INTERNATIONAL JOURNAL OF INNOVATIVE RESEARCH & DEVELOPMENT

ISSN 2278 – 0211 (Online)

An Overview of Fuzzy Database Approaches in Handling Imprecision

Ugochukwu M. ENE
Ph.D. Student, Department of Computer Science, University of Nigeria, Nsukka, Nigeria
Dr. M. C. Okoronkwo
Senior Lecturer, Department of Computer Science, University of Nigeria, Nsukka, Nigeria

Abstract:
Many approaches have been proposed in order to represent imprecise information in databases. This paper gives an overview of different database approaches in which different fuzzy relational and object-oriented database models are discussed. Finally, an overview of Fuzzy Object-Oriented Database (FOOD) model including applications, limitations and improvement is given.

Keywords: *Fuzzy database models, Fuzzy Object-Oriented Database, Fuzzy Database approaches*

1. Introduction

Enterprise systems are being developed to be adaptable, intelligent, flexible and efficient in order to meet up with the ever-changing business environments. According to [23], decision making in the real word takes place in an environment in which goals, constraints and outcomes are not precisely known. Decision making at the Management level are usually based on imprecise or incomplete information.

Information from real life systems are sometimes vague, incomplete, imprecise and ambiguous. In [1], the author observed distinct types of uncertainty in fuzzy information as follows:

- Uncertainty – impossibility of determining whether an assertion in the model is true or false.
- Imprecision – the information in this model is not specific or could be unknown.
- Vagueness – the elements in the model are vague.
- Inconsistency – the model contains two or more assertions that cannot be true at the same time.
- Ambiguity – some elements of the model lack complete semantics which could lead to several interpretations.

In real life, many abstract concepts such as bad, good, cheap, expensive and so forth are used. Ignoring imprecise information in real life applications will lead to elimination of very vital information. Hence, imprecise information is very useful in building applications. Conventional databases are unable to handle fuzzy data.

There are several attempts to represent imprecise information without fuzzy logic on databases such as introduction of NULL values by Codd (Codd Approach), default values by Date, interval values by Grant, statistical and probabilistic databases. An overview of these attempts is given in section 3. The basic model of Fuzzy databases is discussed in section 4. Several models have been built to incorporate fuzzy logic into databases. These models are discussed in section 5. Basic databases are unable to handle impreciseness. The concept of fuzzy object-oriented data has been evolved to handle complex data that classical databases are unable to handle.

According to [2], fuzzy object-oriented database comprises of the concept of database techniques, object-oriented modelling and application of fuzzy set theory.

Implementation of fuzzy object-oriented database are required in medical diagnosis system, multimedia database, information retrieval and so on. An overview of Fuzzy Object-Oriented Database (FOOD) model is given in Section 6.

2. Related Works

An overview of Fuzzy database models was done in [24] with references to researches and application issues of Fuzzy Relational and Object-Oriented Databases. In [2], an overview of Fuzzy Object-Oriented Database systems (FOODs) with details of different modeling and algebraic expressions that can be performed on it is given. A brief review of logical relational database models and fuzzy set theory was done by [25]. The review concentrated on only relational database models. Fuzzy database approaches such as imprecision without Fuzzy Logic, basic model of fuzzy databases, similarity relations, possibilistic models and Fuzzy Object-Oriented Database models are discussed in [5]. In [12], the strengths and weaknesses of Fuzzy Relational and Fuzzy Object-Oriented Database models are compared as well as their data model representations and data manipulation languages.

3. Imprecision Without Fuzzy Logic

There have been attempts to represent imprecise information without utilizing fuzzy logic. The first attempt to represent imprecise information in databases was made by E.F Codd in 1979[3]. It is commonly referred to as "Codd Approach". It was further expanded (Codd 1986, 1987, and 1990). Fuzzy set theory was not used in the model. According to [4], NULL is a special marker in a Structured Query Language (SQL) to show that a data value does not exist in a database. In other words, NULL shows the lack of a value which is different from zero (0). A three-valued logic was adopted in [3] for extracting data from databases that may contain null values. The Greek lowercase character, omega (ω) was used to represent unknown truth value while (T) represents "True" and (F), "False". The following truth tables represents the three-valued logic:

AND	F	ω	T	**OR**	F	ω	T	**NOT**	
F	F	F	F	F	F	ω	T	F	T
ω	F	ω	ω	ω	Ω	ω	T	ω	Ω
T	F	ω	T	T	T	T	T	T	F

Table 1 : Truth Table for Tri-Valued Logic: True, False and Unknown [3]

In [5], the details of the two marks added to differentiate the NULL value: the "A-mark" and the "I-mark" are summarized. The A-mark represents a missing or unknown but applicable value mark while an I-mark represents absent and inapplicable value mark [6]. For instance, an A-mark would be assigned to a salesman who has a missing but applicable sales volume while and I-mark would be assigned to a salesman with an inapplicable sales volume (probably because such salesman does not sell X product anymore).
The tetra- value logic is shown in Table 2.

