开放大学中一门为期 20 周的在线课程。共有 1527 人最终完成课程学习并参与了期末考试,所以本研究的样本量为 1527。该课程共 12 个学习活动模块:介绍、公告、学习内容、资源、作业与测验、讨论、常见问题、实验指南、小组学习、学习绩效、模拟考试、课程评价。课程中,有 3 次以不同学习内容为主题的小组学习活动。课程分为两个阶段,前 16 周为学习周,后 4 周为考试周。这两个阶段将数据分为两个数据集:学习周数据集、考试周数据集。

本研究中,12 个学习活动就是 HMM 提到的可观察的显性学习活动。本研究的主要目标是以这 12 个显性学习活动为基础,挖掘其背后的隐性学习状态,以及由这些隐性学习状态组成的潜在学习活动的模式。HMM 输出结果是由不同隐性学习状态组成的具有时序特征的潜在学习模式。HMM 分别对两个数据集进行潜在模式挖掘,在此基础上,探究这两个课程阶段中学习者活动模式的异同。

(二)不同阶段的潜在活动模式

通过 HMM 算法分别对学习周和考试周两个不同的数据集进行挖掘,学习周和考试周的最佳隐性状态数量均为 4。这里仅以学习周数据集为例,梳理 HMM 分析的全过程。

如图 1 所示,学习周数据集的最佳隐性状态为 4。以此为基础,表 1 中学习周的初始向量 π 显示了学习者对于每个隐性状态的投入程度,如学习者对于第一个隐性状态的投入度是 30%。表 2 显示了各个隐性状态之间转移的概率。因为学习过程是流动的,学习者不可避免地会处于多个学习状态的变化之中。例如,表 2 中的第 2 行显示的是学习者在隐性状态 1 与本状态和其他三个状态之间流动的可能性。表 3 显示的是每个隐性的学习状态中包含的显性学习活动及其比例。例如,学习周中,隐性状态 1 中主要包含了“学习内容”“作业与测验”等 2 个显性的学习活动,第 2 列显示了每个学习活动所占用该状态的时间,“学习内容”占状态 1 全部时间的 46.1%，“作业与测验”

占状态 1 全部时间的 30.4%。

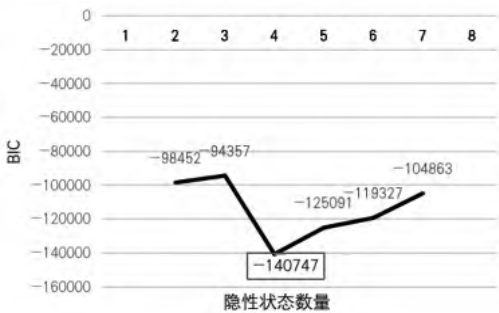


图 1 学习周数据集 BIC 与隐性状态数量

表 1 学习周初始概率向量 π

初始概率向量	状态 1	状态 2	状态 3	状态 4
π	0.3	0.045	0.635	0.02

表 2 学习周转移概率矩阵 A

转移概率矩阵 A	状态 1	状态 2	状态 3	状态 4
状态 1→	0.638	0.018	0.34	0.004
状态 2→	0.564	0.052	0.287	0.097
状态 3→	0	0.065	0	0.935
状态 4→	0	0.97	0.03	0

表 3 学习周输出概率矩阵 B

输出概率矩阵 B	状态 1↓	状态 2↓	状态 3↓	状态 4↓
介绍	0.019	0	0	0
公告	0.018	0.001	0.001	0.002
学习内容	0.461	0	0	0
资源	0.076	0.325	0	0
作业与测验	0.304	0	0	0
讨论	0	0.066	0	0.989
常见问题	0	0.001	0	0
实验指南	0.013	0	0	0
小组学习	0.097	0.607	0.02	0.009
学习绩效	0	0	0.979	0
模拟考试	0	0	0	0
课程评价	0.012	0	0	0

综合表 1—3 的内容,我们得出图 2,即学习周中学习者的潜在学习模式。表 1 与表 3 的数据在每个隐性状态节点的内部显示,表 2 的数据在每个状态节点之间通过箭头来标注。

通过图 2 我们可以看出,在学习周,学习者的潜在学习活动的模式由 4 个隐性状态组成。状态 1 占学习者整体投入度的 30%,状态 2 占学习者整体投入度的 4.5%,状态 3 占学习者整体投入度的 63.5%,状态 4 占学习者整体投入度的 2%。在隐性状态 1 中,频繁的显性交互活动有“学习内容”和“作业与测验”两类。学习者对于“学习内容”的投入时间为整个隐性状态

