JOURNAL OF SCIENCE OF HNUE DOI: 10.18173/2354-1075.2015-0063
Educational Sci., 2015, Vol. 60, No. 7A, pp. 157-168
This paper is available online at http://stdb.hnue.edu.vn

PHƯƠNG PHÁP XỬ LÍ TRUY VẤN ĐA ĐIỀU KIỆN TRONG CƠ SỞ DỮ LIỆU HƯỚNG ĐỐI TƯỢNG MỜ DỰA TRÊN NGỮ NGHĨA ĐẠI SỐ GIA TỬ

Nguyễn Tấn Thuận[1], Đoàn Văn Ban[2]
Trương Ngọc Châu[3], Trần Thị Thúy Trinh[1]
[1]*Khoa Công nghệ Thông tin, Đại học Duy Tân;*
[2]*Viện Công nghệ Thông tin, Viện Hàn lâm Khoa học và Công nghệ Việt Nam;*
[3]*Khoa Công nghệ Thông tin, Đại học Bách khoa Đà Nẵng*

Tóm tắt. Bài báo này đề xuất các phương pháp xử lí truy vấn mờ thông qua hướng tiếp cận ĐSGT. Bằng các phương pháp xử lí khác nhau như xác định giá trị chân lí, dựa trên cơ sở đó chúng tôi xây dựng bộ truy vấn mờ cho các trường hợp xử lí đơn điều kiện và đa điều kiện, kết quả được kiểm chứng thông qua lược đồ cơ sở dữ liệu hướng đối tượng mờ.
Từ khóa: ĐSGT,Giá trị chân lí, Đơn điều kiện, Đa điều kiện, truy vấn mờ

1. Mở đầu

Ngôn ngữ truy vấn trong CSDL mờ được các tác giả quan tâm nghiên cứu và đã có nhiều kết quả. Truy vấn theo cách tiếp cận lí thuyết tập mờ [3, 8], truy vấn theo cách tiếp cận quan hệ tương tự [1], truy vấn theo hướng tiếp cận ĐSGT [2]. Hầu hết, các tác giả đều xây dựng ngôn ngữ truy vấn với mong muốn thao tác mềm dẻo, "chính xác" với dữ liệu mờ bằng cách tập trung xây dựng các hàm thuộc, từ đó tùy theo ngữ nghĩa của dữ liệu để chọn các ngưỡng phù hợp khi thao tác dữ liệu.

Tuy nhiên, mỗi ngôn ngữ truy vấn chỉ phù hợp với một mô hình CSDL mờ cụ thể mà không có ngôn ngữ nào tổng quát cho các mô hình. Chẳng hạn, đối với ngôn ngữ truy vấn trên mô hình CSDL mờ theo cách tiếp cận tập mờ, việc xác định giá trị chân lí của điều kiện mờ trong truy vấn thường là khó khăn và kết quả của truy vấn phụ thuộc nhiều vào việc xây dựng các hàm thuộc.

Dựa trên mô hình đã được đề xuất [2] đó, chúng tôi tiếp tục nghiên cứu một số phương pháp mới xử lí truy vấn hướng đối tượng mờ và xem xét việc đưa điều kiện lọc vào trong câu truy vấn nhằm làm cho thao tác tìm kiếm dữ liệu trên mô hình này được mềm dẻo và phù hợp với thực tế.

Bài báo gồm 5 phần. Phần 2 giới thiệu mô hình CSDL hướng đối tượng mờ và đại số gia tử, phần 3 khái quát lớp mờ trong cơ sở dữ liệu HĐT mờ, phần 4 trình bày các phương pháp xử lí truy vấn trong CSDL hướng đối tượng mờ, phần 5 phương pháp truy vấn dữ liệu mờ và phần 6 đánh giá hiệu năng.

Ngày nhận bài: 5/7/2015. Ngày nhận đăng: 10/11/2015.
Liên hệ: Nguyễn Tấn Thuận, e-mail: nguyentanthuan2008@yahoo.com.

2. Nội dung nghiên cứu

2.1. Mô hình cơ sở dữ liệu hướng đối tượng mờ dựa trên đại số gia tử

Một mô hình dữ liêu hướng đối tượng mờ được định nghĩa như một mở rộng mô hình dữ liệu hướng đối tượng rõ bằng cách thay thế các đối tượng bởi các đối tượng mờ, các lớp bởi các lớp đối tượng mờ và các kết hợp bởi các kết hợp mờ trên lược đồ CSDL HĐT.

Một hướng tiếp cận mới để xây dựng mô hình cơ sở dữ liệu hướng đối tượng với thông tin mờ đó là dựa trên đại số gia tử (ĐSGT), một mô hình toán học về cấu trúc thứ tự ngữ nghĩa của miền giá trị của biến ngôn ngữ [2, 4].

