邻域信息系统基于同态的数据压缩

刘晓娟[1] 米据生[1,2] 李仲玲[3]
（河北师范大学数学与信息科学学院 石家庄 050024）[1]
（河北省计算数学与应用重点实验室 石家庄 050024）[2]
（河北师范大学汇华学院数学部 石家庄 050091）[3]

摘 要 用同态的概念作为处理邻域信息系统中数据压缩的工具。给出了邻域信息系统和诱导的邻域信息系统概念，并在此基础上定义了一种新的协调函数，研究了在同态映射下邻域信息系统的一些重要性质，证明了原邻域信息系统与同态象邻域信息系统约简的等价性，从而得到一种通过计算数据量较少的象信息系统的约简来对原系统进行数据压缩的方法。

关键词 邻域信息系统，协调函数，同态，约简
中图法分类号 TP18 **文献标识码** A

Date Compression with Homomorphism in Neighborhood Information Systems

LIU Xiao-juan[1] MI Ju-sheng[1,2] LI Zhong-ling[3]
(College of Mathematics and Information Science, Hebei Normal University, Shijiazhuang 050024, China)[1]
(Key Laboratory of Computational Mathematics with Applications, Shijiazhuang 050024, China)[2]
(College of Mathematics, Huihua College of Hebei Normal University, Shijiazhuang 050091, China)[3]

Abstract The notion of homomorphism is used as a tool to deal with date compression in neighborhood information systems. We reviewed the notions of neighborhood information systems and the induced neighborhood information systems. Then a novel definition of consistent function was proposed, and some important properties of neighborhood information systems under homomorphism were discussed. Furthermore, we proved that the reduction of the original system and its image system is equivalent to each other. By calculating the reduction of the less data image system, a more convenient measure to handle the date compression of the original system was obtained.

Keywords Neighborhood information systems, Consistent function, Homomorphism, Reduction

1 引言

在信息系统中，经常需要处理大量的复杂数据。如何在保持数据结构不变的情况下除去冗余数据，如何发现潜在的有价值信息，比如属性约简、重要性度量等，这些都是我们要关心的问题。而在解决拥有大量复杂数据的信息系统的数据压缩问题时，数据库中信息量的减少就意味着把前后两个信息系统中的数据用多对一的关系联系起来。同态概念，作为研究两个信息系统关系的数学工具，最早是由 Graymala-Busse 在 1986 年提出的[1]。随后有学者研究了在同态意义下信息系统的基本性质，证明了在同态下原系统和象系统的约简是等价的[2-8]。通过这一重要结论，我们得到了一种新的解决信息系统中约简或数据压缩的方法，即一种对数据量较大的原信息系统进行约简或数据压缩，通过计算由多对一的关系映射得到的数据量较少的象信息系统的约简，再根据两个信息系统约简是等价的这一重要性质，最终对原系统进行约简或数据压缩的简便方法。在很多情况下，信息系统是不完备的，由此在论域上会产生一个覆盖，我们称这样的系统为覆盖信息系统[9-11]。随后，一些学者通过定义覆盖信息系统间的同态和相应的协调函数概念研究了覆盖信息系统基于同态的数据压缩[12]。这样，对于用同态的方法解决信息系统中的约简或数据压缩的问题就形成了一套较为完善的理论体系。

本文沿用以上思想，对邻域信息系统引入了同态的概念，并构造了新的协调函数，然后验证了基于同态的邻域信息系统的性质，最后证明了原邻域信息系统与象信息系统的约简是等价的，进而得到了邻域信息系统中基于同态的数据压缩方法。

2 邻域信息系统

本节主要给出邻域信息系统的定义及其诱导的邻域覆盖。

到稿日期：2012-04-14 返修日期：2012-06-21 本文受国家自然科学基金（61170107，60963006），高等学校博士学科点专项科研基金（20101303110004），河北师范大学数学与信息科学学院研究生基金资助。

刘晓娟（1985－），女，硕士生，主要研究方向为人工智能数学基础、粗糙集、粒计算等，E-mail：shuihan507@163.com；**米据生**（1966－），男，博士，教授，主要研究方向为粗糙集、粒计算、数据挖掘与近似推理等；**李仲玲**（1983－），女，硕士，主要研究方向为人工智能数、粗糙集、粒计算、概念格。

