基于在线学习行为分析的个性化学习推荐

陈晋音 方 航 林 翔 郑海斌 杨东勇 周 晓
(浙江工业大学信息工程学院 杭州 310000)

摘 要 随着在线课程和线上学习的普及,大量的在线学习行为数据被积累。如何利用数据挖掘技术分析积累的大数据,从而为教学决策和学习优化提供服务,已经成为新的研究重点。文中分析了在线学习的行为特征,挖掘学习者的性格特征与学习效率的关系,实现个性化学习方法推荐。首先,提取在线学习行为特征,并提出了一种基于 BP 神经网络的学习成绩预测方法,通过分析在线学习行为特征,预测其相应的线下学习成绩;其次,为了进一步分析学习者的在线学习行为与成绩的关系,提出了基于实际熵的在线学习行为规律性分析,通过分析学习者的在线学习行为,定义并计算相应的实际熵值来评估个体的学习行为规律性,从而分析规律性与最终成绩的关系;再次,基于 Felder-Silverman 性格分类法获得学习者的性格特征,对学习者实现基于 K-means 的聚类分析获得相似学习者的类别,将学习成绩较优的学习者的在线学习习惯推荐给同一类别的其他学习者,从而提高学习者的在线学习效率;最终,以某在线课程平台的实际数据为实验对象,分别实现在线学习行为特征提取、线下成绩预测、学习规律性分析和个性化学习推荐,从而验证了所提方法的有效性和应用价值。

关键词 在线学习行为,BP 神经网络,实际熵,Felder-Silverman 性格分析,个性化学习推荐
中图法分类号 TP181 **文献标识码** A

Personal Learning Recommendation Based on Online Learning Behavior Analysis

CHEN Jin-yin FANG Hang LIN Xiang ZHENG Hai-bin YANG Dong-yong ZHOU Xiao
(College of Information Engineering, Zhejiang University of Technology, Hangzhou 310000, China)

Abstract With the wide use of online courses and the population of online learning, massive data of online learning behaviors have been collected. How to take advantages of those accumulated data through novel data mining technology for improving teaching decision and learning efficiency is becoming the research focus. In this paper, online learning behavior features are extracted, relationship between online learner's personality and learning efficiency is modeled and analyzed, and personal learning recommendation is designed as well. First, online learner behavior features were extracted, and BP neural network based academic performance prediction algorithm was put forward, in which offline score was predicted based on accordingly online learning behavior features. Second, in order to further analyze the relationship of online learning behavior and offline practical score, a novel actual entropy based online learning behavior orderness evaluation model was proposed. Each learner's offline academic performance can be predicted on basis of online learning orderness. Third, learners' personalities were estimated through Felder-Silverman method. K-means algorithm was carried out on those personality vectors to achieve clusters of learners with the similar personality. Among those learners clustered into the same class, the top scored learner's learning behavior will be recommended to the rest learners. Finally, tackinga practical online courses platform's data as our experimental subject, plenty of experiments were carried out including online learning behavior feature extraction, offline academic performance evaluation and orderness analysis, perso-nal learning behavior recommendation, and the efficiency and application value of proposed method was proved.

Keywords Online learning behavior, BP neural network, Actual entropy, Felder-Silverman personality analysis, Personal learning recommendation

1 引言

随着信息技术的发展,网络在线学习在当代学生的学习生活所占的比例日益增大。基于在线学习行为产生的大量学习行为数据,如何利用大数据挖掘技术分析在线学习行为,更好地实现教学决策、学习过程优化和个性化学习方法推荐等,已经成为研究重点。

当前,国内外的很多研究都是针对参与者的学习行为进行描述,在选择研究对象和研究样本时存在一定的局限性;同时,有关在线学习行为的一些研究是在研究者的干预下进行的[1]。Strang 等[2]对用户在网络商务课程中的学习行为与反思性学习的影响等级进行了研究。Prior 从学习态度、信息素养和自我效能感 3 方面分析了对在线学习行为的影响程度[3]。Butcher 等研究了不同层次的先验知识和学习者在线

陈晋音(1982-),女,副教授,硕士生导师,主要研究方向为智能计算、优化计算、网络安全等,E-mail:chenjinyin@zjut.edu.cn;**方 航**(1996-),男,硕士生,主要研究方向为聚类分析和深度学习,E-mail:201505910304@zjut.edu.cn;**林 翔**(1995-),男,硕士生,主要研究方向为进化计算和深度学习,E-mail:201403080215@zjut.edu.cn。