Xét một lược đồ CSDL hướng đối tượng mờ trên miền vũ trụ $U = \{CA1, CA2, \ldots, Can\}$. Mỗi thuộc tính Ai được gắn với một miền trị thuộc tính, kí hiệu là Dom(CAi), trong đó một số thuộc tính cho phép nhận các giá trị ngôn ngữ trong lưu trữ hay trong các câu truy vấn và được gọi là thuộc tính mờ. Các thuộc tính còn lại được gọi là thuộc tính kinh điển. Thuộc tính kinh điển Ai được gắn với một miền giá trị kinh điển, kí hiệu là D_{Aj}. Thuộc tính mờ Ai sẽ được gắn một miền giá trị kinh điển D_{Aj} và một miền giá trị ngôn ngữ LD_{Aj} hay là tập các phần tử của một ĐSGT. Xem giá trị ngôn ngữ như là một phần tử của ĐSGT. Để bảo đảm tính nhất quán trong xử lí ngữ nghĩa dữ liệu trên cơ sở thống nhất kiểu dữ liệu của thuộc tính mờ, mỗi thuộc tính mờ sẽ được gắn với một ánh xạ định lượng ngữ nghĩa ĐSGT.

Theo cách tiếp cận này giá trị ngôn ngữ là dữ liệu, không phải là nhãn của các tập mờ biểu diễn ngữ nghĩa của giá trị ngôn ngữ và ưu điểm cơ bản của nó là việc cho phép tìm kiếm, đánh giá ngữ nghĩa của thông tin không chắc chắn chỉ bằng các thao tác dữ liệu kinh điển thường dùng và do đó bảo đảm tính thuần nhất của kiểu dữ liệu trong xử lí ngữ nghĩa của chúng. Vì tất cả các thuộc tính có miền trị chứa giá trị số trong CSDL đều tuyến tính, nên một cách tự nhiên ta giả thiết ĐSGT được sử dụng là ĐSGT tuyến tính, do đó tập $H+$ và $H-$ là tập sắp thứ tự tuyến tính. Như vậy, cho $X = (X, G, H, \leq)$ với $G = \{0, c-, W, c+, 1\}$, $H = H- \cup H+$ với giả thiết $H- = \{h1, h2, .., hp\}$, $H+ = \{h-1, ..., h-q\}$, $h1 > h2 > ... > hp$ và $h-1 < ... < h-q$ là dãy các gia tử. Cho một ĐSGT tuyến tính đầy đủ $AX = (X, G, H, \sum, \Phi, \leq)$, trong đó $Dom(X) = X$ là miền các giá trị ngôn ngữ của thuộc tính ngôn ngữ X được sinh từ tập các phần tử sinh $G = \{0, c-, W, c+, 1\}$ bằng việc tác động các gia tử trong tập $H, \sum$ và Φ là hai phép tính với ngữ nghĩa là cận trên đúng và cận dưới đúng của tập $H(x)$, tức là $\sum x = supremumH(x)$ và $\Phi x = infimumH(x)$, quan hệ $\leq$ là quan hệ sắp thứ tự tuyến tính trên X cảm sinh từ ngữ nghĩa của ngôn ngữ.

2.2. Lớp mờ trong cơ sở dữ liệu hướng đối tượng mờ

Dựa vào các thảo luận trên, chúng ta thấy rằng các lớp trong CSDL hướng đối tượng mờ có thể mờ. Theo đó, trong CSDL hướng đối tượng mờ, một đối tượng thuộc Z thuộc một lớp tùy theo mức k và một lớp là lớp con của một lớp khác cũng theo mức k $(k+)$. Trong CSDL hướng đối tượng, một lớp được định nghĩa bao gồm mối quan hệ kế thừa, thuộc tính và phương thức. Để xác định một lớp mờ, cần bổ sung một số định nghĩa mới. Khi khai báo mối quan hệ kế thừa cần chỉ ra mức mà lớp này là lớp con của lớp cha, trong định nghĩa của một lớp mờ, các thuộc tính mờ có thể được chỉ ra một cách rõ ràng. Về mặt hình thức, định nghĩa của một lớp mờ được thể hiện như sau:

CLASS tên lớp
  INHERITES
    tên lớp cha thứ 1 WITH LEVEL OF mức_1
    ...
    tên lớp cha thứ n WITH LEVEL OF mức_n
  ATTRIBUTES
    tên thuộc tính thứ 1: [FUZZY] DOMAIN dom_1: TYPE OF kiểu_1
    ...
    tên thuộc tính thứ m: [FUZZY] DOMAIN dom_m: TYPE OF kiểu_m
  METHODS ...
END

Đối với các thuộc tính không phải là thuộc tính mờ, các kiểu dữ liệu bao gồm các loại đơn giản như số nguyên, thực, logic, chuỗi, và các loại phức tạp như kiểu tập hợp và kiểu đối tượng. Đối với thuộc tính mờ, các kiểu dữ liệu là kiểu mờ dựa trên các kiểu dữ liệu đơn giản hoặc phức tạp, cho phép biểu diễn thông tin không chính xác [10].

