

MỞ RỘNG MÔ HÌNH CƠ SỞ ĐỐI TƯỢNG XÁC SUẤT VỚI GIÁ TRỊ TÍNH CHẤT LỚP KHÔNG CHẮC CHẮN VÀ KHÔNG CHÍNH XÁC

NGUYỄN HÒA*

1. GIỚI THIỆU

Thực tế đã cho thấy hướng đối tượng là một phương pháp hiệu quả để mô hình hóa, thiết kế và hiện thực các hệ thống lớn. Tuy nhiên, mô hình cơ sở dữ liệu (CSDL) hướng đối tượng truyền thống không biểu diễn và xử lý được thông tin không chắc chắn và không chính xác của các đối tượng trong thực tế. Điều này đã đòi hỏi và thúc đẩy việc nghiên cứu và phát triển các mô hình CSDL hướng đối tượng xác suất và mờ. Tuy nhiên cho đến nay ít có mô hình kết hợp được cả hai yếu tố không chắc chắn và không chính xác trên một nền tảng lý thuyết chặt chẽ. Hơn nữa, thật sự không có mô hình nào có thể biểu diễn và xử lý hết mọi khía cạnh không chắc chắn và không chính xác về thông tin trong thế giới thực ([2], [3]). Vì vậy, các mô hình CSDL xác suất và mờ vẫn được tiếp tục nghiên cứu để đáp ứng các mục tiêu ứng dụng khác nhau ([4], [5], [6]).

Gần đây, Eiter và cộng sự (2001) đã giới thiệu một mô hình cơ sở đối tượng xác suất gọi là POB (Probabilistic Object Base). Mô hình POB được xây dựng dựa trên cơ sở toán học vững chắc và nhất quán với mô hình cơ sở dữ liệu hướng đối tượng truyền thống, có khả năng biểu diễn và truy vấn thông tin không chắc chắn về các đối tượng trong thế giới thực. Tuy nhiên, thiếu sót chính của mô hình này là chưa cho phép biểu diễn các giá trị thuộc tính không chính xác và các phương thức của một lớp.

Bài báo này giới thiệu một mở rộng mô hình POB thành mô hình cơ sở đối tượng xác suất mờ FPOB (Fuzzy Probabilistic Object Base) với hai đặc tính chính: (1) các giá trị không chắc chắn và không chính xác của một *thuộc tính lớp* (class attribute) được biểu diễn bởi một khoảng phân bố xác suất trên một tập các giá trị tập mờ; (2) các *phương thức lớp* (class method) với các đối số và giá trị không chắc chắn và không chính xác được tích hợp một cách hình thức vào mô hình mới.

Cơ sở toán học để phát triển FPOB được trình bày trong Phần 2, lược đồ và thể hiện FPOB được giới thiệu trong Phần 3. Phần 4 trình bày về truy vấn trên FPOB và cuối cùng, Phần 5 là một số kết luận và hướng nghiên cứu trong tương lai.

2. CƠ BẢN VỀ XÁC SUẤT VÀ TẬP MỜ

Phần này giới thiệu cơ sở toán để xây dựng FPOB. Các chiến lược kết hợp các khoảng xác suất đã được giới thiệu trong POB. Chúng tôi đề xuất diễn dịch xác suất các quan hệ trên các tập mờ, các bộ ba xác suất mờ và đại số trên các bộ ba xác suất mờ.

2.1 Các chiến lược kết hợp các khoảng xác suất

Cho hai sự kiện e_1 và e_2 với các xác suất tương ứng trong các khoảng $[L_1, U_1]$ và $[L_2, U_2]$. Khi đó các khoảng xác suất biểu diễn cho các sự kiện hội $e_1 \wedge e_2$, tuyển $e_1 \vee e_2$, hiệu $e_1 \wedge \neg e_2$ của hai sự kiện e_1 và e_2 có thể được tính toán bởi các *chiến lược kết hợp xác suất* (probabilistic combination strategy). Bảng 2.1 là một ví dụ về các chiến lược kết hợp xác suất, trong đó $\otimes$, $\oplus$ và $\ominus$ tương ứng biểu thị các phép toán hội, tuyển và trừ.

* Tiến sĩ, Giáo viên cơ hữu Khoa CNTT, Trường Đại học Mở TP. Hồ Chí Minh

Bảng 2.1 Các ví dụ về các chiến lược kết hợp xác suất

Chiến lược	Phép toán
Bỏ qua (Ignorance)	$([L_1, U_1] \otimes_{ig}[L_2, U_2]) \equiv [\max(0, L_1 + L_2 - 1), \min(U_1, U_2)]$ $([L_1, U_1] \oplus_{ig}[L_2, U_2]) \equiv [\max(L_1, L_2), \min(1, U_1 + U_2)]$ $([L_1, U_1] \ominus_{ig}[L_2, U_2]) \equiv [\max(0, L_1 - U_2), \min(U_1, 1 - L_2)]$
Độc lập (Independence)	$([L_1, U_1] \otimes_{in}[L_2, U_2]) \equiv [L_1 \cdot L_2, U_1 \cdot U_2]$ $([L_1, U_1] \oplus_{in}[L_2, U_2]) \equiv [L_1 + L_2 - (L_1 \cdot L_2), U_1 + U_2 - (U_1 \cdot U_2)]$ $([L_1, U_1] \ominus_{in}[L_2, U_2]) \equiv [L_1 \cdot (1 - U_2), U_1 \cdot (1 - L_2)]$
Loại trừ nhau (Mutual Exclusion)	$([L_1, U_1] \otimes_{me}[L_2, U_2]) \equiv [0, 0]$ $([L_1, U_1] \oplus_{me}[L_2, U_2]) \equiv [\min(1, L_1 + L_2), \min(1, U_1 + U_2)]$ $([L_1, U_1] \ominus_{me}[L_2, U_2]) \equiv [L_1, \min(U_1, 1 - L_2)]$

