

基于行为分析的在线课程成绩预测模型

任占广¹, 尚福华²

(1. 重庆文理学院 软件工程学院, 重庆 402160;

2. 东北石油大学 计算机与信息技术学院, 黑龙江 大庆 163318)

摘要:随着大学在线课程占全部课题比重的不断提高, 为了更加科学地分析在线学习行为和准确地预测在线课程成绩, 提出了一种基于行为分析的在线课程成绩预测模型。首先, 对学习行为及成绩预测策略进行了系统分析, 构建了在线平台数据处理、成绩预测算法设计、成绩预测及算法优化的在线成绩预测机制; 其次, 利用数据挖掘技术收集在线学习行为数据, 结合在线用户的操作特点对行为数据进行分析, 提取了与成绩密切相关的10种行为指标数据并存储到数据库中; 最后, 以“玩课网”平台的重庆文理学院“大学生计算机基础”课程后台数据库作为实验数据基础, 结合该课程实施特点, 分析了学生学习行为, 确定了学习行为指标等级, 提取和转换了学生学习行为数据, 并利用神经网络实现了在线课程成绩的预测。实验结果表明成绩预测的准确率较高。

关键词:行为分析; 数据处理; 神经网络; 成绩预测

中图分类号: TP39

文献标识码: A

文章编号: 1673-629X(2019)11-0139-05

doi: 10.3969/j.issn.1673-629X.2019.11.028

Online Course Grade Prediction Model Based on Behavior Analysis

REN Zhan-guang¹, SHANG Fu-hua²

(1. School of Software Engineering, Chongqing University of Arts and Sciences, Chongqing 402160, China;

2. School of Computer & Information Technology, Northeast Petroleum University, Daqing 163318, China)

Abstract: With the increasing proportion of online courses in all subjects, in order to analyze online learning behaviors more scientifically and predict online course grade more accurately, a behavior analysis-based online course grade prediction model is proposed. Firstly, the learning behavior and grade prediction strategy are systematically analyzed, and an online grade prediction mechanism is constructed, which includes data processing of online platform, design of grade prediction algorithm and optimization of grade prediction algorithm. Secondly, data mining technology is used to collect online learning behavior data, and according to the online user's operational characteristics, the behavior data is analyzed, and 10 kinds of behavior indicators are extracted and stored in the database. Finally, with the database of the course "Computer Basis for College" of Chongqing University of Arts and Sciences as the platform of "Wankewang", combined with the characteristics of course implementation, we analyze the learning behavior of students, determine the level of learning behavior indicators, extract and transform the data of students' learning behavior, and utilize the neural network to finish the prediction of online course grade. The experiment shows that the performance prediction is more accurate.

Key words: behavior analysis; data processing; neural network; grade prediction

0 引言

随着互联网技术的飞速发展, 网络课程已经成为继传统课堂之外最主要的学习途径, 因其不受时间和空间等限制, 得到了当代大学生的追捧^[1]。在网络课程的学习过程中将产生大量的用户行为数据, 而如何利用神经网络、机器学习、数据挖掘等技术, 根据在线

学习行为, 建立合理有效的成绩预测模型, 科学分析影响在线课程成绩的相关因素, 已成为大数据挖掘领域的重要研究方向^[2]。目前, 已经有部分学者对行为分析及成绩预测做了研究。例如, 文献[3]对网络商务课程用户的学习行为与反思性学习的影响等级进行研究; 文献[4-5]分别分析了在线学习行为与学习态度、

收稿日期: 2018-12-24

修回日期: 2019-04-25

网络出版时间: 2019-06-27

基金项目: 国家自然科学基金(61170132); 国家科技重大专项资助项目(2017ZX05019005-006); 重庆文理学院校级科研项目(Z2015RJ05)

作者简介: 任占广(1988-), 男, 助理实验师, 硕士研究生, 研究方向为移动计算、机器学习等; 尚福华, 教授, 博士(后), 研究方向为人工智能、软件工程等。

