



PHỤ THUỘC HÀM TRONG HỆ THỐNG THÔNG TIN VÀ CÁC TÍNH CHẤT CỦA XẤP XỈ TRÊN DỰA VÀO HÀM ĐÓNG

Trịnh Thị Nhị, Nguyễn Bá Tường

Trường Đại học Sư phạm Kỹ thuật Hưng Yên

Ngày nhận: 11/5/2016

Ngày sửa chữa: 03/6/2016

Ngày xét duyệt: 20/6/2016

Tóm tắt:

Trong bài này chúng tôi trình bày một số tính chất của sự phụ thuộc của thuộc tính trong một hệ thống thông tin. Chúng tôi đã đề xuất lập biểu đồ đóng và từ đó chỉ ra một số tính chất và công thức xấp xỉ trên cho một hệ thống thông tin.

Đồng thời, trong bài viết này chúng tôi đã chỉ ra được hệ thống thông tin đơn trị và hệ thống thông tin đa trị xác định.

Từ khóa: Phụ thuộc hàm, hệ thống thông tin, xấp xỉ trên, hàm đóng, hệ quyết định.

Mở đầu

Trong [15, 16] chúng ta đã biết rằng, mỗi quan hệ là một hệ thống thông tin theo định nghĩa của Z.Pawlak [2]. Tuy nhiên, mỗi hệ thống thông tin trong [2] có thể không là quan hệ như trong [15, 16].

Vì vậy khái niệm phụ thuộc hàm được định nghĩa như trong [15, 16] nói chung không thể dùng trong hệ thống thông tin. Trong bài này, chúng tôi trình bày khái niệm phụ thuộc hàm dựa vào quan hệ bất khả phân biệt. Từ định nghĩa phụ thuộc hàm, chúng tôi nêu một số tính chất liên quan. Trong bài viết chúng tôi cũng nêu định nghĩa hàm đóng và từ tính chất hàm đóng, chúng tôi có đẳng thức kép giữa các tập xấp xỉ trên và một số tính chất của xấp xỉ trên.

1. Một số khái niệm cơ bản

Định nghĩa 1. Hệ thống thông tin

Hệ thống thông tin (information system) là $S = (U, A)$; trong đó U là tập hữu hạn khác rỗng các đối tượng; A là tập hữu hạn khác rỗng các thuộc tính. Mỗi thuộc tính $a \in A$, V_a là tập giá trị của a và $u \in U$, $a(u)$ là giá trị của u tại thuộc tính a .

Chú ý: Nếu $\forall a \in A, \forall u \in U$ $a(u)$ chỉ có một giá trị thì $S = (U, A)$ là hệ tin đơn trị, ngược lại $S = (U, A)$ gọi là hệ tin đa trị hay hệ tin giá trị tập (set-value information system). Trong bài viết này chúng tôi chỉ xét các hệ thống thông tin xác định đầy đủ, nghĩa là các hệ thống thông tin mà mọi thuộc tính luôn có giá trị (tập giá trị) xác định.

Ví dụ, Bảng 1 là hệ thống thông tin đơn trị, Bảng 2 là hệ thống thông tin đa trị.

Định nghĩa 2. Quan hệ bất khả phân biệt

Cho hệ thống thông tin đơn trị $S = (U, A)$, $B \subseteq A$.

Quan hệ $IND(B) \subseteq U \times U$ được gọi là quan hệ bất khả phân biệt trên U nếu với mọi cặp đối tượng $u, u' \subseteq U$ thì $uIND(B)u'$ khi và chỉ khi $a(u) = a(u')$ với mọi $a \in B$.

Dễ dàng thấy rằng quan hệ $IND(B)$ là quan hệ tương đương trên U . Phân hoạch $U/IND(B) = U/B$ là phân hoạch tương đương.

Chú ý: Chúng ta sẽ ký hiệu U/B là phân hoạch của $U/IND(B)$ và $U/B = \{[o]_B: o \in U\}$ là các nhóm tương đương. Với $[o]_B$ là nhóm các đối tượng quan hệ với nhau.

