

大学英语诊断性练习系统中诊断性评价模型研究^①

吴涛, 张晖, 吴敏

(中国科学技术大学 现代教育技术中心, 合肥 230026)

通讯作者: 吴涛, E-mail: jacksonwu1992@163.com

摘要: 在线学习系统中没有教师对学习者的学习状态进行评价, 大多需要学习者自主设定学习策略、调整学习步伐, 这导致了学习者缺乏个性化的指导, 使得部分学习者的学习效率不高. 为了解决这个问题, 本文在大学英语诊断性练习系统中提出了诊断性评价模型. 该模型使用大量实际数据, 从学习状态、题型关联分析、知识点关联分析和四级成绩预测这四个角度进行数据分析, 并使用 S-P 表分析法、数据挖掘和机器学习等分别建模, 最终将这四个模型结合得到了诊断性评价模型, 并在大学英语诊断性练习系统中进行实现. 实验结果表明, 诊断性评价模型可以有效辅助学习者进行练习, 提高四级成绩.

关键词: 诊断性评价模型; S-P 表; 数据挖掘; 关联分析; 机器学习

引用格式: 吴涛, 张晖, 吴敏. 大学英语诊断性练习系统中诊断性评价模型研究. 计算机系统应用, 2018, 27(11): 90-95. <http://www.c-s-a.org.cn/1003-3254/6644.html>

Diagnostic Evaluation Model of College English Diagnostic Test System

WU Tao, ZHANG Hui, WU Min

(Center of Modern Educational Technology, University of Science and Technology of China, Hefei 230026, China)

Abstract: There are no teachers to evaluate the learning state of the learners in the online learning system, so most of the learners need to set their own learning strategies and adjust their learning steps, which lead to the lack of individualized guidance for the learners and the poor efficiency of the learners. In order to solve this problem, a diagnostic evaluation model was proposed in the College English Diagnostic Test System. The model analyzed the data collected in the database from the four aspects of learners' learning condition, question types association analysis, knowledge point association analysis, and CET-4 grade prediction, and it used Student-Problem Chart analysis method, data mining, and machine learning to build models respectively. Finally, the diagnostic evaluation model was obtained by merging these four sub models and it was implanted to improve the College English Diagnostic Test System. The experimental results showed that the diagnostic evaluation can effectively help learners to practice and improve their grades of CET-4.

Key words: diagnostic evaluation; Student-Problem chart; data mining; correlation analysis; machine learning

在传统的教学活动中, 教师会根据练习测试结果和学习者在课堂内外的具体表现, 对学习者的学习过程和、纠正学习中的偏差^[1]. 在网络在线学习系统中, 学习者在进行学习的同时, 需要自行设定学习目

标、选择学习策略、调整学习步调^[2].

为了帮助越来越多的大学生备考大学英语四级考试, 许多高校和企业推出了很多在线英语练习系统. 但试用后不难发现, 这些系统的使用流程仍然局限于做题和查看解答的传统模式, 提供给学习者的反馈信息

^① 收稿时间: 2018-04-08; 修改时间: 2018-04-27, 2018-05-11; 采用时间: 2018-05-15; csa 在线出版时间: 2018-10-24

也仅仅停留在得分情况、试题解答和题目分析上。这样的反馈信息缺乏针对性和指向性,既不能体现出人工智能时代大数据的优势,更无法为学习者提供客观的诊断性评价和有效的学习建议,这必然会对学习效果造成一定的偏差,学习的效率也不是很高。

在线学习平台应利用实际使用中收集的大量数据来进行数据分析和理解,揭示数据内部蕴藏的信息,才能弥补个人经验的不足和系统的分析缺陷^[3]。本文利用系统使用过程中收集的数据进行数据分析和数据挖掘,从学习状态、题型关联分析、知识点关联分析和四级成绩预测这四个角度分别建模,最终合并得到了诊断性评价模型。

本文组织结构如下,第1节对大学英语诊断性练习系统进行了简介,第2节介绍诊断性评价模型的总体设计,第3节介绍了学习状态评价、题型关联分析、知识点关联分析和四级成绩预测这四个模型的构建和合并后的诊断性评价模型以及诊断性评价模型在系统中的实现,第4节给出诊断性评价模型在某高校的实验结果并进行相关分析,最后进行总结。

