ECONTECHMOD. AN INTERNATIONAL QUARTERLY JOURNAL – 2014, Vol.3, No.3, 99-102

FUZZY MODELS PRESENTATION AND REALIZATION BY MEANS OF RELATIONAL SYSTEMS

FILATOV V.

Kharkov National University of Radio Electronics, Lenin Ave, 14, Kharkov, 61166, Ukraine
filatov_val@ukr.net

Received July 15.2014: accepted August 28.2014

Abstract. The feature of the research is the proposed fuzzy models representation method by means of relational systems, which unlike the known approaches, can solve integrated data mining problems in databases and fuzzy systems.

Key words: fuzzy set, fuzzy logic, fuzzy relational databases, intelligent systems, data mining.

INTRODUCTION

Computer technologies of intelligent algorithms are experiencing their heyday. This is due to the flow of new ideas coming from computer science, which was formed at the intersection of artificial intelligence, statistics and database theory [5]. The traditional mathematical statistics has long claimed to be the main tool for data analysis, but now because of the heterogeneity of the source data, disruption of traditional statistical assumptions, substantial prior uncertainty, etc. it was not efficient enough to process information in many applied problems.

However, the methods of mathematical statistics were useful mainly for testing pre-formulated hypotheses (verification-driven data mining). Methods of multivariate data mining designed to find patterns in multivariate data (or sequence of one-dimensional data). These include: visualization of data sets, cluster analysis, factor analysis, multidimensional scaling, Log-Linear analysis, canonical correlation, stepwise linear and nonlinear regression, correspondence analysis, time series analysis, classification trees and etc. The areas of application of these methods are the various areas of research: technological data, medicine, biomedicine, metallurgy, chemistry, molecular genetics, and others.

Modern technology intellectual Data Mining (discovery-driven data mining) is the process of extracting previously unknown, nontrivial, practically useful and interpretable knowledge from the "raw" unstructured data in large arrays or databases. In general, the problem being solved by Data Mining consist from finding patterns in data of different nature. The composition of the most common problems are such as forecasting, the designing of mathematical models, classification, clustering, rule generation, aggregation, reduction of dimension, visualization of data [4].

One of the areas of Data Mining is an exploratory data analysis, which is used to find systematic relationships between variables in situations where there are no (or are inadequate) a priori ideas about the nature of these connections [2].

Usually, the exploratory analysis takes into account and compared a large number of variables, and for the search of patterns variety of methods is used.

Many scientists research the problem of designing fuzzy databases, particularly on the basis of the relational model. Various versions of fuzzy relational database models have been designed. The results of the surveys, can be found in works of next authors: Buckles B.P., Petty F.E., Shenoi S., Melton A., Vila M.A., Lipski W. Jr., Prade H.

Systems built by combining the databases and fuzzy logic can significantly extend the functionality and range of tasks of data mining. The theory of fuzzy databases is not yet complete from a mathematical point of view, and there are still many issues that require resolution.

DEVELOPMENT OF A RELATIONAL MODEL REPRESENTATION AND STORAGE OF FUZZY DATA

Intelligent technologies are becoming an integral part of modern decision support systems, and the relevant question becomes - what requirements must meet the database to the data and structures in which they are stored to be as focused on methods of processing these data. For example, let's consider the task of designing an integrated intelligence system which combines a relational database technology and fuzzy logic. Systems built by combining the databases and fuzzy logic, can significantly extend the functionality and range of tasks of data mining [6].

To solve the problems of storing fuzzy data we will define a specialized type of relation. The scheme of such relation must meet two conditions: compatibility with the requirements of the classical relational data model and effective storage and representation of linguistic variable model [7]. Now let’s look into it’s graphical representation (Fig.1).

Any line on a coordinate plane can be represented as a binary relation, where Dom R is represented by the values of the x-axis, and Im R - the y-axis.

There are three indicators in the fuzzification problem that should be considered when forming relation.

