文章编号：1001-0920(2009)11-1615-05

一种基于多决策类的贝叶斯粗糙集模型

韩 敏[1]，张俊杰[1]，彭 飞[2]，肖正宇[3]
(1.大连理工大学 电子与信息工程学院，辽宁 大连 116024；2.本钢板材股份有限公司
炼钢厂，辽宁 本溪 117021；3.北京金自天正智能控制股份有限公司，北京 100070)

摘 要：针对传统贝叶斯粗糙集理论只能处理二决策类的不足，提出一种基于多决策类的贝叶斯粗糙集.在此基础上定义一个衡量条件属性对决策属性影响程度的 γ 依赖度函数，并证明了该函数具有随条件属性的增加而单调递增的性质.最后基于 γ 依赖度函数的单调特性，提出一种确定属性权重的算法.以某钢厂 150 t 转炉的实际生产数据为例，仿真结果表明了模型的有效性和实用性.

关键词：贝叶斯粗糙集；多决策类；属性权重；γ 依赖度函数
中图分类号：TP18 **文献标识码**：A

Bayesian rough set model based on multiple decision classes

HAN Min[1], *ZHANG Jun-jie*[1], *PENG Fei*[2], *XIAO Zheng-yu*[3]
(1. School of Electronic and Information Engineering, Dalian University of Technology, Dalian 116024, China; 2. Steelmaking Plant of Benxi Steel Sheet Company Limited, Benxi 117021, China; 3. Beijing ARITIME Intelligent Control Company Limited, Beijing 100070, China. Correspondent: HAN Min, E-mail: minhan@dlut.edu.cn)

Abstract: For the limitation that traditional Bayesian rough set model theory can only deal with the situation of two decision classes, a Bayesian rough set model based on multiple decision classes is proposed, which can deal with the problem of multiple decision classes. On this condition, a γ dependency function is defined to evaluate the condition attributes significance to decision attributes, and is proved that the function is monotonic increase with condition attributes. Finally, an algorithm to compute attribute weight is proposed based on the monotonic property of γ dependency function. The simulation result of the model using the practical data from a steel plant's 150 ton converter shows the effectiveness and practicality of this model.

Key words: Bayesian rough set; Multiple decision classes; Attribute weight; γ dependency function

1 引 言

粗糙集理论(RST)是由波兰学者 Pawlak 于 1982 年提出的，仅利用数据本身提供的信息，无需任何先验的专家知识，因而在实际的决策中得到较为广泛的应用.许多学者在经典的 Pawlak 粗糙集理论的基础上进行改进，主要有以下两方面：一是将模糊集与粗糙集相结合，解决数据的连续性或不确定性等问题[1-4]，如 Dubois 和 Prade 等；二是对不可分辨关系进行扩充[5-8]，如 Ziarko 和 Slezak 等.

经典的 Pawlak 粗糙集理论[9-11]在处理分类关系时过于严格，易将有用的信息也剔出掉，不利于决策分析.Ziarko 等将概率理论与粗糙集理论相结合，提出了变精度粗糙集和贝叶斯粗糙集，在判断集合间隶属关系时，引入了一个不确定度，能够更好地描述集合间的依赖关系，弥补了经典粗糙集的不足.但是，Ziarko 等提出的理论只能处理决策属性是二分类的情况，对于决策属性是多分类的情况则无能为力，从而限制了它的应用范围.

针对上述问题，本文将 Slezak 和 Ziarko 提出的贝叶斯粗糙集扩展到多决策类的情况，提出一个属性间依赖度函数.证明了该依赖度函数具有随属性的增加单调递增的性质，并利用这个性质计算属性权重.采用某钢厂实际生产数据来确定吹氧量各影响因素的权重，仿真结果表明了所提出模型的有效性和实用性.

2 传统贝叶斯粗糙集

收稿日期：2008-12-31；**修回日期**：2009-03-26.
基金项目：国家 863 计划项目(2007AA04Z158)；国家自然科学基金项目(60674073).
作者简介：韩敏(1959—)，女，辽宁大连人，教授，博士，从事复杂工业系统建模与控制、智能技术及优化算法等研究；张俊杰(1984—)，男，吉林延边人，硕士生，从事复杂工业系统建模的研究.