定义 1 设论域$U=\{x_1,x_2,\cdots,x_n\}$，$\forall x_i\in U$，x_i 的邻域为$\delta(x_i)=\{x_j\mid x_j\in U,\Delta(x_i,x_j)\leqslant\varepsilon\}$，其中 ε 为阈值，Δ 为度量函数。

$\delta(x_i)$是包含 x_i 的信息粒，阈值的大小决定了邻域的大小，ε 越大，邻域越大，相应的就有更多的样本落入 x_i 的邻域中。

定义 2 设$\langle U,\Delta\rangle$为一个度量空间，则邻域粒族$\{\delta(x_i)\mid x_i\in U\}$形成一个粒信息系统，这些基元形成了 U 的覆盖，即

(1) $\forall x_i\in U,\delta(x_i)\neq\varnothing$；

(2) $\bigcup_{i=1}^{n}\delta(x_i)=U$。

定义 3 设论域 $U=\{x_1,x_2,\cdots,x_n\}$，令 $\delta=\{\delta(x_1),\delta(x_2),\cdots,\delta(x_n)\}$为 U 上的一个邻域覆盖。此时设 $\delta_x=\bigcap\{\delta(x_i):x\in\delta(x_i)\}$，且 $\mathrm{cov}(\delta)=\{\delta_x:x\in U\}$，则 $\mathrm{cov}(\delta)$也为 U 的一个邻域覆盖，称为由 δ 诱导的邻域覆盖。

对任意 $x\in U$，δ_x 为 $\mathrm{cov}(\delta)$中包含 x 的最小子集。$\mathrm{cov}(\delta)=\delta$ 当且仅当 δ 为 U 的一个划分。对任意 $x,y\in U$，若 $y\in\delta_x$，由 $\mathrm{cov}(\delta)$定义有，δ_y 是包含 y 的最小集合，又 $y\in\delta_x$，则 $\delta_x\supseteq\delta_y$。所以，若 $y\in\delta_x$ 且 $x\in\delta_y$，则 $\delta_x=\delta_y$。

定义 4 令 $\Omega=\{\delta_i:i=1,\cdots,m\}$为 U 的一个邻域族，$\forall x\in U$，令 $\Omega_x=\bigcap\{\delta_{ix}:\delta_{ix}\in\mathrm{cov}(\delta_i),i=1,\cdots,m\}$，$\mathrm{cov}(\Omega)=\{\Omega_x:x\in U\}$，则 $\mathrm{cov}(\Omega)$也为 U 的一个覆盖，称为由 Ω 诱导的邻域覆盖。

明显有，Ω_x 为 Ω 中所有包含 x 的元素交集，所以对任意 $x\in U$，Ω_x 为 $\mathrm{cov}(\Omega)$中包含 x 的最小集合，$\mathrm{cov}(\Omega)$可以看作是 Ω 中覆盖的交集。若 Ω 中每一个覆盖都是划分，则 $\mathrm{cov}(\Omega)$也是一个划分。Ω_x 为包含 x 的等价类，同理，对任意 $x,y\in U$，若 $y\in\Omega_x$，则 $\Omega_x\supseteq\Omega_y$，所以若 $y\in\Omega_x$ 且 $x\in\Omega_y$，则 $\Omega_x=\Omega_y$。

以上给出了邻域信息系统及其诱导的邻域信息系统的定义，下面我们构造邻域信息系统中的协调函数并研究其性质。

3 邻域信息系统中的协调函数及其基本性质

设有两个信息系统论域为 U 和 V，U 的所有覆盖类记作 $C(U)$，V 的所有覆盖类记作 $C(V)$。

定义 5 设 $f:U\to V$ 是 U 到 V 的映射，$\delta=\{\delta(x_1),\delta(x_2),\cdots,\delta(x_n)\}$为 U 的一个邻域覆盖，$\delta_x=\bigcap\{\delta(x_i):x\in\delta(x_i)\}$，且 $\mathrm{cov}(\delta)=\{\delta_x:x\in U\}$，设$[x]=\{y\in U:f(y)=f(x)\}$。若 $\forall x\in U$，有$[x]\subseteq\delta_x$，则称 f 是关于δ 的协调函数。