2.3. Ngôn ngữ truy vấn trong cơ sở dữ liệu hướng đối tượng mờ

2.3.1. Truy vấn dữ liệu mờ

Việc xác định giá trị chân lí của điều kiện mờ trong truy vấn thường là khó khăn và kết quả của truy vấn phụ thuộc nhiều vào việc xây dựng các hàm thuộc. Nếu chúng ta xem miền trị của thuộc tính mờ là một ĐSGT thì việc xây dựng ngôn ngữ truy vấn trên mô hình CSDL HĐT mờ để phù hợp với ngữ nghĩa mới là cần thiết cho việc thao tác và tìm kiếm dữ liệu, giá trị chân lí của điều kiện mờ trong câu truy vấn được xác định đơn giản và hiệu quả. Vấn đề quan tâm ở đây là một ngôn ngữ được đề xuất dùng để truy vấn dữ liệu trong mô hình CSDL HĐT mờ đã được xây dựng trong phần 4.1. Do đó, trước tiên phải đi xây dựng các thuật toán xác định giá trị chân lí của điều kiện mờ để làm cơ sở khi truy vấn dữ liệu, tiếp đến áp dụng các thuật toán vừa xây dựng để thực hiện truy vấn.

2.3.2. Câu lệnh Select

SELECT <danh sách thuộc tính>

FROM <Lớp1 WITH mức_k1, ..., Lớp m WITH mức_km>

WHERE <điều_kiện_truy_vấn WITH mức_k>

Trong đó, <điều kiện truy vấn> là một điều kiện mờ hoặc liên kết các điều kiện mờ có sử dụng các phép toán tuyển và hội, k là số nguyên dương.

2.3.3. Thuật toán xác định giá trị chân lí của điều kiện mờ

Thuật toán 1. Xác định giá trị chân lí của đơn điều kiện mờ với phép toán θ

Vào: Lớp C cùng với các thuộc tính $\{A_1, A_2, \ldots, A_n\}$, tập các đối tượng thuộc lớp C: $\{Oi, i = 1, \ldots, m\}$

Điều kiện Ai θ fvalue, với fvalue là giá trị mờ và A_i là thuộc tính mờ C tính đơn điệu,

$\theta \in \{= k, \neq k, < k, > k\}$.

Ra: Với mọi $t \in O$ thoả mãn điều kiện $t[A_i]\theta$ fvalue

Phương pháp

```
Begin
for each t ∈ O do
  if t[A_i] ∈ D_Ai then t[A_i] = Φ_k(f(t[A_i])) // Xây dựng các P^k dựa vào độ dài các từ.
  k = 1
  while k ≤ p do
  begin
  P^k = Φ
    for each t ∈ O do
      if |t[A_i]| = k then P^k = P^k ∪ {I(t[A_i])}
    k = k + 1
  end
  if θ là phép toán = k then
  begin
    k = p
    while k > 0 do
      begin
        for each Δ^k ∈ P^k do
          begin
          if (I(t[A_i]) ⊆ Δ^k and I(fvalue) ⊆ Δ^k) then
(t[A_i] = k fvalue) = 1
          exit
          end
      k = k − 1
      end
  end
    else
  begin
    k = 1
    while k ≤ p do
      begin
        for each Δ^k ∈ P^k do
          begin
          Case θ of
            ≠k: if I(t[A_i]) ⊄ Δ^k or I(fvalue) ⊄ Δ^k then (t[A_i]θ fvalue) = 1
```

$< k$: if ($t[A_i] \neq k$ fvalue) and $v(t[A_i]) < v$ (fvalue) then ($t[A_i]\theta$ fvalue) = 1
$> k$: if ($t[A_i] \neq k$ fvalue) and $v(t[A_i]) > v$ (fvalue) then ($t[A_i]\theta$ fvalue) = 1
exit
end
end
$k = k + 1$
end
end

Trong trường hợp đa điều kiện mờ với phép toán θ, không mất tính tổng quát, giả sử điều kiện mờ có dạng Ai θ $fvalue_i$ ζ Aj θ_1 $fvalue_j$, với $fvalue_i$, $fvalue_j$ là giá trị mờ, Ai, Aj là thuộc tính mờ có tính đơn điệu; θ, $\theta_1 \in \{= k, \neq k, < k, > k\}$ và ζ là phép toán and hoặc or. Muốn xác định giá trị chân lí trong trường hợp này, trước hết các điều kiện Ai θ $fvalue_i$ và Aj $\theta 1$ $fvalue_j$ được xác định dựa vào Thuật toán 1, tiếp theo tuỳ thuộc vào ζ là phép toán and hay or để kết hợp các giá trị chân lí vừa xác định. Do đó, Thuật toán xác định giá trị chân lí của đa điều kiện mờ được trình bày chi tiết:

Thuật toán 2. Xác định giá trị chân lí của đa điều kiện mờ với phép toán θ.

Vào: Lớp C cùng với các thuộc tính $\{A_1, A_2, \ldots, A_n\}$, tập các đối tượng thuộc lớp C: $\{Oi, i = 1, \ldots, m\}$

Điều kiện Ai θ $fvalue_i$ ζ Aj θ_1 $fvalue_j$.

