Volume 2, No. 4, July-August 2011

ISSN No. 0976-5697

International Journal of Advanced Research in Computer Science

RESEARCH PAPER

Available Online at www.ijarcs.info

Querying Fuzzy Object-Oriented Data Based On Fuzzy Association Algebra

Prof. Doan Van Ban
Integrated Software Systems
Institute of Information Technology, Academy Science and Technology of VietNam
Hanoi City, VietNam
dvban@ioit.ac.vn

Dr.Ho Cam Ha
Information Technology Faculty
Hanoi National University of Education
Hanoi City, VietNam
hahc@hnue.edu.vn

Vu Duc Quang*
Information Technology Faculty
Quang Nam University
Tamky City, VietNam
vuquangqn79@gmail.com

Abstract: Along with the developing fuzzy object-oriented data models for an expression and a processing of uncertain and imprecise data, the fuzzy data query languages should also be studied, is built to perform data queries in the fuzzy data models flexibly and certainly. Fuzzy association algebra is considered as a query algebra for fuzzy object-oriented data models. In this article, based on possibility distribution and the semantic measure of fuzzy data and equivalence degree of two objects, we propose a more general way to define truth values of fuzzy association patterns which mean the degree of suitability of patterns as answers for the queries. In addition, we present an illustrative example of querying fuzzy object oriented data based on fuzzy association algebra.

Keyword: Truth value, fuzzy association algebra, fuzzy OODB, association pattern, equivalence degree.

I. INTRODUCTION

Fuzzy relational algebra or fuzzy relational calculus are considered as the theoretical foundation for querying data in the fuzzy relational data models. In object-oriented database (OODB), several query algebra for the OODB have been developed on the different bases of data models such as association algebra in [12] for an object-oriented semantic association model, the query algebra for ENCORE [13], …. which are the mathematical foundations for their models. In this article, by considering three levels of fuzziness into object classes, the structure of class level, the object instance level and attribute value level, we show how to determine degree of the relationship between objects. Thereby, determining the truth value for the fuzzy association pattern.

The article is organized as follows. The basic knowledge is presented in section 2. The fuzzy object-oriented data model is give in section 3. Section 4 introduces to fuzzy association algebra. The querying a fuzzy OODB is mentioned in section 5. And the final section is the conclusion.

II. BASIC KNOWLEDGE

Fuzzy data can be described with fuzzy sets by Zadeh [13]. Let *U* be a universe of discourse. Then a fuzzy value on *U* is characterized by a fuzzy set *F* in *U*. Here a membership function $\mu_F : U \to [0,1]$ is needed to define the fuzzy set *F*, in which $\mu_F(u)$, for each $u \in U$, denotes the degree of membership of *u* in the fuzzy set *F*. The fuzzy set *F* is hereby represented as follows.

$$F = \{\mu_F(u_1)/u_1, \mu_F(u_2)/u_2, ..., \mu_F(u_n)/u_n\}$$

When the $\mu_F(u)$ above is explained to be a measure of the possibility that a variable *X* has the value *u*, where *X* takes values in *U*, a fuzzy value can be described by a possibility distribution.

$$\pi_X = \{\pi_X(u_1)/u_1, \pi_X(u_2)/u_2, ..., \pi_X(u_n)/u_n\}$$

Here $\pi_X(u_i), u_i \in U$ denotes the possibility that *X* takes value u_i. Let π_X, F be the possibility distribution representation and the fuzzy set representation for a fuzzy value, respectively. It is apparent that is $\pi_X = F$ true.

The semantics of a fuzzy data represented by possibility distribution corresponds to the semantic space. The semantic relationship between two fuzzy data is then described by the relationship between their semantic spaces [15]. The semantic inclusion degree is used to measure semantic inclusion and further semantic equivalence of fuzzy data.

Let π_A and π_B be two fuzzy data, and their semantic spaces be $SS(\pi_A)$ and $SS(\pi_B)$ respectively. Let SID(π_A, π_B) denotes the degree that π_A semantically includes π_B. Then

$$SID(\pi_A, \pi_B) = (SS(\pi_A) \cap SS(\pi_B))/SS(\pi_B)$$

The meaning of $SID(\pi_A, \pi_B)$ is the percentage of the semantic space of π_A which is wholly included in the semantic space of π_B.

Let *U* = $\{u_1, u_2, \ldots, u_n\}$ be the universe of discourse. Let π_A and π_B be two fuzzy data on *U* based on possibility distribution. The degree that π_A semantically includes π_B is defined as follows

$$SID(\pi_A,\pi_B)=\sum_{i=1}^{n}\min_{u_i\in U}(\pi_B(u_i),\pi_A(u_i))\Big/\sum_{i=1}^{n}\pi_B(u_i)$$

Let $U = \{u_1, u_2, ..., u_n\}$ be the universe of discourse. Let Res be a resemblance relation on domain U, α for $0 \le \alpha \le 1$ be a threshold corresponding to Res. The degree that π_A semantically includes π_B is defined as follows

$$SID_\alpha(\pi_A,\pi_B)=\sum_{i=1}^{n}\min_{u_i,u_j\in U;\mathrm{Res}(u_i,u_j)\ge\alpha}(\pi_B(u_i),\pi_A(u_j))\Big/\sum_{i=1}^{n}\pi_B(u_i)$$

Let π_A and π_B be two fuzzy data. Let $SE(\pi_A,\pi_B)$ denote the degree that π_A and π_B are equivalent to each other.

$$SE(\pi_A,\pi_B)=\min\{SID(\pi_A,\pi_B),SID(\pi_B,\pi_A)\}$$

III. FUZZY OBJECT-ORIENTED DATA MODEL

A fuzzy object-oriented data model is defined as an enhanced the object-oriented data model by replacing objects with fuzzy objects, classes with fuzzy classes and associations with fuzzy associations on the essential characteristics of object-oriented paradigm.

