

基于粗糙集对高等数学课程成绩的影响因素分析

刘巧云

青海民族大学数学与统计学院, 青海 西宁

收稿日期: 2021年11月10日; 录用日期: 2021年12月13日; 发布日期: 2021年12月20日

摘要

针对普遍认为的高等数学难的情况, 采集了不同专业120名学生学习高数的情况与高等数学的综合成绩, 利用粗糙集理论中上下近似的相关理论对采集的数据进行分析, 得到数据之间的依赖程度, 最终得到高等数学是否能通过的主要影响因素与得高分的主要影响因素。

关键词

粗糙集, 上下近似, 影响因素, 支持度, 重要性

Influence Factors of Higher Mathematics Course Grades Based on Rough Set Theory

Qiaoyun Liu

College of Mathematics and Statistics, Qinghai Minzu University, Xining Qinghai

Received: Nov. 10th, 2021; accepted: Dec. 13th, 2021; published: Dec. 20th, 2021

Abstract

With the view that higher mathematics is difficult, 120 students' situation from different majors to study high mathematics and the grades of higher mathematics are collected. By using the related theory of upper and lower approximation in rough set theory, the degree of dependence between

the data is obtained, and the main influencing factors of whether higher mathematics can pass or not and the main influencing factors of high scores are finally obtained.

Keywords

Rough Set, Upper and Lower Approximation, Influence Factors, Support Degree, Significance

Copyright © 2021 by author(s) and Hans Publishers Inc.

This work is licensed under the Creative Commons Attribution International License (CC BY 4.0).

<http://creativecommons.org/licenses/by/4.0/>



Open Access

1. 研究背景

粗糙集理论是一种处理信息的方法, 可以从先验信息中发现规律和一些潜在的知识。学习成绩的影响因素具有不确定的特点, 范娟[1]利用粗糙集理论研究了数据结构这门课五个因素对学习成绩的影响, 高丽红[2]基于粗糙集理论对 100 个学生的学习成绩进行了影响因素分析。在这些研究的基础上, 本文针对高等数学的成绩进行了因素分析。

《高等数学》是高等院校理工科专业的专业基础课、必修课。这门课的学习情况关系到后续专业课程的学习。自参加工作以来, 一直在从事高等数学课程的教学工作, 在期末考试结束后, 考试结果不尽人意。从这个事实出发, 选择在高等数学学习过程中对成绩有影响的 8 个因素作为对象, 随机收集了本校不同专业 120 名学生的情况与成绩并进行一系列的分析。本文的安排如下: 第 1 部分简要阐述粗糙集理论中的相关概念; 第 2 部分对采集到的数据进行分析; 第 3 部分给出主要结论。

2. 预备知识

这部分介绍与本文研究内容相关的基本概念。

定义 1 [3] 设四元数组 $S = (U, A, V, F)$, 其中 $U = \{u_1, u_2, \dots, u_{|U|}\}$ 是研究对象的非空有限集合, 称为论域; $A = \{a_1, a_2, \dots, a_{|U|}\}$ 是属性的非空集合; $V = \bigcup_{a \in A} V_a$ 表示 A 的值域, 其中 $a \in A$, V_a 是属性 a 的值域; $f: U \times A \rightarrow V$ 是一个映射, 反映了每个研究对象每个属性的信息。知识表达系统也叫信息系统, $S = (U, A, V, F)$ 可以简化为 $S = (U, A, V, F)$ 。

定义 2 [3] 设 (U, A, V, F) 是一个信息系统, 对于 U 中的一个划分 U/a , 两个对象 $u, v \in U$ 在同一类中当且仅当 $a(u) = a(v)$ 。

定义 3 [3] 对于给定的信息系统 $S = (U, A, V, F)$, 对于每个子集 $X \subseteq U$ 和等价关系 $R \in \text{ind}(K)$, 给出如下定义:

X 的 R 上近似集: $\underline{R}X = \bigcup \{x \in U \mid [x]_R \subseteq X\}$,

X 的 R 下近似集: $\overline{R}X = \bigcup \{x \in U \mid [x]_R \cap X \neq \emptyset\}$ 。

称 $\text{pos}_R(X) = \underline{R}X$ 为 X 的正域, $\text{neg}_R(X) = U - \overline{R}X$ 为 X 的负域。 $\overline{R}X$ 或 $\underline{R}X$ 是根据知识 R 判断 U 中一定属于 X 的元素组成的集合; $\overline{R}X$ 是根据知识 R 判断 U 中可能属于 X 的元素组成的集合; $\text{neg}_R(X)$ 是根据知识 R 判断 U 中一定不属于 X 的元素组成的集合。

定义 4 [3] 当 $A = C \cup D$, $C \cap D = \emptyset$, 那么称集合 C 为条件属性集, 集合 D 为决策属性集。属性 $a \in X$, 则 a 关于 D 的重要性被定义为: $\sigma_{CD}(a) = \gamma_C(D) - \gamma_{C-\{a\}}(D)$, 其中 $\gamma_C(D) = \frac{|\text{pos}_C(D)|}{|U|}$ 。