Ra: Với mọi t $\in$Oi thoả mãn điều kiện ($t[A_i]\theta$ $fvalue_i$ ζ $t[A_j]\theta_1$ $fvalue_j$).

Phương pháp.

Begin
for each $t \in O$ do
begin
if $t[A_i] \in D_{Aj}$ then $t[A_i] = \Phi_k(f(t[A_i]))$
if $t[A_j] \in D_{Aj}$ then $t[A_j] = \Phi_k(f(t[A_j]))$
end
// Xây dựng các P^K_{Ai} và P^K_{Aj} dựa vào độ dài các từ
$k = 1$
while $k \leq p$ do
begin
$P^k_{Ai} = \emptyset$; $P^k_{Aj} = \emptyset$;
for each $t \in O$ do
begin
if $|t[A_i]| = k$ then $P^K_{Ai} = P^K_{Ai} \cup \{I(t[A_i])\}$
if $|t[A_j]| = k$ then $P^K_{Aj} = P^K_{Aj} \cup \{I(t[A_j])\}$
end
$k = k + 1$

end

foreach $t \in O$ do

begin

if (($t[A_i]\theta\ fvalue_i$) = 1) and (($t[A_j]\theta_1\ fvalue_j$) = 1) then (($t[A_i]\theta\ fvalue_i$) and ($t[A_j]\theta_1\ fvalue_j$)) = 1

if (($t[A_i]\theta\ fvalue_i$) = 1) or (($t[A_j]\theta_1\ fvalue_j$) = 1) then (($t[A_i]\theta\ fvalue_i$) or ($t[A_j]\theta_1\ fvalue_j$)) = 1

end

end

Thuật toán 3. Thực hiện truy vấn FOQL mờ trong trường hợp đơn điều kiện.

Vào: Lớp C cùng với các thuộc tính $\{A_1, A_2, \ldots, A_n\}$, tập các đối tượng thuộc lớp C: $\{Oi, i = 1, \ldots, m\}$.

Câu truy vấn dạng select ... from.... O where A_i θ (Ai) fvalue.

Ra: Quan hệ O_{result} thỏa mãn với mọi $t \in O_{result}$ ta có $t[A_i]\theta$ (Ai) fvalue.

Phương pháp

Begin

Cho $G_{Ai} = \{O, c - Ai, W, c + Ai, 1\}$, $H_{Ai} = H^+_{Ai} \cup H^-_{Ai}$.

Trong đó $H^+_{Ai} = \{h1, h2\}$, $H^-_{Ai} = \{h3, h4\}$, với $h_1 < h_2$ và $h_3 > h_4$. Chọn độ đo tính mờ cho các phần tử sinh và gia tử.

D_{Aj} = [minAi, maxAi], với minAi, maxAi: giá trị nhỏ nhất và lớn nhất miền trị Ai.

$LD_{Aj} = H_{Ai}(c^+) \cup H_{Ai}(c^-)$.

$O_{result} = \emptyset$.

$k = 1$

While $k \leq p$ do // mức tương tự lớn nhất $k = p$

begin

Xây dựng các khoảng tương tự mức k: $S_{Aj}(x_1), S_{Aj}(x_2), \ldots, S_{Aj}(x_m)$

$k = k + 1$

end

if fvalue $\in S(x_i)$ then $\Omega_k(fvalue) = S(x_i)$

foreach $t \in O$ do

if $t[A_i] \in LD_{Aj}$ then Xác định lân cận mức k của $t[A_i]$ là $\Omega_k(t[A_i]) = S(x_j)$

for each $t \in O$ do

if θ là phép $= k$ then

begin

if $t[A_i] \in \Omega_k$ (fvalue) then O_{result} = $O_{result} \cup t$

elseif $\Omega_k(t[A_i]) = \Omega_k$ (fvalue) then $O_{result} = O_{result} \cup t$

end

else

begin

Case θ of

$\neq k$: if $\Omega_k(t[A_i]) \neq \Omega_k$ (fvalue) then $O_{result} = O_{result} \cup t$

$< k$: if $\Omega_k(t[A_i]) < \Omega_k$ (fvalue) then $O_{result} = O_{result} \cup t$

$> k$: if $\Omega_k(t[A_i]) > \Omega_k$ (fvalue) then $O_{result} = O_{result} \cup t$

end

return O_{result}

end.

Trong trường hợp đa điều kiện mờ, muốn xác định giá trị chân lí trước hết các điều kiện Ai θ(Ai) $fvalue_i$ và Aj θ_1(Aj) $fvalue_j$ được xác định dựa vào Thuật toán 3, tiếp theo tuỳ thuộc vào ζ là phép toán and hay or để kết hợp các giá trị chân lí vừa xác định. Do đó, Thuật toán được xây dựng chi tiết như sau:

Thuật toán 4. Thực hiện truy vấn FOQL mờ trong trường hợp đa điều kiện.

