Eighth Int. Workshop on GNs, Sofia, 26 June 2007, 66-72

Generalized net model of binary operations over intuitionistic fuzzy OLAP cubes

P. Chountas[1], B. Kolev[2], V.Tasseva[2], K. Atanassov[2]

[1]University of Westminster, School of Computer Science, DKM Group London, UK, chountp@westminster.ac.uk

[2] Centre for Biomedical Engineering, Bulgarian Academy of Sciences, Sofia, Bulgaria
{bik, violeta}@clbme.bas.bg, krat@bas.bg

Abstract

We present our approach for extending the OLAP model to include treatment of value uncertainty as part of a multidimensional model inhabited by concepts and non-rigid hierarchical structures of organization. New multidimensional-cubic operators defined over the IF-Cube are introduced. The emphasis is on operating over data with imprecision either in the facts or in the dimensional attributes organised in the form of kind of hierarchies.

1. Introduction

We introduce the semantics of the Intuitionistic fuzzy cubic representation [6], [7], [8], [9] in contrast to the basic multidimensional-cubic structures [1], [2], [3], [4], [5] and model its functionality with the aid of Generalised Net (GN; see [10]) models. The *union*, *intersection*, *difference* and *join* cubic operators are extended and enhanced with the aid of Intuitionistic fuzzy Logic. With the aid of Intuitionistic fuzzy cubic slices represented fact tables we show how to apply the operations over the multidimensional structure.

An **IFS cube** is an abstract structure that serves as the foundation for the multidimensional data cube model. A cube C is defined as a five-tuple *(D, l, F, O, H)* where:

- *D* is a set of dimensions.
- *l* is a set of levels $l_1, \ldots, l_n,$
    - A dimension $D_i = (l \leq O, l_\perp, l_\top)$ $dom(D_i)$ where $l = l_i$ $i=1 \ldots n.$
      l_i is a set of values and $l_i \cap l_j = \emptyset$,
      $\leq O$ is a partial order between the elements of *l*.
      To identify the level *l* of a dimension, as part of a hierarchy, we use *dl*.
      $l_\perp$: base level, $l_\top$: top level
      for each pair of levels l_i and l_j we have the function

$$\mu_{ij} : l_i \times l_j \rightarrow [0,1]$$
$$\nu_{ij} : l_i \times l_j \rightarrow [0,1]$$
$$0 < \mu_{ij} + \nu_{ij} < 1$$

- F is a set of fact instances with schema $F = \{<x, \mu_F(x), \nu_F(x)> | x \in X\}$, where $x=<att_1, ..., att_n>$ is an ordered tuple belonging to a given universe X, $\{att_1, ..., att_n\}$ is the set of attributes of the elements of X, $\mu_F(x)$ and $\nu_F(x)$ are the degree of membership and non-membership of x in the fact table F respectively.
- H is an object type history that corresponds to a cubic structure (l, F, O, H') which allows us to trace back the evolution of a cubic structure after performing a set of operators i.e. aggregation.

In this paper the fundamental set cubic operators *union*, *intersection*, *difference* and *join* for navigating through the IFS cube are defined. Its functionality is modelled with the aid of GNS .

***Cubic Product* ($\otimes$):** This is a binary operator $C_{i1} \otimes C_{i2}$. It is used to relate two cubes C_{i1} and C_{i2} assuming that $D_1 \subseteq D_2$ and O_1, O_2 are reconcilable partial orders. Thus, l_1, l_2 could lead to l_o being a ragged hierarchy.

Input: $C_{i1} = (D_1, l_1, F_1, O_1, H_1)$ and $C_{i2} = (D_2, l_2, F_2, O_2, H_2)$
Output: $C_o = (D_o, l_o, F_o, O_o, H_o)$ where
$D_o = D_1 \cup D_2$, $l_o = l_1 \cup l_2$, $O_o = O_1 \cup O_2$ $H_o = H_1 \cup H_2$,
$F_o = F_1 \text{ X } F_2$
$=\{<<x, y>, \min(\mu_{f1}(x), \mu_{f2}(y)), \max(\nu_{f1}(x), \nu_{f2}(y),)>|<x, y> \in X \times Y\}$

