集值信息系统基于邻域关系的属性约简

陈子春[1,2] 秦克云[1]
(西南交通大学数学学院 成都 610013)[1] (西华大学数学与计算机学院 成都 610039)[2]

摘 要 集值信息系统是完备信息系统的广义形式,其中的一些对象在某些属性下的取值可能不止一个,反映的是信息的不确定性。在集值信息系统上引入对象的邻域关系,并以每个对象的邻域作为基本集,建立了集值信息系统的粗糙集方法。为了简化知识的表示,进一步在集值信息系统上提出了邻域协调集与邻域约简的概念,并讨论了这种邻域协调集的性质与判定。最后通过区分函数得到了计算集值信息系统邻域约简的方法。

关键词 集值信息系统,邻域,邻域约简,区分函数
中图法分类号 TP18 **文献标识码** A

Attribute Reduction of Set-valued Information Systems Based on Neighborhood Relation

CHEN Zi-chun[1,2] QIN Ke-yun[1]
(School of Mathematics, Southwest Jiaotong University, Chengdu 610013, China)[1]
(School of Mathematics and Computer, Xihua University, Chengdu 610039, China)[2]

Abstract Set-valued information systems, in which some objects may have more than one value for an attribute, are generalized complete information systems. Introducing a neighborhood relation to set-valued information systems, we established a rough set approach based on the neighborhood of every object as elementary set. Furthermore, in order to simplify knowledge representation, we proposed the concept of neighborhood consistent set and neighborhood reduction, and discussed the properties and judgments of neighborhood consistent set. Finally, we obtained the approach to compute neighborhood reduction for set-valued information systems by the discernibility function.

Keywords Set-valued information system, Neighborhood, Neighborhood reduction, Discernibility function

1 引言

粗糙集理论是一种数据分析的有效工具,自1982年由波兰数学家Pawlak[1]首次提出以来,经过20多年的研究与发展,已经在理论和实际应用中取得了长足的进展。粗糙集理论处理数据主要是通过一对可定义的上、下近似给出任一对象集的近似范围,然后从数据库中挖掘出以规则形式表达的知识。粗糙集理论的另一重要应用是知识的属性约简,即从给定的数据库中找出一个最小属性集,使该属性集与数据库中所有属性包含有同样多的信息,从而简化知识的表示。但Pawlak提出的粗糙集模型中的可定义集是基于等价关系形成的,这限制了它对许多实际数据集的处理,因而人们对Pawlak粗糙集模型做了许多推广。一是把等价关系放松到相容关系(自反、对称的)、优势关系(自反、传递的),甚至自反关系[2-10];另一方法是把等价类形成的划分推广到覆盖[11-13],这些推广可以用来处理更为复杂的(如不完备信息系统、集值信息系统等)知识获取与属性约简。

集值信息系统是完备信息系统的广义形式,其中的一些对象在某些属性下的取值不是一个,也不是取空值,而是取几个值,反映的是信息的不确定性。在给定的集值信息系统上,取相同属性值的对象所形成的知识类不是论域的划分,而是一个覆盖。因此本文利用论域的覆盖,在集值信息系统上引入对象的邻域概念,并以每个对象的邻域为基本集,建立了集值信息系统的粗糙集方法。这可以看作是覆盖粗糙集理论的具体化。为了简化集值信息系统的知识表示,进一步提出了邻域协调集与邻域约简的概念,并讨论了这种邻域协调集的性质与判定。最后通过区分函数得到了计算集值信息系统邻域约简的方法。

2 集值信息系统上对象的邻域关系

信息系统是一个有序二元组 $S=(U,AT)$,其中 U 是有限非空对象集,AT 为有限非空属性集。若每个对象 x 在每个属性 a 下都有一个确定的属性值 $a(x)\in V_a$(其中 V_a 是属性 a 的值域),则称 S 是完备信息系统;若存在 $a(x)$ 是未知的,则称 S 是不完备信息系统;若 $a(x)$ 是不确定的,即 $a(x)\in P_0(V_a)$(其中 $P_0(V_a)$ 表示 V_a 的非空子集全体),则称 S 是集值信息系统。对不完备信息系统,若未知属性值是存在的,则可以转化为集值信息系统来处理[14]。

