Tenth Int. Conf. on IFSs, Sofia, 28-29 Oct. 2006
NIFS Vol. 12 (2006), No. 3, 46-54

The cubes' magic and their application in different fields of science

Ivanka Valova
ICSR, BAS, Acad. G. Bonchev Str., P.O. Box 79, Sofia 1113, Bulgaria
e-mail: vania@icsr.bas.bg

Abstract: Subject of this paper is to study the use of a common geometric figure CUBE in various fields of science – fuzzy sets and neural systems, multiprocessor systems and in multidimensional models for analytical data processing. It is shown also how such scientific sectors use some concepts as one-dimensional cube, two-dimensional cube, etc. We present further a classification of software products depending on way of data storage in one-cube or multi-cube structure. An approach to analysis of multidimensional models in OLAP systems as outlined in this paper is application of Intuitionistic Fuzzy Evaluations (IFE), aiming to provide new functional possibilities to developers of systems for on-line analytical data processing and business analysts by utilization of IF evaluations.
Key words: fuzzy sets and neural systems, multiprocessor systems, multidimensional models, IF estimations

1 Introduction

Cube is one of the most frequently used notions in On-line Analytical Data Processing (OLAP) systems (infocube, hypercube, one-dimensional cube, two-dimensional cube, etc., up to N-dimensional cube). In the course of our study on applications of such geometric figure in the systems for analytical data processing and respective operators, used for analysis of multidimensional models, we resolved to step aside and to study the use of this geometric figure also in other modern sciences. In this paper we investigate how the various scientific branches are interpreting and using this common geometric figure.

As a basis for our discussion we assume the following definition:
Hypercube – an indirect graph with the number of vertices always equal to the exact power of 2 (For n-dimensional space this number is 2^n). If n=1 the graph may be presented as a segment on a straight line with coordinates of the ends (0) and (1). If n=2 we have a square with the following coordinates (0,0), (0,1), (1,1) and (1,0). If n=3 again we have the respective interpretation with a cube in space, etc.
Hypercube (n-dimensional cube) is the respective generalization for the n-dimensional space; in this case the coordinates will be placed at n-tuples between 0 and 1 such that if vertices i and j are adjacent, then their respective n-tuples will differ in exactly one position.

The paper has the following structure: first, the use of cubes in neural networks and fuzzy logic systems is discussed; further the issue of processors' linkage using the topology of type "hypercube" is studied; further some issues are discussed related with the calculation ad maintenance of the cube, used in OLAP systems; and finally, a classification of software products depending on multidimensional data structure used is presented.

2 Using cubes in neural networks and fuzzy sets

The term "hypercube" is used in both neural networks and fuzzy systems. Without much detail we will focus the discussion on main points of interest, related with the discussed term.

A set of n-neurons defines a family from fuzzy sets (FS). At each instant the n-vector of the neuronal outputs defines a fuzzy unit (fit vector).

The set of all possible neural outputs corresponds to the set of all n-dimensional fit vectors (the fuzzy power set). The set of all vectors of length n and with coordinates in the unit interval $I= [0, 1]$ *defines of unit hypercube* $I^n = [0, 1]^n = [0, 1] \times .... \times [0, 1]$.

The vertices of unit hypercube I^n, which are 2^n, represent neuronal-output combinations. Many neural networks move initial sets inside the unit cube to nearest vertices. The vertices present the nonfuzzy set of the n elements. The bit value 0 in the i-th position of a bit vector indicates the absence of elements x_i in that subset. The bit value 1 indicates the presence of x_i in the subset. The bit vector (1 0 1 0 0) indicates the subset $\{x_1, x_2\}$ of set $\{x_1, x_2, x_3, x_4, x_5\}$.

Let the set X consists two elements x_1 and x_2. Therefore the set denoted as 2^x contains the four subset - $\{\emptyset, \{x_1\}, \{x_2\}, X\}$. The four non-fuzzy subsets -$\emptyset$ = (0 0); $\{x_1\}$ = (1 0); $\{x_2\}$ = (0 1); X = (1 1) in nonfuzzy set 2^x correspond to the four corners of the two-dimensional cube. Fuzzy set F (2^x) of X corresponds to the unit square (Fig.1).

