TÍCH HỢP KHẢ NĂNG ÁP DỤNG KHÔNG CHẮC CHẮN TÍNH CHẤT LỚP VÀO MÔ HÌNH CƠ SỞ ĐỐI TƯỢNG XÁC SUẤT MỜ

TS. NGUYỄN HÒA *

1. GIỚI THIỆU

Mô hình cơ sở dữ liệu (CSDL) hướng đối tượng truyền thống đã được ứng dụng nhiều để thiết kế và hiện thực các hệ thống lớn. Tuy nhiên, mô hình dữ liệu này không biểu diễn và xử lý được thông tin không chắc chắn và không chính xác của các đối tượng trong thực tế ([4], [5], [7], [18]). Các nghiên cứu mở rộng mô hình CSDL hướng đối tượng truyền thống để xử lý tính không chắc chắn và không chính xác tập trung vào các vấn đề sau: (1) mô hình hóa các quan hệ lớp; (2) định nghĩa mức độ *thành viên* (membership) không chắc chắn của các đối tượng; (3) biểu diễn giá trị thuộc tính không chắc chắn hoặc không chính xác của các đối tượng; (4) biểu diễn và thực thi các *phương thức* (method) lớp; (5) mô hình hóa *khả năng áp dụng không chắc chắn* (uncertain applicability) của các tính chất lớp và xác định cơ chế thừa kế không chắc chắn của các đối tượng.

Đối với quan hệ lớp, có nhiều mô hình định nghĩa các mức độ bao hàm giữa các lớp ([11], [17], [18]). Tuy nhiên, như đã thảo luận trong [3], một tập các lớp với quan hệ bao hàm như vậy thực sự tạo thành một mạng hơn là một phân cấp. Giá trị thuộc tính không chắc chắn và không chính xác dẫn đến tính không chắc chắn của thành viên lớp của các đối tượng. Một số độ đo khác nhau để tính mức độ thành viên lớp đã được đề nghị trong ([5], [8], [19]). Tuy nhiên, các độ đo này chỉ xác định tính thành viên lớp dựa trên mức độ bao hàm mờ mà không phải mức độ xác suất.

Trong [4] và [5], các tác giả đã mô hình hóa tính không chắc chắn của một giá trị mà một thuộc tính có thể nhận. Tuy nhiên, mức độ không chắc chắn để thuộc tính có giá trị trong một tập các giá trị chưa được quan tâm. Trong khi các thuộc tính lớp được mô hình hóa, các phương thức lớp, như là các tính chất để mô hình hóa các hành vi đối tượng, thường bị bỏ qua. Trong [5], các phương thức đã được giới thiệu, nhưng biểu diễn và thao tác của chúng trong mô hình chưa được hình thức hóa. Trong [7] và [19], phương thức đã được định nghĩa một cách hình thức như những câu Horn và được thực thi như là một quá trình suy luận, tuy nhiên, các mô hình này chỉ dựa trên cơ sở logic mà không phải trên cơ sở thủ tục.

Thành viên lớp và khả năng áp dụng tính chất lớp không chắc chắn của các đối tượng dẫn đến thừa kế không chắc chắn các tính chất lớp của chúng. Điều này chưa được quan tâm trong [5], [8] và [19]. Trong [4], mức độ thành viên lớp được sử dụng như một ngưỡng để xác định các tính chất có thể được thừa kế trong lớp. Trong [7], mức độ thành viên lớp và khả năng áp dụng một tính chất lớp của đối tượng được kết hợp thành mức độ để đối tượng thừa kế tính chất đó.

Gần đây, trong [9], các tác giả đã giới thiệu một mô hình dựa trên xác suất để xử lý các đối tượng, được gọi là POB (Fuzzy Probabilistic Object Base). Tuy nhiên, các thiếu sót chính của POB là: (1) không cho phép giá trị thuộc tính không chính xác; (2) không quan tâm đến phương thức lớp; và (3) khả năng áp dụng của các tính chất lớp được giả sử là chắc chắn. Chẳng hạn, để biểu diễn giá trị thuộc tính sun, mô tả nhu cầu về ánh sáng của các thực vật, trong POB đã sử dụng các giá trị liệt kê *mild*, *medium* và *heavy* mà không có bất kỳ một diễn dịch nào. Trong khi đó, thực tế các giá trị như vậy có bản chất là mờ, không chính xác theo mức độ ánh sáng của mặt trời. Hơn

* *Giảng viên khoa Công Nghệ Thông Tin Trường Đại học Mở TP.HCM*

nữa, nếu không có một sự diễn dịch chúng không thể đo được và sự phân bố xác suất của chúng không thể tính toán được.

Trong [6], chúng tôi đã áp dụng lý thuyết tập mờ mở rộng POB để biểu diễn và tính toán với các giá trị không chính xác, được gọi là FPOB (Fuzzy Probabilistic Object Base). Chẳng hạn, các giá trị *mild*, *medium*, và *heavy* của thuộc tính sun ở trên có thể được định nghĩa và xử lý như các tập mờ. Trong [14], chúng tôi mở rộng hơn nữa mô hình FPOB với phương thức lớp để biểu diễn hành vi không chắc chắn của đối tượng. Trong bài báo này, chúng tôi tích hợp thêm khả năng áp dụng và thừa kế không chắc chắn các tính chất lớp vào FPOB, cho phép mô hình hóa hầu hết các khía cạnh không chắc chắn và không chính xác các hệ thống đối tượng trong thực tế.

Cơ sở toán học để phát triển FPOB được trình bày trong Phần 2. Tính chất lớp và lược đồ FPOB được giới thiệu trong Phần 3. Phần 4 trình bày về thể hiện và thừa kế không chắc chắn trong FPOB. Phần 5 trình bày mở rộng cú pháp và ngữ nghĩa của phép chọn trên mô hình mới. Cuối cùng, Phần 6 là kết luận và đề nghị các 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 ([9]). 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 $\$$ tương ứng biểu thị các phép toán hội, tuyển và trừ.

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)	$([L_1, U_1] \otimes_{in} [L_2, U_2]) \equiv [L_1 . L_2, U_1 . U_2]$ $([L_1, U_1] \oplus_{in} [L_2, U_2]) \equiv [L_1 + L_2 - (L_1 . L_2), U_1 + U_2 - (U_1 . U_2)]$ $([L_1, U_1] \ominus_{in} [L_2, U_2]) \equiv [L_1 . (1 - U_2), U_1 . (1 - L_2)]$
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 [2], 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 {=, ≤, <, ⊆, ∈}. *Diễn dịch xác suất* của quan hệ *A* θ *B*, ký hiệu *prob*(*A*θ *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* → *B*, ký hiệu *prob*(*A* → *B*), là một giá trị trong khoảng [0, 1] được định nghĩa bởi:

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

Chúng tôi lưu ý rằng, diễn dịch xác suất trên cũng có thể được định nghĩa cho các tập mờ trên miền liên tục bằng cách sử dụng phép tích phân thay cho phép lấy tổng như trong định nghĩa xác suất có điều kiện mờ trong [1]. Nghĩa là:

$$prob(A \to B) = \int_0^1 \int_0^1 \frac{\Pr({}^x A \cap {}^y B)}{\Pr({}^x A)} dxdy = \int_0^1 \int_0^1 \frac{\left|{}^x A \cap {}^y B\right|}{\left|{}^x A\right|} dxdy,$$

trong đó ${}^xA$ và ${}^yB$ là các lát cắt α của tập mờ *A* và *B* tương ứng với α = *x* và α = *y*.

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) ⟨***X***, α, β⟩ bao gồm một tập hữu hạn ***X***, một hàm phân bố xác suất α trên ***X*** và một hàm β: ***X*** → [0, 1] sao cho α(*x*) ≤ β(*x*). Nếu các phần tử của ***X*** là các tập mờ thì ⟨***X***, α, β⟩ được gọi là một *bộ ba xác suất mờ* (fuzzy probabilistic triple).

