

基于风险最小化的多粒度三支决策模型

史进玲^a, 张全友^a, 杜根远^b

(许昌学院 a. 国际教育学院; b. 信息工程学院, 河南 许昌 461000)

摘要:为减小传统的二支决策中直接接受或拒绝决策带来的决策风险,分析问题决策的多粒度空间,研究了基于风险最小化的多粒度三支决策模型.在三支决策风险代价分析基础上,为寻求最优的粒度空间,结合不同属性特征在粒度空间中具有不同决策权重的特点,以粒化重要度和粒化决策权重为启发式信息,从多个不同的粒度层次出发,寻求风险最小的决策行动.最后针对不承诺选项中一些急需决策的现实问题,给出了基于风险控制的二支决策转化方法,并进行了具体的实例应用.

关键词:分险最小化;粒化;多粒度空间;最优决策

中图分类号:TP181

文献标志码:A

随着互联网、通信技术与云计算技术的迅猛发展,目前信息社会进入了大数据(Big Data)时代^[1].由于大数据往往呈现出大规模、多模态与快速增长等特征,因此使得传统的数据分析理论、方法与技术对海量信息的有效挖掘、处理和决策面临着可计算性、有效性与时效性等严峻挑战.在传统的二支决策中,往往只有接受或拒绝、是或非这两种选项,但是在解决很多实际问题时,强制做出接受或拒绝的决策,会付出不必要的代价或后果^[2].比如在获取的有效信息或证据不足时,在监控视频中判断一个可疑人脸图像是小偷或公司员工都显得不太合适,即接受或拒绝都比不做决策的代价风险大.为进行更优化、风险更小的决策,文献[3-4]在决策信息系统中引入了三支决策理论,即在不能准确决策时将选择不承诺选项.三支决策理论的提出和发展引起了智能信息处理与决策领域研究的广泛关注,目前在医疗诊断、石油勘探、人脸识别、金融投资等领域得到了广泛的应用^[5-9].在不承诺决策中,随着决策信息的收集和更新使准确的决策成为可能,决策的过程可以看成从粗粒度到细粒度的转化,多粒度决策粗糙集理论通过多个二元关系导出多个粒空间进行建模,从中挖掘有用的知识并形成有效决策.近几年,多粒度粗糙集理论得到不断完善和有效应用,由此本文结合多粒度空间分析最小风险化的三支决策方法.

1 三支决策理论

当传统的二支决策面临一些实际问题时,例如因所获取信息不足、掌握证据不够等因素影响而不能准确地决策时而强制进行接受或拒绝决策,将会带来一定的误决策风险,也可能会付出一些不必要的代价或后果.三支决策是在传统的二支决策中引入了不承诺选项,即当面对一些问题不能准确决策时,可以暂缓决策,待收集充分的决策信息后进行准确决策,从而减小直接决策的风险或代价,在实际应用中三支决策方法更符合最小决策代价原则.

定义 1^[10](决策信息表) 设四元组 $I = (U, A = C \cup D, V, f)$ 是一个决策信息表,其中 U 表示对象的非空有限集合,也称为论域; A 表示属性的非空有限集合, $V = \bigcup \{V_r \mid r \in A\}$, V_r 为属性 r 的值域; 设 $f: U \times A \rightarrow V$ 表示从论域 U 到值域 V_r 的信息函数,即对于 $\forall r \in A, x \in U$, 有 $f(x, r) \in V_r$. 对 $\forall S \subseteq A$, 定义等价

收稿日期:2016-06-15; 修回日期:2016-12-13.

基金项目:国家自然科学基金(U1304403); 2016年许昌市科技局基础与前沿计划研究项目.

作者简介(通信作者):史进玲(1982-),女,河南社旗人,许昌学院讲师,研究方向为粒计算、三支决策、生物信息学, E-mail:shijinling126@126.com.

关系 S 的划分 $I(S) = \{(x, y) \in U \times U \mid \forall s \in S, f(x, s) = f(y, s)\}$. 在等价关系 S 下 x 关于属性集合 S 的等价分类为 $[x]_S = \{y \mid f(x, s) = f(y, s), \forall s \in S\}$.

在问题决策中,为简化问题在对事件进行分类决策中,考虑两种分类决策 $DC = \{X, \neg X\}$,假设待分类对象 $x \in D$,对对象 x 进行分类采取的决策行为 $A = \{a_P, a_B, a_N\}$,其中 a_P, a_B, a_N 分别表示将对象决策为正域 $P(X)$ (分类为 X)、边界域 $B(X)$ (延迟决策)、负域 $N(X)$ (分类为 $\neg X$) 三种决策.

