Available online at www.sciencedirect.com
ScienceDirect
Information and Software Technology 49 (2007) 172–180
INFORMATION AND SOFTWARE TECHNOLOGY
www.elsevier.com/locate/infsof

Generalization of strategies for fuzzy query translation in classical relational databases

Z.M. Ma *, Li Yan

College of Information Science and Engineering, Northeastern University, Shenyang, Liaoning 110004, China

Received 26 May 2004; received in revised form 16 May 2006; accepted 17 May 2006
Available online 27 June 2006

Abstract

Users of information systems would like to express flexible queries over the data possibly retrieving imperfect items when the perfect ones, which exactly match the selection conditions, are not available. Most commercial DBMSs are still based on the SQL for querying. Therefore, providing some flexibility to SQL can help users to improve their interaction with the systems without requiring them to learn a completely novel language. Based on the fuzzy set theory and the α-cut operation of fuzzy number, this paper presents the generic fuzzy queries against classical relational databases and develops the translation of the fuzzy queries. The generic fuzzy queries mean that the query condition consists of complex fuzzy terms as the operands and complex fuzzy relations as the operators in a fuzzy query. With different thresholds that the user chooses for the fuzzy query, the user's fuzzy queries can be translated into precise queries for classical relational databases.

Keywords: Classical relational databases; Fuzzy set; Fuzzy query; SQL; Translation

1. Introduction

In real-world applications, information is often vague or ambiguous. Therefore imprecision and uncertainty of information have been extensively introduced into database management systems (DBMSs). This issue covers the storage and handing of imperfect information on the one hand. On the other hand, an important issue is to extend DBMSs functions for the expression of flexible query in order to make these systems able to satisfy some user needs closely [3]. Fuzzy values have been employed to model and handle imprecise information in databases since Zadeh [20] introduced the theory of fuzzy sets [11,16]. In [15], Tahani firstly advocated the use of fuzzy sets for querying conventional databases, where imprecise conditions inside queries were seen as fuzzy sets. SQLf language, which is a fuzzy extension to SQL and was proposed by Bosc and Pivert [3], represents a synthesis of the characteristics and functionalities suggested in other previous proposals of flexible query in classical databases, such as Tahani [15], Bosc et al. [2], Wong and Leung [17], and Nakajima et al. [12]. Also there are some extensions to SQLf [1,9,10]. While most of these works mainly focused on the SELECT statement in SQL, in [8], Galindo et al. developed the FSQL language (Fuzzy SQL), an extension of the SQL language that allows flexible conditions in the queries. The FSQL contains DML (Data Manipulation Language) of FSQL, including SELEC, INSERT, DELETE, and UPDATE, and DML (Data Definition Language) of FSQL, including CREATE, ALTER, and DROP.

Concerning evaluation of fuzzy queries against regular databases, two strategies have been identified in [4]. The first strategy assumes that a threshold (α) is associated with a fuzzy query in order to retrieve its α-level cut. The other strategy is relies on the decomposition of a fuzzy query into elementary operations and the evaluation of these basic operations by means of efficient algorithms. Compared with the second strategy that a completely new systems should be

* Corresponding author. Fax: +86 24 8368 1823.
E-mail address: mazongmin@ise.neu.edu.cn (Z.M. Ma).

0950-5849/$ - see front matter © 2006 Elsevier B.V. All rights reserved.
doi:10.1016/j.infsof.2006.05.002

built in order to implement it, the first strategy can be implemented by using a regular DBMS and developing an additional layer. So some works have been undertaken along the line of the first strategy. In [5,6], a databases skeleton was presented and its application to fuzzy query translation was introduced. In [7], Chen and Jong developed a method for fuzzy query translation based on α-cuts operations of fuzzy numbers. It should be noted that the query condition in a fuzzy query can be described by fuzzy terms or fuzzy relations. Combining fuzzy terms and various of regular relations such as $=, >, <, \neq, \geqslant$, and $\leqslant$ etc., or combing some fuzzy relations such as "(*not*) *close to*/*around*", "(*not*) *at least*", and "(*not*) *at most*" etc. with crisp values, the fuzzy query conditions are formed. The translation of the fuzzy query presented in [7], however, is only useful in the following two situations:

- the fuzzy query condition consists of the regular relation "=" and a fuzzy term;
- the fuzzy query condition consists of the fuzzy relation "(*not*) *close to*" and a crisp value.