Slezak 和 Ziarko 提出的传统贝叶斯粗糙集模型(BRSM)[6],通过比较引入条件后目标集发生的条件概率与引入条件前目标集发生的先验概率之间的关系来分析该条件对决策的影响. 整个过程完全依赖数据本身的性质,并不需要限制参数,从而避免了主观因素的引入.

设 U 为有限论域,R 为 U 上的一个等价关系. U 按等价关系 R 进行不可分辨划分 $U/\mathrm{IND}(R)$,得到 R 的基本集 $E=U/\mathrm{IND}(R)=\{E_1,E_2,\cdots,E_n\}$. 其中 E_i 为按 $\mathrm{IND}(R)$ 划分得到的等价类,对于任意 $i\neq j$,满足 $E_i\cap E_j=\phi$,且 $U=\bigcup_{i=1}^{n}E_i$.

P 为定义在 U 的子集类构成的 σ 代数上的概率测度,对于 U 的任意非空子集 $X\subset U$,都满足 $0<P(X)<1$. 设 X 和 Y 为 U 上非空子集,$P(X\mid Y)$ 表示 Y 发生的条件下 X 发生的概率,可以反映 Y 对 X 的影响程度.

贝叶斯粗糙集模型对目标集 $X\subset U$ 的正域 $\mathrm{POS}^*(X)$,负域 $\mathrm{NEG}^*(X)$ 和边界域 $\mathrm{BNR}^*(X)$ 的定义如下:

$$\mathrm{POS}^*(X)=\bigcup\{E_i\in E:P(X\mid E_i)>P(X)\},\tag{1}$$

$$\mathrm{NEG}^*(X)=\bigcup\{E_i\in E:P(X\mid E_i)>P(X)\},\tag{2}$$

$$\mathrm{BNR}^*(X)=\bigcup\{E_i\in E:P(X\mid E_i)>P(X)\}.\tag{3}$$

3 改进的贝叶斯粗糙集

传统的贝叶斯粗糙集可以处理二决策的情况,但不能用于处理多决策类的情况. 设 $S=(U,R)$ 为一个决策表. 其中:U 为非空的有限论域,是所有样本的集合;$R=C\cup D$,且 $C\cap D=\phi$,是非空的等价关系有限集,这里指所有属性的集合. $C=\{c_1,c_2,\cdots,c_s\}$ 为条件属性集,$D=\{d\}$ 为决策属性. U 按等价关系 C 进行不可分辨划分 $U/\mathrm{IND}(C)$,得到 C 的基本集

$$E=U/\mathrm{IND}(C)=\{E_1,E_2,\cdots,E_n\}.$$

在决策属性为二分类 $U/\mathrm{IND}(D)=\{X,\neg X\}$,即有两个目标集的情况下,根据 Slezak 和 Ziarko 提出的传统贝叶斯粗糙集,当 $E_i\in\mathrm{POS}^*(X)$ 时,有

$$P(\neg X\mid E_i)=1-P(X\mid E_i)<1-P(X)=P(\neg X),$$

则 $E_i\notin\mathrm{POS}^*(\neg X)$. 因此,任何 E_i 都不会同时划分到 $\mathrm{POS}^*(X)$ 和 $\mathrm{POS}^*(\neg X)$ 中. 而在决策属性是多分类 $U/\mathrm{IND}(D)=\{X_j\mid j=1,2\cdots,m\}$,即有多个目标集的情况下,其中 m 为按决策属性划分的类别数,X_j 为第 j 个决策类,如果仍按传统贝叶斯粗糙集划分各区域,则可能产生同一个 E_i 被划分到不同决策类的正域中的情况,即在同一个条件下,却产生了多个决策. 这与经典 Pawlak 粗糙集理论相悖.

