TẬP THỌ & ĐÁNH GIÁ HỆ THÔNG TIN NỀN

Hà Quang Thùy
Đại học Khoa học tự nhiên - DHQG Hà Nội

ho hệ thông tin [6] $S = \{ \Omega, A, V, r \}$, trong đó $\Omega$ là tập điểm trường, $A$ tập tén thuộc $V$ là tập các giá trị còn $r$ là ảnh xạ từ $\Omega \times A$ tới $V$.

ô mỗi $a \in A$, xây dựng quan hệ thông trường $a^* \in \Omega : xa^*y \iff r(x, a) = r(y, a)$.

trong trường theo quan hệ $S^* = \bigcap_{a \in A} a^*$ được gọi là tập cơ bản trong $S$. Trong bài này, thường dùng chữ cái $e$ để chỉ tập cơ bản.

áp hợp tất cả các tập cơ bản trong $S$ được kí hiệu là $E_S$ (dựng viết tắt là $E$). diễn tượng hợp, dũng kí hiệu $(\Omega, E)$ để chỉ hệ thống tin $S$.

áp con của $\Omega$ là hợp của một số tập cơ bản trong $S$ được gọi là tập mở và được.


uy vậy, trong nhiều trường hợp không chỉ giới hạn xem xét các tập mở và được còn quan tâm đến các tập hợp con khác trong $\Omega$ - nay sinh vấn đề xấp xỉ một át kỳ quá tập mở và được. Có một số cách xây dựng xấp xỉ trong hệ thống tin. Lẽ này được định hướng theo khái niệm "tập thổ". Các khái niệm được trình bày đầy đủ được dùng để hình thức hóa quá trình xấp xỉ các tập không mở và được thông ảnh tập mở và được.

1. TẬP GOURI HAN

ho X là tập con của $\Omega$. Các tập giới hạn liên quới $X$ được định nghĩa như nhau:

để: $X^* = \bigcup_{e \in E, e \not\in X} e$; $X_+ = \bigcup_{e \in E, e \in X}$

hướng dẫn: Chúng ta chứng minh trực tiếp $X^*$ (trong hợp $X$, là trường tự). $t = \bigcup_{e \in E, e \not\in X} e$. Hiển nhiên, $X^* \subseteq B^*$. Cần kiểm chứng chỉ xây ra dâu đăng thực.

ngược lại có nghĩa là $\exists x \in B \setminus X^*$. Theo cách xấp xỉ tập $B$, ta phải có tập cơ
nảo đó để cho $r(\cdot, x) = r(x, \cdot)$ khác $\not\in X^*$. Suy ra $r(x, \cdot) \not\in X^* = \emptyset$.

$X^*$ mở và được nên $\exists e^* \subseteq X^*$, hay cùng xấp xỉ $x \in X^*$ cấu tạo với giá thiết
chứng. Vậy ta có $B = X^*$ (dịu phối chắc chắn).

t: Xét hệ thống hình cái bệnh án của một số bệnh nhân trẻ em cho trong bằng

13
<table>
<thead>
<tr>
<th>Đối tượng</th>
<th>Dịch não</th>
<th>BK dịch</th>
<th>Thần niet</th>
<th>BK dịch</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>tùy lần 1</td>
<td>đã dạy</td>
<td>não tủy</td>
<td></td>
</tr>
<tr>
<td>A</td>
<td>vàng</td>
<td>âm</td>
<td>rất cao</td>
<td>homo án</td>
</tr>
<tr>
<td>B</td>
<td>trong</td>
<td>dương</td>
<td>nhẹ</td>
<td>homo du</td>
</tr>
<tr>
<td>C</td>
<td>dục</td>
<td>âm</td>
<td>cao</td>
<td>homo án</td>
</tr>
<tr>
<td>D</td>
<td>vàng</td>
<td>âm</td>
<td>nhẹ</td>
<td>homo án</td>
</tr>
<tr>
<td>E</td>
<td>vàng</td>
<td>âm</td>
<td>cao</td>
<td>homo án</td>
</tr>
<tr>
<td>F</td>
<td>dục</td>
<td>âm</td>
<td>cao</td>
<td>homo án</td>
</tr>
<tr>
<td>G</td>
<td>vàng</td>
<td>dương</td>
<td>hơi cao</td>
<td>homo án</td>
</tr>
<tr>
<td>H</td>
<td>vàng</td>
<td>âm</td>
<td>rất cao</td>
<td>homo án</td>
</tr>
<tr>
<td>I</td>
<td>dục</td>
<td>âm</td>
<td>cao</td>
<td>homo án</td>
</tr>
<tr>
<td>J</td>
<td>vàng</td>
<td>âm</td>
<td>nhẹ</td>
<td>homo án</td>
</tr>
</tbody>
</table>

