

基于粗糙集的大学生心理健康特征分析

刘 峰, 魏 明

(南通大学 校长办公室, 江苏 南通 226019)

摘 要:文章提出一种基于粗糙集的大学生心理健康特征关联性挖掘模型,分析大学生的心理健康程度与学生的基本属性、家庭基本情况、学习情况、社交活动等方面的关联性,以及各属性变化对大学生心理健康的影响程度,生成它们之间潜在的关联规则或知识。最后,以南通大学学生为例,开展问卷调查,分析大学生心理健康的总体情况,为学校学生管理工作做好大学生心理健康工作提供智能决策。

关键词:大学生;心理健康;特征选取;粗糙集

中图分类号:G444

文献标识码:A

文章编号:1671-9891(2017)02-0087-05

0 引言

目前,大学生的心理健康是我国高校管理过程重点关注的热点问题之一。^[1]大学生的心理素质不仅影响到学生自身的学习、工作和生活,而且也关系到整个民族素质的提高。因此,如何提高大学生的心理健康水平已成为社会各界所共同关注的重要课题。影响大学生心理健康的因素很多,涉及大学生的基本属性、家庭基本情况、学习情况、社交活动等方面,各因素的重要性也会有很大的差别。对于大学生心理健康特征分析,传统方法离不开人的因素,如:定量方法在确定各因素的权重中充分利用了专家的主观经验,定性方法通过专家打分法得到每个指标的评价值,存在一定的主观性和片面性的问题。^[2-4]

粗糙集是一种处理不确定问题的数据挖掘工具,与传统的多元线性回归、结构方程等方法相比,具备处理不完备、不精确、不一致数据方面等优点。^[5-7]本文提出一种基于粗糙集的大学生心理健康特征关联性分析方法,通过分析与其相关联的大学生自身的基本属性、家庭基本情况、学习情况、社交活动等方面之间因果关系,从客观的数据中得出了各主要因素的客观权重,并挖掘一系列评价规则。

1 大学生心理健康的特征数据调查问卷设计

影响大学生心理健康的因素众多,笔者将大学生心理健康特征的调查内容主要分为四部分:(1)大学生的基本属性,包括:性别、年龄、血型、性格类型、气质类型等;(2)大学生家庭基本信息,包括:是否独生子女、是否单亲家庭、贫困情况、家庭居住地、生源地等;(3)学习情况,包括:获奖学金情况、不及格门数、参加课外科技活动情况等;(4)社交活动,包括:参加社团活动、担任学生干部、恋爱情况、人际关系等。^[8-10]每个新学年的开始,学校都会对一年级和三年级学生进行心理普测,我们将心理普测的结果,即心理健康状态与问卷调查的数据相关联。大学生心理健康调查问卷如表1所示。

2 基于粗糙集的大学生心理健康特征分析方法

2.1 粗糙集基本概念

粗糙集理论(Rough Set)是一种不确定数据处理工具,它不需要研究对象的任何先验知识,就可以有效地分析和处理其不确定性,已经在军事、电信等诸多领域有了广泛的应用。

定义1:定义 U 为论域,将 R 视为 U 上一个等价关系。若 $P \subseteq R$ 且 $P \neq \emptyset$,则 $\cap P$ 是一等价关系,称为 P 的

收稿日期:2017-03-22

基金项目:中国交通教育研究会教育科学研究立项课题“基于粗糙集的大学生心理健康特征分析”(项目编号:交教研1402-62);2013年南通大学学生工作专项研究课题“基于粗糙集的大学生心理健康特征分析”(项目编号:13xgzx14)。

作者简介:刘峰(1981—),男,江苏南通人,南通大学校长办公室讲师,硕士。

表 1 大学生心理健康调查问卷

纬度	属性	类别
基本属性 a_1	性别 a_{11}	男/女
	年龄 a_{12}	17-28
	血型 a_{13}	A/B/O/AB
	性格类型 a_{14}	外倾型/内倾型/混合型
家庭基本情况 a_2	气质类型 a_{15}	多血质/粘液质/胆汁质/抑郁质
	独生子女 a_{21}	是/否
	单亲家庭 a_{22}	是/否
	贫困情况 a_{23}	特困/普困/一般贫困/不贫困
	家庭居住地 a_{24}	城市/农村
	生源地 a_{25}	东部地区/西部地区/中部地区
学习情况 a_3	获奖学金情况 a_{31}	有/无
	不及格门数 a_{32}	0-30
	课外科技活动 a_{33}	参加/偶尔参加/不参加
社交活动 a_4	参加社团活动 a_{41}	参加/偶尔参加/不参加
	担任学生干部 a_{42}	是/否
	恋爱情况 a_{43}	有/没有
	人际关系 a_{44}	融洽/一般/不好
心理健康状态 a_5	总体评价	健康/不健康

