INTUITIONISTIC FUZZY (IF) EVALUATIONS OF MULTIDIMENSIONAL MODEL

Ivanka Valova
Sofia University, Sofia, Bulgaria

Abstract

There are different logical methods for data structuring, but no one is perfect enough. Multidimensional model-MD of data is presentation of data in a form of cube (referred also as infocube or hypercube) with data or in form of "star" type scheme (referred as multidimensional scheme), by use of F-structures (Facts) and set of D-structures (Dimensions), based on the notion of hierarchy of D-structures. The data, being subject of analysis in a specific multidimensional model is located in a Cartesian space, being restricted by D-structures. In fact, the data is either dispersed or „concentrated", therefore the data cells are not distributed evenly within the respective space. The moment of occurrence of any event is difficult to be predicted and the data is concentrated as per time periods, location of performed business event, etc. To process such dispersed or concentrated data, various technical strategies are needed. The basic methods for presentation of such data should be selected. The approaches of data processing and respective calculations are connected with different options for data representation. The use of intuitionistic fuzzy evaluations [1]- IFE provide us new possibilities for alternative presentation and processing of data, subject of analysis in any OLAP application. The use of IFE at the evaluation of multidimensional models will result in the following advantages: analysts will dispose with more complete information for processing and analysis of respective data; benefit for the managers is that the final decisions will be more effective ones; enabling design of more functional multidimensional schemes.

The purpose of this work is to apply intuitionistic fuzzy evaluations of multidimensional model of data.

Key words: On-line Analytic Processing, Intuitionistic Fuzzy Evaluations.

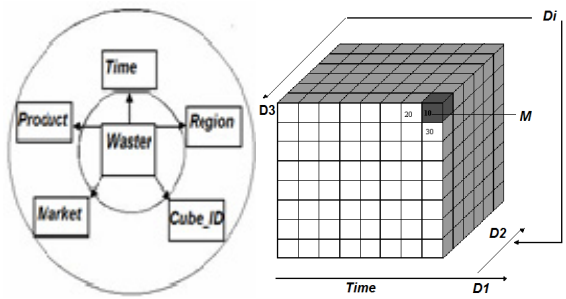
INTRODUCTION

The main task at development of MD model referred as OLAP models is, on the one hand to make the data schemas more understandable for end users and on the other hand to improve the performance in case of queries placed by such users. For such purposes the data schemes are simplified in a manner to contain only most important things (i.e. fact that is to analyzed and respective D-structures, which will be involved in data analysis). These schemas are very close to the concept for performance of data analysis and require the use of specific types of queries, enabling easy adjustment of the system to the needs of users, aiming the obtaining of better response times and producing of result in case of specific query placed.

DESCRIPTION OF THE MODEL

Three different levels of detail have been distinguished in the establishment of a data model, intended for OLAP. On a low level there are measures, which can be grouped into cells if they refer to the same fact (F-structure). The cells further are grouped into different classes, which can be drawn as n-dimensional cubes (on a medium level of detail), owing to which the different dimensions, defining the cube are functionally independent. Eventually, on a high level several cubes representing the same type of fact on different levels of aggregation are grouped into one Fact.

The OLAP model must be able to present a global view of data, including full support for hierarchies of D-structures and multiple hierarchies of data, subject to analysis.



1a) 1b)

Figure 1: Example of OLAP model. 1a) Multidimensional scheme-one F-structure (fact) and several Dimensions (D-structures), 1b) Infocube.

On Fig. 1 is specified example of an OLAP model, where the fact is Wasted product and D-structures are Time, Info Package, Cube_ID, Region, etc. In Fig.1 the Fact is shown in the middle of the scheme, and D-structures (dimensions), being used for analysis of the Fact, are shown around it. The scheme presented in Fig. 1 is also termed "multidimensional scheme" due to the large number of D-structures, which may be involved. The fact in a multidimensional scheme is the object, which contains measures.

Organization of data in a "star" type schema or its representation in form of a cube is very close to the concepts for performance of analysis and facilitates the use of data by end users.

Description of the D-structures

In multidimensional models the information is divided in facts (F-structures) and dimensions (D-structures) [2].

D-structure, (Dimension) – different initial viewpoints at data selection, which will be used during fact analyzing. D-structures contain mainly description attributes.