Định nghĩa 2.3.2 Cho ***U*** = {⟨*V*, α, β⟩| *V* ⊆ *U*} là một tập các bộ ba xác suất mờ khác rỗng trên tập *U* khác rỗng. Nếu ***A*** = (*U*, o_1, …, o_n) là một đại số với các phép toán o_1, …, o_n trên *U*, thì ***A*** = (***U***, $\boldsymbol{o}_1$, …, $\boldsymbol{o}_n$) là một đại số, được gọi là *đại số các bộ ba xác suất mờ* (fuzzy probabilistic triple algebra) trên ***U***, với các phép toán $\boldsymbol{o}_1$, …, $\boldsymbol{o}_n$ trên ***A*** được suy dẫn từ ***A*** như sau:

$\boldsymbol{o}_j(\langle V_1, \alpha_1, \beta_1\rangle, \langle V_2, \alpha_2, \beta_2\rangle, \ldots, \langle V_{m_j}, \alpha_{m_j}, \beta_{m_j}\rangle) = \langle V, \alpha, \beta\rangle$, trong đó $V = \{v = o_j(v_1, v_2, \ldots, v_{m_j}) \mid v_i \in V_i, i = 1, \ldots, m_j\}$, m_j là số đối số của o_j và

$[\alpha(v), \beta(v)] = \oplus_{me:\, v_1 \in V_1, v_2 \in V_2, \ldots, v_{m_j} \in V_{m_j}, v = o_j(v_1, v_2, \ldots, v_{m_j})} [\alpha_1(v_1), \beta_1(v_1)] \otimes_j [\alpha_2(v_2), \beta_2(v_2)] \otimes_j \ldots \otimes_j [\alpha_{m_j}(v_{m_j}), \beta_{m_j}(v_{m_j})]$, với mọi $v \in V$ và $\otimes_j$ là một chiến lược hội xác suất và ∀ *j* = 1,…, *n*.

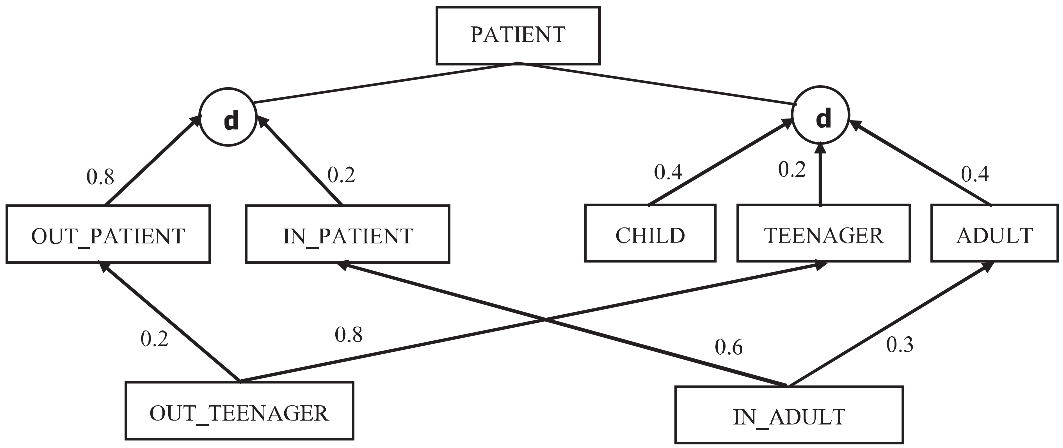
Ví dụ, nếu o_j là các phép toán + và × các số mờ trên tập các số thực theo *nguyên lý mở rộng* (extension principle), thì $\boldsymbol{o}_j$ tương ứng là các phép toán + và × các bộ ba xác suất mờ ⟨*V*, α, β⟩ và ta có một đại số các bộ ba xác suất mờ trên tập số thực.

3. TÍNH CHẤT CỦA CÁC ĐỐI TƯỢNG XÁC SUẤT MỜ

Trong FPOB, *tính chất* (nghĩa là thuộc tính và phương thức) của lớp không chỉ được mở rộng để có thể nhận giá trị tập mờ, mà còn có thể được mô hình hóa cho khả năng áp dụng không chắc chắn của các đối tượng.

3.1 Phân cấp lớp FPOB

Đối với FPOB, chúng tôi vẫn duy trì mô hình phân cấp lớp như trong POB. Hình 3.1 là một ví dụ về phân cấp lớp các *bệnh nhân* (patient) được phân loại như là *trẻ nhỏ* (child), *trẻ vị thành niên* (teenager) hoặc *người lớn* (*adult*) hay như là *bệnh nhân ngoại trú* (out_patient) hoặc *bệnh nhân nội trú* (*in_patient*).



Hình 3.1: Một ví dụ phân cấp lớp trong FPOB

Các lớp con của một lớp cùng liên kết với một nút **d** là loại trừ lẫn nhau. Trong ví dụ này lớp PATIENT có hai nhóm lớp con là {OUT_PATIENT, IN_PATIENT} và {CHILD, TEENAGER, ADULT}. 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 PATIENT có 80% khả năng thuộc về OUT_PATIENT trong khi chỉ có 20% khả năng còn lại thuộc về IN_PATIENT.

3.2 Thuộc tính và phương thức lớp FPOB

Như trong mô hình hướng đối tượng truyền thống, trong [6] mỗi lớp được đặc trưng bởi một số thuộc tính mà các giá trị của chúng có các kiểu tương ứng nào đó. Ở đây, chúng tôi mở rộng mô hình trong [6] hơn nữa với phương thức và sử dụng khái niệm *tính chất* (property) để thống nhất chỉ chung cho cả thuộc tính và phương thức. Mỗi một tính chất có kiểu và giá trị của nó. Đối với một phương thức, kiểu và giá trị của nó được xác định phụ thuộc vào kiểu và giá trị của các đối số của hàm định nghĩa phương thức đó. Hơn nữa, mỗi tính chất được kết hợp với một khoảng xác suất biểu diễn khả năng áp dụng không chắc chắn của nó đối với lớp trong đó nó được định nghĩa.

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* (atomic type). Các *kiểu* được định nghĩa một cách đệ qui như sau:

Mỗi kiểu cơ sở trong $\mathcal{T}$ là một kiểu.

1. Nếu τ là một kiểu, thì $\{\tau\}$ là một kiểu, được gọi là *kiểu tập mờ* (fuzzy set type) của τ.

Nếu $P_1, P_2, \ldots, P_k$ là các tính chất đôi một khác nhau 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, $[l_i, u_i]$ với mọi *i* từ 1 đến *k* là các khoảng con của khoảng [0, 1], thì $\tau = [P_1(\tau_{11}, \tau_{12}, \ldots, \tau_{1n_1}): \tau_1[l_1, u_1], P_2(\tau_{21}, \tau_{22}, \ldots, \tau_{2n_2}): \tau_2[l_2, u_2], \ldots, P_k(\tau_{k1}, \tau_{k2}, \ldots, \tau_{kn_k}): \tau_k[l_k, u_k]]$ là một kiểu được gọi là *kiểu bộ* trên $\{P_1, P_2, \ldots, P_k\}$. Chúng tôi gọi $P_1, P_2, \ldots, P_k$ là các *tính chất mức cao nhất* (top-level property) của τ. Các ký hiệu $\tau.P_i$ và $[\tau.P_i]$ được sử dụng để biểu thị lần lượt τ_i và $[l_i, u_i]$.

Trong định nghĩa trên, các τ_{ij} biểu diễn tập 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. Đối với mỗi P_i, nếu $l_i = u_i = 1$ thì $[l_i, u_i]$ có thể được bỏ qua.

Ví dụ 3.2.1 Trong CSDL các bệnh nhân ở trên, các tính chất có thể là name, age, address, check_date, medical_history, disease mô tả thông tin về tên, tuổi, địa chỉ, ngày khám, lịch sử bệnh và loại bệnh của mỗi bệnh nhân. Một số tính chất khác có thể là duration, check_again, cost_per_day, total_cost định nghĩa thời gian điều trị, ngày tái khám, chi phí điều trị mỗi ngày và tổng chi phí điều trị của mỗi bệnh nhân. Ở đây total_cost là một phương thức tính toán tổng chi phí điều trị mà bệnh nhân phải trả trong thời gian điều trị tại bệnh viện. Một số kiểu cơ sở có thể là **integer**, **real**, **string** và **datetype.** Các kiểu tập mờ và kiểu bộ là {**real**} và [name: **string**, age: {**real**}, address: **string**, check_date: **datetype**, medical_history: {**string**}[0.8, 1], disease: **string**, duration: {**real**}, cost_per_day: {**real**}, total_cost([duration: {**real**}], [cost_per_day: {**real**}]): {**real**}]. Ở đây, medical_history: {**string**}[0.8, 1] biểu diễn ít nhất 80% bệnh nhân có tiền sử bệnh.

Trong FPOB, giá trị tính chất của đối tượng không chỉ được mở rộng đối với các tập mờ như trong [6] mà mỗi giá trị còn được kết hợp với một khoảng xác suất biểu diễn mức độ áp dụng của các tính chất có thể nhận giá trị này.

