

基于网络教学在线学习平台的大数据分析研究

杨子琪 杨 杉

四川大学锦城学院计算机与软件学院 四川 成都 611731

【摘要】以成都市某高校的网络教学平台为研究对象。利用 spss 分析工具对任务点完成数,视频任务点完成数,作业完成数,考试完成数,章节学习次数,学生观看视频总时长等六个重要指标进行了决策树分析和 k 均值分析研究。分析结果能直观看出线上线下不同时期学生的学习情况,从而依靠结论更有目的性,更有针对性的进行管理和提高。

【关键词】网络教学; 大数据分析

1 引言

随着网络教育的形式不断流行化,各种网络教育平台也渐渐发展起来。新的网络技术也使在线教育形式也越来越精彩,与此同时也成为目前以至于未来的重要组成部分^[1]。尤其是2019年年末开始,网络教学成为全国各地从幼儿园到大学疫情期间主要的教学方式^[2],在这段自己内展现的作用是非常大的,同时也累积了大量的学生在线学习数据,在学生回到校园上课,对于网络教学的需求降低,在这段时间也产生了与之前的不同的数据,两种不同时期的数据为研究学生的在线学习行为提供了丰富的素材^[3]。

通过对网络教学平台上导出的线上线下不同时期学生的信息数据进行数据分析画像^[6],通过分析大量的数据能够直观地看出线上线下不同时期学生的学习情况,从而依靠结论更有目的性,更有针对性的进行管理和提高^[4]。

2 研究思路

以成都市某高校的网络教学平台为研究对象,对该平台于2019.5-2020.5的学生在线学习数据进行数据分析。利用 spss 分析工具对任务点完成数,视频任务点完成数,作业完成数,考试完成数,章节学习次数,学生观看视频总时长等重要指标进行了决策树分析和 k 均值分析。

3 数据说明

3.1 数据来源

数据来源于成都市某高校在线学习平台,覆盖了2019.5-2020.5各个学院的学生在线学习数据,数据表中包含了学生姓名、学生账号、院系、专业、入学年级、任务点完成数,视频任务点完成数,作业完成数,考试完成数,章节学习次数,学生观看视频总时长等52个字段共100000条数据,51M。

3.2 数据清洗

数据表中,有很多数据是对于分析没有意义的,比如回帖总数,发帖总数等,因此对数据表中的数据进行了数据预处理^[5],筛选出了研究需要的数据,删除其他数据。

4 线上线下不同时期网络教学平台重要指标研究

4.1 对2018级同学线上线下不同时期重要指标做决策树分析

4.1.1 数据预处理

提取出线上线下不同时期2018级的同学数据,放至两个新表中。添加新列:视频任务点完成数等级、任务点完成数等级、作业完成数等级、考试完成数等级、章节学习次数等级。通过各个属性的四分位数进行等级划分为:差(0)、中(1)、优(2)。

4.1.2 对2018级学生线上教学时期重要指标分析

1. 准确率

其中“章节学习次数等级=2”准确率高于75%,为:76.58%。准确率高于75%的规则集分析



图1 规则集

规则1用于2.0:具有以下特征的人章节学习等级为2:任务点完成数为2。说明,任务点完成的优秀的人大部分章节学习也完成的好。规则2用于2.0:具有以下特征的人章节学习等级为2:视频任务点完成数为1、任务点完成数为1。说明,能够不错的完成视频任务点与任务点的人大部分章节学习也完成的好。规则3用于2.0:具有以下特征的人章节学习等级为2:视频任务点完成数为0、作业完成数等级为2,考试完成数为1。说明,视频任务点完成较差、作业完成较好并且考试完成数中等的同学,大多数章节学习等级较好。规则4用于2.0:具有以下特征的人章节学习等级为2:视频任务点完成数为2。说明,能较好的完成视频任务点的同学大多数都能较好的完成章节学习。规则5用于2.0:具有以下特征的人章节学习等级为2:作业完成数等级为2,考试完成数为0。说明,作业完成好并且考试完成成绩较差的同学大多数都能较好的完成章节学习。规则6用于2.0:具有以下特征的人章节学习等级为2 作业完成数等级为2,视频任务点完成数为1。说明,作业完成好并且视频任务点完成一般的同学大多数都能较好的完成章节学习。规则7用于2.0:具有以下特征

