Hans Journal of Data Mining 数据挖掘, 2011, 1, 7-12
doi:10.4236/hjdm.2011.11002 Published Online July 2011 (http://www.hanspub.org/journal/hjdm/)

Rough Set Based on Logical AND and OR of Equivalence Relations*

Weihua Xu, Xiantao Zhang, Qiaorong Wang
School of Mathematics and Statistics, Chongqing University of Technology, Chongqing
Email: chxuwh@gmail.com
Received: Jul. 20th, 2011; revised: Aug. 4th, 2011; accepted: Aug. 6th, 2011.

Abstract: We popularize the classical rough set model in this paper and study rough set in the view of logical operation of equivalence relations. The logical AND rough set model and the logical OR rough set model are proposed on the basis of the operations logical AND and logical OR of equivalence relations. Furthermore, the connection between the logical AND rough set model and Pawlak's classical rough set model is illustrated. Important properties are discussed in depth and several measures are defined in OR-RS. An example is employed to explain OR-RS.

Keywords: Rough Set; Logical Operation of Equivalence Relations; AND-RS; OR-RS

基于等价关系逻辑“与”“或”的粗糙集*

徐伟华，张先韬，王巧荣
重庆理工大学数学与统计学院，重庆
Email: chxuwh@gmail.com
收稿日期：2011 年 7 月 20 日；修回日期：2011 年 8 月 4 日；录用日期：2011 年 8 月 6 日

摘　要：本文从关系逻辑运算的角度研究粗糙集，对经典的粗糙集进行了推广。对多个等价关系进行逻辑“与”和逻辑“或”运算，提出了逻辑“与”粗糙集模型和逻辑“或”粗糙集模型。说明了逻辑“与”粗糙集模型和 Pawlak 经典粗糙集的关系，并详细研究了逻辑“或”粗糙集模型的重要性质，定义了逻辑“或”粗糙集模型中的若干度量，举例验证了该模型。

关键词：粗糙集；等价关系逻辑运算；逻辑“与”粗糙集；逻辑“或”粗糙集

1. 引言

粗糙集理论是波兰数学家 Z. Pawlak 于 1982 年提出的一种数据分析理论[1]，是一种处理不确定性信息和知识的数学工具，也是一种有效的软计算工具，其主要思想是在保持分类能力不变的前提下，通过知识约简，得出问题的分类规则[2-4]。该理论在决策与分析、数据挖掘、过程控制、模式识别、机器学习与知识发现等方面已被成功应用并得到拓展，在医学、化学、材料学、地理科学、管理学和金融学等其他学科取得了显著的成果[5-7]。

Pawlak 经典粗糙集模型是以等价关系为基础的，在实际应用中受到不同程度的限制。根据应用的需求，研究者们对经典的粗糙集模型进行了推广，提出了多种粗糙集模型[5-10]。在 Pawlak 粗糙集模型中，属性集对应的不可区分关系是一个等价关系，Pawlak 粗糙集模型就是基于一个等价关系的粗糙集模型．我们将其进行推广，从关系逻辑运算的角度，由多个等价关系进行逻辑运算来定义粗糙集的上、下近似。本文从 Pawlak 粗糙集模型的定义出发，将近似算子依等价关系的逻辑“与”“或”运算进行推广，提出了基于等

*基金项目：中国博士后基金(20100481331)，中国国家自然科学基金(11001227，71071124)。

价关系逻辑“与”“或”运算的粗糙集模型，研究了模型中近似算子的重要性质，进一步探讨了逻辑“或”粗糙集模型中概念的粗糙性度量、分类精度和分类质量的度量，并用实例验证了有关结论。

2. Pawlak 粗糙集

为本文后续内容的叙述方便，本节给出一些后文用到的基本知识。

设 $S=(U,AT,V,f)$ 是一个信息系统[7]，其中：

$U=\{u_1,u_2,\cdots,u_n\}$ 为论域，

$AT=\{a_1,a_2,\cdots,a_m\}$ 为属性集，

$V=\bigcup_{a\in AT}V_a$，V_a 是属性 a 的值域，

$f:U\times AT\to V$ 为信息函数，$\forall u\in U,a\in AT$，有 $f(u,a)\in V_a$。

若 $AT=C\bigcup D$，$C\bigcap D=\varnothing$，C 为条件属性集，D 为决策属性集，则称 S 为目标信息系统，也称为决策表[6]。

R_A 为属性子集 $A\subseteq AT$ 对应的等价关系，对论域构成的划分记为 $U/R_A=\left\{[u]_{R_A}\middle|u\in U\right\}$，其中 $[u]_{R_A}=\left\{v\in U\middle|(u,v)\in R_A\right\}$ 为对象 u 在 R_A 下的等价类[6,7]。