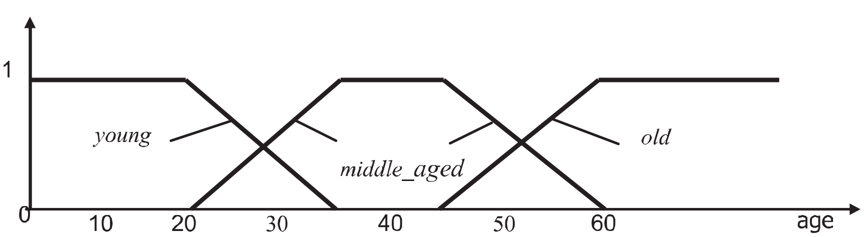
Định nghĩa 3.2.2 Giả sử mỗi kiểu cơ sở $\tau \in \mathcal{T}$ có một miền xác định *dom*(τ) kết hợp với nó. *Giá trị* (value) được định nghĩa một cách đệ qui như sau:

Với mỗi kiểu cơ sở $\tau \in \mathcal{T}$, thì mọi $v \in$ *dom*(τ) là một giá trị kiểu τ.

1. Với mỗi $\tau \in \mathcal{T}$, thì mọi tập tập mờ trên *dom*(τ) là một giá tri kiểu {τ}
2. Nếu $P_1, \ldots, P_k$ là các tính chất đôi một khác nhau trong **P** và với mọi *i* từ 1 đến *k*, v_i là giá trị của τ_i, $[l'_i, u'_i]$ và $[l_i, u_i]$ là các khoảng con của [0, 1], thì $[P_1: v_1[l'_1, u'_1], \ldots, P_k: v_k[l'_k, u'_k]]$ là một giá trị kiểu $\tau = [P_1: \tau_1[l_1, u_1], \ldots, P_k: \tau_k[l_k, u_k]]$.

Trong Định nghĩa 3.2.1, mỗi khoảng xác suất $[l_i, u_i]$ lượng hóa khả năng áp dụng của P_i đối với tất cả các đối tượng có kiểu τ nói chung. Nghĩa là, xác suất cho một đối tượng ngẫu nhiên kiểu τ có tính chất P_i là giữa l_i và u_i. Nhưng trong Định nghĩa 3.2.2 mỗi khoảng xác suất $[l'_i, u'_i]$ chỉ ra khả năng áp dụng của P_i đối với một đối tượng cụ thể kiểu τ. Vì vậy, $[l_i, u_i]$ chỉ là giá trị mặc định dựa trên cơ sở thống kê và không nhất thiết phải bằng $[l'_i, u'_i]$.

Ví dụ 3.2.2 Giả sử *young*, *middle_aged*, *old* là các nhãn ngôn ngữ của các số mờ trên *dom*(**real**) như Hình 3.2.2, thì [name: *Ho V.T*, age: *middle_aged*, address: *Cantho*, check_date: 21-3-08, medical_history: {*cholecystitis*}[0.7, 1], disease: *hepatitis*] là một giá trị kiểu [name: **string**, age: {**real**}, address: **string**, check_date: **datetype**, medical_history: {**string**}[0.8, 1], disease: **string**].



Hình 3.2.2: Các giá trị tập mờ của thuộc tính age

Định nghĩa giá trị không chắc chắn và không chính xác trong [6] bây giờ được mở rộng cho giá trị cũng như các đối số của phương thức.

Định nghĩa 3.2.3 Giả sử $P_1, \ldots, P_k$ là các tính chất đôi một phân biệt trong $\mathcal{P}$, V_i và V_{ij} tương ứng 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 trên V_i và V_{ij}, với mọi *i* từ 1 đến *k* và *j* từ 1 đến n_i, $[l'_i, u'_i] \subseteq [0, 1]$ với $1 \le i \le k$. Thì *fptv* = $[P_1(\langle V_{11}, \alpha_{11}, \beta_{11}\rangle, \ldots, \langle V_{1n_1}, \alpha_{1n_1}, \beta_{1n_1}\rangle): \langle V_1, \alpha_1, \beta_1\rangle[l'_1, u'_1], \ldots, P_k(\langle V_{k1}, \alpha_{k1}, \beta_{k1}\rangle, \ldots, \langle V_{kn_k}, \alpha_{kn_k}, \beta_{kn_k}\rangle): \langle V_k, \alpha_k, \beta_k\rangle[l'_k, u'_k]]$ là một *giá trị bộ xác suất mờ* (fuzzy-probabilistic tuple value) kiểu $[P_1(\tau_{11}, \ldots, \tau_{1n_1}): \tau_1[l_1, u_1], \ldots, P_k(\tau_{k1}, \ldots,$

τ_{kn_k}): $\tau_k[l_k, u_k]$] trên $\{P_1,\ldots,P_k\}$. Các ký hiệu *fptv*.P_i và [*fptv*.P_i] tương ứng được sử dụng để biểu thị $\langle V_i, \alpha_i, \beta_i\rangle$ và $[l'_i, u'_i]$.

Ở đây, mỗi P_i, $[l_i, u_i]$ hoặc $[l'_i, u'_i]$ biểu diễn tính không chắc chắn của P_i, trong khi $[\alpha_i, \beta_i]$ biểu diễn tính không chắc chắn về giá trị của P_i trên V_i khi P_i được áp dụng.

Ví dụ 3.2.3 Giả sử, một bệnh nhân có tên là Hồ V.T, tuổi là *trung niên* (middle_aged) và địa chỉ là Cần Thơ được khám bệnh và bác sĩ chưa biết chắc chắn bệnh nhân bị bệnh gì. Tuy nhiên, qua các triệu chứng bệnh của bệnh nhân và bằng kinh nghiệm thực tế, ông có thể đưa ra phán đoán là 50% khả năng bệnh nhân này bị bệnh *viêm gan* (hepatitis) hoặc *xơ gan* (cirrhosis). Hơn nữa, với một số thông tin về sức khỏe do bệnh nhân cung cấp, bác sĩ cho rằng ít nhất 70% khả năng bệnh nhân này từng có tiền sử bệnh, và bệnh đó có thể là viêm *túi mật* (cholecystitis) hoặc *sỏi mật* (gall-stone). Ngoài ra, nếu chi phí điều trị mỗi ngày là khoảng 60 nghìn đồng và thời gian điều trị được ước đoán là 30 hoặc 32 ngày với xác suất 40% -60%, thì thông tin này có thể được biểu diễn bởi một giá trị bộ xác suất mờ [name: ⟨{*Ho V.T*}, *u*, *u*⟩, age: ⟨{*middle_aged*}, *u*, *u*⟩, address: ⟨{*Cantho*}, *u*, *u*⟩, medical_history: ⟨{{*cholecystitis*}, {*gall-stone*}}, *u*, *u*⟩[0.7, 1], disease: ⟨{*hepatitis*, *cirrhosis*}, *u*, *u*⟩, duration: ⟨{30, 32}, 0.8*u*, 1.2*u*⟩, cost_per_day: ⟨{*about_60*}, *u*, *u*⟩], trong đó *about_60* là một giá trị tập mờ.

3.3 Lược đồ FPOB

Lược đồ FPOB được mở rộng từ lược đồ POB với phương thức lớp và khả năng áp dụng không chắc chắn các tính chất lớp của các đối tượng.

Định nghĩa 3.3.1 Một *lược đồ FPOB* là một bộ sáu **S** = ($\mathcal{C}$, τ, $\Rightarrow$, *me*, $\wp$, f), trong đó:

1. $\mathcal{C}$ 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 $c \in \mathcal{C}$ với một kiểu bộ $\tau(c)$ biểu diễn các tính chất, kiểu và khả năng áp dụng của chúng đối với các đối tượng trong lớp đó.

3. $\Rightarrow$ là một quan hệ hai ngôi trên $\mathcal{C}$ sao cho ($\mathcal{C}$, $\Rightarrow$) là một đồ thị có hướng không có chu trình, mỗi một đỉnh của ($\mathcal{C}$, $\Rightarrow$) là một lớp trong $\mathcal{C}$, mỗi cạnh $c1 \Rightarrow c2$ biểu diễn $c1$ là một lớp con trực tiếp của $c2$.

4. *me* là một ánh xạ đặt tương ứng mỗi lớp $c \in \mathcal{C}$ 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 $d \Rightarrow c$ trong ($\mathcal{C}$, $\Rightarrow$) với một số hữu tỉ $\wp(d, c)$ trong khoảng [0, 1] như là xác suất có điều kiện để một đối tượng được chọn một cách ngẫu nhiên của lớp *c* là thuộc về lớp *d* sao cho tổng $\Sigma_{d\in\mathcal{P}}\wp(d, c) \le 1$, $\forall\, c \in \mathcal{C}$, $\forall \mathcal{P} \in me(c)$.

f là một ánh xạ đặt tương ứng mỗi tính chất $P_i(\tau_{i1}, \tau_{i2}, \ldots, \tau_{in_i})$: τ_i 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 $f(P_i)$ định nghĩa P_i như là một phương thức xác suất mờ, có tập đối số và giá trị trả về là các bộ ba xác suất mờ.