2.2 Diễn dịch xác suất của các quan hệ trên tập mờ

Diễn dịch xác suất các quan hệ hai ngôi trên các tập mờ (probabilistic interpretation of binary relations on fuzzy sets) là cơ sở để tính toán xác suất của các quan hệ giữa các giá trị tính chất của đối tượng được biểu diễn bởi các tập mờ trong FPOB. Dựa trên cơ sở *phép gán khối* (mass assignment) trong [1], chúng tôi đề nghị các diễn dịch (nghĩa là độ đo) xác suất của các quan hệ hai ngôi trên các tập mờ như sau.

Định nghĩa 2.2.1 Giả sử A, B là các tập mờ tương ứng trên các miền U và V , θ là một quan hệ hai ngôi từ $\{=, \leq, <, \subseteq, \in\}$. *Diễn dịch xác suất* của quan hệ $A \theta B$, ký hiệu $prob(A \theta B)$ là một giá trị trong khoảng $[0, 1]$ được định nghĩa bởi:

$$\sum_{s \subseteq U, T \subseteq V} \Pr(u \theta v \mid u \in S, v \in T) \cdot m_A(S) \cdot m_B(T).$$

Định nghĩa 2.2.2 Giả sử A và B là hai tập mờ trên một miền U . *Diễn dịch xác suất* của quan hệ $A \rightarrow B$, ký hiệu $prob(A \rightarrow B)$, là một giá trị trong khoảng $[0, 1]$ được định nghĩa bởi:

$$\sum_{s, T \subseteq U} \Pr(u \in T \mid u \in S) \cdot m_A(S) \cdot m_B(T).$$

2.3 Đại số các bộ ba xác suất mờ

Bộ ba xác suất mờ và đại số các bộ ba xác suất mờ, là cơ sở để biểu diễn và thao tác trên các giá trị tập mờ không chắc chắn của các đối tượng, được định nghĩa như sau.

Định nghĩa 2.3.1 Một *bộ ba xác suất* (probabilistic triple) $\langle X, \alpha, \beta \rangle$ bao gồm một tập hữu hạn X , một hàm phân bố xác suất α trên X và một hàm $\beta: X \rightarrow [0, 1]$ sao cho $\alpha(x) \leq \beta(x)$. Nếu các phần tử của X là các tập mờ thì $\langle X, \alpha, \beta \rangle$ được gọi là một *bộ ba xác suất mờ* (fuzzy probabilistic triple).

Định nghĩa 2.3.2 Cho $\mathcal{V} = \{\langle V, \alpha, \beta \rangle \mid V \subseteq U\}$ là một tập các bộ ba xác suất mờ khác rỗng trên tập U khác rỗng. Nếu $A = (U, o_1, \dots, o_n)$ là một đại số với các phép toán $o_1, \dots, o_n$ trên U , thì $\mathcal{A} = (\mathcal{V}, o_1, \dots, o_n)$ là một đại số, được gọi là *đại số các bộ ba xác suất mờ* (fuzzy probabilistic triple algebra) trên $\mathcal{V}$, với các phép toán $o_1, \dots, o_n$ trên $\mathcal{A}$ được suy dẫn từ A như sau:

1. $o_j(\langle V_1, \alpha_1, \beta_1 \rangle, \langle V_2, \alpha_2, \beta_2 \rangle, \dots, \langle V_{m_j}, \alpha_{m_j}, \beta_{m_j} \rangle) = \langle V, \alpha, \beta \rangle$, trong đó $V = \{v = o_j(v_1, v_2, \dots, v_{m_j}) \mid v_i \in V_i, i = 1, \dots, m_j\}$, m_j là số đôi số của o_j và
2. $[\alpha(v), \beta(v)] = \bigoplus_{me: v_1 \in V_1, v_2 \in V_2, \dots, v_{m_j} \in V_{m_j}, v = o_j(v_1, v_2, \dots, v_{m_j})} [\alpha_1(v_1), \beta_1(v_1)] \otimes_j [\alpha_2(v_2), \beta_2(v_2)] \otimes_j \dots \otimes_j [\alpha_{m_j}(v_{m_j}), \beta_{m_j}(v_{m_j})]$, với mọi $v \in V$ và $\otimes_j$ là một chiến lược hội xác suất và $\forall j = 1, \dots, n$.

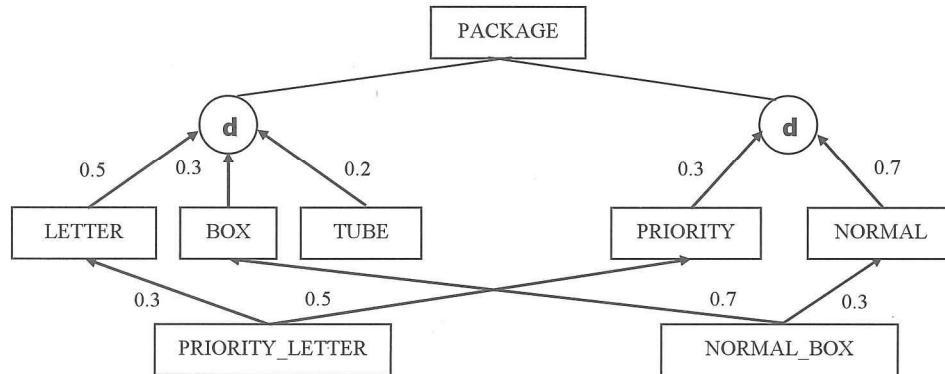
Ví dụ, nếu o_j là các phép toán + và $\times$ các số mờ trên tập các số thực theo *nguyên lý mở rộng* (extension principle), thì o_j tương ứng là các phép toán + và $\times$ các bộ ba xác suất mờ $\langle V, \alpha, \beta \rangle$ và ta có một đại số các bộ ba xác suất mờ trên tập số thực.