定义 2.1[6] 设 $S=(U,AT,V,f)$ 为信息系统，$A\subseteq AT$，$X\subseteq U$。X 关于 A 的下近似和上近似分别定义为：

$$\underline{R_A}(X)=\bigcup\left\{[u]_{R_A}\in U/R_A\middle|[u]_{R_A}\subseteq X\right\},$$

$$\overline{R_A}(X)=\bigcup\left\{[u]_{R_A}\in U/R_A\middle|[u]_{R_A}\cap X\neq\varnothing\right\},$$

当 $\underline{R_A}(X)=\overline{R_A}(X)$ 时称 X 为 R_A 可定义集；当 $\underline{R_A}(X)\neq\overline{R_A}(X)$ 时称 X 为 R_A 粗糙集。

有关粗糙集的更多内容读者可以参考文献[6,7]，这里不再详述。

3. 基于等价关系逻辑“与”运算的粗糙集

基于上一节内容，根据经典粗糙集模型中近似集的定义形式，我们首先给出基于等价关系逻辑“与”运算的粗糙集模型。

记 R_{A_1},R_{A_2} 分别为属性子集 $A_1,A_2\subseteq AT$ 对应的等价关系，A_1,A_2 的逻辑“与”为不可区分关系 $ind(A_1,A_2)$，记 $ind(A_1,A_2)=R_{A_1\cup A_2}=\mathrm{AND}(A_1,A_2)$。

定义 3.1 设 $S=(U,AT,V,f)$ 为信息系统，$A_1,A_2\subseteq AT$，$X\subseteq U$。X 关于 A_1,A_2 逻辑“与”下近似和上近似定义为：

$$\underline{R_{A_1\cup A_2}}(X)=\bigcup_{u\in U}\left\{\bigcap_i[u]_{R_{A_i}}\middle|\wedge\left([u]_{R_{A_i}}\subseteq X\right),i\in\{1,2\}\right\}$$

$$\overline{R_{A_1\cup A_2}}(X)=\bigcup_{u\in U}\left\{\bigcup_i[u]_{R_{A_2}}\middle|[u]_{R_{A_i}}\cap X\neq\varnothing,i\in\{1,2\}\right\}$$

当 $\underline{R_{A_1\cup A_2}}(X)=\overline{R_{A_1\cup A_2}}(X)$ 时，称 X 关于 A_1,A_2 为逻辑“与”可定义集；当 $\underline{R_{A_1\cup A_2}}(X)\neq\overline{R_{A_1\cup A_2}}(X)$ 时，称 X 关于 A_1,A_2 为逻辑“与”粗糙集；该粗糙集模型是逻辑“与”粗糙集模型，记为 AND-RS。

对于 $A=\left\{A_i\subseteq AT\middle|i\in I\right\}$，$I=\{1,2,\cdots,m\}$，任意子集 $X\subseteq U$ 的逻辑“与”近似集可以从上面定义在等价关系的数量上进行推广。

定义 3.2 设 $S=(U,AT,V,f)$ 为信息系统，$A=\left\{A_i\subseteq AT\middle|i\in I\right\}$，$I=\{1,2,\cdots,m\}$，$X\subseteq U$ 关于 A 逻辑“与”的下近似和上近似定义为：

$$\underline{R_{\cup A_i}}(X)=\bigcup_{u\in U}\left\{\bigcap_i[u]_{R_{A_i}}\middle|\wedge\left([u]_{R_{A_i}}\subseteq X\right),i\in I\right\},$$

$$\overline{R_{\cup A_i}}(X)=\bigcup_{u\in U}\left\{\bigcup_i[u]_{R_{A_i}}\middle|[u]_{R_{A_i}}\cap X\neq\varnothing,i\in I\right\},$$

当 $\underline{R_{\cup A_i}}(X)=\overline{R_{\cup A_i}}(X)$ 时称 X 关于 A 为逻辑“与”可定义集；当 $\underline{R_{\cup A_i}}(X)\neq\overline{R_{\cup A_i}}(X)$ 时称 X 关于 A 为逻辑“与”粗糙集。

由 $ind(A_1,A_2)=R_{A_1\cup A_2}$，有 $[u]_{ind(A_1,A_2)}=[u]_{R_{A_1\cup A_2}}=[u]_{R_{A_1}}\cap[u]_{R_{A_2}}$，根据定义 3.1 可以看到，在有多个条件属性的信息系统中，Pawlak 粗糙集模型使用多个等价关系的不可区分关系对概念作近似，就是通过多个等价关系逻辑“与”运算实现的，此时逻辑“与”粗糙集模型和 Pawlak 粗糙集模型是一致的．由此，本文中不再详细探讨 AND-RS 与定理 2.1 对应的性质，仅给出 Pawlak 粗糙集模型与 AND-RS 的关系。

