

文章编号: 1001-7402(2014)01-0178-06

多源环境下学生综合素质的决策分析*

徐伟华, 罗术群, 张先韬, 孔莎莎

(重庆理工大学 数学与统计学院, 重庆 400054)

摘 要: 在现实生活中, 多源环境下大学生综合素质分析是一个非常重要的课题。针对此问题, 本文从粒计算的角度出发, 结合粗糙集决策的方法, 构建了多源环境下大学生综合素质决策信息系统, 用 Matlab 对该系统进行约简, 并根据约简进行规则提取, 对获得的规则进行分析和验证, 得出了大学生综合素质决策结论。本文是多粒度粗糙集应用的一个范本, 具有重要的理论和实践意义。

关键词: 大学生综合素质; 多粒度粗糙集; 多源环境; 决策分析; 信息系统

中图分类号: TP181 文献标识码: A

1 引言

粗糙集理论^[7-8]是一种处理模糊和不确定性问题的新型数学工具。它已被成功地应用于人工智能、数据挖掘、模式识别与智能信息处理等多个领域^[2,4-6,9]。经典的粗糙集理论是以等价关系为基础对论域进行划分, 划分越细, 获得的信息越充分。

知识约简是粗糙集理论的核心问题之一。众所周知, 知识库中知识(属性)并不是同等重要的, 甚至其中有些属性是冗余的。所谓知识约简, 就是在保持知识库分类能力不变的条件下, 删除其中不相关或不重要的属性。

从粒计算的角度来看, 论域上的一个等价关系为一个粒度, 论域上的一个划分可看作是一个粒空间^[6], 经典粗糙集模型是一个单粒度粗糙集模型。在多粒度粗糙集中, 一个概念通过属性集的划分来进行近似, 这些划分由多个等价关系诱导。本文使用多粒度粗糙集方法, 在多源环境下, 从多个视角对大学生综合素质进行定性的决策分析。

2 预备知识

设 $I = (U, A, V, f)$ 是一个信息系统, U 为对象集, A 为属性集, F 为 U 和 A 的关系集, V 为属性的值域集, $B \subseteq A$, $x \in U$, 属性 B 对应的 x 等价类为:

$$[x]_B = \{y \in U \mid f(x, a) = f(y, a), \forall a \in B\}$$

对于任意 $X \subseteq U$, 集合 $\underline{B}(X)$ 和 $\overline{B}(X)$ 分别称为 X 关于 B 的下近似集和上近似集, 其中 $\underline{B}(X) = \{x \in U \mid [x]_B \subseteq X\}$, $\overline{B}(X) = \{x \in U \mid [x]_B \cap X \neq \emptyset\}$ 。

* 收稿日期: 2012-06-27; 修订日期: 2012-11-05
基金项目: 国家自然科学基金资助项目(61105041); 重庆市自然科学基金资助项目(cstc2011jjA40037); 重庆市教委科技项目(KJ090612)

作者简介: 徐伟华(1979-), 男, 重庆理工大学数学与统计学院教授, 博士, 研究方向: 粗糙集理论与应用; 罗术群(1988-), 女, 重庆理工大学数学与统计学院研究生, 研究方向: 人工智能的数学基础; 张先韬(1986-), 男, 重庆理工大学数学与统计学院研究生, 研究方向: 人工智能的数学基础; 孔莎莎(1988-), 女, 重庆理工大学数学与统计学院学生。

X 称为是可定义集当且仅当 $\underline{B}(X) = \bar{B}(X)$, 否则 X 称为粗糙集^[4, 7-8]。

为了表述广义多粒度粗糙集, 我们首先介绍一个特征函数, 这个函数称为支持特征函数。

定义 2.1^[12] 设 $I = (U, A, V, f)$ 是一个信息系统, $X \subseteq U, p_1, p_2, \dots, p_l \subseteq A$, 设

$$S_X^{P_i}(x) = \begin{cases} 1, & [x]_{P_i} \subseteq X \\ 0, & \text{else} \end{cases} \quad (i \leq l)$$

称 $S_X^{P_i}(x)$ 为 x 关于 X 的支持特征函数, 用来描述等价类 $[x]_{P_i}$ 和概念 X 的包含关系, 表示对象 x 关于 P_i 是否精确支持 X 。