在完备信息系统中,每个对象在每个属性下的取值都是唯一的,取相同属性值的对象所形成的知识块是对象集的一

到稿日期:2009-05-14 返修日期:2009-07-27 本文受国家自然科学基金资助项目(60474022),四川省重大科技专项配套资金项目(2008GZ0118)资助。
陈子春(1970—),男,博士生,副教授,主要研究方向为粗糙集理论及应用,E-mail:czclph@163.com;秦克云(1962—),男,教授,博士生导师。

个划分。而在集值信息系统中，由于某些对象在某些属性下的取值不是唯一的，因此取相同属性值的对象所形成的知识块一般不是对象集的划分，而是一个覆盖。为了便于理解，下面类似于完备信息引入一些记号。

设 $S=(U,AT)$ 是给定的集值信息系统，$A\subseteq AT$。$\forall a\in A, v\in V_a$，称 (a,v) 是 S 的一个原子，而称 $\bigwedge_{a\in A}(a,v)$ 是 A 的一个描述，A 的所有描述记为 $des(A)$；若 t 是一个 A 描述，(a,v) 是出现在 t 中的一个原子，则记为 $(a,v)\in t$；设 $t\in des(A)$，记 $\|t\|$ 是满足描述 t 的所有对象，即 $\|t\|=\{x\in U:\forall(a,v)\in t, v\in a(x)\}$。

容易看到，$\|(a,v)\|=\{x\in U:v\in a(x)\}$，$\|t\|=\bigcap\{\|(a,v)\|:(a,v)\in t\}$。

定义 1 设 (U,AT) 是集值信息系统，$A\subseteq AT$，记 $des_A(x)=\{t:t\in des(A), x\in\|t\|\}$，则称 $N_A(x)=\bigcap\{\|t\|:t\in des_A(x)\}$ 是对象 x 的邻域。

定理 1 设 (U,AT) 是集值信息系统，$A,B\subseteq AT$，则

(1) $N_A(x)=\bigcap_{a\in A}N_a(x)$，其中 $N_a(x)$ 是 $N_{\{a\}}(x)$ 的简写；

(2) $B\subseteq A\Rightarrow N_A(x)\subseteq N_B(x)$；

(3) $y\in N_A(x)\Leftrightarrow N_A(y)\subseteq N_A(x)$；

(4) $N_A(U)=\{N_A(x):x\in U\}$ 是论域 U 的覆盖。

证明：(1)注意到 $N_a(x)=\bigcap\{\|t\|:t\in des_{\{a\}}(x)\}=\bigcap\{\|(a,v)\|:v\in a(x)\}$，$\|t\|=\bigcap\{\|(a,v)\|:(a,v)\in t\}$，所以由 $y\in N_A(x)\Leftrightarrow\forall t\in des_A(x), y\in\|t\|\Leftrightarrow\forall t\in des(A), x\in\|t\|$ 时，$y\in\|t\|\Leftrightarrow\forall a\in A; x\in\|(a,v)\|$ 时，$y\in\|(a,v)\|\Leftrightarrow\forall a\in A; v\in a(x)$ 时，$y\in\|(a,v)\|\Leftrightarrow\forall a\in A, y\in N_a(x)\Leftrightarrow y\in\bigcap_{a\in A}N_a(x)$，知 $N_A(x)=\bigcap_{a\in A}N_a(x)$ 成立。

(2)由(1)即可得证。

(3)设 $y\in N_A(x)$，则 $\forall t\in des_A(x)$，有 $y\in\|t\|$，即 $\forall t\in des_A(x)$，有 $t\in des_A(y)$。因此，$\forall z\in N_A(y), t\in des_A(x)$，必有 $z\in\|t\|$，从而 $z\in N_A(x)$，$N_A(y)\subseteq N_A(x)$；反之，若 $N_A(y)\subseteq N_A(x)$，则由 $y\in N_A(y)$，即得 $y\in N_A(x)$。

(4)是显然的。

例 1 下面是一个集值信息系统，我们来计算各对象的邻域。

	x_1	x_2	x_3	x_4	x_5	x_6
a_1	{1}	{1,2,3}	{3}	{1,2,3}	{1}	{3}
a_2	{1,2}	{1}	{1}	{1,2}	{1}	{1,2}
a_3	{2}	{1}	{1}	{2}	{1}	{2}
a_4	{1,3}	{1,2}	{2}	{1,2,3}	{1}	{2}

