

基于信息系统属性同态的数据压缩

郝彦彬¹ 郭晓² 杨乃定¹

(西北工业大学管理学院 西安 710129)¹ (西安电子科技大学经济与管理学院 西安 710126)²

摘要 数据压缩是数据处理的一个重要主题,同态是实现数据压缩的一种有效工具。根据信息系统属性集及其上的依赖关系,定义了信息系统属性同构和信息系统属性同态的概念。研究了信息系统属性同态的性质,利用属性等价关系诱导出了信息系统属性理想同态,应用信息系统属性理想同态实现了信息系统属性集的无损压缩。最后通过比较原信息系统与同态像信息系统的距离,给出了度量任意属性同态理想程度的方法。

关键词 函数依赖,信息粒,信息系统结构,属性同态,理想同态

中图分类号 TP182 **文献标识码** A **DOI** 10.11896/j.issn.1002-137X.2016.2.009

Data Compression with Attribute Homomorphism in Information Systems

HAO Yan-bin¹ GUO Xiao² YANG Nai-ding¹

(College of Management, Northwestern Polytechnical University, Xi'an 710129, China)¹

(School of Economics and Management, Xidian University, Xi'an 710126, China)²

Abstract Data compression is an important topic in data processing and homomorphism is considered as an effective tool for data compression. This paper defined the attribute isomorphism and attribute homomorphism for information system based on the functional dependency relation over attributes. It then investigated major properties of the attribute homomorphism. By using the attribute equivalence to acquire the ideal homomorphism, it achieved lossless compression for information system. Finally, this paper provided a method to measure the ideal level of any attribute homomorphism by comparing distance between the original system and the image information system.

Keywords Functional dependency, Information granules, Information system structure, Attribute homomorphism, Ideal homomorphism

1 引言

信息系统是人工智能领域重要的数学模型之一。信息系统同态的概念作为研究信息系统之间关系的一种有效的数学工具,最早在文献[1]中被提出,文献[2]研究了信息系统的自同态,文献[3]揭示了构成信息系统同态的对象映射、属性映射以及值域映射的相互依赖性。文献[4-6]研究了关系信息系统的同态性质,文献[7,8]研究了模糊信息系统的同态性质,这些研究都证明了在同态下原系统和像系统的约简是等价的。文献[9,10]研究了覆盖信息系统及决策信息系统的同态性质,文献[11]研究了领域信息系统基于同态的数据压缩。

已有的信息系统概念强调信息系统的论域对象、属性以及属性值域,而将属性间的依赖关系视为空集。事实上,现实中很多信息系统,其属性间的依赖关系是明确存在的,比如关系数据库理论中的函数依赖关系^[12]。函数依赖关系可以看作信息系统上的语义约束,除了数据库领域外,函数依赖也在其他很多领域有着应用,文献[13]中总结了函数依赖在逻辑编程、动态资源管理、神经网络设计、知识表达、分类系统设

计、语义网等领域的应用。本文认为,信息系统的结构是由属性集及属性间的依赖关系决定的。为此,本文将传统的信息系统定义的四元组 $S=(U,A,V,f)$ 扩展为五元组 $S=(U,A,V,f,\Omega)$,增加了属性集 A 上的函数依赖集合 Ω 。

本文根据属性集及属性间的依赖关系定义了信息系统结构的概念,以此为基础,进一步定义了信息系统属性同构和属性同态的概念。如果属性同态下的像信息系统结构与原信息系统结构相同,则称之为属性理想同态。属性理想同态因为没有改变信息系统结构,可以认为是对信息系统做了无损压缩。本文研究了信息系统属性理想同态存在的条件。一般情况下,具有 N 个属性的信息系统若压缩为 M 个属性的像信息系统($M < N$),可能存在多个属性同态,度量这些属性同态的像信息系统与原信息系统的距离,从而找到(近似)理想同态,是本文研究的另一个主要内容。

本文第2节给出信息系统属性同态的定义及相关概念;第3节给出信息系统属性集同态的基本性质以及理想同态存在的条件;第4节给出信息系统属性同态下系统结构变化的度量方法;第5节给出应用上述理论对信息系统做属性压缩

到稿日期:2015-03-20 返修日期:2015-06-20 本文受国家自然科学基金:相继故障视角下基于风险传播模型的研发网络脆弱性研究(71471146)资助。

郝彦彬(1976—),男,博士生,讲师,主要研究方向为信息系统、数据挖掘、粒计算,E-mail:nwpuhyb@163.com;郭晓(1980—),女,博士,讲师,主要研究方向为管理系统工程、风险管理、项目管理;杨乃定(1964—),男,博士,教授,博士生导师,主要研究方向为管理系统工程、风险管理、项目管理。

的例子;最后总结全文。

2 信息系统属性同态

本文约定用大写字母表示属性的集合,小写字母表示单个属性,用 Xa 表示 $X \cup a$, XY 表示 $X \cup Y$, $X \setminus Y$ 表示 X 与 Y 的差集 $X - Y$, $P(X)$ 表示 X 的幂集。在不引起歧义的情况下,将属性集合 $X = \{1, 2, \dots, n\}$ 简写为 $X = 12 \dots n$ 。

定义 1(信息系统) 信息系统 S 是一个四元组 $S = (U, A, V, f)$ 。这里 U 称为论域,它的元素称为对象; A 称为属性集,它的元素称为属性; U 与 A 都是有限非空集; V 称为属性值域, $V = \{V_a | V_a \text{ 是 } a \text{ 的值域}, a \in A\}$; f 为 $U \times A$ 到 V 的一个映射, $\forall x \in U, a \in A, f(x, a) \in V_a$ 。

本文研究的信息系统同态是在属性集 A 的函数依赖的基础上进行的,为此对定义 1 做了扩展,定义信息系统如下。

定义 2(扩展的信息系统) 信息系统 S 是一个五元组 $S = (U, A, V, f, \Omega)$ 。 U, A, V, f 的含义同定义 1。 Ω 是 A 上函数依赖集合。 $\Omega \subseteq P(A) \times A, \Omega = \{X \rightarrow a | X \in P(A), a \in A\}$ 。 $X \rightarrow a$, 也记为 $a \leq X$ 。在不涉及 U, V, f 3 个要素的情况下,信息系统可简写为 $S = (A, \Omega)$ 。

设 $A = \{1, 2\}, \Omega = \{1 \rightarrow 2\}$, 则 $S = (A, \Omega)$ 是一个信息系统。

定义 3(属性集闭包) 设 Ω 为属性集 A 上的一组函数依赖, $X \subseteq A, \Omega(X+) = \{a | X \rightarrow a \text{ 或者 } a \leq X \text{ 能由 } \Omega \text{ 根据 Armstrong 公理导出}^{[12]}\}$, 称为属性集 X 关于函数依赖集 Ω 的闭包,可以简写为 $X+$, A 的所有属性子集的 Ω 闭包的集合记为 $\Omega+, \Omega+ = \{G | G = G+, G \subseteq A\}$ 。 A 的空子集 $\emptyset$ 的闭包为 $\emptyset, \emptyset+ = \emptyset, A$ 的闭包为 $A, A+ = A$ 。

