Visualization of Database Query Improved by Fuzzy Logic Using Virtual Reality

Vanessa Avelino Xavier de Camargo, Marcos Wagner de Souza Ribeiro

Federal University of Goias
vanessaxcamargo@gmail.com, marcos_wagner@yahoo.com.br

Abstract – Fuzzy Logic and Fuzzy Set Theory have been widely exploited on account of their feature to translate uncertain, ambiguous and vague information - typical of human beings thoughts - to numerical data. The scope of fuzzy applications is actually large and it may be increasing. One of its applications is in Relational Databases, promoting an improvement in information retrieval. In this context, this paper presents a convenient and simple way to treat linguistic variables in database queries with the aim to make data retrieval closer to imprecise human reasoning. A trigger, using Structured Query Language (SQL) and MySQL Database Management System, has been created to make fuzzy membership degrees based on trapezoidal fuzzy numbers. With the purpose to illustrate and test the advantages of this approach, a case study based on age variables - child, young, middle-age, elderly - has been done, assuming an insurance company interested in setting values according to the age of clients. However, this approach presented a negative aspect: Data Visualization, which has been solved using Virtual Reality.

Keywords – Fuzzy logic, linguistic variables, membership degree, database, SQL, Virtual Reality.

1 Introduction

Databases and Databases Management Systems have become essential components of daily life in modern society. During the day, most of us face activities that involve any interaction with databases. These interactions are examples of what Elmasri [1] calls traditional applications of database, in which most of information that is stored and accessed is presented by text or numerical formats. In recent years, technological advances have generated interesting and innovative applications of database systems [1].

A DBMS is a data collection interrelated and a set of programs to access these data. Its main goal is provide an efficient and convenient way to retrieve information of the database. [2] But, due to the fact that information is stored by numerical format, sometimes classical queries cannot return desired results when decisions have inaccuracies through linguistic variables like "almost", "much", "few", "young", "old", "low", "high" [9].

The traditional way to do a query in databases is using numbers to translate users' feelings to the computers. Unfortunately, these queries do not handle vague, incomplete and also contradictory data, causing losses in information retrieval. In this context, fuzzy logic can be an effective method to reduce or avoid such losses and Virtual Reality in Data Visualization presents a new dynamism in systems with a lot of data.

Therefore, purposes of this work are show how fuzzy sets and Virtual Reality can be used to optimize SQL queries involving these indefinites expressions and view the data. To illustrate this, it has been done a case study based on an insurance company interested in setting values according to the age of clients

1.1 Fuzzy Logic Approach

The fuzzy set theory was conceived as a generalization of classic set theory in a 1965 paper by Lotfi A. Zadeh. In the classic theory, once there is a subset of the set , there are only two possibilities for a given element in : or is in the set or isn't in the set . That is, membership is given by a function in which value 0 means that the element is not in the set while value 1 means that the element is in the set .

As it can be seen in [3], Zadeh introduced a membership function in which the image is within the unitary interval of the real numbers, so that now it can have certain "shades of gray" between 0 (no membership at all of the element to the set) and 1 (total membership of the element to the set). An important consequence is that it can have an element lying in a set with certain degree of membership and also lying in the complement of this set with another degree. This is nonsense in terms of classic set theory, but it works in terms of the generalized fuzzy theory.

A fuzzy set is a pair consisting of a "crisp" subset of the set and a membership function which gives the "membership level" of any element of relative to the set . A linguistic variable is a variable whose values are

words or sentences [4]. The main role of a linguistic variable in fuzzy systems is to propose logical consideration of imprecise and vague concepts. For instance, let us take the example of the age: it is a vague concept that we can describe using one fuzzy set (usually a fuzzy number) for the value "child", one for the value "teen", one for "middle age" and so on.

A fuzzy number is a fuzzy set with special properties that it can be define for some operations like plus or multiplication by a real number, as in [6]. Among the most used ones, there are the triangular, trapezoidal and Gaussian fuzzy numbers. This paper deals with the first two of them.