Trong hệ thống tin này, các tập cơ bản là:

\[ e_1 = \{A, H\}, \quad e_2 = \{B\}, \quad e_3 = \{C, F, I\}, \quad e_4 = \{D, J\}, \quad e_5 = \{E\}, \quad e_6 = \{G\}. \]

Giá standoff, các bệnh nhân A, C, E, H, I, J đã được sử bì lao màng não cơn các bệnh khác chưa thể kết luận. Nếu kỳ hiệu X để chỉ tập bệnh nhân lao màng não, thì cả giới han liên quan đến X là:

\[ X_* = e_1 \cup e_5 \quad X^* = e_1 \cup e_3 \cup e_4 \cup e_5 \]

II. TẬP THƠ

Khái niệm tập thơ được A. Marek và Z. Pawlak đưa ra trong [7,8]:

Cho \((Ω, E)\) là hệ thống tin với tập đối tượng Ω và tập các tập cơ bản E. Cho tập con của Ω. Tập thơ (rough set) đối với tập X được định nghĩa như dưới đây:

Theo cách biểu diễn tập thơ:

\[ R_*(X) = \bigcup_{e \in X} e \quad \text{tập thơ chắc chắn (thyoutube mạnh)} \]

\[ R^*(X) = \bigcup_{e \not\in X} e \quad \text{tập thơ có thể} \ (\text{thyoutube yếu)} \]

hoặc theo cách biểu diễn chỉ bằng duy nhất hàm thành viên \(f_X\) (tương ứng với Ω vào \(R^+\) thì tập thơ được xác định như sau: \(∀x ∈ Ω, x ∈ e \ (e \in E)\), ta có

\[ f_X(x) = \begin{cases} 
0 & \text{nếu } e \cap X = \emptyset \\
1 & \text{nếu } e \subseteq X \text{ hay } x \in R_*(X) \\
1/2 & \text{trường hợp còn lại hay } x \in R^*(X) - R_*(X) 
\end{cases} \]

Theo cách biểu diễn tập thơ theo ngôn ngữ hàm thành viên, khi đối sánh với định nghĩa tập mở của Zadeh, theo Pawlak [7], định nghĩa tập thơ tùy có một số điều tương
cùng có những khác biệt quan trọng, nhất là đối với một số phép toán thao tác các tập.

Lưu quan niệm hàm thành viên đối với tập mở chỉ cần thỏa mãn các điều kiện ng hoặc s-dàng thì các khác biệt đó không phải là khác biệt cơ bản.

Mà khác, tồn tại sự tương ứng giữa tập hỗn độn với tập là hợp các tập conforme tính năng và các xúc suất cơ sở của lý thuyết Dempster-Š (hay cùng vậy, tập hỗn độn có thể → xúc suất trên) [4].

III. TẬP THỔ TRONG KHÔNG GIAN DO ĐƯỢC

để diễn tập thổ

Trong trường hợp hệ thống tin trên tập đôi tương là không gian do được, chúng ra ra một công thức "min hồn" cho định nghĩa tập thổ (theo ngôn ngữ hàm thành viên).

Cho \((\Omega, \mu)\) là không gian do được với do do \(\mu\), hợp các tập \(\mu\)-đo được được kiểu hiệu \(\sigma\-dài sở toàn tập do được). Ta giả thiết với mọi tập cơ bản \(e\) là \(\nu\)-do được và \(\nu > 0\), giả thiết \(X\) là \(\nu\)-do được. Định nghĩa tập thổ (theo thuật ngữ hàm thành viên) là bộ công thức như sau (giả sử \(x \in e\) tập cơ bản nào đó):

\[
fx(x) = \frac{\mu(X \cap e)}{\mu(e)}
\]

\(\Omega\) là hửu hàn thì ta có

\[
fx(x) = \frac{n(X \cap e)}{n(e)}
\]

đó \(n(\cdot)\) là số lượng phần tử có trong tập đôi sở.

Về hình thức, định nghĩa này cho biểu diễn gọn hơn so định nghĩa ban đầu. Về nội, trong định nghĩa hàm thành viên xấp xỉ min hồn.

Nhu vậy, với mọi tập con \(X\) là \(\nu\)-do được trên \(\Omega\), có một hàm xấp xỉ tương ứng \(\Omega\) cũng được gọi là tập thổ.