的人章节学习等级为 2：视频任务点完成数为 1，考试完成数为 2。说明，视频任务点完成一般并且考试完成较好的同学大多数都能较好的完成章节学习。小结：视频任务点完成数为优的同学章节完学习完成的都较好；能够中等完成两样学习标准的同学章节学习都完成的较好；完成的多项学习指标中，有完成优秀的指标的同学章节学习都完成得

预测变量重要性

目标：章节学习次数。

重要性等级由强到弱：视频任务点完成数，作业完成数，任务点完成数，考试完成数。

4.1.3 对 2018 级学生线下教学时期重要指标分析

1. 准确率

其中“章节学习次数等级=2”准确率高于 85%，为：89.43%。准确率高于 85% 的规则集分析：



图 2 规则集

规则 1 用于 2.0：具有以下特征的人章节学习等级为 2：任务点完成数为 2、作业完成数为 2。说明，任务点完成优秀并且作业完成优秀的同学大多数都能较好的完成章节学习。规则 2 用于 2.0：具有以下特征的

人章节学习等级为 2：视频任务点完成数为 2、作业完成数等级为 2。说明，视频任务点完成优秀并且作业完成数优秀的同学大多数都能较好的完成章节学习。规则 3 用于 2.0：具有以下特征的人章节学习等级为 2：任务点完成数为 2、考试完成数等级为 2。说明，视频任务点完成优秀并且考试完成数优秀的同学大多数都能较好的完成章节学习。小结：在所有的学习指标中，有两项指标完成优秀的同学大多数都能够较好的完成章节学习。

2. 预测变量重要性

目标：章节学习次数。

重要性等级由高到低：视频任务点完成数，任务点完成数，作业完成数，考试完成数。

4.1.4 线上线下不同时期对比分析总结

分析得出：无论是线上还是线下教学，对于章节学习次数影响最大的是视频任务点完成数，影响最小的是考试完成数。并且它们都存在一个同样的规律：所有学习指标中，如果出现一个或多个学习指标等级为优的话，那么这位同学的章节学习等级大概率为优秀。

4.2 对全校学生线上线下不同时期重要指标做 k 均值分析

4.2.1 K 均值建模处理

将除大四以外数据表导入 Modeler 中；添加类型节点，并选择重要指标作为输入角色，其他字段角色设置无；从“建模”选项中添加 K-Means 模型，将聚类数设置为 4；通过“表”和“均值”输出结果并分析，其中在“均值”选项中将分组字段设置为聚类结果，将聚类的六个重要指标设置为测试字段。

4.2.2 对全校学生线上教学时期重要指标分析

1. 聚类结果与分析



图 3 线上教学时期聚类结果人数占总数的比例及预测变量的重要性

结果显示，凝聚和分离的轮廓测量大于 0.5，说明聚类结果比较好。通过右图预测变量的重要性可以看出，除“作业完成数”外，其他字段的重要性都为 1，说明这五个变量对这次聚类结果起较为重要的作用。通过左

图可以看出四个聚类结果人数占总数的比例，其中可以看出最后一个聚类结果只有一个样本且其视频任务点学生观看时长和章节学习次数异常高于其他三类，所以通过导出数据，定位到该同学并对他进行成绩分析。

表 1 学习记录

视频观看时长（分钟）	章节测验完成数	章节测验平均分	作业平均分	考试平均分
11311	10	81.6	90	91.5

通过学习记录可以看出，这位同学在章节测验、作业、考试平均分中表现为中上水平，可以排除极端努力学习的可能，所以可能

导致这种现象的原因可能是，该账号学生可能被老师选做线上测试账号。

2. 均值结果与分析

表 2 线上教学时期均值结果

单元格内容：平均值					
字段	聚类-1*	聚类-2*	聚类-3*	聚类-4*	重要性
作业完成数	4.34	1	3.925	3.215	1
考试完成数	1.088	1	10.834	0.953	1
视频观看时长(分钟)	118.282	11311	221.667	777.408	1
章节学习次数	102.361	3965	187.414	175.027	1
视频任务点完成数	11.951	50	32.639	55.337	1
任务点完成数	16.809	66	64.755	102.431	1