定理 3.1 设 $S=(U,AT,V,f)$ 为信息系统，$A_1,A_2\subseteq AT$，$X\subseteq U$。Pawlak 粗糙集模型与 AND-RS 有如下关系：

$$\underline{R_{A_1\cup A_2}}(X)=\underline{R_{A_1}}(X)\cap\underline{R_{A_2}}(X),$$

$$\overline{R_{A_1\cup A_2}}(X)=\overline{R_{A_1}}(X)\cup\overline{R_{A_2}}(X)。$$

证明：由定义 3.1 和定义 2.1 即证。

4. 基于等价关系逻辑“或”运算的粗糙集

根据上一节的内容，对应的我们可以通过等价关系逻辑“或”运算来近似概念并定义对应的粗糙集模型，用各个属性或者属性子集分别来近似刻画一个概念，为知识发现和模式识别提供更好的的处理方法，下面我们给出详细叙述。

对属性子集 $A_1,A_2\subseteq AT$，我们用 $R_{A_1+A_2}$ 表示属性子集 A_1,A_2 对应的等价关系进行逻辑“或”运算的结果，记 $R_{A_1+A_2}=\mathrm{OR}(A_1,A_2)$。

定义 4.1 设 $S=(U,AT,V,f)$ 为信息系统，$A_1,A_2\subseteq AT$。$X\subseteq U$ 关于 A_1,A_2 逻辑“或”的下近似和上近似定义为：

$$\underline{R_{A_1+A_2}}(X)=\bigcup_{u\in U}\left\{\bigcup_i [u]_{R_{A_i}}\ \middle|\ [u]_{R_{A_i}}\subseteq X, i\in\{1,2\}\right\},$$

$$\overline{R_{A_1+A_2}}(X)=\bigcup_{u\in U}\left\{\bigcap_i [u]_{R_{A_i}}\ \middle|\ \wedge\left([u]_{R_{A_i}}\cap X\neq\varnothing\right), i\in\{1,2\}\right\},$$

当 $\underline{R_{A_1+A_2}}(X)=\overline{R_{A_1+A_2}}(X)$ 时，称 X 关于 A_1,A_2 为逻辑“或”可定义集；当 $\underline{R_{A_1+A_2}}(X)\neq\overline{R_{A_1+A_2}}(X)$ 时，称 X 关于 A_1,A_2 为逻辑“或”粗糙集；该粗糙集模型是逻辑“或”粗糙集模型，记为 OR-RS。

令 $A=\{A_i\subseteq AT\,|\,i\in I\}$，$I=\{1,2,\cdots,m\}$，我们记 m 个属性子集对应的等价关系进行逻辑“或”运算为 $+A_i=A_1+A_2+\cdots+A_m$，任意子集 $X\subseteq U$ 关于 A 的逻辑“或”上、下近似也可以在等价关系的数量上进行推广。

定义 4.2 设 $S=(U,AT,V,f)$ 为信息系统，$A=\{A_i\subseteq AT\,|\,i\in I\}$，$I=\{1,2,\cdots,m\}$。$X\subseteq U$ 关于 A 的逻辑“或”下近似和上近似定义为：

$$\underline{R_{+A_i}}(X)=\bigcup_{u\in U}\left\{\bigcup_i [u]_{R_{A_i}}\ \middle|\ [u]_{R_{A_i}}\subseteq X, i\in I\right\},$$

$$\underline{R_{+A_i}}(X)=\bigcup_{u\in U}\left\{\bigcup_i [u]_{R_{A_i}}\ \middle|\ [u]_{R_{A_i}}\subseteq X, i\in I\right\},$$

当且仅当 $\underline{R_{+A_i}}(X)=\overline{R_{+A_i}}(X)$ 时 X 关于 A 为逻辑“或”可定义集；当且仅当 $\underline{R_{+A_i}}(X)\neq\overline{R_{+A_i}}(X)$ 时 X 关于 A 为逻辑“或”粗糙集。

由定义 2.3 可以得到，$\forall u\in U$，当 $[u]_{A_1}=[u]_{A_2}$ 时逻辑“或”粗糙集模型退化为 Pawlak 粗糙集模型，此时即有 $A_1=A_2$。下面我们研究 $A_1\neq A_2$ 时，OR-RS 和 Pawlak 粗糙集模型、AND-RS 的关系。

