Available online at www.sciencedirect.com

ScienceDirect

Procedia Computer Science 115 (2017) 14–21

Procedia Computer Science

www.elsevier.com/locate/procedia

7th International Conference on Advances in Computing & Communications, ICACC-2017, 22-24 August 2017, Cochin, India

An Approach for Augmenting Selection Operators of SQL Queries using Skyline and Fuzzy-Logic Operators

Ajay Agarwal[a], Apeksha Aggarwal[b]*, Apoorv Agarwal[b]

[a]*KIET, Ghaziabad, 201206, India*
[b]*Dr B.R. Ambedkar NIT, Jalandhar, 144011, India*

Abstract

In this work, expressivity two operators fuzzy and skyline is primarily focused upon. Stated two known operators have been used here for augmenting selection operator of SQL. Since SQL is unable to understand linguistic concepts of natural language there is essence of a preference based solution so as to return answers according to user preferences. Both these operators can return a different answer to the same query. Additionally in this paper, performance of fuzzy and skyline processing time and the number of returned answers by means of an experimental study is analyzed.

Peer-review under responsibility of the scientific committee of the 7th International Conference on Advances in Computing & Communications.

Keywords: Fuzzy; Skyline; Preprocessor; Operator;

1. Introduction and motivation

SQL has several limitations in order to express linguistic terms as well as user preferences. The ORDER BY clause is a way to express these preferences, but it is complicated to manage criteria that involve three or more conditions. In order to specify these preferences in a simple way, we propose to extend database systems by two operations skyline operation and fuzzy Logic based operations.

* Corresponding author. Tel.: +91-963-924-1481;
E-mail address: apeksha.aggarwal785@gmail.com, apu.grwl@gmail.com

1877-0509
Peer-review under responsibility of the scientific committee of the 7th International Conference on Advances in Computing & Communications
10.1016/j.procs.2017.09.071

The skyline operator [1], is introduced for relational database as extended SQL which filters out a set of interesting points from a potentially large set of data points. A point is interesting if it is not dominated by any other point. Another operator based on fuzzy logic uses some standard mathematical functions in order to simulate the formula of selection operation of a relational algebra is discussed in section 2. This augmentation permits the fuzzy predicates such as "large", "much-greater-than", etc. to be used in a formula. Using fuzzy predicate is semantically equivalent to define a fuzzy set.

Borzsonyi et al. [1] discussed skyline operator and Aggarwal et al. [2] discussed various algorithms on skyline computation in detail. Out of various skyline algorithms, distributed vertical decompositions of skyline is one of the most important works that is discussed in Trimponias et al [3].

Application of fuzzy logic in SQL is been a decade old process. One of the early researches have been done by Bosc et al. [4]. Later application of fuzzy on relational database are given in several works [5-8]. Identifying set of k-dominating tuples from a relation is considered in Santoso et al. [9]. Other works includes ranking of SQL queries using skyline [10] computation method.

In this work, a software module has been proposed which provides an interface to the user to make fuzzy queries and preprocesses these queries. These fuzzy queries are able to express linguistic terms along with SQL basic language. Preprocessing of these fuzzy queries involves identification of imprecise terms (such as "approximately-equal-to", "medium" etc.) and eliciting of threshold values for definition of these fuzzy operators.

After this preprocessing the module retrieves all records satisfying these fuzzy relationships. This paper form a basis for augmenting the basic capability of any relational query language which in turn relieves the database users of the constraint of having to formulate queries to the database in precise terms. As an example, Employee relational database is chosen, details are provided in further sections.

Section 2 describes the skyline and fuzzy queries. Section 3 and 4 explains the proposed fuzzy query preprocessor and algorithm used to process both fuzzy and skyline queries respectively. Section 5 includes results and discussions. Section 6 presents conclusion and future work.

2. Skyline and fuzzy operator

The skyline [1] is defined as set of those points that are not dominated by any other point. A point dominates another point if it is as good or better in all dimensions and better in at least one dimension. Skyline SQL query extension includes various operators. Operators like MIN, MAX, and DIFF specify whether the value in that dimension should be minimized, maximized, or simply be different. Syntactically, the SKYLINE OF clause is similar to the ORDER BY clause and a_1, …, a_n, named dimensions, are the attributes over which user preferences ranges.

A membership degree [5-8] is a real number between [0, 1]. In extreme cases, if the degree is 0 the element does not belong to the set, and if 1 the element belongs 100% to the set. Let *X* denote a universal set. Then, the membership function mdA(x) by which a fuzzy set A is usually defined has the form as shown in (1).

$$MdA(x): x \rightarrow [0,1] \tag{1}$$

Where [0, 1] denotes the interval of real numbers from 0 to 1, inclusive. A fuzzy set is a method of encoding and using knowledge that does not have clearly defined boundaries. Many of the real world phenomena we encounter fall into this class such as, heavy, high, low, fast, slow, full, empty, many, few, light and dark etc. are often associated with variables that do not have accurately defined boundaries. A fuzzy set maps a value in the data to its degree of membership in a set of ill-defined elements. Fuzzy set takes a concept and expresses it through a mapping of data to membership functions. Piecewise Linear Membership function of *AGE* (small, large) should be defined as given in Fig. 1.

This piecewise linear membership function for a fuzzy set ‘small’ on the universe of discourse *age* is defined as mdlarge(*age*) →[0,1], where each element of age is mapped to a value between 0 and 1 as shown in (2). This value, called membership value of degree of membership quantifies the grade of membership of the element in age to a fuzzy set ‘small’. Similarly for fuzzy set ‘large’ over universe of discourse *age* with membership degrees, mdsmall(*age*) →[0,1] are mapped to a value between 0 and 1 as shown in (3). Membership functions allow us to graphically represent a fuzzy set. Fig. 1 shows membership map of ‘small’ and ‘large’ *age*. The x-axis represents the universe of discourse i.e. *age*, whereas the y-axis represents the degrees of membership in the [0,1] interval. Similarly piecewise linear