定理 1 设 $f:U\to V$ 是 U 到 V 的映射，$\delta=\{\delta(x_1),\delta(x_2),\cdots,\delta(x_n)\}$为 U 邻域覆盖，若 f 为关于δ 的协调函数，则 $f(\delta(x_i))\cap f(\delta(x_j))=f(\delta(x_i)\cap\delta(x_j))$，$\forall i,j\leqslant n$。

证明：先证若 $\delta(x_i)\cap\delta(x_j)=\varnothing$，就有 $f(\delta(x_i))\cap f(\delta(x_j))=\varnothing$。（反证）假设 $f(\delta(x_i))\cap f(\delta(x_j))\neq\varnothing$，则存在 $u\in f(\delta(x_i))\cap f(\delta(x_j))$，即存在 $u\in f(\delta(x_i))$ 且 $u\in f(\delta(x_j))$。所以就存在 $x\in\delta(x_i)$，$y\in\delta(x_j)$，使得 $f(x)=u$，$f(y)=u$，即$[x]=[y]$。由 δ_x,δ_y 的定义有 $\delta_x\subseteq\delta(x_i)$，$\delta_y\subseteq\delta(x_j)$。因为 f 是协调函数，则有$[x]\subseteq\delta_x$，$[y]\subseteq\delta_y$，那么就有$[x]\subseteq\delta(x_i)$，$[y]\subseteq\delta(x_j)$，即$[x]=[y]\subseteq\delta(x_i)\cap\delta(x_j)$，这与 $\delta(x_i)\cap\delta(x_j)=\varnothing$ 矛盾，因此 $f(\delta(x_i))\cap f(\delta(x_j))=\varnothing$。

下证当 $\delta(x_i)\cap\delta(x_j)\neq\varnothing$，有 $f(\delta(x_i)\cap\delta(x_j))=f(\delta(x_i))\cap f(\delta(x_j))$。显然有 $f(\delta(x_i)\cap\delta(x_j))\subseteq f(\delta(x_i))\cap f(\delta(x_j))$恒成立。因为对于任意 $u\in f(\delta(x_i)\cap\delta(x_j))$，存在 $x\in\delta(x_i)\cap\delta(x_j)$，使得 $f(x)=u$。所以就有 $x\in\delta(x_i)$，使得 $f(x)\in f(\delta(x_i))$，且 $x\in\delta(x_j)$使得 $f(x)\in f(\delta(x_j))$，即 $f(x)\in f(\delta(x_i))\cap f(\delta(x_j))$。反之证明 $f(\delta(x_i))\cap f(\delta(x_j))\subseteq f(\delta(x_i)\cap\delta(x_j))$。对任意 $u\in f(\delta(x_i))\cap f(\delta(x_j))$，有 $u\in f(\delta(x_i))$且 $u\in f(\delta(x_j))$。所以就存在 $x\in\delta(x_i)$，$y\in\delta(x_j)$，使得 $f(x)=u$，$f(y)=u$，即$[x]=[y]$。又因为 $\delta_x\subseteq\delta(x_i)$，$\delta_y\subseteq\delta(x_j)$，且 f 为协调函数，则有$[x]\subseteq\delta_x$，$[y]\subseteq\delta_y$，所以就有$[x]\subseteq\delta(x_i)$，$[y]\subseteq\delta(x_j)$，即$[x]=[y]\subseteq\delta(x_i)\cap\delta(x_j)$，那么就有 $f([x])=f([y])\subseteq f(\delta(x_i)\cap\delta(x_j))$。又因为 f 是协调函数，$f([x])=f([y])=f(x)=f(y)=u$，即 $u\in f(\delta(x_i)\cap\delta(x_j))$，所以 $f(\delta(x_i))\cap f(\delta(x_j))\subseteq f(\delta(x_i)\cap\delta(x_j))$。综上所述，定理得证。

定理 2 设 $f:U\to V$ 是 U 到 V 的映射，$\delta=\{\delta(x_1),\delta(x_2),\cdots,\delta(x_n)\}$为 U 的邻域覆盖，若 f 为协调函数，则 $\forall\delta(x_i)\in\delta$，有 $f^{-1}(f(\delta(x_i)))=\delta(x_i)$。