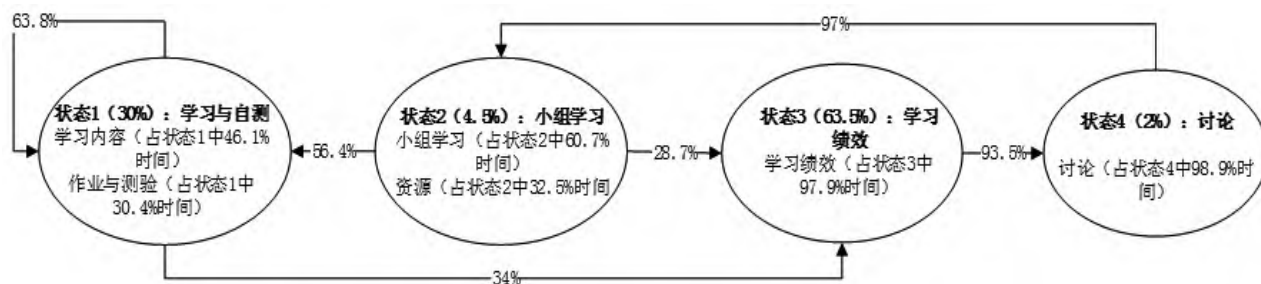


图2 学习周中学习者的潜在学习模式

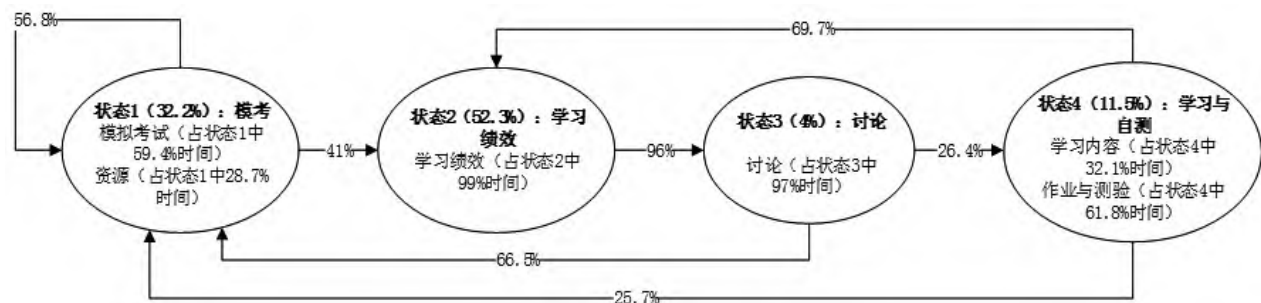


图3 考试周中学习者的潜在学习模式

1 的 46.1%，对于“作业与测验”的投入时间占状态 1 的 30.4%。基于这两个主要的显性活动，我们可以将隐性状态 1 定义为“学习与自测”状态。同样，隐性状态 2 可以定义为“小组学习”状态，因为学习者投入 60.7% 的时间来进行小组学习，投入了 32.5% 的时间来使用各种资源辅助自身的学习。隐性状态 3 是“学习绩效”状态。隐性状态 4 是“讨论”状态。这四个连贯的隐性学习状态连在一起，就构成了学习者在学习周的交互活动模式。图 2 中的箭头指示了在这个潜在的交互活动模式中，学习者在各个状态之间的流动和转移的可能性。如图 2 所示，在学习周，学习者从状态 1 开始他们的学习进程，之后有 63.8% 的可能性留在状态 1 继续学习课程内容并进行作业与测验，有 34% 的可能转移到状态 3 中查看自己和同伴的学习绩效。学习者进入状态 3 之后，有 93.5% 的可能转移到状态 4 中进行各种讨论。之后，学习者会进入状态 2。进行小组学习之后，学习者有 56.4% 的可能性会返回状态 1 展开进一步学习，有 28.7% 的可能性进入状态 3 查看学习绩效。

同样，图 3 显示了考试周中学习者的潜在交互活动模式。在考试周，学习者的潜在学习活动的模式也由 4 个隐性状态组成。状态 1 占学习者整体投入度的 32.2%，状态 2 占学习者整体投入度的 52.3%，状态 3 占学习者整体投入度的 4%，状态 4 占学习者整体投入度的 11.5%。在隐性状态 1 中，频繁的显性交互活动有“模拟考试”和“资源”两类。学习者对于“模拟考试”的投入时间为整个隐性状态 1 的 59.4%，对于“资