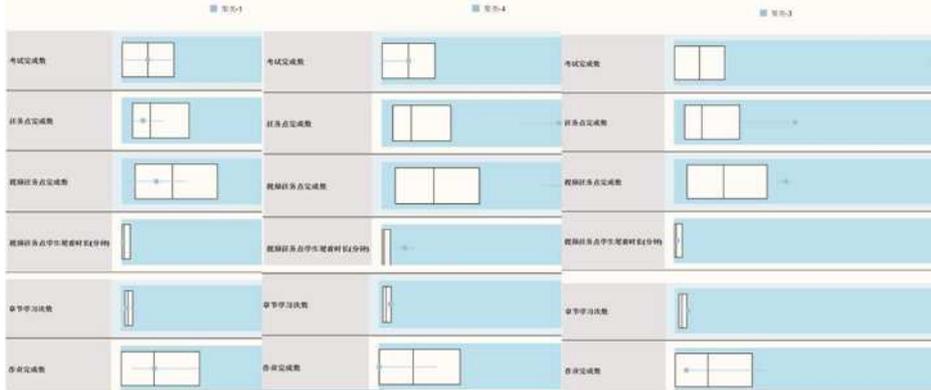


图 4 线上教学时期重要指标箱形图

通过均值分析、箱形图得出（除第二类学生）：第一类学生虽然在“作业完成数”中高于其他类学生，但差值非常小，且在其他指标中表现较为不好，所以可以将第一类学生划分成“学习任务较为轻松、学习氛围较为紧张，且努力完成老师布置的任务”的学生，这类学生占全校的 80.7%。第三类学生在“考试完成数”上表现突出，可能说明该类学生的考试数量较多，学习任务较为紧张；同时在“章节学习次数”上表现也比较突出，所以可以将第三类学生划分成“学习任务相对较紧、学习氛围紧张，且学习比较努力”的学生，这类学生占全校的 1.7%。第四类学生在“视频任务点学生观看时长”、“视频任务点完成数”和“任务点完成数”中表现比较

突出，可以说明在各类任务点上完成数比较多，所以可以将第四类学生划分成“学习任务紧、学习氛围较为轻松，且学习努力”的学生，这类学生占全校的 17.6%。

总结如下表：

表 3 线上教学学生的情况分布占比图

学生分类	学习任务	学习氛围	学习努力程度	占比
第一类	1	2	1	80.70%
第三类	2	3	3	1.70%
第四类	3	1	2	17.60%

4. 2. 3 对全校学生线下教学时期重要指标分析

1. 聚类结果与分析

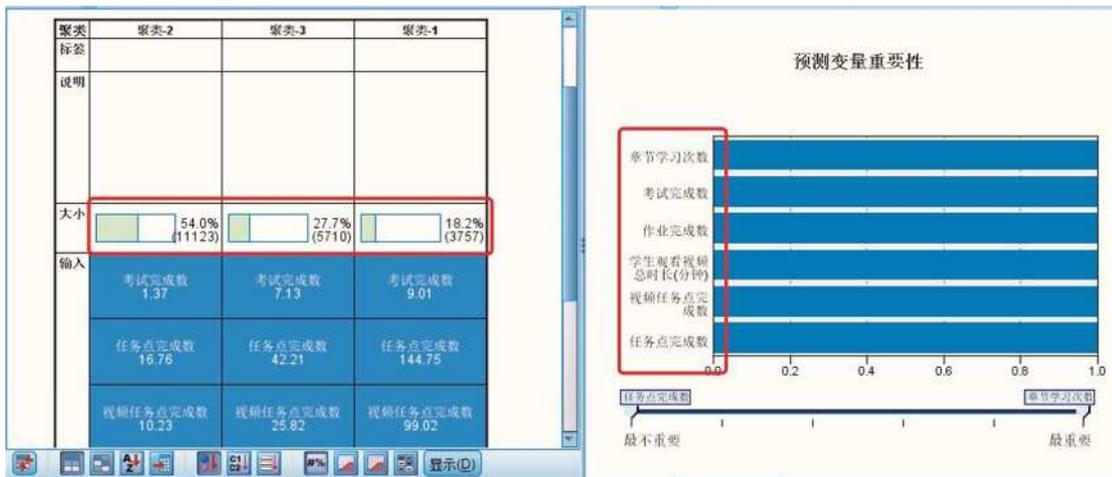


图 5 线下教学时期聚类结果人数占总数的比例及预测变量的重要性

结果显示，凝聚和分离的轮廓测量约等于 0.5，说明聚类结果尚好。通过上图预测变量的重要性可以看出：

所有字段的重要性都为 1, 说明这些变量对这次聚类结果起较为重要的作用。通过上图可以看出三个聚类结果

人数占总数的比例, 分别为 54%、27.7% 和 18.2%。

2. 均值结果与分析

表 4 线下教学时期均值结果

单元格内容: 平均值				
字段	聚类 -1*	聚类 -2*	聚类 -3*	重要性
任务点完成数	144.745	16.761	42.206	1
视频任务点完成数	99.022	10.232	25.824	1
作业完成数	30.385	4.172	20.641	1
考试完成数	9.009	1.372	7.133	1
章节学习次数	491.659	66.089	232.848	1
视频观看时长(分钟)	1183.416	114.454	337.074	1