乐观多粒度粗糙集和悲观多粒度粗糙集是两种多粒度粗糙集模型的推广, 我们将提出一个带有参数 $\beta \in (0.5, 1]$ 的新的多粒度粗糙集模型。我们引入这个参数来实现概念刻画时在多数粒度下支持概念的对象被包含进来, 那些可能刻画概念的对象低于相应水平而被忽略掉。这个新的模型表述如下。

定义 2.2^[12] 设 $I = (U, A, V, f)$ 是一个信息系统, $X \subseteq U, p_1, p_2, \dots, p_l \subseteq A, S_X^{P_i}(x)$ 是 x 关于 X 的支持特征函数。对任意 $\beta \in (0.5, 1]$, X 关于 P 的下近似和上近似定义如下。

$$\underline{P}(X)_\beta = \left\{ x \in U \mid \frac{\sum_{i=1}^l S_X^{P_i}(x)}{l} \geq \beta \right\}$$

$$\bar{P}(X)_\beta = \left\{ x \in U \mid \frac{\sum_{i=1}^l (1 - S_X^{P_i}(x))}{l} > 1 - \beta \right\}$$

X 是可定义的当且仅当 $\underline{P}(X)_\beta = \bar{P}(X)_\beta$, 否则 X 是粗糙集。我们将这个广义多粒度粗糙集模型记为 GMGRS, 称 β 为关于 P 的信息水平。

多粒度粗糙集^[10-12] 是经典粗糙集的推广, 由于信息系统中的若干属性对决策作用能够产生不同影响, 当这些作用不能同时进行, 而是各自分开单独进行的时候, 我们就不能使用经典粗糙集理论将这些属性作为一个整体通过一个不可区分关系^[11] 来进行系统的约简和规则的提取。为此, 按照粗糙集决策的一般过程, 我们可以得到多粒度粗糙集决策的具体步骤, 为多源环境下大学生综合素质决策分析提供理论模型。

3 多源环境下大学生综合素质决策信息系统的构建

本节以构建多源环境下大学生综合素质决策信息系统为主要目标, 综合考虑和选择大学生素质表现^[1, 3] 的因素集, 选取决策因素, 将这些因素采用等级考核形式给予赋值, 并对这些因素进行分类, 从而成功构建多源环境下的大学生综合素质决策信息系统。

综合素质测评作为对大学生的一种指标性评价, 测评内容应当包括德育测评、智育测评、体育测评、基本能力测评、拓展素质测评等方面, 根据这些方面的表现, 即道德文化素质、专业素质、心理素质、身体素质等表现, 对大学生进行一个比较全面的整体评测, 能够较为客观的对大学生素质进行考查。

根据上面分析, 为了研究方便, 我们仅进行小规模数据问题的研究, 对属性取值进行简化, 各个条件属性全部采用等级表示分为 A、B、C、D、E 五个等级, 分别对应优秀、良好、一般、差、很差; 决策分为好 (Good)、一般 (Common) 两个决策结果, 这五个条件属性即为五个粒度, 即 $P_1 = \{a_1\}, P_2 = \{a_2\}, P_3 = \{a_3\}, P_4 = \{a_4\}, P_5 = \{a_5\}$ 。

由老师对随机抽取的 6 名同学进行等级评定, 这些老师对随机抽取的同学都比较了解, 经过长期的交往和相处, 其评价结果可以认为构成一个小型的专家系统。

通过这种评定形式, 我们获取到 6 份有效数据, 将 6 个对象构成表 1 的决策信息系统, 表中使用 x_j 来表示学生, 用 P_i 表示各自对应的属性, 具体如表 1。

表1 多源环境下大学生综合素质决策信息系统

U	P_1	P_2	P_3	P_4	P_5	d
x_1	A	A	C	B	C	G
x_2	B	A	C	C	D	Co
x_3	C	D	E	B	D	Co
x_4	B	B	A	B	A	G
x_5	A	B	D	A	B	G
x_6	C	C	C	D	E	Co

注:我们主要是查看约简及决策规则,设计多粒度粗糙集决策的步骤,得到多粒度粗糙集决策的实例应用范本,并不进行大规模的数据详尽探讨,不设定测试集对获取的规则进行测试,这将是我们的下一步的研究工作。

