

doi: 10.16104/j.issn.1673-1891.2022.03.022

# 基于聚类算法的线上学习行为分析 ——以 Java 面向对象程序设计课程为例

贾丙静, 赵海燕

(安徽科技学院信息与网络工程学院, 安徽 蚌埠 233000)

**摘要:** 为了有效利用线上学习平台记录的学生学习行为数据, 从多方面刻画学习者画像, 充分发挥线上教学的作用, 以安徽科技学院物联网专业 2017 级 80 名学生为研究对象, 对其开展了 Java 面向对象程序设计课程线上教学。在人工智能技术的驱动下, 首先统计学生学习时间和视频内容数据, 分析他们的观看习惯和对重点、难点内容的重视程度; 然后在 K-Means++ 聚类算法的基础上, 分析课程视频、章节检测、学习次数、作业和签到等特征对学习效果的影响。结果表明: 上述方法可以帮助总结学生的学习态度、偏好和习惯, 将相似学习风格的学生聚为一类。老师可以通过线上学习行为的分析调整教学内容, 改进教学方法, 从而改善线上教学效果。

**关键词:** Java 面向对象程序设计; 学习行为数据; K-Means++ 聚类算法; 线上教学

**中图分类号:** TP312-4; G434 **文献标志码:** A **文章编号:** 1673-1891(2022)03-0119-04

## Analysis of Online Learning Behavior Based on Clustering Algorithm: Take the Java Object-oriented Programming Course as an Example

JIA Bingjing, ZHAO Haiyan

(School of Information & Network Engineering, Anhui Science and Technology University,  
Bengbu, Anhui 233000, China)

**Abstract:** In order to effectively use the student behavior data recorded by the online learning platform, depict the portraits of learners from various aspects, and take full advantage of the online teaching and learning, 80 students of grade 2017 majoring in Internet of Things in Anhui University of Science and Technology are taken as the research objects, and the online teaching of Java object-oriented programming course is carried out. With the help of artificial intelligence, firstly, we count the students' learning time and collect their video content data of Java object-oriented programming course, and analyze their viewing habits and their attention paid to key and difficult content. Secondly, we analyze the influence of course videos, chapter tests, study and sign-in times, and homework performance on students' study effects based on the K-Means++ clustering algorithm. The results show that the above methods help summarize students' learning attitudes, preferences, and habits to categorize students with similar learning styles. In conclusion, teachers can choose teaching contents and adjust teaching methods to improve online teaching effects based on the analysis of online learning behavior.

**Keywords:** Java object-oriented programming; learning behavior data; K-Means++ clustering; online teaching

### 0 引言

2021 年教育部部长陈宝生在全国教育工作会议上提出:“普及数字校园建设与应用, 扩大优质教育资源覆盖面”<sup>[1]</sup>。加强线上教育资源建设与应用,

利用信息技术更新教育理念, 变革教育模型, 因此, 线上教育在大数据和人工智能时代应用越来越广泛, 它的上课模式新颖、开放、自由, 学习者不受时间和地点的约束, 可以充分利用碎片时间获取学习内容, 弥补了传统上课模式呆板、不灵活、缺乏个性

收稿日期: 2022-04-07

基金项目: 安徽科技学院质量工程项目(X2019034)。

作者简介: 贾丙静(1982—), 女, 山东曹县人, 副教授, 博士, 研究方向: 计算机程序设计、机器学习。

化引导等问题,满足新时代终身学习的需求<sup>[2]</sup>。

然而,目前线上教学平台比较注重教学资源建设,没有综合考虑学习者的需求、兴趣、习惯、能力等因素,学生综合成绩和学生实际学习情况并不能一一对应。伴随着线上教学平台和人工智能的普及,大量的学习日志被记录下来,通过学习日志,利用算法分析学习者的行为,已成为当前人工智能助力下线上教育主要的发展趋势<sup>[3]</sup>。陈逸菲等<sup>[4]</sup>首先分析了不同分数段的学生行为,然后对比月、周和各时间段下学生的章节学习次数,得到优秀学生的学习主动性强,最后成绩和章节学习次数直接相关;张晓颖等<sup>[5]</sup>基于数据挖掘对学生行为进行预测和解释,从而优化学生管理;郭阳等<sup>[6]</sup>在多层兴趣模型的基础上,描述学习者对各因素的兴趣度,构建关系网,从而完成课程推荐;秦晓安等<sup>[7]</sup>对学习成效进行关联分析,发现线上平台任务设置不合理,学生学习积极性不高等问题,并给出合理化的建议;张丽园等<sup>[8]</sup>使用探索性分析方法,分析有利于实现个性化学习的行为。以上研究主要集中在学习数据的累加,缺乏深入的分析,没有充分发挥算法的优势,因此,本文研究 K-Means++ 聚类算法在线上学习平台上的应用,针对安徽科技学院物联网专业 2017 级 80 名学生 Java 面向对象程序设计课程线上学习特征进行多维度分析,建立合理的模型,分析各种行为特征对学习效果的影响,从而建立以学生为中心,教师为辅助的教学策略,帮助教师调整教学策略,学生进行个性化学习。