Mathematical notation: $C_{i1} \otimes C_{i2} = C_o$

***Join* (Θ):** The join operator relates two cubes having one or more dimensions in common, and having identical mappings from common dimensions to the respective attribute sets of these dimensions.
This operation can be expressed using Cubic Product operation. $C_{i1} = (D_1, l_1, F_1, O_1, H_1))$ and $C_{i2} = (D_2, l_2, F_2, O_2, H_2)$ are candidates to join if $D_1 \cap D_2 \neq 0$

Input: $C_{i1} = (D_1, l_1, F_1, O_1, H_1)$ and $C_{i2} = (D_2, l_2, F_2, O_2, H_2)$
Output: $C_o = (D_o, l_o, F_o, O_o, H_o)$

Mathematical notation: $C_{i1} \,\Theta\, C_{i2} = \sigma_p(C_{i1} \otimes C_{i2})$

***Union* ($\cup$):** The union operator is a binary operator that finds the union of two cubes. C_{i1} and C_{i2} have to be union compatible. The operator also coalesces the value-equivalent facts using the minimum membership and maximum non-membership.

Input: $C_{i1} = (D_1, l_1, F_1, O_1, H_1)$ and $C_{i2} = (D_2, l_2, F_2, O_2, H_2)$
Output: $C_o = (D_o, l_o, F_o, O_o, H_o)$ where
$D_o = D_1 = D_2$, $l_o = l_1 = l_2$, $O_o = O_1 = O_2$, $H_o = H_1 = H_2$,
$F_o = F_1 \cup F_2 = \{ < x, max(\mu F_1(x), \mu F_2(x)), min(\nu F_1(x), \nu F_2(x)) > | x \in X \}$

Mathematical notation: $C_{i1} \cup C_{i2} = C_o$

***Difference* (-):** The difference operator is a binary operator that the difference of two cubes. It is similar to the difference operator in relational algebra. C_{i1} and C_{i2} have to be union compatible. The difference operator removes the portion of the cube C_{i1} that is common to both cubes.

Input: $C_{i1} = (D_1, 1_1, F_1, O_1, H_1)$ and $C_{i2} = (D_2, l_2, F_2, O_2, H_2)$
Output: $C_o = (D_o, l_o, F_o, O_o, H_o)$ where

$$D_o = D_1 = D_2, \quad l_o = l_1 = l_2, \quad O_o = O_1 = O_2, \quad H_o = H_1 = H_2,$$
$$F_o = F_1 \cap F_2 = \{ < x, \min(\mu F_1(x), \mu F_2(x)), \max(\nu F_1(x), \nu F_2(x)) > \mid x \in X \}$$

Mathematical notation: $C_{i1} - C_{i2} = C_o$

In the next section a GN model for the simulation needs of a generic set cubic operations is constructed and discussed

2. Generalized net model

The constructed GN-model describes a general approach for performing of the operations *union*, *intersection*, *difference* and *join* over two cubes using a sort-merge algorithm. Let the cube key be an ordered n-tuple of all dimension attributes, i.e. the key is defined over $dom_{dim(1)} x\ dom_{dim(2)} x \ldots x\ dom_{dim(n)}$. In other words, a key value is an ordered n-tuple of the coordinates of a cube cell. Let a "<" relation be defined over the key domain and the cubes are union compatible. Then a comparison between key values from different cubes is possible. The sort-merge algorithm iterates through the cells of the both cubes in ascending order of key values. At each step the algorithm moves to the next cell of that cube, whose current key value is less than or equal to the current key value of the other cube. Thus the algorithm passes through all matching and all non-matching pairs of key values. Depending on the type of operation, it does different things on each step (assuming that KV_1 is the current key value of the first cube, KV_2 is the current key value of the second cube):

OP type	$KV_1 < KV_2$	$KV_1 > KV_2$	$KV_1 = KV_2$
Union	Add the cell to the result cube	Add the cell to the result cube	Add the cell to the result cube and calculate its IF degree of membership
Intersection	-	-	Add the cell to the result cube and calculate its IF degree of membership
Difference	-	If on the previous step $KV_1 < KV_2$, add the cell to the result cube	Add the cell to the result cube and calculate its IF degree of membership
Join	-	-	Add the pair of cells to the result cube and calculate its IF degree of membership