It is clear that the fuzzy queries against regular relational databases in [7] only investigated two simple fuzzy query conditions. Therefore, the corresponding translation is incomplete. In this paper, the generic fuzzy queries against regular relational databases are introduced and the translation of fuzzy queries is developed based on the fuzzy set theory and the α-cut operation of fuzzy number. Here the generic fuzzy queries mean that the query condition consists of complex fuzzy terms as the operands and complex fuzzy relations as the operators in a fuzzy query. With different thresholds that the user chooses for the fuzzy query, the user's fuzzy queries can be translated into precise queries for regular relational databases.

The rest of this paper is organized as follows. In Section 2, we briefly review some basic knowledge of fuzzy sets theory. In Section 3, we investigate the generic fuzzy queries. In Section 4, we develop the methods for fuzzy query translation based on the α-cuts operations of fuzzy numbers. Section 5 concludes this paper.

2. Fuzzy sets theory

Fuzzy data is originally described as fuzzy set by Zadeh [20]. Let U be a universe of discourse. A fuzzy value on U is characterized by a fuzzy set F in U. A membership function

$$\mu_F : U \rightarrow [0,1] \tag{1}$$

is defined for the fuzzy set F, where $\mu_F(u)$, for each $u \in U$, denotes the degree of membership of u in the fuzzy set F. Thus the fuzzy set F is described as follows:

$$F = \{\mu_F(u_1)/u_1, \mu_F(u_2)/u_2, \ldots, \mu_F(u_n)/u_n\} \tag{2}$$

When U is an infinite set, then the fuzzy set F can be represented by

$$F = \int_{u \in U} \mu_F(u)/u \tag{3}$$

Let A and B be fuzzy sets on the same universe of discourse U with the membership functions μ_A and μ_B, respectively. Then we have

Union. The union of fuzzy sets A and B, denoted $A \cup B$, is a fuzzy set on U with the membership function $\mu_{A\cup B}$: $U \rightarrow [0,1]$, where

$$\forall u \in U, \mu_{A\cup B}(u) = \max(\mu_A(u), \mu_B(u)). \tag{4}$$

Intersection. The intersection of fuzzy sets A and B, denoted $A \cap B$, is a fuzzy set on U with the membership function $\mu_{A\cap B}$: $U \rightarrow [0,1]$, where

$$\forall u \in U, \mu_{A\cap B}(u) = \min(\mu_A(u), \mu_B(u)). \tag{5}$$

Complementation. The complementation of fuzzy set $\bar{A}$, denoted by $\bar{A}$, is a fuzzy set on U with the membership function $\mu_{\bar{A}}$: $U \rightarrow [0, 1]$, where

$$\forall u \in U, \mu_{\bar{A}}(u) = 1 - \mu_A(u). \tag{6}$$

Definition: A fuzzy set F of the universe of discourse U is convex if and only if for all u_1, u_2 in U,

$$\mu_F(\lambda u_1 + (1-\lambda)u_2) \geqslant \min(\mu_F(u_1), \mu_F(u_1)) \tag{7}$$

where $\lambda \in [0,1]$.

Definition: A fuzzy set F of the universe of discourse U is called a normal fuzzy set if $\exists u \in U$, $\mu_F(u) = 1$.

Definition: A fuzzy number is a fuzzy subset in the universe of discourse U that is both convex and normal.

Now several notions related to fuzzy number are discussed. Let U be a universe of discourse and F a fuzzy number in U with the membership function μ_F: $U \rightarrow [0,1]$. We have then the following notions

Support. The set of the elements that have non-zero degrees of membership in F is called the support of F, denoted by

$$\mathrm{supp}(F) = \{u | u \in U \text{ and } \mu_F(u) > 0\}. \tag{8}$$

Kernel. The set of the elements that completely belong to F is called the kernel of F, denoted by

$$ker(F) = \{u | u \in U \text{ and } \mu_F(u) = 1\}. \tag{9}$$

α-Cut. The set of the elements which degrees of membership in F are greater than (greater than or equal to) α, where $0 \leqslant \alpha < 1$ ($0 < \alpha \leqslant 1$), is called the strong (weak) α-cut of F, respectively, denoted by

$$F_{\alpha+} = \{u | u \in U \text{ and } \mu_F(u) > \alpha\} \text{ and}$$
$$F_{\alpha} = \{u | u \in U \text{ and } \mu_F(u) \geqslant \alpha\}. \tag{10}$$

The relationships among the support, kernel, and α-cut of a fuzzy set can be illustrated in Fig. 1.

It is clear that the α-cut of a fuzzy number corresponds to an interval. Let A and B be the fuzzy numbers of the universe of discourse U and let A_α and B_α be the α-cuts of the fuzzy numbers A and B, respectively, where

$$A_\alpha = [x_1, y_1] \text{ and } B_\alpha = [x_2, y_2]. \tag{11}$$