A. Fuzzy Object and Fuzzy Class

Some objects having the same structural and behavioral properties are grouped together to form an object class. Object classes are categories into primitive classes and non-primitive classes. A primitive class represents a class of self-named objects serving as a domain for defining other object classes, such as a class of symbols or numerical values. A non-primitive class represents a set of objects, each of which is an assigned OID and its data are explicitly entered in a database by the user. The structural properties of an object class are represented by descriptive data which define states of objects and association data which specify relationships between its objects of some related classes.

In fuzzy object-oriented database, there are three levels of fuzziness into object classes. These three levels of fuzziness are defined as follows

- The first level (the structure of class level) in which attributes of class may be fuzzy, i.e., they have a membership degree to their class.

- The second level (object instance level) related to the fuzzy occurrences of objects. Even though the structure of an entity is crisp, it is possible that an instance of the class belongs to the class with degree of membership.

- The third (attribute value level) concerns in the values of attributes of the instance of the class. An attribute in a class defines a value domain. When this domain is a fuzzy subset or a set of fuzzy subsets, the fuzziness of an attribute value appears.

An fuzzy object of the fuzzy class C_i is represented by a pair $(o_{ij},\mu_{C_i}(o_{ij}))$, where o_{ij} is j[th] object in fuzzy class C_i, $\mu_{C_i}(o_{ij})$ is a degree of membership of object o_{ij} in fuzzy class C_i, $0<\mu_{C_i}(o_{ij})\le 1$.

For extensional database, a fuzzy class C_i can be defined as the following: $C_i=\{(o_{ij},\mu_{C_i}(o_{ij}))\}$

B. Equivalence of Two Objects Within a Class

In the context of the fuzzy object-oriented databases, objects may be fuzzy. In order to denote the possibility degree that two fuzzy objects o_1 and o_2 refer to the same real world object, the notion of equivalence degree of o_1 and o_2 is in [0, 1]. Formally, let the equivalence degree of o_1 and o_2 be $\mu(o_1,o_2)$. In the following, we investigate how to calculate $\mu(o_1,o_2)$ based on the relationships between two classes that o_1 and o_2 belong to.

a. Equivalence of Two Objects within a Class

Let C be a class with attributes $\{A_1, A_2, ..., A_n\}$ and o_1 and o_2 be two objects belong to the same class C. Also assume that $o_1.A_i$ and $o_2.A_i$ are allowed to be fuzzy values. Then o_1 and o_2 are fuzzy objects.

In order to compare objects o_1 and o_2 and calculate $\mu(o_1,o_2)$, we first need to compare their corresponding attributes. For each pair of values of the same attribute (say A_i ($1\le i \le n$), we obtain their equivalence degree, denoted $\mu_{A_i}(o_1,o_2)$,($0 \le \mu_{A_i}(o_1,o_2) \le 1$). Here

$\mu_{A_i}(o_1,o_2)$ = SE($o_1.A_i$, $o_2.A_i$), where *SE($o_1.A_i$, $o_2.A_i$)* is computed in a similar way to [4].

On the basis, we then combine the equivalence degrees that are obtained in the values of their attributes.

Considering that different attributes play different roles in object comparison, and some may be dominant and some may be non-dominant, a weight w_i is assigned to each attribute of C according to its importance such that $0 < w_i \le 1$.

Formally, the equivalence degree of o_1 and o_2, denoted $\mu(o_1,o_2)$, is expressed as follows.

$$\mu(o_1,o_2)=\frac{\sum_{i=1}^{n}\{SE(o_1.A_i,o_2.A_i)\times w(A_i)\}}{\sum_{i=1}^{n}w(A_i)} \qquad (1)$$

If $\mu(o_1,o_2)$ = 0, o_1 and o_2 do not refer to the same real- world object; if $\mu(o_1,o_2)$ = 1, o_1 and o_2 refer to the same real world object. If 0 < $\mu(o_1,o_2)$ < 1, o_1 and o_2 refer to the same real-world object to some extent.

b. Equivalence of Two Objects of Superclasse and Subclass

Let C be a class with attributes $\{A_1, A_2, ..., A_n\}$ and C' be a subclass of C with attributes $\{A_1, A_2, ..., A_k, A'_{k+i}, ..., A'_m, A_{m+1},..., A_n\}$. Here attributes A'_{k+i}, ..., A'_m are overridden from A_{k+1}, ..., A_m and attributes A_{m+1},..., A_n are special. Let o_1 be a fuzzy object of C_1 and o_2 be a fuzzy object of C_2.

In order to compare objects o_1 of C and o_2 of C' and calculate $\mu(o_1,o_2)$, generally speaking, we also need to compare their corresponding attributes and then combine their equivalence degrees together with the consideration of attribute weights. For each pair of values of the same attribute (say A_i ($1\le i \le k$)), we obtain their equivalence degree, denoted $\mu_{A_i}(o_1,o_2)$ $(0\le\mu_{A_i}(o_1,o_2)\le 1)$. For each pair of values of the attribute and its overridden attribute (say A_j and A'_j ($k + 1 \le j \le m$)), we obtain their equivalence degree, denoted $\mu_{A_j}(o_1,o_2)$ $(0\le\mu_{A_j}(o_1,o_2)\le 1)$. Here

$\mu_{A_j}(o_1, o_2) =$ SE(o_1.A_i, o_2.A_i), where *SE(o_1.A_i, o_2.A_i)* is computed in a similar way in [4].

The values of the attributes A_p (m+1 $\le p \le$ n) in o_2 do not need to be considered because they have no counterpart in o_1. Finally we have:

$$\mu(o_1,o_2) = \frac{\sum_{i=1}^{k} SE(o_1.A_i, o_2.A_i) \times w(A_i) + \sum_{j=k+1}^{m} SE(o_1.A_i, o_2.A_j^{'}) \times w(A_j^{'})}{\sum_{i=1}^{k} w(A_i) + \sum_{j=k+1}^{m} w(A_j^{'})} \quad (2)$$

C. Fuzzy Association

A fuzzy association is a relationship with fuzzy semantics of two classes. For examples, '*may be* 'association,' fluent associations are fuzzy associations. At an object level, a fuzzy association is defined as an association having a degree of relationship $R\ (0 \le R \le 1)$ between two objects. A fuzzy association (FA) is defined as the followings [11]:

$$FA_{im}(k) = \{((o_{ij}, o_{mn}), R(k)(o_{ij}, o_{mn}))\}$$

Such that

$$\forall i,n\ \ (o_{ij}, \mu_{C_i}(o_{ij})) \in C_i, (o_{mn}, \mu_{C_m}(o_{mn})) \in C_m)$$

and $R(k)(o_{ij}, o_{mn}) \in [0,1]$

where $R(k)(o_{ij}, o_{mn}) \in [0,1]$ is a relationship value between two objects o and o_{nm}. k is an identifier of association.