Định nghĩa 3. Xấp xỉ của tập hợp

Cho hệ thống thông tin đơn trị $S = (U, A)$; $B \subseteq A$; $X \subseteq U$.

Xấp xỉ dưới của X , ứng với phân hoạch U/B , ký hiệu X_B và $X_B = \cup \{[o]_B: o \in U \text{ và } [o]_B \subseteq X\}$.

Xấp xỉ trên của X , ứng với phân hoạch U/B , ký hiệu X^B và $X^B = \cup \{[o]_B: o \in U \text{ và } [o]_B \cap X \neq \emptyset\}$.

Định nghĩa 4. Hệ quyết định

Hệ quyết định là hệ thống thông tin S mà trong tập thuộc tính A có tập thuộc tính quyết định D .

Vậy hệ quyết định $T = (U, A)$; trong đó $A = C \cup D$; $C \cap D \neq \emptyset$. Tập C được gọi là tập thuộc tính điều kiện, D là thuộc tính quyết định.

Ví dụ:

Bảng 1. Hệ quyết định đơn trị

U	Mã Bệnh nhân	Thân nhiệt	Ho	Sốt	Ho có đờm	Kết luận
u_1	1	40	Nhiều	Cao	Không	Viêm họng
u_2	2	37	Ít	Thấp	Không	Bình thường

u_3	3	40	Nhiều	Cao	Không	Viêm họng
u_4	4	41	Nhiều	Cao	Có	Viêm phổi cấp
u_5	5	38	Không	Thấp	Không	Bình thường
u_6	6	38	Không	Thấp	Không	Bình thường
u_7	7	38	Không	Thấp	Không	Bình thường

Bảng 2. Hệ quyết định đa trị

U	Mã NV	Học Hàm	Học vị	Chuyên ngành	Ngoại ngữ	Kết luận
u_1	1	PGS	TS	Cơ khí	{Anh, Pháp}	Giảng viên cao cấp
u_2	2	PGS	TS	CNTT	{Nga, Anh, Pháp}	Giảng viên cao cấp
u_3	3	GS	TS	Cơ khí	{Nga}	Giảng viên cao cấp
u_4	4	GS	TSKH	Điện tử	{Nga, Anh}	Giảng viên cao cấp
u_5	5	PGS	TS	Điện tử	{Anh, Pháp}	Giảng viên cao cấp
u_6	6	0	Ths	Điện tử	{Pháp}	0
u_7	7	GS	TS	CNTT	{Nga}	Giảng viên cao cấp

Chú ý: Trong hệ quyết định đa trị $\forall o \in U$ $o[D]$ chỉ có một giá trị.

Định nghĩa 5. Phụ thuộc hàm trong hệ thống thông tin

Cho hệ thống thông tin đơn trị $S = (U, A)$; $B, B' \subseteq A$; Ta nói B xác định phụ thuộc hàm B' , ký hiệu $B \rightarrow B'$ nếu và chỉ nếu $IND(B) \subseteq IND(B')$.

2. Một số tính chất của phụ thuộc hàm

2.1. Tính phản xạ:

$\forall B \subseteq A$ thì $B \rightarrow B$ và nếu $B' \subseteq B$ thì $B \rightarrow B'$.

2.2. Tính mở rộng hai vế:

Nếu $B \rightarrow B'$ thì $BC \rightarrow B'C$

2.3. Tính bắc cầu:

Nếu $B \rightarrow B'$ và $B' \rightarrow C$ thì $B \rightarrow C$

2.4. Tính tựa bắc cầu:

Nếu $B \rightarrow B'$ và $B'C \rightarrow C'$ thì $BC \rightarrow C'$

2.5. Tính mở rộng trái thu hẹp phải:

Nếu $B \rightarrow B'$ thì $BC \rightarrow B' \setminus C'$

2.6. Tính cộng đầy đủ:

Nếu $B \rightarrow C$ và $B' \rightarrow C'$ thì $BC \rightarrow B'C'$

2.7. Tính tích lũy:

Nếu $B \rightarrow C$ và $C \rightarrow B'C'$ thì $B \rightarrow BB'C'$

Chú ý: Trong Cơ sở dữ liệu, BC là hợp của hai tập B và C hay $BC = B \cup C$.