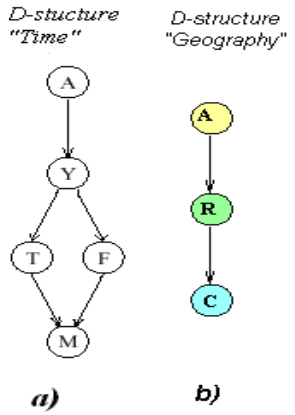


Figure 2: a) D-structure "Time" , A-All, Y- Year, T-three month, F-four month, M –Month b) D- structure "Geography", A-All, R- Region, C-City.

D-structure represents a connected directed graph, provided that each node of the graph corresponds to a given aggregation level, and its arcs reflect "part-whole" relations between the objects within the aggregation levels.

The hierarchy of certain D-structure is linked with determined data consolidation path (Fig. 2). The consolidation path is composed of series of levels or consolidation steps, where each level represents a higher degree of data consolidation. For instance, "business enterprise" may include as consolidation path the levels "business area", "branch", "department", "project", "task" and "employee" [3]. The software product SAP BIW supports 16 D-structures, provided that three of it are mandatory. These are: "Time", "Units", which provide meaning of the values for key indicators and Info packages.

Description of the F-structures

F-structure (Fact) – it represents the data, subject to analysis. The fact contains numerical attributes. The fact in a multidimensional scheme is the object, which contains measures. Measures evaluate attributes of fact [4].

The fact wasted products is evaluated by using the following measures: costs for invested labor, costs for materials used, transportation costs, number of persons involved in the process of wasting and evaluation of the respective products and total costs and other depending on the point of view of the developer and on the endmost needs of analysts and users.

We have evaluated D-structure Product (the degree of preference for whether an article is liked or not), and to make our evaluation more complete it is necessary to consider the Region, where these goods are wasted, the Chain of stores in which the largest quantities have been wasted (and to try to answer whether the wasted goods are a result from poor management), the Vendor who has wasted the goods (if wasted goods are in found in greatest quantity in the same vendor) and if we also consider the Time as a D-structure, we can evaluate in which periods the quantity of wasted goods is the greatest and thus we can evaluate the degree of consumption depending on different seasons.

The subject matter of analysis (fact) is wasted products. Typical D-structures of this application are: product, which is wasted, the store in which it is wasted (e.g. Sofia, Z street) and the date of the respective operation. These D-structures form one hypercube.

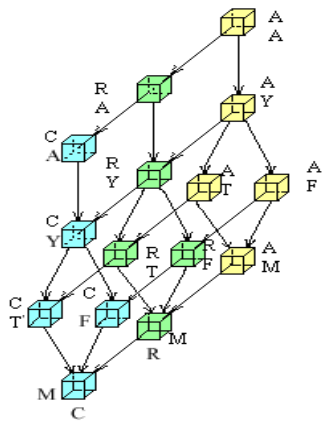


Figure 3: Example of F-structure.

In order to simplify the drawing of model only two D-structures are used – Time Fig. 1a), Geography –Fig. 1b). It is shown on Fig. 3 the structure of Fact, composed of several cubes (boxes). The cubes containing letter C may be used for the purposes of analysis of goods in certain city for different time periods – month, quarter..... year. The cubes containing letter R are used for analysis of goods per regions, again for month, quarter..... year. The box MC shows goods for a month in given city. The cube CT shows certain goods in given city and for time period - quarter. The cube MR reflects the certain goods for a month, but in given region, etc.

If in any F- structure, intended for OLAP, exist All-level and atomic level and the remaining levels are segmented (i.e. contain more than one object or more than one cube), than the graph is a lattice. In the F-structure, intended for OLAP, exists unique level called atomic, which contains elementary objects. Each object at atomic level has one part and this object is called elementary. In D – and F-structures, intended for OLAP, exists a unique level called All- level. This aggregation level has exactly one object.

INTRODUCTION OF IF ASSESSMENTS

We use A_1, A_2,..,.A_n for marking of articles (attributes of certain product). Each symbol may be indicated by lower index for marking of different names from same type.