For fuzzy associations, imprecise and vague linguistic terms such as 'young' can be represented, for examples, 'Young People" association can be expressed as a set of fuzzy associations between object o_{ij} (an object in class Person) and an integer in the domain of the age such as $\{((o_{ij},25),0.9),((o_{ij},26),0.7),((o_{ij},28),0.6),((o_{ij},30),0.5),((o_{ij},31),0.3)\}$

This possibility distribution can be decomposed with 5 fuzzy associations, such as 'one person relates to object o_{ij} is 25 years old' with relationship value 0.9, 'one person relates to object o_{ij} is 26 years old' with relationship value 0.7 or as so on. The association with $R(k)(o_{ij}, o_{mn}) = 0$ is not needed to be represented explicitly in database. A conventional association of OO database can be extend to an association having value 1.

D. Representations of fuzzy OO Databases

The fuzzy OODB is represented by two extended fuzzy graphs for both extensional and intensional databases. A fuzzy schema graph represents an intensional database and a fuzzy object graph represents an extensional database.

Fuzzy Schema Graph (FSG): The FSG is defined as $FSG(C, FA)$, where $C = \{C_i, \forall i\}$, is a set of vertices representing fuzzy classes and $FA = \{FA_{im}(k) \mid \forall i, m\}$ is a set of edges of fuzzy associations between two fuzzy classes.

Fuzzy Object Graph (FOG): The FOG is defined as $FOG(FO, FE)$, where $FO = \{(o_{ij}, \mu_{C_i}(o_{ij}))\}$ is a set of vertices with truth values representing fuzzy object instances, and $FE = \{((k): o_{ij} \_ o_{mn}, R(o_{ij}, o_{mn}))\}$ is a set of edges with truth values representing fuzzy associations between two fuzzy object instances.

When one object instance is connected with another in FOG, a regular-edge (solid line) is drawn between the corresponding vertices as $o_{ij} \xrightarrow{R(o_{ij}, o_{mn})} o_{mn}$ which specifies that j[th] object instance in fuzzy class C_i is related to n[th] object instance in fuzzy class C_m through the association of classes C_i and C_m with relationship value $R(o_{ij}, o_{mn})$. If two object instance o_{ij}, o_{mn} is not connected in the FOG but their classes C_i and C_m in the corresponding FSG is directly connected, a complement-edge (dotted line) is drawn between them and is denoted by $o_{ij} \overset{R(o_{ij}, o_{mn})}{\dashrightarrow} o_{mn}$

In general OO model, an object may participate in several classes (e.g., in generalization hierarchy). Its representation in a class is called an object instance. Since in this article, "object" and "object instance" can be used interchangeably without any ambiguity.

For example, a FSG and a FOG for a human resource database are illustrated in Fig. 1 and Fig. 2.

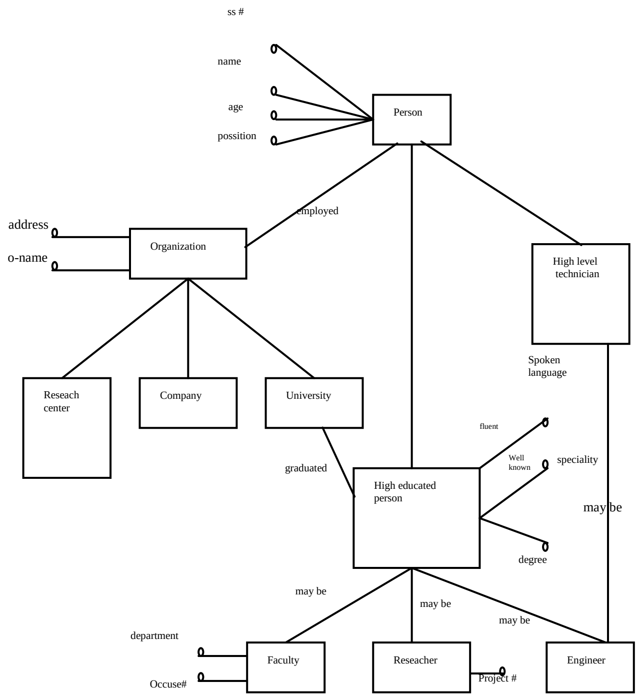


Figure. 1 A Fuzzy schema graph of human resource database

IV. FUZZY ASSOCIATION ALGEBRA

Fuzzy association algebra is represented following as some fuzzy enhancement of Association algebra [12]. In the Fuzzy association algebra, all operators are defined to operate on fuzzy association patterns of homogeneous as well as heterogeneous structures.

A. Fuzzy Association Pattern

In order to represent a fuzzy association pattern, we use a pair of a pattern and a truth value of the pattern, algebraically represented by $(P, T(P))$. The truth value presents how compatible to a query a pattern is. As the value is closer to 1, the pattern is most likely a true answer pattern.