为了弥补以上不足,本文对原始贝叶斯粗糙集进行改进. 引入一个评价函数[6]

$$g(X\mid Y)=\frac{P(X\mid Y)-P(X)}{P(X)},\tag{4}$$

并用该函数衡量 Y 对 X 的影响程度. 各决策类 X_j 的正域 $\mathrm{POS}^M(X_j)$,负域 $\mathrm{NEG}^M(X_j)$ 和边界域 $\mathrm{BNR}^M(X_j)$ 定义为

$$\mathrm{POS}^M(X_j)=\bigcup\{E_i\in E:g(X_j\mid E_i)=\max(g(X_1\mid E_i),\cdots,g(X_m\mid E_i))>0\},\tag{5}$$

$$\mathrm{NEG}^M(X_j)=\bigcup\{E_i\in E:g(X_j\mid E_i)=\min(g(X_1\mid E_i),\cdots,g(X_m\mid E_i))<0\},\tag{6}$$

$$\mathrm{BNR}^M(X_j)=\bigcup\{E_i\in E:\min(g(X_1\mid E_i),\cdots,g(X_m\mid E_i))\leqslant g(X_j\mid E_i)\leqslant\max(g(X_1\mid E_i),\cdots,g(X_m\mid E_i))\}.\tag{7}$$

并定义一个决策 D 的全局边界域 $\mathrm{GBNR}^M(D)$ 为

$$\mathrm{GBNR}^M(D)=\bigcup\{E_i\in E:g(X_1\mid E_i)=g(X_2\mid E_i)=\cdots=g(X_m\mid E_i)=0\}.\tag{8}$$

全局边界域是一种特殊的区域,对于某些决策表可能不存在.

根据改进后的贝叶斯粗糙集的定义,可以得到以下命题:

命题 1 $\max(g(X_1\mid E_i),\cdots,g(X_m\mid E_i))\geqslant 0$,$\min(g(X_1\mid E_i),\cdots,g(X_m\mid E_i))\leqslant 0$,当且仅当 $E_i\in\mathrm{GBNR}^M(D)$ 时等号成立,即

$$\max(g(X_1\mid E_i),\cdots,g(X_m\mid E_i))=\min(g(X_1\mid E_i),\cdots,g(X_m\mid E_i))=0.$$

证明 因为

$$\sum_{j=1}^{m}(P(X_j\mid E_i)-P(X_j))=\sum_{j=1}^{m}(P(X_j\mid E_i)-\sum_{j=1}^{m}(P(X_j)=1-1=0,$$

故必然存在 u 和 v,使得 $P(X_u\mid E_i)-P(X_u)\geqslant 0$,$P(X_v\mid E_i)-P(X_v)\leqslant 0$. $g(X_j\mid E_i)$ 是在 $P(X_j\mid E_i)-P(X_j)$ 的基础上,除以一个大于零的分母,不影响最终的符号. 因此,$\max(g(X_1\mid E_i),\cdots,g(X_m\mid E_i))\geqslant 0$ 且 $\min(g(X_1\mid E_i),\cdots,g(X_m\mid E_i))\leqslant 0$. 当 $E_i\in\mathrm{GBNR}^M(D)$ 时,对于任意 X_j 都满足 $g(X_j\mid E_i)=0$,得到 $\max(g(X_1\mid E_i),\cdots,g(X_m\mid E_i))=\min(g(X_1\mid E_i),\cdots,g(X_m\mid E_i))=0$.

命题 2 $\exists E_i\in\mathrm{GBNR}^M(D)$,则 E_i 同时属于每个决策类 X_j 的边界域 $\mathrm{BNR}^M(X_j)$.

命题 3 $\forall E_i\notin\mathrm{GBNR}^M(D)$,都划分到某个决

策类的正域，且同一个 E_i 不会划分到多个决策类的正域；$\forall E_i \notin \mathrm{GBNR}^M(D)$，不会划分到任何一个决策类的正域.