Chúng ta cần xác định (lâm rồ) tập thổ, song thực chất chưa biết đầy đủ thông tin về xác định nó. Như vậy, việc đưa ra các lớp tương ứng trong \(\Omega\), cho phép thay thế các deficiency có đây đủ thông tin về giá trị hàm thành viên đó với mỗi đối tương cụ

bằng một thông tin chung cho lớp nào đó chứa đối tương đó, đủ rằng việc phân loại tương ứng với tập có thể một cách "thờ" nào đó.

Trong y học, người ta quan tâm đến một nhóm các căn bệnh (một số bệnh có liên

với nhau), mỗi căn bệnh tương ứng với một tập người (bệnh nhân) là một tập con tập các đối tương \(\Omega\) (tập tất cả mọi người). Có thể có mỗi tập người bị một căn

nào đó là một tập chưa thể xác định. Bởi vậy, với từng căn bệnh, có thể phân lớp tương ứng theo những cách min hay thơ nào đó mà thường quan tâm đến các thông tin

ta thu, nghề nghiệp, tình trạng kinh tế, địa đạo sinh sống, thời quen sinh hoạt, các

chứng kiến quan đến căn bệnh v.v. Đây chính là các tiêu chuẩn đầu tiên để phân

Nhu vậy, các đối tương trong một lớp tương ứng e được đánh giá là đồng nhất

hà năng nhóm một căn bệnh nào đó (cùng điều kiện để nhóm bệnh). Các thông

tế tiêu chung làm sàng cho ta một cách phân hoach để tạo ra các tập cơ bản (xây

hệ thống tin).

Diểm qua các tính chất của ảnh xạ

\[
F : \Sigma \rightarrow \{\text{các hàm hỗn độn}\}
\]

tính chất này nhận được một cách tự nhiên từ định nghĩa của hàm hỗn độn và tính chất đó do.
a. \( f_X + f_Y = f_{X \cup Y} + f_{X \cap Y} \)

b. Tính đơn điều dời theo phép hợp: \( \max\{f_X, f_Y\} \leq f_{X \cup Y} \leq f_X + f_Y \)

c. Tính đơn điều dời theo phép giao: \( \min\{f_X, f_Y\} \geq f_{X \cap Y} \)

Nhu vậy một số tính chất của hàm thành viên trong tập mô không còn dùng tập thủ, chẳng hạn hai tính chất đặc trưng sau (m biểu thị hàm mô):

\[
\begin{align*}
    m(X \cup Y) &= \max\{m(X), m(Y)\} \\
    m(X \cap Y) &= \min\{m(X), m(Y)\}
\end{align*}
\]

b. Tính giá trị hàm thủ

Đang tiếp, trong cả hai định nghĩa (1) và (2), một số điều chung được làm rõ. X dạng được xem xét nội chung là chung xác định được (dù có thể quan sát được, số phần tử của nó). Điều đó cũng có nghĩa cách xác định giá trị của hàm thủ trước hai định nghĩa mang tính chất áp đặt. Lẽ tự nhiên, chúng ta có thể giải trình, tính định nghĩa (1) cũng như (2), các dạng giá trị trong so sánh tập hợp hoặc dời đó là có thực ứng dụng và thay vì những giá trị dũng, chúng ta chọn các giá trị ứng dụng. Ứng dụng này có thể nhận được hàng không thông kế toán học hoặc những kết quả một quá trình thủ thay ý kiến chuyên gia hoặc cả hai biện pháp. Qua cách tiến hành như vậy, đối với một tập X dường như không thể (hoặc chưa thể) xác định hoàn được thì để biết diễn nổ, ta xác định các tập thủ tương ứng.

IV. HỆ THÔNG TIN NỀN

Không gian các đối tượng \( (\Omega) \) của hệ thông tin có thể cơ là vụ trừ không thấy vặt tập thủ tồn tại khác quan. Như vậy, có thể quan niệm phân hoá các tập cơ mơi chỉ là mỗi cách làm thủ trong quá trình xác định các tập chung xác định. Với tập chung xác định, có thể có nhiều cách làm thủ “làm thủ” nó. Đặt ra vấn đề, có thể hay khả năng so sánh các cách làm thủ mới trên? Để trả lời câu hỏi này, chúng ta xem dụng định nghĩa các cách làm thủ tốt nhất để hướng tới cách xem dụng hệ thống tin có ý nghĩa trong ứng dụng. Một trong những vấn đề là đề xuất so sánh cần đưa ra các đối chung. Như cùng với tập mà, chúng ta có một múi đối chung (có thể quan niệm tương tự như niệm mồi quan trắc trong thống kê). Cho \( \{x_i, V(x_i) > V(x_i) \in \{0,1\}\} \), trong đó tập mô (thủ) đang được quan tâm.

