THE PROPOSAL OF APPLYING TEMPORAL - PROBABILISTIC DEPENDENCIES INTO DATABASE SYSTEM

ZYGMUNT MAZUR, PIOTR MICHALEC
Institute of Applied Informatics
Wroclaw University of Technology
Skwer Idaszewskiego 1
50-370 Wroclaw
POLAND

Tel.: (+48 71) 320 35 16,
Fax: (+48 71) 321 10 18
E-mail: {zygmunt.mazur, piotr.michalec}@pwr.wroc.pl

MAREK WODA
Institute of Engineering Cybernetics
Wroclaw University of Technology
Janiszewskiego 11-17
50-370 Wroclaw
POLAND

Tel.: (+48 71) 320 29 69,
Fax: (+48 71) 321 26 77
E-mail: mwoda@ict.pwr.wroc.pl

Abstract. There are many database applications, which manipulate data and queries with uncertainty about the occurrence time of an event. These applications are naturally represented by object-oriented models, which allow creating easy in understanding description of database system. Features of that model are parameterized by time (time dependencies) and uncertainty (probabilistic dependencies). The data model allows to associate with each event a set of time points, and with each time point, an interval for the probability that event occurred. Furthermore, there is presented definition of temporal - probabilistic description of transaction and priority assigning algorithm, which works basing of that description.

Key words: database models, database transaction processing

1. INTRODUCTION

The working out of object paradigm [1, 12] caused fundamental change in the way of data perception and procedures working with them. Traditionally data and procedures where stored separately: data and relationships among them in database and procedures in application programs. In object-oriented model there is combination of procedures and data, it is considered as an advantage in data management subjects. Entities are independent, and can be easily moved and re-used, their behaviour is predictable. These and other features of object-oriented model [9, 11, 12, 14] predestine it to store information for systems supporting designing, production (CAD, CAE, CASE i CAM), knowledge databases (e.g. advising systems, geographical information systems), multimedia systems and other. Many of these applications works with data and expressions where exist uncertainty about the occurrence time of an event. They are naturally represented by object - oriented model, and time (temporal dependencies) and uncertainity (probabilistic dependencies) parametrize that model. In section 2. there will be presented temporal - probabilistic - object model.

The real-time database system is integration of database management system and real-time system. It inherits many features from database systems as well as real-time systems. The increasing interest in that systems results from the fact, that rise frequency of needs for application, which process huge amounts of data, and that applications have to meet time requirements (where values of data depends on time, or have restricted valid time). Applications of real - time database systems include air - traffic control systems, production systems, military systems, telecommunication and computer network systems, and many financial systems. Because all these systems are distributed systems, their primary task is processing distributed transactions. Moreover, in real-time systems, transaction processing is connected with time restrictions. The time restrictions can concern both transaction processing and time of valid data. System has to process transactions and return result before deadline. The main aim of real-time database system is increasing the number of transactions, which are completed before their deadlines. In section 3., there is presented introduction of temporal - probabilistic description of transaction into priority assigning algorithm, in order to increase the number of successfully completed transactions in real-time database systems.

2. TEMPORAL – PROBABILISTIC - OBJECT DATABASE MODEL

There is small number of scientific studies about integration of temporal and probabilistic information in databases. In the first paper in this subject [20], there is presented approach to temporal uncertainty in the relational database model, based on probability calculus. Author of this pioneer paper made very important first step to solve the problem of representation temporal-probabilistic information in databases. Next papers [4], [5] introduce temporal - probabilistic dependencies into relational database systems and basing on that in [2] there is contructed temporal - probabilistic - object database. However, it is very similar to [4] and it is rather object-relational database than pure object

database. Of course there are other studies in the field of temporal uncertainty which use e.g. fuzzy sets [6], [7].
Object - oriented database model is very important as a starting point for research and development of database systems. In this section there will be presented definition of temporal - probabilistic - object database, based on object - oriented database model.