NOT	
T	F
A	A
I	I
F	T

AND	T	A	I	F
T	T	A	I	F
A	A	A	I	F
I	I	I	I	F
F	F	F	F	F

OR	T	A	I	F
T	T	T	T	T
A	T	A	A	A
I	T	A	I	F
F	T	A	F	F

Table 2 : Truth Table for Tetra-Valued Logic [5]

In [6], other approaches are summarized like "default values" approach presented by C.J Date in 1982 as an alternative to NULL values, the "internal values" approach by Grant (1980) who expands the relational model to allow a possible range of values/intervals to be stored in one attribute, in addition to precise value as well as NULL value (where no information is provided) and statistical and probabilistic databases. The proposal relating to statistical database was done by Harry K.T Wong in 1992 in which cases of incomplete information can be statistically represented and compared. Barbara, Garcia-Molina and Porter in 1992 published probabilistic databases in which uncertainties were modelled as probabilities, where the probabilities are measures of imprecision in data [22]. The sum total of probabilities of all probable value is one (1) and the missing probabilities of is one minus the sum the known probabilities.

4. Basic Model of Fuzzy Databases

The basic model of fuzzy database is a simple one which entails the adding of a grade in the range of [0, 1] to each instance or tuple. This ensures that data homogeneity exists. The grade could have different meanings: membership grade, dependence strength level, fulfillment degree, importance degree [5, 7]
The challenge with these fuzzy models is that imprecise information cannot be accurately represented about a certain attribute of a specific entity (such as "dark" or "light" for colour attributes). In addition, the assignment of fuzzy character globally to each instance makes it impossible to know the exact contribution from each attribute.

5. Classification of Models Of Fuzzy Databases

The leading approaches for representing imprecise information are broadly classified into

	Classification	**Models**
(i)	Similarity relations model	Buckles-Petry model
(ii)	Possibilistic model	Prade-Testemale model, Umano-Fukami model, Zemankova-Kaendel model, GEFRED model

Table 3: Classification of Models of Fuzzy Databases

5.1. Similarity Relations Model

A similarity relation is a fuzzy relation whose membership function expresses similarity degree or resemblance between every two elements of the domain. In this approach, query results contain both exact that satisfy search conditions and data similar to the exact data [8].

The Buckles- Petry model proposed by Buckles and Petry utilized the similarity model [14]. The data type allowed in this model are finite set of scalars (labels), finite set of numbers and fuzzy number set. The similarity

values are normalized in the interval [0, 1], where 0 represents "totally different" and 1 "totally alike" or "similar". The fuzzy relation in this model is defined as a subset of the following Cartesian product: $P(D_1)X\ldots P(D_i)$ represents the part set of a D_i domain including all the subsets that could belong to the D_i domain.

5.2. Possibilistic Models

In possibilistic models, possibility theory is used to model imprecise information. The models in this group are: Prade-Testemale model, Umano-Fukami model, Zemankova-Kaendel model and GEFRED model.

Prade and Testemale published a Fuzzy Relational Database (FRDB) model that integrates incomplete data using possibility theory [15,16,17,18]. In this model, attribute value of a tuple can be a possibility distribution. A possibility distribution $\pi_{A(m)}$ about D ∪{e}represents all available information about the values of A for an m object(where e is a special element denoting the case in which A is not applied to m. $\pi_{A(m)}$ is an application that goes from D ∪{e}to the [0,1] interval. This implies that all the value types adopted by this model can be represented. In every possibilistic model, the value d ∈ D must be taken account of. Hence, if $\pi_{A(m)}(d) = 1, \forall d \in D$ then d value is completely possible for $A_{(m)}$.

The Umano-Fukami model proposed by [13] also utilized the possibility distribution in order to model information knowledge. The model was called possibility-distribution fuzzy relational model. During the query session, the model solves the query problem by dividing the set of instances into three subsets:

- The first subset contains instances completely satisfying the query.
- The second subset contains instances that may satisfy the query.
- The third subset contains instances that do not satisfy the query.

The presentation of information in Prade-testemale and Umano-Fukami models is shown in Table 4.