定理 4.1 OR-RS 和 Pawlak 粗糙集模型、AND-RS 有如下关系：

(a) $\underline{R_{A_1+A_2}}(X)=\underline{R_{A_1}}(X)\cup\underline{R_{A_2}}(X)$，
$\overline{R_{A_1+A_2}}(X)=\overline{R_{A_1}}(X)\cap\overline{R_{A_2}}(X)$；

(b) $\underline{R_{A_1+A_2}}(X)\subseteq\underline{R_{A_1\cup A_2}}(X)$，
$\overline{R_{A_1+A_2}}(X)\supseteq\overline{R_{A_1\cup A_2}}(X)$。

证明：(a) 由定义 2.1 和定义 4.1，有

$$\begin{aligned}\forall u\in\underline{R_{A_1}}(X)&\Leftrightarrow[u]_{R_{A_1}}\subseteq X\Rightarrow u\in\underline{R_{A_1+A_2}}(X)\\&\Rightarrow\underline{R_{A_1}}(X)\subseteq\underline{R_{A_1+A_2}}(X)\end{aligned};$$

同理有 $\underline{R_{A_2}}(X)\subseteq\underline{R_{A_1+A_2}}(X)$。

$$\begin{aligned}\forall u\in\underline{R_{A_1+A_2}}(X)&\Leftrightarrow\left([u]_{R_{A_1}}\subseteq X\right)\vee\left([u]_{R_{A_2}}\subseteq X\right)\\&\Rightarrow\left(u\in\underline{R_{A_1}}(X)\right)\vee\left(u\in\underline{R_{A_2}}(X)\right)\\&\Rightarrow u\in\underline{R_{A_1}}(X)\cup\underline{R_{A_2}}(X)\\&\Rightarrow\underline{R_{A_1+A_2}}(X)\subseteq\underline{R_{A_1}}(X)\cup\underline{R_{A_2}}(X);\end{aligned}$$

由上证得 $\underline{R_{A_1+A_2}}(X)=\underline{R_{A_1}}(X)\cup\underline{R_{A_2}}(X)$，同理可证 $\overline{R_{A_1+A_2}}(X)=\overline{R_{A_1}}(X)\cap\overline{R_{A_2}}(X)$。

(b) 由定义 2.1 和定义 3.1 我们有 $\forall u\in U$，$[u]_{R_{A_1\cup A_2}}=[u]_{R_{A_1}}\cap[u]_{R_{A_2}}\subseteq[u]_{R_{A_i}}$，$i=1,2$。

从而 $\forall u\in\underline{R_{A_1+A_2}}(X)$ 有 $[u]_{R_{A_1}}\subseteq X$ 或 $[u]_{R_{A_2}}\subseteq X$，即有 $[u]_{R_{A_1\cup A_2}}\subseteq[u]_{R_{A_i}}\subseteq X$ $(i=1\,or\,2)$，故 $u\in\underline{R_{A_1\cup A_2}}(X)$，从而即证得 $\underline{R_{A_1+A_2}}(X)\subseteq\underline{R_{A_1\cup A_2}}(X)$ 成立；

$$\forall u\in\overline{R_{A_1\cup A_2}}(X)$$

有 $[u]_{R_{A_i}}\cap X\supseteq[u]_{R_{A_1\cup A_2}}\cap X\neq\varnothing$ $(i=1,2)$，即有 $[u]_{R_{A_1}}\cap X\neq\varnothing$ 且 $[u]_{R_{A_2}}\cap X\neq\varnothing$，故 $u\in\overline{R_{A_1+A_2}}(X)$，即证得 $\overline{R_{A_1+A_2}}(X)\supseteq\overline{R_{A_1\cup A_2}}(X)$ 成立。

对比 Pawlak 粗糙集模型，OR-RS 成立如下性质。

定理 4.2 设 $S=(U,AT,V,f)$ 为信息系统，$A_1,A_2\subseteq AT$，$X\subseteq U$。X 关于 A_1,A_2 逻辑“或”的下近似和上近似满足以下性质：

(O1) $\underline{R_{A_1+A_2}}(X)=\sim\overline{R_{A_1+A_2}}(\sim X)$，
$\overline{R_{A_1+A_2}}(X)=\sim\underline{R_{A_1+A_2}}(\sim X)$；

(O2) $\underline{R_{A_1+A_2}}(X)\subseteq X\subseteq\overline{R_{A_1+A_2}}(X)$；

(O3) $\underline{R_{A_1+A_2}}(\varnothing)=\overline{R_{A_1+A_2}}(\varnothing)=\varnothing$，
$\underline{R_{A_1+A_2}}(U)=\overline{R_{A_1+A_2}}(U)=U$；