3. 主要分析过程

3.1. 描述性分析

从采集到的 120 个样本数据看,得到下面的统计数据:认为自己数学基础好的有 10 个,占总样本的 8.3%;认为数学基础中等水平的有 84 个,占总样本的 70%;认为数学基础不好的有 26 个,占总样本的 21.7%。为了描述更简洁,列表如下表 1。

Table 1. Sample data statistics table

表 1. 样本数据统计表

因素	情况			因素	情况		
	人数与占比	人数与占比	人数与占比		人数与占比	人数与占比	人数与占比
数学基础	好	中	差	课后作业	独立	基本独立	抄袭或不写
	10	84	26		55	62	3
	8.3%	70%	21.7%		45.8%	51.7%	2.5%
课前预习	经常	偶尔	从不	干部、成员	是	否	-
	13	98	9		95	25	-
	10.8%	81.7%	7.5%		79.2%	20.8%	-
上课效率	高	中	低	兴趣	有	没有	-
	11	94	15		105	15	-
	9.2%	78.3%	12.5%		87.5%	12.5%	-
课后复习	经常	偶尔	从不	做题情况	是	否	-
	17	97	6		44	76	-
	14.2%	80.8%	5%		36.7%	63.3%	-

从表 1 可以看出,学生普遍课前不预习,导致在课堂上漫无目的的学习,学习效率不高;大部分都是社团协会成员或者班干部,本校学生参加活动的热情与积极性较高;很大一部分学生对数学都是有兴趣的;大部分学生只做了课上的例题和课后作业,没有做其他的题目。

从通过率来看,在这 120 个样本中,有 21 个综合测评成绩没有达到 60,占调查人数的 17.5%。从成绩的分布来看,16 人成绩在 85 以上,占 13.3%,48 人成绩在 75~85,占 40%,35 人成绩在 60~75,占 29.2%。成绩分布基本合理。

3.2. 影响考试通过的主要因素分析

利用传统粗糙集理论中的属性约简方法,将考试成绩作为决策属性,研究列出来的 8 个因素对最后的考试成绩的影响程度。

首先,需要对数据做预处理。将成绩分为 5 段:85~100 分,75~85 分,60~75 分,50~59 分,50 以下,并分别用 1、2、3、4、5 来进行表示,即决策属性 $D_1 = \{1, 2, 3, 4, 5\}$ 。将通过考试(成绩高于 60)用 1 表

示, 没有通过考试(低于 60)用 0 表示, 即决策属性 $D_2 = \{0, 1\}$ 。将影响因素数学基础、课前预习、听课效率、课后复习、完成作业方式、班团干部或者协会成员、对数学的兴趣、是否做其他题目分别记为 $x_1, x_2, x_3, x_4, x_5, x_6, x_7, x_8$, 即条件属性 $C = \{x_1, x_2, x_3, x_4, x_5, x_6, x_7, x_8\}$ 。取值情况列表如下表 2。

其次, 计算等价类 U/C (表示调查对象根据所有的条件属性进行划分等价类)、 U/D_1 (表示调查对象根据 D_1 即考试是否通过这一决策属性进行划分等价类)、 $U/(C - \{x_i\}) (i = 1, 2, 3, 4, 5, 6, 7, 8)$ (表示调查对象依次去掉某一条件属性后进行划分等价类), 并计算出每一个条件属性的重要性。

Table 2. Value table

表 2. 取值情况表

属性	取值情况			属性	取值情况		
	1	2	3		1	2	3
x_1	好	中	差	x_5	独立	基本独立	抄袭
x_2	经常	偶尔	从不	x_6	是	否	-
x_3	高	中	低	x_7	有	没有	-
x_4	经常	偶尔	从不	x_8	是	否	-

经计算, 得 $\gamma_C(D_1) = \frac{|\text{pos}_C(D_1)|}{|U|} = \frac{93}{120} = 0.775$, 说明考试通不通过与这 8 个因素的依赖度为 77.5%。

$$\sigma_{CD_1}(x_1) = \gamma_C(D_1) - \gamma_{C-\{x_1\}}(D_1) = \frac{93}{120} - \frac{81}{120} = 0.100$$

$$\sigma_{CD_1}(x_2) = \gamma_C(D_1) - \gamma_{C-\{x_2\}}(D_1) = \frac{93}{120} - \frac{81}{120} = 0.100$$

$$\sigma_{CD_1}(x_3) = \gamma_C(D_1) - \gamma_{C-\{x_3\}}(D_1) = \frac{93}{120} - \frac{75}{120} = 0.150$$

$$\sigma_{CD_1}(x_4) = \gamma_C(D_1) - \gamma_{C-\{x_4\}}(D_1) = \frac{93}{120} - \frac{89}{120} \approx 0.033$$