证明　由命题 1 可得，对于 $\forall E_i \notin \mathrm{GBNR}^M(D)$，都存在 u 使得 $g(X_u \mid E_i) = \max(g(X_1 \mid E_i), \cdots, g(X_m \mid E_i)) > 0$，并将 E_i 划分到 $\mathrm{POS}^M(X_u)$. 由式(5)的定义可知，E_i 不会划分到其他决策类的正域. 这个性质也符合实际要求，对于一种情况最终只能对应一种决策. 当 $E_i \in \mathrm{GBNR}^M(D)$ 时，$\max(g(X_1 \mid E_i), \cdots, g(X_m \mid E_i)) = 0$，而正域的定义要求其大于 0，所以这时 E_i 不会划分到任何一个决策类的正域. □

为了衡量条件属性集 C 对决策属性 D 的影响程度，将 C 对每个决策类 X_j 的影响程度求和，定义为 D 对 C 依赖度 $\gamma(D \mid C)$，即

$$\gamma(D \mid C) = \sum_{j=1}^{m} \sum_{E_i \in \mathrm{POS}^M(X_j)} P(E_i) g(X_j \mid E_i). \quad (9)$$

命题 4　$\gamma(D \mid C)$ 可改写为

$$\gamma(D \mid C) = \sum_{i=1}^{n} \max(P(E_i \mid X_1), \cdots, P(E_i \mid X_m)) - 1. \quad (10)$$

证明　由贝叶斯定理可得

$$P(E_i) g(X_j \mid E_i) = P(E_i) \frac{\dfrac{P(E_i \mid X_j) P(X_j)}{P(E_i)} - P(X_j)}{P(X_j)} = P(E_i \mid X_j) - P(E_i).$$

对于 $\forall E_i \notin \mathrm{GBNR}^M(D)$，满足 $\max(g(X_1 \mid E_i), \cdots, g(X_m \mid E_i)) = 0$，则

$$\gamma(D \mid C) = \sum_{j=1}^{m} \sum_{E_i \in \mathrm{POS}^M(X_j)} P(E_i) g(X_j \mid E_i) + \sum_{E_i \in \mathrm{GBNR}^M(D)} P(E_i) \max(g(X_1 \mid E_i), \cdots, g(X_m \mid E_i)).$$

再根据定义(5)，可得

$$\gamma(D \mid C) = \sum_{j=1}^{m} \sum_{E_i \in \mathrm{POS}^M(X_j)} P(E_i) \max(g(X_1 \mid E_i), \cdots, g(X_m \mid E_i)) + \sum_{E_i \in \mathrm{GBNR}^M(D)} P(E_i) \max(g(X_1 \mid E_i), \cdots, g(X_m \mid E_i)).$$

由命题 3 可知，对于 $\forall E_i \in E$，不是唯一地划分到某个决策类的正域，就是划分到 $\mathrm{GBNR}^M(D)$. 由此可得

$$\gamma(D \mid C) = \sum_{E_i \in \mathrm{GBNR}^M(D)} P(E_i) \max(g(X_1 \mid E_i), \cdots, g(X_m \mid E_i)) + \sum_{E_i \in \mathrm{GBNR}^M(D)} P(E_i) \max(g(X_1 \mid E_i), \cdots, g(X_m \mid E_i)) = \sum_{i=1}^{n} P(E_i) \max(g(X_1 \mid E_i), \cdots, g(X_m \mid E_i)) = \sum_{i=1}^{n} \max((P(E_i \mid X_1) - P(E_i)), \cdots, (P(E_i \mid X_m) - P(E_i))) = \sum_{i=1}^{n} \max(P(E_i \mid X_1), \cdots, P(E_i \mid X_m)) - 1.$$

命题 5　对于 $\forall B \subseteq C$，满足如下关系：

$$\gamma(D \mid B) \leqslant \gamma(D \mid C). \quad (11)$$

等号成立的条件是任意决策类 X_j 满足

$$\mathrm{POS}_C^M(X_j) \subseteq \mathrm{POS}_B^M(X_j).$$

证明　设 $U/\mathrm{IND}(B) = \{F_1, F_2, \cdots, F_k\}$，可以看出每个 F_j 都是由一些 E_i 组成的，即

$$F_j = \bigcup_{i=1}^{n} \{E_i : E_i \subseteq F_j\}.$$

要证明命题 5，只需证明下面不等式成立：

$$\max(P(F_j \mid X_1), \cdots, P(F_j \mid X_m)) \leqslant \sum_{E_i \subseteq F_j} \max(P(E_i \mid X_1), \cdots, P(E_i \mid X_m)).$$