Một *đường đi có hướng* (directed path) trong ($\mathcal{C}$, $\Rightarrow$) là một dãy các lớp $c_1, c_2, \ldots, c_k$ sao cho $c_1 \Rightarrow c_2 \Rightarrow \ldots \Rightarrow c_k$ với $k \ge 1$. Ký hiệu $c \Rightarrow^* d$ được sử dụng để biểu thị cho sự tồn tại một đường đi từ *c* đến *d*.

Ví dụ 3.3.1 Một lược đồ FPOB cho cơ sở đối tượng các bệnh nhân đã nêu trên có thể được định nghĩa như sau:

$\mathcal{C}$ = {PAT IENT, OUT_PATIENT, IN_PATIENT, CHILD, TEENAGER, ADULT, OUT_TEENAGER, IN_ADULT}.

τ được cho trong Bảng 3.3.1, với lưu ý, các lớp con có thể thừa kế tính chất từ các lớp cha của chúng.

Bảng 3.3.1: Ánh xạ gán kiểu τ

c	$\tau(c)$
PATIENT	[name: **string**, age: **{real}**, address: **string**, check_date: **datetype**, disease: **string,** duration:**{integer}**, cost_per_day: **{real}**, total_cost: **{real}**]
OUT_PATIENT	[check_again: **datetype**[0.95, 1]]
IN_PATIENT	[bed_no: **string]**
CHILD	[medical_history: **{string}**[0.4, 0.6]]
TEENAGER	[medical_history: **{string}**[0.6, 0.8]]
ADULT	[medical_history: **{string}**[0.8, 1]]
OUT_TEENAGER	[]
IN_ADULT	[]

$(\mathcal{C}, \Rightarrow)$, me, $\wp$ được cho như trong Hình 3.1, và f định nghĩa phương thức total_cost sử dụng đại số các bộ ba xác suất mờ trong Định nghĩa 2.3.2 như sau:

PATIENT: total_cost(duration: $\langle V_1, \alpha_1, \beta_1\rangle$, cost_per_day: $\langle V_2, \alpha_2, \beta_2\rangle$): $\langle V, \alpha, \beta\rangle$

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** $\langle V, \alpha, \beta\rangle$

4. THỰC HIỆN VÀ THỪA KẾ KHÔNG CHẮC CHẮN TRONG FPOB

4.1 Các thể hiện FPOB

Cho một lược đồ FPOB, một thể hiện FPOB trên lược đồ này là một cơ sở đối tượng được kết hợp với các giá trị bộ xác suất mờ tương ứng với kiểu của mỗi lớp của chúng như định nghĩa sau.

Định nghĩa 4.1.1 Giả sử ***S*** = $(\mathcal{C}, \tau, \Rightarrow, me, \wp, f)$ là một lược đồ FPOB và **O** là một tập các *định danh đối tượng* (object identifier - oid), *một thể hiện FPOB* (FPOB-instance) trên ***S*** là một cặp (π, v), trong đó:

1. π: $\mathcal{C} \rightarrow 2^{\mathbf{O}}$ là ánh xạ đặt tương ứng mỗi lớp *c* thuộc $\mathcal{C}$ với một tập con hữu hạn của tập **O**, sao cho $\pi(c_1) \cap \pi(c_2) = \emptyset$, $\forall$ $c_1 \neq c_2 \in \mathcal{C}$.

2. v là ánh xạ đặt tương ứng mỗi $o \in \pi(c)$ với một giá trị bộ xác suất mờ $[P_1: \langle V_1, \alpha_1, \beta_1\rangle[l_1, u_1], \ldots, P_k: \langle V_k, \alpha_k, \beta_k\rangle[l_k, u_k]]$ kiểu τ(*c*), với mọi $c \in \pi(\mathcal{C})$.

Ví dụ 4.1.1 Một thể hiện của lược đồ FPOB trong Ví dụ 3.3.1 được chỉ ra trong các Bảng 4.1.1 và 4.1.2. Ở đây, *about_60* = {59: 0.5, 60: 1, 61: 0.5} là một số mờ thể hiện chi phí điều trị xấp xỉ mỗi ngày của một bệnh nhân, được biểu diễn bởi o_4, *about_1800* = 30 × *about_60* = {1770: 0.5, 1800: 1, 1830: 0.5} và *about_2400* = 40 × *about_60* = {2360: 0.5, 2400: 1, 2440: 0.5}là các số mờ biểu diễn tổng chi phí điều trị của bệnh nhân này. Trong khi đó medical_history: ⟨{{*cholecystitis*}}, *u*, *u*⟩[0.7, 1] biểu diễn khả năng ít nhất 70% o_4 có tiền sử bệnh viêm túi mật. Đơn vị tiền chi phí điều trị mỗi ngày và tổng chi phí phải trả của các bệnh nhân trong Bảng 4.1.2 là một nghìn đồng Việt Nam.

Bảng 4.1.1: Ánh xạ đối tượng π và π^*

c	$\pi(c)$	$\pi^*(c)$
PATIENT	$\{o_1\}$	$\{o_1, o_2, o_3, o_4\}$
OUT_PATIENT	{}	$\{o_2, o_3\}$
IN_PATIENT	{}	$\{o_4\}$
CHILD	{}	{}
TEENAGER	{}	$\{o_2, o_3\}$
ADULT	{}	$\{o_4\}$
OUT_TEENAGER	$\{o_2, o_3\}$	$\{o_2, o_3\}$
IN_ADULT	$\{o_4\}$	$\{o_4\}$

Bảng 4.1.2: Ánh xạ gán giá trị ν

oid	ν(*oid*)
o_1	[name: ⟨{*Le H*}, *u*, *u*⟩, age: ⟨{45}, *u*, *u*⟩, address: ⟨{*Saigon*}, *u*, *u*⟩, check_date: ⟨{20-3-08}, *u*, *u*⟩, medical_history: ⟨{{*bronchitis*}}, *u*, *u*⟩[0.32, 0.4], disease: ⟨{*lung cancer*, *tuberculosis*}, 0.8*u*, 1.2*u*⟩, duration: ⟨{400, 500}, *u*, *u*⟩, cost_per_day: ⟨{300}, *u*, *u*⟩, total_cost: ⟨{120,000, 150,000}, *u*, *u*⟩]
o_2	[name: ⟨{*Tran T.L*}, *u*, *u*⟩, age: ⟨{16}, *u*, *u*⟩, address: ⟨{*Binhduong*}, *u*, *u*⟩, check_date: ⟨{20-3-08}, *u*, *u*⟩, disease: ⟨{*flu*}, *u*, *u*⟩, duration: ⟨{7}, *u*, *u*⟩, cost_per_day: ⟨{30}, *u*, *u*⟩, total_cost: ⟨{210}, *u*, *u*⟩]
o_3	[name: ⟨{*Nguyen H.L*}, *u*, *u*⟩, age: ⟨{*young*}, *u*, *u*⟩, address: ⟨{*Hanoi*}, *u*, *u*⟩, check_date: ⟨{21-3-08}, *u*, *u*⟩, disease: ⟨{*angina*}, *u*, *u*⟩, duration: ⟨{10}, *u*, *u*⟩, check_again: ⟨{27-3-08}, *u*, *u*⟩[0.9, 1], cost_per_day: ⟨{160, 170}, 0.8*u*, *u*⟩, total_cost: ⟨{1600, 1700}, 0.8*u*, *u*⟩]
o_4	[name: ⟨{*Ho V.T*}, *u*, *u*⟩, age: ⟨{*middle_aged*}, *u*, *u*⟩, address: ⟨{*Cantho*}, *u*, *u*⟩, check_date: ⟨{21-3-08}, *u*, *u*⟩, medical_history: ⟨{{*cholecystitis*}}, *u*, *u*⟩[0.7, 1], disease: ⟨{*hepatitis*, *cirrhosis*}, *u*, *u*⟩, duration: ⟨{30, 40}, *u*, *u*⟩, bed_no: {*A35*}, *u*, *u*⟩, cost_per_day: ⟨{*about_60*}, *u*, *u*⟩, total_cost: ⟨{*about_1800*, *about_2400*}, *u*, *u*⟩]

4.2 Phạm vi xác suất của lớp FPOB

Trong cơ sở đối tượng truyền thống, *phạm vi* (extent) của một lớp bao gồm tất cả các đối tượng thuộc về lớp đó. Trong cơ sở đối tượng mờ hoặc xác suất, do tính thành viên và khả năng áp dụng tính chất lớp là không chắc chắn ([6], [7], [9], [14]) nên phạm vi của một lớp là không chắc chắn. Đối với FPOB, phạm vi của một lớp chỉ xác suất để mỗi đối tượng thuộc lớp đó và được định nghĩa như sau.