2.1. TYPES AND VALUES

Before we define our database model it is necessary to define set of types and set of values. Firstly, set of atomic types is defined as usual: Θ={integer, string, Boolean, float, etc} with standard domains for these types. Additionally, it is necessary to take into consideration temporal atomic type (after [13]) specified as calendar τ, which domain dom(τ) consists of all allowed in τ time points S_τ.
Assuming that A is a set of attribute names, and Θ is a finite set of atomic types, the set of all types is defined by induction:
1. Every atomic type from set Θ is a type
2. If θ is a type, than $\{\theta\}$ is a type, which is called complex type or tuple type
3. If $A_1,\ldots,A_K$ are attributes mutually independent (independence means that value of every attribute can be changed regardless of other attributes) from A and $\theta_1,\ldots, \theta_k$ are types, so $[A_1: \theta_1,\ldots, A_K: \theta_k]$ is a type too.
Every atomic type $\theta\in\Theta$ has standard domain dom(θ). Let W be set of all values, W is defined by induction:
1. For all atomic types $\theta\in\Theta$, every value $w \in$ dom(θ) is value of type θ.
2. If $w_1,\ldots w_k$ are values of type θ, so $\{w_1,\ldots,w_k\}$ is value of type $\{\theta\}$.
3. If $A_1,\ldots,A_K$ are attributes mutually independent from A and $w_1,\ldots,w_k$ are values of types $\theta_1,\ldots \theta_k$, then $[A_1: w_1, \ldots A_k: w_k]$ is a value of type $[A_1: \theta_1,\ldots,A_K: \theta_k]$.
In object-oriented databases type describe the structural part of class. Class contains structure and specific behavior: especially it is creating and deleting of objects. Values with the same structure has the same type, and object has are the same class.

2.2. TEMPORAL - PROBABILISTIC OBJECT

The temporal - probabilistic object is defined as ID, time interval and probabilistic mass distribution specified on that time interval.
Definition of temporal-probabilistic object.
Object is defined as six:

$$(i, k, t, p, \sigma, v),$$

where:
i is ID,
k is the object class,
t is time interval,
p is probability of object occurence,
σ is mass distribution function,
v is a type value of class.
ID of object i unambiguous distinguish object in the objects spase, and k defines class of objects group. Time interval t stores object's valid time, and mass distribution σ of probability p in the interval t allows for calculating of occurence certainty of object in every point of time from time interval t. Value v means the object's actual state.
Because i, k, t, p and σ are also values, then the class itself is a type too, and it can be used as a type in the structure of another class.

2.3. TEMPORAL – PROBABILISTIC - OBJECT DATABASE

After introducing all these concepts we can define now the schema of database, which assigns types to classes according to class hierarchy:
Temporal-probabilistic-object database schema

$$(K, IsA, \tau, I)$$

where:
K is a finit set of classes,
IsA is a hierarchical arrangement of classes, i.e. if k IsA k' is true, then k is a subclass of k', a k' is a supclass of k,
ω is a function $\omega(K)$, which assigns tuple type to every class from the K,
I is a set of objects, which have assigned base class from the set K.
Classes are arranged as hierarchy, in which the most important are relationships IsA among participating classes. If class k is a subclass of k', then every object k is also the object of class k'. This hierarchy is a base of inheritance.

3. INTRODUCTION OF TEMPORAL - PROBABILISTIC DESCRIPTION INTO TRANSACTION PROCESSING

With transactions in real-time database systems are connected following attributes [Harista01]:
- arrival time A
- deadline D
- execution time E
- maximum slack time S
- value function V
- critical value C
- priority P

The arrival time A and deadline D define time restrictions of transaction processing. The execution time E is approximate value and usually it is not known before the transaction is completed. Attribute S informs about maximum allowed delaying of beginning of transaction processing. The value function V defines validity of transaction ending in time point, value of this function for deadline is called critical value C. The

priority P is assigned to transaction to optimize effectiveness of transaction processing.
conflicts of accessing to system resources. System can use different strategies of priority assigning for different types of resources. If conflict occurs, assigned priority is used for solving that conflict. However, the standard transaction scheduling algorithms are inappropriate for real - time database systems, because in real - time database systems transactions have to be ranked with respect to critical value or deadline.

3.1. THE PRIORITY ASSIGNING ALGORITHMS

The main aim of transaction scheduling in real-time database applications is executing all transaction before occurring their deadlines. Because access to resources is exclusive, in one time point only one transaction can access the resources. Execution of single transaction keeps correct state of real - time database, but execution of many transactions needs coordination. To this end scheduling procedures with priority mechanism are used. The priority value has influence for concurrent transaction processing to optimize effectiveness parameters of system. Presented attributes of transaction form value of priority in the following way:

- critical value - if critical value of transaction is bigger, the priority is higher too
- deadline - if deadline is occurring earlier, the priority is higher
- arrival time - transaction, which is oldest should have higher priority
- slack time - if maximum slack time is shorter, the priority is higher
- time already spent for transaction processing - transaction, which is processed longer, has higher priority. Database systems, despite exclusive access, require also rollback a transaction if it is not finished, so sometimes it is better to execute transaction to the end.