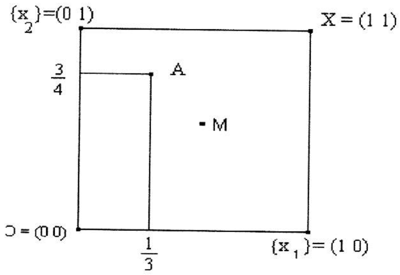


Fig.1. Fuzzy set F (2^x) of X corresponds to the unit square (two-dimensional cube). X contains x_1, x_2 and 4 non fuzzy subsets.

Fuzzy set F (2^x) of X corresponds to the unit cube (Fig.2) when X = $\{x_1, x_2, x_3\}$. Fuzzy subset B = (1/4, 1/2, 1/3) contains infinitely many fuzzy subsets and they define the shaded hyper rectangle. $S(A, B) < 1$, since A lies outside the hyper rectangle. B* denotes the subset of B closest to A. $B^* = (1/4, 1/3, 1/6)$, which also equals the pair wise minimum of A and B. $B^* = A \cap B$

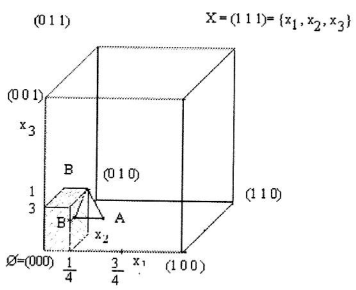


Fig.2. Fuzzy set F (2^x) of X corresponds to the unit cube [4]. X contains x_1, x_2, x_3 and 8 non fuzzy subsets

Subesethood (S), reveals the relation between fuzziness and randomness. Randomness equals the uncertainty that arises when a non-fuzzy set B is partially contained in one of its own nonfuzzy subset A.

$S(A, B) = 1$, but $0 < S(B, A) < 1$

The properties of the middle point of the cube are remarkable ones. It possesses maximal fuzzy entropy. The middle point of the cube corresponds to the maximal fuzzy set M (M equals its own opposite M^c) and is valid the following relationship: $M = M \cap M^c = M \cup M^c = M^c$, which maximally violates the bivalent laws of non-opposition and exclusion of median values. The phenomenon of middle point is related with the classical paradoxes of logic and theory of sets, but it is not a subject of discussion in this article.

When discussing FS issues we could not skip the research made in the field of Intuitionistic Fuzzy Set - IFS [1]. IFS is defined as follow:

$A=\{\langle x, \mu_A(x), \nu_A(x)\rangle / x \in E\}$,

Where E is fixed set, functions $\mu_A:E \to [0,1]$ and $\nu_A:E \to [0,1]$ give degree of membership and non-membership of the element $x \in E$ to set A. Set A is subset to E and $\forall\ x \in E$: $0 \le \mu_A(x)+\nu_A(x) \le 1$.

Value $\pi_A(x)=1-\mu_A(x)-\nu_A(x)$ gives the degree of non-determinacy of the element x:E to the set A.

Very useful for our work were the studies of Eulaila Szmidt and Jim F. Baldwin [5] and all other papers, presented at the Eight International Conference, dedicated on IFS issues. The triangle in Fig.3 represents a plane, which elements may be described by three coordinates (μ, ν, π), i.e. each element belonging to IFS may be presented as a point, which belongs to the triangle. The point with coordinates (1,0,0) represents elements belonging to IFS, and the point with coordinates (0,1,0) represents elements not belonging to IFS. It is interesting the behavior of point with coordinates (0,0,1), and after the respective transformations from 3-dimensional space into 2-dimensional one, it is projected in a point with coordinates (0,0). For further in-depth knowledge and use by the future specialists in this field we recommend the work [2].