$$\sigma_{CD_1}(x_5) = \gamma_C(D_1) - \gamma_{C-\{x_5\}}(D_1) = \frac{93}{120} - \frac{73}{120} \approx 0.167$$

$$\sigma_{CD_1}(x_6) = \gamma_C(D_1) - \gamma_{C-\{x_6\}}(D_1) = \frac{93}{120} - \frac{83}{120} \approx 0.083$$

$$\sigma_{CD_1}(x_7) = \gamma_C(D_1) - \gamma_{C-\{x_7\}}(D_1) = \frac{93}{120} - \frac{92}{120} = 0.008$$

$$\sigma_{CD_1}(x_8) = \gamma_C(D_1) - \gamma_{C-\{x_8\}}(D_1) = \frac{93}{120} - \frac{91}{120} = 0.017$$

3.3. 影响考试成绩高低的因素分析

计算等价类 U/D_2 (将调查对象根据成绩高低进行划分等价类), 并计算出 8 个因素整体与各个因素的重要性, 得到下面的结果:

$$\gamma_C(D_2) = \frac{|\text{pos}_C(D_2)|}{|U|} = \frac{54}{120} = 0.45$$

$$\sigma_{CD_2}(x_1) = \gamma_C(D_2) - \gamma_{C-\{x_1\}}(D_2) = \frac{54}{120} - \frac{41}{120} = \frac{13}{120} \approx 0.108$$

$$\sigma_{CD_2}(x_2) = \gamma_C(D_2) - \gamma_{C-\{x_2\}}(D_2) = \frac{54}{120} - \frac{39}{120} = \frac{15}{120} = 0.125$$

$$\sigma_{CD_2}(x_3) = \gamma_C(D_2) - \gamma_{C-\{x_3\}}(D_2) = \frac{54}{120} - \frac{40}{120} = \frac{14}{120} = 0.117$$

$$\sigma_{CD_2}(x_4) = \gamma_C(D_2) - \gamma_{C-\{x_4\}}(D_2) = \frac{54}{120} - \frac{39}{120} = \frac{15}{120} = 0.125$$

$$\sigma_{CD_2}(x_5) = \gamma_C(D_2) - \gamma_{C-\{x_5\}}(D_2) = \frac{54}{120} - \frac{37}{120} = \frac{17}{120} \approx 0.142$$

$$\sigma_{CD_2}(x_6) = \gamma_C(D_2) - \gamma_{C-\{x_6\}}(D_2) = \frac{54}{120} - \frac{36}{120} = \frac{18}{120} = 0.15$$

$$\sigma_{CD_2}(x_7) = \gamma_C(D_2) - \gamma_{C-\{x_7\}}(D_2) = \frac{54}{120} - \frac{42}{120} = \frac{12}{120} = 0.10$$

$$\sigma_{CD_2}(x_8) = \gamma_C(D_2) - \gamma_{C-\{x_8\}}(D_2) = \frac{54}{120} - \frac{49}{120} = \frac{5}{120} = 0.042$$

4. 结论

从 2.1 描述性分析来看, 任课教师应该在激发学生的学习积极性方面更加努力, 比如怎么让学生能在课前主动进行预习、在课后及时巩固复习等。目前我校正在采用线上线下结合的模式教学, 希望学生能动起来, 不管是高数还是其他课程期末考试成绩都能提高。

从 2.3 粗糙集理论分析结果看, 上述 8 个因素对高等数学课程的影响程度为 77.5%, 总体对成绩影响较大。从单个因素看, 这八个因素的重要性从数据看都比较小, 可能我们问问题的方式不太准确, 学生在回答的时候也有一定的隐瞒, 不太符合真实情况。数学的基础对能否通过高数考试与考高分的影响不大。当学生干部等对能否通过考试影响不大, 但是对考高分有一定的影响。课前预习与课后复习这两个因素对考试能否通过与考高分影响也较大。从这些样本分析得到做题多少对二者影响都不大, 这个可能与期末考试题目偏简单有一定关系, 在以后的出题过程中要注意这个问题。是否独立完成课后作业对二者影响都较大。

综合起来, 以后在教学过程中要注意对学生学习积极性的培养, 利用各种教学平台让学生课前预习、课后能复习并独立完成作业, 在期末考试题目的难度方面, 适当增加一些难度。

基金项目

青海民族大学校级项目(2021XJGH26)。

参考文献

- [1] 范娟. 基于粗糙集理论对影响高校学生成绩因素的分析[J]. 电脑迷, 2017(5): 64-65.
- [2] 高丽红. 基于粗糙集理论的大学生学习成绩影响因素分析[J]. 科学技术与工程, 2007, 7(4): 521-524.
- [3] 张文修, 等, 编著. 粗糙集理论与方法[M]. 北京: 科学出版社, 2001.