源”的投入时间占状态 1 的 28.7%。基于这两个主要的显性活动，我们将隐性状态 1 定义为“模考”状态。同样，隐性状态 2 可以定义为“学习绩效”状态，因为学习者几乎投入了全部的时间来查看学习绩效。隐性状态 3 是“讨论”状态。隐性状态 4 是“学习与自测”状态。这四个连贯的隐性学习状态连在一起，就构成了学习者在考试周的交互活动模式。图 3 中的箭头标注了在这个潜在的交互活动模式中，学习者在各个状态之间的流动和转移的可能性。如图 3 所示，在考试周，学习者从状态 1 开始他们的课程活动，之后有 56.8% 的可能性留在状态 1 继续利用各种资源进行模拟考试，有 41% 的可能转移到状态 2 中查看自己和同伴的学习绩效。之后，学习者有 96% 的可能转移到状态 3 中进行各种讨论。讨论之后，学习者会有 26.4% 的可能性进入状态 4 对学习内容和作用测验进行复习；有 66.5% 的可能性重新返回状态 1，再进行模考训练。而进入状态 4 的学习者，会有 69.7% 的可能性转移到状态 2 中查看自己和同伴的学习绩效，有 25.7% 的可能性回到状态 1 中，进行模拟考试。

HMM 清晰地挖掘出了在一段长时间的学习过程中，学习者的隐性学习状态和由这些状态组成的潜在学习交互活动的模式。这类研究结果对于我们把握学习者的学习活动和梳理课程的进展是非常有帮助的。

(三) 不同阶段的模式比较

基于图 2 与图 3 所示，以及上文对于学习者在不同课程阶段的交互活动模式的详细阐述，可以从以下

几个方面来梳理学习者在这两个阶段中学习模式的异同:

1. 潜在模式的构成

本研究中,学习周和考试周的潜在交互活动的模式都由4个隐性状态组成,但是构成两个模式的隐性状态和序列是不同的。学习周模式由“学习与自测”“小组学习”“学习绩效”“讨论”四个状态组成,与课程阶段主要相关的状态是“学习与自测”,占整个模式投入度的30%。考试周模式由“模考”“学习绩效”“讨论”“学习与自测”四个状态组成,与课程阶段主要相关的状态是“模考”,占整个模式投入度的32.2%。

2. 隐性状态的比较

图2与图3也明确显示,无论学习周还是考试周,“学习绩效”都是学习者投入度最高的活动,占到整个阶段性投入度的50%以上。学习者不仅频繁查看自己在课程中的行为绩效,也非常关注同伴的学习绩效,尤其是小组学习中同组成员的学习绩效。“讨论”也是两个阶段中学习者都投入的学习活动,但是与构成模式的其他隐性状态相比,“讨论”活动的投入度占比是最低的。此外,“小组学习”和“学习与自测”是学习者在学习周和考试周较为次要的不同活动。虽然在学习周有三次强制的小组学习活动,但是学习者对于“小组学习”的投入度仅占整个学习周的4.5%。

3. 学习者转移路径的分析

学习周活动模式的起始状态是“学习与自测”。作为学习周模式的主要活动,学习者用46.1%的时间用来进行课程学习,用30.4%的时间完成作业与自测。学习者在这一状态的黏性较大,有63.8%的可能性继续留在本状态进行学习,而另外34%的可能性会转到最具吸引力的“学习绩效”活动中去。从“学习绩效”到“讨论”,再到“小组学习”,是学习者在学习周从起始状态之后的主要转移路径。考试周活动模式的起始状态是“模考”,作为考试周的主要活动,“模考”的黏性也比较大,学习者有56.8%的可能性会留在本状态进行复习,但是也有41%的可能性转移到“学习绩效”模块,而且这个转移的概率比学习周从起始状态到“学习绩效”模块的转移概率要大。从“学习绩效”到“讨论”,再到“学习与自测”,是考试周学习者从起始状态之后主要的转移路径。通过比较,我们可以发现,两个阶段中,学习者从起始状态到“学习绩效”“讨论”,再到另外一个隐性状态的转移路径是相对类似的,但是该隐性状态之后的转移路径出现了较大的差异。这种差异显示了学习者不同的学习需求和策略,需要引起教师和课程设计者的关注。

四、讨论与总结

本研究侧重于交互活动的层面,采用一种探索性的学习分析方法——HMM对MPOCs中非实验性的日常课程的不同阶段进行了学习模式挖掘和分析。研究结果可以从以下几个方面提供一些价值和反思:首先,本研究给出了一个具体的利用学习分析技术进行教育大数据研究的例子,HMM的分析结果说明了学习分析的方法对于挖掘教育大数据的有效性和创新性,即不同课程阶段的交互活动模式确实有所不同,而且传统的定性或者统计的方法不能够充分地描述和定量地辨别这种潜在的差异。其次,两个课程阶段的不同模式的比较和分析为进行有效教学和促进学习提供了可参考的信息。