Vào: Lớp C cùng với các thuộc tính $\{A_1, A_2, \ldots, A_n\}$, tập các đối tượng thuộc lớp C: $\{Oi, i = 1, \ldots, m\}$.

Câu truy vấn dạng select ... from O where Ai θ(Ai) $fvalue_i$ ζ Aj θ_1(Aj) $fvalue_j$, trong đó ζ là phép toán and hoặc or.

Ra: Đối tượng O_{result} thỏa mãn với mọi $t \in O_{result}$ ta có $t[A_i]\theta$ (Ai) $fvalue_i$ ζ $t[A_j]\theta_1$(Aj) $fvalue_j$.

Phương pháp

Begin

Cho $G_{Ai} = \{0, C^-_{Ai}, W, C^+_{Ai}, 1\}, H_{Ai} = H^+_{Ai} \cup H^-_{Ai}$.

Trong đó $H^+_{Ai} = \{h1, h2\}, H^-_{Ai} = \{h3, h4\}$, với $h_1 < h_2$ và $h_3 > h_4$.

Chọn độ đo tính mờ cho các phần tử sinh và gia tử.

Cho $G_{Aj} = \{0, C^-_{Aj}, W, C^+_{Aj}, 1\}, H_{Aj} = H^+_{Aj} \cup H^-_{Aj}$.

Trong đó $H^+_{Aj} = \{h1, h2\}, H^-_{Aj} = \{h3, h4\}$, với $h_1 < h_2$ và $h_3 > h_4$.

Chọn độ đo tính mờ cho các phần tử sinh và gia tử.

$D_{Aj} = [\min Ai, \max Ai]$, với $\min Ai$, $\max Ai$: giá trị nhỏ nhất và lớn nhất miền trị A_i.

$D_{Aj} = [\min Aj, \max Aj]$, với $\min Aj, \max Aj$: giá trị nhỏ nhất và lớn nhất miền trị A_j.

$LD_{Aj} = H_{Ai}(C^+_{Ai}) \cup H_{Ai}(C^-_{Ai})$.

$LD_{Aj} = H_{Aj}(C^+_{Aj}) \cup H_{Aj}(C^-_{Aj})$.

$O_{result} = \emptyset$.

// Phân hoạch D_{Aj} và D_{Aj} thành các khoảng tương tự mức k.

$k = 1$

While $k \leq p$ do

begin

Xây dựng các khoảng tương tự mức k : $S_{Aj}(x_1)$, $S_{Aj}(x_2)$, ...,$S_{Aj}(x_m)$

Xây dựng các khoảng tương tự mức k : $S_{Aj}(y_1)$, $S_{Aj}(y_2)$, ...,$S_{Aj}(y_m)$

$k = k + 1$

end

// Xác định lân cận mức k của $fvalue_i$ và $fvalue_j$

if $fvalue_i \in S_{Aj}(x_i)$ then Ω_k ($fvalue_i$) = $S_{Aj}(x_i)$

if $fvalue_j \in S_{Aj}(y_i)$ then Ω_k ($fvalue_j$) = $S_{Aj}(y_i)$

//Duyệt các đối tượng trong O để tìm các thuộc tính thỏa mãn điều kiện

$t[A_i]$ θ (Ai) $fvalue_i$ ζ

$t[A_j] = \theta_1$(Aj) $fvalue_j$

// Trường hợp ζ là phép toán and

for each $t \in O$ do

if $\{(t[A_i]\theta(Ai)\ fvalue_i)$ = 1} and $\{(t[A_j]\theta_1$ (Aj) $fvalue_j)$ = 1} then $O_{result} = O_{result} \cup t$

if $\{(t[A_i]\theta(Ai)\ fvalue_i) = 1\}$ or $\{(t[A_j]\theta_1(Aj)\ fvalue_j) = 1\}$ then $O_{result} = O_{result} \cup t$

Return O_{result}.

End.

2.4. Phương pháp truy vấn dữ liệu mờ

Các ngôn ngữ truy vấn trong mô hình CSDL HĐT được nhiều tác giả quan tâm nghiên cứu và mở rộng trong mô hình CSDL HĐT mờ như: đại số quan hệ mờ, ngôn ngữ FOQL. Vì thế, tương tự trong CSDL HĐT, cấu trúc của câu lệnh FOQL trong CSDL HĐT mờ được xem xét như trong mục trên. Như vậy, câu lệnh FOQL trong CSDL HĐT mờ có thể được tổng quát hoá sau: (1): Thực hiện phép tích Descartes giữa các quan hệ tham gia truy vấn. (2): Xác định giá trị chân lí của các điều kiện mờ (Sử dụng thuật toán 1 - 4) và liên kết các giá trị chân lí vừa xác định. (3): Chọn các bộ dữ liệu thoả mãn bước (2). Do đó, vấn đề quan trọng của câu lệnh FOQL trong CSDL HĐT mờ chính là xác định giá trị chân lí của điều kiện mờ và liên kết các giá trị chân lí đó.

