Volume 5, Issue 9, September 2015 ISSN: 2277 128X
International Journal of Advanced Research in Computer Science and Software Engineering
Research Paper
Available online at: www.ijarcsse.com

Overview of Fuzzy Database Conceptual Modelling in Commercial Databases

Deepti Dalal
Worked as an Assistant Professor, KIIT,
Gurgaon, India

Abstract— Industrial data management requires database technique support as data in these is frequently permeated with imperfect information. Industrial applications have exclusive characteristics which makes the management process a difficult task. There are some techniques which are used frequently like the AI, mining and web technology in the applications of organizations. Due to the frequent use of these innovative and recent technologies there is a need of database management in these industries to improve their standards and help in their business. Hence, database systems must be able to deal with such type of information. On the whole, to incorporate the necessary requirements required to deal with such data types that are imperfect or inconsistence the commercial database systems seems to be slow. A number of fuzzy database models are used to handle enterprise-data uncertainty. The conceptual data models represent the various complex objects with uncertainty in the element values and complex relationships can be modeled between attributes.

Keywords— IFO, ExIFO, EXPRESS-G, Fuzzy, EER, IDEF1X, GQL, Fuzziness

I. INTRODUCTION

In real-world applications database systems model our knowledge which is repeatedly permeated with imperfect information. Consequently, database systems have to deal with such type of information. Various possible type of imperfect information in database systems is inconsistency, imprecision, vagueness, uncertainty and ambiguity. Certainly, in most of the database systems there is a capability to deal with such imperfect information like some allow replacing the missing values with the null value, some allow dealing with the pattern matching and allowing the users to query. Some other system also allows substitution mechanism, but all these capabilities are not enough to deal with the type of uncertainty that we encounter in daily routine. With the conventional database modeling it is not so easy to deal with such type of information so to deal with such data types where there is uncertainty in the data is more some fuzziness is to be added to the data. In the business and various application areas where data is collected in the databases such type of data is common. Various difficulties are encountered to deal with such data types so to deal accurately and help the business process there is a need of management of this imperfect information. Fuzzy database is introduces to deal with such imperfect data type which handles the data in the databases. There are various models to deal with fuzzy database; these are broadly classified as conceptual and logical data models. Logical database models can further be classified as relational and object oriented database models. By mapping the conceptual database models we can get logical database models. The various complex objects are represented by the object oriented database model without dividing the aggregate data into small chunks and the various relationships that exists between the attributes is also kept maintained. Conceptual database model is almost same as object oriented database model; it added fuzziness to the values of the attributes to represent complex object. A number of fuzzy logical and conceptual data models are surveyed by Yazici and *et al*., 1992; the models including IFO, ER and relational are surveyed for the management of uncertainty in the data. In business applications and all the areas where there is a need to store data in the database system we can use the data models that suit the particular area to deal with the uncertain data types. These data models are used widely to deal with uncertain data proved to be very useful by enhancing the management capability of the database management system.

II. IMPERFECT INFORMATION AND FUZZY SETS THEORY

In a general term we use a word uncertainty for the data values that are null, not available, contain multiple values for the same attribute, but the imperfect information may be classified as four different types:

Imprecision

Information is said to be imprecise if we do not know what the exact value of an attribute is and we have to select a value from a set. Consider an example to understood this a color of car may be red, white, black or silver we don't know the color of the car; in this case the value is given by a set{red, white, black, silver}. An another example to understand this concept is when a range of values is given for a particular attribute like marks of a student lies in the range 50-70 This type of attribute values where we have to select a value from the available set or range are called imprecise values.

Inconsistency

Sometimes there are some attributes whose values are assigned different at different places, may be in the same database or in different databases. This may be due to typing error or may be due to anomaly in updation that a data value is updated or changed only at some place s not at all places the values are updated. For example if in a database the phone number of an employee is changed then it might be possible that the updation can be done only in some files not in all files where this information is stored so we have multiple values for the same data which is correct. This type of data values that contain multiple values is called the in consistence values.

Ambiguity

There are some cases for which we don't have much information available to interpret the result correctly and we cannot find the value of the attribute. For example if in a database it is mentioned that the student is pass and we have no other information available for this attribute like the percentage or if only marks obtained is given we do not know the total marks, then it is difficult to find whether the student is passed in first grade or in second grade and etc. This type of information where we have a number of possible values to interpret for the attribute but we are lack of the exact value is called ambiguity.