A triangular fuzzy number A has a membership function   as the equation:

$$\overline{\qquad}$$

$$\overline{\qquad} \tag{1}$$

Graphically, following [4] the membership function of a triangular fuzzy number has triangular form with the interval   as base and the point   as the other vertice. Also by [4] and [6] it can see that a fuzzy number is said to be trapezoidal if its membership function   is as the equation:

$$\overline{\qquad}$$

$$\overline{\qquad} \tag{2}$$

As it could be expect, the graphic of a trapezoidal fuzzy number has trapezoidal form.

1.2 Database Approach

A database is a related data collection. The data are events which have an implicit meaning and can be stored. It has the following properties:

- It represents some real world aspects, being called, sometimes, of Universe of Discourse (UoD). Changes are reflected in a database.

- It is a logical collection of data with one meaning. A random organization cannot be interpreted as a database.

- It is designed, built and supplemented by data, answering a specific purpose. It purchases definite users set and some applications, according to the interest of this users set [1].

A database system provides a Data Definition Language (DDL) to specialize the database schema and a Data Manipulation Language (DML) to express database queries and updates. Actually, these languages are not separated, together they are parts of only one database language, as the famous Structured Query Language (SQL).

A Data Manipulation Language allows users to access or manipulate data as organized by the appropriate data modeling. While Data Definition Language is used to define a database framework [2].

A standard relational database consists of a group of relations expressed as tables made of columns and rows. The names of the columns are called attributes. The cells in a column are the domain of the attribute. The rows called tuples contain records or entries each occupying a cell. Several tables having common domains connected together represent a relational database [4].

2 System Architecture

Figure 1 represents the architecture used to develop the system. The database is on the top of the architecture, communicating with fuzzy logic layer and virtual environment.
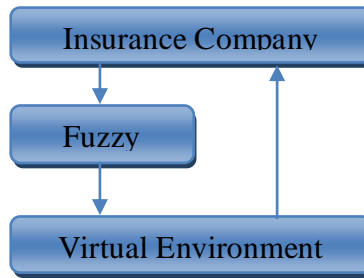
Figure 1. System Architecture.

3 Methodology

For the development of SQL queries using fuzzy logic, it has been chosen the DBMS MySQL Server 5.1 because it provides flexibility, scalability, strong data protection and management ease, even being open source software [7]. It has been run in a local machine with SQL standard.

3.1 Data Description

It has been chosen a case study of a fictitious company which wants to do a promotion targeted to a specific age group. First, it has been listed all people and their respective ages in a table called *pessoa*, as shown in Figure 1.

Figure 2. List of registered people and their ages (table *pessoa*).

The intention of the company is select people whom can be classifying by a term "middle age", which is the specific age target. If it is done a standard query, selecting people with age between forty and sixty years old (Figure 2), results are going to exclude people with almost these ages, but which would be potential aim.

Figure 3. People query by age interval.

When compared to Figures 3 and 4, can be observed that person named João Eduardo Alves Ferreira, whom has nearly forty years and can be called with "middle-age", is not in the query result, since the range is accurate and does not allow flexibility.

As previously mentioned, fuzzy logic gives an important tool to solve these problems and improve most of database queries[5].

3.2 Fuzzy Implementation

Beyond to storing these data, the database also has other tables, forming the fuzzy knowledge base. One table has been created to shown the fuzzy inference resulting. It tuples has been generate through a trigger – a named database object that is associated with a table, and that activates when a particular event occurs for the table [7] – based on IF-THEN rule system. The pseudo-code presented in Figure 3 illustrates it.

```
CREATE TRIGGER gerapert AFTER INSERT ON pessoa FOR EACH ROW
BEGIN
    DECLARE pert_1, pert_2, pert_3, pert_4, m FLOAT;
    IF (idade > 14) THEN pert_1 = 0;
        ...
    IF (idade > 45) THEN pert_n = 1;
        ...

    INSERT INTO fidade VALUES (cod_pessoa, pert_1, pert_n);
END;
```

Figure 4. Trigger called *gerapert*.