由描述的定义知，$des_{AT}(x_1)=\{t_{11},t_{12},t_{13},t_{14}\}$，其中，$t_{11}=(a_1,1)\wedge(a_2,1)\wedge(a_3,2)\wedge(a_4,1)$，$t_{12}=(a_1,1)\wedge(a_2,1)\wedge(a_3,2)\wedge(a_4,3)$，$t_{13}=(a_1,1)\wedge(a_2,2)\wedge(a_3,2)\wedge(a_4,1)$，$t_{14}=(a_1,1)\wedge(a_2,2)\wedge(a_3,2)\wedge(a_4,3)$，$\|t_{11}\|=\|t_{12}\|=\|t_{13}\|=\|t_{14}\|=\{x_1,x_4\}$，于是 $N_{AT}(x_1)=\|t_{11}\|\cap\|t_{12}\|\cap\|t_{13}\|\cap\|t_{14}\|=\{x_1,x_4\}$。另一方面，由 $N_{a_1}(x_1)=\|(a_1,1)\|=\{x_1,x_2,x_4,x_5\}$，$N_{a_2}(x_1)=\|(a_2,1)\|\cap\|(a_2,2)\|=\{x_1,x_4,x_6\}$，$N_{a_3}(x_1)=\|(a_3,2)\|=\{x_1,x_4,x_6\}$，$N_{a_4}(x_1)=\|(a_4,1)\|\cap\|(a_4,3)\|=\{x_1,x_4\}$，也可以得到 $N_{AT}(x_1)=\bigcap_{a\in AT}N_a(x_1)=\{x_1,x_4\}$。

类似地，可以计算出

$N_{a_1}(x_2)=\{x_2,x_4\}$，$N_{a_2}(x_2)=U$，$N_{a_3}(x_2)=\{x_2,x_3,x_5\}$，$N_{a_4}(x_2)=\{x_2,x_4\}$，$N_{AT}(x_2)=\{x_2\}$；

$N_{a_1}(x_3)=\{x_2,x_3,x_4,x_6\}$，$N_{a_2}(x_3)=U$，$N_{a_3}(x_3)=\{x_2,x_3,x_5\}$，$N_{a_4}(x_3)=\{x_2,x_3,x_4,x_6\}$，$N_{AT}(x_3)=\{x_2,x_3\}$；

$N_{a_1}(x_4)=\{x_2,x_4\}$，$N_{a_2}(x_4)=\{x_1,x_4,x_6\}$，$N_{a_3}(x_4)=\{x_1,x_4,x_6\}$，$N_{a_4}(x_4)=\{x_4\}$，$N_{AT}(x_4)=\{x_4\}$；

$N_{a_1}(x_5)=\{x_1,x_2,x_4,x_5\}$，$N_{a_2}(x_5)=U$，$N_{a_3}(x_5)=\{x_2,x_3,x_5\}$，$N_{a_4}(x_5)=\{x_1,x_2,x_4,x_5\}$，$N_{AT}(x_5)=\{x_2,x_5\}$；

$N_{a_1}(x_6)=\{x_2,x_3,x_4,x_6\}$，$N_{a_2}(x_6)=\{x_1,x_4,x_6\}$，$N_{a_3}(x_6)=\{x_1,x_4,x_6\}$，$N_{a_4}(x_6)=\{x_2,x_3,x_4,x_6\}$，$N_{AT}(x_6)=\{x_4,x_6\}$。

3 集值信息系统基于邻域关系的粗糙集方法

由定理 1 知，$N_A(U)=\{N_A(x):x\in U\}$ 是论域 U 的一个覆盖，因此以 $N_A(U)=\{N_A(x):x\in U\}$ 作为基本集，可以讨论集值信息系统中对象子集的近似。

定义 2 设 (U,AT) 是集值信息系统，$A,B\subseteq AT$，$X\subseteq U$，则 X 关于 $N_A(U)$ 的下近似和上近似可以定义如下

$\underline{C_A}(X)=\{x\in U:N_A(x)\subseteq X\}$，$\overline{C_A}(X)=\{x\in U:N_A(x)\cap X\neq\emptyset\}$