学习行为间的关系，研究结果表明更高层次的先验知识会引发学习者进行更深层次的学习行为[4]。

从分析方法及构建分析模型的角度出发，一些学者提出了自己的构想。Chyan 等[5]构建了一个多层次的潜类模型来探讨平台用户的行为模式及区域差异。Park 等[6]采用教育数据挖掘中的聚类方法提取韩国某高等教育学院的 612 门课程学习者的行为特征并进行了聚类分析。Shimada 等[7]采用机器学习和分类方法对非正式学习的学习行为和课程测试中有效性之间的关系进行了分析。Hwang 等[8]研究了合作规划学习行为和学习绩效的关系，并给出了学生在合作学习活动中的 6 种不同类别的学习行为。Fakir 等[9]借助研究数据算法对存储在 Moodle 中的数据进行分析，从而理解学习者在线学习行为的模式。

除此之外，有学者从不同的角度提出了新方法来分析与挖掘在线学习的行为[10]。Linan 等[11]利用教育数据挖掘与学习分析两种理论方法，分析和解释教育数据的相似性和差异性，以及大数据与在线课程的关系。Durksen 等[12]利用贝叶斯网络对在线学习平台中的学习者建立了一个心理需求的概率分析模型。Fitoussi 等[13]揭示了学习者在线学习过程中学习行为是如何发生的，并提出大数据、学习分析等方法技术能够对平台中的学习数据进行有效的分析和预测。Callum 等[14]通过增添 3 个变量扩展了技术接受模型来测量用户对在线学习的接受程度。

在学习者学习行为数据分析方法与建模方面，也有学者从不同角度进行了优化与创新。樊超等[15]依据人类动力学研究了学习者在线学习行为的特性，揭示出学习者在线学习行为呈现出显著的异质性。宗阳等[16]利用逻辑回归方法揭示了在线学习行为与效果之间的关联关系，找出了影响在线学习效果的关键指标。

综上所述，国内外关于学习分析技术的研究尚处于初步探索阶段，大多数研究是以文献研究问卷调查为主的。国内的在线学习分析技术主要研究概念、影响因素等理论，研究基于网络在线平台的数据分析的实证很少，且对在线数据的采集和数据的挖掘不够深入，没有对数据进行合理的可视化分析，不利于发现隐藏信息。因此，在线学习行为与学生线下成绩的关系，以及有效的学习指导方法，尤为值得关注。本文围绕以下 3 个研究问题展开研究。

RQ1：根据在线学习行为客观评价其学习效果，即如何利用在线学习行为预测其线下真实的成绩。

RQ2：在线学习行为的较规律性评价，并且分析这种规律性是否与成绩正相关。

RQ3：分析学习者的性格差异及其对应的在线学习行为，并为不同的性格人群推荐有效的在线学习方法。

2　在线学习行为分析

在线学习平台积累了大量的用户学习数据，包括用户的基本信息、在线学习行为、习惯统计数据、在线学习效果评估等。其中，在线学习效果评估是在线教育的重要环节，但在线教育的松散结构以及远程教学环境的开放性导致对学习者进行客观评估的难度非常大[17]。教育数据挖掘工作也将不仅仅只是基于各大网络学习平台参与情况的综合性分析，简单设计系统进行数据处理以提供可视化的信息。因此，本文针对上述问题进行学习行为特征的有效提取，并使用 BP 神经网络模型进行线下成绩的预测。

2.1　基于 BP 神经网络模型的成绩预测

BP 神经网络算法是一种有监督分类方法，其主要思想是：输入学习样本，使用反向传播算法对网络的权值和偏差进行反复的调整训练，使输出的向量与期望向量尽可能地接近。当网络输出层的误差平方和小于指定的误差时，训练完成，保存网络的权值和偏差。分类模型完成训练。

对用户的在线学习行为数据进行预处理等操作后，得到用户行为特征。利用行为特征来进行 BP 神经网络的训练，从而得到分类模型。

2.2　线上学习行为的规律性分析

在当前的研究环境下，定量分析学习者的行为方式和学习者的学术表现之间的关系是个性化教育的一个重要步骤。实验证明，一些学习者的个性特点（如责任心、真诚度等）与学术表现都具有很强的相关性，特别是自觉性。在实验中可描述为：学术表现的相关度是最大的[18-20]。因此，将线上学习的规律性作为一个重要的维度去预测学生学术表现，以增加模型的精准度。

为了衡量学生线上学习的规律性和学习成绩之间的关系，本文设计了两种方法：1）设计学生登录时间差散点模型；2）设计实际熵函数。

2.2.1　学生登录时间差散点模型

为了刻画学生登录学习时间的规律性，本文提出了相近两次登录之间的时间差散点模型：分别计算学生每两次登录之间的时间间隔，再加上一个实验周期中的学生登录次数，就可以将学生的线上学习时间分布刻画出来；将学习者的学习时间映射到二维坐标系中进行可视化，并制成用户学习时间散点图。由此，统计出各个用户的登录时间间隔的最大值、最小值、平均时间间隔以及去掉最大值和最小值的截尾平均值，并分析出它们与成绩之间的关系。选用相关性最高的加入到模型中，以提升模型的预测精度。