This trigger runs after any insertion on table *pessoa*. For each age inserted, it calculates its membership function based on trapezoidal fuzzy statement (Figure 2).

The membership functions for variable age have been intuitively structured, considering the authors' opinions. This variable is formed by a terms set ordered 1 to 4, saved on table *termos* shown in Figure 4, where *cod_termo* 1 refers to linguistic variable "child"; *cod_termo* 2 refers to "young"; *cod_termo* 3 refers to "middle-age"; and *cod_termo* 4 refers to "elderly".
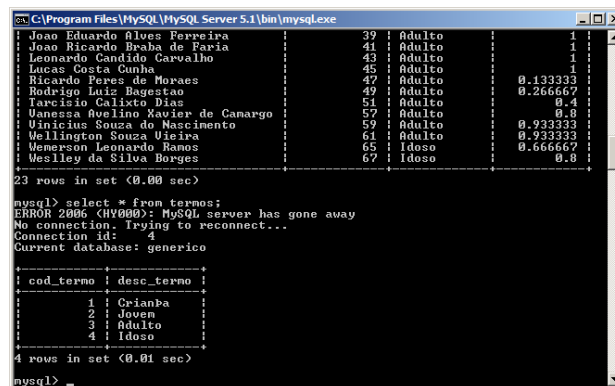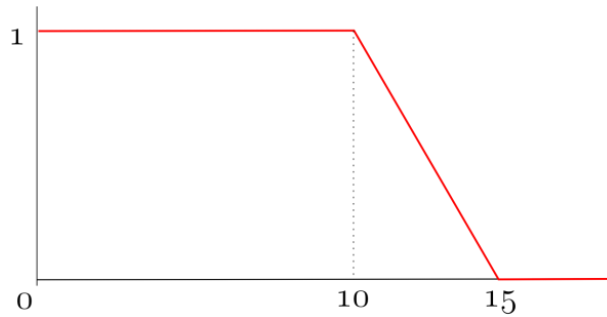

Figure 5. Table *termos*.

The interval of each term is represented by a graphic based on the membership function previously calculated, whose domain is
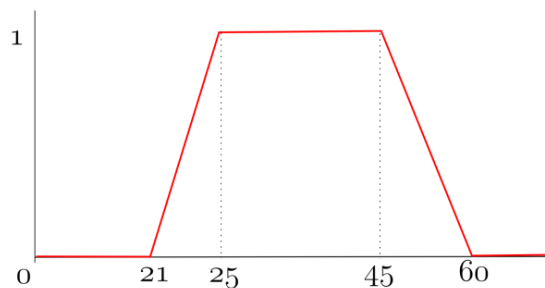
Graphic 1. Interval representing term "child".

Graphic 1 is showing membership degree of term "child", where ages up to 10 are considered child for sure (membership degree 1) and ages between 11 and 14 have membership degree given by ——— . Older than this, means that it is not a child (membership degree 0).
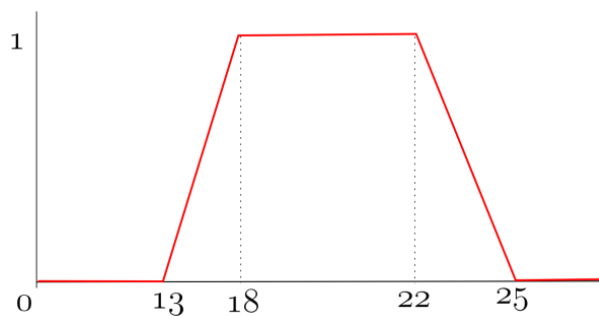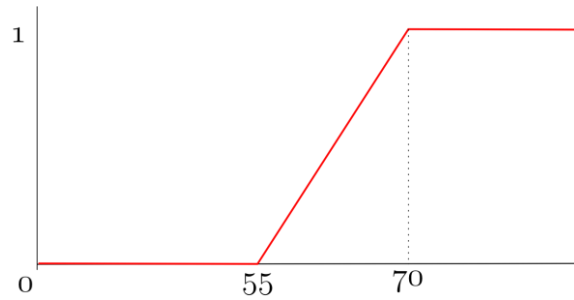

Graphic 2. Interval representing term "young".