Định nghĩa 6. Hàm đóng

Cho U là tập bất kỳ, $P(U)$ là họ các tập con của U .

Hàm $f: P(U) \rightarrow P(U)$ là hàm đóng nếu f thỏa mãn 3 điều kiện sau:

- (1) Tính phản xạ: $\forall X \in P(U) \quad X \subseteq f(X)$
- (2) Tính đồng biến: $\forall X, Y \in P(U)$ nếu $X \subseteq Y$ thì $f(X) \subseteq f(Y)$
- (3) Tính lũy đẳng: $\forall X \in P(U) \quad f(f(X)) = f(X)$

3. Một số tính chất của hàm đóng

3.1. Hàm đóng của hợp các tập chứa hợp của các hàm đóng

$\forall X, Y \in P(U) \quad f(XY) \supseteq f(X)f(Y)$.

3.2. Hàm đóng của giao hai tập được chứa trong giao các hàm đóng của hai tập đó

$\forall X, Y \in P(U) \quad f(X \cap Y) \subseteq f(X) \cap f(Y)$

3.3. Đẳng thức

$\forall X, Y \in P(U) \quad f(XY) = f(f(X)Y)$ và $f(XY) = f(Xf(Y))$.

3.4. Đẳng thức kép

$\forall X, Y \in P(U) \quad f(XY) = f(f(X)Y) = f(Xf(Y)) = f(f(X)f(Y))$.

4. Một số tính chất của xấp xỉ trên

Cho hệ thống thông tin đơn trị và đầy đủ $S = (U, A)$; $B \subseteq A$.

Trên $P(U)$ ta xây dựng hàm $f: P(U) \rightarrow P(U)$ xác định như sau:

$$\forall X \in P(U) \quad f(X) = X^B$$

Ta dễ dàng thấy rằng f là hàm đóng vì f thỏa ba điều kiện của hàm đóng, đó là tính phản xạ: $X \subseteq X^B$, tính đồng biến: nếu $X \subseteq Y$ thì $X^B \subseteq Y^B$, tính lũy đẳng: $X^B = X^{BB}$

Theo các tính chất của hàm đóng ta có các tính chất của xấp xỉ trên như sau

1. $\forall X, Y \in P(U) \quad (XY)^B \supseteq X^B Y^B$
2. $\forall X, Y \in P(U) \quad (X \cap Y)^B \subseteq X^B \cap Y^B$
3. $\forall X, Y \in P(U) \quad (XY)^B = (X^B Y)^B$
4. $\forall X, Y \in P(U) \quad (XY)^B = (X Y^B)^B$
5. $\forall X, Y \in P(U) \quad (XY)^B = (X^B Y)^B = (X Y^B)^B = (X^B Y^B)^B$

Định nghĩa 7. Vùng dương của hai tập thuộc tính B, B'

Cho hệ thống thông tin đơn trị $S = (U, A)$, $B, B' \subseteq A$.

Vùng dương của B và B' , ký hiệu $POS(B, B')$ và $POS(B, B') = \cup \{[o]_{B'} : [o]_B \subseteq [o]_{B'} \text{ và } o \in U\}$.

Định nghĩa 8. Phụ thuộc hàm với độ phụ thuộc k(B, B')

Cho hệ thống thông tin $S = (U, A)$, $B, B' \subseteq A$.

Tập B' được gọi là *phụ thuộc hàm độ k(B, B')* vào B , ký hiệu $B \xrightarrow{k(B, B')} B'$ nếu

$$k(B, B') = \frac{Card(POS(B, B'))}{Card(U)}$$

Định lý 1. Cho hệ thống thông tin $S = (U, A)$, $B, B' \subseteq A$.

$B \xrightarrow{1} B'$ khi và chỉ khi $IND(B) \subseteq IND(B')$.