3. LƯỢC ĐỒ VÀ THỂ HIỆN CỦA MÔ HÌNH FPOB

Các khái niệm lược đồ và thể hiện trong FPOB được mở rộng từ các khái niệm tương ứng trong POB với tập mờ và phương thức lớp.

3.1 Mô hình ý niệm FPOB

Trong FPOB thuộc tính và phương thức đối tượng có thể nhận giá trị tập mờ. Chẳng hạn, một cơ sở dữ liệu về các gói bưu kiện được vận chuyển bởi một công ty vận tải có thể được mô hình hóa bởi FPOB. Trong cơ sở dữ liệu này, thời gian vận chuyển của mỗi gói bưu kiện không phải luôn luôn được xác định một cách chắc chắn. Tương tự, kích thước của mỗi gói (chiều dài, rộng, cao) cũng như diện tích, thể tích và do đó cả cước phí vận chuyển cũng không phải luôn luôn được đo và được tính toán một cách chính xác. Các giá trị như thế có thể là mờ và không chắc chắn. Hình 3.1 cho thấy một ví dụ đơn giản về sự phân cấp các lớp gói bưu kiện trong mô hình FPOB.



Hình 3.1: Một ví dụ phân cấp lớp trong FPOB

Các lớp con của một lớp liên kết với một nút **d** là loại trừ lẫn nhau (nghĩa là một đối tượng không thể thuộc về hai lớp tại cùng một thời điểm). Trong ví dụ này lớp PACKAGE có hai nhóm lớp con là {LETTER, BOX, TUBE} và {PRIORITY, NORMAL}.

Các giá trị số trong khoảng $[0, 1]$ trên các cung liên kết giữa một lớp với lớp con trực tiếp của nó biểu diễn xác suất có điều kiện để một đối tượng thuộc lớp cha là thuộc lớp con của nó. Chẳng hạn, sự phân cấp này chỉ ra rằng một đối tượng bất kỳ của

PACKAGE có 70% khả năng thuộc về NORMAL trong khi chỉ có 30% khả năng còn lại thuộc về PRIORITY.

3.2 Kiểu, giá trị và lược đồ FPOB

Đối với FPOB, hệ thống kiểu và giá trị của chúng được mở rộng từ POB. Trong đó, không chỉ khái niệm kiểu tập hợp được mở rộng thành kiểu tập hợp mờ và giá trị kiểu này thành giá trị tập hợp mờ, mà khái niệm kiểu bộ cũng được mở rộng để cho phép tích hợp các phương thức vào các lớp đối tượng như trong các định nghĩa dưới đây.

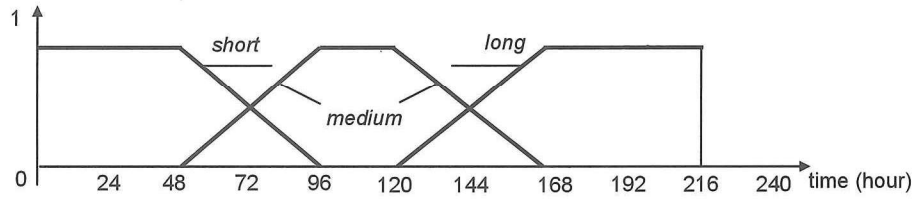
Định nghĩa 3.2.1 Giả sử $\mathcal{P}$ là một tập các tính chất và $\mathcal{T}$ là một tập các kiểu cơ sở. Các kiểu được định nghĩa một cách đệ quy như sau:

1. Mỗi kiểu cơ sở trong $\mathcal{T}$ là một kiểu.
2. Nếu τ là một kiểu, thì $\{\tau\}$ là một kiểu, được gọi là *kiểu tập mờ* (fuzzy set type) của τ .
3. Nếu $P_1, P_2, \dots, P_k$ là các tính chất đôi một phân biệt trong $\mathcal{P}$, τ_i và τ_{ij} với mọi i từ 1 đến k và j từ 1 đến n_i là các kiểu, thì $\tau = [P_1(\tau_{11}, \tau_{12}, \dots, \tau_{1n_1}): \tau_1, P_2(\tau_{21}, \tau_{22}, \dots, \tau_{2n_2}): \tau_2, \dots, P_k(\tau_{k1}, \tau_{k2}, \dots, \tau_{kn_k}): \tau_k]$ là một kiểu được gọi là *kiểu bộ* (tuple type) trên $\{P_1, P_2, \dots, P_k\}$. Chúng tôi sử dụng $\tau.P_i$ để biểu thị τ_i và gọi $P_1, P_2, \dots, P_k$ là các *tính chất mức cao nhất* (top-level properties), có kiểu tương ứng $\tau_1, \tau_2, \dots, \tau_k$ của τ .

Như trong [6], ở định nghĩa trên, các τ_{ij} biểu diễn các đối số của P_i khi nó là một phương thức và chúng là rỗng khi P_i là thuộc tính. Tuy nhiên, trong một ngữ cảnh nào đó, các đối số τ_{ij} có thể được bỏ qua nếu thấy không nhất thiết phải kể đến chúng.