性质 1 属性集闭包运算性质

- (i) 扩展性: $X \subseteq X+$;
- (ii) 单调性: 若 $X \subseteq Y$, 则 $X+ \subseteq Y+$;
- (iii) 幂等性: $(X+)+ = X+$;
- (iv) 若 $X \rightarrow Y$, 则 $Y \subseteq X+$;
- (v) 若 $X \rightarrow Y$, 则 $X+ = (X \cup Y)+$;
- (vi) 若 $X \subseteq Y+$, 则 $X+ \subseteq Y+$;
- (vii) 交封闭性: $X \in \Omega+, Y \in \Omega+$, 则 $X \cap Y \in \Omega+$ 。

定义 4(属性集等价) A 是信息系统 S 的属性集, Ω 是 A 上函数依赖集合。对于 $X, Y \subseteq A$, 并且 $X \setminus Y \neq \emptyset, Y \setminus X \neq \emptyset$ 。如果 $X \rightarrow Y$ 同时 $Y \rightarrow X$ 能由 Ω 根据 Armstrong 公理导出, 称 X 等价于 Y , 记做 $X \approx Y$ 。显然 $\approx$ 是等价关系。如果 $X = \{a\}, Y = \{b\}, X \approx Y$ 可以简记为 $a \approx b$, 称属性 a 等价于属性 b 。若 $X \approx Y$, 显然有 $X+ = Y+$; 反之亦然。

设 $A = \{1, 2, 3, 4\}, \Omega = \{1 \rightarrow 2, 1 \rightarrow 3, 1 \rightarrow 4, 4 \rightarrow 1\}, S = (A, \Omega)$ 是信息系统。因为 $1 \rightarrow 4, 4 \rightarrow 1$ 同时成立, 所以有 $1 \approx 4$, 计算 $1+ = A = 4+$ 。

定义 5(信息系统结构) $S = (A, \Omega)$ 是一个信息系统。 $\Omega+ = \{G | G = G+, G \subseteq A\}$, 称为信息系统 S 的结构, 简称信息系统结构。 $\forall G \in \Omega+, G$ 称为属性集信息粒, 简称信息粒。 S 上信息粒的数目 $\alpha_S = |\Omega+|$ 称为信息系统结构复杂度, 简称复杂度, α_S 也可以写作 $S(\alpha)$, 在 S 明确的情况下, 可以简记为 α 。称 $\gamma_S = \alpha_S / |2^A|$ 为信息系统结构相对复杂度, 简称相对复杂度, γ_S 也可以写作 $S(\gamma)$, 在 S 明确的情况下, 可以简记为 γ 。

设 $S = (A, \Omega)$ 是信息系统, $A = \{1, 2\}, \Omega = \{1 \rightarrow 2\}$ 。容易算出 $\Omega+ = \{\emptyset, 2, 12\}, \alpha = 3, \gamma = 3/4$ 。因为 $\Omega+$ 是 A 的幂集

的子集, $\Omega+ \subseteq P(A)$, 根据集合的包含于关系, $(\Omega+, \subseteq)$ 成为一个偏序集。

$\forall G_1, G_2 \in \Omega+, G_1 \vee G_2 = (G_1 \cup G_2)+ \in \Omega+, G_1 \wedge G_2 = (G_1 \cap G_2)+ \in \Omega+$ (性质 1(vii) 交封闭性), 所以 $\Omega+$ 是格。

定义 6(信息系统属性同构) 设信息系统 $S = (A_S, \Omega_S), T = (A_T, \Omega_T)$, 如果存在 A_S 到 A_T 的双射 h_A , 使得 $\forall X \subseteq A_S, h_A(X) = Y \subseteq A_T$, 同时 $h_A(X_S+) = Y_T+$ 。称 S 同构于 T , 记作 $S \cong T$ 。这里 X_S+ 表示属性集 X 关于函数依赖集 Ω_S 的闭包, Y_T+ 表示属性集 Y 关于函数依赖集 Ω_T 的闭包。 h_A 称为 S 到 T 的同构映射。显然, 若 h_A 是 S 到 T 的同构映射, 则 h_A^{-1} 是 T 到 S 的同构映射。如果 $S(A) = T(A)$, 称为信息系统自同构。

同构的两个信息系统, 它们的属性集一一对应, 并且各自的属性集信息粒也一一对应。

设 $S = (A_S, \Omega_S), A_S = \{1, 2, 3\}, \Omega_S = \{1 \rightarrow 2, 1 \rightarrow 3\}$ 。 $T = (A_T, \Omega_T), A_T = \{a_1, a_2, a_3\}, \Omega_T = \{a_1 \rightarrow a_2, a_1 \rightarrow a_3\}$ 。则存在 A_S 到 A_T 的双射 $h_A: i \rightarrow a_i, i = 1, 2, 3$, 使得 $S \cong T$ 。另外设 $R = (A_R, \Omega_R), A_R = \{1, 2, 3\}, \Omega_R = \{3 \rightarrow 1, 3 \rightarrow 2\}$ 。存在 A_S 到 A_R 的双射 $h_{A1}: 1 \rightarrow 3, 2 \rightarrow 1, 3 \rightarrow 2$, 使得 $S \cong R$ 。因为 $A_S = A_R$, 所以 h_{A1} 是信息系统自同构。

文献[14]中利用两个相同属性集上的闭包系统定义了数据库距离的概念, 并研究了数据库距离的性质, 本文应用此概念来度量不同信息系统结构之间的距离。

定义 7(信息系统结构距离^[14]) 具有相同属性集但函数依赖集不同的两个信息系统 $S = (A, \Omega_1), T = (A, \Omega_2), \Omega_1+$ 是 S 的系统结构, Ω_2+ 是 T 的系统结构, 则 S 与 T 信息系统结构间的距离定义为 $d(S, T) = d = |\Omega_1+ \setminus \Omega_2+ \cup \Omega_2+ \setminus \Omega_1+|$ 。 d 是信息系统 S 和 T 各自系统结构的对称差中元素的个数。

利用定义 7 可以比较两个信息系统结构的差异。

设 $S = (A, \Omega_1), T = (A, \Omega_2), A = \{1, 2, 3\}, \Omega_1 = \{1 \rightarrow 2, 1 \rightarrow 3\}, \Omega_2 = \{1 \rightarrow 2, 2 \rightarrow 3\}$ 。计算出 Ω_1+, Ω_2+ 分别为: $\Omega_1+ = \{\emptyset, 2, 3, 23, 123\}, \Omega_2+ = \{\emptyset, 3, 23, 123\}$ 。则 $d(S, T) = d = |\Omega_1+ \setminus \Omega_2+ \cup \Omega_2+ \setminus \Omega_1+| = 1$ 。

定义 8(信息系统属性同态) 设信息系统 $S = (A_S, \Omega_S), T = (A_T, \Omega_T)$, 令 $\varphi: A_S \rightarrow A_T$ 是一个从信息系统 S 的属性集到信息系统 T 的属性集的满射, 若 $\{y \leq X\}$ 或 $\{X \rightarrow y\} \subseteq \Omega_S$, 则 $\{\varphi(y) \leq \varphi(X)\}$ 或 $\{\varphi(X) \rightarrow \varphi(y)\} \subseteq \Omega_T$ 。称 φ 是信息系统 S 到 T 的一个属性同态。如果 $A_T \subseteq A_S$, 称 φ 是信息系统 S 的一个属性自同态。