Ví dụ 1. Cho một lớp "nhân viên": như sau:

Class NhanVien{

Oid: allID

Họ tên: string;

Sức khỏe: string;

Tuổi: [fuzzy] domain [18.. 60]: int

lương: [fuzzy] domain [0..7,5]: float

}

Như vậy tập các đối tượng thuộc lớp NhanVien: {Oi, i=SOCM, HOTEN, TUOI, LUONG}.

Và đối tượng Luong_tuoi được xác định như sau:

Bảng 1. Cho các đối tượng của lớp "Nhân viên"

Oid	SOCM	HOTEN	TUOI	LUONG
1	12345	Phạm Trọng	31	2.800.000
2	23456	Nguyễn Văn A	85	Cao

3	34567	Trần Tiến	32	2.000.000
4	45678	Vũ Hoàng	45	500.000
5	56789	An Thuyên	41	rất cao
6	67890	Thuận Yến	61	thấp
7	78912	Văn Cao	59	ít cao
8	89123	Thanh Tùng	75	2.500.000
9	90123	Nguyễn Cường	25	khá thấp

Tìm những cán bộ có TUOI =1 khả năng trẻ và LUONG =1 cao, sử dụng thuật toán 4.3 ta có:

(a). Xét XTuoi = (XTuoi, GTuoi, HTuoi, $\leq$) là ĐSGT, với GTuoi = {trẻ, già}, H+Tuoi = {rất, hơn}, H-Tuoi = {khả năng, ít}, với rất > hơn và ít > khả năng. Wtuoi = 0.4, fm(trẻ) = 0.4, fm(già) = 0.6, fm(rất) = 0.3, fm(hơn) = 0.15, fm(khả năng) = 0.25, fm(ít) = 0.3. Ta có fm(rất trẻ) = 0.12, fm(hơn trẻ) = 0.06, fm(ít trẻ) = 0.12, fm(khả năng trẻ) = 0.1. Vì rất trẻ < hơn trẻ < trẻ < khả năng trẻ < ít trẻ nên I(rất trẻ) = [0,0.12], I(hơn trẻ) = [0.12,0.18], I(khả năng trẻ) = [0.18,0.3], I(ít trẻ) = [0.3,0.4].

Nếu chọn ψ1= 100 $\in$ Xtuoi khi đó f(31) = 0.31. Vì 0.31$\in$I(ít trẻ) nên Φ2(0.31) = ít trẻ.

Xác định điều kiện (TUOI =1 có thể trẻ) = 1. Vì v(ít trẻ) $\in$ I(trẻ) và v(có thể trẻ) $\in$ I(trẻ) nên ta có (ít trẻ =1 có thể trẻ) = 1, (có thể trẻ =2 có thể trẻ) = 1, nên suy ra (có thể trẻ =1 có thể trẻ) = 1.

(b). Xét XLuong = (Xluong, Gluong, Hluong, $\leq$) là ĐSGT, với Gluong = {cao, thấp}, H+luong = {rất, hơn}, H-luong = {khả năng, ít}, rất > hơn và ít > khả năng.

Wluong = 0.6, fm(thấp) = 0.6, fm(cao) = 0.4, fm(rất) = 0.25, fm(khá) = 0.25, fm(khả năng) = 0.25, fm(ít) = 0.25. Ta có fm(rất thấp) = 0.15, fm(hơn thấp) = 0.15, fm(ít thấp) = 0.15, fm(khả năng thấp) = 0.15. Vì rất thấp < hơn thấp < thấp < khả năng thấp < ít thấp nên I(rất thấp) = [0,0.15], I(hơn thấp) = [0.15,0.3], I(khả năng thấp) = [0.3,0.45], I(ít thấp) = [0.45,0.6].

Ta có fm(rất cao) = 0.1, fm(khá cao) = 0.1, fm(ít cao) = 0.1, fm(khả năng cao) = 0.1. Vì ít cao < khả năng cao < cao < hơn cao < rất cao nên I(ít cao) = [0.6,0.7], I(khả năng cao) = [0.7,0.8], I(hơn cao) = [0.8,0.9], I(rất cao) = [0.9,1]. Nếu chọn ψ2 = rất rất cao $\in$ Xluong và ψ1=3.000.000, ta có v(rất rất cao) = 0.985, khi đó f(2.800.000) = 0.92 và f (2.000.000) = 0.65 nên Φ2(0.92) = rất cao và Φ2(0.65) = ít cao.

Xác định điều kiện (LUONG =1 cao) = 1. Vì v(rất cao) $\in$ I(cao) và v(ít cao) $\in$ I(cao) nên ta có (rất cao =1 cao) = 1 và (ít cao =1 cao) = 1.

