TẬP SAN TIN HỌC QUẢN LÝ
Tập 03, số 1, 2014, 05-14

MỘT SỐ ỨNG DỤNG LÝ THUYẾT TẬP THÔ TRONG KINH TẾ VÀ TÀI CHÍNH

Nguyễn Đức Thuần [1]

***Abstract**: Rough set theory provides a useful mathematical foundation for developing automated computational systems that can help understand and make use of imperfect knowledge. Recently, Rough set theory and its extensions have been widely applied to many problems, including decision analysis, data mining, intelligent control and pattern recognition, economic and financial problems: risk management, economic and financial forecasting, evaluating marketing performance, decision support... This paper presents an outline of the basic concepts of rough sets and their major extensions. It also shows the diversity of successful applications these theories in Economy and Finance.*

***Tóm tắt**: Lý thuyết tập thô cung cấp một nền tảng toán học hữu ích cho việc phát triển hệ thống tính toán tự động, khai phá và phát hiện tri thức của các hệ thống không đầy đủ. Trong thời gian gần đây, lý thuyết tập thô đã được ứng dụng rộng rãi giải quyết nhiều bài toán: phân tích ra quyết định, khai thác dữ liệu, điều khiển thông minh và nhận dạng mẫu, các bài toán kinh tế và tài chính như: quản lý rủi ro, dự báo thị trường tài chính, đánh giá năng lực hoạt động các mạng lưới tiếp thị sản phẩm, hỗ trợ ra quyết định…. Bài viết này trình bày các khái niệm cơ bản của tập thô cổ điển và mở rộng, sự ứng dụng hữu hiệu lý thuyết tập thô trong lĩnh vực kinh tế và tài chính.*

***Từ khóa**: Tập thô, Kinh tế, Tài chính, Quản lý rủi ro, Dự báo, Đánh giá năng lực thị trường.*

1 MỞ ĐẦU

Lý thuyết tập thô hiện nay đã trở thành một công cụ toán học quen thuộc để xử lý dữ liệu không đầy đủ và mơ hồ. Các kết quả đạt được và tiềm năng của lý thuyết tập thô đã thu hút sự chú ý giới học thuật trên toàn thế giới nghiên cứu và phát triển. Kết quả mở rộng lý thuyết tập thô ngày càng nhiều và được ứng dụng trong nhiều lĩnh vực khác nhau. Bài viết này cố gắng trình bày một số khái niệm cơ bản của lý thuyết tập thô và một số ứng dụng trong lĩnh vực kinh tế, tài chính.

2 LÝ THUYẾT TẬP THÔ

Lý thuyết tập thô (lt. tập thô) là sự mở rộng lý thuyết tập hợp, nhằm phát hiện tính chất đặc trưng xác định 1 tập hợp. Để xấp xỉ tính chất đặc trưng này, lý thuyết tập thô gắn liền với 2 phép xấp xỉ, gọi là xấp xỉ dưới và xấp xỉ trên. Hai phép xấp xỉ này trong lý thuyết tập thô cổ điển được định nghĩa dựa trên một phân hoạch gắn liền với một quan hệ tương đương. Từ 2 phép xấp xỉ này, lt.tập thô xác định được mối quan hệ giữa các tập thuộc tính của tập các đối tượng khảo sát, định lượng được tầm quan trọng các thuộc tính, phát hiện được các đặc trưng của dữ liệu, rút gọn các thuộc tính, loại bỏ đối tượng dư thừa, hỗ trợ việc rút trích tập luật quyết định..

---

[1] Khoa Công Nghệ Thông Tin – ĐH Nha Trang

Các mở rộng của lý thuyết tập thô tập trung vào việc định nghĩa lại các phép xấp xỉ dưới và xấp xỉ trên. Việc lai tập thô và lý thuyết tập mờ, tập thô và mạng neuron..cũng nhằm định nghĩa lại các phép xấp xỉ này nhằm mở rộng phạm vi ứng dụng lt.tập thô trong nhiều lĩnh vực khác nhau.