网络出版地址: <http://kns.cnki.net/kcms/detail/61.1450.TP.20190627.1111.064.html>

自我效能感和先验知识的关系。以上针对学习行为的研究,在对象和样本的选择上都有一定的局限性,同时,在线学习行为分析的主观干预度比较强,不具有一定的普遍性。文献[6]提出了基于 PCA-RBF 网络的学生写作成绩预测模型,虽然该模型提高了预测的准确率,但是收敛速度较慢。文献[7]提出了基于决策树 C4.5 算法的学生成绩预测模型,侧重分析了影响成绩的因素,但是准确率不高。文献[8]提出了基于频繁模式聚类课程关联分析和成绩预测模型,对课程分类及较差课程的预测的准确性较高,具有一定的参考价值。

综上所述,国内关于学习行为和成绩预测的研究还处在起步阶段,对在线学习行为数据的采集和挖掘还不够深入。因此,文中提出一种基于行为分析的在线课程成绩预测模型,分析并提取了在线学习行为数据,构建了成绩预测模型,完成了在线课程成绩预测,达到了预期的效果。

1 模型构建

成绩预测模型是在线课程学习平台的重要组成部分,是从数据采集、数据分析、数据处理到成绩预测的整个过程^[9-10]。数据采集是指利用数据库、网页抓取等技术获取用户在线学习过程中的所有行为数据,如鼠标的操作、课程课件或者视频的观看、键盘的录入等。由于采集的数据过于多样化、过于冗余,需要进一步对数据进行除杂、降噪等规范化、标准化处理,一方面压缩了数据总量,另一方面便于存入到关系型数据库中。数据标准化后,结合在线学习平台的特点、课程本身的安排等,提取与成绩密切相关的行为数据。最后,利用神经网络算法,以在线课程指标及在线课程真实成绩作为类标,完成预测算法的设计,并根据预测效果、在线课程方案等进一步的调整,使其更加适合在线学习平台。在线课程成绩预测模型如图 1 所示。

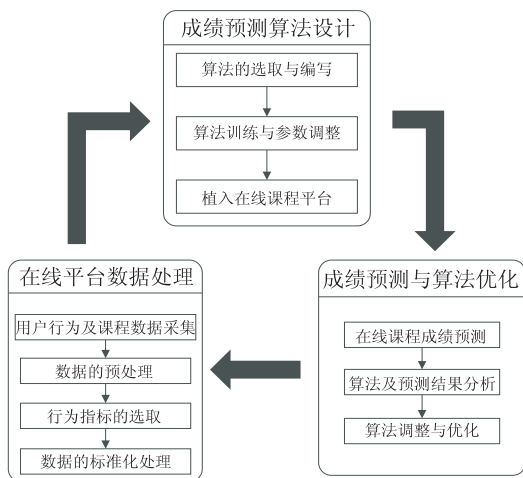


图 1 基于在线行为分析的成绩预测模型

2 在线学习行为分析

在线用户即网络用户,根据 CNNIC(中国互联网信息中心)的定义,网络用户是指“在最近 6 个月使用互联网的 6 岁以上的公民”,是网络使用者的总称^[11]。在线学习行为即在线用户利用网络学习平台进行学习的所有操作行为,例如提交作业、在线观看等。在线学习行为分析就是通过数据挖掘、大数据分析等手段对学习行为数据进行处理,主要包括三个阶段:行为数据准备、行为数据分析、行为数据结果输出,如图 2 所示。

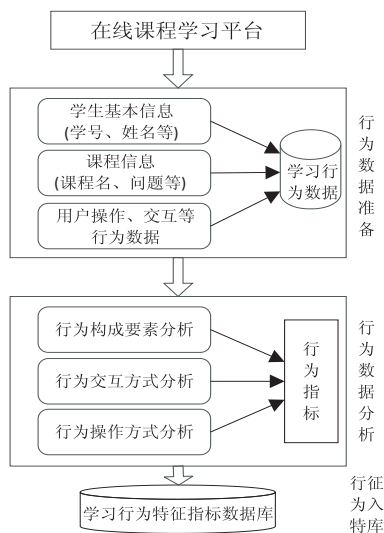


图 2 在线学习行为分析模型

在线学习行为分析主要包括三个方面:行为构成要素分析、行为交互方式分析以及行为操作方式分析^[12-13]。

(1)行为构成要素分析即对在线学习平台的行为主体和行为客体的分析^[14-16]。在线学习平台的主体是大学生,也就是研究的对象。行为客体是指与行为主体进行交互的客观对象,比如学习系统、网络环境等。文中的行为主体分析就是对大学生的认知结构、学习动机、学习习惯、学习规律的分析,行为客体分析主要是对在线学习资源类型、构成等的分析。而学习资源作为承载学习内容、知识信息的重要载体,它的设计与安排的合理性、创新性将直接影响到大学生在线学习的效果。