定义 2^[3](决策代价矩阵) 令 $\lambda_{PP}, \lambda_{BP}, \lambda_{NP}$ 表示 x 真实属于 X 时分别做出 a_P, a_B, a_N 三种决策所对应的风险代价值, $\lambda_{PN}, \lambda_{BN}, \lambda_{NN}$ 表示 x 真实属于 $\neg X$ 时分别做出 a_P, a_B, a_N 三种决策所对应的风险代价值,其对应的决策代价矩阵 $M_C = \begin{bmatrix} \lambda_{PP} & \lambda_{BP} & \lambda_{NP} \\ \lambda_{PN} & \lambda_{BN} & \lambda_{NN} \end{bmatrix}$.

通常在实际问题决策时,往往可以认为正确决策的代价风险 $\lambda_{PP} = \lambda_{NN} = 0$,故有简化的决策矩阵 $M_C = \begin{bmatrix} 0 & \lambda_{BP} & \lambda_{NP} \\ \lambda_{PN} & \lambda_{BN} & 0 \end{bmatrix}$.

定义 3^[3](三支决策) 对 $\forall S \subseteq A$,设 $P(X \mid [x]_S), P(\neg X \mid [x]_S)$ 分别表示等价类 $[x]_S$ 属于 $X, \neg X$ 的条件概率,根据贝叶斯最小风险决策理论和简化的代价矩阵,对对象 x 进行 a_P, a_B, a_N 三种决策行为的期望风险代价值 $R(a_P \mid [x]_S), R(a_B \mid [x]_S), R(a_N \mid [x]_S)$ 分别为:

$$\begin{aligned} R(a_P \mid [x]_S) &= \lambda_{PP}P(X \mid [x]_S) + \lambda_{PN}P(\neg X \mid [x]_S) = \lambda_{PN}P(\neg X \mid [x]_S); \\ R(a_N \mid [x]_S) &= \lambda_{NP}P(X \mid [x]_S) + \lambda_{NN}P(\neg X \mid [x]_S) = \lambda_{NP}P(X \mid [x]_S); \\ R(a_B \mid [x]_S) &= \lambda_{BP}P(X \mid [x]_S) + \lambda_{BN}P(\neg X \mid [x]_S). \end{aligned}$$

其中,条件概率 $P(X \mid Y) = \frac{|X \cap Y|}{|Y|}$. 根据贝叶斯最小化风险代价决策原则,可以得到 $C_o(x, DC) =$

$\text{Min}\{R(d \mid [x]_S), \forall d \in A_c\}, A_c = \{a_P, a_B, a_N\}$. 下面给出具体的三支决策规则:

- (1) 若 $C_o(x, DC) = R(a_P \mid [x]_S)$, 则有 $x \rightarrow a_P$.
- (2) 若 $C_o(x, DC) = R(a_B \mid [x]_S)$, 则有 $x \rightarrow a_B$.
- (3) 若 $C_o(x, DC) = R(a_N \mid [x]_S)$, 则有 $x \rightarrow a_N$.

由分析可知, $P(X \mid [x]_S) + P(\neg X \mid [x]_S) = 1$, 通常进行正确分类决策的风险代价值最小, 而延迟分类的风险代价值小于错误分类的决策风险, 因此可以做出一个合理的假设, 即 $\lambda_{PP} \leq \lambda_{BP} \leq \lambda_{NP}, \lambda_{NN} \leq \lambda_{BN} \leq \lambda_{PN}$. 故可推理得 $\alpha = \left(1 + \frac{\lambda_{BP} - \lambda_{PP}}{\lambda_{PN} - \lambda_{BN}}\right)^{-1}, \beta = \left(1 + \frac{\lambda_{NP} - \lambda_{BP}}{\lambda_{BN} - \lambda_{NN}}\right)^{-1}$, 其中阈值对 (α, β) 满足 $0 \leq \beta < \alpha \leq 1$ 由此得到简化的决策规则: (1) 若 $P(X \mid [x]_S) \geq \alpha$, 则有 $x \rightarrow a_P$. (2) 若 $\beta < P(X \mid [x]_S) < \alpha$, 则有 $x \rightarrow a_B$. (3) 若 $P(X \mid [x]_S) \leq \beta$, 则有 $x \rightarrow a_N$.