2.2.2　实际熵函数的设计

假定学习时间周期按周进行计算，作业也按周进行提交，因此本文按周作为时间段进行划分，将一周时间分成 7 等份，每一份的跨度为 1 天，一周编码为 1 到 7（例如，周一为 1，周二为 2，如此类推）。然后将学习者每次登录学习的时间映射到这个离散的时间序列（1，2，3，4，5，6，7）中，如果时间分布跨越在两天之中，那么就将此时间映射到时间长的时间序列上。例如，如果一个用户的开始学习时间是周一晚上21：00，结束时间为周二凌晨1：00，那么计算得到在周一学习时间为 3 小时，周二学习时间为 1 小时，这段学习时间映射到 1 上。使用实际熵函数[21-22]来衡量用户登录学习的时间规律性系数，定义如式（1）：

$$S_{\varepsilon}=x^{2}\left(\frac{1}{n}\sum_{i=1}^{n}\Lambda_{i}\right)^{-1}\ln n \tag{1}$$

其中，x 代表映射的时间序列所占用的星期数；n 表示映射的序列数，为了准确提取学习者的学习规律性，这里的 n 必须取连续的序列；Λ_i 代表第 i 个开始的前面没出现过的最短序列的长度。

例如，某用户 i 学习某课程映射到时间序列之后的序列为{1，4，1，4，5，1，4，1，4，1，4}，另一个用户 j 的学习时间序列为{1，1，3，1，3，3，2，1，3，2}，用户 k 的学习时间序列为{1，2，3，4，5，6，7，1，2，3}。用户 i 对应的 Λ 序列为{1，1，3，3，5，5，0，0，0，0}，j 对应的 Λ 序列为{1，2，1，3，2，2，1，3，0，0}，k 对应

的 Λ 序列为{1,1,1,1,1,1,1,0,0,0}。i 的 x 为 5,j 的 x 为 7,k 的 x 为 2。计算出用户 i 的实际熵值为 31.9803,用户 j 的实际熵值为 75.2178,用户 k 的实际熵值为 13.1576。

结果表明,用户 k 的实际熵值要小于前面两个用户,若序列所占用的时间长度越短,则熵的效果就越明显。若以实际熵作为评估指标,对学习者的学习规律性进行评估,则实际熵越小,学习的时间间隔即学习的时间规律性越强。

2.2.3 基于聚类的学生性格分析

不同的学生对应于不同的性格特征,对应会有不同的学习方法,促使个性化教育和网络在线教育联系起来。本文采用 Felder-Silverman 性格分类法对学习者进行性格统计分析,并使用 K-means 聚类算法对学生进行性格聚类,继而对不同性格的学生提出个性化的推荐学习方法。

2.2.4 Felder-Silverman 性格分类法

Felder 和 Silver 从学生差异的角度出发,围绕信息加工、感知信息、输入信息及内容理解 4 个方面,在每个方面将学习者划分成两种不同类型的学习者[23],最终得到 16 种不同的学习风格。Felder-Silverman 性格分类法在实际实验研究中不同于 FFM 人格五因素性格测试法需要结合人脸识别技术实时判断学习者的情绪[24];相比于 Cattle 16 性格分析方法,其不只是单纯地对学生进行特点的分类,而是明确其特点与网络在线学习行为间的相关联系[25]。因此,本研究最终采用了针对性强且更适用于网络在线学习平台的 Felder-Silverman 性格分类法。信息处理维度有活跃型(active)和沉思型(reflective);感知维度中有感悟型(sensing)和直觉型(intuitive)学习者;信息输入维度分为视觉型(visual)和言语型(verbal);内容理解维度下有序列型(sequential)和综合型(global)学习者 。

3 实验结果及分析

实验中的操作系统为 Windows10,集成开发环境为 Anaconda4.4.0,python IDE 为 Pycharm;硬件条件:CPU 为 Intel(R)Core(TM)i5-4200H CPU@2.80GHz,内存为 4GB。