Fuzzy association patterns can be represented by representation graphical or representation algebraic. Five primitive patterns and a complex pattern are defined in [11] as follows:

Fuzzy Inner Association Pattern is a single vertex (or fuzzy object) in FOG represented by $((a_i),T(a_i))$, where truth value $T(a_i)$ is defined by membership degree of a_i belong to class C, $T(a_i)=\mu_C(a_i)$

representation graphical *representation algebraic*

$a_i(T)$ (a_i,T)

Fuzzy Inter Association Pattern is composed of two vertices and an edge between two vertices represented by $((a_i,b_j),T(a_ib_j))$, where truth value $T(a_ib_j)$ is defined followings as $T(a_ib_j) = Min(T(a_i),T(b_j),R(a_ib_j))$, where $T(a_i),T(b_j)$ are truth values of two fuzzy associations a_i and b_j, respectively. $R(a_ib_j)$ is relationship value between two fuzzy associations a_i and b_j.

representation graphical *representation algebraic*

a_i —— b_j (T) $((a_ib_j),T)$

Fuzzy Complement Association Pattern is composed of two vertices and a complement edge between two vertices represented by $((a_ib_j)',T(a_ib_j)')$ where $T(a_ib_j)' = Min(T(a_i),T(b_j),1-R(a_ib_j))$, $R(a_ib_j)'$ is a complement relationship of $R(a_ib_j)$. Thus $R(a_ib_j)'=1-R(a_ib_j)$.

representation graphical *representation algebraic*

a_i - - - - - - b_j (T) $((a_ib_j)',T)$

Fuzzy Derived Inter Association Pattern is composed of two nonadjacent vertices and a derived edge between two vertices represented by $((a_i,b_j)\sim,T(a_ib_j)\sim)$. It is possible that there exists many sequences between a_i, and b_j. Let $c_{1k},c_{2k},c_{3k},...,c_{nk}$ be kth sequence of adjacent vertices between a_i and b_j, $T_k(a_ib_j)\sim$ be a truth value of the sequence. Then $T_k(a_ib_j)\sim =$

$Min(T(a_i,c_{1k}),T(c_{1k},c_{2k}),T(c_{2k},c_{3k}),...,T(c_{nk},b_j))$

$T(a_ib_j)\sim$ is the supreme of $T_k(a_ib_j)\sim$, for all k, i.e. $T(a_ib_j)\sim = Sup_k(T_k(a_ib_j)\sim)$.

representation graphical *representation algebraic*

a_i b_j (T) $((a_ib_j)\sim,T)$

Fuzzy *Derived Complement Association Pattern* is composed of two nonadjacent vertices and a derived complement edge between two vertices represented by $((a_ib_j)\sim',T(a_ib_j)\sim')$, where $T(a_ib_j)\sim' = Sup_k(T_k(a_ib_j)\sim')$ for all *k*.

representation graphical *representation algebraic*

a_i b_j (T) $((a_ib_j)\sim',T)$

Complex pattern is generated by composing of the primitive patterns:

representation graphical *representation algebraic*

a_i b_j c_l (T) $((a_ib_j,b_jc_l,b_jd_k),T)$

d_k

Fuzzy object-oriented data model exists many difference relationships such as Object-class relationship, Aggregation relationship, Generalization relationship, etc. Determining the degree of relationship between objects in the relationships allows us to determine the truth value of the fuzzy association patterns when performing the query data based on fuzzy association algebra.

B. Truth Value of Fuzzy Association Pattern

a. Fuzzy Object-Class Relationship

In the OODB, determining if an object belongs to a class depends on if its attribute values are respectively included in the corresponding attribute domains of the class. Similarly, the membership degree of an object to the class in a fuzzy object-class relationship is calculated using the inclusion degree of object values with respect to the class domains, and the weight of attributes.

Let *C* be a class with attributes $\{A_1,A_2,...,A_n\}$, o be an object on attribute set $\{A_1,A_2,...,A_n\}$, and $o.A_i$ denotes the attribute value of o on *A*.

In *C*, each attribute A_i is connected with a domain denoted $dom(A_i)$. In the OODB, $dom(A_i)$ is a set of crisp values and may be a set of fuzzy subsets in fuzzy databases. Therefore, in a uniform OODB for crisp and fuzzy information modeling, $dom(A_i)$ should be the union of these two components, $dom(A_i)=cdom(A_i)\cup fdom(A_i)$, where $cdom(A_i)$ and $fdom(A_i)$ respectively denote the sets of crisp values and fuzzy subsets.

The inclusion degree of $o.A_i$ with respect to $dom(A_i)$ is denoted *ID(dom(A$_i$), o.A$_i$)* and the evaluation of *ID(dom(A$_i$), o.A$_i$)* bases on two cases followings:

Case 1: o.A$_i$ is a fuzzy value. Let $fdom(A_i)=\{f_1,f_2,...,f_m\}$, where $f_i\,(1\le i\le m)$ is a fuzzy value, and $cdom(A_i)=\{c_1,c_2,...,c_k\}$, where c_l (1 < l < k) is a crisp value. Then $ID(dom(A_i),o(A_i))$

$$=\max(ID(cdom(A_i),o.A_i),ID(fdom(A_i),o.A_i)) \quad (3)$$
$$=\max(SID(\{1.0/c_1,1.0/c_2,...,1.0/c_k\},o.A_i),$$
$$\max(SID(f_i,o.A_i)))$$

Case 2: $o.A_i$ is a crisp value. Then

$ID(dom(A_i),o(A_i))=1$ if $o(A_i)\in cdom(A_i)$

Else

$ID(dom(A_i),o(A_i))=ID(fdom(A_i),\{1.0/o.A_i\}$

The membership degree of the object *o* to the class *C* is calculated as follows, where $w(A_i(C))$ denotes the weight of attribute A_i to class *C*.

$$\mu_C(o)=\frac{\sum_{i=1}^{n}ID(dom(A_i),o.A_i))\times w(A_i(C))}{\sum_{i=1}^{n}w(A_i(C))} \quad (4)$$

Truth value of Fuzzy Inner Association Pattern a_i is defined as follows:

- If a_i is an object instance of a primitive class then $T(a_i)$ is the inclusion degree of a_i with respect to its value domain. $T(a_i)=$ *ID(dom(a$_i$), a$_i$)*, where *dom(a$_i$)* denotes value domain of the attribute which has value a_i
- If a_i is an object instance of a non-primitive class *C* then $T(a_i)=\mu_C(a_i)$ (calculating $\mu_C(a_i)$ in a similar way to eq. (4))

For example, consider fuzzy class Young students with attributes Age and Height and an object *o*. Weight of Age and weight of Height are 0.9 and 0.2, respectively. Assume $cdom(Age)=\{5-20\}$, $fdom(Age)=\{\{1.0/20,$