Trong phần này, chúng tôi chỉ trình bày tóm tắt các khái niệm của tập thô, chi tiết có thể tham khảo [9][18]

2.1 Hệ thông tin và bảng quyết định

2.1.1 Hệ thống thông tin

Hệ thống thông tin là một cặp I = (U, A), U là một tập hữu hạn khác rỗng các đối tượng gọi là tập vũ trụ hay là tập phổ dụng, A là một tập hữu hạn khác rỗng các thuộc tính. Với mỗi $u \in U$ và $a \in A$, ta ký hiệu u(a) là giá trị của đối tượng u tại thuộc tính a. Nếu gọi Ia là tập tất cả giá trị của thuộc tính a, thì $u(a) \in Ia$ với mọi $u \in U$. Bây giờ, nếu B = {b1, b2, ..,bk} $\subseteq$ A, ta ký hiệu bộ các giá trị u(bi) bởi u(B).

2.1.2 Bảng quyết định

Bảng quyết định là một hệ thống thông tin có dạng I = (U, A), trong đó A $=C \cup D$, với $C \cap D = \emptyset$, C được gọi là tập thuộc tính điều kiện, còn D là tập thuộc tính quyết định.

2.2 Quan hệ không phân biệt được

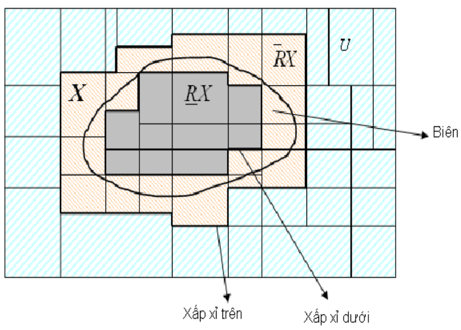
Xét hệ thống thông tin I = (U, A), với mỗi tập thuộc tính B $\subseteq$ A tạo ra một quan hệ hai ngôi trên U, ký hiệu IND(B),

$$\text{IND(B)}=\{(u,v) \in U \times U \mid u(a) = v(a), \forall a \in B\}$$

IND(B) được gọi là quan hệ B_không phân biệt được. Dễ kiểm chứng đây là một quan hệ tương đương trên U. Với mọi đối tượng $u \in U$, lớp tương đương của u trong quan hệ IND(B) được kí hiệu bởi [u]B. Tập thương xác định bởi quan hệ IND(B) được ký hiệu U/IND(B) hay U/B, tức là U/IND(B) = U/B = {[u]B | $u \in U$}

2.3 Tập thô

Để biểu diễn một tập hợp bằng tri thức được cho xác định bởi một tập thuộc tính, người ta định nghĩa hai phép xấp xỉ:



Cho một hệ thống thông tin I = (U, A), với mỗi tập con X $\subseteq$ U và B $\subseteq$ A, ký hiệu R = IND(B), ta có 2 tập con sau

$$\underline{R}(X) = \{u \in U \mid [u]_B \subseteq X\}$$

$$\overline{R}(X) = \{u \in U \mid [u]_B \cap X \neq \emptyset\}$$

$\underline{R}(X)$, $\overline{R}(X)$ lần lượt gọi là R-xấp xỉ dưới và R-xấp xỉ trên của tập X.

Từ xấp xỉ trên, xấp xỉ dưới người ta định nghĩa các tập:

$BN_B(X) = \overline{R}(X) - \underline{R}(X)$: *B- miền biên của X.*

$POS_B(X) = \underline{R}(X)$: *B-vùng dương của* X.

$NEG_B(X) = U - \overline{R}(X)$: *B-vùng âm của* X.

Trong trường hợp $BN_B(X) \neq \emptyset$, X được gọi là tập thô, ngược lại X được gọi là tập rõ.