3.1 实验数据集以及预处理

本文从玩课网平台所提供的开放数据接口采集学生在线学习的行为操作数据。根据系统的设计要求,学生的操作行为主要包括以下 6 种。

1)登录时间:每次的系统登录时间(已筛选出超过 30 分钟的时间为有效时间)。

2)浏览教学资源:行为数据是在线时间长短,必须配合鼠标键盘操作,在浏览视频资源过程中,不包含快进、认证失败等操作的学习行为才能作为有效的行为数据。

3)坛数据:主要是发贴和回帖的数量与质量,将字数、浏览回复率等作为有效数据。

4)在线笔记:记录学生笔记的数量。

5)课后作业完成情况:记录学生的作业总数,以及作业完成的数量和质量。

6)在线测试:包括每次提交的最高分和最低分。

鉴于玩课网平台与院校合作的特点,我们使用用户的真实期末考试成绩以及任课老师的评价作为类标进行综合评定,采用以下分类方式:

$$G=g\times 70\%+e\times 30\% \tag{2}$$

其中,g 代表学生期末考试的真实成绩,e 代表任课老师评定的成绩,G 代表综合评定成绩。

将综合评定成绩分成 3 个等级,60 分以下记为不及格,60~80 分记为良好,80~100 记为优秀。从线上学习平台的数据库我们共得到 5238 名同学的线上学习行为特征和真实成绩类标。实验中,为了进行交叉验证,我们将一次测试分为 10 组,每组随机选取 238 份数据作为测试集,剩下的 5000 份数据作为训练集,并将 10 组结果的平均值作为最终的实验结果。在训练之前,我们还将所有的数据在各个维度上按 z-score 标准进行归一化处理,处理后得到的数据值域为[-2,3],标准化的处理将学习行为多个维度的数据值统一控制在一个范围内,以便之后更好地进行处理与分析。

另外,结合玩课网平台提供的用户在线学习行为数据,锁定 110 位学习者,将其作为实验组,并对其进行性格调研,获得 101 位有效的在线学习者的性格问卷(所罗门性格调查问卷)调查数据。问卷共由 44 道单项选择题组成,这些题目包含了学习风格分类的 4 个维度,每道题的选项设计主要从每个维度对应的两种学习风格偏向出发,设置与其学习风格紧密相关的学习行为,且只有 a,b 两个选项[26]。在本研究中,我们采用计分的形式为:选 a 记 1 分,选 b 记-1 分。由此获得每个学习者性格问卷 4 维的得分。

3.2 评价指标

学生成绩预测实验采取准确率(P)作为评价指标。准确率是所有分类的学生中与分类结果吻合的学生所占的比率,定义为:

$$P=\frac{\text{分类正确的学生数}}{\text{实际分类的学生数}}\times 100\% \tag{3}$$

为了衡量学生线上学习的时间规律性与其学习成效之间的关系,我们使用斯皮尔曼(Spearman)相关系数进行定量描述。通过斯皮尔曼相关系数计算它们之间的相关度,从而确定两者的关系。斯皮尔曼相关度如式(4)所示:

$$r_s=1-\frac{6\times\sum d_i}{n(n^2-1)} \tag{4}$$

其中,$d_i=rg(X_i)-rg(Y_i)$ 为成绩排序等级和时间差排序等级的等级差。若大量实验样本的等级相同,则不能使用式(5)来计算,这里需要用定义式(即排名变量之间的斯皮尔曼相关系数)来计算,如式(5)所示:

$$r_s=\rho_{rg_X,rg_Y}=\frac{\mathrm{cov}(rg_X,rg_Y)}{\sigma_{rg_X}\sigma_{rg_Y}} \tag{5}$$

其中,rg_X 表示真实成绩的等级排序,rg_Y 表示时间差排序,$\mathrm{cov}(rg_X,rg_Y)$ 是等级变量的协方差,σ_{rg_X} 和 σ_{rg_Y} 是等级变量的标准偏差。

3.3 实验结果及对比分析

为了对学生成绩进行预测并对学生性格进行聚类分析,实验分成 3 个部分完成,并将所提方法与其他算法进行对比。最后,综合 3 个部分的实验结果,对相应的学生做出个性化的学习方法推荐。

3.3.1 利用 BP 神经网络预测成绩及分析

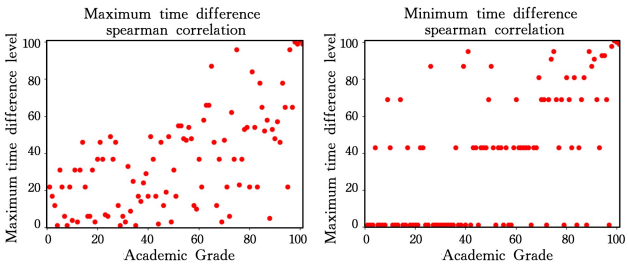
学习率设置为 0.001,采用了 sigmoid 激活函数,在迭代 200000 后,损失函数开始逐渐振荡。使用该神经网络分类模型进行教学评估时,精确度高达 69.748%,因此,该神经网络分类器可以用来参考评估玩课网线上学习者的行为。

对于 RQ1,RQ2 和 RQ3,实验结果给予回答。本文采用的分类类标是来自大学院校的真实学习者期末在该学科的综合测评,由式(2)可知,综合评估成绩不仅包括真实期末考试