(2)行为交互方式分析即对行为主体与行为客体的交互和行为主体与网络平台的交互的分析^[14-16]。前者即为大学生与在线课程内容的交互,如浏览学习内容、检索学习内容、分享学习内容等。后者即为大学生与在线学习平台的交互,如作业的完成与提交、问题的提出、添加学习笔记等。文中主要分析大学生提出问题的次数、在线观看学习视频的时长、登陆在线学习平台的次数、作业提交次数以及提交的时间间隔等。

(3)行为操作方式分析即行为主体在网络平台上的所有行为轨迹的分析,也就是对在行为交互过程中

的所有操作行为的分析^[14-16]。行为操作数据是学习行为数据的最主要的组成部分,同时也是研究的重点和难点。在一定程度上,行为轨迹数据反映了在线学习用户的学习习惯、学习时间安排、学习兴趣等。文中主要分析与在线课程学习密切相关的键盘输入、鼠标操作等操作行为。

表 1 在线课程学习行为指标

行为指标	描述	编码
在线注册人数	--	P ₁
课程登陆次数	--	P ₂
课件观看时长	单位:分钟	L ₁
课件观看完成度	L ₁ /总时长	L ₂
课件重复观看次数	--	L ₃
结课后课件观看次数	--	L ₄
问题提问次数	--	L ₅
作业提交次数	--	E ₁
作业提交与发布时间差	作业提交时间-作业发布时间(单位:天)	E ₂
作业提交时间间隔	两次作业提交的时间差(单位:天)	E ₃

文中以在线学习平台“玩课网”的用户网络操作日志及在线课程数据库作为原始数据,首先对操作日志数据、课程数据进行分析,找到它们之间存在的相关性,并对其进行整合,形成学习行为特征数据,并存入数据库;其次,提取与成绩密切相关的全部学习行为指

通过以上三个方面的分析,结合在线学习平台的交互和操作方式、在线学习资源的构成与安排、在线学习用户本身的特点、任课老师的意见等,利用调查问卷、在线课程学习讨论会等多种形式最终确定了 10 种学习行为指标,并对这 10 种行为指标进行了编码,如表 1 所示。

标数据;再次,结合表 1 的在线课程学习行为指标,对行为指标数据进行相关度分析,并确定相应的学习行为指标;最后,将确定好的行为指标数据进行标准化处理,并将标准化数据存入到成绩预测支撑数据库中,如图 3 所示。

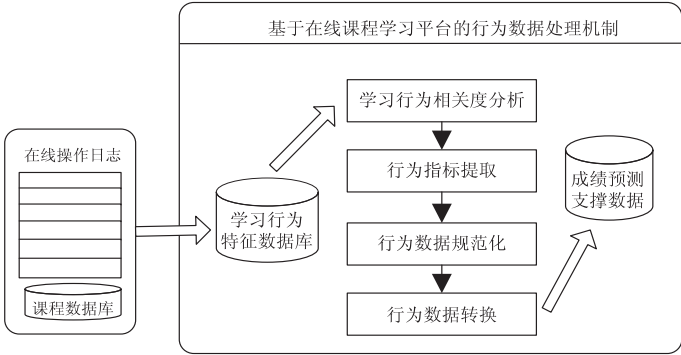


图 3 在线行为数据分析与处理机制

3 实验

3.1 实验数据分析与处理

实验选取“玩课网”平台的重庆文理学院“大学生计算机基础”课程后台数据库作为实验数据基础,该课程是针对大一理工类学生开设的公共必修课程,采用线上自主学习、线下实际操作的教学方式。由于是必修课程,学生只有完成平台信息注册,并通过任课教师的审核通过,才能进入所在的班级学习并通过最终的期末上机实操考试获得相应的学分。学习者的在线学习平台的注册行为是一种强制的、被动的行为,因此在本次实验中,不应将在线注册人数 P₁ 作为输入元素;当学期末结束课程后,在线学习班级自动解散,结