例 1 $A = \{1, 2, 3, 4, 5\}, \Omega = \{1 \rightarrow 2, 2 \rightarrow 3, 14 \rightarrow 5\}$ 。容易算出 $\Omega+ = \{\emptyset, 3, 4, 5, 23, 34, 35, 45, 123, 234, 235, 345, 1235, 2345, A\}$ 。

$\alpha = |\Omega+| = 15, \gamma = 15/32$ 。定义 $\varphi_i: A \rightarrow A, i = 1, 2, 3$ 。

$\varphi_1(j) = j$ 对 $i = 1, 4, 5; \varphi_1(2) = \varphi_1(3) = 2, \Omega_1 = \{1 \rightarrow 2, 14 \rightarrow 5\}$

$\varphi_2(j) = j$ 对 $i = 1, 2, 4; \varphi_2(3) = \varphi_2(5) = 3, \Omega_2 = \{1 \rightarrow 2, 2 \rightarrow 3\}$

$\varphi_3(j) = j$ 对 $i = 2, 3, 5; \varphi_3(1) = \varphi_3(4) = 1, \Omega_3 = \{1 \rightarrow 2, 2 \rightarrow 3, 1 \rightarrow 5\}$

根据定义, 容易验证 φ_i 都是 A 上的属性自同态。

3 信息系统属性同态性质

若 φ 是信息系统 S 到 T 的属性同态, S 的系统结构 $\Omega_S +$ 与 T 的系统结构 $\Omega_T +$ 具有怎样的关系? 下面的定理 1 表明, 若 φ 是属性同态, 则存在 $\Omega_S +$ 与 $\Omega_T +$ 之间的并同态映射。根据信息系统属性同态的定义, 可以诱导出一个从原系统结构 $\Omega_S +$ 到像系统结构 $\Omega_T +$ 的映射 θ , θ 满足定理 1 的结论。

定理 1 若 φ 是 A_S 到 A_T 的属性同态, 定义 $\theta: \Omega_S + \rightarrow \Omega_T +$, $\theta(G) = \varphi(G) +$, $\forall G \in \Omega_S +$ 。

(i) θ 是 $\Omega_S +$ 到 $\Omega_T +$ 的序同态;

(ii) θ 是 $\Omega_S +$ 到 $\Omega_T +$ 的并同态。

证明: (i) 显然 $(\Omega_S +, \subseteq)$ 和 $(\Omega_T +, \subseteq)$ 是偏序集。 $\forall G_1, G_2 \in \Omega_S +$, 且 $G_1 \subseteq G_2$, 则 $\varphi(G_1) \subseteq \varphi(G_2)$ 。由 θ 的定义, $\theta(G_1) = \varphi(G_1) +$, $\theta(G_2) = \varphi(G_2) +$, 根据性质 1(ii), 因为 $\varphi(G_1) \subseteq \varphi(G_2)$, 所以 $\varphi(G_1) + \subseteq \varphi(G_2) +$ 。推出 $\theta(G_1) \subseteq \theta(G_2)$ 成立。

(ii) 根据定义, $\theta(G_1 \vee G_2) = \varphi((G_1 \cup G_2) +)$, $\theta(G_1) \vee \theta(G_2) = \varphi(G_1) + \vee \varphi(G_2) + = (\varphi(G_1) \cup \varphi(G_2) +) +$, 只需证明 $\varphi((G_1 \cup G_2) +) + = (\varphi(G_1) \cup \varphi(G_2) +) +$ 即可。

先证 $\varphi(G_1) \cup \varphi(G_2) + \subseteq \varphi((G_1 \cup G_2) +)$ 。

因为 $G_1 \subseteq G_1 \cup G_2 \subseteq (G_1 \cup G_2) +$, 所以 $\varphi(G_1) \subseteq \varphi((G_1 \cup G_2) +)$, 由性质 1(ii), $\varphi(G_1) + \subseteq \varphi((G_1 \cup G_2) +) +$; 同理 $\varphi(G_2) + \subseteq \varphi((G_1 \cup G_2) +) +$, 所以 $(\varphi(G_1) + \cup \varphi(G_2) +) \subseteq \varphi((G_1 \cup G_2) +) +$, 根据性质 1(vi), 直接得到 $(\varphi(G_1) + \cup \varphi(G_2) +) + \subseteq \varphi((G_1 \cup G_2) +) +$ 。

再证 $\varphi((G_1 \cup G_2) +) + \subseteq (\varphi(G_1) + \cup \varphi(G_2) +) +$ 。

设 $\forall y \in \varphi((G_1 \cup G_2) +)$, 存在 $x \in (G_1 \cup G_2) +$, 使 $\varphi(x) = y$, 根据属性集闭包定义 3, $x \subseteq (G_1 \cup G_2)$, 根据信息系统同态定义 8, $\varphi(x) \subseteq \varphi(G_1 \cup G_2) = \varphi(G_1) \cup \varphi(G_2)$, 因为 $\varphi(G_1) \subseteq \varphi(G_1) +$, 同时 $\varphi(G_2) \subseteq \varphi(G_2) +$, 所以 $\varphi(G_1) \cup \varphi(G_2) \subseteq \varphi(G_1) + \cup \varphi(G_2) +$, 于是 $y = \varphi(x) \subseteq \varphi(G_1) + \cup \varphi(G_2) +$, 所以有 $y \in (\varphi(G_1) + \cup \varphi(G_2) +) +$, 从而推出 $\varphi((G_1 \cup G_2) +) \subseteq (\varphi(G_1) + \cup \varphi(G_2) +) +$, 根据性质 1(vi), 推出 $\varphi((G_1 \cup G_2) +) + \subseteq (\varphi(G_1) + \cup \varphi(G_2) +) +$ 。

综上有 $\varphi((G_1 \cup G_2) +) = (\varphi(G_1) + \cup \varphi(G_2) +) +$, 即 $\theta(G_1 \vee G_2) = \theta(G_1) \vee \theta(G_2)$ 。

所以 θ 是 $\Omega_S +$ 到 $\Omega_T +$ 的并同态。证毕。

利用同态对信息系统进行压缩, 将一个大系统压缩成一个小系统, 是同态的一个重要应用^[8,10]。已有文献中的同态概念, 主要是指数据对象集(信息系统的行)的同态, 经过同态压缩后, 冗余的数据对象(行)被去除, 进而在较小的像系统中做属性约简等工作。本文所讨论的同态压缩, 则是直接压缩属性集(削减信息系统的列), 避免了大数据集情况下对大量数据对象(行)的处理所带来的较大的开销。

那么, 如何将冗余的属性从系统中移除且不改变系统的结构, 或者尽可能小地改变系统的结构, 就成了属性同态压缩下的一个关键问题。为此, 需要研究以下几个问题:

问题 1 什么条件下, 经同态压缩后的系统结构与压缩前的系统结构相比不变?

问题 2 什么条件下系统结构改变, 如何度量系统结构变化的程度?

问题 3 如何找到结构变化最小的属性同态?