设

$$P(F_j \mid X_r) = \max(P(F_j \mid X_1), \cdots, P(F_j \mid X_m)) = \sum_{E_i \subseteq F_j} P(E_i \mid X_r) \leqslant \sum_{E_i \subseteq F_j} \max(P(E_i \mid X_r), \cdots, P(E_i \mid X_m)),$$

因此 $\gamma(D \mid B) \leqslant \gamma(D \mid C)$ 成立. 换言之，γ 依赖度函数随着条件属性的增加是单调递增的. 当 $\forall E_i \subseteq F_j$，$P(E_i \mid X_r) = \max(P(E_i \mid X_1), \cdots, P(E_i \mid X_m))$ 成立时，$\gamma(D \mid B) = \gamma(D \mid C)$. 可以看出，它是在条件属性 C 的各决策类正域划分的基础上，可能将一些 $E_i \in \mathrm{GBNR}^M(D)$ 也划分到某些决策类的正域. 即对于任意决策类 X_j，满足

$$\mathrm{POS}_C^M(X_j) \subseteq \mathrm{POS}_B^M(X_j).$$ □

利用命题 5 可以进行属性约简、确定属性权重等. 下面具体介绍属性权重的确定过程.

4　属性权重确定算法

利用本文提出的 γ 依赖度函数随条件属性的增加单调递增的性质，可以确定各条件属性的权重. 在计算属性权重的过程中，完全依赖数据本身的特征，不需引入人为的因素，使结果更加客观. 因为粗糙集理论无法处理连续的属性数据，因此引入模糊 c 均值聚类算法，将连续的属性离散化.

4.1　模糊 c 均值聚类算法

模糊 c 均值(FCM)聚类算法是基于目标函数的

聚类算法，目标函数[12]

$$J_m = \sum_{i=1}^{c}\sum_{k=1}^{n}(u_{ik})^m(d_{ik})^2. \tag{12}$$

其中：c 为分类个数；m 为加权指数；n 为样本个数；u_{ik} 为第 k 个样本对于第 i 类的隶属度；d_{ik} 为第 k 个样本与第 i 类的典型样本之间的失真度，通常用两个矢量间的距离来衡量. 最终使目标函数达到最小值，并作为最终的分类结果.

FCM 算法步骤如下：

Step1：确定聚类类别数 c，$2 \leqslant c < n$，n 为数据个数. 设定迭代停止阈值 ε，初始化聚类原型模式 $P^{(0)}$，设置迭代计数器 $b = 0$.

Step2：计算出新的隶属度矩阵 $U^{(b+1)}$. 矩阵中各元素 $u_{ik}^{(b+1)}$ 可通过下式求出：

$$u_{ik}^{(b+1)} = \frac{1}{\sum\limits_{j=1}^{c}\left(\dfrac{d_{ik}^{(b)}}{d_{jk}^{(b)}}\right)^{\frac{2}{m-1}}}. \tag{13}$$

如果存在 i 和 k，使得 $d_{ik}^{(b)} = 0$，则 $u_{ik}^{(b+1)} = 1$，且对于 $l \neq k$，$u_{il}^{(b+1)} = 0$. 可以证明各元素对应于各分类的隶属度之和为 1.

Step3：更新聚类原型模式 $P^{(b+1)}$，矩阵中各分类的新聚类中心可通过下式计算：

$$p_i^{(b+1)} = \frac{\sum\limits_{k=1}^{n}(u_{ik}^{(b+1)})^m \cdot x_k}{\sum\limits_{k=1}^{n}(u_{ik}^{(b+1)})^m}, \tag{14}$$

式中 x_k 为第 k 条样本数据.

Step4：如果 $\| P^{(b)} - P^{(b+1)} \| < \varepsilon$，则算法停止；否则，令 $b = b+1$，转 Step2 继续执行.