The access conflicts are solved to higher priority transaction advantage, it means, that transaction with low priority is restarted if the higher priority transaction Transaction scheduling algorithms usually assign priorities to transactions and uses methods which solve occurs with demand for the same resources (Figure 1). This mechanism has also some disadvantages. There is possible that transaction with higher priority restarts the low priority transaction and then it is restarted by the next transaction with highest priority. Another, disadvantageous situation is where transaction with higher priority restarts the low priority transaction and then it crosses its deadline, and the restarted transaction also crosses its deadline (because of enforced restart).

3.2. THE PROPOSITION OF PRIORITY ASSIGNING ALGORITHM WITH TEMPORAL - PROBABILISTIC DESCRIPRION OF TRANSACTION

The loss of time caused by situations described above can be limited by introducing temporal - probabilistic description of transaction. The temporal - probabilistic description of transaction is defined by probability p of completing transaction according to mass distribution function σ on time interval t and represented as three

$$(t, p, \sigma),$$

where:
t - is time interval,
p - probability of completing transaction,
σ - probability mass distribution function.
Graphical illustration of temporal - probabilistic description of transaction is presented on Figure 2.
The priority assigning algorithm with temporal-probabilistic description of transaction is defined by formula:

$$P = max(p_i(t_i)),$$

where: $p_i(t_i)$ - probability of completing transaction in time t_i, $t_i \leq D$
The transaction control system can skip transactions, which have low probability of transaction completing before their deadlines (in the time interval from arrival time to deadline).

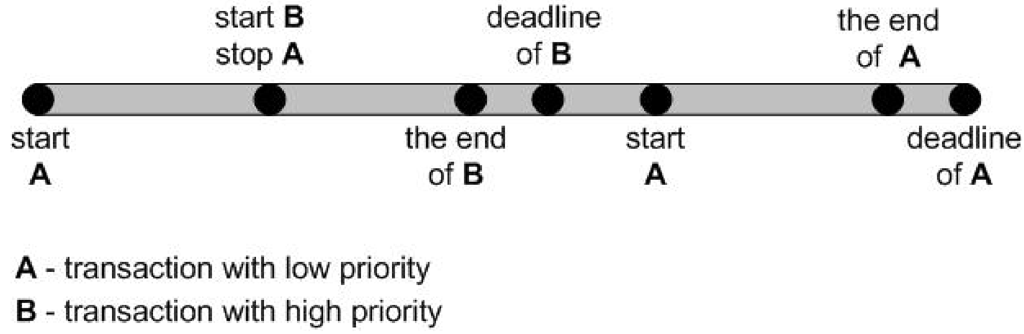

Figure 1. Transaction Processing In System With Priority Mechanism

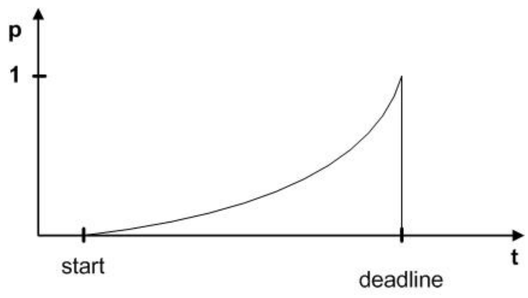

Figure 2. An Example Of Probabilistic Distribution For Transaction Execution

Limitation of time loss, and increased number of successfully completed transaction is achieved by elimination of transactions, which have small (or none) chances for successful completing. Transaction with low priorities are not restarted by transactions, which do not have chances for completing before their deadlines, thus time designed for one transaction increase. There is possible adaptation this system, to admit transactions with small probability for successful completing but with low priority. Then transactions keep their chances for being processed, and they do not lock other transactions with larger probabilities.
Applying temporal - probabilistic description of transaction and classification according to probability of successful completing criterion can be used as additional information for known, more advanced concurrent transaction processing algorithms.