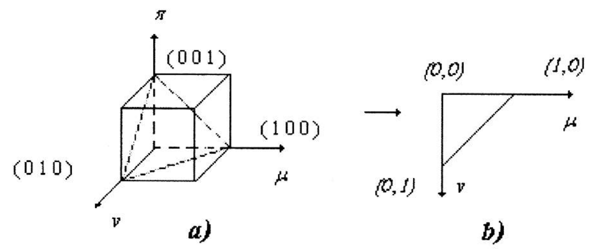


Fig.3. IFS- transformations from 3-dimensional space into 2-dimensional one

3 Hypercube topology

In this section we will discuss the manner of usage in multiprocessor systems of such notions as one-dimensional, two-dimensional... n-dimensional cube. The application of FS and IFS in this area would be very useful in process of development of respective algorithms. We will focus here our attention on architecture schemes, and software interpretation will be subject of our further research papers.

In a Hypercube interconnection topology, the number of processor is 2^n. For example, a hypercube with dimensions - n=3 (fig.4c) will have eighth processors numbered 000, 001, 010, 011, 100, 101, 111. In this kind of topology each processor has a number whose binary representation has ***n*** bits. The interconnection scheme is simple. Processors with binary number *i* will have direct connections to all processors with binary number j, such that j differs from *i* in exactly one binary digit. The processor 000 (fig.4c) has direct connections to the following processors: 001, 010,100 [3]. Also, the distance between processors is equal to the number of bit positions in which their processor numbers differ. For example processor 000 is two steps from processor 011. For processor 000 to communicate with 011, it will first send the message to its neighbour 001, which will forward the message to its neighbour 011. The total communication from 000 to 011 takes two steps. For n=4, each processor will have four immediate neighbors and the maximum distance between processors is four steps (fig.5& fig.6).

By constructing a hypercube with dimensions n+1, the following recursive sequence. First, an exact copy of a hypercube with dimensions n should be created, including the number of processors; then, a direct link should be created between processors with the same numbers in the copy and the original; finally, a binary 1 should be attached on the left of every processor number in the copy or a binary 0 on the left of every processor number in the original. This recursive construction rule will be useful as an aid in designing efficient algorithms for execution on Hypercube multiprocessors [6].

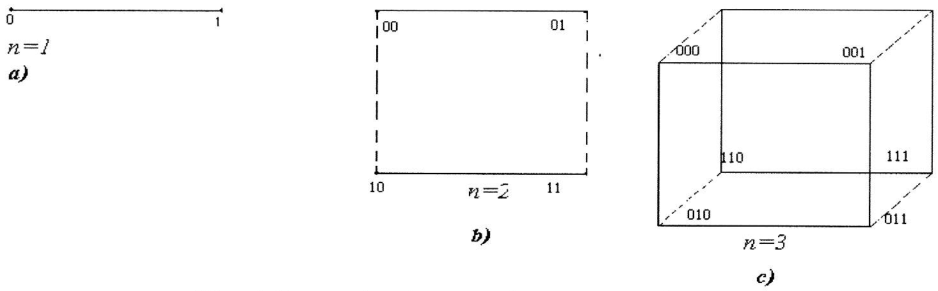


Fig. 4.Recursive hypercube construction

In Figures 5 and 6 a 4-dimensional cube is shown as constructed using 3-dimensional cubes connected in the respective vertices.

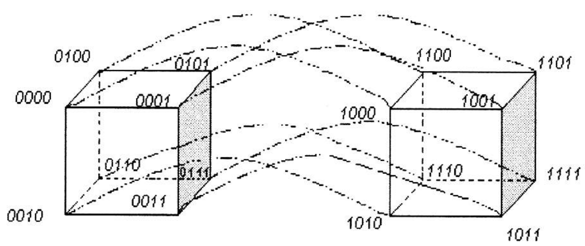


Fig. 5. 4D-hypercube

A five –dimensional cube can be made by cloning the structure of fig.5 and connecting corresponding vertices to form a block of four cubes. To go to six-dimensional cube the block of four cubes can be replicated and corresponding vertices have to connect.

A 4-dimensional cube may be presented as two cubes (one placed in the other) – Fig. 6 regardless of the manner of drawing, again the rule for the number of connections between processors is applied. We should not exclude the option to use four coordinate axes to draw a 4-dimensional cube.