每个课程阶段学习者对于不同学习活动投入度的量化确定,可以提高研究者对学习者学习状况的把握程度。研究显示:不同的课程阶段,学习者都有特定的主要学习状态,这些学习状态一般由几个表面上相对独立的学习活动组成,学习者对这些活动的投入也有所不同。学习者在不同状态之间的转移也有一定的规律可循,例如,从主要学习状态转到绩效的查看和论坛的讨论,然后是对次要投入状态的关注和对主要学习状态的回归。又如,HMM的挖掘结果显示,“学习绩效”是学习者在两个阶段中投入度最高的活动,这有力地证明了学习者关心自己或同伴在整个群体中的排名,可能会影响学生的参与和学习的努力程度。因此,课程设计者可以在学习绩效活动中设计一些相关模块和接口,提供资源和反馈,引导学习者积极地投入到课程学习中。基于这些清晰立体的量化描述和分析,教师和学习者都可以更好地梳理学习过程,提高教学和学习的效果。

另一个值得讨论的方面是小组学习和讨论活动对于课程学习的有效性。众所周知,在线讨论和小组学习是许多在线课程中设计的典型活动^[19-21],支持者们认为在线讨论可以支持主动学习和知识构建^[22]。同样,许多关于协作学习的研究也表明,协作学习可以帮助学习者培养更高阶思维能力,建立更多的支持性关系以缓解孤立状态^[23-24]。然而,与之前研究结果不同,本研究中,发现学习者很少参与讨论和小组学习,这两个活动几乎都没有影响学习过程。由于样本局限在一门课程中,因此,研究结果不会导致对在线讨论和小组学习有效性的根本性挑战。但值得注意的是,在MPOCs背景下的论坛讨论和小组学习,可能有一些特殊的属性有待今后深入地进行研究。

与以往的实验性研究不同,本研究的结果更接近在日常现实状态下的真实状况,这有助于教师更好地掌握学习者真实的交互活动的特点和规律,以学习者为中心设计和开发课程。同时,研究仍存在部分局限性,其分析数据只来源于一所开放大学的一门课程,并且追踪的时间仅为一个学期,虽然样本数量和交互行为的数据量级较大,但是在更广阔的范

围内,本研究结果还不足以得出某种活动模式就是典型的有代表性的模式。在本研究的基础上,后续的研究中会继续拓展研究对象的范围,增加其多样性,最终通过对大量数据长期追踪和分析,不断挖掘典型的学习活动模式,建立不同的推荐系统,及时地为学习者提供支持和反馈,真正达到大规模个性化教学的目的。

[参考文献]

- [1] DANIEL B K. Big Data in higher education: the big picture [M]// DANIEL B K. Big data and learning analytics in higher education. Switzerland: Springer International Publishing, 2017: 19-28.
- [2] 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,2015,36(9): 54-61,69.
- [3] KRUMM A E, WADDINGTON R J, TEASLEY S D, et al. A learning management system-based early warning system for academic advising in undergraduate engineering [M]// LARUSSON J A, WHITE B. Learning analytics: from research to practice. New York: Springer, 2014: 103-119.
- [4] 祝智庭,沈德梅.学习分析学:智慧教育的科学力量[J].电化教育研究,2013,34(5):5-12,19.
- [5] 陈丽.远程学习的教学交互模型和教学交互层次塔[J].中国远程教育,2004(3):24-28.
- [6] 石磊,程罡,刘志敏,冯立国.大规模私有型在线课程建设模式及其质量保障机制——以国家开放大学网络课程建设过程为例[J].中国远程教育,2018(8):10.
- [7] PERERA D, KAY J, KOPRINSKA I, et al. Clustering and sequential pattern mining of online collaborative learning data [J]. IEEE transactions on knowledge and data engineering, 2009, 21 (6): 759-772.
- [8] MARTINEZ R, YACEF K, KAY J, et al. Analysing frequent sequential patterns of collaborative learning activity around an interactive tabletop [C/OL]. 4th International Conference on Educational Data Mining, 2011:111-120 [2021-07-08]. https://www.academia.edu/1408235/Analysing_frequent_sequential_patterns_of_collaborative_learning_activity_around_an_interactive_tabletop.
- [9] JEONG H, BISWAS G, JOHNSON J, et al. Analysis of productive learning behaviors in a structured inquiry cycle using hidden Markov models [C/OL]. 3rd International Conference on Educational Data Mining, 2010:81-90 [2021-06-13]. https://educationaldatamining.org/EDM2010/uploads/proc/edm2010_submission_59.pdf.
- [10] CHIU C, LIN C. Sequential pattern analysis: method and application in exploring how students develop concept maps [J]. The Turkish online journal of educational technology, 2012, 11 (1): 145-153.
- [11] KINNEBREW J S, LORETZ K M, BISWAS G. A contextualized, differential sequence mining method to derive students' learning behavior patterns[J]. JEDM-Journal of educational data mining, 2013, 5 (1): 190-219.
- [12] VENANT R, SHARMA K, VIDAL P, et al. Using sequential pattern mining to explore learners' behaviors and evaluate their correlation with performance in inquiry-based learning [C/OL]. European Conference on Technology Enhanced Learning, 2017:286-299 [2021-09-15]. <https://oatao.univ-toulouse.fr/19113/>.
- [13] FATAHI S, SHABANALI-FAMI F, MORADI H. An empirical study of using sequential behavior pattern mining approach to predict learning styles[J]. Education and information technologies, 2018, 23 (4): 1427-1445.
- [14] BAH I L, BROWN P F, DE SOUZA P V, et al. Maximum mutual information estimation of hidden Markov model parameters for speech recognition [J]. Proceedings of the IEEE-IECEJ-AS International Conference on Acoustics, Speech, and Signal Processing, 1986 (1): 49-52.
- [15] BEN-YISHAI A, BURSHTEN D. A discriminative training algorithm for hidden Markov models [J]. IEEE transactions on speech and audio processing, 2004, 12 (3): 204-217.
- [16] LI C, BISWAS G. A Bayesian approach for structural learning with hidden Markov models [J]. Scientific programming, 2002, 10 (3): 201-219.
- [17] BAUM L E, PETRIE T, SOULES G, et al. A maximization technique occurring in the statistical analysis of probabilistic functions