的卷面成绩，还包括任课老师对学生平时表现的评价，相对客观公正，真实反映了学习的水平。因此我们得出的结论是：线上表现好的学习者的真实成绩未必好。实验中发现，有几个组数据在测试集中程序预测类标为 1 类（优秀），但真实成绩为 2 类（一般）；还有两组数据根据线上表现为 3 类（不及格），但真实成绩为 2 类（一般），这说明存在线上学习较少但在平时课堂教学中认真的学习者。

3.3.2　线上学习规律性与学习成绩之间的关系及分析

图 1—图 3 中，横坐标表示用户的登录时间，纵坐标表示用户相邻两次登录的时间间隔。针对每位所选实验对象，取得时间差的最大值、最小值、平均值，以及去掉两极端值的截尾平均数，分别将其作为一个规律性的维度，得到用户登录周期数据表；分别对表中数据进行等级排序，并对用户进行真实成绩等级排序，画出斯皮尔曼相关性散点图。



(a)　　(b)

图 1　最大时间差和最小时间差的相关性散点图

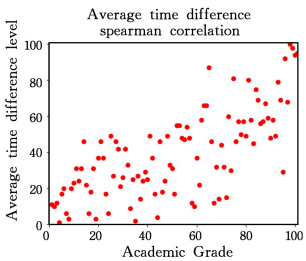


图 2　平均时间差的相关性散点图

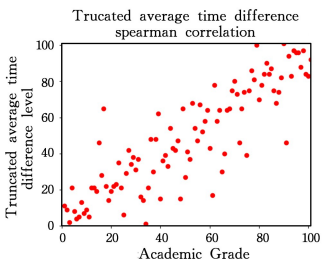


图 3　截尾平均时间相关性散点图

在计算最大时间差和最小时间差时，由于出现大量实验样本等级相同的情况，因此不能使用式(4)来计算，这里需要用定义式(5)即排名变量之间的 Pearson 相关系数来计算。

通过直观地观察得到上面 4 幅图中成绩与时间差等级的相关性，显然，图 3 中去掉最值影响的排序相关性更好。通过具体计算斯皮尔曼系数值来精准衡量相关性。最大时间差 $r_{s1}=0.34875$，最小时间差 $r_{s2}=0.51582$，平均时间差 $r_{s3}=0.67751$，截尾平均时间相关系数 $r_{s4}=0.87028$，与观察到的结果基本一致。通过计算学习者的实际熵函数，并给出相应的排序，再根据学习者实际成绩的排序，画出实际熵与成绩排序斯皮尔曼相关性散点图，如图 4 所示。

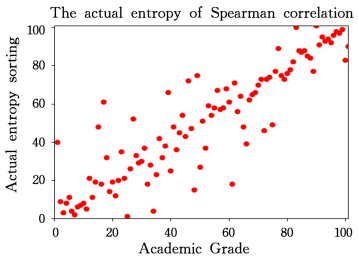


图 4　实际熵函数斯皮尔曼相关性散点图

由式(4)计算得出斯皮尔曼相关度 $r_s=0.89449$ 具有明显的正相关特性。直观地，具有较高有序性的学习者可能更自律，表现在学习上即为周期性地去复习所学知识，从而提高 GPA。

结合上述时间差等级相关性和实际熵函数的等级相关性，我们将相关度较高的实际熵函数加入到评估模型中，并进行模型的相关训练，最终使得测试集的正确率上升到 74.7900%。通过充分挖掘线上学习者的学习日志记录，提升了模型的预测精度。

3.3.3　基于聚类的性格分析

使用热图将上述性格得分进行可视化。热图是将学习者性格的距离相似度矩阵映射到二维坐标中，然后通过 RGB 显示出来。根据问卷数据，我们得到 101 位用户的性格得分，分别计算用户 i 和用户 j 之间的相似度距离，即使用公式：

$$\cos\theta_{i,j}=\frac{x_{i1}x_{j1}+x_{i2}x_{j2}+x_{i3}x_{j3}+x_{i4}x_{j4}}{\sqrt{x_{i1}^2+x_{i2}^2+x_{i3}^2+x_{i4}^2}\sqrt{x_{j1}^2+x_{j2}^2+x_{j3}^2+x_{j4}^2}} \tag{6}$$

得到的 $\cos\theta_{i,j}$ $(0\leqslant i\leqslant 100, 0\leqslant j\leqslant 100)$ 值域范围为 $[-1,1]$，将其映射到 $[0,255]$ 中，得到 101×101 的相似度矩阵。

将处理得到的性格维度数据进行 K-means 聚类分析，从而得到性格分类结果。K-means 聚类结果显示，当 $K=3$ 时，聚类结果最好，即各点距离各类质心的测量函数值最小。根据性格数据，101 位学习者的性格属性被分为 3 类。