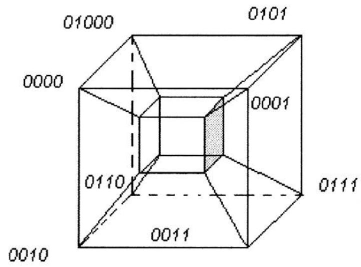
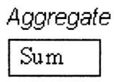


Fig. 6. Four-dimensional hypercube

4 Hypercubes in analytical data processing systems.

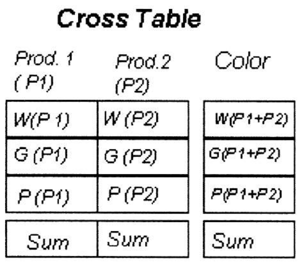
Many of the data visualization and analysis instruments in OLAP systems present the data set as N-dimensional space. Visualization instruments provide 2-dimensional and 3-dimensional sub-planes of such space in the form of 2D or 3D objects.

Figure 7 shows the concept for aggregation to 3-D structures. The GROUP BY operator generates an N-dimensional nucleus of a data cube. We have the lower N -1-dimensional set when outside the data nucleus of the cube there are points, lines, planes, cubes or hypercubes. The operator of the data cube – *Cube*, constructs a table comprising aggregated values.



A 0-dimensional cube in OLAP systems represents one cell (one entry).

A 1-dimensional cube in analytical data processing systems is a column of cells. In the example, the Group by operator was used to group products by colors and the Sum operator was used to determine their total value.



A 2-dimensional cube in cube in analytical data processing systems is a data table, also referred to as a cross table.

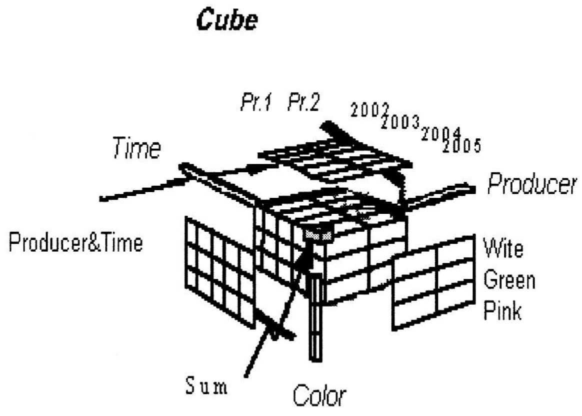


Fig.7 3D cube for OLAP systems is a cube with three 2D cross tables [7].

The CUBE operator is supported for the first time by SQL Server 6.5. In order to make possible an easy and dynamic cube update, the users of such systems define triggers for the connected tables so that is the tables change, the data in the cube is dynamically updated.

There are still problems related both to the calculation of the cube and to its maintenance. The main problems occur depending on the type of function. It is easy to calculate the maximum value in a cube because *max* is a distribution function. However, any of the functions *delete* or *update* changes the greatest value in the basic table, then the 2^N elements of the cube should be re-calculated and the new global maximum should be found. Therefore, the whole cube should be re-calculated. The *max* function is distributive for SELECT and INSERT, but holistic one for DELETE.

To exhaust the issue, we will present a classification of software products depending on the manner in which data us structured.

Hypercubes – These products have a one-cube logistic structure. The basic model for compression of scattered data is quite complicated. The data values are introduced for each combination of D-structure members. Representatives of this type of multi-dimensional structures include Essabase (version 4.1), FCS-Multi. Hyperion Enterprise and Climer Comshare FDC have a similar structure. Additional D-structures may be added to the one-cube system.

Multicubes – In these products the database is segmented in one set of multi-dimensional structures and each structure comprises a subset of the common D-structure sub-cubes. These structures are more flexible and more efficient in processing scattered data. Software products of this class include Microsoft OLAP Services, Express, TM1, SAP BIW, etc.

Creation of new programming languages and new mathematical algorithms for processing of multidimensional structures has been subject of intensive research in recent years [8,9].

In this paper we will not analyze the mathematical model itself, but we will briefly discuss how IFS may be used for analysis of multidimensional model and how figure 3b will be introduced for use.