2 多粒度三支决策

通常人们在对问题分析和决策时主要从全局和局部出发, 即从不同的粒度层次将复杂问题分解为若干子问题, 通过子问题分析求解、子问题间的关联性分析等方法进行复杂问题的简化求解. 因此, 人们在问题分析与解决中可以从不同的粒度出发或通过多粒度层次之间转换构建多粒度分析模型. 在三支决策方法中, 当决策信息不足时, 无法获取准确决策的完备信息, 执行延迟决策(边界域决策), 即需要决策者进一步完善决策信息, 使下一步的准确决策成为可能, 从而降低整体决策代价风险, 由此从多粒度分析有助于进行有效的三支决策. 例如, 在人脸识别中, 发现人们所获得的目标图像随着镜头与图像之间距离逐步缩小, 使信息由粗粒度逐步变为细粒度, 即从一个图像的轮廓逐渐清晰地分辨出目标物. 由此可知, 在图像识别过程中, 面对粗粒度图像时, 对图像的决策可暂列入边界域决策行列, 将镜头逐步转向边界域, 在收集更细粒度信息的过程中人脸特征逐步清晰, 直到进行准确的识别, 获得确定的决策结果.

定义 4(粒的表示) 设 $I = (U, A = C \cup D, V, f)$ 是一个决策信息表, 令 $a \in A, v \in V$, 二元对 (a, v) 或表达式 a_v 可以作为对象 x 在关于属性 a 的信息描述, 在逻辑语义下 a_v 定义为一个原子公式, 令 $m(a_v) = \{x \mid f(x, a) = v, \forall x \in U\}$, 表示 U 上所有满足原子公式 a_v 的对象集, 称 $g = (a_v, m(a_v))$ 为决策信息系统

中的一个粒表示.

定义 5(粒度的定义) 在决策信息表 $I = (U, A = C \cup D, V, f)$ 中,对 $\forall a \in A, \forall v \in V$,则有任意的原子公式 a_v 通过有限次逻辑合取(\wedge)运算得到合取公式 φ_k, k 表示 φ_k 中包含的原子公式个数, $k | A | (| A |$ 表示集合 A 中属性个数), $m(\Phi_k) = \{x | \Phi_k = f(x, a_{1v} \wedge a_{2v} \wedge \dots \wedge a_{iv} \wedge a_{mv}), i = 1, 2, \dots, n(n = | S |),$ 对 $\forall x \in U. \forall a_i \in S, \forall v \in V$,有 x 满足合式公式 $\Phi_k\}$. 对 $\forall S \subseteq A, r = 1, 2, \dots, | S |$, 令 $g(r) = (\Phi_r, m(\Phi_r))$, 表示论域 U 中的任意 r 层粒, 由此可得到属性集合对论域的 r 层划分粒度 $GM(r) = \{m(\Phi_r), \forall a \in S, | S | = r\}$.

性质 1 假设 $| U | \geq i \geq k \geq 1$ 表示决策信息表 I 中的粒度层次, 则有 $| m(\Phi_k) | \geq | m(\Phi_i) |$.

证明 设 a_{v_1}, a_{2v} 上为决策信息表 I 中的原子公式, Φ_i, Φ_j, Φ_k 分别代表 i, j, k 层粒空间中粒对应的任意公式, $\Phi_k = a_{1v} \wedge a_{2v}, \Phi_i = a_{1v} \wedge a_{2v} \wedge \Phi_j$. 对于 $\forall x \in U$, 若 $x \in m(\Phi_i)$, 则有 $x | \approx a_{1v} \wedge a_{2v} \wedge \Phi_j$, 即有 $f(x, a_1) = v_1$ (对象 x 关于属性 a_1 的属性值) 且 $f(x, a_2) = v_2$ (对象 x 关于属性 a_2 的属性值) 且 $x | \approx \Phi_j$, 显然有 x 满足 $f(x, a_1) = v_1$ 且 $f(x, a_2) = v_2$, 即 $x | \approx a_{1v} \wedge a_{2v}$, 因此有 $x \in m(a_{1v} \wedge a_{2v}) = m(\Phi_k)$. 反之, 若对 $\forall x \in m(\Phi_k)$, 则有 $x \in m(\Phi_i)$ 或 $x \notin m(\Phi_i)$. 因此证得 $| m(\Phi_k) | \geq | m(\Phi_i) |$.

性质 2 在决策信息表 $I = (U, A = C \cup D, V, f)$ 中, $\forall S \subseteq A, 1 \leq r \leq | S |, GM(r)$ 表示决策信息表 I 中第 r 层粒度空间.

性质 3 在决策信息表中, $A \subseteq R$, 则有属性集 A 对论域的等价类划分 $I(A) = \{m(\Phi_{|A|}) | \forall G_r = (\Phi_{|A|}, m(\Phi_{|A|})) \in GM(| A |)\}$.