Chọn hệ thống tin \( (\Omega, E) \) mà \( \Omega \) có hứa han phân tử, có các đại lượng:

\[
\begin{align*}
    n(e) &= ||x_i \in e|| \\
    \rho(e) &= ||x_i \in e, V(x_i) = 1||/n(e)
\end{align*}
\]

Chọn hai hệ thống tin \( (\Omega, E_1) \) và \( (\Omega, E_2) \). Nói hệ thống tin \( (\Omega, E_2) \) là nhỏ hơn so \( (\Omega, E_1) \). Nếu thỏa mãn điều kiện một tập có bxn trong hệ thống tin \( (\Omega, E_1) \) dược là tất cả được trong \( (\Omega, E_2) \). Nếu \( (\Omega, E_2) \) là nhỏ hơn so với \( (\Omega, E_1) \). thì gối \( (\Omega, E_1) \) là thủ hệ với \( (\Omega, E_2) \).

Chọn hai hệ thống tin \( (\Omega, E_1) \), \( (\Omega, E_2) \) với \( (\Omega, E_2) \). là nhỏ hơn \( (\Omega, E_1) \) và một lợp tập chung tương mình trên \( \Omega \).

Gọi hệ thống tin \( (\Omega, E_2) \). là thủ tốt hơn hệ thống tin \( (\Omega, E_1) \) nếu như:

\[
\Sigma n(e)|f_{E_2}(e) - \rho(e)| \leq \Sigma n(e')|f_{E_1}(e') - \rho(e')|
\]

trong đó \( f_{E_2}, f_{E_1}, \) đề chỉ cùng một hàm thủ trên các hệ thống tin \( (\Omega, E_2) \), \( (\Omega, E_1) \) tương ứng.

Chủ ý: Trong định nghĩa có thể bổ qua điều kiến về tính min của một hệ đối với h lại.
đi thắc độ định nghĩa: Trong định nghĩa, các hệ số $r(e)$, $n(e')$ để đánh giá trong số
thò đối với các tập cơ sở. Giả thành xấp xỉ (làm thô) không chi phụ thuộc vào
của xấp xỉ mà còn phụ thuộc vào giá trị của mỗi tập cơ bản (không phải các tập
1 cũng có một trong số khi xem xét). Việc thực hiện, có thể đạt được là càng làm
§ thông tin nên thì càng nhận được thông tin tốt hơn.

ật khác, ban than trong định nghĩa tập thơ là bảo đảm ý nghĩa là càng làm "thô"
bao nhiêu thì càng tốt bởi nhiều chi cần thôi mảnh kiến đảm bảo tính xấp
thiệt. Như vậy việc làm quá minh họa thông tin không phải là hường đi duy nhất
vi như thế có thể dẫn đến điều gì thì thu nhận được không tương xứng với công
ra khi làm min. Và như thế, lại có thể đạt ra vấn đề là: làm thơ hệ thống
nước do nào và theo những tiêu chuẩn nào để đảm bảo được giả định của việc làm
theo hai khả năng: giả thành và độ tin cậy. Bái toàn đạt ra là hoàn toàn hợp lý
thức thế thông qua việc sau đây:
theo các nghiên cứu trong [1], khi theo dõi bệnh lão màng não của trẻ nhỏ, chứng
án được các thông tin (với khoảng vài chục lôp như trong bảng thông kê [1]) về
nhan được mô hình qua một hệ thống tin $S$. Các thông tin trong hệ phỏng sử dụng $S$
tập hợp từ nhiều nguồn, theo các điều kiện phần cấp (các tuyển khám bệnh, xét
m và điều trị). Tuy vậy, ngày càng trong cấp độ cao nhất (có đõ các thuốc bệnh)
tin về bệnh lão màng não cùng chửa đầy đủ. Và để đạt ra, có mối liên hệ
đối giữa các tiêu chuẩn (các thông tin về bệnh nhân) với bệnh. Thông qua mối liên
ý, bài viết [1] cho một phương pháp nhân biết thông tin tình trạng bệnh thông qua
việc chứng. Phương pháp sử dụng trong bài viết đó là tính thông thông kế các giá
trong quan.