(O4) $\underline{R_{A_1+A_2}}(X)=\underline{R_{A_2+A_1}}(X)$，

$\overline{R_{A_1+A_2}}(X)=\overline{R_{A_2+A_1}}(X)$。

证明：(O1)由定理 4.1，有

$$\begin{aligned}\overline{R_{A_1+A_2}}(\sim X)&=\overline{R_{A_1}}(\sim X)\cap\overline{R_{A_2}}(\sim X)\\&=\left(\sim\underline{R_{A_1}}(X)\right)\cap\left(\sim\underline{R_{A_2}}(X)\right)\\&=\sim\left(\underline{R_{A_1}}(X)\cup\underline{R_{A_2}}(X)\right)\\&=\sim\underline{R_{A_1+A_2}}(X)\end{aligned};$$

$\underline{R_{A_1+A_2}}(X)=\sim\overline{R_{A_1+A_2}}(\sim X)$ 得证，用 $\sim X$ 代替 X 即证得 $\overline{R_{A_1+A_2}}(X)=\sim\underline{R_{A_1+A_2}}(\sim X)$。

(O1)由定义 4.2，有

$$\begin{aligned}\forall u\in\underline{R_{A_1+A_2}}(X)&\Leftrightarrow\left([u]_{R_{A_1}}\subseteq X\right)\vee\left([u]_{R_{A_2}}\subseteq X\right)\\&\Rightarrow u\in X\Rightarrow\underline{R_{A_1+A_2}}(X)\subseteq X\end{aligned};$$

根据(O1)对偶性可证得 $X\subseteq\overline{R_{A_1+A_2}}(X)$，性质 $\underline{R_{A_1+A_2}}(X)\subseteq X\subseteq\overline{R_{A_1+A_2}}(X)$ 即证。

(O3)由定义 4.1，有

$\underline{R_{A_1+A_2}}(\varnothing)=\cup\left\{\cup[u]_{R_{A_i}}\middle|[u]_{R_{A_i}}\subseteq\varnothing\right\}=\cup\varnothing=\varnothing$，

$\overline{R_{A_1+A_2}}(\varnothing)=\cup\left\{\cap[u]_{R_{A_i}}\middle|[u]_{R_{A_i}}\cap\varnothing\neq\varnothing\right\}=\varnothing$，

$\underline{R_{A_1+A_2}}(\varnothing)=\overline{R_{A_1+A_2}}(\varnothing)=\varnothing$ 即证；

由(O1)对偶性，得

$\underline{R_{A_1+A_2}}(U)=\sim\overline{R_{A_1+A_2}}(\sim U)=\sim\overline{R_{A_1+A_2}}(\varnothing)=U$，

$\overline{R_{A_1+A_2}}(U)=\sim\underline{R_{A_1+A_2}}(\sim U)=\sim\underline{R_{A_1+A_2}}(\varnothing)=U$，

$\underline{R_{A_1+A_2}}(U)=\overline{R_{A_1+A_2}}(U)=U$ 即证。

(O4)由定义 4.1 直接可证。

定理 4.3 设 $S=(U,AT,V,f)$ 为信息系统，$A_1,A_2\subseteq AT$，$X,Y\subseteq U$．X,Y 关于 A_1,A_2 逻辑“或”的下近似和上近似满足以下性质：

(O5) $\underline{R_{A_1+A_2}}(X\cap Y)\subseteq\underline{R_{A_1+A_2}}(X)\cap\underline{R_{A_1+A_2}}(Y)$，

$\overline{R_{A_1+A_2}}(X\cup Y)\supseteq\overline{R_{A_1+A_2}}(X)\cup\overline{R_{A_1+A_2}}(Y)$；

(O6) $\underline{R_{A_1+A_2}}(X\cup Y)\supseteq\underline{R_{A_1+A_2}}(X)\cup\underline{R_{A_1+A_2}}(Y)$，

$\overline{R_{A_1+A_2}}(X\cap Y)\subseteq\overline{R_{A_1+A_2}}(X)\cap\overline{R_{A_1+A_2}}(Y)$；

(O7) $X\subseteq Y\Rightarrow\underline{R_{A_1+A_2}}(X)\subseteq\underline{R_{A_1+A_2}}(Y)$，

$\overline{R_{A_1+A_2}}(X)\subseteq\overline{R_{A_1+A_2}}(Y)$。

证明：(O5) $\forall u\in\underline{R_{A_1+A_2}}(X\cap Y)\Rightarrow[u]_{A_1}\subseteq X\cap Y$ 或 $[u]_{A_2}\subseteq X\cap Y$