由例 1 可以看出, 满足属性同态定义的映射很多, 但是哪一个才是比较“理想的”同态映射, 这正是下文要研究的内容。

下面的定理 2 给出问题 1 的答案, 将具有等价关系的属性“压缩”成一个属性, 得到的像系统结构与原系统结构同构, 即可认为系统结构未发生改变。

定理 2 设 $S = (A, \Omega)$, A 根据属性间的等价关系 $\approx$, 可以生成一个划分 $A/\approx$ 。定义映射 $\varphi: \varphi(a) = [a] \approx, \forall a \in A$ 。 $\forall a, b \in A$, 定义 $[a] \approx \subseteq [b] \approx$ iff $a \subseteq b$ 。令 $T = (\varphi(A), \varphi(\Omega))$, $\varphi(\Omega) = \{\varphi(X) \rightarrow \varphi(a) \mid X \rightarrow a \in \Omega\}$, 则:

(i) φ 是 S 到 T 的属性同态;

(ii) $\Omega + \cong \varphi(\Omega) +$ 。

证明: (i) 设 $\forall y \in A$, $\varphi(y) = [y] \approx$, 则 $\forall X \subseteq A$, $\varphi(X) = [X] \approx = \bigcup [x] \approx_{(x \in X)}$, 若 $y \subseteq X$, 有 $[y] \approx \subseteq [X] \approx$, 即 $\varphi(y) \subseteq \varphi(X)$, 所以 φ 是 S 到 T 的属性同态。

(ii) $\forall G \in \Omega +$, 令 $\varphi(\Omega +) = \{\varphi(G) = [G] \approx \mid G \in \Omega +\}$ 。我们先证明 $\varphi(\Omega +) = \varphi(\Omega) +$, 然后再证明 $\Omega + \cong \varphi(\Omega) +$ 。

为了证明 $\varphi(\Omega +) = \varphi(\Omega) +$, 只需证明 (a) $\forall Y \subseteq \varphi(A)$, $Y + \in \varphi(\Omega) +$, 有 $Y + \in \varphi(\Omega +)$ 成立; 同时 (b) $\forall \varphi(G) \in \varphi(\Omega +)$, $\exists Y \subseteq \varphi(A)$, 使得 $Y + = \varphi(G) \in \varphi(\Omega) +$ 。

先证明 (a)。 $\forall Y \subseteq \varphi(A)$, 设 X 是 Y 在 φ 下的一个原像。即 $\varphi(X) = Y$, 令 $X + = G$ 。下面证明 $Y + = \varphi(G)$ 。首先, $\forall y \in Y +$, 有 $Y \rightarrow y$, 设 $x \in \varphi^{-1}(y)$, 即 $\varphi(x) = y$, 根据定理 2 的条件, 有 $X \rightarrow x$, 于是 $x \in X + = G$, 推出 $y = \varphi(x) \in \varphi(G)$, $Y + \subseteq \varphi(G)$; 其次 $\forall y \in \varphi(G)$, $\exists x \in G = X +$, $\varphi(x) = y$, $X \rightarrow x$, 因为 $\varphi(X) = Y$, 所以有 $Y \rightarrow y$, 于是 $y \in Y +$, 即 $\varphi(G) \subseteq Y +$, 所以有 $Y + = \varphi(G) \in \varphi(\Omega +)$ 。(a) 得证。

再证明 (b)。 $\forall \varphi(G) \in \varphi(\Omega +)$, 因为 G 是 A 中的信息粒, 所以存在 $X \subseteq A$, 使得 $G = X +$, 令 $Y = \varphi(X)$, 根据前面 (a) 的证明过程可知 $Y + = \varphi(G) \in \varphi(\Omega) +$ 。(b) 得证。

由 (a)、(b) 知 $\varphi(\Omega +) = \varphi(\Omega) +$ 。

现在证明 $\Omega +$ 与 $\varphi(\Omega +)$ 是格同构, $\Omega + \cong \varphi(\Omega +)$ 。根据定义 5, $\Omega +$ 与 $\varphi(\Omega +)$ 是格。为证明 $\Omega +$ 与 $\varphi(\Omega +)$ 是格同构, 先证明 φ 是 $\Omega +$ 到 $\varphi(\Omega +)$ 的格同态, 再证明 φ 是 $\Omega +$ 到 $\varphi(\Omega +)$ 的双射。

设 $\forall G_1, G_2 \in \Omega +$, $\varphi(G_1 \vee G_2) = \varphi((G_1 \cup G_2) +)$, $\varphi(G_1) \vee \varphi(G_2) = (\varphi(G_1) \cup \varphi(G_2) +) +$ 。设 $\forall y \in \varphi((G_1 \cup G_2) +)$, $\exists x \in (G_1 \cup G_2) +$, 使得 $\varphi(x) = y$, 且 $(G_1 \cup G_2) \rightarrow x$, 于是 $\varphi(G_1 \cup G_2) \rightarrow y$, 而 $\varphi(G_1 \cup G_2) = \varphi(G_1) \cup \varphi(G_2)$, 所以 $y \in (\varphi(G_1) \cup \varphi(G_2) +) +$, 即 $\varphi(G_1 \vee G_2) \subseteq \varphi(G_1) \vee \varphi(G_2)$; 反过来, 设 $\forall y \in (\varphi(G_1) \cup \varphi(G_2) +) +$, 则有 $\varphi(G_1) \cup \varphi(G_2) \rightarrow y$, 即 $\varphi(G_1 \cup G_2) \rightarrow y$, 于是 $\exists y$ 的原像 x 使得 $\varphi(x) = y$, 且 $(G_1 \cup G_2) \rightarrow x$, 所以 $x \in (G_1 \cup G_2) +$, 推出 $y = \varphi(x) \in \varphi((G_1 \cup G_2) +)$, 即 $\varphi(G_1) \cup \varphi(G_2) \subseteq \varphi(G_1 \vee G_2)$, 所以 φ 是 $\Omega +$ 到 $\varphi(\Omega +)$ 的并同态。

由性质 1(vii), $G_1 \wedge G_2 = G_1 \cap G_2$, $\varphi(G_1 \wedge G_2) = \varphi(G_1 \cap G_2)$, 同时 $\varphi(G_1) \wedge \varphi(G_2) = \varphi(G_1) \cap \varphi(G_2)$ 。因为 $G_1 \cap G_2 \subseteq G_1$, 所以 $\varphi(G_1 \cap G_2) \subseteq \varphi(G_1)$, 同理 $\varphi(G_1 \cap G_2) \subseteq \varphi(G_2)$, 于是推出 $\varphi(G_1 \cap G_2) \subseteq \varphi(G_1) \cap \varphi(G_2)$, 即 $\varphi(G_1 \wedge G_2) \subseteq \varphi(G_1) \wedge \varphi(G_2)$; 设 $\forall y \in \varphi(G_1) \cap \varphi(G_2)$, $\exists x_1 \in G_1, x_2 \in G_2$, 使得 $\varphi(x_1) = y$, $\varphi(x_2) = y$, 由 φ 的定义知, $x_1 \approx x_2$, 于是 $x_1 \in G_2, x_2 \in G_1, x_1, x_2 \in G_1 \cap G_2$, 即 $y \in \varphi(G_1 \cap G_2)$, 推出 $\varphi(G_1) \cap \varphi(G_2) \subseteq \varphi(G_1 \cap G_2)$, 即 $\varphi(G_1) \wedge \varphi(G_2) \subseteq \varphi(G_1 \wedge G_2)$, 所以 φ 是 $\Omega +$ 到 $\varphi(\Omega +)$ 的交同态。