通过对相似度矩阵的处理，对用户进行矩阵的行列交换，将上述聚在同一个类的用户交换到一起。如此操作，3 个类的用户都集中在一起，在热图的可视化中将看到在对角线上有 3 块对角块，说明聚类效果明显。图 5 为聚类前后的用户相似度矩阵热图可视化对比图。

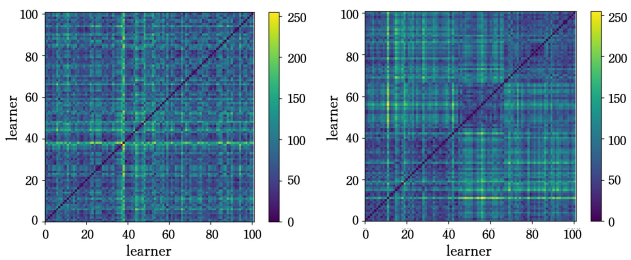


图 5　聚类前后用户相似度矩阵热图的可视化对比

在这 3 类中获取学习成绩最好的同学，即成绩类标为等级 1 的学生，抽取他们的共同线上学习特征。

性格分类为第一类的学习者属于沉思型、感悟型、视觉型和综合型学习者，但是其分数偏差并没有特别大。这类性格的学习者数量最多。我们筛选出这一类中学习效果较好的用户，并且对比他们的学习方法，得出其学习行为特征总结为视频浏览时间占整个课程学习时间的 55%以上，而文本模块时间所占比值为 15%左右，笔记记录量为 8 项左右，发帖量为 3 贴左右，测试次数都到达了 4 次以上；并且这类学习者偏好深

入思考问题，喜欢在安静环境下学习，善于全局思考问题，思维发散。建议这类学习者多思考、多感悟、多应用。

性格分类为第二类的学习者在感知维度、输入维度和理解维度的分数都很高，属于感悟型、视觉性、序列型学习者，同时属于活跃型学习者，其最佳学习行为主要特征参加第 8 类学习者学习行为，视频浏览时间占整个课程学习总时间的 65%以上，而文本模块时间所占比值小于 10%，笔记记录量为 2 项，发帖量为 8 贴，测试完成时间与测试次数值均较高，分别为 135 min 以上和 5 次。建议这类学习者充分利用自己观察能力强的特点记住所看到的信息的特点，在学习过程中多借助图片去加深自己对知识点的记忆，比如记英语单词时可采用小卡片的方式；或者通过流程图的形式去掌握知识结构。

性格分类为第三类的学习者在输入维度的分数偏高，属于视觉型学习者；在处理维度、感知维度和理解维度偏向于活跃型、直觉型和综合型，但分数不是很高。将这种性格类型的学习者的学习行为特点总结为视频观看时间占学习课程总时间的 65%，论坛浏览时间占总学习时间的 10%左右，发帖量较少，只有 2～3 贴，测试次数为 2～3 次。相对而言，该类学习者喜欢活跃的学习氛围，喜欢与别人讨论，在相互讨论的过程中听取不同人对知识的理解，从而加深自己的理解。建议此类学习者多加入团队合作，主动承担起团队的核心任务，通过团队协作提升自己的能力。

3.3.4 对比实验

为了验证使用 BP 神经网络来做学生成绩预测相比其他预测方法的优越性，我们设置了对比实验，分别与 ADTree[27]，ZeroR，NBM[28]方法进行横向对比。同时，为了验证加入时间规律性维度之后模型的分类情况，对每个模型分别进行包含规律性维度和不包含规律性维度的训练与测试，并将两次的实验结果进行纵向对比。

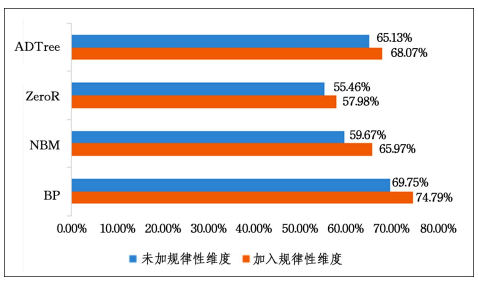


图 6 不同分类器对比实验结果的柱状图

同样地，使用在浙江省高等学校在线开放课程平台上收集到的包括 C++、电力电子技术等 5 门课程在内的学生在线行为数据进行实验，结果如图 7 所示。

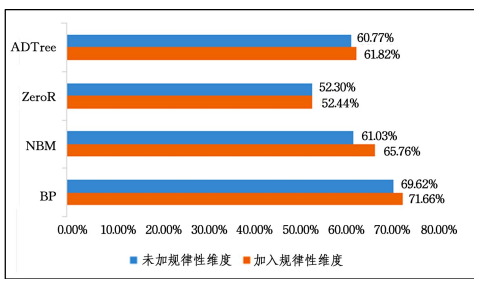


图 7 开放课程平台上对比实验结果的柱状图

从图 6 和图 7 可以看出，使用 BP 神经网络得到的成绩预测结果均优于其他模型，并且比 3 种对比算法中的最优结果高出了 4%～8%。此外，在纵向对比实验中可以发现，加入规律性维度取得的预测结果均优于未加入规律性维度时取得的预测结果，这说明规律性维度确实能有效挖掘在线学习数据的隐藏特征，并提高不同模型的预测能力。