束课程以后学生无法继续观看相关学习资料,因此,结课后视频/课件观看次数 L₅ 也符合本次实验的行为指标。除去学习行为指标 P₁ 和 L₅,剩下的 8 个学习行为指标数据作为本次实验输入,结合该课程教学、考核方案等,将 8 个行为指标分为 A、B、C、D 四个等级,如表 2 所示。

按照表 2 的指标等级,收集了 2018-2019 第一学年的 1 000 名学生的 8 个行为指标数据(根据学习行为指标等级进行了分类处理,便于后续实验数据的录入和神经网络的训练)和真实成绩(真实成绩分为四个等级:优秀 A、一般 B、及格 C、不及格 D),由相应的任课老师提供,用 G 表示,如表 3 所示。

表 2 行为指标等级

指标	A 级	B 级	C 级	D 级
P ₂	>80	80-50	49-20	<20
L ₁	>500	500-400	399-300	<300
L ₂	>0.85	0.85-0.75	0.74-0.6	<0.6
L ₃	>15	15-10	9-5	<5
L ₅	>20	20-15	14-9	<9
E ₁	>10	9-7	6-3	<3
E ₂	<2	3-4	5-7	>7
E ₃	<7	8-10	11-14	>14

表 3 处理后学生行为数据及真实成绩

序号	P ₂	L ₁	L ₂	L ₃	L ₅	E ₁	E ₂	E ₃	G
1	A	A	B	A	A	A	A	A	A
2	A	B	B	C	B	B	B	A	B
3	B	B	C	C	C	B	B	B	B
4	C	C	B	C	C	D	D	D	C
5	A	A	A	A	A	A	B	B	A
6	B	A	C	C	C	C	C	C	C
7	A	A	B	B	A	B	B	B	B
8	C	C	B	D	D	C	C	C	B
9	B	B	A	A	A	A	A	B	A
10	D	C	C	C	B	C	C	A	C
...

3.2 算法实现

考虑到行为指标的输入与成绩预测结果的输出都是四类等级值,因此,本次实验采用前向神经网络完成前期的训练和后期的成绩预测。记该前向神经网络的神元输入个数为 p ,输出个数为 q ,隐藏个数为 s ,则前向神经网络结构如 4 所示。

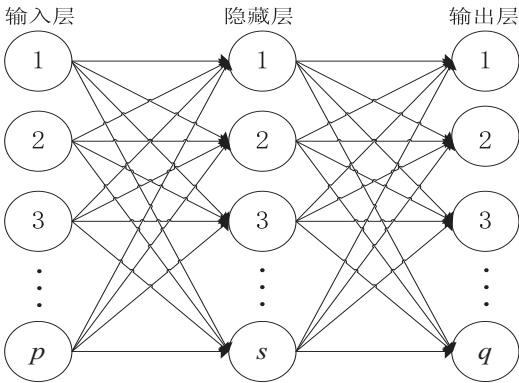


图 4 前向神经网络结构

因为影响成绩的行为指标共 8 个,而成绩结果有 4 个,所以该网络的输入神元个数 $p=8$,输出神元的个数 $q=4$,而隐藏神元个数 s 的确定是该网络的关键,具体步骤如下:

Step1:将 s 的值设置无穷大;

- Step2:利用样本集 R 训练网络,直到达到预期值;
- Step3:将能够识别的样本构成集合 R' ;
- Step4:输入 R' ,计算每个隐藏神经元的激活值 H_{ij} ;
- Step5:将 H 沿每一行重新排列,得到排序后的矩阵 H' ,记每行排序映射为 f_i ;
- Step6:对 H' 的每行分别假定一个切入点将该行分割成 2 块,计算分割后的信息增益矩阵 N ,选择 N 中的最大值 N_{\max} ,构建决策树的一个节点;
- Step7: $P' = P' - f_i$,若 P' 不为空,则转到 Step4 继续执行;
- Step8:搜索整棵决策树,寻找所有重要的节点。
- 通过计算共寻找到重要节点 6 个,所以隐藏神元个数 $s=6$ 。

3.3 成绩预测

实验在内存为 4 GB、主频为 3 GH、操作系统为 Win7 的台式电脑上进行,使用 Matlab 将处理过的标准数据集导入到前向神经网络中。学习行为数据集共选取了 1 000 名大一学生的行为数据,其中 500 名学生行为数据作为训练子集,200 的行为数据作为验证子集,300 的行为数据作为测试子集,测试子集作为最