不可区分关系,记为 $ind(P)$,且有: $[x \in U]_{ind(P)} = \bigcap_{R \in P} [x \in U]_R$ 。

定义 2:对论域 U 上的任意等价关系 $\forall r \in R$,如果 $ind(R) = ind(R - \{r\})$,则称 r 为 R 中不必要关系,否则为必要关系。设 Q 和 P 为论域 U 的两个等价关系, $Q \subseteq P$,如果 Q 是独立的,且 $ind(Q) = ind(P)$,则称 Q 为 P 的一个约简。

由上可知,粗糙集理论的核心在于将问题 $S = (U, C, D, V, f)$ 抽象为一个信息系统,根据知识简约原理,利用 $R = \{C \rightarrow D\}$ 将 U 等价划分为 n 类,即 $U/R = \{[x_1]_R, [x_2]_R, \dots, [x_n]_R\}$,其中: $\forall x_i \in U, [x_i]_R$ 表示包含 $\forall x_i \in U$ 的 R 等价类。 C 和 D 分别为条件和决策属性集,满足 $U = C \cap D \neq \emptyset; V = \cup \{V_r | r \in R\}, V_r$ 为属性在关系 $r \in R$ 的值域; $f: U \times (C \cup D) \rightarrow V$ 是一个信息函数,它为 $\forall x \in U$ 在关系 $r \in R$ 赋予一个信息值 $f(x, r) \in V_r$ 。

2.2 大学生心理健康特征的粗糙集挖掘过程

(1)大学生的心理健康特征决策表建立。在构建所有大学生的心理健康特征决策表时,将所有学生调查问卷视为论域 U ,单独每个学生的问卷数据 x_i 就是 U 中的一个对象,总共 n 个样本。将大学生心理健康的影响因素构成条件属性集 C ,如:学生的基本属性 a_1 、家庭基本情况 a_2 、学习情况 a_3 、社会活动 a_4 等,将大学生心理健康的类别构成决策属性集 D ,如:心理健康总体评价 a_5 等,作为决策属性集 D ,得到如表 2 所示的二维决策表 $S = (U, C, D, V, f)$ 。

表 2 决策信息表(离散前)

U	C				D
	a_1	a_2	a_3	a_4	a_5
x_1	(男,19,A,外倾型,多血质)	(是,否,不贫困,城市,东部地区)	(有,无,参加)	(参加,是,恋爱过,融洽)	健康
x_2	(女,20,B,混合型,粘液质)	(否,否,普困,农村,中部地区)	(无,无,参加)	(不参加,否,没有恋爱过,融洽)	健康
...
x_3	(男,21,O,内倾型,抑郁质)	(否,是,特困,农村,西部地区)	(无,5,不参加)	(偶尔参加,否,没有恋爱,一般)	不健康

在上述决策表中,每一行某列的数值为某个调查样本相应调查项的结果。

(2)离散化处理。由于粗糙集仅擅长处理离散型属性数据,无法较好地处理语言类或连续性属性数据,故有必要对上述属性值离散化。由上可知,某些属性是语言类型,如: a_1 的性别、血型、性格类型、气质类型, a_2 的独生子女、单亲家庭、贫困情况、家庭居住地、生源地, a_3 的获奖学金情况、课外科技活动, a_4 的社团活动、担任学生干部情况、恋爱情况、人际关系, a_5 的总体评价,这些属性可以利用自然数0、1、2、3、4、...刻画它们的类别;某些属性可以是具体数值,如: a_1 的年龄、 a_3 的不及格门数等,采用等距离划分方法,根据属性的量化范围,进行某属性所有连续值的离散归一化。离散化后的决策信息表见表3所示。

表3 决策信息表(离散后)

U	C				D
	a_1	a_2	a_3	a_4	a_5
x_1	(1,3,1,1,1)	(1,2,4,1,1)	(1,2,2)	(1,1,1,1)	1
x_2	(0,4,2,3,2)	(2,2,2,2,3)	(2,2,1)	(3,2,2,1)	1
...
x_3	(1,5,3,2,4)	(2,1,1,2,2)	(2,5,3)	(2,2,2,2)	2

(3)大学生心理健康特征的属性重要度分析。利用启发式属性约简方法,引入“粒度函数”概念,计算条件属性如何影响决策属性分类的程度,可将大学生心理健康的特征属性选择与其总体评价联系起来。