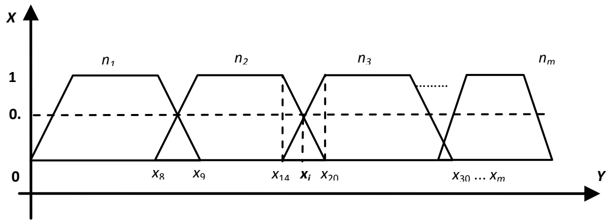

Fig.1. Graph of the values of linguistic variable

Define the fuzzy variable as a set (N, X, Y), where N - is the name of a variable, X - the area of research, Y - a fuzzy set on X. Using this definition, we define the three domains corresponding to elements of the variable.

Assume that:

$N = \{n_1, \ldots, n_m\}$,

$Y = \{0, 0.1, \ldots, 1\}$,

$X = \{x_0, \ldots, x_k\}$.

The *X* and *Y* correspond to the selected scale sampling of the coordinate axes and represent the region belonging to the parameter *N*. For this particular case we define the respective domains in order to present the values of fuzzy variable [3]:

$D_1 = \{n_1, n_2, n_3, \ldots, n_m\}$,

$D_2 = \{x_0, \ldots, x_8, x_9, \ldots, x_{14}, \ldots, x_i, \ldots, x_{20}, \ldots, x_{30}, \ldots, x_m\}$,

$D_3 = \{0, 0.1, \ldots, 1\}$.

We define a set of domain names and mapping to formulate a set of attributes: for a set of names $A = \{A_1, A_2, A_3\}$, - mapping $\rho : (A_1 \to D_1;\ A_2 \to D_2;\ A_3 \to D_3)$ determines the set attributes $A = \{A_1, A_2, A_3\}$ (for example Fig.2), which corresponds to the relation scheme:

$$S(A_1, A_2, A_3).$$

Thus, in general, we can talk about a universal relation, which includes the full set of tuple resulted by domains of Cartesian product $D_1 \times D_2 \times D_3$ [1].

A_1	A_2	A_3
n_2	x_{14}	0
n_3	x_{14}	1
n_2	x_i	0.9
n_3	x_i	0.9
n_2	x_{20}	1
n_3	x_{20}	0
...	...	…

Fig.2. Fragment of the values of linguistic variable

On this basis, we can conclude that the key of such relation will be a set of all attributes:

$$K = \{A_1, A_2, A_3\}.$$

Obviously, the informative of tuples is defined by the values in a chart of fazzification. Based on the conditions of the problem, must take into account another parameter - the set of values of sample from the database, for which diagram was made. That is, it is necessary to establish relation between the corresponding domains of fazzification relation - R^f and integrated into the database.

DEVELOPING A MODEL OF INTEGRATION FUZZY RELATION AND RELATIONAL DATABASE

Let's look at the problem in general. Let U $(R_1,\ldots, R_n)$ – database that stores basic data about the studied domain, $R^f(A_1, A_2, A_3)$ – fuzzy relation. The problem will make sense if there is an attribute in U database, fuzzification is based on.

In order to provide simultaneous work of two databases U and R; we need to formalize the process of integration, based on the gradual normalization technique. The structure of U relation is based on basic functional dependencies

$$F = \{M_i \rightarrow N_i\}$$

where: $M_i, N_i \in U$.

Now select one of the dependencies, which includes an attribute with fuzzy parameters, and define it as a $W \rightarrow V$, where W and V, in global case, may be sets. Relation R^f contains one dependence:

$$F' = \{A_1, A_2, A_3 \rightarrow A_1, A_2, A_3\}.$$

Now we can obtain an equivalent set:

$$F' = \{A_1, A_2, A_3 \rightarrow A_1; A_1, A_2, A_3 \rightarrow A_2; \\ A_1, A_2, A_3 \rightarrow A_3\}$$

using output axioms.

Let's assume that the parameter corresponds to the fuzzy attribute A_2, then to determine the type of communication we need to get a set $F = F \cup F'$ and study two cases affecting the rules of normalization.

1. $A_2 \in W$ - search for parent relations: if functional relations and $\xi \rightarrow \zeta$, and $\omega \rightarrow \zeta$ take place - then the dependence of $\omega \rightarrow \zeta$ is incomplete.

2. $A_2 \in V$ - search for transitively dependent elements: if the functional relations $\xi \rightarrow \omega$ and $\omega \rightarrow \zeta$ take place - then an element ζ is transitive dependent. The existence of such relations will allow to make correct decomposition and to establish connection between U and R^f databases. If $A_2 = W$ or $A_2 = V$, then the process of decomposition leads to the second or third normal form. If equation is not satisfied, then it's impossible to organize support of connected data uniqueness, because the association between relations will be "many-to-many".