Table 1. Operation types and the corresponding actions

In the constructed GN-model (Figure 1) the two operand cubes of the binary operation are represented by different types of tokens – α_1 and α_2. They enter the net through places l_1 and l_2 with initial characteristics the cubes' data. Token α_0, representing the result of the operation, enters the GN-model in place l_3. Initially it has an empty characteristic. One β token, representing the type of the operation, enters the GN in place l_4. The intuitionistic fuzzy estimation criterion is represented by a γ token, which enters the net in place l_5.

Transition Z_1 corresponds to the sort-merge algorithm. The α_1 token iterates through place l_6 for each key value of the first cube and leaves the GN-model through place l_8 when the iteration finishes. Analogously, the α_2 token iterates through place l_7 for each key value of the first cube and leaves the model through place l_9 when the iteration finishes. The β and the γ tokens leave the net through places l_{10} and l_{11} respectively.

On each step, the current key values of α_1 and α_2 tokens (KV_1 and KV_2), which are currently in places l_6 and l_7, are compared and the α_0 token moves:

- to l_{12}, if $KV_1 = KV_2$;
- to l_{13}, if $KV_1 < KV_2$;
- to l_{14}, if $KV_1 > KV_2$.

The characteristic functions of l_{12}, l_{13} and l_{14} may add a cube cell to the characteristic of the α_0 token, depending on the operation type, as described in Table 1.

Transition Z_1 ahs the following form:

$$Z_1 = \langle\{l_1, l_2, l_3, l_4, l_5, l_6, l_7, l_{21}\}, \{l_6, l_7, l_8, l_9, l_{10}, l_{11}, l_{12}, l_{13}, l_{14}, l_{15}\}, r_1\rangle,$$

where:

$r_1=$

	l_6	l_7	l_8	l_9	l_{10}	l_{11}	l_{12}	l_{13}	l_{14}	l_{15}
l_1	true	false	false	false	false	false	false	false	false	false
l_2	false	true	false	false	false	false	false	false	false	false
l_3	false	false	false	false	false	false	W_{3_12}	W_{3_13}	W_{3_14}	false
l_4	false	false	false	false	true	false	false	false	false	false
l_5	false	false	false	false	false	true	false	false	false	false
l_6	W_{6_6}	false	W_{6_8}	false	false	false	false	false	false	false
l_7	false	W_{7_7}	false	W_{7_9}	false	false	false	false	false	false
l_{21}	false	false	false	false	false	false	W_{21_12}	W_{21_13}	W_{21_14}	W_{21_15}

,

where:

W_{3_12} = "The current key values of the both cubes are equal",
W_{3_13} = "The current key value of the first cube is less then the one of the second cube",
$W_{3\_14} = \neg W_{3\_12} \wedge \neg W_{3\_13}$,
W_{6_6} = "There are still cells from the first cube to iterate through",
$W_{6\_8} = \neg W_{6\_6}$,
W_{7_7} = "There are still cells from the second cube to iterate through",
$W_{7\_9} = \neg W_{7\_7}$,
$W_{21\_12} = W_{3\_12}$,
$W_{21\_13} = W_{3\_13}$,
$W_{21\_14} = W_{3\_14}$,
$W_{21\_15} = \neg (W_{6\_6} \vee W_{7\_7})$.

α_1 and α_2 tokens extend their characteristics in places l_6 and l_7 with information about current status of the cubes. α_1, α_2, β and γ tokens do not obtain any new characteristics in places l_8, l_9, l_{10} and l_{11}. Entering places l_{12}, l_{13} and l_{14} tokens obtain respectively the following characteristics:

"$KV_1 = KV_2$",
"$KV_1 < KV_2$",
"$KV_1 > KV_2$".

The received in place l_{15} characteristic is:

"result of the applied binary operation over cubes α_1 and α_2".