Đối với một hệ thống thông tin I= (U, A), B, $D \subseteq A$, ký hiệu R = IND(B), người ta gọi B-miền khẳng định dương của D là tập được xác định như sau

$$POS_B(D) = \bigcup_{V \in U/D} (\underline{R}(V)) \tag{1}$$

2.4 Phụ thuộc thuộc tính và trọng số

Một vấn đề quan trọng trong phân tích dữ liệu là phát hiện sự phụ thuộc giữa các thuộc tính. Trong lý thuyết tập thô, phụ thuộc thuộc tính được định nghĩa như sau:

Với P, $Q \subseteq A$, Q phụ thuộc độ k vào P ($0 \leq k \leq 1$), ký hiệu $P \Rightarrow_k Q$, nếu:

$$k = \gamma_P(Q) = \frac{|POS_P(Q)|}{|U|} \tag{2}$$

Ở đây, $|S|$ là lực lượng của S. Nếu k=1, Q là phụ thuộc đầy đủ vào P khi đó ký hiệu $P \Rightarrow Q$; nếu ($0 < k < 1$) Q là phụ thuộc bộ phận (độ k) vào P; nếu k = 0, Q không phụ thuộc vào P.

Bằng cách tính sự thay đổi độ phụ thuộc, khi 1 thuộc tính bị xóa bỏ thể hiện được trọng số của thuộc tính này đối với sự phụ thuộc. Định nghĩa trọng số của thuộc tính đối với sự phụ thuộc như sau: cho 2 tập thuộc tính P, Q và thuộc tính $a \in P$, trọng số của thuộc tính a theo Q được xác định:

$$\gamma_P(Q, a) = \gamma_P(Q) - \gamma_{P-\{a\}}(Q) \tag{3}$$

Một số tính chất của phụ thuộc thuộc tính:

Mệnh đề 1

1. *Nếu* $P \Rightarrow_k Q$ *và* $R \Rightarrow_l Q$ *thì* $P \cup R \Rightarrow_m Q$, *với* $m \geq \max(k, l)$
2. *Nếu* $P \cup R \Rightarrow_k Q$ *thì* $P \Rightarrow_l Q$ *và* $R \Rightarrow_m Q$, *với* $l, m \leq k$
3. *Nếu* $P \Rightarrow_k Q$, *và* $P \Rightarrow_l R$ *thì* $P \Rightarrow_m Q \cup R$, *với* $m \leq \min(k, l)$
4. *Nếu* $P \Rightarrow_k Q \cup R$, *thì* $P \Rightarrow_l Q$ *và* $P \Rightarrow_m R$, *với* $l, m \geq k$
5. *Nếu* $P \Rightarrow_k Q$, *và* $Q \Rightarrow_l R$ *thì* $P \Rightarrow_m R$, *với* $m \geq l + k - 1$

2.5 Rút gọn tập thuộc tính

Nhiều bài toán ứng dụng cần rút gọn hệ thống thông tin, tìm biểu diễn tối thiểu của hệ thống thông tin ban đầu, nhất là các bản quyết định. Đối với điều này, khái niệm rút gọn và tập thuộc tính điều kiện tối thiểu được định nghĩa như sau:

Tập thuộc tính *P* được gọi là tập rút gọn tối thiểu của tập thuộc tính điều kiện *C* ứng với tập thuộc tính quyết định *D* xác định như sau:

$$P = \{a \in C \mid \gamma_P(D) = \gamma_C(D), \gamma_{P-\{a\}}(D) \neq \gamma_P(D)\} \quad (4)$$

Giao tất cả tập rút gọn tối thiểu của *C*, gọi là nhân của C ký hiệu *Core(C).* Có nhiều thuật toán tìm rút gọn tập thuộc tính điều kiện, phổ biến nhất là thuật toán QuickReduct.

2.6 Luật quyết định [18]

Cho bảng quyết định **I**= (U, $C \cup D$), giả sử $U/C = \{X_1, X_2,.., X_m\}$ và $U/D = \{Y_1, Y_2,.., Y_n\}$. Nếu $X_i \cap Y_j \neq \emptyset$, ký hiệu des($X_i$), des($Y_j$) lần lượt là các mô tả của các lớp tương đương tương ứng với X_i, Y_j. Một luật quyết định xác định bởi X_i, Y_j có dạng Z_{ij}: $des(X_i) \rightarrow des(Y_j)$.