证明：只需证 $f^{-1}(f(\delta(x_i)))\subseteq\delta(x_i)$即可。设 $x\in f^{-1}(f(\delta(x_i)))$，则 $f(x)\in f(\delta(x_i))$，$\exists y\in\delta(x_i)$使得 $f(x)=f(y)$，即 $\delta_y\subseteq\delta(x_i)$，又由 f 为协调函数，$[x]=[y]\subseteq\delta_y\Rightarrow[x]=[y]\subseteq\delta(x_i)$，所以 $x\in[x]\subseteq\delta(x_i)$，因此 $f^{-1}(f(\delta(x_i)))\subseteq\delta(x_i)$，得证 $f^{-1}(f(\delta(x_i)))=\delta(x_i)$。

注：若 f 为关于δ 的协调函数，我们定义 f^{-1} 为如下形式：$f^{-1}(f(x))=\{y:f(y)=f(x),x,y\in U\}$。

定义 6 设论域$U=\{x_1,x_2,\cdots,x_n\}$，$\delta_1,\delta_2\in C(U)$，$\delta_1\cap\delta_2=\{\delta_1(x)\cap\delta_2(x):\delta_{ix}\in\mathrm{cov}(\delta_i),i=1,2,\forall x\in U\}$，则称 $\delta_1\cap\delta_2$ 为 δ_1 和 δ_2 的交集。

通过定义 6 可知，$\forall x\in U$，$\delta_{1x}\cap\delta_{2x}$是 $\delta_1\cap\delta_2$ 中包含 x 的最小集合，显然 $\delta_1\cap\delta_2$ 也为 U 的一个邻域覆盖。

定理 3 设 $f:U\to V$ 是 U 到 V 的映射，$\delta_1,\delta_2\in C(U)$，若 f 相对于δ_1 和 δ_2 都协调，则 f 关于 $\delta_1\cap\delta_2$ 也协调。

证明：对任意 $x\in U$，设 δ_x 为 $\delta_1\cap\delta_2$ 中包含 x 的最小集合，其中 δ_{1x}为 $\mathrm{cov}(\delta_1)$中包含 x 的最小集合，δ_{2x}为 $\mathrm{cov}(\delta_2)$中包含 x 的最小集合，则由定义 6 知，$\delta_{1x}\cap\delta_{2x}=\delta_x$。因为 f 为关于 δ_1 和 δ_2 的协调函数，则有$[x]\subseteq\delta_{1x}$且$[x]\subseteq\delta_{2x}$，则$[x]\subseteq\delta_{1x}\cap\delta_{2x}=\delta_x$，所以 f 关于 $\delta_1\cap\delta_2$ 也协调。

4 覆盖映射的基本性质

本节将把两个经典集合映射的概念拓展到论域的幂集，下面介绍覆盖映射的概念及其性质。

定义 7 设 $f:U\to V$ 是 U 到 V 的满射，f 可以诱导一个从 $C(U)$到 $C(V)$的映射和 $C(V)$到 $C(U)$的映射，即：

$\hat{f}:C(U)\to C(V),\delta\to f(\delta)\in C(V),\forall\delta\in C(U)$

$\hat{f}(\delta)=\{f(\delta(x_i)):\delta(x_i)\in\delta\}$

$\hat{f}^{-1}:C(V)\to C(U),\delta'\to f^{-1}(\delta')\in C(U),\forall\delta'\in C(V)\hat{f}^{-1}(\delta')=\{f^{-1}(\delta'(y_i)):\delta'(y_i)\in\delta'\}$

注：在不混淆定义的情况下，下面的讨论中用 f 和 f^{-1}代替$\hat{f}$和$\hat{f}^{-1}$。

定理 4 设 $\delta\in C(U)$，若 f 为关于 δ 的协调函数，则 $f^{-1}(f(\delta))=\delta$。

证明：设 $\delta(x_i)\in\delta$，因为 f 为 U 上关于 δ 的协调函数，由定理 2 知，$f^{-1}(f(\delta(x_i)))=\delta(x_i)$，对任意 $\delta_i\in\Omega$，则有 $f^{-1}(f(\delta))=\delta$。