定义3:对论域上的一个等价关系 $U/R = \{[x_1]_R, [x_2]_R, \dots, [x_n]_R\}$, R 的知识粒度定义为 $GD(R) = (| [x_1]_R |^2 + | [x_2]_R |^2 + \dots + | [x_n]_R |^2) / |U|$, $|X|$ 表示集合 X 的基数。

定义4:对论域 $U = C \cup D$, 将 D 相对于 C 在论域 U 上的分辨能力视为粒度函数 $\psi_C(D) = GD(C) - GD(C \cup D)$ 。显然,若 $\psi_C(D)$ 变小,则 D 相对于 C 对 U 中对象的分辨力减弱;否则,分辨力增强。

为了揭示大学生心理健康特征的条件属性 C 对其决策属性 D 的影响,根据定义3和4,在大学生心理健康特征的决策表中,主要进行以下两项工作。

①计算大学生心理健康特征的任意条件属性 $\forall a \in C$ 对其总体评价决策属性集 D 的重要度,记为 $Sig(a, C, D) = \psi_{C-[a]}(D) - \psi_C(D)$ 。 $Sig(a, C, D)$ 刻画条件属性 C 若缺少属性 a 后不能准确分类的对象所占的比例。若 $Sig(a, C, D) = 0$, 则属性 a 对分类无影响;否则, $Sig(a, C, D)$ 的值越大决定属性 a 对分类的影响越大。

②若 $C_0 \subseteq C$, 计算大学生心理健康的某条件属性 $\forall A \in C - C_0$ 关于属性集 C_0 对 D 的重要度,记为 $Sig(a, C_0, D) = \psi_{C_0}(D) - \psi_{C_0 \cup \{a\}}(D)$ 。 $Sig(a, C_0, D)$ 描述 C_0 中添加属性 a 后其分类能力的提高程度。若 $Sig(a, C_0, D)$ 变大,则属性 $\forall A \in C - C_0$ 影响属性集 C_0 对 D 的分类作用越大。也就是说,当 $Sig(a, C_0, D) = 0$ 时,属性 a 关于 C_0 对 D 的分类无作用,无需将属性 a 添加到 C_0 中。

(4)影响大学生心理健康特征因素的选择。在分析大学生心理健康属性重要性基础上,进行属性约简。

①根据 $Sig(a, C, D)$ 计算大学生心理健康特征分析决策表的相对核,将其作为求最小约简的起点,令 $C_0 = Core_C(D) = \{a \in C | Sig(a, C, D) > 0\}$ 。

②针对表1的所有可能影响大学生心理健康的属性因素,将 $Sig(a, C_0, D)$ 最大的属性 a 都添加到 C_0 中,直到 $\psi_{C_0}(D) = \psi_C(D)$, 得到大学生心理健康分类的最简影响因素。

启发式属性约简特征选择算法的实现如下。输入为:决策信息系统 $S = (U, C, D, V, f)$ 。输出为:该 S 的一个 C 和 D 最小相对约简。

步骤1:针对某属性 $\forall a \in C$, 计算其重要度 $Sig(a, C, D)$, 选取所有 $Sig(a, C, D) > 0$ 的属性组成 $Core_C(D)$, 令 $C_0 = Core_C(D)$, 重复 step2 和 step3, 直至 $\psi_{C_0}(D) = \psi_C(D)$;

步骤2:对条件属性集 $C - C_0$ 中的所有属性 $\forall a \in C$, 分别计算它们的 $Sig(a, C_0, D)$;

步骤3:选择使 $Sig(a, C_0, D)$ 最大的属性 $\forall a \in C$ (若同时有多个属性满足 $Sig(a, C_0, D)$ 最大,则任一个选择使 $\psi_{C_0 \cup \{a\}}(D)$ 最小的属性 $\forall a \in C$), 令 $C_0 = C_0 \cup \{a\}$;

步骤4:将 C_0 输出,其为决策信息系统 S 的条件属性集的最小相对约简。

3 实验结果与分析

以南通大学为例,采取以开展问卷调查为主、实地调查访问为辅的调查方式,收集广大大学生心理健康问卷的数据,利用 matlab 编写求解该问题的粗糙集程序,分析影响大学生心理健康特征的各个属性重要度,挖掘心理健康特征与学生之间的关联规则,据此分析其原因,并针对突出和潜在的问题进行政策反馈。