4 多源环境下大学生综合素质的决策分析

本节在前面理论准备的基础上用多粒度粗糙集进行大学生综合素质决策。首先利用广义多粒度粗糙集模型对多源环境下大学生综合素质决策信息系统进行约简,再根据求得的约简进行规则提取、分析,最后给出多源环境下大学生综合素质决策结论。

从表1中多源环境下大学生综合素质决策信息系统的数据出发,依各个粒度对学生进行分类,分类结果如下: $U/P_1 = \{\{x_1, x_5\}, \{x_2, x_4\}, \{x_3, x_6\}\}$; $U/P_2 = \{\{x_1, x_2\}, \{x_3\}, \{x_4, x_5\}, \{x_6\}\}$; $U/P_3 = \{\{x_1, x_2, x_6\}, \{x_3\}, \{x_4\}, \{x_5\}\}$; $U/P_4 = \{\{x_1, x_3, x_4\}, \{x_2\}, \{x_5\}, \{x_6\}\}$; $U/P_5 = \{\{x_1\}, \{x_2, x_3\}, \{x_4\}, \{x_5\}, \{x_6\}\}$; $U/d = \{\{x_1, x_4, x_5\}, \{x_2, x_3, x_6\}\}$ 。

取信息水平 $\beta = 0.7 (0.6 < \beta \leq 1)$ 和取信息水平 $\beta = 0.6 (0.5 < \beta \leq 0.6)$,信息表中的对象对各个决策类的下近似,上近似支持特征矩阵分别为:

$$\begin{array}{cccc}
 D_1, D_2 & D_1, D_2 & D_1, D_2 & D_1, D_2 \\
 \begin{array}{c} x_1 \\ x_2 \\ x_3 \\ x_4 \\ x_5 \\ x_6 \end{array} \begin{pmatrix} 0 & 0 \\ 0 & 0 \\ 0 & 1 \\ 0 & 0 \\ 1 & 0 \\ 0 & 1 \end{pmatrix} & \begin{array}{c} x_1 \\ x_2 \\ x_3 \\ x_4 \\ x_5 \\ x_6 \end{array} \begin{pmatrix} 1 & 1 \\ 1 & 1 \\ 0 & 1 \\ 1 & 1 \\ 1 & 0 \\ 0 & 1 \end{pmatrix} & \begin{array}{c} x_1 \\ x_2 \\ x_3 \\ x_4 \\ x_5 \\ x_6 \end{array} \begin{pmatrix} 0 & 0 \\ 0 & 0 \\ 0 & 1 \\ 1 & 0 \\ 1 & 0 \\ 0 & 1 \end{pmatrix} & \begin{array}{c} x_1 \\ x_2 \\ x_3 \\ x_4 \\ x_5 \\ x_6 \end{array} \begin{pmatrix} 1 & 1 \\ 1 & 1 \\ 0 & 1 \\ 1 & 0 \\ 1 & 0 \\ 0 & 1 \end{pmatrix}
 \end{array}$$

即 D_1, D_2 的上下近似分别为:

$$\begin{aligned}
 \underline{P}(D_1)_{0.7} &= \{x_5\}, \quad \bar{P}(D_1)_{0.7} = \{x_1, x_2, x_4, x_5\}; \\
 \underline{P}(D_2)_{0.7} &= \{x_3, x_6\}, \quad \bar{P}(D_2)_{0.7} = \{x_1, x_2, x_3, x_4, x_6\}; \\
 \underline{P}(D_1)_{0.6} &= \{x_4, x_5\}, \quad \bar{P}(D_1)_{0.6} = \{x_1, x_2, x_4, x_5\}; \\
 \underline{P}(D_2)_{0.6} &= \{x_3, x_6\}, \quad \bar{P}(D_2)_{0.6} = \{x_1, x_2, x_3, x_6\}.
 \end{aligned}$$

在保持分类不变的前提下,用 Matlab 计算得到决策信息系统的全部约简如下:当取信息水平 $\beta = 0.7 (0.6 < \beta \leq 1)$ 时,约简为 $\{a_1, a_2, a_3, a_4, a_5\}$;当信息水平 $\beta = 0.6 (0.5 < \beta \leq 0.6)$ 时,约简为 $\{a_1, a_2, a_3, a_5\}$ 和 $\{a_2, a_3, a_4, a_5\}$ 。