由定义 5 容易证得性质 2 和性质 3, 证明略.

定义 6(粒化精度) 在决策信息表 $I = (U, A = C \cup D, V, f)$ 中, $S \subseteq C$, 定义属性集 S 在论域 U 中关于决策集 D 的粒化精度为 $G_{CA}(D | S) = (| m(\Phi_{|S|}) |^2 - | m(\Phi_{|S \cup D|}) |^2) / | U |^2$.

定义 7 在决策信息表 $I = (U, A = C \cup D, V, f)$ 中, $S \subseteq C$, 属性 a 在论域 U 中关于决策集 D 的粒化重要度为 $S_i(a, S, D) = G_{CA}(D | S) / \{G_{CA}(D | S \cup \{a\})\} \times | m(\Phi_{|A|}) |$.

由性质 1 可知, $G_{CA}(D | S) / G_{CA}(D | S \cup \{a\}) \geq 1$. 如果 $S_i(a, S, D) = 1 / | U |$ 反映了在 C 中增加属性特征 a 后, 决策属性集合的粒化精度没有发生改变; 反之, 如果 $S_i(a, S, D) > 1 / | U |$ 反映了在 C 中增加属性特征 a 后, 决策属性集合的粒化精度发生了改变, 表明 a 相对于属性集合 D 具有一定的粒化重要度.

在目前的多粒度空间分析研究中, 主要考虑不同属性特征对论域的粒化数量, 而不同的粒化程度对决策的主导权重考虑较少. 但是在实际的问题分析、决策中, 不同的粒度或不同的特征属性对决策的重要性并不是等同的. 例如在医学诊断中, 医生根据患者表现的主要特征如发病突然、发热、出血、骨骼及关节病变诊断病人是否患白血病, 但是在实际诊断中, 医生往往会根据临床经验、疾病发病情况等确定出血、肝脾肿大等在白血病诊断中占有重要的决策权重.

因此, 在决策信息表 $I = (U, A = C \cup D, V, f)$ 中, 令 $S \subseteq C$, 定义属性 a 在论域 U 中关于决策集 D 的综合决策权重为 $W(a, S, D) = S_i(a, S, D) + \epsilon_a$. ϵ_a 表示粒化的决策权重, 粒化的决策权重可以根据领域专家的经验或结合问题的具体情况综合分析给出.

由于在现实世界中人们所获取的信息系统往往具有不确定性、不完备、冗余等特征, 在不同的粒度空间下, 对信息系统的描述、有效数据的挖掘或决策通常是不同的, 因此, 在决策过程中可以对不同的粒度空间分析, 选择合适的粒度层次, 使采取的决策行动的风险代价最小.

假设 $I = (U, A = C \cup D, V, f)$ 是一个决策信息表, $A = \{a_1\} \cup \{a_2\} \cup \dots \cup \{a_n\}, \forall a_j \in A, j = 1, 2, \dots, n. a, v_{ji}$ 描述了决策信息表 I 中 1 层粒度下关于任意属性特征 a_j 的一个粒表示, 其中原子公式 v_{ji} 代表对象 x_i 关于 a_j 的属性值. 该决策信息表 I 的粒表示如表 1 所示. 在表 1 中对 $\forall x_i \in U$, 有 $f(x_i, D) = d_{il} \in \{D_1, D_2, l = 1, 2\}$ 表示对象 x_i 关于属性集合 D 的取值. 以 x_i 为例, 它在 n 个属性特征下的一个粒表示为 $\Phi_n d_{il} = a_1 v_{1i} \wedge a_2 v_{2i} \wedge \dots \wedge a_n v_{ni} \wedge d_{il}$.

根据条件属性集合, 从多粒度空间出发, 不断增加粒度层次, 直至 n 层粒度, 可以得到表 1 对应的多层粒度下的粒度矩阵如表 2 所示. 为便于描述, 在表 2 中假设 $\forall a_{ij} \in A, (i, j = 1, 2, \dots, n), \forall v_{ij} \in V$ 表示 $\forall x \in U$ 在 a_{ij} 下的属性值.