1.0/21, 0.7/22, 0.5/23}, {0.4/22, 0.6/23, 0.8/24, 1.0/25, 0.9/26, 0.8/27, 0.6/28}, {0.6/27, 0.8/28, 0.9/29, 1.0/30, 0.9/31, 0.6/32, 0.4/33, 0.2/34}} and $dom(Height) = cdom(Height) = [60,\ 210]$. Let $o(Age)$ = {0.6/25, 0.8/26, 1.0/27, 0.9/28, 0.7/29, 0.5/30, 0.3/31}, $o(Height)$ = 182. According to the definition above, we have

$$ID(dom(Height), o(Height)) = 1$$

$$ID(cdom(Age), o(Age)) = SID(1.0/5,\ 1.0/6,\ ...,\ 1.0/19,\ 1.0/20\},\ o(Age)) = 0$$

$$ID(fdom(Age), o(Age)) = \max(SID(\{1.0/20, 1.0/21, 0.7/22, 0.5/23\}, o(Age)),\ SID(\{0.4/22, 0.6/23, 0.8/24, 1.0/25, 0.9/26, 0.8/27, 0.6/28\}, o(Age)),\ SID(\{0.6/27, 0.8/28, 0.9/29, 1.0/30, 0.9/31, 0.6/32, 0.4/33, 0.2/34\}, o(Age))) = \max(0, 0.58, 0.60) = 0.60$$

Therefore, $ID(dom(Age), o(Age)) =$

$$\max(ID(cdom(Age), o(Age), ID(fdom(Age), o(Age))) = 0.60$$

Therefore, we have:

$$T(o) = \mu_{Young\_Students}(o)$$

$$= (0.9 \times 0.6 + 0.2 \times 1.0)/(0.9 + 0.2) = 0.67$$

Truth value of *Fuzzy Inter Association Pattern* $(a_i b_j)$, where a_i is an object instance of a primitive class, b_j is an object instance of a non-primitive class C, computed as follows

$T(a_i b_j) = Min(T(a_i), T(b_j), R(a_i b_j))$ where $R(a_i b_j)$ is degree of membership of attribute has value a_i to the class C.

b. Fuzzy Generalization Relationship

A new class, called subclass, is produced from another class, called superclass, by means of method of inheriting. Because a subclass is the specialization of the superclass, any one object belonging to the subclass must belong to the superclass.

In fuzzy OODB, classes may be fuzzy. A class produced from a fuzzy class must be fuzzy. The subclass-superclass relationship is also fuzzy relationship. Therefore, the relationship between an object of superclass and an object of subclass is a fuzzy. Degree of relationship between them is the possibility degree that the object of subclass is the specialization of an object of superclass. The possibility degree can also be determined based on the degree of equivalence between the two objects and the degree of membership of each object to the its class.

Therefore, we have truth value of *Fuzzy Inter Association Pattern* $(a_i b_j)$ is computed as follows

$$T(a_i b_j) = Min(T(a_i), T(b_j), R(a_i b_j))$$

where a_i, b_j are the objects of superclass *C* and subclass *C'* respectively. $R(a_i b_j)$ is computed as follows.

$R(a_i b_j) = \min(SE(a_i, b_j) \times \mu_C(a_i),\ SE(a_i, b_j) \times \mu_{C'}(b_i))$, SE($a_i$, b_j) is competed in similar way to eq. 2

Similarly, the truth of *Fuzzy Complement Association Pattern* $(a_i b_j)\sim$ is computed as follows

$$T(a_i b_j)' = Min(T(a_i), T(b_j), 1 - R(a_i b_j))$$

c. Fuzzy Aggregation Relationship

An aggregation captures a whole-part relationship between an aggregate and a constituent part. These constituent parts can exist independently. Therefore, every instance of an aggregate can be projected into a set of instances of constituent parts. Let A be an aggregation of constituent parts B_1, B_2, …, B_n.

For $o \in$ A, the projection of o to B_i is denoted by $o\downarrow_{B_i}$. Then we have $o\downarrow_{B_1} \in B_1, o\downarrow_{B_2} \in B_2, \ldots, o\downarrow_{B_n} \in B_n$

A class aggregated from fuzzy constituent parts must be fuzzy. The relationship between an aggregate class and a constituent class is called fuzzy aggregation relationship.

Let *A* be a fuzzy aggregation of fuzzy class sets B_1, B_2, …, B_n. Here , B_i has a set of attributes $A_i^1, A_i^2, \ldots A_i^{m_i}$ with weights respectively are $w_i^1, w_i^2, \ldots, w_i^{m_i}$. Let *o* be an object of class *A*, the projection of *o* to B_i is denoted by $o\downarrow_{B_i}$. Then we have $o\downarrow_{B_1} \in B_1, o\downarrow_{B_2} \in B_2, \ldots, o\downarrow_{B_n} \in B_n$

The membership degree of the object *o* to the class *A* is calculated follows, where $w_i^{j_i}$ denotes the weight of attribute $A_i^{j_i}$ to class B_i.

$$\mu_A(o) = \frac{\sum_{i=1}^{n}\sum_{j=1}^{m_i} ID(dom(A_i^{j_i}), (o \downarrow B_i).A_i^{j_i}) \times w_i^{j_i}}{\sum_{i=1}^{n}\sum_{j=1}^{m_i} w_i^{j_i}} \quad (5)$$

Where $(o \downarrow B_i).A_i^{j_i}$ is a value of $A_i^{j_i}$ of object $o\downarrow_{B_i}$. $ID(dom(A_i^{j_i}), (o \downarrow B_i).A_i^{j_i})$ is the inclusion degree of attribute $A_i^{j_i}$ of object $o\downarrow_{B_i}$ with respect to value domain $dom(A_i^{j_i})$.

Truth value of *Fuzzy Inner Association Pattern* a_i is defined as follows:

$T(a_i) = \mu_A(a_i)$, here calculating $\mu_C(a_i)$ in a similar way to eq. (5), a_i is an object of *A*.