$\Rightarrow$(I) $\left([u]_{A_1}\subseteq X\right)\vee\left([u]_{A_2}\subseteq X\right)\Rightarrow u\in\underline{R_{A_1+A_2}}(X)$，i.e.，$\underline{R_{A_1+A_2}}(X\cap Y)\subseteq\underline{R_{A_1+A_2}}(X)$；

(II) $\left([u]_{A_1}\subseteq Y\right)\vee\left([u]_{A_2}\subseteq Y\right)\Rightarrow u\in\underline{R_{A_1+A_2}}(Y)$，i.e.，$\underline{R_{A_1+A_2}}(X\cap Y)\subseteq\underline{R_{A_1+A_2}}(Y)$；

证得 $\underline{R_{A_1+A_2}}(X\cap Y)\subseteq\underline{R_{A_1+A_2}}(X)\cap\underline{R_{A_1+A_2}}(Y)$ 成立。由(O1)对偶性，可证得另一性质成立。

(O6) $\forall u\in\underline{R_{A_1+A_2}}(X)\cup\underline{R_{A_1+A_2}}(Y)\Rightarrow u\in\underline{R_{A_1+A_2}}(X)$

或 $u\in\underline{R_{A_1+A_2}}(Y)\Rightarrow\left([u]_{A_1}\subseteq X\right)\vee\left([u]_{A_2}\subseteq X\right)$

或 $\left([u]_{A_1}\subseteq Y\right)\vee\left([u]_{A_2}\subseteq Y\right)$，i.e.，$[u]_{A_1}\subseteq X\cup Y$

或 $[u]_{A_2}\subseteq X\cup Y\Rightarrow u\in\underline{R_{A_1+A_2}}(X\cup Y)$；

证得 $\underline{R_{A_1+A_2}}(X\cup Y)\supseteq\underline{R_{A_1+A_2}}(X)\cup\underline{R_{A_1+A_2}}(Y)$ 成立。由(O1)对偶性，可证得另一性质成立。

(O7)由定义 4.1 直接可证。

AND-RS 与 Pawlak 粗糙集模型在本质上和形式上都是一致的，其性质与 Pawlak 粗糙集完全一致．定义 4.2 的近似集只是等价关系在数量上的平行推广，对应性质全部成立，这里不再详细叙述。

5. OR-RS 中的粗糙性度量、近似分类精度和近似分类质量

概念 $X\subseteq U$ 在 OR-RS 中产生粗糙性的原因也是由于边界域的存在，OR-RS 中 X 关于 A_1,A_2 逻辑“或”的边界域定义为：

$$Bn_{R_{A_1+A_2}}(X)=\overline{R_{A_1+A_2}}(X)/\underline{R_{A_1+A_2}}(X).$$

边界域越大，概念越粗糙，OR-RS 中概念的粗糙性度量定义如下。

定义 5.1 设 $S=(U,AT,V,f)$ 为信息系统，$A_1,A_2\subseteq AT$，$X\subseteq U$。OR-RS 中 X 关于 A_1,A_2 逻辑“或”的近似精度和粗糙度分别定义为：

$$\alpha_{R_{A_1+A_2}}(X)=\left|\underline{R_{A_1+A_2}}(X)\right|/\left|\overline{R_{A_1+A_2}}(X)\right|,$$

$$\rho_{R_{A_1+A_2}}(X)=1-\left|\underline{R_{A_1+A_2}}(X)\right|/\left|\overline{R_{A_1+A_2}}(X)\right|。$$

OR-RS 与 Pawlak 粗糙集模型、AND-RS 中的近似精度和粗糙度满足下面关系。

定理 5.1 设 $S=(U,AT,V,f)$ 为信息系统，$A_1,A_2\subseteq AT$，$X\subseteq U$。OR-RS 和 Pawlak 粗糙集模型、