Transition Z_2 takes into account the estimation criteria and lets the α_0 token move from l_{12} to L_{17}, L_{18}, L_{19} or L_{20} depending on the formula for calculating the degree of membership of a cell, which participates in both cubes. The considered formulas are:

$$(\mu_1\mu_2,\ \nu_1 + \nu_2 - \nu_1\nu_2), \tag{1}$$
$$(\min(\mu_1, \mu_2),\ \max(\nu_1, \nu_2)), \tag{2}$$
$$(\max(\mu_1, \mu_2),\ \min(\nu_1, \nu_2)), \tag{3}$$
$$(\mu_1 + \mu_2 - \mu_1\mu_2,\ \nu_1\nu_2). \tag{4}$$

Transition Z_2 has the form:

$$Z_2 = \langle\{l_{12}\}, \{l_{16}, l_{17}, l_{18}, l_{19}\}, r_2\rangle,$$

where:

$$r_2 = \begin{array}{c|cccc} & l_{16} & l_{17} & l_{18} & l_{19} \\ \hline l_{12} & W_{12\_16} & W_{12\_17} & W_{12\_18} & W_{12\_19} \end{array},$$

where:

W_{12_16} = "The estimation criterion specifies formula (1) for calculation of the degree of membership of a cell",
W_{12_17} = "The estimation criterion specifies formula (2) for calculation of the degree of membership of a cell",
W_{12_18} = "The estimation criterion specifies formula (3) for calculation of the degree of membership of a cell",
W_{12_19} = "The estimation criterion specifies formula (4) for calculation of the degree of membership of a cell".

The tokens obtain as new characteristics in places l_{16}, l_{17}, l_{18} and l_{19} the estimated on the current step degrees of membership.

The form of transition Z_3 is the following:

$$Z_3 = \langle\{l_{16}, l_{17}, l_{18}, l_{19}\}, \{l_{20}\}, r_3\rangle,$$

where:

$$r_3 = \begin{array}{c|c} & l_{20} \\ \hline l_{16} & true \\ l_{17} & true \\ l_{18} & true \\ l_{19} & true \end{array}.$$

The tokens do not obtain new characteristics in place l_{20}, they just keep the obtained on the previous step characteristics.

$$Z_4 = \langle \{l_{13}, l_{14}, l_{20}\}, \{l_{21}\}, r_4 \rangle,$$

where:

$$r_4 = \begin{array}{c|c} & l_{21} \\ \hline l_{13} & true \\ l_{14} & true \\ l_{20} & true \end{array} .$$

The α_0 token returns to transition Z_1 via transition Z_4 and place l_{21} in order to process the next cell. It does not obtain any new characteristics in place l_{21}. When the algorithm finishes, the resulting α_0 token leaves the GN through place l_{15}.