Graphic 2 is showing membership degree of term "young", where ages between 18 and 22 are considered young for sure (membership degree 1); ages between 14 and 17 have membership degree given by ——— and ages 23 and 24 have membership degree given by ——— . People younger than 14 and older than 24 are not young (membership degree 0).

Graphic 3. Interval representing term "middle-age".

Graphic 3 is showing membership degree of term "middle-age", where ages between 25 and 45 are considered middle-age for sure (membership degree 1); ages between 22 and 24 have membership degree given by ——— and ages between 46 and 59 have membership degree given by ——— . People younger than 22 and older than 59 are not middle-age (membership degree 0).

Graphic 4. Interval representing term "elderly".

Graphic 4 is showing membership degree of term "elderly", where ages up to 70 are considered elderly for sure (membership degree 1) and ages between 56 and 69 have membership degree given by ———— . Younger than this, means that it is not elderly (membership degree 0).

3.3 Fuzzy Queries

The intervals shown in Graphics 1, 2, 3, 4 form the conditions included in trigger. Each age inserted creates a membership degree for each term about age variable. After that, trigger compares what is its highest membership degree, classifying each age according to respective terms. If an age is in the interval between two terms, in other words, if it generates membership degrees other than 0 in two linguistic terms, the trigger chooses the highest one and classifies this age for this term. For example, in Figure 1, it can be seen that Bruno de Vasconcelos is twenty-three years old. The membership function returns   and   , which means that he is more middle-age than young and will be classifying by middle-age.

A query with every people and their respective classification is illustrated in Figure 5.

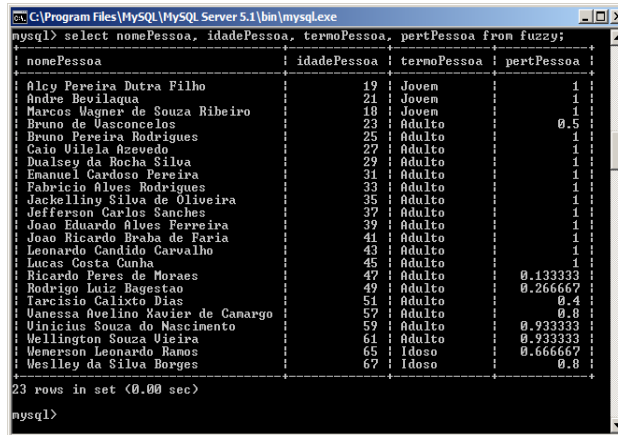**Figure 6.** Overview of all people and their specific classification.

From this, is no longer necessary to define ranges in queries. For instance, it can seek by people who are young, as shown in Figure 6. The search will bring people's name and their respective classification about age variable.

Figure 7. Query by term "young".

Even more, it can be done a query by membership degree, resulting in a table with people's name, their ages, classification and highest fuzzy membership degree (Figure 7).

Figure 8. Overview of all people, their ages, classification and fuzzy degree.