- of Markov chains[J]. Annals of Mathematical statistics, 1970, 41 (1): 164–171.
- [18] KWONG S, HE Q, MAN K F. Training approach for hidden Markov models[J]. Electronics letters, 1996, 32 (17): 1554–1555.
- [19] CHAN J, HEW K F, CHEUNG W S. Asynchronous online discussion thread development: examining growth patterns and peer—facilitation techniques[J]. Journal of computer assisted learning, 2009, 25 (5): 438–452.
- [20] GULBRANDSEN C, WALSH C A, FULTON A E, et al. Evaluating asynchronous discussion as social constructivist pedagogy in an online undergraduate gerontological social work course[J]. International journal of learning, teaching and educational research, 2015, 10 (4): 94–111.
- [21] O'MALLEY C. Computer supported collaborative learning[M]. Berlin: Springer Science & Business Media, 2012.
- [22] HEW K F. Student perceptions of peer versus instructor facilitation of asynchronous online discussions: further findings from three cases[J]. Instructional science, 2015, 43 (1): 19–38.
- [23] BARKLEY E F, CROSS K P, MAJOR C H. Collaborative learning techniques: a handbook for college faculty [M]. San Francisco, CA: John Wiley & Sons, 2014.
- [24] MASON W, WATTS D J. Collaborative learning in networks [J]. Proceedings of the national academy of sciences, 2012, 109 (3): 764–769.

A Study of Interaction Activity Patterns in Massive Private Online Courses Based on Educational Big Data

CHENG Gang¹, SUN Di², SHANG Weiwei³

(1.Department of Learning Resource, The Open University, Beijing 100039;

2.Graduate School of Education, Dalian University of Technology, Dalian Liaoning 116024;

3. College of Marxism, East China Normal University, Shanghai 200062)

[Abstract] As the main form of open higher education in China, the learning management system of massive private online courses (MPOCs) has accumulated a large amount of learner behavior data. The analysis of these data can effectively explore the patterns of teaching and learning and improve the practice and evaluation of open higher education. This paper uses the Hidden Markov model to identify the sequential patterns of learner interaction activities at different course stages in MPOCs. The results indicate that the learners in MPOCs differ significantly in their behavioral patterns during study weeks and examination weeks. They can adjust their learning strategies and focus according to different course phases, but group learning and discussion have little effect on learning effectiveness. This result is totally different from the positive results of traditional mainstream research on cooperative learning and discussion. These findings help learners and teachers to sort out the educational and teaching process dynamically in the teaching and learning process, promote them to adopt different learning strategies to facilitate teaching and learning, as well as to evaluate the teaching and learning of open higher education from a more detailed perspective.

[Keywords] Massive Private Online Course; Interaction Activity; Pattern Analysis and Evaluation; Hidden Markov Model