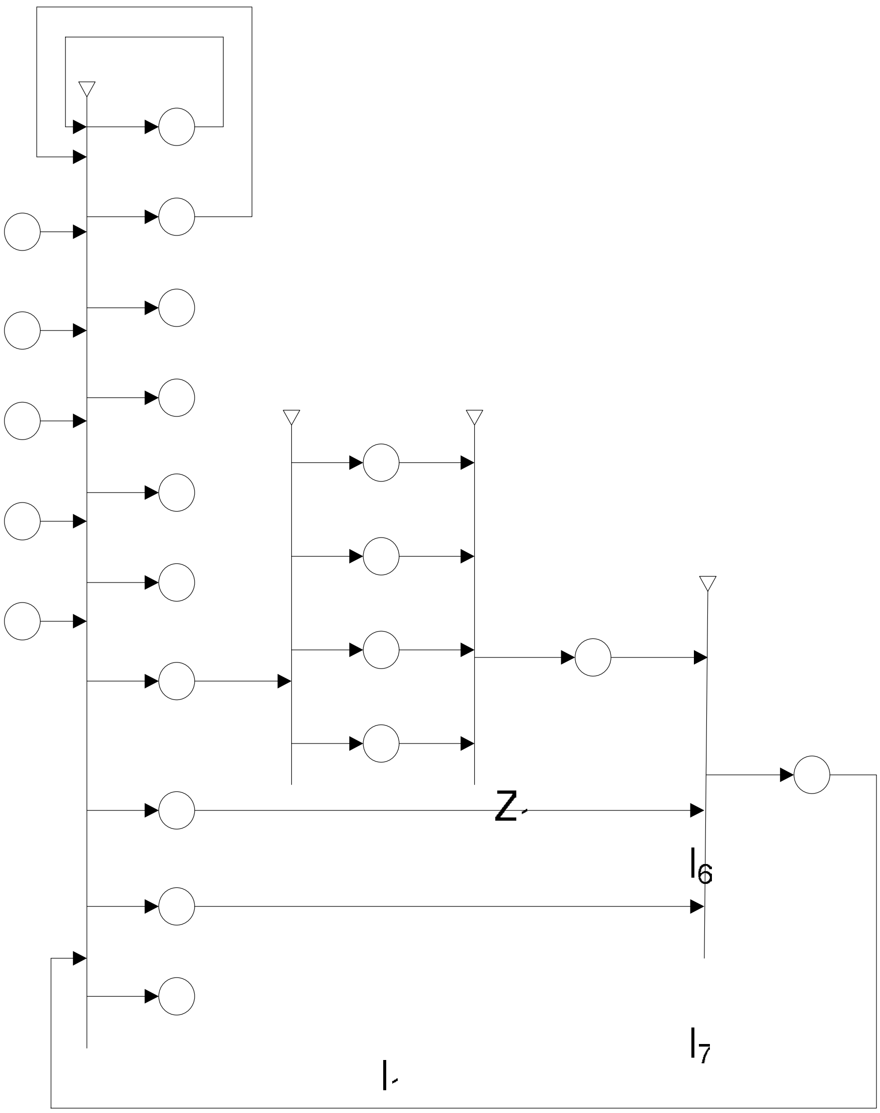

Figure 1. GN Model of binary operations over IF-cubes

3. Conclusion

In this paper we have presented the definition and functionality of the multidimensional-set based operations defined over the IF-cube, with the aid of Generalised Net models. The constructed GN-model describes a general approach for performing of the operations *union*, *intersection*, *difference* and *join* over cubes using a sort-merge algorithm. The main contribution of this new model is that is able to operate over data with imprecision in the facts and the summarisation hierarchies. Classical models impose a rigid structure that made the models to be difficult to merge information from different but still reconcilable sources.

References

[1] Codd, E., "Providing OLAP (On-line Analytical Processing) to user-analysts: An IT mandate," E.F. Codd and Associates, Tech. Rep., 1993.

[2] Gray J., Chaudhuri S., Bosworth, A., Layman, A., Reichart, D., Venkatrao, M., "Data Cube: A relational aggregation operator generalising group-by, cross-tab, and sub-totals," Data Mining & Knowledge Discovery, vol. 1, 1997, pp. 29-53.

[3] Kimball, R., The Data Warehouse Toolkit. New York: John Wiley & Sons, 1996.

[4] Agrawal, R., Gupta, A., Sarawagi, S., "Modeling multidimensional databases." IBM Almaden Research Center, Tech. Rep., September 1995.

[5] Vasilakis, C., El-Darzi, E., Chountas, P., "A Data Warehouse Environment for Storing and Analyzing Simulation Output Data" Winter Simulation, ACM Guide, IEEE Press 2004, pp. 691– 698.

[6] Atanassov, K., "Intuitionistic Fuzzy Sets", Springer-Verlag, Heidelberg, 1999.

[7] Rogova, E., Chountas, P., Kolev, B., "Intuitionistic Fuzzy Knowledge- based OLAP", Notes Intuitionistic Fuzzy Sets, Vol.13, No.2, ISSN 1310-4926, 2007, pp.88-100.

[8] Chountas, P., Kolev B., Tasseva, V., Atanassov, K., "Generalized Nets in Artificial Intelligence, Vol. 4: Generalised Nets, Uncertain Data, and Knowledge Engineering", Publishing House of Bulgarian Academy of Sciences, Sofia, 2007.

[9] Rogova E., Chountas P., "On imprecision Intuitionistic fuzzy sets & OLAP – The case for KNOLAP" IFSA 2007, Springer Verlag, pp. 11-20

[10] Atanassov, K., Generalized Nets, World Scientific, Singapore, New Jersey, London 1991.