终的成绩预测结果。

前向神经网络的收敛情况如图 5 所示,预测结果如表 4 所示。

表 4 预测与真实成绩对比

	A	B	C	D
预测结果	118	65	74	43
真实结果	118	48	91	43

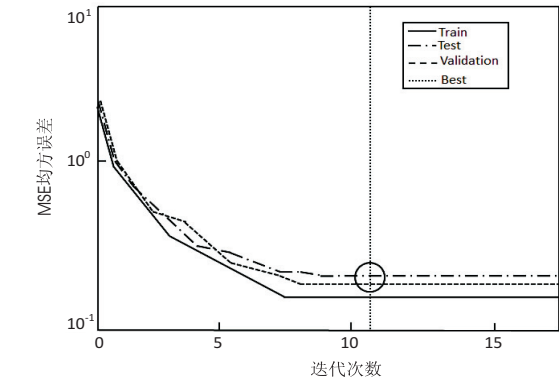


图 5 迭代次数误差均方图

如图 5 所示,当迭代到第 11 次时,验证子集的误差率达到最低,当迭代到第 17 次时,算法结束。如表 4 所示,优秀和不及格的学生全部预测正确,只有一般或者及格的 34 人预测发生错误,预测准确率高达 88.67%,说明该模型比较适合在线课程成绩的预测。

4 结束语

文中提出了一种基于行为分析的在线课程成绩预测模型,以“玩课网”平台在线数据为基础,结合重庆文理学院“大学生计算机基础”课程用户行为数据,分析并找到了与在线课程成绩密切相关的学习行为指标数据,实现了在线课程成绩的预测,为在线课程的实施、发展、改革等提供了很好的参考作用。虽然该模型整体预测的准确率较高,但是当学习行为指标数据越接近指标等级临界点时,预测的错误的概率就越高。因此,还需要获取更多的样本集合,对神经网络进行训练,调整隐藏神经元个数,从而降低误差率。

参考文献:

[1] 陈晋音,方航,林翔,等.基于在线学习行为分析的个性化学习推荐[J].计算机科学,2018,45(11A):422-426.

[2] 吴强,方睿,韩斌,等.基于决策树-LMBP神经网络的学生成绩分析及预测模型的研究[J].成都信息工程大学学报,2018,33(3):274-280.

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

[4] PRIOR D D,MAZANOV J,MEACHEAM D,et al. Digital literacy and self efficacy:flow-on effects for online learning behavior[J]. The Internet & Higher Education,2016,29:91-97.

[5] 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.

[6] 胡帅,顾艳,姜华.基于PCA-RBF网络的学生写作成绩预测模型[J].计算机与现代化,2016(1):69-72.

[7] 王黎黎,刘学军.决策树C4.5算法在成绩分析中的应用[J].河南工程学院学报:自然科学版,2014,26(4):69-73.

[8] 何楚,宋健,卓桐.基于频繁模式谱聚类的课程关联分类模型和学生成绩预测算法研究[J].计算机应用研究,2015,32(10):2930-2933.

[9] 刘志妮.基于决策树算法的学生成绩的预测分析[J].计算机应用与软件,2012,29(11):312-314.

[10] 王法玉,姜妍.基于自组织神经网络和模糊聚类的校园无线网用户学习兴趣度行为分析[J].计算机应用研究,2018,35(1):186-189.

[11] 罗海艳,杨勇,王珏,等.基于云计算的移动用户上网行为分析系统[J].控制工程,2018,25(2):218-223.

[12] 肖宇航.行为分析技术在用户管理系统中的应用研究[D].北京:北方工业大学,2018.

[13] 王子清.大学生学习行为分析研究与应用[D].苏州:苏州大学,2018.

[14] 王玮.在线学习行为分析及应用研究[D].武汉:华中师范大学,2018.

[15] 张大胤.基于互联网环境下学习行为的数据挖掘研究[D].桂林:广西师范大学,2018.

[16] 周剑,薛景,韩崇,等.基于BP神经网络的学生成绩预测方法研究[J].计算机时代,2018(12):71-74.