φ 既是并同态, 又是交同态, 所以 φ 是格同态。

根据 $\varphi(\Omega +) = \{\varphi(G) \mid G \in \Omega +\}$ 知 φ 是 $\Omega +$ 到 $\varphi(\Omega +)$ 的

满射,现在证明 φ 是 $\Omega+$ 到 $\varphi(\Omega+)$ 的单射。设 $\forall G_1, G_2 \in \Omega+, G_1 \neq G_2$, 不妨设 $x \in G_1 \setminus G_2$ (若 $G_1 \setminus G_2$ 为空集,则可设 $x \in G_2 \setminus G_1$), 令 $\varphi(x) = y$, 则 $y \in \varphi(G_1)$, 现在证明 $y \notin \varphi(G_2)$ 。反证,若 $y \in \varphi(G_1)$ 同时 $y \in \varphi(G_2)$, 根据前面交同态的证明过程,可以推出 $x \in G_2$, 矛盾,所以 $\varphi(G_1) \neq \varphi(G_2)$, φ 是 $\Omega+$ 到 $\varphi(\Omega+)$ 的单射。 φ 是满射,又是单射,所以 φ 是双射。 φ 是格同态,又是双射,所以 φ 是格同构。证毕。

例2 设 $A = \{1, 2, 3, 4\}$, $\Omega = \{1 \rightarrow 2, 1 \rightarrow 3, 1 \rightarrow 4, 4 \rightarrow 1\}$, 则 $\Omega+ = \{\emptyset, 2, 3, 23, 1234\}$ 。因为 $1 \approx 4$, 定义 $\varphi: A \rightarrow A$, $\varphi(1) = \varphi(4) = 1, \varphi(2) = 2, \varphi(3) = 3$, 此时 $\varphi(A) = \{1, 2, 3\}$, $\varphi(\Omega) = \{1 \rightarrow 2, 1 \rightarrow 3\}$ 。 $\varphi(\Omega+) = \varphi(\Omega+) = \{\emptyset, 2, 3, 23, 123\}$ 。图1是 $\Omega+$ 和 $\varphi(\Omega+)$ 各自的哈斯图,从图中容易看出 $\Omega+$ 与 $\varphi(\Omega+)$ 同构。

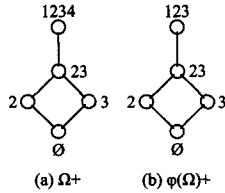


图1 $\Omega+$ 和 $\varphi(\Omega+)$ 的哈斯图

定义9 设 $S = (A, \Omega)$, 如果存在 S 上的属性同态 φ 使得 $\Omega+ \cong \varphi(\Omega+)$, 则称 φ 是信息系统的一个理想同态。

理想同态,对于信息系统属性集做了无损压缩,即没有改变原信息系统的结构。根据定理2知道,理想同态存在的条件是属性集 A 中至少存在一对属性 a, b 并且 $a \approx b$, 如果系统中不存在这样相互等价的属性,则同态像系统的结构跟原系统结构比较会发生改变。下面回答前面提出的问题2,如何度量系统结构变化的程度。

4 属性同态下信息系统结构变化的度量

实际中的信息系统未必存在理想的属性同态,这时判断信息系统的一个同态性质如何,可以考虑该自同态的像信息系统结构与原信息系统结构间的改变程度,根据这个变化程度来度量自同态的理想程度。

根据定义7,可以度量具有相同属性集而函数依赖集不同的两个信息系统之间的距离。但是同态像系统属性集的属性个数少于原信息系统属性集的属性个数(压缩),属性数目不一样,无法直接利用定义7度量它们之间的系统结构距离。

重新考察例1中的同态映射 φ_1 , 如果增加函数依赖 $3 \rightarrow 2$ 到 Ω 中,使 Ω 变为 Ω_m , 记此时的信息系统为 $T(A, \Omega_m)$, $\Omega_m = \{1 \rightarrow 2, 2 \rightarrow 3, 3 \rightarrow 2, 14 \rightarrow 5\}$, 显然 $2 \approx 3$, 而 $\varphi_1(A) = \{1, 2, 4, 5\}$, $\varphi_1(\Omega) = \varphi_1(\Omega_m) = \{1 \rightarrow 2, 14 \rightarrow 5\}$, 根据定理2, $T(A, \Omega_m)$ 的系统结构 Ω_m+ 与 $\varphi_1(\Omega+)$ 格同构。因为同构的系统可以看作相同系统的不同形式,所以度量 φ_1 的像系统 $R(\varphi_1(A), \varphi_1(\Omega))$ 与原系统 $S(A, \Omega)$ 的结构距离,等价于度量 $T(A, \Omega_m)$ 与原系统 $S(A, \Omega)$ 的结构距离,而 $T(A, \Omega_m)$ 与 $S(A, \Omega)$ 具有相同的属性集,可以根据定义7直接度量两者之间的结构距离 $d(S, T)$ 。

上面的情况容易推广到任意的属性同态。如果 φ 是信息系统 $S(A, \Omega)$ 到 $R(\varphi(A), \varphi(\Omega))$ 的属性同态, $\forall y \in \varphi(A)$, 令 $X_y = \{x | \varphi(x) = y\}$ 是 y 的原像的集合,令 $\Omega_y = \bigcup \{x_i \rightarrow x_j, x_j \rightarrow x_i | (\forall x_i, x_j \in X_y, x_i \neq x_j)\}$, $\Delta\Omega = \bigcup \Omega_y (\forall y \in \varphi(A))$, $\Omega_m = \Omega \cup \Delta\Omega$, 则根据定理2有 $\Omega_m \cong \varphi(\Omega+)$ 。再根据定义9, φ 是

$T(A, \Omega_m)$ 到 $R(\varphi(A), \varphi(\Omega))$ 的理想同态,为了度量 S 与 R 之间的结构距离,只需要度量 S 与 T 之间的距离即可。

下面定理3给出当信息系统 $S(A, \Omega)$ 增加特定函数依赖成为 $T(A, \Omega_m)$ 时,其系统结构的变化规律。

定理3 信息系统 $S(A, \Omega)$ 中,设 $A = \{1, 2, \dots, n\}$, $i, j \in A$, 令 $\Omega_m = \Omega \cup \{i \rightarrow j\}$, 增加函数依赖后的信息系统为 $S_n = (A, \Omega_m)$, 则

- (i) $\Omega_m+ \subseteq \Omega+$;
- (ii) $S_n(a) \subseteq S(a)$;
- (iii) $\forall X \subseteq A, \Omega(X+) \subseteq \Omega_m(X+)$;
- (iv) 定义从 $\Omega+$ 到 Ω_m+ 的映射 $\varphi: \Omega(X+) \rightarrow \Omega_m(X+)$, $\forall X \subseteq A$, 则 φ 是并同态映射。

证明:先证明(i), 设 $\forall G \in \Omega_m+$, 如果 $i \notin G$, 有 $\Omega(G+) = \Omega_m(G+) = G$, 所以 $G \in \Omega+$; 如果 $i \in G$, 因为 $i \rightarrow j \in \Omega_m$, 必然有 $j \in G$, 于是有 $\Omega(G+) = \Omega_m(G+) = G$, 推得 $G \in \Omega+$ 。因此(i)成立。