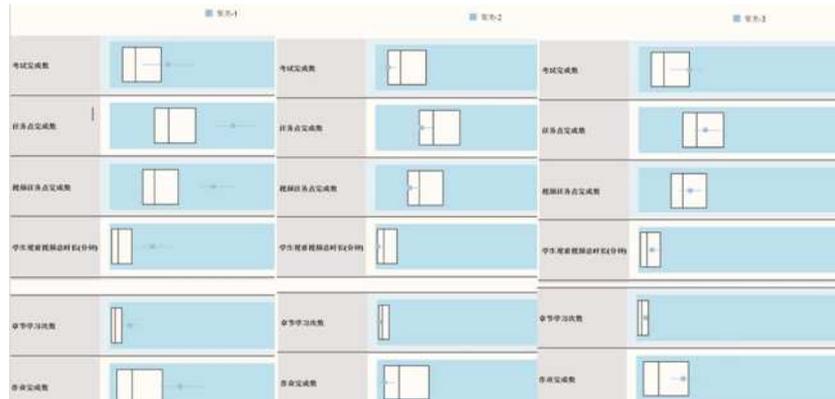


图 6 线下教学时期重要指标箱形图

通过上图得出: 第一类学生在所有指标上表现的都是最突出的, 所以可以将第一类学生划分成“学习任务紧张、学习氛围紧张, 学习努力”的学生, 这类学生占全校的 18.2%。第二类学生在所有指标上表现的都不太好, 所以可以将第二类学生划分成“学习任务轻松、学习氛围轻松, 努力完成老师任务”的学生, 这类学生占全校的 54%。第三类学生在所有指标上的表现均为中等水平, 所以可以将第三类学生划分成“学习任务较为轻松、学习氛围较为轻松, 且学习较为努力”的学生, 这类学生占全校的 27.7%。

总结如下表:

表 5 线下教学学生的情况分布占比图

学生分类	学习任务	学习氛围	学习努力程度	占比
第一类	3	3	3	18.20%
第二类	1	1	1	54.00%
第三类	2	2	2	27.70%

4.2.4 线上线下不同时期对比分析总结

学习任务: 线下教学时期学习任务轻松的人数变少, 学习任务中等的人数变多。

学习氛围: 线下教学时期学习氛围轻松和紧张的人数变多, 学习氛围中等的人数变少。

努力程度: 线下教学时期努力程度中等和努力的人数变多, 努力程度轻松的人数变少。

线上教学时老师可能主要以线上教学、考试、作业为主, 学习任务点设置的较为少, 而线下教学时则以学习任务点为主。线上教学时较为依赖同学们的自觉主动性进行学习, 所以在线上教学时大部分同学们的学习

努力情况都不太好, 而在返校学习后有部分同学可能恢复了学习状态, 使得线下教学时有一半左右的同学努力情况都呈中等及以上。

5 结论及建议

线上线下学习的方式不同, 所以对于网络教育平台的利用有所不同; 但无论是上线还是线下, 对于章节学习次数, 视频任务点完成数的影响是最大的, 考试完成数是影响是最小的。所以提高视频任务点的完成数, 对章节学习次数提高有效。由于之前是网络教学, 那么学习任务点会少一点, 而回到学校是线下教学, 则以学习任务点为主, 又由于在家学习的原因大部分同学们的学习努力情况都不太好, 而在返校学习后, 由线上教学转化为线下教学, 同学们的努力情况好些。

【参考文献】

- [1] 冯广, 何雅萱, 贺敏慧. 基于校园大数据的学生画像系统应用研究 [J], 软件, 2020 (08): 40-42
- [2] 严正宇. 用“大数据+微服务”构建学生综合画像 [J], 信息系统工程, 2019 (12): 105-106
- [3] 黄山. 学生画像技术在高校学生事务精准管理的运用 [J], 科学咨询 (教育科研), 2019 (12): 37
- [4] 程泽凯, 余星星, 谢宁宇. 网络教学平台学生学习数据分析 [J], 常州工学院学报, 2019 (04): 77-80
- [5] 李光耀, 宋文广, 谢艳晴. 智慧校园学生画像方法研究 [J], 现代电子技术, 2018 (06): 161-163