4.2 属性权重确定过程

决策表中各条件属性的权重求解过程如下：

Step1：对决策表中的连续属性离散化. 当属性连续时，采用 FCM 聚类算法按该属性对数据聚类，初步选择分类数的范围，并将取各分类数时的分类结果代入有效性评价函数[13]，以使该函数值最小的分类数作为最佳的分类数. 有效性评价函数具体形式如下：

$$V_{XB} = \frac{\sum\limits_{i=1}^{c}\sum\limits_{j=1}^{N}\mu_{ij}^2 \| x_j - v_i \|^2}{n(\min\limits_{i\neq k}\| v_i - v_k \|^2)}. \tag{15}$$

其中：c 为分类个数，N 为数据个数，x_j 为第 j 条数据，v_i 为第 i 个聚类中心，μ_{ij} 为第 j 条数据对应第 i 类的隶属度. 最后将各条数据划分到其隶属度最大的那个分类.

Step2：根据式(9)计算决策属性 D 对所有条件属性 C 的依赖度 $\gamma(D \mid C)$，以及缺失某一条件属性 c_i 的依赖度 $\gamma(D \mid C - \{c_i\})$.

Step3：计算各条件属性 c_i 的重要度 $\mathrm{SIG}(c_i)$. 因为 γ 依赖度函数具有单调性，所以可通过下式进行计算：

$$\mathrm{SIG}(c_i) = \gamma(D \mid C) - \gamma(D \mid C - \{c_i\}). \tag{16}$$

Step4：对各重要度进行归一化处理，得到各条件属性权重

$$w(c_i) = \frac{\mathrm{SIG}(c_i)}{\sum\limits_{j=1}^{s}\mathrm{SIG}(c_j)}. \tag{17}$$

5 计算实例

表 1 为某钢厂 150 t 转炉实际生产数据. 被控制量为吹氧量，是连续属性. 铁水碳含量、铁水硅含量和目标钢种是影响吹氧量的 3 个因素，其中铁水碳含量和铁水硅含量是连续属性，目标钢种是离散属性.

表 1 转炉生产数据

序号	条件属性			决策属性
	铁水碳含量 (0.01%)	铁水硅含量 (0.01%)	钢 种	吹氧量 /m³
1	450	100	DC01	8692
2	438	50	Q235B	7564
3	487	58	DC01	8931
⋮	⋮	⋮	⋮	⋮
18	478	80	Q235B	8814
19	430	68	Q235B	8432
20	456	119	DC01	8781

设条件属性集 $C = \{c_1, c_2, c_3\}$，分别代表铁水碳含量、铁水硅含量和钢种，决策属性 D 为吹氧量. 采用模糊 c 均值聚类算法分别对连续条件属性 c_1，c_2 和决策属性 D 进行离散化，分类数选取范围为 2～4 类. 根据式(15)的有效性评价函数确定的最佳分类数分别为 3，3，4. 对于离散属性 c_3，标记 Q235B 为 1，标记 DC01 为 2. 离散化后的结果如表 2 所示.

按决策属性 D 进行划分，可得

$U/\mathrm{IND}(D) =$

$\{\{1,3,10,12,18,20\},\{2,11,13,14,16\},$

$\{4,8,17,19\},\{5,6,7,9,15\}\}$；

按所有条件属性 C 进行划分，可得

$U/\mathrm{IND}(C) =$

$\{\{1,20\},\{2\},\{3,9\},\{4,6,13,15,17\},\{5,7\},$

$\{8\},\{10\},\{11,19\},\{12,18\},\{14\},\{16\}\}$；

删除条件属性 c_1 后进行划分，可得

$U/\mathrm{IND}(C - \{c_1\}) =$

$\{\{1,20\},\{2,16\},\{3,4,6,9,13,14,15,$

表 2　离散化后的决策表

序号	条件属性			决策属性
	c_1	c_2	c_3	d
1	3	1	2	4
2	2	3	1	1
3	1	3	2	4
4	3	3	2	3
5	3	2	1	2
6	3	3	2	2
7	3	2	1	2
8	3	1	1	3
9	1	3	2	2
10	3	2	2	4
11	2	2	1	1
12	1	2	1	4
13	3	3	2	1
14	2	3	2	1
15	3	3	2	2
16	3	3	1	1
17	3	3	2	3
18	1	2	1	4
19	2	2	1	3
20	3	1	2	4

17},{5,7,11,12,18,19},{8},{10}};

删除条件属性 c_2 后进行划分，可得

$U/\mathrm{IND}(C-\{c_2\}) =$

{{1,4,6,10,13,15,17,20},{2,11,19},

{3,9},{5,7,8,16},{12,18},{14}};

删除条件属性 c_3 后进行划分，可得

$U/\mathrm{IND}(C-\{c_3\}) =$

{{1,8,20},{2,14},{3,9},{4,6,13,15,

16,17},{5,7,10},{11,19},{12,18}}.