### 1 线上学习行为数据统计

Java 面向对象程序设计课程是计算机相关专业非常重要的一门基础课程,由于课程涵盖知识点多,并且比较分散,需要结合线上教学平台给学生更多的时间学习,同时也方便老师动态掌握学生的学习情况。超星学习通具有资源丰富、不同学习环境的数据同步和社交功能强等优点,被很多高校选做在线学习平台。在该平台上,老师可以直播课程、发布视频、课堂测试、发布签到和分享资源等,同时会记录学生每个视频的学习时长及次数、线上测试、作业、讨论和签到情况等,通过这些行为数据的分析,可以挖掘隐藏在背后的学生兴趣和态度,以及各视频知识点被关注程度,从而帮助老师根据学生特点调整教学资源,调整教学方法。本研究收集安徽科技学院 2019—2020 学年第 2 学期物联网 2017 级 80 名学生 Java 面向对象程序设计课程真实线上教学平台学习数据进行分析。该课程共 13 章,

64 学时,每周 4 学时,授课时间为 2—6 月;采用网上直播方式授课,并将一些重要知识点做成短视频供学生课后学习,共 18 个短视频。收集学生直播观看、签到、课程回放、短视频观看、作业、小测验、讨论、期末成绩等数据。

由于线上学习的特殊性,教师缺乏和学生及时沟通,这就需要学生提高学习的主动性和参与性,明确自己的主体地位,深入参与线上教学的各项活动,完成签到、直播观看、访问、作业、测试等,学生的主要数据图谱如图 1 所示。

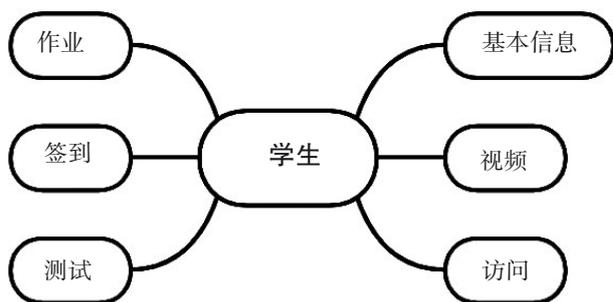


图 1 学生线上数据图谱

对这些特征数据的统计和分析,能反映学生的学习态度、偏好、规律和学习的难点,学生只有深入参与到线上学习的各项活动中,线上教学才会取得应有的效果。同时,对学生行为数据的统计也能帮助老师客观地认识学生,调整教学过程,对学生进行个性化指导。以 4 月份学生的线上学习数据为例,发现有些章节学生反复观看课程回放和课程相关的短视频,对这些课程和视频内容进行分析,发现这部分知识点本身就是整个课程中的难点和重点,如多态、抽象类和接口这 3 个知识点,说明大部分同学的学习态度比较端正,对于难的知识点反复观看,这也需要老师根据学生行为数据改进教学方法。

### 2 聚类算法

聚类算法是一种无监督学习方法,不需要大规模的标记数据进行训练,可以根据数据的特征自动把相似的数据聚为一簇。为了探索学生的学习风格,拟采用 K-Means++ 聚类算法分析图 1.1 统计的学生行为数据。K-Means++ 是在 K-Means 算法的基础上改进而来,早期的 K-Means 算法的主要思想是首先给定聚类个数  $K$ ,采用目标函数最小化的方法将原始数据  $\mathbf{X}=(x_1, x_2, \dots, x_i)$  分为  $K$  类,其中  $\mathbf{X}$  是  $d$  维向量。该方法在聚类的过程中,需要不断地修正与调整,初始聚类中心的选择会严重影响聚类的效

率和时间,聚类中心的选择太近或太远都会影响最终的效果,因此Arthur等<sup>[9]</sup>针对上述问题,提出了K-Means++算法,主要是改进K-Means初始中心选择方法,使K个聚类中心相互之间距离越远越好,其算法实现步骤如下。