自然 X 关于 $N_A(U)$ 的边界定义为：$Bn_A(X)=\overline{C_A}(X)-\underline{C_A}(X)$。

定理 2 设 (U,AT) 是集值信息系统，$A,B\subseteq AT$，$X\subseteq U$，则

(1) $\underline{C_A}(\emptyset)=\overline{C_A}(\emptyset)=\emptyset$，$\underline{C_A}(U)=\overline{C_A}(U)=U$；

(2) $\underline{C_A}(X)\subseteq X\subseteq\overline{C_A}(X)$，$\underline{C_A}(\underline{C_A}(X))=\underline{C_A}(X)$，$\overline{C_A}(X)=\overline{C_A}(\overline{C_A}(X))$；

(3) $\underline{C_A}(X)=\sim\overline{C_A}(\sim X)$，$\overline{C_A}(X)=\sim\underline{C_A}(\sim X)$；

(4) $\underline{C_A}(X\cap Y)=\underline{C_A}(X)\cap\underline{C_A}(Y)$，$\overline{C_A}(X\cup Y)=\overline{C_A}(X)\cup\overline{C_A}(Y)$；

(5) $\underline{C_A}(X)=\bigcup\{N_A(x):N_A(x)\subseteq X\}$，$\overline{C_A}(X)\subseteq\bigcup\{N_A(x):N_A(x)\cap X\neq\emptyset\}$；

(6) $B\subseteq A\Rightarrow\underline{C_B}(X)\subseteq\underline{C_A}(X)$，$\overline{C_A}(X)\subseteq\overline{C_B}(X)$，$Bn_A(X)\subseteq Bn_B(X)$；

(7) $Y\subseteq X\Rightarrow\underline{C_A}(Y)\subseteq\underline{C_A}(X)$，$\overline{C_A}(Y)\subseteq\overline{C_A}(X)$。

证明：由定理 1 及定义 1 验证即可得证。

例 2 （续例 1）取 $X=\{x_1,x_3,x_4\}$，则由定义 1 计算得

$\underline{C_{AT}}(X)=\{x_1,x_4\}$，$\overline{C_{AT}}(X)=\{x_1,x_3,x_4,x_6\}$，$Bn_{AT}(X)=\{x_3,x_6\}$

4 集值信息系统基于邻域关系的属性约简

对于集值信息系统上的基本集 $N_{AT}(U)=\{N_{AT}(x):x\in U\}$，可确定它并不需要 AT 中的所有属性，人们希望找出 AT 的一个最小属性子集就能确定出知识 $N_{AT}(U)$，这就是知识的属性约简。通过约简，可以深化人们对知识的认识。因此，接下来讨论集值信息系统在邻域关系下的属性约简。

定义 3 称 $A\subseteq AT$ 是集值信息系统 (U,AT) 的邻域协调集，若 $\forall x\in U, N_A(x)=N_{AT}(x)$。若 A 是邻域协调集，且对任意 $B\subset A$，B 不是邻域协调集，则称 A 是集值信息系统的一个邻域约简。

由定理 1 知，$A\subseteq AT$ 是集值信息系统的邻域协调集，等价于 $\forall x\in U, N_A(x)\subseteq N_{AT}(x)$。

设 $x,y\in U$，$A\subseteq AT$，类似于文献[13]，对于 x 的邻域 $N_A(x)$ 与 y 的邻域 $N_A(y)$，从集合的角度来看有 3 种关系：一是

$N_A(x)=N_A(y)$；二是 $N_A(x)\subset N_A(y)$ 或 $N_A(y)\subset N_A(x)$；三是 $N_A(x)\neq N_A(y)$，且 $N_A(x)\not\subset N_A(y)$，$N_A(y)\not\subset N_A(x)$，即 $N_A(x)$ 与 $N_A(y)$ 互不包含。

定理 3 设 (U,AT) 是集值信息系统，$A\subseteq AT$，$x,y\in U$，则

(1) $N_A(x)=N_A(y)\Leftrightarrow\forall a\in A,N_a(x)=N_a(y)$；

(2) $N_A(x)\subset N_A(y)\Leftrightarrow\forall a\in A,N_a(x)\subseteq N_a(y)$，且 $\exists b\in A,N_b(x)\subset N_b(y)$；