Định nghĩa 3.2.2 Giả sử $P_1, P_2, \dots, P_k$ là các tính chất đôi một khác nhau trong $\mathcal{P}$, V_i và V_{ij} là các tập hữu hạn các giá trị kiểu τ_i và τ_{ij} , $[\alpha_i, \beta_i]$ và $[\alpha_{ij}, \beta_{ij}]$ là các cặp phân bố xác suất tương ứng trên V_i và V_{ij} , với mọi i từ 1 đến k và j từ 1 đến n_i . Thì $fptv = [P_1(\langle V_{11}, \alpha_{11}, \beta_{11} \rangle, \langle V_{12}, \alpha_{12}, \beta_{12} \rangle, \dots, \langle V_{1n_1}, \alpha_{1n_1}, \beta_{1n_1} \rangle): \langle V_1, \alpha_1, \beta_1 \rangle, P_2(\langle V_{21}, \alpha_{21}, \beta_{21} \rangle, \langle V_{22}, \alpha_{22}, \beta_{22} \rangle, \dots, \langle V_{2n_2}, \alpha_{2n_2}, \beta_{2n_2} \rangle): \langle V_2, \alpha_2, \beta_2 \rangle, \dots, P_k(\langle V_{k1}, \alpha_{k1}, \beta_{k1} \rangle, \langle V_{k2}, \alpha_{k2}, \beta_{k2} \rangle, \dots, \langle V_{kn_k}, \alpha_{kn_k}, \beta_{kn_k} \rangle): \langle V_k, \alpha_k, \beta_k \rangle]$ là một *giá trị bộ xác suất mờ* (fuzzy-probabilistic tuple value) kiểu $[P_1(\tau_{11}, \tau_{12}, \dots, \tau_{1n_1}): \tau_1, P_2(\tau_{21}, \tau_{22}, \dots, \tau_{2n_2}): \tau_2, \dots, P_k(\tau_{k1}, \tau_{k2}, \dots, \tau_{kn_k}): \tau_k]$ trên $\{P_1, P_2, \dots, P_k\}$. Ký hiệu $fptv.P_i$ được sử dụng để biểu thị $\langle V_i, \alpha_i, \beta_i \rangle$.

Ví dụ 3.2.2 Trong cơ sở đối tượng các gói bưu kiện đã giới thiệu ở trên, gọi *short*, *medium*, *long* là các nhãn ngôn ngữ của các tập mờ trên tập số thực có đồ thị hàm thành viên như trong Hình 3.2.1. Giả sử, có một gói sẽ được chuyển từ Sài Gòn đến Hà Nội nhưng chúng ta chỉ biết rằng 30%-60% khả năng nó sẽ được chuyển trong thời gian ngắn hoặc trung bình mà không biết chắc chắn và chính xác nó được vận chuyển trong bao lâu. Ngoài ra, chúng ta cũng không chắc chắn chiều dài của nó là 36 hay 38, chiều rộng của nó là 18 hay 19, thì các thông tin của gói này có thể được biểu diễn bởi giá trị bộ xác suất mờ [origin: $\langle \{Saigon\}, u, u \rangle$, destination: $\langle \{Hanoi\}, u, u \rangle$, time: $\langle \{short, medium\}, 0.6u, 1.2u \rangle$, size: $\langle \{[length: 36, width: 18], [length: 38, width: 19]\}, u, u \rangle$, trong đó u là hàm phân bố đều.



Hình 3.2.1: Các giá trị tập mờ của thuộc tính **time**

Khái niệm giá trị bộ xác suất mờ cho phép biểu diễn một cách thích hợp thông tin không chắc chắn và không chính xác của các đối tượng trong FPOB. Bây giờ lược đồ FPOB là mở rộng lược đồ POB để có thể biểu diễn được các tính chất lớp cùng với định nghĩa hàm của chúng như sau.

Định nghĩa 3.2.3 Một lược đồ FPOB (FPOB-schema) là một bộ sáu $S = (C, \tau, \Rightarrow, me, \wp, f)$, trong đó:

1. C là một tập hữu hạn các lớp (đó là các lớp được kết hợp với FPOB).
2. τ là một ánh xạ đặt tương ứng mỗi lớp $c \in C$ với một kiểu bộ $\tau(c)$, biểu diễn các tính chất và kiểu của chúng trong lớp đó.
3. $\Rightarrow$ là một quan hệ hai ngôi trên C sao cho $(C, \Rightarrow)$ là một đồ thị có hướng không có chu trình, mỗi một đỉnh của $(C, \Rightarrow)$ là một lớp trong C , mỗi cạnh $c_1 \Rightarrow c_2$ biểu diễn c_1 là một lớp con trực tiếp của c_2 .
4. me là một ánh xạ đặt tương ứng mỗi lớp $c \in C$ với một phân hoạch của tập tất cả các lớp con trực tiếp của c sao cho các lớp trong mỗi nhóm của $me(c)$ là tách rời và loại trừ lẫn nhau.
5. $\wp$ là một ánh xạ đặt tương ứng mỗi cạnh $c \Rightarrow d$ trong $(C, \Rightarrow)$ với một số hữu tỉ trong khoảng $[0, 1]$ sao cho $\sum_{c \in \wp} \wp(c, d) \leq 1, \forall d \in C, \forall \wp \in me(d)$.
6. f là một ánh xạ đặt tương ứng mỗi tính chất $P_i(\tau_{i1}, \tau_{i2}, \dots, \tau_{in_i})$: τ_i trong C với một hàm từ các tích Descartes của các bộ ba xác suất mờ kiểu τ_{ij} đến các bộ ba xác suất mờ kiểu τ_i . Về trực giác, $f(P_i)$ là định nghĩa hàm của P_i .