Như vậy, câu truy vấn SQL mờ: select * from Luong_tuoi where (TUOI =1 có thể trẻ) and (LUONG =1 cao) sẽ cho kết quả sau:

Bảng 2. Kết quả thực hiện truy vấn "Nhân viên tuổi có thể trẻ và lương cao"

Oid	SOCM	HOTEN	TUOI	LUONG
1	12345	Phạm Trọng	31	2.800.000
3	34567	Trần Tiến	32	2.000.000

2.5. Đánh giá

Để cung cấp sơ bộ về đánh giá hiệu suất thực hiện xử lí truy vấn đã được đề xuất dựa trên đại số đối tượng mờ [14]. Chúng tôi định nghĩa bộ ba câu truy vấn xử lí mệnh đề điều kiện trích lọc dữ liệu cho hai trường hợp đơn điều kiện, đa điều kiện và thực hiện chúng với dữ liệu có kích thước khác nhau. Tệp dữ liệu rating_final.csv sử dụng cho việc đánh giá có tham chiếu đến web site UCI.

Câu truy vấn thứ nhất xử lí trích lọc dữ liệu cho trường hợp đơn điều kiện. Yêu cầu bộ máy xử lí truy vấn trả lại tất cả các đánh giá cuối cùng có lại xếp hạng là khá tốt. Như vậy câu truy vấn được viết như sau.

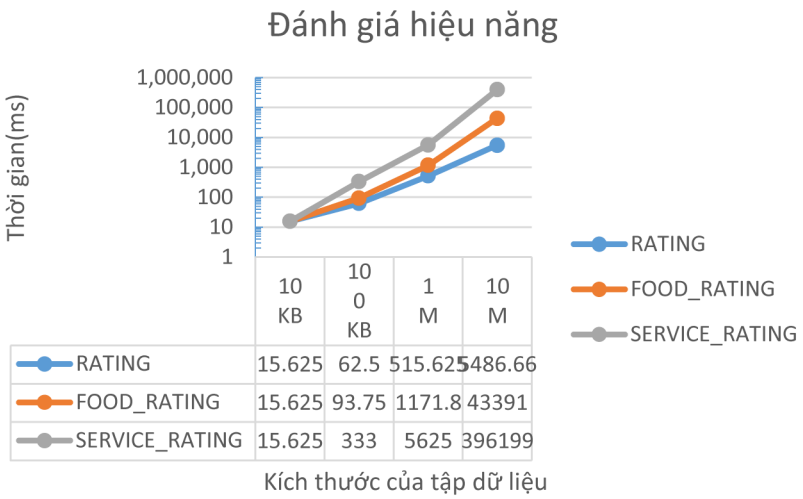
select * from rating_final where (RATING =khá tốt THOLD 7)

Câu truy vấn thứ hai xử lí trích lọc dữ liệu cho trường hợp đơn điều kiện. Yêu cầu bộ máy xử lí truy vấn trả lại tất cả các đánh giá cuối cùng có loại xếp hạng ẩm thực là rất tốt. Như vậy câu truy vấn được viết như sau.

select * from rating_final where (FOOD_RATING = rất tốt THOLD 9)

Câu truy vấn thứ ba xử lí trích lọc dữ liệu cho trường hợp đơn điều kiện. Yêu cầu bộ máy xử lí truy vấn trả lại tất cả các đánh giá cuối cùng có lại xếp hạng là khá tốt và xếp hạng ẩm thực là rất tốt. Như vậy câu truy vấn được viết như sau.

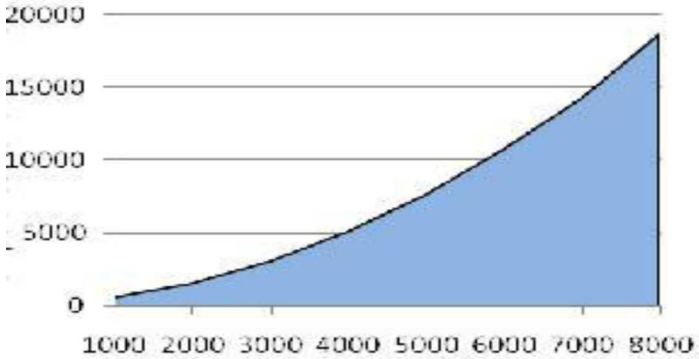
select * from rating_final where (RATING = khá tốt THOLD 7 and SERVICE_RATING=rất tốt THOLD 9)



Hình 1. Hiệu suất truy vấn

Ba câu truy vấn được thực hiện đối với bốn bộ dữ liệu kích thước khác nhau. Truy vấn thứ nhất xử lí trên file có dung lượng là 10KB và chứa 56 đối tượng; thứ hai file có dung lượng file khoảng 100KB và chứa 596 đối tượng; thứ ba file có dung lượng vào khoảng 1MB chứa 6052 đối tượng; thứ tư file có dung lượng vào khoảng 10MB và chứa 60.972 đối tượng. Kiểm thử đánh giá hiệu năng được thực hiện trên một máy tính intel(R) Core(TM) i3 CPU M 370 @ 2.40GHz, bộ nhớ RAM 512MB chạy Windows 7.