Degree of relationship between an object of aggregate class and an object of constituent class is the possibility degree that the projection of object of aggregate class to constituent class is an object of constituent class. The possibility degree can also be determined based on the degree of equivalence between the two objects and the degree of membership of object of aggregate class and an object of constituent class to the their classes.

Therefore, truth value of *Fuzzy Inter Association Pattern* $(a_i b_j)$ is computed as follows

$$T(a_i b_j) = Min(T(a_i), T(b_j), R(a_i b_j))$$

where a_i, b_j are the objects of classes *A* and B_i,

$$R(a_i b_j) = \min(SE((a_i \downarrow B_i), b_j) \times \mu_A(a_i), SE((a_i \downarrow B_i), b_j) \times \mu_{B_i}(b_j))$$

$SE((a_i \downarrow B_i), b_j)$ is computed in similar way to eq. 1

Similarly, the truth value of *Fuzzy Complement Association Pattern* $(a_i b_j)'$ is computed as follows

$$T(a_i b_j)' = Min(T(a_i), T(b_j), R(a_i b_j)')$$

where $R(a_i b_j)' = 1 - R(a_i b_j)$

d. Fuzzy Association Relationship

For three levels of fuzziness within the above object class, an association relationship between two object classes is fuzzy because of some following reasons.

- First, an association relationship fuzzily exists in two associated classes C_1, C_2, this association relationship occurs with a degree of possibility $\chi\ (0 \le \chi \le 1)$.

- Second, an object belongs to associated classes with membership degree.

The possibility that the two fuzziness mentioned may occur in an association relationship simultaneously.

For example, an association relationship between Person and Organization class (in F.g 1) is represented by fuzzy UML as follows:

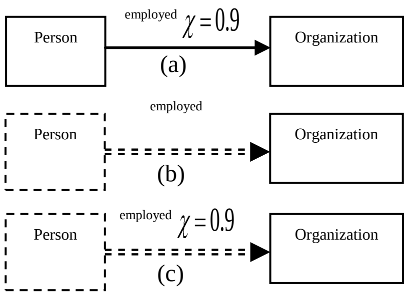


Figure. 2 An Fuzzy Association relationship

Figure 2 shows fuzzy association relationships. In part (a), it is uncertain if the person is employed in the organization, and the possibility is 0.9. Classes Person and Organization have the association relationship *'employed'* with an 0.8 membership degree. In part (b), it is certain that the person is employed in the organization, and the possibility is 1.0. Classes Person and Organization have an association relationship '*employed'* with 1.0 membership degree. But at the level of object instances, there exists the possibility that the instances of class Person may or may not have the association relationship '*employed*'. In part (c), two kinds of fuzzy association relationships in parts (a) and (b) arise simultaneously.

Let two fuzzy classes C_1, C_2. The object o_1 of C_1 is one with membership degree $\mu_{C_1}(o_1)$, the object o_2 of C_2 is one with membership degree $\mu_{C_2}(o_2)$. Assume that association relationship between C_1 and C_2 occurs with a degree of possibility χ. Degree of association relationship between o_1 and o_2 is denoted $\mu(Ass(o_1,o_2))$, is defined as follows:

$$\mu(Ass(o_1,o_2)) = \min(\mu_{C_1}(o_1), \mu_{C_2}(o_2), \chi).$$

Therefore, truth value of *Fuzzy Inter Association Pattern* $(a_i b_j)$ is computed as follows

$$T(a_i b_j) = Min(T(a_i), T(b_j), R(a_i b_j))$$

where a_i, b_j are the objects of associated classes C_1 and C_2, $R(a_i b_j) = \mu(Ass(a_i, b_j))$. Therefore,

$$T(a_i b_j) = Min(\mu_{C_1}(a_i), \mu_{C_2}(b_j), \chi)$$

Similarly, the truth of *Fuzzy Complement Association Pattern* $(a_i b_j)'$ is computed as follows

$$T(a_i b_j)' = Min(T(a_i), T(b_j), R(a_i b_j)')$$

where $R(a_i b_j)' = 1 - R(a_i b_j)$.

V. QUERYING A FUZZY OODB

In order to query fuzzy data in fuzzy OODB, the first we introduced some fuzzy association operators in [11]. The notations that will be used in the definition of operators. A, B,..., K denote fuzzy classes. C_i denotes a variable for a fuzzy class, which can be explicitly named by an attribute. $[R(C_i, C_j)]$ denotes the fuzzy association between two classes. {[R (FC, FC,)]} denotes the fuzzy set of fuzzy inter patterns having the association denoted by $[R(C_i, C_j)]$. a_i denotes i^{th} fuzzy pattern of fuzzy class A. $P(a_i)$ denotes the pattern of fuzzy pattern a_i. $T(a_i)$ denotes the truth value of fuzzy pattern a_i. @ denotes a fuzzy inner pattern variable. $\alpha, \beta, \gamma, \ldots$ Denote sets of fuzzy associations. α_i Denotes i^{th} fuzzy pattern of fuzzy association set. $P(\alpha_i)$ Denotes a set of primitive patterns in α. {W}, {X}, {Y}, ... denote sets of fuzzy classes. $\alpha_{\{X\}}$ Represents fuzzy association set α which has fuzzy inner patterns from the class in {X}.