Uuncertainty

When we are not certain about the data value then it is called uncertain data. For example if we say that the salary of the employee may be around 50,000 or we can say that his salary lies in the scale 2 and we don't know about the increments, from this sort of information we are unable to calculate the exact value for the salary attribute. This type of information where approximation, percentage or degree of truth of attribute values is considered for data collection is considered as uncertain data.

In a database either the data is fuzzy or the query is imprecise. So to represent this type of imprecision in the data there is a need for fuzziness in the data modeling to properly and correctly representing the information. Fuzzy membership function is used to represent Fuzzy sets. In the fuzzy membership function a number of attributes are taken as an input and are assigned values between 0 and 1 as a fuzzy membership output value

III. HISTORIC FUZZINESS

In 1973, according to Galbraith uncertainty is considered as the difference that exists between the available information and the required information which is essential to perform a specified task. In true world we encounter different types of uncertainty in the data which can directly or indirectly affect the process of the production. It was Plato who gives the basis for fuzzy logic by signifying that there is a third possible value further than true or false i.e. neither true nor false. It was the early 1900's that a three valued logic was proposed by Lukasiewicz, and this third value along with the two values true and false has a value that is numerical in term lies between true and false. Later he give four valued logics then five valued logics and consequently he come to a result that there exists infinite number of values between true and false and explored an infinite valued logics.

It was not until quite a moment ago that the idea of an infinite-valued logic took place. In 1965 the published seminal work of Lotfi A. Zadeh "fuzzy sets" gives a complete new set of calculus and several operations that are applied to data which a generalisation of the classic logic. Three levels are specified by Medina and *et al.,* 1995 for the implementation of imperfect information in the database. The three levels are: database system level which is associated with extended data manipulation languages, Database level is related to the internal storage of the imperfect information and the third level meta knowledge base level is concerned with the meaningful definition of different relations and classes that exists in the fuzzy database. Bahri and *et al.,* 2005 add a new model base level of implementing imperfect information in database to the three levels proposed by Medina and *et al*., 1995. The newly added model base level is to group the function definitions that are used for the computing of the membership degrees and also group those function definitions that are associated with the different data types that are used to generate their possibility distribution. A similar study of Popat and *et al*., 2004 gives a classification of fuzzy data that can be used as a framework for understanding the origins of fuzzy data and the impact of fuzzy data in database management systems.

IV. FUZZY CONCEPTUAL DATA MODELS

In conceptual data models of the fuzzy logic the various objects which are complex in nature are represented by the fuzziness in the values associated with the attributes. We use the conceptual data models to model the information in those cases where there is a high level of abstraction is used and for the data manipulation level. To model a database we generally start the modeling at conceptual level i.e. called conceptual data model and after the conceptual data model is completed we proceed to the logical data model which is formed by mapping with the conceptual data model. There are different conceptual data models which are available for the modeling of the databases that represent fuzziness in the data in their databases. Here some conceptual data models are discussed in brief.

A. IFO Data Model

According to Abiteboul and Hull, 1987 we can define an IFO model mathematically that represent basic principles in a framework which is graph based for a semantic database modeling. In an IFO model the schema which represents atomic, constructed objects and ISA relationship is defined as a graph with a different type of vertices and edges. A schema is considered as comprising these elements in it. Atomic objects are singular items that cannot be partitioned

further and are not derivable from the other objects. In an IFO data model there are basically three types of atomic objects first is of free type, second abstract type and the third one is printable type. We can construct complex objects by combining two or more atomic objects which may be of any type from the three mentioned. There are basically two types of constructs that can be used to create a complex object, by grouping the atomic objects or by aggregating the atomic objects.

An extenton of IFO model is given by Yazici and Cinar, 1998. They proposed ExIFO2 model in which the uncertain and complex information in the database is represented by object oriented approach. A similar study of Yazici and *et al*., 1999 also extend the IFO model to ExIFO i.e. the extended version of IFO data model. The ExIFO model is the semantic data model in which conceptual design level is used to represent the data and the various uncertainties are represented at the level of attributes, at the object level and at the class level. The ExIFO model is then converted into Logical data model. In it the three different attribute types are used these are fuzzy valued attribute, incomplete valued attributes and the null valued attributes. The fuzzy valued attributes are those in which the correct value of the attribute can be given from a set of values. An incomplete valued attributes are those for which the values are unknown and the null values attributes are those for which the information is not precise.