步骤1:输入学生线上学习行为数据  $X = \{x_1, x_2, \dots, x_i\}$ 。

步骤2:从数据集中随机选取一个样本作为初始聚类中心  $c_1$ 。

步骤3:计算每个样本与当前已有聚类中心之间的最短距离,用  $D(x_i)$  表示;接着计算每个样本被选为下一个聚类中心的概率  $D(x_i)^2 / \sum (x_i \in X) D(x_i)^2$ ;最后,按照轮盘法选择出下一个聚类中心。

步骤4:重复步骤3,直到选择K个聚类中心。

步骤5:根据每个簇的聚类中心,计算每个样本与这些聚类中心的距离,并根据最短距离重新划分对象。

步骤6:重新计算每个(有变化)簇的聚类中心。

步骤7:循环步骤4~5,直到每个聚类中心不再发生变化为止。

步骤8:输出学生的聚类结果。

### 3 实验结果与分析

#### 3.1 数据预处理

在学生的线上数据中,选取学生学号、课程视频观看进度、章节测验平均分(满分100分)、学习次数(满分100分)、作业(满分100分)、签到(满分100)和卷面成绩(满分100分)进行分析。为了更好地完成聚类,需要使用相同的度量尺度来表示这些特征,因此先进行归一化处理,从而消除奇异值的影响,结果如表2所示。

#### 3.2 聚类结果分析

对表2中学生6个维度的线上学习行为数据,采用K-Means++算法进行聚类分析,总结学生的风格,并找出每类学生的关键特征,以便后期调整教学方法,优化教学资源。K-Means++重要的思想就是初始聚类中心之间的距离尽可能远,为了确定聚类个数,获得最优的聚类结果,结合肘部法来统计簇内误差平方和随着聚类簇数的变化,如图2所示。由图1可知,聚类簇数为4的时候,簇内误差平方和趋于稳定,因此最优聚类簇数为4。

K-Means++算法将80名学生聚为4类,为了展示聚类效果,可借助TSNE方法进行降维,将高维数据降低到2维或3维,结果如图3所示。由图3可

表2 物联网2017级Java面向对象程序设计课程学习数据

学号	课程视频	章节检测	学习次数	作业	签到	卷面成绩
1108170116	1.00	0.94	0.88	0.38	0.89	0.90
2603150126	1.00	0.72	1.00	0.70	1.00	0.89
2704170101	0.97	0.93	1.00	0.46	1.00	0.92
2704170104	1.00	0.72	0.75	0.32	0.95	0.79
2704170105	0.97	0.99	1.00	0.56	1.00	0.96
2704170106	1.00	0.93	0.79	0.59	1.00	0.79
2704170107	1.00	0.94	1.00	0.34	1.00	0.68
2704170108	1.00	0.99	1.00	0.61	1.00	0.94
2704170109	1.00	0.94	1.00	0.47	0.95	0.89
2704170110	1.00	0.94	1.00	0.60	1.00	0.90
...	...	...	...	...	...	...

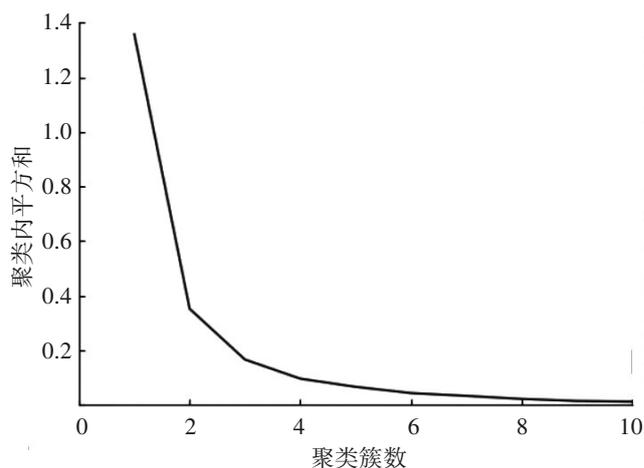


图2 肘部法求最优聚类簇数

知,上述80名学生明显地被分为4类:(1)第1类学生为紫色倒三角聚集处,占学生总人数的22.8%,这类学生视频观看时间长,学习次数多,出勤率高,保质保量地完成老师发布的测试和作业,总体学习积极,态度端正,卷面成绩高,属于“优等生”;(2)第2类学生为蓝色五角星聚集处,占学生总人数的10%,该类学生虽然观看课程视频、章节检测、学习次数、作业和签到得分都不是很高,但是卷面成绩比较理想。这些学生短期学习爆发力比较强,可以参与竞赛或老师的课题,属于“竞赛生”;(3)第3类学生为绿色圆点聚集处,在总人数中占比42%,能基本完成老师布置的任务,但是缺乏探索精神,课上不活跃,卷面成绩中等,属于“良好生”;(4)第4类