推论 1 设 $f:U\to V$ 是 U 到 V 的映射，$\Omega=\{\delta_i:i=1,\cdots,m\}$ 为 U 的一个邻域覆盖，若对任意 $\delta_i\in\Omega$，f 是关于 δ_i 的协调函数，则 $f^{-1}(f(\bigcap\Omega))=\bigcap\Omega$。

定理 5 设 $f:U\to V$ 是 U 到 V 的映射，$\delta_1,\delta_2\in C(U)$，若 f 是相对于 δ_1 和 δ_2 的协调函数，则 $f(\delta_1\cap\delta_2)=f(\delta_1)\cap f(\delta_2)$。

证明：对任意 $x\in U$，由定义 6、定理 1 和定理 2，用反证法可得 $f(\delta_{1x})\cap f(\delta_{2x})$ 是 $f(\delta_1)\cap f(\delta_2)$ 中包含 $f(x)$ 的最小集合。因为 $f(\delta_{1x})\cap f(\delta_{2x})=f(\delta_{1x}\cap\delta_{2x})=f(\delta_x)$，则有对任意 $x\in U$，$\delta_x\in\delta_1\cap\delta_2$，则有 $f(\delta_x)\in f(\delta_1\cap\delta_2)$，即有 $f(\delta_1\cap\delta_2)\subseteq f(\delta_1)\cap f(\delta_2)$。又有 $f(\delta_{1x})\cap f(\delta_{2x})$ 为 $f(\delta_1)\cap f(\delta_2)$ 中包含 $f(x)$ 的最小集合，则 $f(\delta_{1x})\cap f(\delta_{2x})=f(\delta_x)\in f(\delta_1\cap\delta_2)$，即 $f(\delta_{1x})\cap f(\delta_{2x})\in f(\delta_1\cap\delta_2)$，则 $f(\delta_1\cap\delta_2)\supseteq f(\delta_1)\cap f(\delta_2)$。因此 $f(\delta_1\cap\delta_2)=f(\delta_1)\cap f(\delta_2)$，定理得证。

推论 2 设 $f:U\to V$ 是 U 到 V 的映射，$\Omega=\{\delta_i:i=1,\cdots,m\}$ 为 U 的一个邻域覆盖，若对任意 $\delta_i\in\Omega$，f 是关于 δ_i 的协调函数，则 $f(\bigcap_{i=1}^{m}\delta_i)=\bigcap_{i=1}^{m}f(\delta_i)$。

5 基于同态的属性约简

本节中用同态的概念作为处理邻域信息系统中数据压缩的工具。通过用同态进行数据压缩，可以得到一个和原系统数据库有相同属性约简的较小的象系统。首先介绍邻域信息系统中同态的概念。

定义 8 设 U,V 是有限论域，$f:U\to V$ 是 U 到 V 的满射，$\Omega=\{\delta_i:i=1,\cdots,m\}$ 为 U 的一个邻域系统，则 (U,Ω) 可以看成是一个邻域信息系统，$(V,f(\Omega))$ 称为由 f 诱导的 (U,Ω) 上的邻域信息系统。

定义 9 设 (U,Ω) 是一个邻域信息系统，$(V,f(\Omega))$ 为由 f 诱导的 (U,Ω) 上的邻域信息系统，若对任意 $\delta_i\in\Omega$，f 为关于 δ_i 的协调函数，则称 f 是从 (U,Ω) 到 $(V,f(\Omega))$ 的同态映射。

定义 10 设 (U,Ω) 是一个邻域信息系统，$\delta_i\in\Omega$，若 $\bigcap(\Omega-\delta_i)=\bigcap\Omega$，则称 δ_i 是冗余的，否则称 δ_i 为必要的。必要元素的交集称为 Ω 的核，记为 $core(\Omega)$。设 $P\subseteq\Omega$，称 P 为 Ω 的约简，若满足：