Galindo and *et al*., 2001introduced a new system in which flexible constraints are expressed, these flexible constraints use EER i.e. enhanced entity relationship to model the database. The restrictions have been proposed using fuzzy quantifiers. Ma and Ma, 2000 proposed a new model IF_2O by extending IFO data model to model fuzziness at different levels. In their model they mainly discussed the various types of fuzzy printable along with the relations that exist between the entities by fuzzy ISA relationship.

B. IDEF1X Data Model

According to Kusiak *et al*., 1997 it is the IDEF1X methodology in which most constructs are same as in the entity relationship models. The constructs that are used in the IDEF1X model are entities, relationships between entities and the attributes associated with the entities. In an IDEF1X data model there is a classification of attributes that is made, it may be either identifier dependent or not dependent on the identifier. When the requirement are known and it is decided that a relational database is o be made then for the logical design of the database IDEF1X is considered to be the most useful model. The relationship that exists in the IDEF1X data model may be of any type from the three which is defined in the model. The first type of relationship type that is defined in the model is connection relationship, the second one is nonspecific relationship and the third relation defined in this model is categorization relationship.

C. EXPRESS Model

A standard is developed by the ISO (International Organization for Standards) for Product Model Data which is used for the sharing of the data and to exchange the product data. This standard is known as STEP and it gives a way of describing the product model of the product data all the way through its lifecycle. It describes how the data should be exchanged between the different units of the product. EXPRESS is a conceptual schema language which is defined in STEP that can be used as description methods. It is a language that can be used for the purpose of modeling of product design; it can be frequently used for the manufacturing purpose and for the purpose of production of the data.

EXPRESS model is mainly used in the field of engineering for the modeling purpose. In this modeling scheme the description of the objects are created to represent the information in an efficient manner. The descriptions of the data types we are using are described in this model along with the explanations of all the statements that are used in the product data. This explanation tells what the statements are supposed to do, these are the executable statements and the various expressions are also described. EXORESS model also describe the various interfaces, and the schemas that are used in the modeling process. Two types of specifications are used in this model so as to minimize the redundancies that exist among different schemas; these specifications are USE and REFERENCE.

D. EXPRESS-G Model

EXPRESS model is further improved by adding graphical representation to it and it is called the EXPRESS-G model. In this model not all the language that was used in the EXPRESS model is used, it just includes only a subset of its language that was originally used the EXPRESS model. In this model symbols are defined to represent the various elements the model, it defines the symbols for definition, relation and the third symbol is for composition. Definition symbol defined in this model is basically used to define the contents that are used in the model; the symbol used for relation describes the structure within the model and the symbol used for composition allows the expansion of the diagram to several pages.

E. Fuzzy EER Model

Fuzzy logic was introduced in the Entity-Relationship model first by Zvieli and Chen in 1986. In this model the fuzziness is added to the values of the various attributes in the entities. In 2006 a new concept was introduced by the Galindo and *et al*. in which they introduce the concept of modelling the database using relational databases to represent the modelled fuzzy knowledge in the fuzzy EER model. The capabilities of fuzzy logic are introduced in the SQL language. Various descriptions of specifications are described in this model along with the description of how to implement the concept of fuzzy logic. The extension of fuzzy logic in SQL is given a name FSQL which allows using the extensions of the FEER model by allowing users to use various flexible conditions to be included in the queries.

F. Fuzzy Formal Concept Analysis

A new approach is given in which by using the information acquired from uncertain data, concept hierarchy is generated automatically and is called conceptual clustering. This approach is known as fuzzy Formal Concept Analysis (FCA) given by Quan and *et al*., in 2004. In conceptual clustering the data is clustered into groups called conceptual clusters. These clusters are then used to construct the hierarchy of the concepts i.e. concept hierarchy. The various steps included in the approach are:

1. First of all a fuzzy concept lattice is formed.
2. Then after constructing the lattice, the lattice is clustered into conceptual clusters.
3. At last the hierarchy of concepts is constructed.