由(i)直接推得(ii)成立。

(iii) 设 $\forall x \in \Omega(X+)$, 则在 S 中有 $X \rightarrow x$ 成立, 即 $X \rightarrow x$ 可以由 Ω 导出, 而 $\Omega_m = \Omega \cup \{i \rightarrow j\}$, 所以 $X \rightarrow x$ 必然可以由 Ω_m 导出, 即 $x \in \Omega_m(X+)$, 因此(iii)成立。

(iv) 设 $\forall X, Y \subseteq A$, 令 $\Omega(X+) = G_1, \Omega(Y+) = G_2, \varphi(G_1 \vee G_2) = \varphi(\Omega((G_1 \cup G_2)+)) = \Omega_m((G_1 \cup G_2)+), \varphi(G_1) \vee \varphi(G_2) = \Omega_m((\varphi(G_1) \cup \varphi(G_2))+)$, 只需证明 $\Omega_m((G_1 \cup G_2)+) = \Omega_m((\varphi(G_1) \cup \varphi(G_2))+)$ 即可。

设 $\forall G \in \Omega+$, 根据 G 是否含 i 且不含 j , 可以将其中的信息粒分为两类。(a) 如果 $i \notin G$, 或者 $i \in G$ 同时 $j \in G$, 此时 $i \rightarrow j$ 在 G 中没有作用, G 在 φ 下的像等于 G , $\varphi(G) = \Omega_m(G+) = \Omega(G+) = G$; (b) 如果 $i \in G$ 且 $j \notin G$, 此时 $\varphi(G) = \Omega_m(G+)$, 因为 $i \rightarrow j \in \Omega_m$, 根据性质1(v)有 $\Omega_m(G+) = \Omega_m((G \cup \{j\})+)$, 此时 $i, j \in G_j$, 所以有 $i, j \in \Omega(G_j+)$, G_j 是 $G \cup \{j\}$ 的简写, 与情况(a)类似, 有 $\Omega(G_j+) = \Omega_m(G_j+)$ 。于是 $\varphi(G) = \Omega(G_j+)$ 。根据 G 的像是其自身(情况(a)), 或 G 的像是 $\Omega(G_j+)$ (情况(b)), 可以将 G_1, G_2 分为4种情况分别证明 $\Omega_m((G_1 \cup G_2)+) = \Omega_m((\varphi(G_1) \cup \varphi(G_2))+)$ 成立。

1. G_1, G_2 都属于情况(a), 即 $\varphi(G_1) = G_1, \varphi(G_2) = G_2$, 所以 $\Omega_m((\varphi(G_1) \cup \varphi(G_2))+) = \Omega_m((G_1 \cup G_2)+)$, 于是(iv)成立。

2. G_1 属于情况(a), G_2 属于情况(b), 即 $\varphi(G_1) = G_1, \varphi(G_2) = \Omega(G_2 j+)$, 等式右边 = $\Omega_m((\varphi(G_1) \cup \varphi(G_2))+) = \Omega_m((G_1 \cup \Omega(G_2 j+))+)$, 此时 $i, j \in \Omega((G_1 \cup \Omega(G_2 j+))+)$ 属于情况(a), 所以 $\Omega_m((G_1 \cup \Omega(G_2 j+))+) = \Omega((G_1 \cup \Omega(G_2 j+))+)$; 而等式左边 = $\Omega_m((G_1 \cup G_2)+)$, 因为 $i \in G_2$, 所以 $i \in G_1 \cup G_2$, 若 $j \notin G_1 \cup G_2$, $\Omega_m((G_1 \cup G_2)+) = \Omega_m((G_1 \cup G_2 \cup \{j\})+)$, 因为 $i, j \in G_1 \cup G_2 \cup \{j\}$, 所以 $\Omega_m((G_1 \cup G_2 \cup \{j\})+) = \Omega((G_1 \cup G_2 \cup \{j\})+)$ 。若 $j \in G_1 \cup G_2$, $\Omega_m((G_1 \cup G_2)+) = \Omega((G_1 \cup G_2)+) = \Omega((G_1 \cup G_2 \cup \{j\})+)$, 即等式左边 = $\Omega((G_1 \cup G_2 \cup \{j\})+)$ 。于是只需证明 $\Omega((G_1 \cup G_2 \cup \{j\})+) = \Omega((G_1 \cup \Omega(G_2 j+))+)$ 即可。

为证明等式左右两边相等, 只需要证明左边和右边相互包含即可。因为 $G_2 \cup \{j\} = G_2 j \subseteq \Omega(G_2 j+)$, 所以 $G_1 \cup G_2 \cup \{j\} \subseteq G_1 \cup \Omega(G_2 j+)$, 根据性质1(ii), 有 $\Omega((G_1 \cup G_2 \cup \{j\})+) \subseteq \Omega((G_1 \cup \Omega(G_2 j+))+)$; 反过来, 因 $G_1 \subseteq \Omega((G_1 \cup G_2 \cup \{j\})+)$

$+$), $\Omega(G_2j+) \subseteq \Omega((G_1 \cup G_2 \cup \{j\})+)$, 故 $G_1 \cup (G_2j+) \subseteq \Omega((G_1 \cup G_2 \cup \{j\})+)$, 根据性质 1(vi), $\Omega((G_1 \cup \Omega(G_2j+)) +) \subseteq \Omega((G_1 \cup G_2 \cup \{j\})+)$ 。等式两边相互包含, 所以等式左边等于右边成立, 即 $\Omega_m((G_1 \cup G_2) +) = \Omega_m((\varphi(G_1) \cup \varphi(G_2)) +)$, (iv) 成立。

3. G_2 属于情况(a), G_1 属于情况(b), 证明过程与 2 类似, 有(iv)成立。

4. G_1, G_2 都属于情况(b), 即 $\varphi(G_1) = \Omega(G_1j+)$, $\varphi(G_2) = \Omega(G_2j+)$, 于是等式右边 $= \Omega_m((\varphi(G_1) \cup \varphi(G_2)) +) = \Omega_m((\Omega(G_1j+) \cup \Omega(G_2j+)) +) = \Omega((\Omega(G_1j+) \cup \Omega(G_2j+)) +)$; 等式左边 $= \Omega_m((G_1 \cup G_2) +)$, $i \in G_1 \cup G_2, j \notin G_1 \cup G_2$, $\Omega_m((G_1 \cup G_2) +) = \Omega_m((G_1 \cup G_2 \cup \{j\}) +) = \Omega((G_1 \cup G_2 \cup \{j\}) +)$, 于是只需证明 $\Omega((G_1 \cup G_2 \cup \{j\}) +) = \Omega((\Omega(G_1j+) \cup \Omega(G_2j+)) +)$ 即可。利用属性集闭包性质 1(ii) 和 (vi), 仿照 2 证明过程容易证得。

综合以上 4 种情况, 有 $\varphi(G_1 \vee G_2) = \varphi(G_1) \vee \varphi(G_2)$ 成立, 即 φ 是 Ω 到 Ω_m 的并同态映射。