Chứng minh:

Giả sử $B \xrightarrow{1} B' \Rightarrow k(B, B') = 1 \Rightarrow POS(B, B') = \cup \{[o]_{B'} : [o]_B \subseteq [o]_{B'} \text{ và } o \in U\} = U \Rightarrow IND(B) \subseteq IND(B')$.

Tương tự giả sử $IND(B) \subseteq IND(B')$ ta dễ dàng thử lại rằng $POS(B, B') = U$ và khi đó $B \xrightarrow{1} B'$.

5. Một số tính chất cơ bản của vùng dương**Tính chất 1. Sự bao nhau của các nhóm trên các tập thuộc tính bao nhau**

Cho hệ thống thông tin $S = (U, A)$. Nếu $B \subseteq B' \subseteq A$ thì mọi $o \in U$ ta luôn có $[o]_{B'} \subseteq [o]_B$.

Chứng minh: Lấy $o' \in [o]_{B'}$ khi đó vì o' và o giống nhau (bất khả phân biệt) trên B' và $B \subseteq B'$ nên o và o' giống nhau trên B hay $o' \in [o]_B$ nên $[o]_{B'} \subseteq [o]_B$.

Tính chất 2. Cho hệ thống thông tin $S = (U, A)$. Với mọi $o \in U$ thì $o \in POS(B, B')$ khi và chỉ khi $[o]_B \subseteq [o]_{B'}$.

Chứng minh: tính chất 2 được suy trực tiếp từ định nghĩa vùng dương.

Tính chất 3. Biểu diễn vùng dương qua xấp xỉ dưới

Nếu đặt $E = U / B = \{E_1, E_2, \dots, E_k\}$; $AprE = (U, E)$ và $P = U / B' = \{P_1, P_2, \dots, P_j\}$; $AprP = (U, P)$ thì $POS(B, B') = \bigcup_{P_j \in P} (P_j)_E$ và

$$POS(B, B') = \bigcup_{E_i \in E} (E_i)_P.$$

Chứng minh: Tính chất 3 được suy trực tiếp từ định nghĩa vùng dương và xấp xỉ dưới.

Tính chất 4. Số các nhóm đối tượng liên quan đến các tập thuộc tính

Cho hệ thống thông tin $S = (U, A)$.

Nếu B và B' là hai tập thuộc tính thỏa mãn $B \subseteq B'$ thì $card(U/B) \leq card(U/B')$.

Chứng minh: Vì mỗi nhóm của U/B' là một nhóm con của U/B nên số nhóm của U/B không thể vượt quá số nhóm của U/B' .

Tính chất 5. Sự đồng biến của hàm độ đo phụ thuộc

Cho hệ quyết định $T = (U, C \cup D)$. Hàm $k(B, D): 2^C \rightarrow [0, 1]$ với 2^C là họ các tập con của C

và $k(B, D) = \frac{Card(POS(B, D))}{Card(U)}$ là hàm đồng biến.

Chứng minh: Để chứng minh tính chất 5, ta chỉ cần chứng minh với mọi cặp tập thuộc tính điều kiện B, B' mà $B \subseteq B'$ thì $POS(B, D) \subseteq POS(B', D)$.

Lấy $o \in POS(B, D)$ khi đó $[o]_B \subseteq [o]_D$. Mặt khác vì $B \subseteq B'$ nên theo tính chất 1 ta có $[o]_{B'} \subseteq [o]_B$. Vậy $[o]_{B'} \subseteq [o]_D$ hay $o \in POS(B', D)$.

Tính chất 6. Cho hệ quyết định $T = (U, C \cup D)$.

Nếu đặt $w(c) = k(\{c\}, D)$ là trọng số của thuộc tính $c \in C$ và $w(B) = k(B, D)$ là trọng số của tập thuộc tính $B(B \subseteq C)$ thì $w(c) \leq w(B)$ với mọi $c \in B$.

Chứng minh tính chất 6 suy từ tính chất 5.