Ví dụ 3.2.3 Một lược đồ FPOB cho cơ sở đối tượng các gói bưu kiện đã nêu trên có thể được định nghĩa như sau:

$C = \{\text{PACKAGE, LETTER, BOX, TUBE, PRIORITY, NORMAL, PRIORITY_LETTER, NORMAL_BOX}\}$.

τ được cho trong Bảng 3.2.1, với lưu ý là các lớp con có thể thừa kế tính chất từ các lớp cha của chúng.

$(C, \Rightarrow)$, me và $\wp$ được cho như trong Hình 3.1.

f định nghĩa các phương thức **area** và **volume** tương ứng trong các lớp LETTER và BOX bằng cách áp dụng đại số bộ ba xác suất mờ trong Định nghĩa 2.3.2 như sau:

LETTER: $\text{area}([\text{length}: \langle V_1, \alpha_1, \beta_1 \rangle], [\text{width}: \langle V_2, \alpha_2, \beta_2 \rangle]): \langle V, \alpha, \beta \rangle$

1. $\langle V, \alpha, \beta \rangle := \langle V_1, \alpha_1, \beta_1 \rangle \times \langle V_2, \alpha_2, \beta_2 \rangle$
2. **return** $\langle V, \alpha, \beta \rangle$.

BOX: volume(length: $\langle V_1, \alpha_1, \beta_1 \rangle$, width: $\langle V_2, \alpha_2, \beta_2 \rangle$, height: $\langle V_3, \alpha_3, \beta_3 \rangle$): $\langle V, \alpha, \beta \rangle$

1. $\langle V, \alpha, \beta \rangle = \langle V_1, \alpha_1, \beta_1 \rangle \times \langle V_2, \alpha_2, \beta_2 \rangle \times \langle V_3, \alpha_3, \beta_3 \rangle$

2. **return** $\langle V, \alpha, \beta \rangle$.

Bảng 3.2.1: Ánh xạ gán kiểu τ trong FPOB

c	$\tau(c)$
PACKAGE	[origin: string, destination: string, time: {real}]
LETTER	[length: {real}, width: {real}, area([length: {real}], [width: {real}]): {real}]
BOX	[length: {real}, width: {real}, height: {real}, volume([length: {real}], [width: {real}], [height: {real}]): {real}]
TUBE	[diameter: {real}, height: {real}, volume([diameter: {real}], [height: {real}]): {real}]
PRIORITY	[priority_level: integer]
NORMAL	[stop_over: string]
PRIORITY_LETTER	[]
NORMAL_BOX	[]

3.3 Thừa kế và thể hiện FPOB

FPOB mở rộng *chiến lược thừa kế* (inheritance strategy) trong POB cho các phương thức lớp. Với mỗi lược đồ FPOB $S = (C, \tau, \Rightarrow, me, \wp, f)$ có một lược đồ $S^* = (C, \tau^*, \Rightarrow, me, \wp, f^*)$ thừa kế S chỉ khác S ở phép gán kiểu τ^* và định nghĩa hàm f^* của các phương thức lớp. Cụ thể, với mỗi $c \in C$, $\tau^*(c) = [P_1: \tau(d_1).P_1, \dots, P_k: \tau(d_k).P_k]$ và $f^*(P_i) = f_{d_i}(P_i)$, trong đó P_i , với mọi $i = 1, \dots, k$, tương ứng là tính chất mức cao nhất của các lớp cha $d_1, \dots, d_k$ của c , và f_{d_i} là diễn dịch hàm thu hẹp của f trên các tính chất của lớp d_i .

Cho một lược đồ S , một thể hiện FPOB trên S được định nghĩa như một cơ sở đối tượng phù hợp với các kiểu dữ liệu và chiến lược thừa kế trên lược đồ này. Một thể hiện trên lược đồ FPOB cho thấy được tình trạng, quan hệ thừa kế của một tập đối tượng trong FPOB. Thể hiện của cơ sở đối tượng xác suất mờ được định nghĩa như sau.

Định nghĩa 3.3.1 Giả sử $S = (C, \tau, \Rightarrow, me, \wp, f)$ là một lược đồ FPOB và O là một tập các định danh đối tượng, một *thể hiện FPOB* (FPOB-instance) trên S là một cặp (π, ν) , trong đó:

1. $\pi: C \rightarrow 2^O$ là ánh xạ đặt tương ứng mỗi lớp c thuộc C với một tập con hữu hạn của tập O , sao cho $\pi(c_1) \cap \pi(c_2) = \emptyset, \forall c_1 \neq c_2 \in C$.
2. ν là ánh xạ đặt tương ứng mỗi $o \in \pi(C)$ với một giá trị bộ xác suất mờ kiểu $\tau(c)$ sao cho $o \in \pi(c)$.

Ví dụ 3.3.1 Một thể hiện $I = (\pi, \nu)$ trên lược đồ FPOB trong Ví dụ 4.3.1 được chỉ ra trong Bảng 3.3.1. Trong đó $\pi(\text{PACKAGE}) = \{o_1\}$, $\pi(\text{PRIORITY_LETTER}) = \{o_2\}$ và $\pi(\text{NORMAL_BOX}) = \{o_3\}$.