3.4 Virtual Environment

Using XML, all queries generate an X3D code that represents tridimensional information distinguishing the linguistic variables (child, young, middle-age and elderly) through avatars and the membership degrees through different colors, being that in bar graphic format. Besides this tridimensional information, the virtual environment also provides to users other feelings, like the sound generated by approximation of the columns. The system primarily presents an interface with all four kinds of avatars, as shown in Figure 9. Users can select one of the avatars to obtain the wished query, for example, if the term "young" has been chosen, the system will return a column with some stacked avatars. This column can has a lot of stacks. First stack represents the membership degree "1" and its color is green. According to the change of membership degree, the size of the stacks and the color will change.
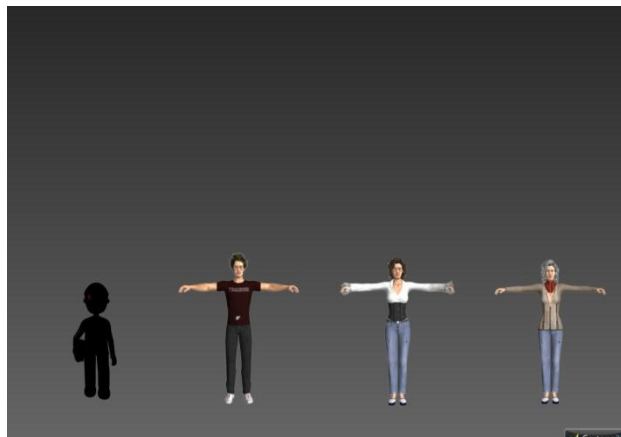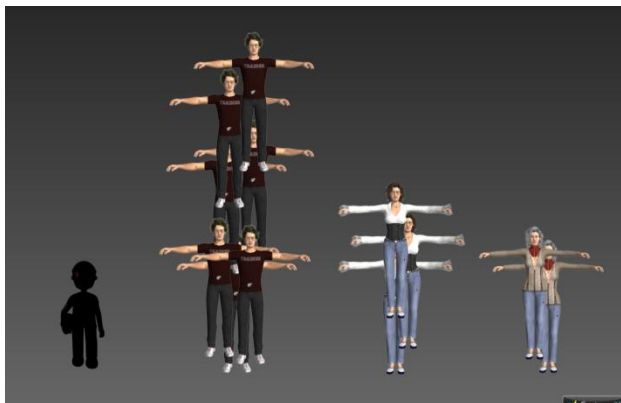

Figure 9. Initial interface.

Figure 10. Modified interface.

4 Conclusion

This article has shown how fuzzy logic can be used to improve database queries, bringing results closer to inaccurate reality. From the tests performed, a fuzzy query compared to a standard query promotes many advantages. First, queries using fuzzy sets ease the results returned to the user, bringing it closer to reality. These types of fuzzy queries are especially useful when involve linguistic values like "high", "almost", "much", "low", "few", "old", etc.

In a nutshell, fuzzy logic, which is already widely used in controlling and modeling [8], has a huge potential to be applied in database technology, as information retrieval and data mining. However, to validate the proposed application would be necessary to test it on larger databases and perform a statistical study to demonstrate the analysis.

Besides that, Virtual Reality is an efficient way to provide Data Visualization in this kind of application, which works with abstract data, commonly represented by bidimensional values.

Future works can elucidate complex queries involving several variables and bigger databases, which would generate a rule-based fuzzy system, helping in decision making for stocks and funds databases.

5 References

[1] Elmasri, R. & Navathe, S. B., Fundamentals of database systems (São Paulo: Pearson Addison Wesley, 2005).

[2] Silberschatz, A. & Sudarshan, H. F. K., Database system concepts (Rio de Janeiro: Elsevier, 2006).

[3] Zadeh, L. A., Fuzzy sets. Information and control 8th Edition, p.338-353, 1965.

[4] G. Bojadziev & M. Bojadziev, Fuzzy logic for business, finance, and management. (2nd Edition). Advances in Fuzzy Systems: Applications and Theory — Vol. 23. World Scientific Publishing Co. Pte. Ltd., Singapore 2007.

[5] Fernanda Bessani Leite Penteado: Método de Filtragem Fuzzy Para Avaliação de Bases de Dados Relacionais. (Dissertação de mestrado). EESC/USP São Carlos, 2009.

[6] Barros, L. C. e Bassanezi, R. C.: Tópicos de lógica fuzzy e biomatemática. Primeira edição. Coleção IMECC Textos Didáticos 5. Editora da Unicamp, Campinas - SP, 2006.

[7] Sun Microsystems, Inc MySQL AB. Site: http://www.mysql.com, 2008

[8] Simões, M. G & Shaw, I. S., Controle e modelagem fuzzy (São Paulo: Blucher: FAPESP, 2007).

[9] Peres, S. M. & Boscarioli, C., Sistemas gerenciadores de banco de dados relacionais fuzzy: uma aplicação em recuperação de informação. Online document: http:// www.periodicos.uem.br, 2002.