表 1 粒表示下的决策信息表

U	a_1	a_2	\dots	a_j	\dots	a_n	d
x_1	$a_1 v_{11}$	$a_2 v_{21}$	\dots	$a_j v_{j1}$	\dots	$a_n v_{n1}$	$\Phi_{1n} d_{11}$
x_2	$a_1 v_{12}$	$a_2 v_{22}$	\dots	$a_j v_{j2}$	\dots	$a_n v_{n2}$	$\Phi_{2n} d_{21}$
\vdots	\vdots	\vdots	\vdots	\vdots	\vdots	\vdots	\vdots
x_i	$a_1 v_{1i}$	$a_2 v_{2i}$	\dots	$a_j v_{ji}$	\dots	$a_n v_{ni}$	$\Phi_{in} d_{i1}$
\vdots	\vdots	\vdots	\vdots	\vdots	\vdots	\vdots	\vdots
x_m	$a_1 v_{1m}$	$a_2 v_{2m}$	\dots	$a_j v_{jm}$	\dots	$a_n v_{nm}$	$\Phi_{mn} d_{m1}$

表 2 粒度矩阵

粒度	粒逻辑语义表示	粒度表示
1	$a_{i1} v_{i1}$	$m(a_{i1} v_{i1})$
2	$a_{i1} v_{i1} a_{i2} v_{i2}$	$m(a_{i1} v_{i1} a_{i2} v_{i2})$
\vdots	\vdots	\vdots
j	$a_{i1} v_{i1} a_{i2} v_{i2} \dots a_{ij} v_{ij}$	$m(a_{i1} v_{i1} a_{i2} v_{i2} \dots a_{ij} v_{ij})$
\vdots	\vdots	\vdots
n	$a_{i1} v_{i1} a_{i2} v_{i2} \dots a_{in} v_{in}$	$m(a_{i1} v_{i1} a_{i2} v_{i2} \dots a_{in} v_{in})$

对 $\forall x \in m(\Phi_i)$ 表示在 i 层粒度下对象 x 满足公式 Φ_i 的等价类, $i = 1, 2, \dots, |S|, S \subseteq C$, 为寻求合适的粒度空间, 令 $a \in S$, 按照 $W(a, S, D)$ 最大的准则, 从 i 层粒度出发, 逐步跳转到论域最细的粒度层次 ($GM(|S|)$), 定义对象 x 进行 a_P, a_B, a_N 三种决策行为的期望风险代价 $R(a_P | m(\Phi_i)), R(a_N | m(\Phi_i)), R(a_B | m(\Phi_i))$ 分别为:

$$R(a_P | m(\Phi_i)) = \lambda_{PP} P(X | m(\Phi_i)) + \lambda_{PN} P(\neg X | [m(\Phi_i)] = \lambda_{PN} P(X | m(\Phi_i));$$

$$R(a_N | m(\Phi_i)) = \lambda_{NP} P(X | m(\Phi_i)) + \lambda_{NN} P(\neg X | m(\Phi_i)) = \lambda_{NP} P(X | m(\Phi_i));$$

$$R(a_B | m(\Phi_i)) = \lambda_{BP} P(X | m(\Phi_i)) + \lambda_{BN} P(\neg X | m(\Phi_i)).$$

由性质 1 可知在决策信息表中属性特征对论域的划分粒度随着属性特征集合的增加, 粒度大小单调递减, 即粒表示下公式 Φ_i 所合取的属性个数越多, $\forall x \in m(\Phi_i)$ 所对应的 x 等价类包含对象个数越少. 另外, 决策信息表通常包含一些对决策重要的粒度属性, 因此为寻求合适的粒度层次, 需要考虑各个属性的粒度重要性. 首先在 1 层粒度下选择所有满足 $W(a, S, D) > 1/|U|$ 的属性构成粒度的主导属性特征集合, 假设这些特征集合对应的公式为 Φ_i , 则对 $\forall x \in m(\Phi_j)$, 假设 $|U| \geq i \geq j \geq k \geq 1$, 以属性的粒化决策权重为启发式信息, 从 k 层粒度到 i 层粒度, 则有 $|m(\Phi_j)| \geq |m(\Phi_i)|$, 直到 $j = i$ 满足 $x \in m(\Phi_i)$ 且 $x \notin m(\Phi_{i+1})$. 由此得到从粒度 k 到粒度 i 不同层次下对对象 x 的三支决策规则, 即对 $\forall x \in m(\Phi_j)$, 给定阈值对 (α, β) 满足 $0 \leq \beta < \alpha \leq 1$, 则有: (1) 若 $P(X | m(\Phi_j)) \geq \alpha$, 则有 $x \rightarrow a_P$. (2) 若 $\beta < P(X | m(\Phi_j)) < \alpha$, 则有 $x \rightarrow a_B$. (3) 若 $P(X | m(\Phi_j)) \leq \beta$, 则有 $x \rightarrow a_N$.