Bảng 3.3.1: Ánh xạ gán giá trị v trong FPOB

oid	$v(oid)$
o_1	[origin: $\langle\{Hanoi\}, u, u\rangle$, destination: $\langle\{Saigon\}, u, u\rangle$, time: $\langle\{24, 48, \dots, 240\}, u, u\rangle$]
o_2	[origin: $\langle\{Saigon\}, u, u\rangle$, destination: $\langle\{Hue\}, u, u\rangle$, length: $\langle\{30, 32\}, 0.8u, 1.8u\rangle$, width: $\langle\{22\}, u, u\rangle$, priority_level: $\langle\{1\}, u, u\rangle$, time: $\langle\{24, 48\}, 0.9u, 1.2u\rangle$, area: $\langle\{660, 704\}, 0.8u, 1.8u\rangle$]
o_3	[origin: $\langle\{Saigon\}, u, u\rangle$, destination: $\langle\{Hue\}, u, u\rangle$, length: $\langle\{40, 42\}, 0.8u, 1.4u\rangle$, width: $\langle\{25\}, u, u\rangle$, height: $\langle\{about_25\}, u, u\rangle$, stop_over: $\langle\{Nhatrang, Quinhon\}, 0.8u, 1.6u\rangle$, time: $\langle\{medium, long\}, 0.8u, 1.6u\rangle$, volume: $\langle\{about_25000, about_26250\}, 0.8u, 1.4u\rangle$]

4. TRUY VẤN TRÊN FPOB

Để truy vấn thông tin trong cơ sở đối tượng xác suất mờ, các phép toán đại số trên FPOB như *chọn* (selection), *chiếu* (projection) vv..., đã được xây dựng. Phần này giới thiệu phép chọn và các truy vấn thông qua phép chọn. Trước hết, các khái niệm biểu thức đường đi, biểu thức chọn và điều kiện chọn lần lượt được mở rộng từ POB như sau.

Định nghĩa 4.1 Giả sử $\tau = [P_1: \tau_1, \dots, P_k: \tau_k]$ là một kiểu bất kỳ. *Biểu thức đường đi* (path expression) được định nghĩa một cách đệ quy cho mọi i từ 1 đến k như sau:

1. P_i là một biểu thức đường đi cho τ .
2. Nếu λ_i là một biểu thức đường đi cho τ_i thì $P_i.\lambda_i$ là một biểu thức đường đi cho τ .

Các biểu thức chọn mờ trên FPOB, như là những phát biểu các ràng buộc trên các giá trị của các tính chất cuối cùng trong các biểu thức đường đi, được mở rộng từ các biểu thức chọn trong POB bằng cách sử dụng các quan hệ hai ngôi trên các tập mờ như định nghĩa dưới đây.

Định nghĩa 4.2 Giả sử $S = (C, \tau, \Rightarrow, me, \wp, f)$ là một lược đồ FPOB và X là một tập các biến đối tượng. Các *biểu thức chọn mờ* (fuzzy selection expression) được định nghĩa một cách đệ quy và có một trong các dạng sau:

1. $x \in c$, trong đó $x \in X$ và $c \in C$.
2. $x.\lambda \theta v$, trong đó $x \in X$, λ là một biểu thức đường đi, θ là một quan hệ hai ngôi thuộc $\{=, \neq, \leq, <, \subseteq, \in, \rightarrow\}$ và v là một giá trị.
3. $x.\lambda_1 =_{\otimes} x.\lambda_2$, trong đó $x \in X$, λ_1 và λ_2 là hai biểu thức đường đi phân biệt và $\otimes$ là một chiến lược hội xác suất kết hợp các xác suất để $x.\lambda_1 = v_1$ và $x.\lambda_2 = v_2$ sao cho $v_1 = v_2$.
4. $E_1 \otimes E_2$, trong đó E_1 và E_2 là các biểu thức chọn trên cùng một biến đối tượng, $\otimes$ là một chiến lược hội xác suất kết hợp các xác suất để E_1 và E_2 đúng.
5. $E_1 \oplus E_2$, trong đó E_1 và E_2 là các biểu thức chọn trên cùng một biến đối tượng, $\oplus$ là một chiến lược tuyển xác suất kết hợp các xác suất để E_1 và E_2 đúng.

Trong FPOB, mỗi điều kiện chọn mờ là sự kết hợp logic của các biểu thức chọn mờ cùng với các khoảng xác suất cần được thỏa mãn như định nghĩa sau.

Định nghĩa 4.3 Giả sử $S = (C, \tau, \Rightarrow, me, \wp, f)$ là một lược đồ FPOB, các *điều kiện chọn mờ* (fuzzy selection condition) được định nghĩa một cách đệ quy như sau:

1. Nếu E là một biểu thức chọn mờ, L và U là các số thực trong khoảng $[0, 1]$, $L \leq U$ thì $(E)[L, U]$ là một điều kiện chọn mờ.
2. Nếu ϕ và ψ là các điều kiện chọn mờ trên cùng một biến đối tượng thì $\neg\phi$, $(\phi \wedge \psi)$, $(\phi \vee \psi)$ là các điều kiện chọn mờ.