由于本实验的类标来源于用户在线测试的测试成绩，在线测试题目的难度比真实测试的难度小，用户在线学习比课堂学习的影响因素小。因此，加上规律性维度后模型的预测效果增长幅度不大，且整体预测的正确率比玩课网的正确率低。通过对比实验得到，BP 神经网络模型无论是否加上规律性维度，其正确率都要明显高于其他算法。通过横向对比发现，加上时间规律性维度后，各种分类模型的分类效果均有所提高，由此回答了 RQ2。

结束语 网络在线教育为学习者提供了一种新的接受教育的途径，其教育模式的多样性，充分调动了学习者的学习主动性和积极性；其教育数据的精确收集，能让教育管理者及时掌握学习者的学习状况，并进行及时的干预和指导。因此，对在线学习数据的挖掘工作显得尤为重要。

本文首先对网络在线教育评估方法进行综述，突出传统的评估方法分为主观和客观两种。客观的评估方法本质上是考试的模拟过程，回避了传统考试的一些弊端，忽略了在线教育平台跟踪学习者学习状况进行数据采集的技术优势；而主观的评估方式并不能准确地评估学习者的真实能力水平。鉴于此，提出线上线下相结合的评估方式来综合评估学习效果。本文通过采集在线教育平台用户的学习日志数据，将预处理后的数据按照8∶1∶1 的比例做成训练集、验证集和测试集，再将采集到的真实学校成绩进行分类作为分类的类标，然后采用 BP 神经网络训练分类模型，从而对学习者的学习效果进行评估。

为了提升对学习者线上学习成效评估的精度，我们对在线数据进行深度挖掘，得到了在线行为的有序性对学习成绩的影响。

计算每位学习者的登录学习时间的间隔差，并进行排序，通过求得的真实成绩等级之间的斯皮尔曼相关度去判断登录的间隔与成绩等级的关系；另外，为了使评估登录的有序性更加合理，文中还设计了实际熵函数进行更加精确化的表达。

最后，我们对在线学习者进行性格调研分析，主要应用 Felder-Silverman 学习风格理论对网络在线学习行为进行具体的分析研究，并给予用户指导与建议；主要提出将聚类算法 K-means 运用于在线学习者性格分类与学习行为分类中，分析两者之间的关系，以及性格各维度属性值偏差对学习行为差异的影响，给出个性化的学习建议，以达到高效学习的目的。

本文研究还存在一些问题和不足：因为教育数据的缺乏，对学习效果的评估不够准确；对不同性格的学习者进行的学习方法推荐不能做到完全准确，存在一定的误差，需要日后再进行大量数据的实验研究以不断修正，使误差达到最小化；同时，分析研究学习者学习行为采集数据的类型不够全面，还需继续发掘出更多能反映出学生在线学习行为特点的属性数据，以完善模型。

参 考 文 献

[1] COATES H. Student engagement in campus-based and online education: University connections[OL]. http://www.bokus.com/cgi-bin/product_search.cgi?authors=Hamish%20Coates.

（下转第 452 页）

[11] FALANDYSZ J. Review:On published data and methods for selenium in mushrooms[J]. Food Chemistry,2013,138(1):242-250.

[12] YANG L,LUO P,CHEN C L,et al. A large-scale car dataset for fine-grained categorization and verification[C]// Computer Vision and Pattern Recognition. IEEE,2015:3973-3981.

[13] SHASHA D. Open Field Tic-Tac-Toe[J]. Communications of the Acm,2017,60(1):112.

[14] JONAS A. DieGSVP-Agenturen als Forschungsobjekt[M]// Das Governance-System der GSVP: Die Rolle des EU-Satellitenzentrums und der Europäischen Verteidigungsagentur. Nomos Verlagsgesellschaft mbH & Co. KG,2015:133-177.

[15] 赵继军,郭昆,冯楠,等. 基于 RSVP—TE 的有向泛洪 IRWA 算法研究[J]. 光通信研究,2013(5):8-11.

(上接第 426 页)

[2] STRANG K. How student behavior and reflective learning impact grades in online business courses[J]. Journal of Applied Research in Higher Education,2016,8(3):390-410.

[3] PRIOR D D,MAZANOV J,MEACHEAM D,et al. Attitude, digital literacy and self efficacy:Flow-on effects for online learning behavior[J]. Internet & Higher Education, 2016,29:91-97.