(3) $N_A(x)$ 与 $N_A(y)$ 互不包含 $\Leftrightarrow\exists c\in A$，使 $N_c(x)\neq N_c(y)$，且 $N_c(x)\not\subset N_c(y)$，$N_c(y)\not\subset N_c(x)$，或者 $\exists a,b\in A$，使 $N_a(x)\subset N_a(y)$，$N_b(y)\subset N_b(x)$。

证明：(1)"$\Leftarrow$"：显然成立。

"$\Rightarrow$"：若 $\exists a\in A$，$N_a(x)\neq N_a(y)$，则必有 $y\notin N_a(x)$ 或 $x\notin N_a(y)$（否则由 $y\in N_a(x)$，$x\in N_a(y)$，就有 $N_a(x)=N_a(y)$，矛盾）。从而有 $y\notin N_A(x)$ 或 $x\notin N_A(y)$，这与题设条件 $N_A(x)=N_A(y)$ 矛盾。

(2)"$\Leftarrow$"：若 $\forall a\in A$，有 $N_a(x)\subseteq N_a(y)$，则 $N_A(x)\subseteq N_A(y)$ 成立。又由 $\exists b\in A$，$N_b(x)\subset N_b(y)$ 知 $y\notin N_b(x)$，这表明 $N_A(x)\subset N_A(y)$ 成立。

"$\Rightarrow$"：设 $N_A(x)\subset N_A(y)$，则 $x\in N_A(y)$，所以 $\forall a\in A$，$x\in N_a(y)$，即 $N_a(x)\subseteq N_a(y)$ 成立。又由 $N_A(x)\subset N_A(y)$ 知，$y\notin N_A(x)$，从而 $\exists b\in A$，使 $y\notin N_b(x)$，再结合 $N_b(x)\subseteq N_b(y)$，即知 $N_b(x)\subset N_b(y)$ 成立。

(3)"$\Leftarrow$"：若 $\exists c\in A$，使 $N_c(x)\neq N_c(y)$，且 $N_c(x)\not\subset N_c(y)$，$N_c(y)\not\subset N_c(x)$，则有 $x\notin N_c(y)$，$y\notin N_c(x)$，从而有 $y\notin N_A(x)$，$x\notin N_A(y)$，即 $N_A(x)$ 与 $N_A(y)$ 互不包含。

又若 $\exists a,b\in A$，使 $N_a(x)\subset N_a(y)$，$N_b(y)\subset N_b(x)$，则有 $y\notin N_a(x)$，$x\notin N_b(y)$，即 $y\notin N_A(x)$，$x\notin N_A(y)$，这表明 $N_A(x)$ 与 $N_A(y)$ 仍互不包含。

"$\Rightarrow$"：设 $N_A(x)$ 与 $N_A(y)$ 互不包含，则由 $N_A(x)\neq N_A(y)$ 及(1)知，$\exists c\in A$，使 $N_c(x)\neq N_c(y)$，若有 $N_c(x)\not\subset N_c(y)$，$N_c(y)\not\subset N_c(x)$ 成立，则问题得证；若有 $N_c(x)\subset N_c(y)$ 或 $N_c(y)\subset N_c(x)$，则必有 $\exists a,b\in A$，使 $N_a(x)\subset N_a(y)$，$N_b(y)\subset N_b(x)$，否则由(2)就有 $N_A(x)\subset N_A(y)$ 或 $N_A(y)\subset N_A(x)$ 成立，与题设矛盾。

为了找到利用区分矩阵与区分函数[15]计算邻域约简的方法，根据定理 3，类似于文献[13]可以构建两个对象 x 和 y 的区分属性集，从而得到邻域协调集的等价描述。但构建的区分属性集在表达上比较复杂，因此，下面直接从邻域关系的定义出发，建立对象的区分属性集与邻域协调集的等价描述。

定义 4 设 (U,AT) 是集值信息系统，$x,y\in U$，记

$$D_{xy}=\{a\in AT:y\notin N_a(x)\}$$

则称 D_{xy} 是对象 x 和 y 的区分属性集；并称 $D=\{D_{xy}:(x,y)\in U^2\}$ 是集值信息系统的区分矩阵。

注意到，对任意 $x\in U$，$a\in AT$，有 $x\in N_a(x)$，因此由定义知集值信息系统区分矩阵 D 的主对角线上的元素全为 $\emptyset$。