接下来,我们根据上面获取的约简进行规则提取,对规则的决策精度和规则精度给出量化结果。

当信息水平 $\beta = 0.7 (0.6 < \beta \leq 1)$ 时,系统的属性约简为 $\{a_1, a_2, a_3, a_4, a_5\}$,此时可以按照系统进

行规则提取,每个对象对应的都是一条决策规则,系统中的信息不能够由简化的属性来表示,这时候我们提取的规则是直接的,规则如下:

- $r_1: (P_1 A) \wedge (P_2 A) \wedge (P_3 C) \wedge (P_4 B) \wedge (P_5 C) \rightarrow (d G \{0.6 \downarrow\});$
- $r_2: (P_1 B) \wedge (P_2 A) \wedge (P_3 C) \wedge (P_4 C) \wedge (P_5 D) \rightarrow (d Co \{0.6 \downarrow\});$
- $r_3: (P_1 C) \wedge (P_2 D) \wedge (P_3 E) \wedge (P_4 B) \wedge (P_5 D) \rightarrow (d Co \{0.6 \downarrow\});$
- $r_4: (P_1 B) \wedge (P_2 B) \wedge (P_3 A) \wedge (P_4 B) \wedge (P_5 A) \rightarrow (d G \{0.6 \downarrow\});$
- $r_5: (P_1 A) \wedge (P_2 B) \wedge (P_3 D) \wedge (P_4 A) \wedge (P_5 B) \rightarrow (d G \{0.6 \downarrow\});$
- $r_6: (P_1 C) \wedge (P_2 C) \wedge (P_3 C) \wedge (P_4 D) \wedge (P_5 E) \rightarrow (d Co \{0.6 \downarrow\})。$

这些规则在表 1 决策信息系统中总是有效的且是唯一的,每条规则的决策精度均为 1/3,规则精度均为 1。这种情形下评定一个学生需要全部信息,对于我们简化决策过程和含有未知信息时的决策作用不大。

当信息水平 $\beta = 0.6 (0.5 < \beta \leq 0.6)$ 时,约简为 $\{a_1 \mu_2 \mu_3 \mu_5\}$ 和 $\{a_2 \mu_3 \mu_4 \mu_5\}$ 。该系统信息得到了简化,不需要全部属性,只需要部分属性的信息即可对整个系统进行知识表达。简化的决策信息系统数据表如表 2 和表 3。

表 2 约简后的多源环境下大学生综合素质决策信息系统(一)

U	P_1	P_2	P_3	P_5	d
x_1	A	A	C	C	G
x_2	B	A	C	D	Co
x_3	C	D	E	D	Co
x_4	B	B	A	A	G
x_5	A	B	D	B	G
x_6	C	C	C	E	Co

表 3 约简后的多源环境下大学生综合素质决策信息系统(二)

U	P_2	P_3	P_4	P_5	d
x_1	A	C	B	C	G
x_2	A	C	C	D	Co
x_3	D	E	B	D	Co
x_4	B	A	B	A	G
x_5	B	D	A	B	G
x_6	C	C	D	E	Co

表 2 提取决策规则如下:

- $r_1: (P_1 A) \wedge (P_2 A) \wedge (P_3 C) \wedge (P_5 C) \rightarrow (d G \{0.5 \downarrow 0.6\});$
- $r_2: (P_1 B) \wedge (P_2 A) \wedge (P_3 C) \wedge (P_5 D) \rightarrow (d Co \{0.5 \downarrow 0.6\});$
- $r_3: (P_1 C) \wedge (P_2 D) \wedge (P_3 E) \wedge (P_5 D) \rightarrow (d Co \{0.5 \downarrow 0.6\});$
- $r_4: (P_1 B) \wedge (P_2 B) \wedge (P_3 A) \wedge (P_5 A) \rightarrow (d G \{0.5 \downarrow 0.6\});$
- $r_5: (P_1 A) \wedge (P_2 B) \wedge (P_3 D) \wedge (P_5 B) \rightarrow (d G \{0.5 \downarrow 0.6\});$
- $r_6: (P_1 C) \wedge (P_2 C) \wedge (P_3 C) \wedge (P_5 E) \rightarrow (d Co \{0.5 \downarrow 0.6\})。$