1 大学英语诊断性练习系统

大学英语诊断性练习系统是一个以大学英语四级考试为背景、向学习者提供个性化的指导建议、帮助学习者提升英语水平的在线学习系统。在对英语考试知识点分类和对题目知识点赋值的基础上,该系统向学习者提供多种练习模式,能够根据学习者以往的练习情况指出其弱项知识点,使得学习者既可以充分享受网络教育带来的便捷,又可以根据自身特点接受个性化指导,针对性地进行学习,从而高效地提高英语四级成绩。

系统模块按照用户角色分为学生模块、教师模块和管理员模块三个部分,其中学生模块包括需完成作业、自主练习模式以及直观化的统计结果等部分;教师模块包括自主出题、布置作业、成绩管理和学生管理等功能;管理员模块则包含教师管理和班级管理等功能。

从上述描述可以看出,大学英语诊断性练习系统中对于学生模块的设计仍是传统的做题和查看分数统计的功能,提供给学习者的反馈信息只有统计和试题解答等,这导致了学习者缺乏详细的后期评估、准确的学习指导和精准的试题推荐来帮助他们快速提高成

绩并且通过四级考试,所以利用数据挖掘和机器学习等技术对诊断性评价模型进行研究和实现势在必行。

2 诊断性评价模型的设计

对学习者的教育评价分为形成性评价和总结性评价。形成性评价是通过诊断学习者的学习情况、学习过程中存在的问题,为学习者正在进行的学习活动提供反馈信息,以提高正在进行的学习活动的质量;总结性评价则是对学习活动效果做出价值判断。

本文所探讨的诊断性评价,是帮助学习者在其自主学习的过程中动态诊断其学习情况和学习障碍,并据此优化学习过程的一种反馈信息,因此诊断性评价更多地属于形成性评价。

对于自主学习者来说,诊断性评价可以帮助他们调整学习策略,改善学习方法。因此,诊断性评价作为完善学习者学习过程的评价,要对学习者在学习过程中出现的问题和学习障碍做出反馈,从而辅助学习者改善学习策略和调整学习步调。

为了给自主学习的学习者提供有效的帮助,诊断性评价应当具备以下三个方面的作用:

(1) 诊断

诊断性评价,顾名思义需要有诊断的功能。评价不仅要对学习者的能力、状态做出鉴定,帮助学习者全面了解自身的学习情况,更应当明确找到学习者在学习过程中所面临的问题和存在的缺点。学习者只有在客观了解自身学习状态、明确学习中的漏洞之后,才能高效地提高学习效率。

(2) 调节

诊断性评价作为一种学习活动的反馈信息,应当注重学习者的个性化调节。网络教育应该因材施教,根据不同学习者的不同特性提供个性化的诊断建议,这样才有利于学习者更加有效地自主调控学习策略,改进学习者的个体学习方法,更适应学习者自身的发展。

(3) 预警

在传统教学中,教师通过作业和练习测试查看学习者的学习状态,但由于传统班级人数众多等原因,教师往往无法关注所有同学的学习状态,从而导致部分学习者的成绩下滑。诊断性评价模型通过查看学习者练习数据,可以动态诊断其近期学习状态和学习障碍,从而做到预警的功能,时刻督促学习者查漏补缺。

本文提出的诊断性评价模型是一种旨在以系统自

动评价的形式诊断学习者的学习情况和学习障碍、指导学习者的学习策略、激发学习者的学习兴趣的动态评价模型。据此,诊断性评价的框架如图1所示。

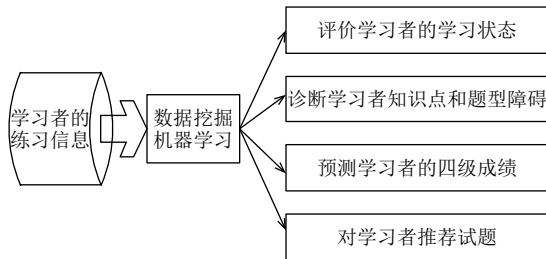


图1 诊断性评价的框架

具体的操作流程如下:

(1) 统计学习者的练习信息, 并进行分析计算

从系统现有的数据库中提取学习者的测试、练习信息, 经过处理后使用诊断性评价模型进行分析计算。

(2) 分析评价学习者的整体学习状态

从学习者的知识点得分、题型得分等方面出发, 以学习者的角度对练习测试信息进行分析, 从得分率和稳定程度两个方面来评定学习者的学习状态。通过对学习者学习状态的公正判断, 系统将对状态不稳定、学习进度较落后的学习者进行预警, 从而激发学习者的学习动力。

(3) 诊断学习者的知识点和题型障碍

使用知识点关联规则表和题型关联规则表分别对学习者的知识点和题型进行关联分析, 推断学习者的知识点和题型障碍, 帮助学习者找到自身的缺陷, 快速有效地提高成绩。

(4) 预测学习者的四级成绩

对学习者的测试、练习信息进行特征提取, 使用随机森林和多元线性回归两个模型并加以融合构建得到四级成绩预测模型, 让学习者了解自身的英语水平, 督促学习者练习试题。

3 诊断性评价模型的构建和实现

根据第二节的设计, 需要对学习状态评价模型、知识点关联分析、题型关联分析和四级成绩预测模型分别进行研究和设计, 最后整合成诊断性评价模型并在大学英语诊断性练习系统中进行了实现。

3.1 学习状态评价模型

学习者学习状态的评价, 应从学习者的能力水平和稳定程度两个维度出发, 对学习状态进行评定。

Student-Problem Chart (S-P 表) 是藤田 (Takahiro Sato) 教授根据统计学提出的一种分析方法, 它将学生和问题相对应并以视觉化的图表进行统计分析, 对学习者给予全面的评价^[4]。在本文中 will 使用 S-P 表分析法来构建学习者学习状态评价算法。其中使用注意系数来评价学习者的学习稳定程度, 知识点题型得分率来评价学习者的掌握情况。

学习状态评价算法主要解决学习者知识点题型掌握情况和注意系数, 并且判定学习者的学习类型, 为学习者提供针对性个性化的指导。具体算法步骤如下:

(1) 读取学习者的知识点和题型得分率

从数据库中读取学习者的知识点和题型得分率, 并以矩阵的形式存储。设有 N 位学习者, 每位学习者的知识点和题型得分率为 M 项, 矩阵中的元素 x_{ij} 表示第 i 名学习者的得分率 j 大小, 矩阵的公式如下:

$$[x_{ij}] = \begin{pmatrix} x_{11} & \cdots & x_{1n} \\ \vdots & \ddots & \vdots \\ x_{m1} & \cdots & x_{mn} \end{pmatrix}$$

(2) 处理连续数据

得分率是区间在 $[0, 1]$ 之间的小数, 而学习状态评价模型只能对二元离散数据进行操作, 因此需要对学习者的得分率进行二元离散处理。

(3) S-P 表行列计算

计算学习者 i 的得分和 x_i , 知识点和题型的学习者得分和 x_j 。

(4) 计算学习者的注意系数

设 y_{ij} 是第 i 名学习者的得分率 j 的数值, y_i 是学习者 i 的总得分, y_j 是知识点题型 j 的答对数, μ 是试题的平均答对数, 则学习者注意系数 CS_i 计算公式如下:

$$CS_i = 1 - \frac{\sum_{j=1}^n (y_{ij})(y_j) - (y_i)(\mu)}{\sum_{j=1}^n y_j - (y_i)(\mu)}$$

(5) 计算学习者的知识点和题型掌握情况

学习者的知识点题型掌握情况的计算公式如下:

$$\text{掌握情况} = \frac{\text{知识点题型得分率大于等于均值的个数}}{\text{知识点题型总数}}$$

(6) 输出结果

输出学习者的知识点题型得分率、注意系数。

以上就是学习状态评价模型的操作流程。

3.2 题型关联分析和知识点关联分析

知识点和知识点之间存在一定的关联, 即某种或

某几种知识点的得分率高可能会和其他知识点的得分率高有所关联, 所以需要对知识点的关联分析进行探究. 相应的, 对于题型的研究也是如此.