定理 4 设 (U,AT) 是集值信息系统，$A\subseteq AT$，记 $D_0=\{D_{xy}\neq\emptyset:x,y\in U\}$，则 A 是集值信息系统的邻域协调集 $\Leftrightarrow\forall D_{xy}\in D_0,A\cap D_{xy}\neq\emptyset$。

证明："$\Rightarrow$" 设 A 是邻域协调集，即 $\forall x\in U$，$N_A(x)=N_{AT}(x)$。若 $D_{xy}\in D_0$，则由定义知，存在 $a\in AT$，使 $y\notin N_a(x)$，因此 $y\notin N_{AT}(x)=N_A(x)$，从而 $A\cap D_{xy}\neq\emptyset$ 成立。

"$\Leftarrow$"要证 A 是邻域协调集，仅需证明 $\forall x\in U$，$N_A(x)\subseteq N_{AT}(x)$ 成立即可。设 $y\notin N_{AT}(x)$，则存在 $a\in AT$，使 $y\notin N_a(x)$，因此 $D_{xy}\neq\emptyset$，$D_{xy}\in D_0$，于是由题设条件知 $A\cap D_{xy}\neq\emptyset$，这表明 $\exists b\in A$，使 $y\notin N_b(x)$，从而 $y\notin N_A(x)$。

定义 5 设 (U,AT) 是集值信息系统，$D_{xy}\in D_0$。令 $\vee D_{xy}$ 表示 D_{xy} 中所含属性的布尔析取，则称 $F=\wedge\{\vee D_{xy}:D_{xy}\in D_0\}$ 是集值信息系统的区分函数。

定理 5 设 (U,AT) 是集值信息系统，$A\subseteq AT$，则 A 是集值信息系统的一个邻域约简当且仅当 $\wedge A$ 是区分函数 F 的极小析取范式中的一个合取式。

证明：由定理 4 和极小析取范式的定义即可得证。

例 3 （续例 1）计算例 1 所示集值信息系统的邻域约简。

由例 1 有，$N_{a_1}(x_1)=\{x_1,x_2,x_4,x_5\}$，$N_{a_2}(x_1)=\{x_1,x_4,x_6\}$，$N_{a_3}(x_1)=\{x_1,x_4,x_6\}$，$N_{a_4}(x_1)=\{x_1,x_4\}$，所以 $x_2\notin N_{a_2}(x_1)$，$x_2\notin N_{a_3}(x_1)$，$x_2\notin N_{a_4}(x_1)$，从而由区分属性集定义就有 $D_{12}=\{a_2,a_3,a_4\}$，其中 D_{12} 是 $D_{x_1x_2}$ 的简写。

类似地可以计算出其它的 D_{ij}，于是可以求出集值信息系统的区分矩阵为

$$D=\begin{pmatrix}\emptyset & a_2,a_3,a_4 & a_1,a_2,a_3,a_4 & \emptyset & a_2,a_3,a_4 & a_1,a_4\\ a_1,a_3,a_4 & \emptyset & a_1,a_4 & a_3 & a_1,a_4 & a_1,a_3,a_4\\ a_1,a_3,a_4 & \emptyset & \emptyset & a_3 & a_1,a_4 & a_3\\ a_1,a_4 & a_2,a_3,a_4 & a_1,a_2,a_3,a_4 & \emptyset & a_1,a_2,a_3,a_4 & a_1,a_4\\ a_3 & \emptyset & a_1,a_4 & a_3 & \emptyset & a_1,a_3,a_4\\ a_1,a_4 & a_2,a_3 & a_2,a_3 & \emptyset & a_1,a_2,a_3,a_4 & \emptyset\end{pmatrix}$$

从而区分函数的极小析取范式为

$$\begin{aligned}F&=(a_2\vee a_3\vee a_4)\wedge(a_1\vee a_2\vee a_3\vee a_4)\wedge(a_1\vee a_4)\wedge(a_1\vee a_3\vee a_4)\wedge a_3\wedge(a_2\vee a_3)\\&=(a_1\wedge a_3)\vee(a_3\wedge a_4)\end{aligned}$$