Each attribute A should be related with definition area dom(A).
Let:
$k(A_i, R_i, T_i)$ – to be the quantity of the objects of the relevant product.
$p(A_i, R_i, T_i)$ – the quantity of the sold articles
$q(A_i, R_i, T_i)$- the quantity of the waste articles.
T_i, R_i – specify time and location of the relevant article

$$\{ \langle A_i, R_i, T_i \rangle, \mu(A_i, R_i, T_i), \nu(A_i, R_i, T_i) \mid \langle A_i, R_i, T_i \rangle \in E \}$$

π - the produced articles, but still remaining unsold, i.e. some of them may be scrapped, therefore there is indefiniteness.
μ - produced and sold articles
$V_\nu(All)$ - gives the correlation of the waste articles towards the overall production.

$$V_\nu(All) = \frac{\sum_{i=1}^{n} q(A_i, R_i, T_i)}{\sum_{i=1}^{n} k(A_i, R_i, T_i)} \quad (1)$$

$V_\nu(A_i, R_i, T_i)$ - assigns the degree of the waste articles.

$$V_\nu(A_i, R_i, T_i) = \frac{q(A_i, R_i, T_i)}{k(A_i, R_i, T_i)} \quad (2)$$

$V_\mu(All)$ - give preference to the relevant article as a whole

$$V_\mu(All) = \frac{\sum_{i=1}^{n} p(A_i, R_i, T_i)}{\sum_{i=1}^{n} q(A_i, R_i, T_i)} \quad (3)$$

$V_\mu(A_i, r_i, t_i)$ – assigns the degree of preference, whether certain article is preferred or not.

$$V_\mu(A_i, R_i, T_i) = \frac{p(A_i, R_i, T_i)}{k(A_i, R_i, T_i)} \quad (4)$$

From (1) and (3) it follows that

$$V_\mu(All) + V_\nu(All) = \frac{\sum_{i=1}^{n} p(A_i, R_i, T_i) + q(A_i, R_i, T_i)}{\sum_{i=1}^{n} k(A_i, R_i, T_i)} \le \frac{\sum_{i=1}^{n} k(A_i, R_i, T_i)}{\sum_{i=1}^{n} k(A_i, R_i, T_i)} = 1 \quad (5)$$

Therefore $\langle V_\mu(All), V_\nu(All) \rangle$ is an IF estimation.

From (2) and (4) it follows that

$$V_\nu(A_i, R_i, T_i) + V_\mu(A_i, R_i, T_i) = \frac{q(A_i, R_i, T_i) + p(A_i, R_i, T_i)}{k(A_i, R_i, T_i)} \le \frac{k(A_i, R_i, T_i)}{k(A_i, R_i, T_i)} = 1 \quad (6)$$

for each A_i, R_i, T_i.
Therefore $\langle V_\mu(A_i, R_i, T_i), V_\nu(A_i, R_i, T_i) \rangle$ is an IF estimation.

For evaluation of the model we can use Fig. 4. We introduce β, α - Fig. 4. If the aggregate evaluation falls within the trapezoid αODB we have the optimal variant. In case that the respective evaluation falls within the trapezoid αECA, it serves as a sign that the production is ineffective. If the aggregate evaluation falls within the triangle EOD, it means that the production output exceeds the minimum and is sold above the minimum, and on other side the goods become slow-moving ones, which serves as indicator that it is produced more than needed.

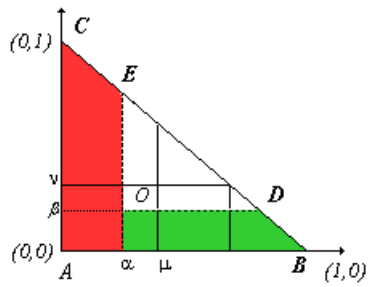


Figure 4: Intuitionistic fuzzy evaluations.

The use of IFE at the evaluation of multidimensional models will result in the following advantages: analysts will dispose with more complete information for processing and analysis of respective data; benefit for the managers is that the final decisions will be more effective ones; enabling design of more functional multidimensional schemes.

CONCLUSIONS

In this paper we represented the use of IFL in assessment of the MD models for analytical data processing.

REFERENCES

[1] Atanassov K., Intuitionistic Fuzzy Sets, Springer, Heidelberg, 1999.
[2] Valova I., M. Noirhomme-Fraiture, Processing of Large Data Sets: Evolution, Opportunities and Challenges, PCaPAC'08, Ljubljana, Slovenia, 2008, 198-200.
[3] Codd E., S. B. Codd, C. T. Salley. Providing OLAP to user-analysts: An IT mandate. Technical report, E. F. Codd & Associates, 1993.
[4] Weidner Jens -Integrierte Data Mining Lösung mit SAP BI, (SAP BW3.5 / SAP CRM4.0), CRM Marketing & Analytics SAP.
[5] Kimball R., Data Warehouse Tool Kits. John Wiley & Sons, Toronto, 1996.