[4] BUTCHER K R,SUMNER T. How Does Prior Knowledge Impact Students' Online Learning Behaviors? [J]. International Journal of Cyber Behavior Psychology & Learning,2011,1(4):1-18.

[5] YANG C,HSIEH T. Regional differences of online learning behavior patterns[J]. Electronic Library,2013,31(2):167-187.

[6] PARK Y,YU J H,JO I H. Clustering blended learning courses by online behavior data:A case study in a Korean higher education institute[J]. Internet & Higher Education,2016,29:1-11.

[7] SHIMADA A,OKUBO F,YIN C,et al. Informal Learning Behavior Analysis Using Action Logs and Slide Features in E-Textbooks[C]// International Conference on Advanced Learning Technologies. IEEE,2015:116-117.

[8] HWANG W Y,SHADIEV R,WANG C Y,et al. A pilot study of cooperative programming learning behavior and its relationship with students' learning performance[J]. Computers & Education,2012,58(4):1267-1281.

[9] TOUYA K,FAKIR M. Mining Students' Learning Behavior in Moodle System[J]. Journal of Information Technology Research (JITR),2014,7(4):12-26.

[10] YE C,KINNEBREW J S,SEGEDY J R,et al. Learning Behavior Characterization with Multi-Feature, Hierarchical Activity Sequences[C]// Proceedings of the 8th International Conference on Educational Data Mining. 2015:380-383.

[11] LINAN L C,ANGEL ALEJANDRO JUAN PEREZ. Educational data mining and learning analytics: differences, similarities and time evolution[J]. Ruse Revista De Universidad Y Sociedad Del Conocimiento,2015,12(3):98-112.

[12] DURKSEN T L,CHU M W,AHMAD Z F,et al. Motivation in a MOOC:a probabilistic analysis of online learners' basic psychological needs[J]. Social Psychology of Education,2016,19(2):241-260.

[13] FITOUSSI J P,VELUPILLAI K. Technology for Mining the Big Data of MOOCs[J]. Research & Practice in Assessment,2014,9:29-37.

[14] MAC CALLUM K,JEFFREY L. Factors Impacting Teachers' Adoption of Mobile Learning[J]. Journal of Information Technology Education Research,2014,13(13):141-162.

[15] 樊超,宗利永. MOOC 在线学习行为的人类动力学分析[J]. 开放教育研究,2016,22(2):53-58.

[16] 宗阳,孙洪涛,张享国,等. MOOCs 学习行为与学习效果的逻辑回归分析[J]. 中国远程教育,2016,36(5):14-22.

[17] 肖建忠,陈小娟,贾秀险. 高等教育评估多元化研究[J]. 高教探索,2013(1):13-15.

[18] O'CONNOR M C,PAUNONEN S V. Big Five personality predictors of post-secondary academic performance[J]. Personality & Individual Differences,2007,43(5):971-990.

[19] POROPAT A E. A meta-analysis of the five-factor model of personality and academic performance[J]. Psychological Bulletin,2009,135(2):322-328.

[20] VEDEL A. The Big Five and tertiary academic performance:A systematic review and metaanal-ysis[J]. Personality & Individual Differences,2014,71(2):66-76.

[21] KONTOYIANNIS I,ALGOET P H,SUHOV Y M,et al. Nonparametric entropy estimation for stationary processes and random fields,with applications to English text[J]. IEEE Transactions on Information Theory,1998,44(3):1319-1327.

[22] CAO Y,GAO J,LIAN D,et al. Orderness Predicts Academic Performance: Behavioral Analysis on Campus Lifestyle[J]. eprint arXiv:1704.04013.

[23] TOKTAROVA V I,PANTUROVA A A. Learning and Teaching Style Models in Pedagogical Design of Electronic Educational Environment of the University[OL]. http://www.mcser.org/journal/index.php/mjss/article/view/6874.

[24] 倍智人才研究院. 大五人格心理学:The big five[M]. 北京:企业管理出版社,2015.

[25] PERRY T W. 16-Cattle Finishing Systems[OL]. http://doi.org/10.1016/B978-012552052-2150019-6.

[26] 王晨煜,管明辉,殷传涛,等. 基于 FelderS-ilverman 学习风格模型的网络学习风格研究[J]. 重庆理工大学学报,2017,31(2):102-109.

[27] FREUND Y, MASON L. The Alternating Decision Tree Learning Algorithm[C]// Machine Learning:Sixteenth International Conference. 1999:124-133.

[28] MOZINA M,DEMSAR J,KATTAN M,et al. Nomograms for Visualization of Bayesian Classifier[C]// European Conference on Principles of Data Mining & Knowledge Discovery. 2004:337-348.