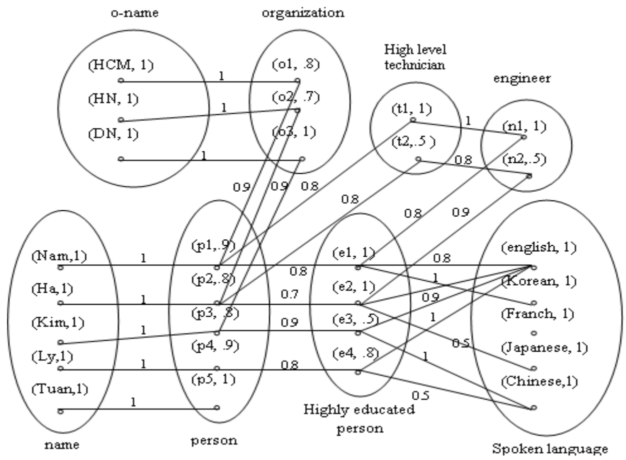


Figure. 3 A Fuzzy object graph of human resource database

A. *Operators:*

a. *Associate (*)*

$$\gamma = \alpha * [\mathrm{R(A, B)}]\, \beta = \{(\mathrm{P}(\gamma_k),\ \mathrm{T}(\gamma_k)) \mid \mathrm{P}(\gamma_k) = (\mathrm{P}(\alpha_i), \mathrm{P}(\beta_j), a_m b_n),$$
$$T(\gamma_k) = \min(T(\alpha_i), T(\beta_j), T(a_m b_n)),$$
$$where\ \ a_m \in P(\alpha_i)\ and\ b_n \in P(\beta_j), T(\gamma_k) > 0\}$$

b. *Complement (|)*

$$\gamma = \alpha \,|\, [\mathrm{R(A, B)}]\, \beta = \{(\mathrm{P}(\gamma_k),\ \mathrm{T}(\gamma_k)) \mid \mathrm{P}(\gamma_k) = (\mathrm{P}(\alpha_i), \mathrm{P}(\beta_j), a_m b_n{}'),$$
$$T(\gamma_k) = \min(T(\alpha_i), T(\beta_j), T(a_m b_n)'),$$
$$where\ \ a_m \in P(\alpha_i)\ and\ b_n \in P(\beta_j), T(\gamma_k) > 0\}$$

c. *Select (σ)*

$$\gamma = \sigma(\alpha)[P\wedge, C\wedge] = \{(\mathrm{P}(\gamma_k),\ \mathrm{T}(\gamma_k)) \mid \mathrm{P}(\gamma_k) = \mathrm{P}(\alpha_i),$$
$$T(\gamma_k) = 1, where\ \ P\wedge(P(\alpha_i)) = true,$$
$$C\wedge(P(\alpha_i)) = true\}$$

When only P^ is specified

$$\gamma = \sigma(\alpha)[P\wedge] = \{(\mathrm{P}(\gamma_k),\ \mathrm{T}(\gamma_k)) \mid \mathrm{P}(\gamma_k) = \mathrm{P}(\alpha_i),$$
$$T(\gamma_k) = T(\alpha_i),$$
$$where\ \ P\wedge(P(\alpha_i)) = true, T(\alpha_i) > 0\}$$

When only C^ is specified,

$$\gamma = \sigma(\alpha)[C\wedge] = \{(\mathrm{P}(\gamma_k),\ \mathrm{T}(\gamma_k)) \mid \mathrm{P}(\gamma_k) = \mathrm{P}(\alpha_i),$$
$$T(\gamma_k) = 1,$$
$$C\wedge(T(\alpha_i)) = true\}$$

The predicate C^ = $T\theta c$. T is a variable of truth values. θ is one of comparison operators $=, >, <, \geq, \leq, \neq$ and c is a value in [0,l]. The predicate $P\wedge = T_1 \theta_1 T_2 \theta_2 \ldots \theta_{n-1} T_n$, $\theta_i\ (i = 1,2,\ldots, n-1)$ is a boolean operator (and or or). $T_i\ (i = 1,2,\ldots, n)$ is a term.

d. *Project (π)*

$$\pi(\alpha)[E\wedge, D\wedge] = \{(\mathrm{P}(\gamma_k),\ \mathrm{T}(\gamma_k)) \mid \mathrm{P}(\gamma_k) = \mathrm{P}(\alpha_i^s),$$
$$T(\gamma_k) = Max(T(\alpha_i), T(\alpha_j), \ldots, T(\alpha_n))$$
$$where\ P(\alpha_i), P(\alpha_j), \ldots, P(\alpha_n) \text{ are projected to a same pattern } \mathrm{P}(\gamma_k)\}$$

where $P(\alpha_i^s)$ is a subpattern of $P(\alpha_i)$ which contains subpatterns as subexpressions of E^ and the paths derived expressions of D^ between two subpatterns.

e. Intersect (•)

$$\gamma = \alpha_{\{X\}} \bullet \{W\} \beta_{\{Y\}} = \{(P(\gamma_k), T(\gamma_k)) \mid P(\gamma_k) = (P(\alpha_i), P(\beta_j)), \; T(\gamma_k) = Min(T(\alpha_i), T(\beta_j))$$

$$\forall P(@) \text{ such that } C_n \in \{W\}, P(@) \in P(C_n) \wedge P(@) \in P(\alpha_i), P(@) \in P(\beta_j)$$

$$\forall P(@) \text{ such that } C_n \in \{W\}, P(@) \in P(C_n) \wedge P(@) \in P(\beta_j), P(@) \in P(\alpha_i) \}$$

f. Union (+)

$$\gamma = \alpha + \beta = \{(P(\gamma_k), T(\gamma_k)) \mid \gamma_k = \alpha_i \vee \gamma_k = \beta_j \vee P(\gamma_k) = P(\alpha_i) \vee P(\gamma_k) = P(\beta_j) \text{ and } (\gamma_k) = Max(T(\alpha_i), T(\beta_j)) \text{ when } P(\alpha_i) = P(\beta_j)\}$$

g. Difference (-)

$$\gamma = \alpha - \beta = \{(P(\gamma_k), T(\gamma_k)) \mid \gamma_k = \alpha_i \text{ or } P(\gamma_k) = P(\alpha_i), \; \gamma_k) = Min(T(\alpha_i), 1 - T(\beta_j)) \text{ when } P(\alpha_i) = P(\beta_j)\}$$

B. Query Example:

A major goal of this section that use fuzzy association operators to query fuzzy object-oriented data. The truth values of the result patterns from implementing fuzzy association operators is computed by using methods of determining truth value mentioned. We consider a followed fuzzy query example on the FSG of Human Resource database in Fig. 1 and the FOG in Fig.3. "List the o-names of organizations and the names of persons who are highly educated, high level technicians and speak fluent English".