表 3 提取决策规则如下:

- $r_7: (P_2 A) \wedge (P_3 C) \wedge (P_4 B) \wedge (P_5 C) \rightarrow (d G \{0.5 \downarrow 0.6\});$

$$\begin{aligned}
 r_8 &: (P_2 A) \wedge (P_3 C) \wedge (P_4 C) \wedge (P_5 D) \rightarrow (d, Co \{0.5 \rho. 6\}); \\
 r_9 &: (P_2 D) \wedge (P_3 E) \wedge (P_4 B) \wedge (P_5 D) \rightarrow (d, Co \{0.5 \rho. 6\}); \\
 r_{10} &: (P_2 B) \wedge (P_3 A) \wedge (P_4 B) \wedge (P_5 A) \rightarrow (d, G \{0.5 \rho. 6\}); \\
 r_{11} &: (P_2 B) \wedge (P_3 D) \wedge (P_4 A) \wedge (P_5 B) \rightarrow (d, G \{0.5 \rho. 6\}); \\
 r_{12} &: (P_2 C) \wedge (P_3 C) \wedge (P_4 D) \wedge (P_5 E) \rightarrow (d, Co \{0.5 \rho. 6\})。
 \end{aligned}$$

由于我们的数据规模较小,约简后的信息系统中没有重复的数据,从而得到的决策规则为12条,当数据规模比较大时,约简可以使很多重复数据合并,有效地减少决策规则的数量,提高决策精度和规则精度。

经计算可知,上面12条决策规则的决策精度均为1/6,规则精度均为1。虽然决策精度降低了,但是规则精度没有降低,约简简化了规则检验时的数据验证量,可以不用进行大规模的数据验证即可获得决策结论。

根据上面获取的规则,可从表1的决策信息系统中,得到多源环境下大学生综合素质决策的研究的结论。由决策规则的表达形式,各个决策结果如表4所示,表中(*)表示对应取值随意,即在表中指定条件下取任何值对决策结果没有影响。

表4 多源环境下大学生综合素质决策信息系统的结论

序号	决策因素	决策结果	决策正确率	信息水平 β
1	a_1 -优秀, a_2 -优秀, a_3 -一般, a_4 -良好, a_5 -一般	G	1	(0.6, 1]
2	a_1 -良好, a_2 -优秀, a_3 -一般, a_4 -一般, a_5 -差	Co	1	(0.6, 1]
3	a_1 -一般, a_2 -差, a_3 -很差, a_4 -良好, a_5 -差	Co	1	(0.6, 1]
4	a_1 -良好, a_2 -良好, a_3 -优秀, a_4 -良好, a_5 -优秀	G	1	(0.6, 1]
5	a_1 -优秀, a_2 -良好, a_3 -差, a_4 -优秀, a_5 -良好	G	1	(0.6, 1]
6	a_1 -一般, a_2 -一般, a_3 -一般, a_4 -差, a_5 -很差	Co	1	(0.6, 1]
7	a_1 -优秀, a_2 -优秀, a_3 -一般, a_4 -(*) , a_5 -一般	G	1	(0.5, 0.6]
8	a_1 -良好, a_2 -优秀, a_3 -一般, a_4 -(*) , a_5 -差	Co	1	(0.5, 0.6]
9	a_1 -一般, a_2 -差, a_3 -很差, a_4 -(*) , a_5 -差	Co	1	(0.5, 0.6]
10	a_1 -良好, a_2 -良好, a_3 -优秀, a_4 -(*) , a_5 -优秀	G	1	(0.5, 0.6]
11	a_1 -优秀, a_2 -良好, a_3 -差, a_4 -(*) , a_5 -良好	G	1	(0.5, 0.6]
12	a_1 -德育一般, a_2 -一般, a_3 -一般, a_4 -(*) , a_5 -很差	Co	1	(0.5, 0.6]
13	a_1 -(*) , a_2 -优秀, a_3 -一般, a_4 -良好, a_5 -一般	G	1	(0.5, 0.6]
14	a_1 -(*) , a_2 -优秀, a_3 -一般, a_4 -一般, a_5 -差	Co	1	(0.5, 0.6]
15	a_1 -(*) , a_2 -差, a_3 -很差, a_4 -良好, a_5 -差	Co	1	(0.5, 0.6]
16	a_1 -(*) , a_2 -良好, a_3 -优秀, a_4 -良好, a_5 -优秀	G	1	(0.5, 0.6]
17	a_1 -(*) , a_2 -良好, a_3 -差, a_4 -优秀, a_5 -良好	G	1	(0.5, 0.6]
18	a_1 -(*) , a_2 -一般, a_3 -一般, a_4 -差, a_5 -很差	Co	1	(0.5, 0.6]