因此, 根据不同粒度层次的风险代价, 选择最优的粒度空间. 在任意的粒度 j 层中, 对象 x 进行 a_P 决策行为的期望风险最小 ($P(X | m(\Phi_j)) \geq \alpha$), 则有 $x \in P(X)$; 若对象 x 进行 a_B 决策行为的期望风险最小 ($\beta < P(X | m(\Phi_j)) < \alpha$), 则有 $x \in B(X)$; 如果对象 x 进行 a_N 决策行为的期望风险最小 ($P(X | m(\Phi_j)) \leq \beta$), 则有 $x \in N(X)$.

在多粒度三支决策中, 如果 $x \in P(X)$ 或 $x \in N(X)$, 则进行确定的决策. 如果 $x \in B(X)$, 此时将要进行延迟决策. 但是, 通常在面对一些急需决策的问题时, 延迟决策可能也会带来一定的风险^[11]. 例如在对急性疾病的诊断治疗或进行股票决策选择等需要立即决策的情况下, 延迟决策可能需要承担暂时不进行决策处理的风险, 往往会影响疾病的治疗或错过好的股票决策时机. 在上述决策形势下, 需要对延迟决策进行风险分析. 图 1 给出基于风险最小化的多粒度三支决策模型.

假设此时处于延迟决策状态 (执行 a_B 决策), 令 λ'_{NB} 表示 x 在 $B(X)$ 中进行 a_N 决策行为的期望风险代价值 $VaR(\text{Value at Risk})$, 分析在延迟决策中随时间推移决策特征的变化、新特征的出现等情况采用风险代价值

VaR 计算方法进行综合分析 λ'_{NB} , 假设 δ 为实际期望的风险代价阈值, $P(X | x)$ 为正确决策的概率, $P(\neg X | x)$ 为错误决策的概率, $P(X | x) + P(\neg X | x) = 1$. 则有:

对 x 进行拒绝决策的风险 $R(a_N) = \lambda'_{NB}P(\neg X | x)$, 如果 $R(a_N | x) \leq \delta$, 则有 $x \in N(X)$; 反之, 若 $R(a_P | x) > \delta$, 则有 $x \in P(X)$.

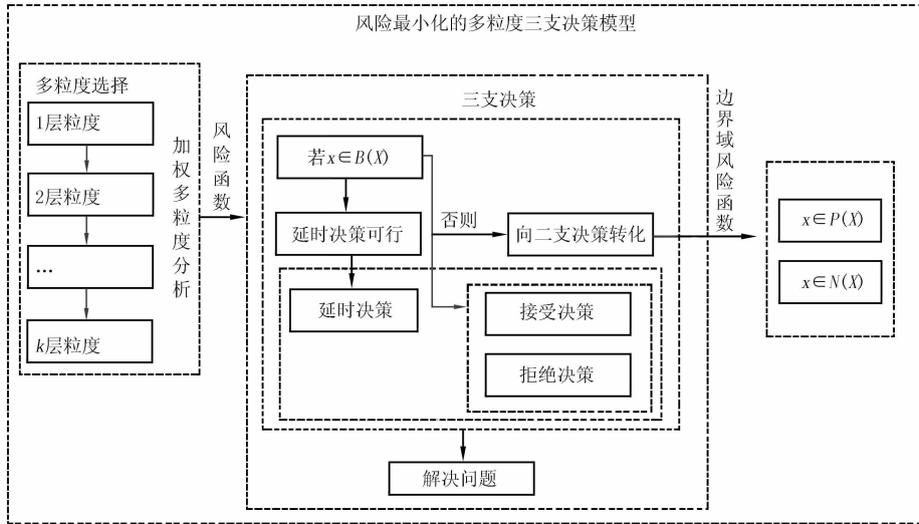


图1 基于风险最小化的多粒度三支决策模型

3 基于风险最小化的多粒度三支决策

结合基于风险最小化的多粒度三支决策模型, 多粒度三支决策主要包括多粒度分析和三支决策两个阶段, 针对待决策问题从单个知识粒度层次出发, 以决策权重为启发式信息, 寻求最优的粒度层次使决策行动风险最小, 决策行动主要包括接受决策、拒绝决策和不承诺决策. 但是针对现实世界中面对一些急需决策的问题, 如果在决策中执行不承诺决策, 可能会错过决策的最佳时期, 因此在基于风险最小化的多粒度三支决策中增加了决策转换阶段, 通过分析决策随时间、决策粒度改变等因素进行综合决策, 最终选择最小风险的决策粒度.

3.1 基于风险最小化的多粒度三支决策实现的步骤

输入: 决策信息系统 $I = (U, A = C \cup D, V, f)$, $S = \Phi$, 结合具体决策形势, 定义阈值对 (α, β) 和期望风险阈值 δ , 对对象 x 进行三支决策.