Thí nghiệm tương tự đã được thực hiện bằng cách sử dụng đại số gia tử, kết quả đạt được khẳng định rằng hiệu suất của phương pháp này là rất tốt. Như một ví dụ, chúng tôi đánh giá truy vấn theo hướng tiếp cận ĐSGT từ biểu đồ đường các kết quả truy vấn thể hiện trong hình 2. Trong trường hợp này, đồ họa cho thấy một sự gia tăng thời gian tuyến tính đối với các kích thước dữ liệu khác nhau.



Hình 2.

3. Kết luận

Trong bài báo này chúng tôi đề xuất một hướng tiếp cận mới về phương pháp xử lí truy vấn trong cơ sở dữ liệu hướng đối tượng mờ theo cách tiếp cận ĐSGT. Dựa trên mô hình cơ sở dữ liệu hướng đối tượng mờ ĐSGT đã được đề xuất, từ đó, xây dựng các thuật toán xác định giá trị chân lí của điều kiện mờ cho truy vấn vấn dữ liệu. Tiếp theo, chúng tôi đề xuất phương pháp xử lí truy vấn cho điều kiện lọc thông qua thuật toán xử lí đơn điều kiện và đa điều kiện dựa trên thuật toán xác định giá trị chân lí. Cuối cùng, áp dụng các thuật toán vừa xây dựng để thực hiện truy vấn.

TÀI LIỆU THAM KHẢO

[1] Nguyễn Cát Hồ, Lê Xuân Vinh, Nguyễn Công Hào, 2009.*Thống nhất dữ liệu và xây dựng quan hệ tương tự trong cơ sở dữ liệu ngôn ngữ bằng đại số gia tử*. Tạp chí Tin học và Điều khiển học, Tập 25, Số 4, 314-332.

[2] Nguyễn Công Hào, Trương Thị Mỹ Lệ, 2012. *Mô hình CSDL hướng đối tượng mờ dựa trên ngữ nghĩa đại số gia tử*. Tạp chí Tin học và Điều khiển, Tập 28, Số 3, 129-140.

[3] Doan Van Ban, Ho Cam Ha, Vu Duc Quang, "Querying Fuzzy Object-Oriented Data Based on Fuzzy Association Algebra," IEEE Transactions, 2011, pp. 40-47.

[4] Đoàn Văn Ban, Đoàn Văn Thắng, 2010. *Mô hình CSDL hướng đối tượng dựa trên ngữ nghĩa đại số gia tử*. Tạp chí khoa học và công nghệ, Đại học Đà Nẵng, Số 6(41).

[5] Nguyễn Công Hào, Trương Thị Mỹ Lệ, 2012.*Truy vấn trong cơ sở dữ liệu hướng đối tượng mờ dựa trên ngữ nghĩa đại số gia tử*. Tạp chí Khoa học, Đại học Huế, Tập 74b, số 5, 39-5.

[6] Bosc P, Pivert O, 1995 *SQLf: a relational database language for fuzzy query*, IEEEransaction on Fuzzy Systems, Vol. 3, No.1, 1-19.

[7] Cristina-Maria Vladarean, 2006*Extending object-oriented databases for fuzzy information modeling*, S.C. WATERS Romania S.R.L, Romai J., 2, 225-237.

[8] G. Bordogna, G. Pasi, and D. Lucarella, 1999. *A fuzzy object-oriented data model for managing vague and uncertain information*. International Journal of Intelligent Systems, Vol. 14, 623-651

[9] L.A. Zadeh, 1983 *A Computational approach to fuzzy quantifiers in natural languages*. Computers and Mathematics with Applications, Vol 9. No. 1, 149-184.

[10] M. Umano, T. Imada, I. Hatono, and H. Tamura, 1998 *Fuzzy object-oriented databases and implementation of its SQL-type data manipulation language*. In Proceedings of the 7th IEEE International Conference on Fuzzy Systems, Vol. 2, 1344- 1349.

[11] Patrick Bosc, 1995 *Subqueries in SQLf, a Fuzzy Database Query Language*. IEEE, 3636-3641.

[12] Z. M. Ma, 2004. *Advances in Fuzzy Object-Oriented Databases: Modeling and Applications*. Idea Group Publishing.

[13] Z. M. Ma and Li Yan, 2008. *A Literature Overview of Fuzzy Database Models*. Journal Of Information Science And Engineering 24, 189-202.

[14] Trương Ngọc Châu, Nguyễn Tân Thuận, 2013 *Một hướng tiếp cận mới trong đại số đối tượng mờ*. Kỉ yếu, hội thảo Quốc Gia lần thứ XVI một số vấn đề chọn lọc của CNTT & TT - Đà Nẵng, 204-209.

ABSTRACT

Method of multi-condition query processing in a fuzzy oriented object database based on a Hedge algebra-HA approach

In this paper the author proposes the use of the fuzzy query method when processing with a Hedge Algebra-HA approach. Different processing methods are used to determine truth value on which we build fuzzy queries for single and multiple condition cases for processing, with proven results obtained with a fuzzy object oriented database schema.

Keywords: Algebra-HA, Truth value, Single condition, Multi condition, FOQL.