AND-RS 中近似精度和粗糙度关系如下：

$$\alpha_{A_i}(X)\le\alpha_{R_{A_1+A_2}}(X)\le\alpha_{R_{A_1\cup A_2}}(X),$$

$$\rho_{A_i}(X)\ge\rho_{R_{A_1+A_2}}(X)\ge\rho_{R_{A_1\cup A_2}}(X),$$

其中 $i=1,2$。

证明：由定理 4.1 和定义 5.1 即证。

两个概念 X,Y 粗糙性可以通过粗糙度或精度的数值大小来比较。对于一个已知的可信知识，同样可以以其为标准来衡量 A_1,A_2 逻辑“或”近似的好坏。令 $\mathcal{F}=\{Y_1,Y_2,\cdots,Y_n\}$ 是论域 U 的一个已知的划分，是可信的专家方案或经验结论，该知识独立于 $A_1,A_2\subseteq AT$ 对应的知识，$\mathcal{F}$ 关于 A_1,A_2 逻辑“或”的近似集分别为：

$$\underline{R_{A_1+A_2}}(\mathcal{F})=\left\{\underline{R_{A_1+A_2}}(Y_k)\middle|Y_k\in\mathcal{F}\right\};$$

$$\overline{R_{A_1+A_2}}(\mathcal{F})=\left\{\overline{R_{A_1+A_2}}(Y_k)\middle|Y_k\in\mathcal{F}\right\}。$$

下面定义近似分类精度和近似分类质量来度量 A_1,A_2 逻辑“或”近似分类的好坏。

定义 5.2 A_1,A_2 逻辑“或”根据 $\mathcal{F}$ 的近似分类精度定义为：

$$\tau_{R_{A_1+A_2}}(\mathcal{F})=\sum_{k=1}^{n}\left|\underline{R_{A_1+A_2}}(Y_k)\right|\Big/\sum_{k=1}^{n}\left|\overline{R_{A_1+A_2}}(Y_k)\right|;$$

A_1,A_2 逻辑“或”根据 $\mathcal{F}$ 的近似分类质量定义为：

$$\upsilon_{R_{A_1+A_2}}(\mathcal{F})=\sum_{k=1}^{n}\left|\underline{R_{A_1+A_2}}(Y_k)\right|\Big/|U|。$$

近似分类精度 $\tau_{R_{A_1+A_2}}$ 表达的是当使用 OR-RS 对概念进行刻画时，概念可能的描述中精确描述的百分比；近似分类质量 $\upsilon_{R_{A_1+A_2}}$ 表示的是应用 OR-RS 能确切的划入 $\mathcal{F}$ 类的对象占所有对象的百分比。

在目标信息系统 S 中，设 $D=\{d\}$，即单个决策的情形，记 $U/d=\{D_1,D_2,\cdots,D_t\}$，在 OR-RS 中定义知识的依赖度如下。

定义 5.3 设 $S=(U,C\cup\{d\},V,f)$ 为目标信息系统，$A_1,A_2\subseteq AT$。在 OR-RS 中决策属性 d 对 A_1,A_2 逻辑“或”的依赖度定义为：

$$\gamma\left(R_{A_1+A_2},d\right)=\sum_{p=1}^{t}\left|\underline{R_{A_1+A_2}}(D_p)\right|\Big/|U|。$$

定理 5.2 在目标信息系统 $S=(U,C\cup\{d\},V,f)$ 中。$A_1,A_2\subseteq AT$。OR-RS 中依赖度和 Pawlak 模型、AND-RS 中依赖度有如下关系：

$$\gamma\left(R_{A_i},d\right)\le\gamma\left(R_{A_1+A_2},d\right)\le\gamma\left(R_{A_1\cup A_2},d\right),\quad i=1,2。$$

证明：由定义 5.3 和定理 4.1 即证。

上面这些度量可以考查和衡量 OR-RS 近似分类的好坏，可以通过使用 OR-RS 进一步寻找信息系统的约简，并进行规则提取的研究，从信息系统中发掘有价值的信息。

6. 算例研究

本节通过一个例子来说明 OR-RS 的应用，并验证有关性质和结论．我们使用一个项目投资目标信息系统——表 1(文献[8])，验证 OR-RS。

此处，仍取 $X=\{e_1,e_2,e_6,e_8\}$，在表 1 中，我们使用如下符号代表对应属性集：

L-{Locus}；P-{Population density}；D-{Decision}；

由各个属性得到的划分为：

$$U/R_L=\{\{e_1,e_7\},\{e_2,e_3,e_4,e_5,e_6\},\{e_8\}\}$$

$$U/R_P=\{\{e_1,e_2\},\{e_3,e_4,e_5\},\{e_6,e_7,e_8\}\}$$

$$U/R_{L\cup P}=\{\{e_1\},\{e_2\},\{e_3,e_4,e_5\},\{e_6\},\{e_7\},\{e_8\}\}$$

经过计算得到：

$\underline{R_L}(X)=\{e_8\}$，$\overline{R_L}(X)=U$，$\underline{R_P}(X)=\{e_1,e_2\}$，