Ví dụ 4.3 Gọi S là lược đồ của thể hiện trong Ví dụ 3.3.1, thì “chọn tất cả các gói có thể tích lớn hơn 20,000 và được vận chuyển trong thời gian trung bình với xác suất từ 0.3 đến 0.5” có thể được thực hiện bởi điều kiện chọn

$$(x.\text{volume}(x.\text{length}, x.\text{width}, x.\text{height}) > 20,000 \otimes x.\text{time} \rightarrow \text{medium})[0.3, 0.5].$$

Định nghĩa 4.4 Giả sử $\tau = [P_1: \tau_1, \dots, P_k: \tau_k]$ là một kiểu bộ, *diễn dịch* (interpretation) của một biểu thức đường đi λ đối với τ theo giá trị $v = [P_1: v_1, \dots, P_k: v_k]$, biểu thị là $v.\lambda$, được định nghĩa đệ qui như sau:

1. Nếu $\lambda = P_i$ thì $v.\lambda = v_i$
2. Nếu $\lambda = P_i.\lambda_i$ trong đó λ_i là một biểu thức đường đi cho τ_i thì $v.\lambda = v_i.\lambda_i$

Đối với FPOB, mỗi biểu thức chọn mờ trong một truy vấn là một sự kiện mờ có xác suất. Xác suất của các biểu thức chọn, được kết hợp với các đối tượng, có thể được đo bởi diễn dịch của nó và được định nghĩa như sau.

Định nghĩa 4.5 Giả sử $S = (C, \tau, \Rightarrow, me, \wp, f)$ là một lược đồ FPOB, $I = (\pi, v)$ là một thể hiện trên S , x là một biến đối tượng và $o \in \pi(C)$. *Diễn dịch xác suất* (probabilistic interpretation) của các biểu thức chọn mờ theo S , I và o , được biểu thị bởi $prob_{S,I,o}$, là một ánh xạ bộ phận từ tập tất cả các biểu thức chọn mờ đến tập tất cả các khoảng con đóng của khoảng $[0, 1]$ và được định nghĩa đệ qui như sau:

1. $prob_{S,I,o}(x \in c) = [\min(ext(c)(o)), \max(ext(c)(o))]$.
2. $prob_{S,I,o}(x.\lambda \theta v) = [\sum_{u \in V} \alpha(u).prob(u.\lambda' \theta v), \min(1, \sum_{u \in V} \beta(u).prob(u.\lambda' \theta v))]$, trong đó $\lambda = P.\lambda'$ và $v(o).P = \langle V, \alpha, \beta \rangle$.
3. $prob_{S,I,o}(x.\lambda_1 =_{\otimes} x.\lambda_2)$
 $= [\sum_{u \in V} \alpha(u).prob(u_1.\lambda_1' = u_2.\lambda_2'), \min(1, \sum_{u \in V} \beta(u).prob(u_1.\lambda_1' = u_2.\lambda_2'))]$,
 trong đó $\lambda_1 = P_1.\lambda_1'$, $v(o).P_1 = \langle V_1, \alpha_1, \beta_1 \rangle$, $\lambda_2 = P_2.\lambda_2'$, $v(o).P_2 = \langle V_2, \alpha_2, \beta_2 \rangle$,
 $[\alpha(u), \beta(u)] = [\alpha_1(u_1), \beta_1(u_1)] \otimes [\alpha_2(u_2), \beta_2(u_2)]$, $\forall u = (u_1, u_2) \in V = V_1 \times V_2$.
4. $prob_{S,I,o}(E_1 \otimes E_2) = prob_{S,I,o}(E_1) \otimes prob_{S,I,o}(E_2)$.
5. $prob_{S,I,o}(E_1 \oplus E_2) = prob_{S,I,o}(E_1) \oplus prob_{S,I,o}(E_2)$.

Về trực giác, $prob_{S,I,o}(x \in c)$ là khoảng xác suất để đối tượng o thuộc về lớp c (với lưu ý $ext(c)(o)$ là xác suất để o thuộc về lớp c như trong [5]); $prob_{S,I,o}(x.P.\lambda' \theta v)$ là khoảng xác suất để tính chất P của đối tượng o có giá trị u sao cho $u.\lambda' \theta v$ và $prob_{S,I,o}(x.P_1.\lambda_1' =_{\otimes} x.P_2.\lambda_2')$ là khoảng xác suất để tính chất P_1 và P_2 của o tương ứng có giá trị u_1 và u_2 sao cho $u_1.\lambda_1' = u_2.\lambda_2'$.

Trên cơ sở các khái niệm đã được giới thiệu, sự thỏa mãn các điều kiện chọn và phép chọn trong FPOB được định nghĩa như sau.

Định nghĩa 4.6 Giả sử $S = (C, \tau, \Rightarrow, me, \wp, f)$ là một lược đồ FPOB, $I = (\pi, v)$ là một thể hiện trên S và $o \in \pi(C)$. *Sự thỏa mãn* (satisfaction) các điều kiện chọn mờ của o theo diễn dịch xác suất $prob_{S,I,o}$ được định nghĩa như sau:

1. $prob_{S,I,o} \models (E)[L, U]$ nếu và chỉ nếu $prob_{S,I,o}(E) \subseteq [L, U]$.
2. $prob_{S,I,o} \models \neg\phi$ nếu và chỉ nếu $prob_{S,I,o} \not\models \phi$ không thỏa.
3. $prob_{S,I,o} \models \phi \wedge \psi$ nếu và chỉ nếu $prob_{S,I,o} \models \phi$ và $prob_{S,I,o} \models \psi$.
4. $prob_{S,I,o} \models \phi \vee \psi$ nếu và chỉ nếu $prob_{S,I,o} \models \phi$ hoặc $prob_{S,I,o} \models \psi$

Định nghĩa 4.7 Giả sử $S = (C, \tau, \Rightarrow, me, \wp, f)$ là một lược đồ FPOB, $I = (\pi, \nu)$ là một thể hiện trên S và ϕ là một điều kiện chọn mờ trên biến đối tượng x . *Phép chọn* trên I theo ϕ , được ký hiệu $\sigma_\phi(I)$, là một thể hiện $I' = (\pi', \nu')$ trên S , trong đó:

1. $\pi'(c) = \{o \in \pi(c) \mid prob_{S,I,o} \models \phi\}$ và
2. $\nu' = \nu \mid \pi'(C)$ (nghĩa là ánh xạ ν thu hẹp trên $\pi'(C)$).