Định nghĩa 4.2.1 Giả sử ***I*** = (π, ν) là một thể hiện trên một lược đồ FPOB ***S*** = (C, τ, $\Rightarrow$, *me*, $\wp$, f). Với mỗi lớp $c \in C$, *phạm vi xác suất* (probabilistic extent) của c, được ký hiệu *ext*(c), là một ánh xạ đặt tương ứng mỗi $o \in \pi(C)$ với một tập các số hữu tỉ trong khoảng [0, 1] sao cho:

1. Nếu $o \in \pi^*(c)$, thì *ext*(c)(o) = {1}.
2. Nếu $o \in \pi^*(d)$ và ε (d) ∩ ε (c) = ∅ với mọi mô hình ε của ***S*** thì *ext*(c)(o) = {0}.

3. Ngược lại, $ext(c)(o) = \{p \mid p$ là tích của các xác suất các cạnh trên một đường đi từ c đến một lớp $d \in \mathbf{C}$, ở đó d là lớp nhỏ nhất sao cho $o \in \pi^*(d)$ và $c \Rightarrow^* d\}$.

Như vậy, khoảng xác suất để đối tượng o là thành viên của lớp c là $[min(ext(c)(o)), max(ext(c)(o))]$.

Ví dụ 4.2.1 Giả sử ***I*** là thể hiện của lược đồ FPOB trong Ví dụ 4.1.1, phạm vi xác suất của lớp IN_ADULT được xác định như sau:

ext(IN_ADULT)(o_1)={0.12} *ext*(IN_ADULT)(o_2)={0.0}.

ext(IN_ADULT)(o_3)={0.0} *ext*(IN_ADULT)(o_4)={1.0}.

4.3 Thừa kế không chắc chắn của các đối tượng FPOB

Trong [9] và [6], những mô hình cơ sở đối tượng xác suất hoặc xác suất mờ, đã sử dụng cùng một chiến lược thừa kế cổ điển để xác định các thuộc tính được thừa kế cho một lớp đối tượng. Cơ chế này chỉ xét quan hệ thừa kế của một lớp đối với lớp cha của nó. Theo đó, mỗi lớp đối tượng chắc chắn thừa kế mọi tính chất của *lớp cha nhỏ nhất* (minimal super-class) của nó. Tuy nhiên, trong các mô hình này, mỗi đối tượng không chắc chắn là thành viên của một lớp. Do đó, một cơ chế thừa kế như vậy là còn hạn chế.

Đối với FPOB, không chỉ thành viên lớp không chắc chắn mà khả năng áp dụng tính chất lớp cũng không chắc chắn. Hệ quả là thừa kế và đa thừa kế trong FPOB cũng không chắc chắn. Trong FPOB, chúng tôi áp dụng cơ chế thừa kế trong [3] nhưng với các chiến lược hội xác suất thay vì lấy tích của mức độ thành viên lớp của đối tượng và mức độ khả năng áp dụng tính chất lớp như trong mô hình này.

Định nghĩa 4.3.1 Giả sử $[l, u]$ là khoảng xác suất biểu diễn khả năng áp dụng của tính chất P đối với lớp c và $[x, y]$ là khoảng xác suất biểu diễn mức độ thành viên của đối tượng o đối với c, thì khoảng xác suất biểu diễn *khả năng áp dụng* của P đối với o được định nghĩa là $[l, u] \otimes [x, y]$, trong đó $\otimes$ là một chiến lược hội xác suất.

Như vậy, với một lớp c có mức độ áp dụng tính chất P là $[l, u]$, thì *mức độ thừa kế* (inheritance degree) P của đối tượng o là $[l, u] \otimes [x, y]$ nếu $[x, y]$ là khoảng xác suất để o là thành viên của c. Chúng tôi gọi cách tính toán mức độ thừa kế như vậy là *một cơ chế thừa kế không chắc chắn* (uncertain inheritance mechanism) các tính chất lớp của các đối tượng trong FPOB.

Đối với các lược đồ FPOB, một mở rộng của chiến lược thừa kế các lược đồ POB được áp dụng để tính toán tính chất của các lớp cha được thừa kế bởi các lớp con của chúng. Theo đó, một lược đồ FPOB $\boldsymbol{S} = (\mathcal{C}, \tau, \Rightarrow, me, \wp, f)$ khi áp dụng cơ chế thừa kế này sẽ dẫn đến một lược đồ FPOB mới $\boldsymbol{S}^* = (\mathcal{C}, \tau^*, \Rightarrow, \text{me}, \wp, f^*)$ chỉ khác $\boldsymbol{S}$ phép gán kiểu τ^* và diễn dịch hàm f^*. Cụ thể hơn, với mỗi $c \in \mathcal{C}$, $\tau^*(c) = [P_1: \tau(d_1).P_1[l_1', u_1'], \ldots, P_k: \tau(d_k).P_k[l_k', u_k']]$ và $f^*(P_i) = f_{d_i}(P_i)$, trong đó P_i với khả năng áp dụng $[l'_i, u'_i]$, $\forall i = 1, \ldots, k$, tương ứng là tính chất mức cao nhất của các lớp $d_1, \ldots, d_k$ được thừa kế bởi c, ký hiệu f_d là diễn dịch hàm thu hẹp của f trên các tính chất của lớp d. Một lược đồ $\boldsymbol{S}$ là được thừa kế đầy đủ nếu và chỉ nếu $\boldsymbol{S} = \boldsymbol{S}^*$.

Ví dụ 4.3.1 Giả sử ***S*** là lược đồ cơ sở các bệnh nhân trong Ví dụ 3.3.1 với kiểu của mỗi lớp trong Bảng 3.3.1, thì kết quả của áp dụng chiến lược thừa kế trên ***S*** theo chiến lược hội độc lập $\otimes_{in}$ là một lược đồ mới ***S**** với một thể hiện, trong đó các đối tượng có thể thừa kế không chắc chắn, được cho trong Bảng 4.1.2. Chẳng hạn, bệnh nhân, được biểu diễn bởi o_4, thừa kế thuộc tính medical_history với khoảng xác suất [0.8, 1] và ít nhất 70% khả năng áp dụng tính chất này. Trong khi đó, các đối tượng o_2 và o_3 thừa kế medical_history với khoảng xác suất [0.6, 0.8] nhưng không áp dụng tính chất này. Hơn nữa, trong mô hình FPOB, một đối tượng của lớp cha có thể thừa kế không chắc chắn một tính chất của lớp con. Chẳng hạn, đối tượng o_1 thừa kế thuộc tính medical_history của lớp ADULT với xác suất [0.32, 0.4] và khả năng áp dụng tính chất này cũng là [0.32, 0.4].

Khả năng áp dụng cùng với khả năng nhận giá trị không chắc chắn của một tính chất của đối tượng là khả năng nhận giá trị cụ thể của tính chất đó. Vì vậy, chúng tôi kết hợp khả năng áp dụng và khả năng nhận giá trị không chắc chắn của một tính chất đối với một đối tượng thành khả năng nhận giá trị không chắc chắn hợp nhất của tính chất cho đối tượng như định nghĩa dưới đây.

Định nghĩa 4.3.2 Giả sử $\langle V, \alpha, \beta\rangle$ là bộ ba xác suất mờ biểu diễn giá trị không chắc chắn của tính chất P của đối tượng o và $[l, u]$ là khoảng xác suất biểu diễn khả năng áp dụng của P đối với o. *Bộ ba xác suất mờ được suy dẫn* (derived fuzzy-probabilistic triple) kết hợp khả năng áp dụng và khả năng nhận giá trị của P đối với o, ký hiệu $\langle\langle V, \alpha, \beta\rangle[l, u]\rangle$, là $\langle V, \alpha', \beta'\rangle$ trong đó $[\alpha'(v), \beta'(v)] = [\alpha(v), \beta(v)] \otimes [l, u]$ với mọi $v \in V$.

5. PHÉP CHỌN TRÊN FPOB

5.1 Cú pháp của điều kiện chọn

Cũng như trong cơ sở đối tượng truyền thống, phép chọn là một phép toán đại số cơ bản trên FPOB. Nói một cách không hình thức, kết quả của một truy vấn chọn trên một thể hiện ***I*** của một lược đồ ***S*** là một thể hiện ***I'*** trên ***S*** sao cho các đối tượng của các lớp trong ***I'*** có các giá trị thuộc tính và phương thức thỏa mãn điều kiện chọn của truy vấn này. Trước khi định nghĩa hình thức các *điều kiện chọn* (selection condition) và *phép chọn* (selection), chúng tôi giới thiệu cú pháp của các biểu thức đường đi như một mở rộng của các biểu thức đường đi trong [6]. Đó là một dãy các tính chất được kết hợp với một khoảng xác suất biểu diễn khả năng áp dụng của chúng.