The query can be represented by fuzzy association algebraic as follows:

π(σ((o-name*organization * person * name) • (person * highly educated person * spoken language) • (person * high level technician)) [spoken language = English]) [o-name, name, o-name:name]

Processing this query includes 4 steps as follows:

HCM o1 p1 Nam (0.8)
HN o2 p2 Ha (0.7)
DN o3 p3 Kim (0.8)

Step 1: Three patterns are processed by associate operator.

Pattern 1 = o-name*organization * person * name

Pattern 2 = person * highly educated person * spoken language

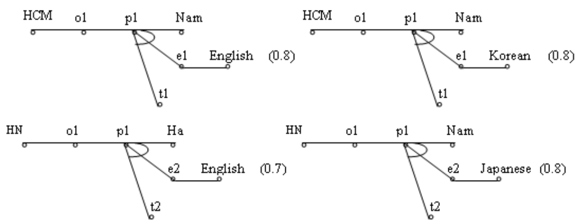
p1 e1 Korean (0.8)
p2 e2 Japanese (0.5)
p3 e3 Chines (0.5)
p4 e4 Chines (0.5)
p1 e1 English (0.8)
p2 e2 English (0.7)
p3 e3 English (0.8)
p4 e4 English (0.8)

Pattern 3 = person * high level technician

p1 t1 (0.8) p2 t2 (0.5)

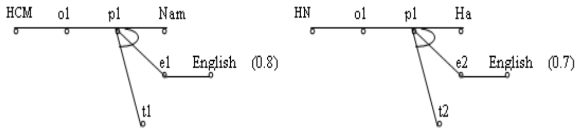
Step 2: Three Pattern 1, Pattern 2 and Pattern 3 are combined to form Pattern 4 by intersect operator.

Pattern 4 = Pattern 1 • Pattern 2 • Pattern 3



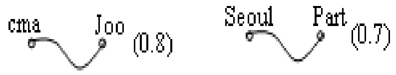
Step 3: Selection the patterns satisfied the condition, person is associated with 'spoken language = English', is processed from Pattern 4

Pattern 5 = σ(Patterns 4)[spoken language = English]



Step 4: Project the Pattern 5 to two subpatterns o-name and name and then make derived pattern from o-name node to name node.

Result_Pattern = π(Pattern 5)[o-name, name, o-name:name]



The Result_Pattern is represented by algebraic representation as follows:

((HCM, Nam)~, 0.8), ((HN, Ha)~, 0.7)

For the query, the result fuzzy patterns ((HCM, Nam)~, 0.8) mean proposition 'Nam employed in HCM is highly educated, high level technicians and speaks fluent English.' is true with degree 0.8. The result fuzzy pattern ((HN, Ha)~, 0.7) means proposition 'Ha employed in HN is highly educated, high level technicians and speak fluent English.' is true with degree 0.7. By the truth values of two patterns in Result_Pattern, we know ((HCM, Nam) ~, 0.8), is a more suitable answer for the query.

VI. CONCLUSION

To extract the necessary information in the database, we need to build a data query language with the expectation that it manipulates flexibly, versatilely and precisely. In this article, we proposed a method of determining the degree of relationship between two objects considered in the relationships between classes in the fuzzy OODB model as the basis for constructing fuzzy association algebra. In addition, we also presented an example of querying data in fuzzy OODB based on fuzzy association algebraic operators. In further studies, we will present the properties of the association algebra operators and its application in the analysis and optimization of query data.

VII. REFERENCES

[1]. B.Bouchon-Meunier, Ho Thuan, Dang Thanh Ha, Fuzzy Logic and Applications, Vietnam National University, Hanoi, 2007.

[2]. Bosc P, Pivert O, SQLf: a relational database language for fuzzy query, IEEE Transaction on Fuzzy System

model, 3 (1), 1995, 1 – 19.

[3]. Christophe Lécluse, Philippe Richard, Fernando Velez, 0_2 - an Object-Oriented Data Model, Proceedings of the ACM SIGMOD, 17(3), june 1988, 424 – 433

[4]. Doan Van Ban, Ho Cam Ha, Vu Duc Quang, Normalizing object classes in fuzzy object-oriented database schema, Journal of computer science and cybernetics, 28 (2), 2011.

[5]. GShaw, S.Zdonic, A object algebra for Object-Oriented databases, In Proc. 6th Int. Conf. on Data Engineering, 1990, 154 – 162.

[6]. Ho Cam Ha, An Approach to extending the relational database model for handling incomplete information, Proceedings of the third international conference on intelligent technologies and third VietNam-Japan symposium on fuzzy systems and applications InTech/VJFuzzy, 2002, 109-115.

[7]. M.Carey, D.Dewitt, S.Vandenberg, A data model and query language for EXODUX, In Proc. ACM SIGMODE Int. Conf on Management of Data, 1988, 413 – 423.

[8]. Roland R. Wagner, Helmut Thoma, Database and expert systems applications: 7th International Conference, DEXA ’96, 1996, 500 – 509.

[9]. S.Cluet, C.Delobe1, A general framework for the optimization of Object-Oriented queries, In Proc. ACM SIGMOD Int. Conf. on Managemeet of Data, 1992, 383 – 392.

[10]S.L.Vandenberg, D.Dewitt, Algebraic support for complex object with array, identity, and inheritance, In Proc. ACM SIGMODE Int. Conf. on Management of Data, 1991, 158 – 167.

[11] Selee Na and Seg Park, A Fuzzy Association Algebra based on A fuzzy Object Oriented Data Model, IEEE, 1996, 276 – 281.

[12] Stanley Y.W.Su, Mingsen Guo, Herman Lam, Association Algebra: A Mathematical Foundation for Object- oriented Databases, IEEE Tran. on Knowledge and Data Engineering, 5 (5), 1993, 775 – 798.

[13] Zadeh, L. A., Fuzzy sets, Information & Control, 8 (3), 1965, 338-353

[14] Zongmin Ma, Fuzzy Database Modeling with XLM, Springer, NewYork, 2005.

[15] Z. M. Ma, W. J. Zhang and W. Y. Ma, “Assessment of data redundancy in fuzzy relational databases based on semantic inclusion degree”, Information Processing Letters, 72 (1-2), 1999, 25-29.