所以 $\{a_1,a_3\}$ 与 $\{a_3,a_4\}$ 是集值信息系统的两个邻域约简。

结束语 本文基于覆盖粗糙集理论在集值信息系统上引入了对象的邻域概念，并以每个对象的邻域作为基本集，建立了集值信息系统的粗糙集方法。为了简化知识的表示，我们进一步在集值信息系统提出了邻域协调集与邻域约简的概念，并讨论了这种邻域协调集的性质与判定，最后通过区分函数得到了计算集值信息系统邻域约简的方法。下一步将考虑集值目标信息系统基于邻域关系的规则提取与属性约简。

参 考 文 献

[1] Pawlak Z. Rough set [J]. International Journal of Computer and Information Science, 1982(11): 341-356

[2] Kryszkiewicz M. Comparative study of alternative type of knowledge reduction in inconsistent systems [J]. International Journal of Intelligent Systems, 2001, 16(1): 105-120

[3] Kryszkiewicz M. Rough set approach to incomplete information systems [J]. Information Sciences, 1998, 112(1): 39-49

[4] Pawlak Z, Skowron A. Rough sets: some extension [J]. Information Sciences, 2007, 177: 41-73

[5] Kryszkiewicz M. Rules in incomplete information systems [J]. Information Sciences, 1999, 113: 271-292

[6] Zhang Wen-xiu, Mi Ju-sheng. Incomplete information system and its optimal selections [J]. Computers and Mathematics with Applications, 2004, 48: 691-698

（下转第 280 页）

$$MSE=\frac{1}{MN}\sum_{i=1}^{M}\sum_{j=1}^{N}(f_a(i,j)-f_b(i,j))^2 \quad (10)$$

在应用中，均方误差经常根据与它等价的倒数度量，即峰值信噪比 PSNR(peak signal to noise ratio)来引用：

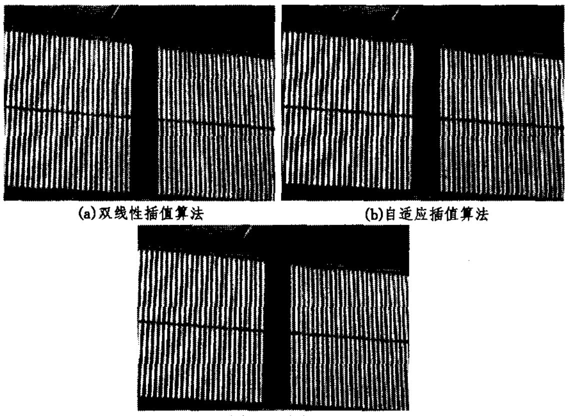
$$PSNR=10*\log10(255^2/MSE) \quad (11)$$

其中，M 和 N 代表图像的尺寸为 M 行 N 列，$f_a(i,j)$ 和 $f_b(i,j)$ 分别代表原始图像和重构图像在第 i 行、第 j 列的强度分量。这里，PSNR 是单个通道的值，即真彩色图像可以分成 R,G,B 三通道独立比较，它反映了原始图像与重构图像相符合的程度，PSNR 越大，则两者越接近。表 2 给出了 3 种不同插值算法对实验图片进行处理后的 PSNR 值。

表 2　3 种算法性能比较

差值方法	B 分量 PSNR 值	G 分量 PSNR 值	R 分量 PSNR 值
双线性插值算法	30.58	36.18	31.44
自适应插值算法	33.24	39.86	36.48
改进的差值算法	33.53	41.18	36.86

本文给出了不同算法对同一图片的插值结果，其采用 MATLAB 软件仿真实现，如图 2 所示。结合实验数据和图片效果可以看出，普通的双线性插值算法效果较差，会产生比较多的伪彩色条纹；基于边界的自适应插值算法由于考虑了颜色的相关性，质量有很大改善；而本文提出的算法与自适应算法相比，在对每个像素点做插值运算时都参照了整个领域内的 8 个像素点，色彩的过渡更平滑，从而进一步减少了伪彩色斑纹(例如图 2(a)中间两扇窗分界部分)；仔细对比图 2(c)和图 2(b)可以看出，图 2(c)上的细密窗叶纹路更清晰，证明了本文算法对图像高频部分恢复效果更好。