G. Graph-based semantic Data Model (GDM)

In Graph-based semantic Data model the information is organized in an efficient way which is useful in the storage of the data and then retrieval of the data. This is a good model of its type which link the gap that exists between the various computing infrastructure needs and the perception of human for an enterprise. In this model the database is considered in a layered organization and in the form of a Graph (V, E). The various advantages of the two different data models are included in this model; the two models are relational data model and the semantic data model. The good features of both of these data models are included in GDM. Construction of graph for the database is carried out in bottom up approach. The vertices, V of the graph are the data instances and are called primary semantic group (PSG), or it may be a functionally abstracted module called secondary semantic group (SSG). The edges of the graph 'E' is used to represent the various relationships that exists among the SSG. A different language GQL (Graph Query Language) is used in this model for the purpose of querying the database, which is similar to SQL

V. CONCLUSIONS

With reference to the fuzzy conceptual data modelling of industrial information, in addition to the ER/EER data models, the IDEF1X data model and the EXPRESS data model are extended for fuzzy industrial data modelling. The mappings from the fuzzy IDEF1X model to the fuzzy relational databases and from the fuzzy EXPRESS-G model to the fuzzy nested relational databases are developed in addition to the mappings from the fuzzy ER model to the fuzzy relational databases and from the EER model to the fuzzy object-oriented databases. IFO data model is a defined mathematically as the conceptual data model which is extended to the ExIFO (Extended IFO) model as the semantic data model and IF_2O to model fuzziness at different levels.

REFERENCES

[1] Abiteboul S. and Hull R. (1987). "IFO: A formal semantic database model," *ACM Transactions on Database Systems*, Vol. 12, pp. 525-565.

[2] Bahri A., Chakhar S., Naïja Y. (2005). "Implementiong Imperfect Information in Fuzzy Databases". The International Symposium on Computational Intelligence and Intelligent Informatics, Hammamet, Tunisia.

[3] Galbraith J., (1973). "Designing Complex Organizations". Addison-Wesley, Reading, MA.

[4] Galindo J., Urrutia A., Carrasco R.A., Pittani M. (2001). "Fuzzy Constraints using the Enhanced Entity Relationship Model", in: XXI International Conference of Chilean Computer Science Society (SCCC'01)", Punta Arenas, Chile, pp. 86-94.

[5] Galindo J., Urrutia A., Piattini M. (2006). "Fuzzy Databases: Modelling Design and Implementation". IDEA Group, Hershey, PA.

[6] Kusiak, A., Letsche, T. and Zakarian, A. (1997), Data modeling with IDEF1X, International Journal of Computer Integrated Manufacturing, 10: 470-486.

[7] Ma Z.M., Ma W.Y. (2000). "An Extended Conceptual Model for Fuzzy Data Modeling", in: First International Conference on Web Information Systems Engineering, Hong Kong, China, pp. 75-80.

[8] Medina J.M., Vila M.A., Cubero J.C. and Pons O. (1995). "Towards the implementation of a generalized fuzzy relational database model". Fuzzy Sets and Systems, Vol. 75, pp. 273-289.

[9] Popat D., Sharda H. and Taniar D. (2004). "Classification of Fuzzy Data in Database Management System". *M. Gh. Negotia et. al. (Eds.): Springer-Verlag Berlin Heidelberg*. Pp. 691-697.

[10] Quan T.T., Hui S.C. and Cao T.H. (2004). "A Fuzzy FCA-based Approach to Conceptual Clustering for Automatic Generation of Concept Hierarchy on Uncertainty Data". *V. Snášel, R. Bělohlávek (Eds.): CLA*, pp. 1–12.

[11] Yazici A. Buckles B.P. and Petry F.E. (1992). " A survey of conceptual and logical data models for uncertainty management". Fuzzy Logic for Management of Uncertainty, John Wiley and Sons Inc., NY, pp. 607-643.

[12] Yazici A., Cinar A. (1998). "Conceptual Design of Fuzzy Object Oriented Databases", in Second International Conference on Knowledge Based Intelligent Electronic Systems", Adelaide, SA, Australia, pp. 299-305.

[13] Yazici, Buckles B.P., Petry F.E. (1999). "Handling Complex and Uncertain Information in the ExIFO and NF2 Data Model", *IEEE Transactions on Fuzzy Systems*, Vol. **7**, No. 6, pp. 659-676.

[14] Zvieli A., Chen P. (1986). "ER modeling and fuzzy databases". In: Proc. 2nd Intl. Conf. Data Engineering (Los Angeles, CA), pp. 320-327.