Định nghĩa 5.1.1 Giả sử $\tau = [P_1: \tau_1[l_1, u_1], \ldots, P_k: \tau_k[l_k, u_k]]$ là một kiểu bộ. *Biểu thức đường đi* (path expression) được định nghĩa một cách đệ qui cho mọi i từ 1 đến k như sau:

1. P_i là một biểu thức đường đi cho τ, và $[l_i, u_i]$ là khoảng xác suất được kết hợp với biểu thức đường đi này.

2. Nếu λ_i là một biểu thức đường đi cho τ_i, $[l_{\lambda_i}, u_{\lambda_i}]$ là khoảng xác suất kết hợp với nó, thì $P_i.\lambda_i$ là một biểu thức đường đi τ, và $[l_i, u_i] \otimes [l_{\lambda_i}, u_{\lambda_i}]$ là khoảng xác suất được kết hợp với biểu thức đường đi này.

Chúng tôi ký hiệu $[\lambda]$ để biểu thị khoảng xác suất được kết hợp với biểu thức đường đi λ.

Ví dụ 5.1.1 Giả sử τ = [name: **string**, age: {**real**}, address: [street: **string**, postcode: **string**[0, 0.1]], medical_history: {**string**}[0.4, 0.6]], thì medical_history và address.postcode là các biểu thức đường đi cho τ với các khoảng xác suất kết hợp tương ứng theo chiến lược hội độc lập là [0.4, 0.6] và [0, 0.1].

Bây giờ các biểu thức chọn trong FPOB, như trong [14], được mở rộng từ [6] để không chỉ cho phép truy vấn giá trị không chắc chắn, không chính xác mà cả khả năng áp dụng không chắc chắn của các tính chất của các đối tượng.

Định nghĩa 5.1.2 Giả sử ***S*** = $(\mathcal{C}, \tau, \Rightarrow, \text{me}, \text{p}, \text{f})$ là một lược đồ FPOB_E và **X** là một tập các biến đối tượng. Các *biểu thức chọn* được định nghĩa một cách đệ qui và có một trong các dạng sau:

1. $x \in c$, trong đó $x \in \mathbf{X}$ và $c \in \mathbf{C}$.

2. $x.\lambda$, trong đó $x \in \mathbf{X}$, và λ là một biểu thức đường đi.

3. $x.\lambda\ \theta\ v$, trong đó $x \in \mathbf{X}$, λ là một biểu thức đường đi, θ là một quan hệ hai ngôi trong $\{=, \neq, \leq, <, \subseteq, \in, \rightarrow\}$, và v là một giá trị.

4. $x.\lambda_1 =_\otimes x.\lambda_2$, trong đó $x \in \mathbf{X}$, λ_1 và λ_2 là các 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 $v_1 = v_2$

5. $E_1 \otimes E_2$, 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.

6. $E_1 \oplus E_2$, 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.

Ví dụ 5.1.2 Giả sử ***S*** là lược đồ FPOB của thể hiện trong Ví dụ 4.1.1, thì chọn "tất cả các bệnh nhân trẻ có tiền sử bệnh" có thể được biểu diễn bởi *x*.age → *young* ⊗ *x*.medical_history.

Các điều kiện chọn bây giờ được định nghĩa như trong [14]. Đó là các biểu thức chọn phải được thỏa mãn một xác suất trong một khoảng nào đó.

Định nghĩa 5.1.3 Giả sử ***S*** = ($\mathcal{C}$, τ, $\Rightarrow$, *me*, $\wp$, f) là một lược đồ FPOB, các *điều kiện chọn* (selection condition) được định nghĩa một cách đệ qui như sau:

1. Nếu *E* là một biểu thức chọn , *L* và *U* là các số thực trong khoảng [0, 1], $L \le U$ thì (E)[L, U] là một điều kiện chọn.

2. Nếu φ và ψ là các điều kiện chọn trên cùng một biến đối tượng thì $\neg\varphi$, $(\varphi \wedge \psi)$, $(\varphi \vee \psi)$ là các điều kiện chọn.

Ví dụ 5.1.3 Chọn "tất cả các bệnh nhân trẻ với một xác suất ít nhất là 0.4 và có bệnh ung thư phổi với một xác suất ít nhất là 0.8" có thể được thực hiện bởi điều kiện chọn (*x*.age → *young*) [0.4, 1] ∧ (*x*.disease = *lung cancer*) [0.8, 1].

5.2 Ngữ nghĩa của điều kiện chọn

Để định nghĩa ngữ nghĩa cho các điều kiện chọn, trước tiên, chúng tôi giới thiệu ngữ nghĩa của các biểu thức đường đi, biểu thức chọn như là mở rộng các định nghĩa tương ứng của các khái niệm này trong [9] và [6] với khả năng áp dụng không chắc chắn của các tính chất.

Định nghĩa 5.2.1 Giả sử $\tau = [P_1: \tau_1[l_1, u_1], \ldots, P_k: \tau_k[l_k, u_k]]$ là một kiểu bộ và $v = [P_1: v_1[l'_1, u'_1], \ldots, P_k: v_k[l'_k, u'_k]]$ là một giá trị kiểu τ. *Diễn dịch* (interpretation) của một biểu thức đường đi λ đối với τ theo *v* là giá trị và khoảng xác suất kết hợp với đường đi này, tương ứng được biểu thị bởi $v.\lambda$ và $[v.\lambda]$, được định nghĩa một cách đệ qui như sau:

1. Nếu $\lambda = P_i$, thì $v.\lambda = v_i$ và $[v.\lambda] = [l'_i, u'_i]$.

2. Nếu $\lambda = P_i.\lambda_i$, trong đó λ_i là một biểu thức đường đi đối với τ_i và $[v_i.\lambda_i] = [l_{\lambda_i}, u_{\lambda_i}]$, thì $v.\lambda = v_i.\lambda_i$ và $[v.\lambda] = [l'_i, u'_i] \otimes [l_{\lambda_i}, u_{\lambda_i}]$.

Ví dụ 5.2.1 Giả sử τ = [name: **string**, age: **{real}**, address: [street: **string**, postcode: **string**[0, 0.1]], medical_history: {**string**}[0.4, 0.6]], thì diễn dịch của biểu thức đường đi address.postcode theo giá trị *v* = [name: *Nguyen L.A*, age: *young*, address: [street: 135-*XVNT-TP.HCM*, postcode: *2468A*[0, 0.5]], medical_history: {*tonsillitis*}[0.4, 0.6]] là giá trị *2468A* và khoảng xác suất [0, 0.5].

Định nghĩa 5.2.2 Giả sử ***S*** = ($\mathcal{C}$, τ, $\Rightarrow$, *me*, $\wp$, f) là một lược đồ FPOB, ***I*** = (π, ν) là một thể hiện trên ***S***, *x* là một biến đối tượng và $o \in \pi(\mathbf{C})$. *Diễn dịch xác suất* của các biểu thức chọn 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 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. $prob_{s,I,o}(x.\lambda) = [\nu(o).P] \otimes (\oplus_{me:\, w \in V}([\alpha(w), \beta(w)] \otimes [w.\lambda']))$, trong đó $\lambda = P.\lambda'$, $\nu(o).P = \langle V, \alpha, \beta\rangle$ và $[\nu(o).P]$ là khoảng xác suất biểu diễn khả năng áp dụng của *P* đối với *o*; tuyển loại trừ $\oplus_{me:\, w \in V}([\alpha(w), \beta(w)] \otimes [w.\lambda'])$ của $[\alpha(w), \beta(w)] \otimes [w.\lambda']$ là khoảng xác suất biểu diễn khả năng áp dụng của biểu thức đường đi λ' theo tập các giá trị loại trừ lẫn nhau *V* của tính chất *P* của đối tượng *o*.

3. $prob_{s,I,o}(x.\lambda\ \theta\ v) = [\Sigma_{w \in V}\alpha'(w).prob(w.\lambda'\ \theta\ v), min(1, \Sigma_{w \in V}\beta'(w).prob(w.\lambda'\ \theta\ v))]$, trong đó $\lambda = P.\lambda'$, $[\nu(o).P] \otimes [w.\lambda'] = [l, u]$, $\nu(o).P = \langle V, \alpha, \beta\rangle$ và $[\alpha'(w), \beta'(w)] = [\alpha(w), \beta(w)] \otimes [l, u]$ với mọi $w \in V$.