Usually, in practice, the condition of equality is not meets, and for the normalization is necessary to allocate basis F and repeat the process of decomposition. Considering the fact that the structure of the database should not be changed, fazzification relation R^f and U must be linked without restructuring the data schema. Using a diagram of the model "entity-relationship", we represent R^f and U as the main entities.

Thus, in general to analyze data that's stored in relational databases, it is enough to build a fuzzy relation and establish a connection with the attribute (attributes) whose values must be properly analyzed.

To destroy one-to-many relationship «*N*:*M*» let's introduce more cohesive entity that will solve the problem of supporting the integrity of data by defining new types of relationships. The entity relationship will contain one attribute - binding for R^f and *U*, for objective reasons, it will be the key. Fig. 3 shows a diagram of integration explored database and the fazzification relationship supporting the uniqueness of relation.

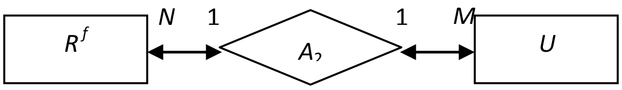

Fig. 3. Graph "entity-relationship"

Based on the description of a conceptual scheme that the correct connections for R^f and *U* needs to design an intermediate relation. This approach ensures data consistency for all types of fazzification settings.

We show that for this problem is quite correct results when the join relations with the association of type «*N*:*M*». Possible values of attribute $A_1 \in U$ can be repeated as many times as given value crosses the boundaries of the Y-axis on fazzification graph. That is, each A_1 attribute value corresponds to the line of unique data. If A_1 is not the key, and the values are repeated, then, by definition of sets, in a row must be at least one different value. In terms of the problem being solved it is needed to analyze all these lines. A_1 attribute's value relation R^f may also be repeated, at all sorts of combinations.

Thus, in a general way, to analyze the data accumulated in a relational database, is enough to design a fazzificationed relation and establish relation with the attribute (attributes) by values of which is needed to be conduct an appropriate analysis/

CONCLUSIONS

Based on analysis of relational data model's features we described the main problems of building integrated information systems that combine the database as a source of primary data mining tools that enable to extract knowledge from data. This article proposes an efficient method for designing relational data model schema to represent the membership functions of linguistic variables. The proposed approach describes data analysis using fuzzy queries.

REFERENCES

1. **Codd E.F.** The Relational Model for Database Management – version 2, reading, MA: Addison-Wesley.
2. **Filatov V.A., Kasatkina N.V., Ponomarenko L.A., Tanyanskiy S.S. 2009.** Postroyeniye sistemi funkcionalnih associativnih pravil na osnove svoystv relyacionnoy modeli dannih / Zbirnik yaukovih prac Viyskovogo institute Kiyivskogo nacionalnogo universitety imeni Tarasa Shevchenka. – Kiyiv. Vip. 23. 34-38.
3. **Filatov V.A., Kasatkina N.V., Tanyanskiy S.S. 2009.** Metodi hraneniya i obrabotki nechetkih dannih v srede relyacionnih system // Avtomatika. Avtomatizaciya. Elektrotehnichni komplekci ta sistemi. Vip. 2(24). 80-86.
4. **Haykin S. 1999.** Neural Networks. A Comprehensive Foundation. – Upper Saddle River, NJ.: Prentice Hall, Inc., 842.
5. **Kasatkina N.V., Ponomarenko L.A., Filatov V.A. 2008.** Informaciyni sistemi ta jih zastosuvannya (monografy). Kiyiv: PP Avers. 142.
6. **Rybytska O., Vovk M. 2014.** An Application of the Fuzzy Set Theory and Fuzzy Logic to the Problem of Predicting the Value of Goods Rests // ECONTECHMOD An International Quarterly Journal On Economics In Technology, New Technologies And Modelling Processes. Vol. 3, №2. 65-69.
7. **Zadeh L.A. 1975.** The concept of a linguistic variable and its application to approximate reasoning // Information Sciences. Vol. 8. 199 - 249.