4. Kết luận

Trong bài viết này chúng tôi đã giới thiệu một số nghiên cứu, tính chất có tính hệ thống, cơ bản của vùng dương, độ phụ thuộc, ràng buộc của các tập thuộc tính trong hệ thống thông tin. Đồng thời trong bài viết này, chúng tôi cũng đã nêu được một số tính chất quan trọng, cơ bản của khái niệm phụ thuộc hàm trong hệ thống thông tin. Trong bài viết các tính chất và một số công thức liên quan đến xấp xỉ trên đã được đề cập tới.

Tài liệu tham khảo

- [1]. Guangming Lang, Quingguo Li, *Data Compression of Dynamic Set-valued Information Systems*, ArXiv: 1209.6509v1 [cs.IT] 28 Sep 2012
- [2]. Pawlak Z. (1991), *Rough sets: Theoretical Aspects of Reasoning About Data*, Kluwer Academic Publishers.
- [3]. Pawlak Z. (1998), "Rough Set Theory and its Applications in Data Analysis", Cybernetics and systems 29, pp. 661-688.

- [4]. Qian Y.H. and Liang J.Y. (2006), “*Combination Entropy and Combination Granulation in Incomplete Information System*”, RSKT 2006, pp. 184-190.
- [5]. Qian Y.H. and Liang J.Y. (2008), “*New Method for Measuring Uncertainty in Incomplete Information Systems*”, International Journal of Uncertainty, Fuzziness and Knowledge-Based Systems.
- [6]. Qian Y.H., Liang J.Y. and Dang C.Y. (2009), “*Knowledge Structure, Knowledge Granulation and Knowledge Distance in a Knowledge Base*”, International Journal of Approximate Reasoning 50, pp. 174-188.
- [7]. Qian Y.H., Liang J.Y., Dang C.Y., Wang F. and Xu W. (2007), “*Knowledge Distance in Information Systems*”, Journal of Systems Science and Systems Engineering, Vol. 16, pp. 434-449.
- [8]. Qian Y.H., Liang J.Y., Li D.Y., Zhang H.Y. and Dang C.Y. (2008), “*Measures of Evaluating the Decision Performace of a Decision Table in Rough Set Theory*”, Information Sciences, Vol.178, pp.181-202.
- [9]. R.López de Mántaras, *A Distance-based Attribute Selection Measure for Decision Tree Induction*, Machine Learning Vol. 6 (1991) 81-92.
- [10]. Simovici D. A. and Jaroszewicz S. (2006), “*A New Metric Splitting Criterion for Decision Trees*”, International Journal of Parallel Emergent and Distributed Systems, Vol. 21 (4), pp. 239-256.
- [11]. Simovici D. A., Jaroszewicz S. (2003), “*Generalized Conditional Entropy and Decision Trees*”, Proceeding of EGC, Lyon, France, pp. 369-380.
- [12]. Sun L., Xu J.C and Cao X.Z (2009), “*Decision Table Reduction Method Based on New Conditional Entropy for Rough Set Theory*”, International Workshop on Intelligent Systems and Applications, pp. 1-4.
- [13]. Thi V.D. (1986), “*Minimal Keys and Antikeys*”, Acta Cybernetica 7, 4, pp. 361-371.
- [14]. Vu Duc Thi, Nguyen Long Giang (2011), “*A Method to Construct Decision Table from Relation Scheme*”, Cybernetics and Information Technologies, Sofia, Bulgarian Academy of Sciences, Volume 11, No 3, 32-41.
- [15]. J.D.Ullman (1998), “*Nguyên lý các hệ cơ sở dữ liệu và cơ sở tri thức*”, NXB Thống kê.
- [16]. Nguyễn Bá Tường (2011), “*Cơ sở dữ liệu quan hệ và ứng dụng*”, NXB Thông tin và truyền thông.

DEPENDENCES ATTRIBUTES OF INFORMATION SYSTEMS AND PROPERTIES OF UPPER APPROXIMATION BASED CLOSE FUNCCTION

Abstract:

This paper investigates some properties of dependence attributes in information systems. In the paper we have been proposed a closed mapping and consequently we have showed some properties and formulas of upper approximation.

Furthermore, it has been shown that single information system and set-value information system determine each other.

Keywords: *Dependence attribute, information system, upper approximation, decision system, close function.*