输出: 寻求最佳粒度层次, 选择最小风险代价的决策行动.

步骤 1 计算每个属性相对于决策集 D 的重要性, 对 $\forall a \in C$, 若 $W(a, C, D) > 1/|U|$, 则令 $S = S \cup a$, $C = C/a$, 即将属性 a 从 C 中删除.

步骤 2 对 $\forall D_k \in U/D, l = 1, 2$. 在 $|S|$ 阶粒度 $GM(S)$ 中, 设定粒度层次 $k = |S|$. 假设三支决策行动集合 $A_c = \{a_P, a_B, a_N\}$, 分析对象 x 在 k 层粒度下关于决策属性集合 D_k 的代价矩阵 M , 确定 (α, β) 的取值.

步骤 3 从 C 中选择 ϵ_a 最大的属性特征 a , 若 $(x \notin m(\Phi_{|S \cup a|}))$, 转步骤 5.

步骤 4 令 $S = S \cup a, C = C/a, |S| = j$, 构建 $k + 1$ 阶粒度空间, $j = k + 1$, 计算 $P(X | m(\Phi_j))$.

步骤 5 在公式 $\Phi_j (i \geq j \geq k)$ 对应的粒度层次下, 比较 $P(X | m(\Phi_j))$. 若 $P(X | m(\Phi_j)) \geq \alpha$, 或若 $\beta < P(X | m(\Phi_j)) < \alpha$, 进行 a_P 决策或 a_N 决策, 转步骤 7. 若 $P(X | m(\Phi_j)) \leq \beta$, 进行 a_B 决策, 若实际问题允许延迟决策执行, 转步骤 7.

步骤 6 对 a_B 进行二支决策转化, 测算风险阈值 δ 的期望取值, 计算延迟决策中进行 a_N 行动的风险代价值 $R(a_N) = \lambda'_{NB}P(\neg X | x)a_B$, 根据阈值 δ 选择最小的风险决策行动, 进行拒绝或接受决策.

步骤 7 输出结果.

3.2 应用实例

张先生和王先生计划在市区繁华商业区内购置门面房1套,房屋购买单价为22 000元/m²,面积为50 m².他们计划将预期购买的门面房进行租赁,目前他们的可用资金为40万元,每月的经济收入总额为10 000元,请根据他们的购买能力和经济收入情况给出具体的投资决策.

采用住房商业贷款方式,针对现有购买能力和房屋总额,按照等额本息还款法(240月),计算目前的购房明细如表3所示.

表3 房屋购买明细

购房清单	价格/万元
房屋总额	110
现有资金	40
利息总额	39.946 6
还款总额	109.946 6
月供金额	0.458 1

对购买房屋所涉及的属性集合(S)包括贷款总额(特征 a)、还款情况(特征 b)、房屋使用年限(特征 c)、预期的租赁价格浮动(特征 d)、租赁年限(特征 e)及租赁行情的变化(特征 f)等分析多粒度决策的风险代价 $C_o(x, DC)$,其中 x 指张先生和王先生购买房屋事件, $DC = \{\text{购买、不购买、延迟决策}\}$.结合该商业区市场行情,确定最小的风险代价阈值对($\alpha = 0.6, \beta = 0.2$).如果 $P(X | m(\Phi_j)) \geq 0.6$,进行购买房屋决策对经济生活影响较小,预期可以收益.如果风险代价阈值 $C_o(x, a_B) \leq 0.2$,进行购买房屋决策对经济生活影响较大,收益可能暂时不太明显或收益很小.如果 $0.2 \leq C_o(x, a_N) \leq 0.6$,决策处于购买与不购买之间,等待进一步的决策.