$\overline{R_P}(X)=\{e_1,e_2,e_6,e_7,e_8\}$，

$\underline{R_{L\cup P}}(X)=\{e_1,e_2,e_6,e_8\}=\overline{R_{L\cup P}}(X)$，

$\underline{R_{L+P}}(X)=\{e_1,e_2\}\cup\{e_8\}=\{e_1,e_2,e_8\}$，

$$\begin{aligned}\overline{R_{L+P}}(X)=&\{\{e_1,e_7\}\cup\{e_1,e_2\}\}\cup\{\{e_2,e_3,e_4,e_5,e_6\}\cap\{e_1,e_2\}\}\\&\cup\{\{e_2,e_3,e_4,e_5,e_6\}\cap\{e_6,e_7,e_8\}\}\\&\cup\{\{e_1,e_7\}\cap\{e_6,e_7,e_8\}\}\cup\{\{e_8\}\cap\{e_6,e_7,e_8\}\}\\=&\{e_1,e_2,e_6,e_7,e_8\}\end{aligned}$$

Table 1. A objective information system
表 1. 某目标信息系统

Project	Locus	Population density	Decision
e_1	Common	Big	Yes
e_2	Bad	Big	Yes
e_3	Bad	Small	No
e_4	Bad	Small	No
e_5	Bad	Small	No
e_6	Bad	Medium	Yes
e_7	Common	Medium	No
e_8	Good	Medium	Yes

通过这一结果，我们可以验证定理 3.1、定理 3.2 和定理 4.1 成立。从表 1 中，我们还有：$U/R_D=\{\{e_1,e_2,e_6,e_8\},\{e_3,e_4,e_5,e_7\}\}$。可以计算得到各个度量：

$$\alpha_{R_{L+P}}(X)=3/5,\ \rho_{R_{L+P}}(X)=2/5;$$

$$\alpha_{R_{L\cup P}}(X)=1,\ \rho_{R_{L\cup P}}(X)=0;$$

$$\gamma(L,d)=1/8,\ \gamma(P,d)=5/8;$$

$$\gamma(L\cup P,d)=1;\ \gamma(L+P,d)=3/4。$$

由此可以验证定理 4.1，定理 4.2。

从上面计算可以看出，OR-RS 对概念的近似和计算的依赖度与 Pawlak 粗糙集模型的结果相比较包含了更多的知识和信息，在这个简单的例子中选用 OR-RS 考察各个属性要比 Pawlak 粗糙集模型效果要好．上面例题仅简单说明该模型，对该模型在实际应用中的效果，可以通过大量数据的实例进行验证，与其他模型进行比较，进一步验证该模型的实际效果．在实际的应用过程中要根据不同的应用背景和要求，选择合适的模型或者多个模型结合使用，在模式识别和数据挖掘中达到更好的效果。

7. 结论

本文从等价关系逻辑“与”运算的角度拓展 Pawlak 粗糙集模型得到逻辑“与”粗糙集模型(AND-RS)，对等价关系进行逻辑“或”运算，提出了逻辑“或”粗糙集模型(OR-RS)，深入研究了 OR-RS 的性质，定义了该模型中的粗糙型度量、近似分类精度和近似分类质量，并举例验证说明了该模型．本文从等价关系逻辑运算的角度推广粗糙集模型，为粗糙集理论研究提出关系逻辑运算这样一种视角，可以使粗糙集的研究按照关系运算的思想进行拓展，对于算法实现和数据挖掘有一定的参考价值。

参考文献 (References)

[1] Z．Pawlak. Rough sets. International Journal of Computer and Information Sciences, 1982, 11(8): 41-356.

[2] G. Gediga, I. Düntsch. Rough approximation quality revisited. Artificial Intelligence, 2001, 102(2): 219-234.

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

[4] 张文修, 吴伟志. 粗糙集理论介绍和研究综述[J]. 模糊系统与数学, 2000, 14(4): 1-12.

[5] 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.

[6] 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.

[7] Y. H. Qian, J. Y. Liang, Y. Y. Yao, et al. MGRS: A multi-granulation rough set. Information Sciences, 2010, 180(6): 949-970.

[8] W. H. Xu，H. Z. Yang, and W. X. Zhang. Uncertainty measures of roughness of knowledge and rough sets in ordered information systems. Lecture Notes in Computer Science, 2007, 4682: 759-769.

[9] W. H. Xu, W. X. Zhang. Measuring roughness of generalized rough sets induced by a covering. Fuzzy Sets and Systems, 2007, 158(22): 2443-2455.

[10] X. Y. Zhang, W. H. Xu. A novel approach to roughness measure in fuzzy rough sets. Advances in Soft Computing, 2007, 40: 775-780.