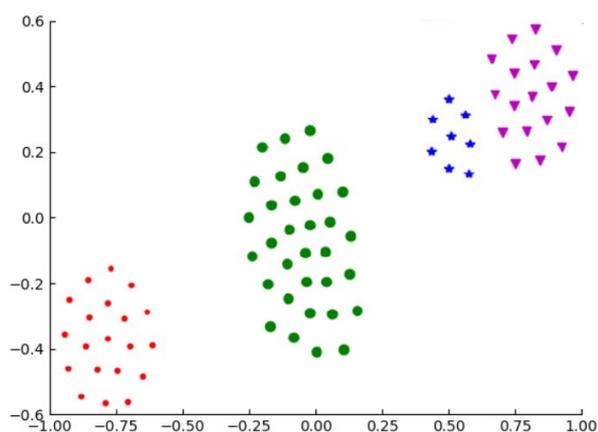


图3 聚类结果

学生为红色圆点处,该类学生占比为25.2%,学生学习习惯一般,课程参与度低,成绩中下等,属于“中等生”。根据上面的聚类结果分析,结合学生个性化特征,需要对Java面向对象程序设计课程的线上教学方案实施改进。具体如下:(1)增加课程的深度和广度,在具体知识点上深度挖掘,通过示例多角度展示知识点,同时增加更多学习资源,展示知识点在各方面的应用,增加课程的广度;(2)增加课

程主观评价得分,引导学生积极参与讨论,分层次设置课程的难度,提高学生学习的积极性和主动性。

#### 4 结语

通过以上学生线上行为数据的分析和挖掘可知,由于学生能力、基础、兴趣和习惯不同,同样的教学资源导致学习效果有很大的差别,K-Means++算法将学生分为4种类别。对不同类别和风格的学生,要想提高线上学习的效果,首先,作为学习者,要树立积极主动的学习态度和明确的学习目标,养成“课前预习-课中参与-课后复习”的学习习惯。其次,作为教育者,教师应根据学生的特点设置合理的线上任务,及时了解学生的学习进度,对学生进行个性化指导,将教学内容碎片化,学习现代教育技术,不断适应人工智能新时代教育的要求。最后,高校应结合自身情况,加强数字化校园建设,改善硬件教学条件,提高教师信息化水平,并提高考核和监察力度,积极鼓励教师进行教学改革,对学生进行个性化指导,从而从多方面来提高线上教学的效果。

#### 参考文献:

- [1] 微言教育.全文! 教育部长陈宝生在2021年全国教育工作会上的讲话[EB/OL].(2021-02-18)[2021-09-12].<http://jyj.fuxin.gov.cn/newsdetail.jsp?id=471823>.
- [2] 钟卓,钟绍春,唐烨伟.人工智能支持下的智慧学习模型构建研究[J].电化教育研究,2021,42(12):71-78.
- [3] 傅钢善,王改花.基于数据挖掘的网络学习行为与学习效果研究[J].电化教育研究,2014,35(9):5.
- [4] 陈逸菲,马辰来,翟慧茹,等.SPOC课程学习行为数据分析——以微机原理及单片机技术课程为例[J].软件导刊,2021,20(9):248-252.
- [5] 张晓颖,荣竹青,曲亮.基于大数据分析的高校学生行为模型的研究[J].长春大学学报,2021,31(2):1-5.
- [6] 郭阳,李全龙,李骥.基于学习者兴趣挖掘的个性化课程推荐方法[J].郑州大学学报(理学版),2021,53(4):77-82.
- [7] 秦晓安,王睿,程鸿芳.线上教学学生学习行为与成效的大数据分析研究[J].安徽商贸职业技术学院学报(社会科学版),2020,19(3):68-71.
- [8] 张丽园,洪如霞.基于数据分析的学生在线学习行为研究[J].豫章师范学院学报,2021(2):87-91.
- [9] ARTHUR D, VASSILVITSKII S.K-Means++: the advantages of careful seeding[C]//Proceedings of the Eighteenth Annual ACM-SIAM Symposium on Discrete Algorithms. New Orleans, Louisiana: ACM,2007.