Ví dụ 4.7 Giả sử $I = (\pi, \nu)$ là thể hiện trên lược đồ S trong Ví dụ 3.3.1. Truy vấn “chọn tất cả các gói có thể tích lớn hơn 20,000 với xác suất ít nhất là 0.7 và được vận chuyển trong thời gian trung bình với xác suất ít nhất là 0.3” có thể được thực hiện bởi phép chọn

$$I' = \sigma_\phi(I)$$

với $\phi = (x.volume(x.length, x.width, x.height) > 20,000)[0.7, 1.0] \wedge (x.time \rightarrow medium)[0.3, 1.0]$.

Vì $prob_{S,I,o_3}(x.volume(x.length, x.width, x.height) > 20,000) = [0.8, 1.0] \subseteq [0.7, 1.0]$ và $prob_{S,I,o_3}(x.time \rightarrow medium) = [0.36, 0.66] \subseteq [0.3, 1.0]$ nên đối tượng o_3 được chọn. Không còn đối tượng nào khác trong I thỏa điều kiện chọn đã cho.

5. KẾT LUẬN

Trong bài báo này, chúng tôi đã giới thiệu mô hình cơ sở đối tượng xác suất mờ FPOB là mở rộng mô hình POB trong [4] với tập mờ và phương thức lớp đối tượng. Mô hình này đã khắc phục được các hạn chế của mô hình POB và cho phép truy vấn thông tin không chắc chắn, không chính xác phổ biến trong thực tế. Ngoài phép chọn, chúng tôi cũng đã xây dựng tất cả các phép toán đại số khác trên FPOB và một hệ quản trị cho FPOB như trong [6].

Trong các bước tiếp theo, chúng tôi sẽ hoàn thiện hệ quản trị FPOB và phát triển một ngôn ngữ cho người sử dụng đầu cuối (end-user), tương tự như OQL trong mô hình CSDL hướng đối tượng truyền thống để biểu diễn, xử lý, và truy vấn thông tin không chắc chắn, không chính xác trên FPOB.

TÀI LIỆU THAM KHẢO

- [1] Baldwin, J.F., Lawry J.M., and Martin, T.P. *A mass assignment theory of the probability of fuzzy events*. Fuzzy Sets and Systems, 1996, 83, 353-367
- [2] Berzal F., Marín N., Pons O., Vila M.A. *A framework to build fuzzy object-oriented capabilities over an existing database system*. In Ma, Z. (Ed.): *Advances in Fuzzy Object-Oriented Database: Modeling and Applications*. Idea Group Publishing, 2005, 177-205.
- [3] Cao T.H., Rossiter J.M. *A deductive probabilistic and fuzzy object-oriented database language*. Fuzzy Sets and Systems. Elsevier Science, 2003, 140, 129 - 150.
- [4] Eiter T., Lu J.J., Lukasiewicz T., Subrahmanian V.S. *Probabilistic object bases*, ACM Transactions on database systems, 2001, 26 264-312.
- [5] Nguyen, H., and Cao, T.H. *Extending probabilistic object bases with uncertain applicability and imprecise values of class properties*. In *Proceedings of the 5th IEEE International Conference on Fuzzy Systems*, London, England, 2007, 487-492.
- [6] Nguyễn Hòa. *Một mô hình cơ sở đối tượng xác suất mờ*. Luận án tiến sĩ, Đại học Bách khoa-Đại học Quốc gia TP.HCM, 2008.

TÓM TẮT

Bài báo này giới thiệu một mô hình cơ sở đối tượng xác suất mờ (FPOB) là mở rộng của mô hình cơ sở đối tượng xác suất (POB) với hai đặc tính chính: (1) các giá trị không chắc chắn và không chính xác của một thuộc tính lớp (class attribute) được biểu diễn bởi một khoảng phân bố xác suất trên một tập các giá trị tập mờ; (2) các phương thức lớp (class method) với các đối số và giá trị không chắc chắn và không chính xác được tích hợp một cách hình thức vào mô hình mới. Một diễn dịch xác suất của các quan hệ trên các giá trị tập mờ và một đại số các bộ ba xác suất mờ được đề nghị để tính toán xác suất của các quan hệ tập mờ và các giá trị của các tính chất đối tượng. Sau đó, lược đồ và thể hiện của cơ sở đối tượng xác suất mờ được nghiên cứu và định nghĩa một cách hình thức để hỗ trợ truy vấn thông tin không chắc chắn, không chính xác trên FPOB.

SUMMARY

This paper introduces a fuzzy and probabilistic object base model (FPOB) that extends the probabilistic object base model (POB) with two key features: (1) uncertain and imprecise attribute values are represented as probability distributions on a set of fuzzy set values; (2) class methods with uncertain and imprecise input and output arguments are formally integrated into the new model. A probabilistic interpretation of relations on fuzzy set values and a fuzzy probabilistic triple algebra are proposed to compute probability degrees of fuzzy set relations and values of object properties. Then, fuzzy-probabilistic object base schemas and instances are formally researched and defined to support queries with the imprecise and uncertain information on FPOB.