4. SUMMARY

In the paper the new model of temporal - probabilistic - object database was proposed, which allows to connect with data the set of time points, and with each time point the probability of appearance the data. There was introduced definition of the temporal-probabilitic object and the schema of temporal - probabilistic - object database. Furthermore, there was presented features and attribures of real - time database, the properties of transaction in real - time database system, and there are pointed out sources of their time limitations. There was also presented short profile of priority assigning algorithms. Next there was presented definition of temporal - probabilistic description of transaction and priority assigning algorithm which uses advantages of temporal - probabilistic description.

REFERENCES

[1] M. Atkinson, F. Banilhom, D. De Witt, K. Dittrich, S. Zdonik, The Object-Oriented Database System Manifesto. *Proc. of the First Object-Oriented and Deductive Database Conf.*, Kyoto, 1989, pp. 40-57.

[2] V. Biazzo, R. Giugno, T. Lukasiewicz, V. Subrahmanian, *Temporal Probabilistic Object Bases,* INFSYS Research Report, Wien, 2002.

[3] C. Date, *Wprowadzenie do systemów baz danych*, Warszawa, WNT, 2000.

[4] A. Dekhtyar, R. Ross, Subrahmanian V., Probabilistic Temporal Databases, part I: Algebra, *ACM Trans. on Database Systems*, vol. 26, 2001, pp. 41-95.

[5] A. Dekhtyar, F. Ozcan, R. Ross, V.Subrahmanian, Probabilistic Temporal Databases, part II: Calculus and Query Processing, *Technical Report - University of Maryland*, Washington, 2002.

[6] D. Dubois, H. Prade, Processing fuzzy temporal knowledge, *IEEE Transactions on Systems, Man, and Cybernetics*, 1989, pp. 729-744.

[7] S. Dutta, Generalized events in temporal databases. *Proceedings of the 5th International Conference on Data Engineering*, 1989, pp. 118-126.

[8] T. Eiter, J. Lu, T. Lukasiewicz, V. Subrahmanian, Probabilistic Object Bases, *ACM Transactions on Database Systems,* vol. 26, 2001, pp. 264-312.

[9] D. Figura, *Obiektowe bazy danych*, Warszawa, Akademicka Oficyna Wydawnicza PLJ, 1996.

[10] J.R. Haritsa, *Performance Analysis of Real_Time Database Systems,* USA, 2001

[11] J. Harrington, *Obiektowe bazy danych dla każdego*, Warszawa, MIKOM, 1991.

[12] W. Kim, *Wprowadzenie do obiektowych baz danych,* Warszawa, WNT, 1996.

[13] S. Kraus, Y. Sagiv, V. Subrahmanian, Representing and integrating multiple calendars, *Technical Reports of University of Maryland*, Washington, 1996.

[14] G. Lausen, G. Vossen, *Obiektowe bazy danych: modele danych i języki,* Warszawa, WNT, 2000.

[15] Z. Mazur, P. Michalec, Temporalno-probabilistyczno - obiektowe bazy danych, *Prace naukowe Wydziałowego Zakładu Informatyki Politechniki Wrocławskiej - Bazy Danych 5,* Wrocław, Oficyna Wydawnicza PWr., 2004, pp. 21-30.

[16] Z. Mazur, P. Michalec, Zastosowanie opisu temporalno - probabilistycznego do harmonogramowania transakcji. *Prace naukowe Instytutu Informatyki Stosowanej Politechniki Wrocławskiej – Bazy Danych 6*, Wrocław, Oficyna Wydawnicza PWr., 2005, pp. 27-34

[17] P. E. O'Neil, K. Ramamritham, C. Pu, *A two-phase Approach to Predictably Scheduling Real-Time Transactions*, New Jersey, Prentice-Hall Pub., 1999.

[18] P. Porkka, *Overview of Real_Time Databases*, USA, 1997.

[19] J.J. Prichard, P.J. Fortier, W. Fay, Flexible Real-Time SQL Transaction, *IEEE Real-Time Systems Symposium,* Puerto Rico, 2000.

[20] R. Snodgrass, C. Dyreson, Supporting Valid-Time Indeterminacy, *ACM Transactions on Database Systems,* vol. 23, 1998, pp. 1-57.

[21] T. Szmuc, R. Klimek, *Systemy czasu rzeczywistego*, Kraków, Katedra Automatyki AGH, 2001.

[22] J. Ullman, J. Widom, *Podstawowy wykład z systemów baz danych,* Warszawa, WNT, 2000.

[23] O. Ulusoy, G.G. Belford, *Real-Time Transaction Scheduling in Database Systems*, Oxford, Elsevier Science, 1998.