ý phải tướng ta quan tâm đến một cách để cấp khắc. Tập thơ em bi bệnh lão màng
lại một tập thơ trên hệ thông tin $S$. Với các bác sĩ ở các tuyển bệnh (chi quan sát
một số tổ chức không trong tập tổ các tiêu chuẩn theo dõi), có nghia hệ thống
ước xem xét là thơ holistic với hệ thống tin của các bệnh viện trong quan. Một vấn
m được xem xét: tập thơ biểu thị bệnh nói trên (nhan được khi xem xét $S$) khi hệ
bi rút gọn nhưng vậy còn sự dụng được không? Việc khẳng định khả năng sử dụng
thu gọn sẽ được đánh giá theo các tiêu chuẩn nào? Liền quan đến mô hình hóa
ống đã cho [1], chúng ta còn phải xem đúng các hệ thống tin với thông tin chua
thu (chưa xác định). Giả quyet cụ thể các vấn đề đó ra sao?

V. HỆ THỐNG TIN THU GỌI VÀ THUỘC TÍNH XÁC ĐỊNH TÁP THƠ

định nghĩa: Cho hệ thống tin $S$ và tập các tập cơ bản $E$. Hệ thống tin $S'$ thơ hon
óc gọi là hệ thu gọn của $S$ đối với lôp tập thơ $F$ nếu thỏa điều kiến: Với mỗi
$ f|s = f|s' $.

định nghĩa của định nghĩa này newValue: Tính phân bố đều của các hàm thơ thuộc $F$
ở các thuốc tình bị hiểu bô.

1: Xét hệ thống tin $S$ là hệ thống tin liên quan đến tất cả các thuốc tình được cho
thống kể tất cả bệnh nhân viêm màng não có trong [1]. Có thể chi ra một $S'$ là
ông tin nhận được từ $S$ khi bỏ đi các thuốc tình chẳng hạn thuộc tình hoa.

lý nghĩa trên định hướng thu gọn hệ thông tin (tính giống nó thông qua việc gian
ng các thuốc tình) để làm đơn giản hoá lại các bài toàn.

hai miền hệ thống thu gọn nhằm đạt được mục đích thu gọn hệ thông tin thông
để bô bót các thuốc tình "vô ích". Tuy nhiên, trong nhiều trường hợp, không chi
đó mức quân lớn đến một lôp tập thơ, mà quan trọng hơn là nhận được các chỉ
phan biệt tập thơ này với tập thơ khác hay cùng vậy nhận biết một tập thơ trong
er cho nào đó.

định nghĩa: Cho $S$ là một hệ thống tin còn $a$ là một thuốc tình $(a \in A)$; cho $L$ là một
lập các tập thể trong $S$ với $f \in L$. Nói rằng $a$ là xác định $f$ nếu như:

$$\forall z = \sigma(x, y) \quad f \rightharpoonup g \mid e \quad \text{hoặc} \quad g \rightharpoonup f \mid e \quad (\forall e \in E_S; \ e \in Z; \ \forall g \in L, \ g \neq f)$$

Kì hiệu $\rightharpoonup$ đề chỉ quan hệ "lớn hơn rất nhiều".

Chú ý rằng, ý nghĩa của định nghĩa này có thể thông tin về tập thể $f$ sẽ được hạn so với các tập thể còn lại.

Ví dụ: Trong hệ thống tin về bệnh viêm màng não của trẻ em thì thuộc tính đặc tulty là $f$ xác định bệnh viêm màng não.

Cho $S$ là một hệ thống tin cơn $S'$ là một hệ thống tin hỗn $S$; cho $L$ là một các tập thể trong $S$ với $f \in L$. Nói rằng $S'$ là xác định $f$ nếu như:

$$f \rightharpoonup g \mid S' \quad \forall g \in L, \ g \neq f, \ g \neq 0$$

(trong cộng thức trên bỏ qua mỗi trường hợp phân số vô nghĩa).

Ví dụ, trong hệ thống các bệnh nhân trẻ em có triệu chứng của bệnh lao màng
Tập các tập chưa xác định trên $S'$ là $L = \{\text{bệnh lao màng não, bệnh sốt cao, bệnh}\}$. Chứng ta chú ý hệ thống tin hỗn $S'$ bao gồm các thuộc tính: sốt cao, dịch não to
1. Lúc đó, $S'$ xác định bệnh lao màng não.

TÀI LIỆU THAM KHẢO


VNU. JOURNAL OF SCIENCE, Nat. Sci., t. XII, n°3, 1996

ROUGH SETS ESTIMATION OF INFORMATION SYSTEMS

Ha Quang Thuy

College of Natural Sciences - VNU

Information systems and the concept of rough sets have been carried out by Pawlak. For information systems, Pawlak gave the question language for problems of se objects which answer is a description set. He also concerned with rough sets.

In this article, we introduce a definition of rough sets in a measurable space objects. Since the determination of rough sets depends on information systems, then the comparison of problem of information systems is dealt with. For a given rough set, it is possible to diminish the size of information systems by saving the case for recognizing the rough set.