3.1 调查数据源

本次调查时间为 2014 年 9 月 14 日至 9 月 24 日,以南通大学为调查地点,随机派发调查问卷,通过运用个别谈话访问的调研方法,监督了受调查者填写问卷的严谨性和真实有效性。总共派发问卷 1526 份,除去无效问卷 181 份。通过对合格问卷的选项进行整理统计和汇总可知:男性和女性占的比例为 48.2%、51.8%,调查对象在年龄段 20 岁以下、20-22 岁、23-25 岁、25 岁以上所占比例为 31.1%、42.4%、20.7%、5.8%,来自城市和农村的大学生分别占 42.3%、57.7%,来自东部地区、西部地区、中部地区的大学生所占比例为 48.2%、21.3%、30.5%,独生子女和非独生子女占比为 58.1%、41.9%,他们中特困家庭占 12.2%、普困家庭占 26.7%、一般贫困家庭占 18.6%、不贫困的家庭占 42.5%。由上可知,所有调查对象具有一定的层次性,样本完全有效支撑分析大学生心理健康的特征。

3.2 大学生心理健康的特征分析

(1)总体特征分析。从表 4 可知,大学生心理健康的水平与他们的性别、年龄等属性关联度不大,与性格类型、气质类型、独生子女、单亲家庭、贫困情况、家庭居住地、生源地、获奖学金情况、不及格门次、课外科技活动、社团活动、担任学生干部情况、恋爱情况和人际关系等属性有关联。

表 4 各个属性的总体评价

	基本属性					家庭基本情况					学习情况			社交情况			
	性别	年龄	血型	性格类型	气质类型	独生子女	单亲家庭	贫困情况	家庭居住地	生源地	获奖学金情况	不及格门次	课外科技活动	参加社团活动	担任学生干部	恋爱情况	人际关系
男生	0.82	0.81	0.82	0.99	0.97	0.98	0.97	0.92	0.90	0.89	0.89	0.92	0.85	0.95	0.95	0.98	0.98
女生	0.84	0.82	0.83	0.97	0.96	0.96	0.95	0.94	0.91	0.87	0.88	0.93	0.87	0.95	0.96	0.99	0.97

(2)属性重要度。若某个条件属性多次出现在一系列规则时,其决定了大学生心理健康特征的情况。根据 $Sig(a, C, D) = \psi_{C-[a]}(D) - \psi_C(D)$ 和 $Sig(a, C_0, D) = \psi_{C_0}(D) - \psi_{C_0 \cup [a]}(D)$ 计算某个条件属性对决策属性的重要度,总共筛选影响大学生心理健康的 14 个重要属性,其中:大学生的基本属性仅与性格类型、气质类型关联性较大;除了大学生基本属性之外,大学生心理健康的影响因素重要度高低依次为恋爱情况、人际关系、独生子女、单亲家庭、担任学生干部情况、社团活动、贫困情况、不及格门次、家庭居住地、生源地、课外科技活动和获奖学金情况。

(3)决策规则。利用上述 80% 的问卷调查数据作为训练样本,根据前文的心理健康决策规则生成方法,生成 2 000 多决策规则,大多数决策规则仅符合一或两个样本,将大学生心理健康水平分为健康、不健康综合评价,从中筛选最好的决策规则,如表 5 所示。