4. $prob_{s,I,o}(x.\lambda_1 =_{\otimes} x.\lambda_2)$

$= [\Sigma_{w \in V}\alpha'(w).prob(w_1.\lambda'_1 = w_2.\lambda'_2), min(1, \Sigma_{w \in V}\beta'(w).prob(w_1.\lambda'_1 = w_2.\lambda'_2))]$, trong đó $\lambda_1 = P_1.\lambda'_1$,

$[v(o).P_1]\otimes[w_1.\lambda'_1] = [l_1, u_1]$, $v(o).P_1 = \langle V_1, \alpha_1, \beta_1\rangle$, $\lambda_2 = P_2.\lambda'_2$, $[v(o).P_2] \otimes [w_2.\lambda'_2] = [l_2, u_2]$, $v(o).P_2 = \langle V_2, \alpha_2, \beta_2\rangle$, và $[\alpha'(w), \beta'(w)] = [\alpha_1(w_1), \beta_1(w_1)] \otimes [\alpha_2(w_2), \beta_2(w_2)] \otimes [l_1, u_1] \otimes [l_2, u_2]$, với mọi $w = (w_1, w_2) \in V = V_1\times V_2$.

5. $prob_{s,I,o}(E_1 \otimes E_2) = prob_{s,I,o}(E_1) \otimes prob_{s,I,o}(E_2)$.

6. $prob_{s,I,o}(E_1 \oplus E_2) = prob_{s,I,o}(E_1) \oplus prob_{s,I,o}(E_2)$.

Về trực giác, $prob_{s,I,o}(x.P.\lambda')$ là khoảng xác suất để biểu thức đường đi $P.\lambda'$ có thể áp dụng cho đối tượng o; $prob_{s,I,o}(x.P.\lambda'\ \theta\ v)$ là khoảng xác suất để biểu thức đường đi $P.\lambda'$ có thể áp dụng cho o đồng thời tính chất P của o có giá trị w sao cho $w.\lambda'\ \theta\ v$. Đó là khoảng xác suất được tích hợp từ khả năng áp dụng không chắc chắn các tính chất trong đường đi $P.\lambda'$ và khả năng nhận giá trị w của P sao cho giá trị tính chất cuối cùng trong đường đi λ' thỏa mãn $w.\lambda'\ \theta\ v$ (Định nghĩa 4.3.2); $prob_{s,I,o}(x.P_1.\lambda'_1 =\otimes x.P_2.\lambda'_2)$ là khoảng xác suất để $P_1.\lambda'_1$ và $P_2.\lambda'_2$ có thể áp dụng cho o, đồng thời P_1 và P_2 của o có các giá trị tương ứng là w_1 và w_2, sao cho $w_1.\lambda'_1 = w_2.\lambda'_2$. Ở đây, các biểu thức đường đi λ', λ'_1 và λ'_2 có thể rỗng. Chúng tôi qui ước $[w.\lambda'] = [1, 1]$ nếu λ' rỗng.

1. Ví dụ 5.2.2 Giả sử ***S*** là lược đồ của thể hiện ***I*** trong Ví dụ 4.1.1, thì $prob_{s,I,o_1}(x.\text{medical\_history} \otimes_{in} x.\text{duration} \geq 500) = prob_{s,I,o_1}(x.\text{medical\_history}) \otimes_{in} prob_{s,I,o_1}(x.\text{duration} \geq 500) = [0.32, 0.4] \otimes_{in} [u(400) \times prob(400 \geq 500) + u(500) \times prob(500 \geq 500), min(1, u(400) \times prob(400 \geq 500) + u(500) \times prob(500 \geq 500))] = [0.32, 0.4] \otimes_{in} [0.5, min(1, 0.5)] = [0.16, 0.2]$, là khoảng xác suất để bệnh nhân, được biểu diễn bởi o_1, có tiền sử bệnh và thời gian điều trị ít nhất là 500 (ngày).

Một đối tượng thỏa mãn một điều kiện chọn nếu xác suất chọn đối tượng này thuộc về khoảng xác suất được yêu cầu trong điều kiện chọn như định nghĩa sau.

Định nghĩa 5.2.3 Giả sử $\boldsymbol{S} = (\mathbf{C}, \tau, \Rightarrow, \text{me}, \wp, \text{f})$ là một lược đồ FPOB, $\boldsymbol{I} = (\pi, v)$ là một thể hiện trên ***S*** và $o \in \pi(\mathbf{C})$. *Sự thỏa mãn* các điều kiện chọn của o theo diễn dịch xác suất $prob_{s,I,o}$ được định nghĩa như sau:

$prob_{s,I,o}\ (E)[L, U]$ nếu và chỉ nếu $prob_{s,I,o}(E) \subseteq [L, U]$.

1. $prob_{s,I,o}\neg\varphi$ nếu và chỉ nếu $prob_{s,I,o}$‘ φ không thỏa.

$prob_{s,I,o}\ \varphi \wedge \psi$ nếu và chỉ nếu $prob_{s,I,o}$‘ φ và $prob_{s,I,o}$‘ ψ.

$prob_{s,I,o}\ \varphi \vee \psi$ nếu và chỉ nếu $prob_{s,I,o}$‘ φ hoặc $prob_{s,I,o}$‘ ψ.

Ví dụ 5.2.3 Gọi *approxi_40* = (35: 0; 40: 1; 45: 0) là một số mờ liên tục có đồ thị hàm thành viên dạng tam giác có tọa độ ba đỉnh là (40, 1), (35, 0) và (45, 0). Giả sử ***S*** là lược đồ FPOB của thể hiện ***I*** trong Ví dụ 4.1.1, thì chọn “tất cả các bệnh nhân tuổi xấp xỉ 40 với xác suất từ 0.2 đến 0.5 và có tiền sử bệnh với xác suất ít nhất là 0.6” có thể được thực hiện bởi điều kiện chọn $\varphi = (x.\text{age} \rightarrow approxi\_40)[0.2, 0.5] \wedge (x.\text{medical\_history})[0.6, 1]$.

Từ định nghĩa hàm thành viên của *approxi_40* ở trên và *middle_aged* trong Ví dụ 3.2.2 ta có các lát cắt tương ứng của chúng là ${}^{\alpha}approxi\_40 = [35+5\alpha, 45-5\alpha]$ và ${}^{\alpha}middle\_aged = [20+15\alpha, 60-15\alpha]$. Theo Định nghĩa 2.2.2 ta có:

$$prob(middle\_aged \rightarrow approxi\_40) = \int_0^1\int_0^1 \frac{\left|{}^{x}middle\_aged \cap {}^{y}approxi\_40\right|}{\left|{}^{x}middle\_aged\right|}dxdy$$

$$= \int_0^1\int_0^1 \frac{\left|[20+15x, 60-15x] \cap [35+5y, 45-5y]\right|}{\left|[20+15x, 60-15x]\right|}dxdy$$

$$= \int_0^1\int_0^1 \frac{10-10y}{40-30x}dxdy = \int_0^1\int_0^1 \frac{1-y}{4-3x}dxdy$$

$$= \int_0^1 \left|\frac{(y-y^2/2)\Big|_0^1}{4-3x}\right|dx = -\frac{1}{2}\int_0^1 \frac{1}{3x-4}dx$$

$$= -\frac{\ln\left|3x-4\right|\Big|_0^1}{6} = \frac{\ln 4}{6} \approx 0.23$$

Xét đối tượng o_4 trong ***I***, thì $prob_{s,I,o_4}$(*x*.age → *approxi_40*) = [*u*(*middle_aged*) × *prob*(*middle_aged* → *approxi_40*), *min*(1, *u*(*middle_aged*) × *prob*(*middle_aged* → *approxi_40*))] = [0.23, *min*(1, 0.23] =[0.23, 0.23] ⊆ [0.2, 0.5] và $prob_{s,I,o_1}$(*x*.medical_history)= [0.7, 1] ⊆ [0.6, 1]. Vậy $prob_{s,I,o_4}$ ‘ φ.

Bây giờ, phép chọn trên các thể hiện FPOB được định nghĩa để cho phép truy vấn theo các điều kiện chọn với khả năng áp dụng không chắc chắn của các tính chất đối tượng như dưới đây.

Định nghĩa 5.2.4 Giả sử ***S*** = (**C**, τ, ⇒, me, ℘, f) là một lược đồ FPOB, ***I*** = (π, ν) là một thể hiện trên ***S*** và φ là một điều kiện chọn trên biến đối tượng *x*. *Phép chọn* trên ***I*** theo φ, được ký hiệu σ_φ(*I*), là một thể hiện ***I'*** = (π', ν') trên ***S***, trong đó:

π'(*c*) = {*o* ∈ π(*c*)| $prob_{s,I,o}$ ‘ φ} và

1. ν' = ν | π'(**C**) (nghĩa là ánh xạ ν thu hẹp trên π'(**C**)).