根据式(9)计算决策属性 D 对各条件属性集的依赖度为

$$\gamma(D\mid C)=2.033,$$
$$\gamma(D\mid (C-\{c_1\}))=1.150,$$
$$\gamma(D\mid (C-\{c_2\}))=1.033,$$
$$\gamma(D\mid (C-\{c_3\}))=1.417.$$

进而求得各条件属性的重要度为

$$\mathrm{SIG}(c_1)=0.883,$$
$$\mathrm{SIG}(c_2)=1.000,$$
$$\mathrm{SIG}(c_3)=0.616.$$

归一化得到各条件属性的权重为 $w_1=0.353$，$w_2=0.400$，$w_3=0.247$. 结果表明，铁水硅含量对吹氧量的影响最大，铁水碳含量次之，钢种影响最小，这完全符合实际生产经验.

6　结　　论

本文提出的改进的贝叶斯粗糙集，将传统贝叶斯粗糙集扩展到了多决策类的情况，大大拓宽了应用范围. 采用引入概率理论定义的 γ 依赖度函数，可以很好地评价条件属性对决策属性的影响程度，而且可利用 γ 依赖度函数随条件属性的增加单调递增的性质，计算属性的权重.

参考文献(References)

[1] Dubois D, Prade H. Rough fuzzy sets and fuzzy rough sets[J]. Int J of General Systems, 1990, 17(2/3): 191-209.

[2] Pal S K, Mitra P. Case generation using rough sets with fuzzy representation[J]. IEEE Trans on Knowledge and Data Engineering, 2004, 16(3): 293-300.

[3] Jensen R, Shen Q. Fuzzy-rough attribute reduction with application to web categorization[J]. Fuzzy Sets and Systems, 2004, 141(3): 469-485.

[4] Greco S, Inuiguchi M, Slowinski R. Fuzzy rough sets and multiple-premise gradual decision rules[J]. Int J of Approximate Reasoning, 2006, 41(2): 179-211.

[5] Ziarko W. Variable precision rough set model[J]. J of Computer and System Sciences,1993, 46(1): 39-59.

[6] Slezak D, Ziarko W. The investigation of the Bayesian rough set model[J]. Int J of Approximate Reasoning, 2005, 40(1): 81-91.

[7] Ziarko W. Probabilistic approach to rough sets[J]. Int J of Approximate Reasoning, 2008, 49(2): 272-284.

[8] Greco S, Matarazzo B, Slowinski R. Parameterized rough set model using rough membership and Bayesian confirmation measures [J]. Int J of Approximate Reasoning, 2008, 49(2): 285-300.

[9] Pawlak Z, Skowron A. Rudiments of rough sets[J]. Information Sciences, 2007, 177(1): 3-27.

[10] Pawlak Z, Skowron A. Rough sets and Boolean reasoning[J]. Information Sciences, 2007, 177(1): 41-73.

[11]张文修，吴伟志，梁吉业，等. 粗糙集理论与方法[M]. 北京：科学出版社，2001.
(Zhang W X, Wu W Z, Liang J Y, et al. Rough set theory and method [M]. Beijing: Science Press, 2001.)

[12] 高新波. 模糊聚类分析及其应用[M]. 西安：西安电子科技大学出版社，2004.
(Gao X B. Fuzzy cluster analysis and its applications [M]. Xi'an: Xidian University Press, 2004.)

[13] Xie X L, Beni G. A validity measure for fuzzy clustering[J]. IEEE Trans on Pattern Analysis and Machine Intelligence, 1991, 13(8): 841-847.