说明: 定理 3 表明当信息系统中增加函数依赖时, 信息系统结构复杂度降低, 同时新的信息系统结构是原信息系统结构的子集和并同态像。

推论 1 信息系统 $S = (A, \Omega)$ 中, 设 $A = \{1, 2, \dots, n\}$, 令 $\Omega_m = \Omega \cup F$, $F = \{X \rightarrow a \mid Xa \subseteq A\}$, 增加函数依赖集 F 后的信息系统为 $S_n = (A, \Omega_m)$, 则定理 3 的结论(i)、(ii)、(iii)、(iv)成立。

根据前面的分析, 给定一个信息系统 $S(A, \Omega)$ 到 $R(\varphi(A), \varphi(\Omega))$ 的属性同态 φ , 存在一个信息系统 $T(A, \Omega_m)$ 与 $R(\varphi(A), \varphi(\Omega))$ 的系统结构格同构, 所以度量 S 与 R 的结构距离等价于度量 S 与 T 的结构距离。即 $d(S, R) = d(S, T) = |\Omega + \setminus \Omega_m + \cup \Omega_m + \setminus \Omega +|$ 。而根据定理 3 和推论 1, $\Omega_m \subseteq \Omega +$, 所以 $\Omega_m + \setminus \Omega + = \emptyset$, 于是推出 $d(S, T) = |\Omega + \setminus \Omega_m +| = |\Omega +| - |\Omega_m +|$, 而 $|\Omega_m +| = |\varphi(\Omega) +|$, 所以有 $d(S, T) = |\Omega +| - |\varphi(\Omega) +|$ 。这样就回答了前面提出的问题 2。

例 3 $A = \{1, 2, 3, 4, 5\}$, $\Omega = \{1 \rightarrow 2, 2 \rightarrow 3, 14 \rightarrow 5\}$ 。 $\Omega + = \{\emptyset, 3, 4, 5, 23, 34, 35, 45, 123, 234, 235, 345, 1235, 2345, A\}$ 。 $\varphi_i (i = 1, 2, 3)$ 的定义同例 1, 下面分别计算 $d(S, R_i)$ 。

根据例 1, $S(A, \Omega)$ 在 φ_1 下的像信息系统为 $R_1(\varphi_1(A), \varphi_1(\Omega))$, $\varphi_1(A) = \{1, 2, 4, 5\}$, $\varphi_1(\Omega) = \{1 \rightarrow 2, 14 \rightarrow 5\}$ 在 Ω 中新增加函数依赖 $3 \rightarrow 2$ 得到信息系统 $T_1(A, \Omega_{m1})$, 这里 $\Omega_{m1} = \{1 \rightarrow 2, 2 \rightarrow 3, 3 \rightarrow 2, 14 \rightarrow 5\}$ 。根据定理 2, 容易验证 φ_1 是 $T_1(A, \Omega_{m1})$ 到 $R_1(\varphi_1(A), \varphi_1(\Omega))$ 的理想同态, 分别计算出 $\Omega_{m1} +$ 和 $\varphi_1(\Omega) +$ 为: $\Omega_{m1} + = \{\emptyset, 4, 5, 23, 45, 123, 234, 235, 1235, 2345, 12345\}$, $\varphi_1(\Omega) + = \{\emptyset, 4, 5, 2, 45, 12, 24, 25, 125, 245, 1245\}$ 。

容易看出 $\Omega_{m1} +$ 和 $\varphi_1(\Omega) +$ 同构, 并且 $\Omega_{m1} \subseteq \Omega +$, 于是有: $d(S, R_1) = d(S, T_1) = |\Omega +| - |\Omega_{m1} +| = |\Omega +| - |\varphi_1(\Omega) +| = 15 - 11 = 4$ 。

同理可计算出:

$$\varphi_2(\Omega) + = \{\emptyset, 4, 3, 23, 34, 123, 234, 1234\}$$

$$d(S, R_2) = |\Omega +| - |\varphi_2(\Omega) +| = 15 - 8 = 7$$

$$\varphi_3(\Omega) + = \{\emptyset, 3, 5, 23, 35, 235, 1235\}$$

$$d(S, R_3) = |\Omega +| - |\varphi_3(\Omega) +| = 15 - 7 = 8$$

计算结果表明, φ_1 是 $\varphi_i (i = 1, 2, 3)$ 中最理想的属性同态。

5 信息系统属性同态的应用

5.1 应用实例

下面通过一个实例说明信息系统属性同态在数据压缩中可能的应用。

例 4 表 1 是一个学生信息系统 $S(A, \Omega)$ 。其中:

$A = \{\text{学号}, \text{所在系}, \text{系地址}, \text{系电话}, \text{课程号}, \text{成绩}\}$

$\Omega = \{\text{学号} \rightarrow \text{所在系}, \text{所在系} \rightarrow \text{系地址}, \text{所在系} \rightarrow \text{系电话}, \text{系地址} \rightarrow \text{系电话}, \text{系电话} \rightarrow \text{系地址}, (\text{学号}, \text{课程号}) \rightarrow \text{成绩}\}$

表 1 学生信息系统 $S(A, \Omega)$

学号	所在系	系地址	系电话	课程号	成绩
2014120301	信息系	管理学院 443	88431703	120321	84
2014120201	工商系	管理学院 343	88431702	120322	75
2014120306	信息系	管理学院 443	88431703	120321	82
...	...	...	...	...	...

因为系地址 $\rightarrow$ 系电话, 系电话 $\rightarrow$ 系地址, 所以系电话与系地址是等价属性。根据定义 9 和定理 2, 若 φ_0 将系地址和系电话映射到其中的一个属性, 假设映射为系地址, 其他属性映射为自身, 则 φ_0 是一个理想自同态。设学生信息系统 $S(A, \Omega)$ 在 φ_0 下的像系统为 $R_0(A_0, \Omega_0)$, 则:

$A_0 = \varphi_0(A) = \{\text{学号}, \text{所在系}, \text{系地址}, \text{课程号}, \text{成绩}\}$

$\Omega_0 = \varphi_0(\Omega) = \{\text{学号} \rightarrow \text{所在系}, \text{所在系} \rightarrow \text{系地址}, (\text{学号}, \text{课程号}) \rightarrow \text{成绩}\}$

$R_0(A_0, \Omega_0)$ 是将表 1 的系电话所在列删除后的表。这相当于对原信息系统 $S(A, \Omega)$ 做了无损压缩(信息系统结构未改变)。如果需要进一步压缩为 4 个属性, 根据例 3 的计算过程, 可以计算出 φ_1 的像系统 $R_1(A_1, \Omega_1)$ 是对 S 的结构改变最小的属性同态, 这里 $\varphi_1(\text{学号}) = \text{学号}$, $\varphi_1(\text{所在系}) = \varphi_1(\text{系地址}) = \text{所在系}$, $\varphi_1(\text{课程号}) = \text{课程号}$, $\varphi_1(\text{成绩}) = \text{成绩}$ 。

$A_1 = \varphi_1(A) = \{\text{学号}, \text{所在系}, \text{课程号}, \text{成绩}\}$

$\Omega_1 = \varphi_1(\Omega) = \{\text{学号} \rightarrow \text{所在系}, (\text{学号}, \text{课程号}) \rightarrow \text{成绩}\}$