The data, being subject of analysis in a specific OLAP application is located in a Cartesian space, being restricted by D-structures of the multidimensional model. 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 business 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 IFE provide us new possibilities for alternative presentation and processing of data, subject of analysis in any OLAP application.

We use $A_1, A_2, \ldots, A_n$ 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 article.

$p(A_i, r_i, t_i)$ – 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

$\{ \langle A_i, r_i, t_i \rangle, \mu(A_i, r_i, t_i), \nu(A_i, r_i, t_i) \, | \langle A_i, r_i, t_i \rangle \in E \}$

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.

$$V_\nu(All) = \frac{\sum_{i=1}^{n} q(A_i, r_i, t_i)}{\sum_{i=1}^{n} k(A_i, r_i, t_i)} \qquad (1)$$

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

$$V_\nu(A_i, r_i, t_i) = \frac{q(A_i, r_i, t_i)}{k(A_i, r_i, t_i)} \qquad (2)$$

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

$$V_\mu(All) = \frac{\sum_{i=1}^{n} p(A_i, r_i, t_i)}{\sum_{i=1}^{n} q(A_i, r_i, t_i)} \qquad (3)$$

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

$$V_\mu(A_i, r_i, t_i) = \frac{p(A_i, r_i, t_i)}{k(A_i, r_i, t_i)} \qquad (4)$$

From (1) and (3) it follows that

$$V_\mu(All) + V_\nu(All) = \frac{\sum_{i=1}^{n} p(A_i, r_i, t_i) + q(A_i, r_i, t_i)}{\sum_{i=1}^{n} k(A_i, r_i, t_i)} \leq \frac{\sum_{i=1}^{n} k(A_i, r_i, t_i)}{\sum_{i=1}^{n} k(A_i, r_i, t_i)} = 1 \qquad (5)$$

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 $< V_\mu(A_i, r_i, t_i), V_\nu(A_i, r_i, t_i) >$ is an IF estimation.

For evaluation of the model we can use Fig. 3.b.

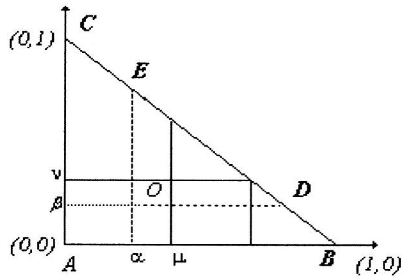


Fig. 9. Intuitionistic Fuzzy Evaluations

We introduce β, α - fig. 9. 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.

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.

5 Conclusions and Future Work

We discussed in this paper the most important aspects pertaining to use of hypercubes in various fields of science. In fact, the cubes' magic nowadays is reigning in a unique way over the trends in business applications and theory across different scientific branches.

More profound research is necessary in the field of OLAP systems related to the calculation and maintenance of hypercubes, creation of new program languages and last but not least unification of the different terminology used by the different software products.

References

[1] Atanassov K., Intuitionistic Fuzzy Sets, Springer, Heidelberg, 1999.

[2] Atanassov K., Remark on a property of the intuitionistic fuzzy interpretation triangle, Notes on Intuitionistic Fuzzy Sets, Vol. 8, No. 1, 34-36, 2002.

[3] Coffin Michel H., Parallel programming a new approach, Prentice Hall, New York, 1992.

[4] Kosko, B., Fuzziness vs. Probability, International Journal of General Systems, Volume 17, Numbers 2-3, Gordon and Breach Science Publishers, 1990.

[5] Szmidt E., J. Baldwin Entropy for Intuitionistic Fuzzy Set Theory and Mass Assignment Theory, Eighth International Conference on IFSs –part 1, June 20-21, Varna, Bulgaria, 15-28, 2004.
[6] Tanenbaum Andrew, Modern Operating Systems, Prentice Hall, 1992.
[7] Technical Report – www.microsoft.com
[8] Valova I., Application of IF logic in systems for analytical data processing, Eighth International Conference on IFSs –part 2, 20-21 June 2004, Varna, Bulgaria, 70-74.
[9] Valova I., Zhechev B., Valov V. Overview of research and software approaches for multidimensional data analysis, Cybernetics and information technologies, vol.4, No1, IIT-BAS, Sofia, 2004, 26-34.