(a)双线性插值算法　(b)自适应插值算法

(c)本文的改进算法

图 2　几种不同算法的仿真结果

3.2　硬件可行性评估

本文拟采用 Altera 公司的 cycloneⅡ系列芯片来实现该算法。从存储容量来说，由于改进的插值法是基于 3×3 掩模窗口来计算的，计算一个像素点会用到相邻 3 行内的像素数据，而图像数据是连续输入的，故需要缓存 3 行数据。输入缓冲可以采用 3 个 16kbit 的片内 RAM(称缓存 1 行数据的 RAM 块为 1 条 line buffer)，用于存储 3 行数据，输出缓冲用 1 个 4kbit 的 FIFO，而为了存储 $flag$ 标志变量，需要再增加 1 条 buffer，加权系数的取值固定，可直接采用查表方式实现。这样，总共需要的片内 RAM 在 70k 以下，不到 cycloneⅡ系列芯片内存总量的 30%，故无须使用外部存储器。在计算复杂性方面，本文提出的算法仅限于四则运算，由于 FPGA 芯片内部提供了大量的乘法器资源，文中所提到的 10 位以内的乘法运算可以顺利完成，不会成为系统运行的速度瓶颈。因此，本文提出的算法在硬件上是完全可行的。

结束语　从第 3 节的分析可以看出，虽然本文在对三角函数的求近似简化方面不甚精确，但相对于只做水平和竖直两个方向插值的自适应插值算法而言，本文提出的算法对图像高频部分恢复效果更好，能进一步减少伪彩色斑纹；同时算法描述简单，容易被硬件理解和实现，基本达到了预期的目的。

参 考 文 献

[1] Bayer B E. Color image array: U. S. Patent No. 3917065[P]. 1976

[2] 刘方，杜平安. 基于 Bayer 彩色滤波阵列插值算法的研究[M]. 西安：电子科技大学，2006

[3] Cok B E. Signal processing method and apparatus for producing interpolated chrominance values in a sampled color image signal: U. S. Patent, No. 4642678[P]. 1987

[4] 刘晓松，杨新，文俊. 一种用于数字图像传感器的彩色插值算法[J]. 中国图像图形学报，2003，8A(5)：516-521

[5] Hamilton J F, Adams J E. Adaptive Color Plan Interpolationin Single Sensor Color Electronic Camera: United States Patent, 5629734[P]. 1997

[6] Gonzalez R C, Woods R E. Digital Image Processing Second Edition[M]. 北京：电子工业出版社，2005：33-43，225-239，467-475

[7] 吴心然，赵群飞，山元康裕. 基于小波的 Bayer 颜色滤波阵列上的插值[J]. 计算机工程与应用，2005，26：64-65

[8] 李奇，徐之海，冯华君. 用于数字相机的 CCD 颜色插值算法研究[J]. 激光与红外，2002，32(5)：300-302

[9] Murean D D, Park T W. Demosaicing using Optimal Recovery [J]. IEEE transactions on Image Processing, 2002

(上接第 207 页)

[7] Guan Yan-yong, Wang Hong-kai. Set-valued information systems [J]. Information Sciences, 2006, 176: 2507-2525

[8] Slowinski R, Vanderpooten D. A generalized definition of rough approximations based on similarity [J]. IEEE Transactions Data Knowledge Engineering, 2000, 2: 331-336

[9] 陈子春，秦克云. 区间值信息系统在变精度相容关系下的属性约简[J]. 计算机科学，2009，36(3)：163-165

[10] 陈子春，秦克云. 集值信息系统基于变精度相容关系的知识约简[J]. 计算机工程与应用，2008，44：27-29

[11] Zhu W, Wang Fei-yue. Some results on the covering generalized rough sets [J]. Pattern Recognition Artificial Intelligence, 2002, 5: 6-13

[12] Zhu W, Wang Fei-yue. Reduction and axiomization of covering generalized rough sets [J]. Information Sciences, 2003, 152: 217-230

[13] Tsang E C C, Chen De-gang, Yeung D S. Approximations and reducts with covering generalized rough sets [J]. Computers & Mathematics with applications, 2008, 56: 279-289

[14] 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

[15] Skowron A, Rauszer C. The discernibility matrices and functions in information systems [M]. Boston: Kluwer Academic Publishers, 1992: 331-362