表 5 决策规则

决策规则	心理健康状况
$(a_{11}=\text{男}) \wedge (a_{14}=\text{外倾型}) \wedge (a_{15}=\text{多血质}) \wedge (a_{21}=\text{是}) \wedge (a_{22}=\text{否}) \wedge (a_{23}=\text{不贫困}) \wedge (a_{24}=\text{城市}) \wedge (a_{25}=\text{东部地区}) \wedge (a_{31}=\text{有}) \wedge (a_{32}=\text{无}) \wedge (a_{33}=\text{经常参加}) \wedge (a_{41}=\text{偶尔参加}) \wedge (a_{42}=\text{是}) \wedge (a_{43}=\text{有}) \wedge (a_{44}=\text{融洽}) \rightarrow (a_5=\text{健康})$	健康
$(a_{11}=\text{女}) \wedge (a_{14}=\text{混合型}) \wedge (a_{15}=\text{粘液质}) \wedge (a_{21}=\text{否}) \wedge (a_{22}=\text{否}) \wedge (a_{23}=\text{普困}) \wedge (a_{24}=\text{农村}) \wedge (a_{25}=\text{西部地区}) \wedge (a_{31}=\text{无}) \wedge (a_{32}=\text{无}) \wedge (a_{33}=\text{偶尔参加}) \wedge (a_{41}=\text{不参加}) \wedge (a_{42}=\text{否}) \wedge (a_{43}=\text{没有}) \wedge (a_{44}=\text{一般}) \rightarrow (a_5=\text{健康})$	
$(a_{11}=\text{男}) \wedge (a_{14}=\text{外倾型}) \wedge (a_{15}=\text{胆汁质}) \wedge (a_{21}=\text{否}) \wedge (a_{22}=\text{是}) \wedge (a_{23}=\text{特困}) \wedge (a_{24}=\text{农村}) \wedge (a_{25}=\text{中部地区}) \wedge (a_{31}=\text{无}) \wedge (a_{32}=3-4 \text{ 门}) \wedge (a_{33}=\text{不参加}) \wedge (a_{41}=\text{不参加}) \wedge (a_{42}=\text{否}) \wedge (a_{43}=\text{没有}) \wedge (a_{44}=\text{不好}) \rightarrow (a_5=\text{不健康})$	不健康
$(a_{11}=\text{女}) \wedge (a_{14}=\text{内倾型}) \wedge (a_{15}=\text{多血质}) \wedge (a_{21}=\text{是}) \wedge (a_{22}=\text{是}) \wedge (a_{23}=\text{不贫困}) \wedge (a_{24}=\text{城市}) \wedge (a_{25}=\text{东部地区}) \wedge (a_{31}=\text{有}) \wedge (a_{32}=\text{无}) \wedge (a_{33}=\text{偶尔参加}) \wedge (a_{41}=\text{不参加}) \wedge (a_{42}=\text{否}) \wedge (a_{43}=\text{没有}) \wedge (a_{44}=\text{一般}) \rightarrow (a_5=\text{不健康})$	

4 结束语

本文研究影响大学生心理健康因素之间的因果关系,利用粗糙集模型,以南通大学为例,从大学生的基本属性、家庭基本情况、学习情况、社交活动等四个方面出发,分析它们对大学生心理健康的影响程度,据此挖掘大学生心理状态健康和亚健康状态下四类关联规则,得出了大学生心理健康的影响因素重要度高低依次为恋爱情况、人际关系、独生子女、单亲家庭、担任学生干部情况、社团活动、贫困情况、不及格门次、家庭居住地、生源地、课外科技活动和获奖学金情况。然而,影响大学生心理健康的部分因素是区间数、模糊不确定等,现有研究将其离散化为一个固定值,与实际不相符合,这将是下一步需开展的研究工作。

参考文献:

- [1]丁养斌,孙晟文.大学生心理健康教育现状调研报告[J].中小企业管理与科技(上旬刊),2017(2):165-166.
- [2]漆振羽.浅谈高校辅导员如何做好大学生心理健康教育工作[J].教育现代化,2017(14):143-144.
- [3]李慧.新形势下大学生心理健康教育进宿舍研究[J].赤峰学院学报(自然科学版),2017(5):186-187.
- [4]琚长庭.团体活动在大学生心理健康教育课中的应用[J].安阳工学院学报,2017(1):121-123.
- [5]郑伟,周志强,马艳丽.一种基于粗糙集理论的特征选择方法[J].河北北方学院学报(自然科学版),2009(1):56-59.
- [6]徐久成,史进玲,孙林.一种基于相对粒度的决策表约简算法[J].计算机科学,2009(3):205-207.
- [7]吕震宇,林永民,赵爽,等.基于类信息的文本特征选择与加权算法研究[J].计算机工程与应用,2008(20):145-147.
- [8]黄冰冰,张雷.大学生心理健康数据库建立的价值、现状及路径[J].中国卫生产业,2017(12):163-165.
- [9]方媛.中国大学生心理健康教育创新体系的构建[J].高教学刊,2017(3):177-178.
- [10]王慧.基于Web的大学生心理健康咨询服务系统的设计与实现[J].西安文理学院学报(自然科学版),2017(1):72-76.

Analysis of College Students' Psychological Health Characteristics Based on Rough Set

LIU Feng, WEI Ming

(President's Office, Nantong University, Nantong 226019, China)

Abstract: This article comes up with the correlation mining model for college students' psychological health characteristics based on Rough Set, analyzes the correlation between college students' psychological health level and their basic attributes, basic family conditions, learning conditions and social activities etc., studies the level of influence of attribute variation on college students' psychological health as well as generates the underlying correlation rules or knowledge. Lastly, taking students from Nantong University as an example, it makes a questionnaire survey and analyzes the overall conditions of college students' psychological health, which is expected to provide college students administrators with intelligent decision-making performing well their duties.

Key words: College students; Psychological health; Characteristic selection; Rough Set