3 MỞ RỘNG LÝ THUYẾT TẬP THÔ

Với lt.tập thô cổ điển, điểm hạn chế là các khái niệm được xây dựng trên các phép xấp xỉ dựa vào phân hoạch hay quan hệ tương đương. Trong thực tế, lớp các quan hệ tương đương là không nhiều, vì vậy một số mở rộng lt.tập thô cổ điển là định nghĩa lại các phép xấp xỉ, hay thay quan hệ tương đương bởi các quan hệ 2 ngôi khác như quan hệ dung sai (*tolerance relation*).

Một vấn đề quan trọng trong phân tích dữ liệu là phát hiện sự phụ thuộc giữa các thuộc tính. Trong lý thuyết tập thô, phụ thuộc thuộc tính được định nghĩa như sau:

3.1 Chuẩn xác tập thô theo tham biến

Chuẩn xác tập thô theo tham biến (*Variable precision rough sets*) [14] là đề xuất mở rộng tập thô bằng cách nới lỏng các phép toán, nhằm để phân tích và xác định các mô hình dữ liệu theo xu hướng thống kê. Ý tưởng chính của đề xuất là cho phép các đối tượng được phân lớp với một độ sai lệch. Xét $X, Y \subseteq U$, sai lệch phân lớp được xác định bởi

$$c(X,Y) = 1 - \frac{|X \cap Y|}{|X|} \quad (5)$$

Nhận xét, $c(X,Y) = 0$ nếu và chỉ nếu $X \subseteq Y$. Quan hệ bao hàm mức β được định nghĩa:

$$X \subseteq_\beta Y \Leftrightarrow c(X,Y) \leq \beta, \text{ với } 0 \leq \beta \leq 0.5$$

Sử dụng $\subseteq_\beta$ thay cho $\subseteq$, phép xấp xỉ β-trên và xấp xỉ β-dưới của tập X được định nghĩa:

$$R\beta = \bigcup \{[x]_R \in U/R \mid [x]_R \subseteq_\beta X \} \quad (6)$$

$$\overline{R}_\beta = \bigcup \{[x]_R \in U/R \mid c([x], X) < 1-\beta \} \quad (7)$$

Chú ý, $\underline{R}_\beta X = \underline{R}X$ khi β=0. Các khái niệm miền dương, miền âm, miền biên, của tập thô cổ điển có thể được mở rộng

$$POSR,\beta\,(X) = \underline{R}_\beta X \quad (8)$$

$$NEGB,\beta\,(X) = U - \overline{R}_\beta X \quad (9)$$

$$BNR,\beta\,(X) = \overline{R}_\beta X - \underline{R}_\beta X \quad (10)$$

β-miền dương của một quan hệ tương đương Q trên U được xác định bởi

$$POS_{R,\beta}(Q) = \bigcup_{X \in U/Q} \underline{R}_\beta X \quad (11)$$

Ở đây, R cũng là một quan hệ tương đương trên U. Đại lượng này có thể dùng tính độ phụ thuộc giữa 2 tập thuộc tính và xác định β-rút gọn tập thuộc tính.

$$\gamma_{R,\beta}(Q) = \frac{|POS_{R,\beta}(Q)|}{|U|} \quad (12)$$

3.2 Tập thô dung sai

3.2.1 Độ đo tương đồng [9][14]

Để xử lý các thuộc tính không chắc chắn, mơ hồ là đưa ra độ đo tương đồng giữa các thuộc tính và định nghĩa lại các phép xấp xỉ theo độ đo tương đồng này.

Độ đo tương đồng được xác định trên mỗi thuộc tính. Độ đo tương đồng chuẩn ứng với thuộc tính a của 2 đối tượng x,y được xác định

$$SIM_a(x,y) = 1 - \frac{|a(x) - a(y)|}{|a_{\max} - a_{\min}|} \quad (13)$$

Ở đây, a(x): giá trị của thuộc tính a ứng với đối tượng x; $a_{\max}$, $a_{\min}$ lần lượt là giá trị lớn nhất và giá trị nhỏ nhất ứng với thuộc tính a.

Có thể mở rộng độ tương đồng khi xét nhiều hơn 1 thuộc tính, giả sử tập P thuộc tính. Hai cách tiếp cận thường gặp

$$(x,y) \in SIM_{P,\tau} \Leftrightarrow \prod_{a \in P} SIM_a(x,y) \geq \tau \quad (14)$$