在房屋购买决策中,对每个特征属性计算其粒化重要度和粒化决策权重,分析贷款总额、租赁价格浮动、租赁年限为主导的粒度特征,其综合决策权重分别为 $W(a, S, D) = 0.389, W(d, S, D) = 0.452, W(e, S, D) = 0.25$,并分析各个特征的粒化权重,从剩余的属性集合中以粒化权重 ϵ_a 的大小顺序 $f > b > c$,分别从 $\{a, d, e\}, \{a, d, e, f\}, \{a, d, e, f, b\}, \{a, d, e, f, b, c\}$ 多个粒度层次综合分析进行三支决策的风险代价,最终确定在粒度层次 $\{a, d, e, b, f\}$ 中进行延迟决策的风险最小, $0.2 < P(X | m(\Phi_j)) < 0.6$,执行决策为“暂时不能准确决策”.但是,由于商业街市场行情、租赁价格、顾客购买能力等不确定因素存在,如果暂时不购买,当到能进行接受或决策时房屋可能已经被出售,而且房屋租赁价格或房价出现提升趋势,那么王先生和张先生的暂缓决策可能会错过投资收益的机会,因此在此情况下暂时不决策可能意味着拒绝决策.为使决策风险代价最小,分析在延迟决策状态下影响风险代价的主导因素为:预期租赁价格为100元/m²,每次租赁年限在3年以上,商业街租赁行情满足非递减趋势,银行利率有可能下降,有一定的还贷压力,房屋价格可能上涨.给定置信水平为95%,投资总额为110万元,实际期望的风险代价阈值 $\delta = 150$ 万元.基于上述因素测算收益率和收益率分布标准差, $P(X | x) = 0.6$.采用风险价值模型(Var)的计算公式得到 $\lambda'_{PB} = 585.25$ 万元,计算 $R(a_N) = \lambda'_{NB} P(\neg X | x) = 585.25 \text{ 万元} \times 40\% = 234.1 \text{ 万元}$,因此,此时应选择购买房屋的决策.

4 结 论

在传统的二支决策中引入不承诺选项,可以避免因人们在不能准确决策时强制决策所带来的代价或后果,考虑到不同的粒度具有不同的决策主导信息,本文针对决策信息表讨论了知识的粒化,以粒度的划分数量和粒化决策权重为主要指标,在不同的粒度层次之间转换以寻求最优的粒度空间,为使决策风险代价最小,探讨了风险控制和多粒度三支决策方法,为实际问题的决策和风险的处理提供了一种新的研究视角.

参 考 文 献

- [1] 王珊,王会举,覃雄派,等.架构大数据:挑战、现状与展望[J].计算机学报,2011,34(10):1741-1752.
- [2] 刘盾.三支决策与粒计算[M].北京:科学出版社,2013.

- [3] Yao Y. Three-way decisions with probabilistic rough sets[J]. *Information Sciences*, 2010, 180(3):341-353.
- [4] Jia X Y, Li W W, Shang L, et al. An Adaptive Learning Parameters Algorithm in Three-Way Decision-Theoretic Rough Set Model[J]. *Acta Electronica Sinica*, 2011, 39(11):2520-2525.
- [5] Liu S, Liu X. An extended three-way decision and its application in member selection[J]. *Journal of Intelligent & Fuzzy Systems*, 2015, 28(5):2095-2106.
- [6] 张里博, 李华雄, 周献中, 等. 人脸识别中的多粒度代价敏感三支决策[J]. *山东大学学报(理学版)*, 2014(8):48-57.
- [7] Liang D, Pedrycz W, Liu D, et al. Three-way decisions based on decision-theoretic rough sets under linguistic assessment with the aid of group decision making[J]. *Applied Soft Computing*, 2015, 29(C):256-269.
- [8] Zhang L, Li H, Zhou X, et al. Cost-Sensitive Sequential Three-Way Decision for Face Recognition[M]. Berlin: Springer, 2014: 375-383.
- [9] 薛占熬, 朱泰隆, 薛天宇, 等. 基于三支决策理论的条件属性权重构造方法[J]. *计算机科学*, 2015, 42(8):265-268.
- [10] Zhao Y, Wong S K M, Yao Y Y. Attribute reduction in decision-theoretic rough set models[J]. *Information Sciences*, 2013, 178(17): 3356 - 3373.
- [11] 李丽红, 李言, 刘保相. 三支决策中不承诺决策的转化代价与风险控制[J]. *计算机科学*, 2016, 43(1):77-80.

Multi-granularity Three-way Decision Model based on Minimum Risks

Shi Jinling^a, Zhang Quanyou^a, Du Genyuan^b

(a. International School of Education; b. College of Information Engineering, Xuchang University, Xuchang 461000, China)

Abstract: A multi-granularity three-way decision model with minimum risk was proposed to remedy the inevitable risks which arise from traditional two-way decision when accepting or refusing directly, by analyzing question decisions in multi-granularity spaces. Firstly, an effective granularity selection criterion was presented with granulating significance and decision weight in considering of different attribute effects for granularity spaces in different granularity level. Then, the particular three-way decision method with minimum risk was described. Lastly, a two-way transformed decision was studied when certain specific decision making was extremely needed and an application example was given to verify the decision efficiency.

Keywords: minimum risk; granulating; multi-granularity space; optimal decision

[责任编辑 陈留院]