Ví dụ 5.2.4 Giả sử ***I*** = (π, ν) là thể hiện FPOB trên lược đồ ***S*** trong Ví dụ 4.1.1. Thì truy vấn "chọn tất cả bệnh nhân có tiền sử bệnh viêm túi mật với một xác suất ít nhất là 0.6 và có thể phải trả tổng chi phí điều trị là khoảng 1800 với một xác suất ít nhất là 0.4" có thể được thực hiện bởi phép chọn

I' = σ_φ(***I***)

trong đó, φ = (*x*.medical_history = {*cholecystitis*})[0.6, 1] ∧ (*x*.total_cost(*x*.duration, *x*.cost_per_day) → *about_1800*)[0.4, 1]. Thể hiện ***I'*** trên ***S*** bao gồm chỉ một đối tượng o_4, theo chiến lược hội độc lập. Thật vậy, vì [α'(*w*), β'(*w*)] = [*u*(*w*), *u*(*w*)] $\otimes_{in}$ [0.7, 1.0] = [0.7*u*(*w*), *u*(*w*)] với mọi *w* ∈ {{*cholecystitis*}}, nên $prob_{s,I,o_4}$(*x*.medical_history = {*cholecystitis*}) = [0.7, 1] ⊆ [0.6, 1]. Mặt khác từ Ví dụ 4.1.1 ta có $prob_{s,I,o_4}$(*x*.total_cost(*x*.duration, *x*.cost_per_day) → *about_1800*) = [*u*(*about_1800*) × *prob*(*about_1800* → *about_1800*) + *u*(*about_2400*) × *prob*(*about_2400* → *about_1800*), *min*(1, *u*(*about_1800*) × *prob*(*about_1800* → *about_1800*) + *u*(*about_2400*) × *prob*(*about_2400* → *about_1800*))] = [0.43, *min*(1, 0.43)] =[0.43, 0.43] ⊆ [0.4, 1]. Do đó, $prob_{s,I,o_4}$ ‘ φ. Kiểm tra tương tự, thấy rằng không còn đối tượng nào khác trong ***I*** thỏa điều kiện chọn này.

6. KẾT LUẬN

Trong bài báo này, chúng tôi đã trình bày mô hình cơ sở đối tượng FPOB, như là một mở rộng mờ của mô hình xác suất được giới thiệu bởi Eiter và CS. FPOB hỗ trợ phương thức lớp và cho phép khả năng áp dụng không chắc chắn các tính chất lớp của các đối tượng. Trong FPOB, giá trị các tính chất đối tượng không chắc chắn và không chính xác được biểu diễn bởi các phân bố xác suất trên một tập các giá trị tập mờ. Tính không chắc chắn của khả năng áp dụng các tính chất đối tượng được biểu diễn bởi một khoảng xác suất. Phương thức lớp như là một hàm có đối số và giá trị là các bộ ba xác suất mờ.

Diễn dịch xác suất của các quan hệ trên các tập mờ được đề nghị, làm cơ sở để kết hợp xác suất và giá trị tập mờ vào trong một mô hình hợp nhất. Một đại số các bộ ba xác suất mờ được phát triển để hỗ trợ cho việc định nghĩa các phương thức lớp. Phép toán chọn các đối tượng được nghiên cứu và định nghĩa một cách hình thức trên mô hình mới cho phép truy vấn cơ sở đối tượng với giá trị thuộc tính và phương thức lớp không chắc chắn và không chính xác.

Từ kết quả nghiên cứu này, trong tương lai, chúng tôi sẽ mở rộng hơn nữa FPOB với các phép toán đại số khác như *chiếu* (projection), *kết* (join), vv., cho phép đa thừa kế không chắc chắn và phát triển một hệ quản trị FPOB cho các áp dụng thực tế.

TÀI LIỆU THAM KHẢO

[1] J.F. Baldwin, J.M. Lawry, T.P. Martin, *A note on probability/possibility consistency for fuzzy events*, in: Proc. 6th International Conference on Information Processing and Management of Uncertainty in Knowledge-Based Systems, IPMU'96, Granada, Spain, 1996, pp. 521-525.

[2] J.F. Baldwin, J.M. Lawry, T.P. Martin, *A mass assignment theory of the probability of fuzzy events*, Int.J. Fuzzy Sets and Systems, 83 (1996), 353-367.

[3] J.F. Baldwin, T.H. Cao, T.P. Martin, J.M. Rossiter, *Toward soft computing object-oriented logic programming*, in: Proc. 9th IEEE International Conference on Fuzzy Systems, 2000, pp. 768-773.

[4] Blanco, N. Marín, O. Pons, M.A. Vila, *Softening the object-oriented database model: imprecision, uncertainty and fuzzy types*, in: Proc 1st International Joint Conference of the International Fuzzy Systems Association and the North American Fuzzy Information Processing Society, 2001, pp. 2323-2328.

[5] G. Bordogna, G. Pasi, D. Lucarella, *A fuzzy object-oriented data model managing vague and uncertain information*, Int. J. Intelligent Systems, 14 (1999), 623-651.

[6] T.H. Cao, H. Nguyen, *Fuzzy and probabilistic object bases*, in: Z. Ma. (Eds.): Advances in Fuzzy Object-Oriented Databases: Modelling and Applications, Idea Group Publisher, 2005, pp. 46-84.

[7] T.H. Cao, J.M. Rossiter, *A deductive probabilistic and fuzzy object-oriented database language*, Int.J. Fuzzy Sets and Systems, 140 (2003), 129-150.

[8] W. Dubitzky et al, *Towards concept-oriented databases*, Int.J. Data & Knowledge Engineering, 30 (1999), 23-55.

[9] T. Eiter, J.J. Lu, T. Lukasiewicz, V.S. Subrahmanian, *Probabilistic object bases*, Int.J. ACM Transactions on Database Systems, 26 (2001), 264-312.

[10] B.R. Gaines, *Fuzzy and probability uncertainty logics*, Int.J. Information and Control, 38 (1978), 154-169.

[11] R. George, B.P. Buckles, F.E. Petry, *Modelling class hierarchies in the fuzzy object-oriented data model*, Inte. J. Fuzzy Sets and Systems, 60 (1993), 259-272.

[12] G.J. Klir, B. Yuan, *Fuzzy Sets and Fuzzy Logic - Theory and Applications*, Prentice Hall PTR, 1995.

[13] L.V.S. Lakshmanan et al., *ProbView: A flexible probabilistic database system*, Int.J. ACM Transactions on Database Systems, 22 (1997), 419-469.

[14] H. Nguyen, T.H. Cao, *Extending probabilistic object bases with uncertain applicability and imprecise values of class properties*, in: Proc 5th IEEE International Conference on Fuzzy Systems, London, England, 2007, pp.487-492.

[15] H. Nguyen, *A fuzzy and probabilistic object base model*. Ph.D thesis, Ho Chi Minh City University of Technology, 2008.

[16] R. Ross, V.S. Subrahmanian, *Aggregate operators in probabilistic databases*, Int. J. ACM, 52 (2005), 54-101.

[17] J-P. Rossazza, D. Dubois, H. Prade, *A hierarchical model of fuzzy classes*, in: R. De Caluwe. (Eds.): Fuzzy and Uncertain Object-Oriented Databases: Concepts and Models, 1997, pp. 21-61.

[18] N. Van Gyseghem, R. De Caluwe, *The UFO database model: dealing with imperfect information*, in: R. De Caluwe. (Eds.): Fuzzy and Uncertain Object-Oriented Databases: Concepts and Models, World Scientific 1997, pp. 123-185.

[19] Yazici, R. George, *Fuzzy database modelling, Studies in Fuzziness and Soft Computing*, Vol. 26, 1999, Physica-Verlag.

TÓM TẮT

Bài báo này giới thiệu một mô hình cơ sở đối tượng xác suất mờ (FPOB) là mở rộng mô hình cơ sở đối tượng xác suất của Eiter và cộng sự (CS) với ba 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; và (3) khả năng áp dụng và thừa kế các tính chất lớp (class property) là không chắc chắn. 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, thể hiện và phép chọn cơ sở đối tượng xác suất mờ được được nghiên cứu và định nghĩa một cách hình thức để hỗ trợ truy vấn thông tin không chắc chắn, khômg chính xác trên FPOB.

Từ khóa: phân bố xác suất; khả năng áp dụng không chắc chắn; tập mờ; phương thức lớp; phép chọn.

SUMMARY

This paper introduces a fuzzy and probabilistic object base model (FPOB) that extending Eiter et al.'s probabilistic object base model with three 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; and (3) applicability of class properties and their inheritance can be uncertain. 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 the syntax and semantics of fuzzy-probabilistic object base schemas, instances, and selection operation are formally researched and defined to support queries with the imprecise and uncertain information on FPOB.

Keywords: probability distribution; uncertain applicability; fuzzy sets; class methods; selection operation.

Tên bài báo tiếng Anh: Integration of Uncertain Applicabilities of Class Properties intoa Fuzzy and Probabilistic Object Base Model