(1) $\bigcap P=\bigcap\Omega$；

(2) $\forall\delta_i\in P$，$\bigcap P\neq\bigcap(P-\delta_i)$。

定理 6 设 (U,Ω) 是一个邻域信息系统，$(V,f(\Omega))$ 为由 f 诱导的 (U,Ω) 上的邻域信息系统，$P\subseteq\Omega$，若 f 是从 (U,Ω) 到 $(V,f(\Omega))$ 的同态映射，则 P 为 Ω 的约简 $\Leftrightarrow f(P)$ 是 $f(\Omega)$ 的约简。

证明：由定义 9 和定义 10、推论 1 和推论 2 可得。

推论 3 设 (U,Ω) 是一个邻域信息系统，$(V,f(\Omega))$ 为由 f 诱导的 (U,Ω) 上的邻域信息系统，$\delta_i\in\Omega$，$P\subseteq\Omega$，若 f 是从 (U,Ω) 到 $(V,f(\Omega))$ 的同态映射，则有：

(1) δ_i 为 Ω 中必要的 $\Leftrightarrow f(\delta_i)$ 是 $f(\Omega)$ 中必要的；

(2) P 在 Ω 中为冗余的 $\Leftrightarrow f(P)$ 在 $f(\Omega)$ 中是冗余的；

(3) 对于前后两个邻域信息系统的核有如下关系：$f(core(\Omega))=core(f(\Omega))$，$f^{-1}(core(f(\Omega)))=core(\Omega)$。

以上证明了在同态下原信息系统与象信息系统属性约简的等价性。这样，在对原邻域信息系统进行数据压缩时就可以通过象信息系统的约简而更快地对原系统进行约简。

结束语 本文在邻域信息系统中引入同态的概念，定义了一种新的协调函数，然后用同态的概念作为处理邻域信息系统中数据压缩的工具。文中研究了邻域信息系统在同态下的一些重要性质，并证明了原邻域信息系统与同态象邻域信息系统约简的等价性。所以若要对数据量较大的原邻域信息系统进行约简或数据压缩，可以先计算由多对一映射得到的数据量较少的象邻域信息系统的约简，再根据两个邻域信息系统约简是等价的这一重要性质，最终得到原邻域系统的约简或数据压缩的简便方法。这种在邻域信息系统中用同态思想进行压缩的方法不仅大大提高了约简的效率，而且可以很好地节省人力和时间。

参考文献

[1] Graymala-Busse J W. Algebraic properties of knowledge representation systems[C]//Proceedings of the International Symposium on Method for Intelligent Systems. 1986:432-440

[2] Li De-yu, Ma Yi-chen. Invariant characters of information systems under some homomorphi-sms[J]. Information Science, 2000, 129 (1-4):211-220

[3] Wang Chang-zhong, Wu Chong-xin, Chen De-gang, et al. Some properties of relation information systems under homomorphisms[J]. Applied Mathematics Letters, 2008, 21(9):940-945

[4] Wang Chang-zhong, Wu Chong-xin, Chen De-gang, et al. Communicating between information Systems[J]. Information Science, 2008, 178:3228-3239

[5] Wang Chang-zhong, Chen De-gang, Zhu Liang-kuan. Homomorphism between fuzzy information systems[J]. Applied Mathematics Letters, 2009, 22:1045-1050

[6] Zhai Yan-hui, Qu Kai-she. On characteristics of information system homomorphisms[J]. Theory Computing Systems, 2009, 44 (3):414-431

[7] Zhu Ping, Wen Qiao-yan. Some improved results on communication between information systems [J]. Information Science, 2010, 180:3521-3531

[8] Zhu Ping, Wen Qiao-yan. Homomorphisms between fuzzy information systems revisited [J]. Information Science, 2011, 24: 1548-1553

[9] Zhu W. Relationship between generalized rough sets based on binary relation and covering[J]. Information Science, 2009, 179: 210-225

[10] Zhu W, Wang F Y. Reduction and axiomization of covering generalized rough sets[J]. Information Science, 2003, 152:217-230

[11] Yao Y Y. Relational interpretations of neighborhood operator and rough set approximation operators[J]. Information Science, 1998, 111:239-259

[12] Wang Chang-zhong, Chen De-gang, Wu Chong, et al. Data compression with homomorphism in covering information systems [J]. International Journal of Approximate Reasoning, 2011, 52: 519-525