5 结论

本文利用多粒度粗糙集决策方法,以多源环境下大学生综合素质决策研究为例进行研究,构建了多源环境下大学生综合素质决策信息系统;使用获取的数据表进行决策分析,得到大学生综合素质决策信

息系统的约简,通过约简提取决策规则并进行规则分析,使用决策规则给出多源环境下大学生综合素质决策的决策结论。这些研究可以为多粒度粗糙集的应用提供一种范本作为参考,为大规模的应用尝试奠定一定基础,为大学生综合素质决策提供一个新的入手点和一种新的处理方法。

参考文献:

- [1] 万远英,尹德志. 大学生综合素质层次分析评价体系及其数学模型 [J]. 西南民族大学学报(社会科学版), 2003, 21(12): 191~193.
- [2] 徐伟华,张先韬,王巧荣. 基于等价关系逻辑“与”“或”的粗糙集[J]. 数据挖掘, 2011, 1: 7~12.
- [3] 张英,冯艳芳. 基于模糊层次分析法的大学生综合素质评价[J]. 武汉理工大学学报(社会科学版) 2007, 20(5): 707~710.
- [4] 张文修,吴伟志,梁吉业,李德玉. 粗糙集理论与方法[M]. 北京:科学出版社, 2001.
- [5] 张文修,梁怡,吴伟志. 信息系统与知识发现[M]. 北京:科学出版社, 2003.
- [6] 张文修,徐伟华. 基于粒计算的认知模型[J]. 工程数学学报, 2007, 24(6): 957~971.
- [7] Pawlak Z. Rough sets: Theoretical aspects of reasoning about data [M]. Boston: Kluwer Academic Publishers, 1991.
- [8] Pawlak Z. Rough sets [J]. Communication of the ACM, 1995, 38(1): 89~95.
- [9] Pawlak Z, Skowron A. Rough sets and Boolean reasoning [J]. Information Sciences, 2007, 177(1): 41~73.
- [10] Xu W H, Wang Q R, Zhang X T. Multi-granulation fuzzy rough sets in a fuzzy tolerance approximation space [J]. International Journal of Fuzzy Systems, 2011, 13(4): 246~259.
- [11] Xu W H, Wang Q R, Zhang X T. Multi-granulation fuzzy rough set model on tolerance relations [C] // Fourth International Workshop on Advanced Computational Intelligence, 2011: 359~366.
- [12] Xu W H, Zhang X T, Wang Q R. A generalized multi-granulation rough set approach [J]. Lecture Notes in Bioinformatics, 2012, 6840: 681~689.

College Students' Comprehensive Qualities Decision Analysis in Multiple Source Environment

XU Wei-hua, LUO Shu-qun, ZHANG Xian-tao, KONG Sha-sha

(School of Mathematics and Statistics, Chongqing University of Technology, Chongqing 400054, China)

Abstract: In real life, college students' comprehensive quality decision analysis in multiple source environment is very important. To solve this problem, combining the methods of granular computing and multi-granulation rough set, college students' comprehensive quality decision information system was established on the basis of multiple source environment. Reductions of the decision system were calculated on the Matlab platform. Moreover, we extracted rules from the reductions, analyzed and checked the rules. College students' comprehensive quality decision conclusions were obtained by the rules. This paper is the application of multi-granulation rough set with theoretical and practical significance.

Key words: College Students' Comprehensive Quality; Multi-granulation Rough Set; Multiple Source Environment; Analysis to Decision-making; Information System