本节通过在系统使用中收集了大量的原始数据, 经过数据处理后, 分别对题型和知识点进行关联分析, 最终得出了比较完善和可靠的题型关联规则表和知识点关联规则表.

其中, 关联分析的具体步骤如下:

(1) 导入数据. 读取所有用户的知识点得分率和题型得分率

(2) 清理数据. 将数据处理成挖掘算法所需要的格式.

(3) 数据挖掘. 使用 Apriori 算法分别对题型和知识点进行关联分析.

(4) 调整参数. 根据挖掘结果调整最小置信度和最小支持度, 得到相对合适的关联规则.

(5) 输出结果. 输出知识点关联规则表和题型关联规则表.

综上所述就是产生关联规则表的操作流程.

3.3 大学英语四级成绩预测模型

当前系统中没有提供大学英语四级分数预测的功能, 而很多刚进入大学的学习者对于四级题型知识点等并不熟悉, 也不能完全预估自身实力, 所以对于能否通过四级并无把握. 本文收集了在系统前期使用过程中的数据和用户留下的四级考试成绩, 经过数据处理、特征选择后, 使用随机森林模型和多元线性回归模型分别进行训练, 通过投票法进行模型融合, 最终得到了的大学英语四级分数预测模型.

随机森林模型和多元线性回归模型都有着生成简单的优点, 所以在本文中使用了这两种模型来构建四级成绩预测模型. 随机森林模型和多元线性回归模型都使用 Python 编程进行自动机器学习, 并输出回归结果. 四级成绩预测模型的具体操作步骤如下:

(1) 导入数据. 使用 pandas 包导入数据集, 并将数据集依据 4:1 的比例划分成训练集和验证集.

(2) 清理数据. 对于缺失数据, 使用其平均值来代替.

(3) 构建交互变量. 对特征集中任意两个不同特征 f_1 、 f_2 , 依次生成 f_1+f_2 、 f_1-f_2 、 f_1*f_2 、 f_1/f_2 这四种特征, 并加入到特征集中.

(4) 训练模型. 使用 sklearn 包来导入多元线性回归模型 (LinearRegression) 和随机森林模型 (Random-

ForestRegressor), 将这两个模型进行训练并使用投票法融合, 其中对随机森林设定训练树为 500.

(5) 输出结果. 根据步骤 (4) 得到的训练模型输出测试集的结果, 并输出多个评价指标.

其中步骤 (4) 中的投票法是将多个模型的输出结果进行线性加权, 而在本文中由于只有两个模型, 故将这两个模型的权值设置为相同.

在本文中使用了均方根误差 (RMSE) 和均方根对数误差 (RMSLE) 这两个评价标准进行验证. 它们的计算公式如下:

$$RMSE = \sqrt{\frac{\sum_{i=1}^n (X_{obs,i} - X_{mdl,i})^2}{n}}$$

$$RMSLE = \sqrt{\frac{1}{n} \sum_{i=1}^n (\log(X_{obs,i} + 1) - \log(X_{mdl,i} + 1))^2}$$

其中, $X_{obs,i}$ 为预测值, $X_{mdl,i}$ 为真实值.

使用验证集对四级成绩预测模型进行验证, $RMSE$ 和 $RMSLE$ 在随机森林、多元线性回归和融合后的模型的数据如表 1 所示.

表 1 各模型的 $RMSE$ 和 $RMSLE$

模型	$RMSE$	$RMSLE$
随机森林	24.325	0.159
多元线性回归	25.624	0.145
融合后的模型	20.541	0.123

从表 1 可以看出, 随机森林模型的 $RMSE$ 在 24.325、 $RMSLE$ 为 0.159, 多元线性回归模型的 $RMSE$ 为 25.624、 $RMSLE$ 为 0.145, 而将这两模型经过投票法融合后的模型的 $RMSE$ 为 20.541、 $RMSLE$ 为 0.123, 由此可以看出融合后的模型在 $RMSE$ 和 $RMSLE$ 两方面都有所提高, 预测模型的准确性也比较高.