$R_1(A_1, \Omega_1)$ 是将表 1 的系电话和系地址所在列删除后得到的表。

5.2 现有信息系统同态方法的比较

1. 近年来对信息系统同态的研究^[4-11]都是建立在对论域对象粒化的基础上, 其同态像可以看作是对原系统论域对象(数据行)的压缩, 可以将这些同态称为对象同态; 本文所提信息系统属性同态则是建立在对属性粒化的基础上, 其同态像可以看作是对原系统属性(数据列)的压缩。

对象同态是根据论域对象在各个属性上的取值, 将信息系统看作是一族划分, 或者一族覆盖, 通过定义从原系统到像系统的论域对象间的一致函数或者协调函数来研究信息系统之间同态的性质。其实质是通过论域对象的粒化, 将一个大的原信息系统压缩为一个小的像信息系统。

属性同态则是根据属性间的函数依赖关系对信息系统属性集进行粒化, 将属性集分为若干“属性分块”, 每个“属性分块”映射为一个新属性, 从而减少像系统的属性数目, 以达到数据压缩的效果。属性理想同态是一种特殊的划分, 要求同一分块中的属性相互等价。

2. 对象同态的像系统能够保持原系统的某些重要特征不变, 比如属性约简、属性核等, 因此可以利用像系统做属性约简, 从而得到原系统的约简; 属性同态的像系统则保持原系统的信息系统结构不变(理想同态)或者信息系统结构变化最小

(近似理想同态)。信息系统结构可以看作是信息系统的语义结构。

3. 计算对象同态的像系统,需要对全体论域对象做扫描比较,其开销会随着信息系统数据量的增加而增加;计算属性同态的像系统,需要计算各种像系统的信息系统结构,以找到近似理想同态,但不需要扫描论域对象,其开销取决于系统属性的个数。一般来说,与论域对象相比较,信息系统的属性较少且更稳定(不易改变),而对象数目会动态增加,变化范围很大,因此对于数据量很大的信息系统,利用属性同态做数据压缩在计算开销上会优于对象同态。

由例4可以看出,应用信息系统属性同态的概念实现属性压缩,其过程与信息系统中论域对象的数目(行数)无关,只需考虑属性上的函数依赖以及由其决定的信息系统结构。当信息系统论域对象的数目较大时,特别是在大数据的环境下,该方法更具有优势。

结束语 本文根据属性集上的函数依赖关系定义了信息系统属性同构与信息系统属性同态的概念。一般而言,信息系统中的论域对象(行)数目远远大于属性(列)的数目,利用本文所提出的属性同态的方法能够直接对信息系统属性集进行压缩,因为没有考虑信息系统的其他要素,并且不涉及对论域对象的扫描比较,某种程度上简化了对信息系统同态问题的研究。

本文通过第3、4节的工作给出了信息系统理想同态存在的条件以及任意属性同态理想程度的度量方法,第5节给出了应用实例。进一步的工作可以更加深入地研究属性同态的性质;将信息系统属性同态与属性约简进行比较研究。

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 homomorphisms[J]. Information Sciences, 2000, 129(1-4): 211-220
- [3] Qu Kai-she, Zhai Yan-hui, Li De-yu, et al. Some Characters of Information System Homomorphism and Invariance of Upper and Lower Approximations[J]. Computer Science, 2005, 32(12): 168-174(in Chinese)
- [4] 曲开社, 翟岩慧, 李德玉, 等. 信息系统同态的性质及上下近似的不变性[J]. 计算机科学, 2005, 32(12): 168-174
- [5] Wang Chang-zhong, Wu Cong-xin, Chen De-gang, et al. Some properties of relation information systems under homomorphisms[J]. Appl. Math. Lett., 2008, 21(9): 940-945
- [6] Wang Chang-zhong, Wu Cong-xin, Chen De-gang, et al. Communicating between information systems[J]. Information Sciences, 2013, 178(16): 3228-3239
- [7] Zhu Ping, Wen Qiao-yan. Some improved results on communication between information systems[J]. Information Sciences, 2010, 180(18): 3521-3531
- [8] Wang Chang-zhong, Chen De-gang, Zhu Liang-kuan. Homomorphisms between fuzzy information systems[J]. Appl. Math. Lett., 2009, 22(7): 1045-1050
- [9] Zhu Ping, Wen Qiao-yan. Homomorphisms between fuzzy information system revisited[J]. Appl. Math. Lett., 2011, 24(19): 1548-1553
- [10] 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(4): 519-525
- [11] Wang Chang-zhong, Chen De-gang, Sun Bai-qing, et al. Communication between information systems with covering based rough sets[J]. Information Sciences, 2012, 216(24): 17-33
- [12] Liu Xiao-juan, Mi Ju-sheng, Li Zhong-ling. Date Compression with Homomorphism in Neighborhood Information Systems[J]. Computer Science, 2012, 39(11): 230-232(in Chinese)
- [13] 刘晓娟, 米据生, 李仲玲. 邻域信息系统基于同态的数据压缩[J]. 计算机科学, 2012, 39(11): 230-232
- [14] Armstrong W W. Dependency structures of data base relationships[C]// Proceedings of IFIP Congress. North-Holland, Amsterdam, 1974: 580-583
- [15] Mora A, Cordero P, Enciso M, et al. Closure via functional dependence simplification[J]. International Journal of Computer Mathematics, 2012, 89(4): 510-526
- [16] Katona G O H, Sali A. On the distance of databases[J]. Annals of Mathematics and Artificial Intelligence, 2012, 65(2/3): 199-216
- [17] (上接第37页)
- [18] Park C-H, Lee J-J, Smith M J, et al. Singular point detection by shape analysis of directional fields in fingerprints[J]. Pattern Recognition, 2006, 39(5): 839-855
- [19] Chua S C, Wong E K, Tan C, et al. Singular point detection in fingerprint images: An investigation on quantization approach[C]// 2014 IEEE Region 10 Symposium. 2014: 606-611
- [20] Bahgat G, Khalil A, Abdel Kader N, et al. Fast and accurate algorithm for core point detection in fingerprint images[J]. Egyptian Informatics Journal, 2013, 14(1): 15-25
- [21] Ma J, Jun Jing X, Zhang B, et al. An effective algorithm for fingerprint reference point detection[C]// 2010 2nd International Conference on Advanced Computer Control (ICACC). 2010: 200-203
- [22] Liu M, Jiang X, Kot A C. Fingerprint Reference-Point Detection[J]. EURASIP J. Adv. Sig. Proc., 2005, 2005(4): 498-509
- [23] Le T H, Van H T. Fingerprint reference point detection for image retrieval based on symmetry and variation[J]. Pattern Recognition, 2012, 45(9): 3360-3372
- [24] Tams B. Absolute fingerprint pre-alignment in minutiae-based cryptosystems[C]// 2013 International Conference of the Biometrics Special Interest Group (BIOSIG). 2013: 1-12
- [25] Areekul V, Boonchaiseree N. Fast focal point localization algorithm for fingerprint registration[C]// 3rd IEEE Conference on Industrial Electronics and Applications, 2008 (ICIEA 2008). 2008: 2089-2094
- [26] Yin J, Zhu E, Yang X, et al. Two steps for fingerprint segmentation[J]. Image and Vision Computing, 2007, 25(9): 1391-1403
- [27] Zhu E, Hancock E, Yin J, et al. Fusion of multiple candidate orientations in fingerprints[M]// Image Analysis and Recognition. Springer, 2011: 89-100