3.4 诊断性评价模型的构建

前几节介绍了四个子模型的构建, 本节将这四个子模型整合, 最终构成了完整的诊断性评价模型. 具体的步骤如下:

(1) 提取数据. 从数据库中提取用户数据.

(2) 处理数据. 清理用户数据, 并将其变换成模型需要的格式.

(3) 评价用户的学习情况和稳定程度. 通过学习状态评价模型计算得到知识点题型掌握情况和注意系数, 判定得到用户的学习情况和稳定程度.

(4) 诊断用户知识点掌握情况. 使用知识点关联规则表得出用户潜在强弱项知识点.

(5) 诊断用户题型掌握情况. 使用题型关联规则表得出用户的潜在强弱项题型.

(6) 预测用户四级分数. 使用大学英语四级分数预测模型预估用户的四级分数.

(7) 输出用户数据.

综上所述就是整体的诊断性评价模型的操作步骤, 为了更好的展现, 诊断性评价模型的流程图如图 2 所示.

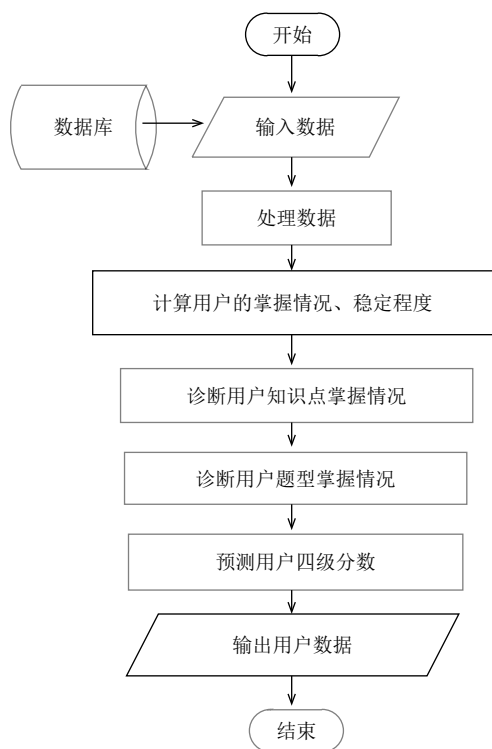


图 2 诊断性评价模型的完整流程图

3.5 诊断性评价模型的实现

诊断性评价模型是大学英语诊断性练习系统的一部分, 所以诊断性评价模型实现所用的技术方案应与大学英语诊断性练习系统的总体技术方案保持一致.

大学英语诊断性练习系统是一个 Web 形式的在线学习系统, 它主要是基于 B/S (Browser/Server) 模式和 ASP.NET MVC 框架. 该系统的客户端运行在 Web 浏览器上, 使用网络通信与服务器端交互. 该系统使用 SQL Server 数据库来存储数据信息.

为了方便调用诊断性评价模型中学习状态评价模型和四级成绩预测模型的计算结果, 这两个模型将采

用 Python 实现并将计算结果存入数据库中供用户查询. 因为 Drools 规则引擎具有方便调整、易于管理的特点, 所以诊断性评价模型中的知识点关联规则和题型关联规则将使用 Drools 规则引擎编写. 具体的诊断性模块调用框架图如图 3 所示.

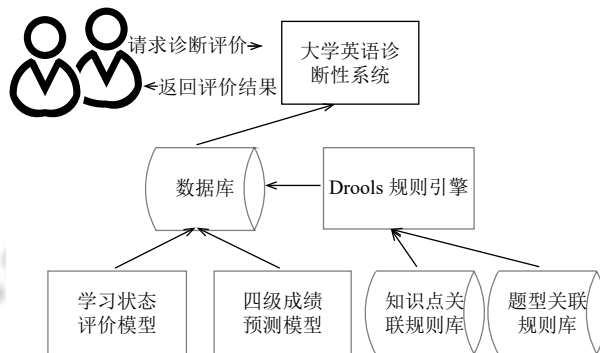


图 3 诊断性模块调用框架图

4 模型验证和分析

为了对诊断性评价模型进行验证, 笔者在使用该系统的高校选择了两个成绩相当并且待考四级的大二班级, 一个作为实验班, 一个作为对照班, 共计 120 人. 其中对照班按照正常的教学安排进行学习, 而实验班除了正常的教学安排还将使用诊断性系统完成 6 套试题.

实验班同学在经过一学期的使用后, 已经非常熟悉系统的各种评价功能, 所以请他们对系统及诊断性评价模型进行满意度点评, 点评数据如表 2 所示.

表 2 诊断性评价模型满意度评分表

评分项目	1	2	3	4	5	均分
诊断性评价满意度	0	3	24	27	6	3.6
学习状态评价满意度	0	2	24	30	4	3.6
知识点诊断满意度	0	4	22	26	8	3.63
题型诊断满意度	0	3	24	27	6	3.6

表 2 中的满意度评分是指用户对此项的满意度, 对分数的定义为: 1 分为非常不满意, 2 分为不满意, 3 分为一般, 4 分为满意, 5 分为非常满意.

表 2 中统计了每个评分的人数. 从表 2 中数据可以看出各项满意度平均分均在 3.6 分以上, 可以看出用户评分在一般和满意之间, 并偏向满意, 这从一定程度上证明了诊断性评价模型的诊断效果.

在本次四级考试成绩公布后, 笔者分别统计了实

验班和对照班的四级平均成绩和四级通过人数, 具体的情况如表3所示.

表3 四级平均成绩和四级通过人数

班级	四级平均成绩	四级通过人数
实验班	379分	10人
对照班	362分	7人

根据表3的统计结果, 我们可以看出实验班此次的四级平均成绩为379分并且有10人通过了此次四级考试, 对照班此次的四级平均成绩为362分并且有7人通过了此次四级考试. 实验班的平均成绩比对照班高出了15分, 四级通过人数也多了3人.

由此可以看出, 实验班和对照班的初始成绩虽然相当, 但实验班在使用诊断性练习系统后四级成绩有了一定的提升, 通过人数也有了增加.

此外, 通过数据库收集的实验班同学的练习数据使用大学英语四级成绩预测模型后得到的预测值的RMSE为20.387, RMSLE为0.126, 也与之前的数据非常相近.

综上所述, 诊断性评价模型在该校使用情况良好, 学生反馈较好, 对于四级成绩的提升也较为明显.

5 总结

本文的主要工作是诊断性评价模型的研究和构建. 首先, 作者介绍了大学英语诊断性练习系统; 其次, 详细探讨了诊断性评价模型的设计细节和设计意义; 接着, 使用S-P表分析法分析设计了学习者学习状态评价模型; 然后, 利用数据挖掘中的Apriori算法对处理

过的数据进行了题型关联分析和知识点关联分析, 得到了比较准确的题型关联分析表和知识点关联分析表; 并且, 利用随机森林模型和多元线性回归模型对处理过的数据设计了大学英语四级分数预测模型, 并使用RMSE和RMSLE两个评价标准验证了该模型; 最后, 对诊断性评价模型进行了总结和用户验证, 从而证明了诊断性评价模型的准确性和可靠性.

本文设计的诊断性评价模型会根据用户的练习情况动态地进行诊断评价、调节和预警, 并且随着用户使用系统进行更多练习测试时, 诊断性评价的准确性也将会提高.

参考文献

- 1 郭秀丽, 刘景慧. 论中学英语课堂现代教育评价体系的建立. 教育教学论坛, 2012, (21): 120-121, 267. [doi: 10.3969/j.issn.1674-9324.2012.21.080]
- 2 Magen-Nagar N, Cohen L. Learning strategies as a mediator for motivation and a sense of achievement among students who study in MOOCs. Education and Information Technologies, 2017, 22(3): 1271-1290. [doi: 10.1007/s10639-016-9492-y]
- 3 Hsieh NT, Huang YJ, Chuang CP. The study of design and performance analysis of adaptive learning and teaching system for junior vocational education students in Taiwan: A cognitive-load perspective. Proceedings of the 3rd International Conference on Education Technology and Training (Volume 1). Wuhan, China. 2010. 6.
- 4 陈洁. S-P表评价法的理论与实践[硕士学位论文]. 武汉: 华中师范大学, 2007.