Information	Prade-Testemale model	Umano-Fukami model
The precise data is known and this is crisp:	$\pi_{A(m)}(e) = 0$ $\pi_{A(m)}(c) = 1$ $\pi_{A(m)}(d) = 0\ \forall d \in D, d \neq c$	$\pi_{A(m)}(d) = \{1/c\}$
Unknown but applicable	$\pi_{A(m)}(e) = 0$ $\pi_{A(m)}(d) = 1\ \forall d \in D$	Unknown = $\pi_{A(m)}(d) = 1\ \forall d \in D$
Not applicable or nonsense	$\pi_{A(m)}(e) = 1$ $\pi_{A(m)}(d) = 0, \forall d \in D$	Undefined = $\pi_{A(m)}(d) = 0, \forall d \in D$
Total ignorance	$\pi_{A(m)}(d) = 1, \forall d \in D \cup\{e\}$	Null = {1/unknown, 1/undefined}
Range[m,n]	$\pi_{A(m)}(e) = 0$ $\pi_{A(m)}(d) = 1$ if $d \in [m,n] \subseteq D$ $\pi_{A(m)}(d) = 0$ in other case	$\pi_{A(m)}(d) = 1$ if $d \in [m,n] \subseteq D$ $\pi_{A(m)}(d) = 0$ in other case
The information available is a possibility distribution, μ_a	$\pi_{A(m)}(e) = 0$ $\pi_{A(m)}(d) = \mu_a(d)\ \forall d \in D$	$\pi_{A(m)}(d) = \mu_a(d)\ \forall d \in D$
The possibility that it may not be applicable is λ and, in case it is applicable the data is μ_a	$\pi_{A(m)}(e) = \lambda$ $\pi_{A(m)}(d) = \mu_a(d)\ \forall d \in D$	Not represented

Table 4: Representation of Information in Prade-Testemale and Umano-Fukami Model [5]

Zemankova-Kandel model was published by Maria Zemankova and Abraham Kandel in 1984 and 1985[19]. The model consists of:

- A value database that stores actual data values organized in a similar way as in the possibilistic models.
- An explanatory database, which stores fuzzy subsets and relations and a set of translating rules for handling fuzzy modifiers, fuzzy connectives or adjectives.

The model adopted possibility/certainty measure in evaluating a query, q unlike the Prade-Testemale model that uses "necessity" with an established relationship with possibility: $N(X) = 1 - P(\neg X)$. The interpretation of certainty degree is not clear and no relationship exists between possibility and certainty. The model poses some limitations making it an incomplete model[5].

The GEFRED (Generalized Fuzzy Relational Database) model was published by Medina J.M, Pons O. and Vila M.A in 1994 and later expanded [20]. The GEFRED model integrates previous models developed in the possibilistic framework. The goal was to develop a model that can handle all fuzzy information. Fuzzy and non-fuzzy domains i.e. numeric values are considered. Table 5 shows the data type handled by this model.

1	A single scalar (e.g., Height = Tall, represented by the possibility distribution 1/Tall).
2	A single number (e.g., Weight = 78, represented by the possibility distribution 1/78).
3	A set of mutually exclusive possible scalars (e.g., distance = {Far, Near}, represented by {1/Far, 1/Near}).
4	A set of mutually exclusive possible numbers (e.g., Age = {25, 26}, represented by {1/25, 1/26}).
5	A possibility distribution in a scalar domain (e.g., Behavior = {0.7/Bad, 1.0/Acceptable}).
6	A possibility distribution in a numeric domain (e.g., Age = {0.6/24, 1.0/25, 0.9/26}, fuzzy numbers or linguistic labels).
7	A real number belonging to [0,1], referring to the degree of matching (e.g., Quality = 0.8).
The Umano-Fukami model data types: Unknown, undefined and NULL are inclusive.	
8	An Unknown value with possibility distribution: Unknown = {1/d : d ∈ D}.
9	An Undefined value with possibility distribution: Undefined = {0/d : d ∈ D}.
10	A NULL value given by: NULL = {1/Unknown,1/Undefined}.

Table 5: Data Types in the GEFRED Model

The model is based on generalized fuzzy domain (D) and generalized fuzzy relation(R) which comprises classic domains and classic relations respectively [5].

The models discussed above apply to Fuzzy Relational databases (FRD). Due to the increasing demand for data driven applications requiring large and complex data sets and the need for reusability and easier maintenance, the classical Fuzzy Relational database is limited in these aspects. Fuzzy Object-Oriented Database (FOOD) model is more suitable and capable of meeting these demands.

6. Fuzzy Object-Oriented Database (Food) Model

Fuzzy Object-oriented database (FOOD) is a model that allows complex objects and hierarchies such as classes, aggregation, inheritance and generalization to be implemented on a database and is capable of handling various types of uncertainties present in the data [26].

In [21], the prototypal implementation of Fuzzy Object-Oriented database model is presented. The model is defined as an extension of graph-based object in order to manage both crisp and imperfect information utilizing fuzzy set and possibility theory.

The model adopts a conceptual scheme of a quintuple: {C, T, A, H, P} in which:

C is a finite set of class names containing crisp and fuzzy classes. Crisp class collects objects with full membership and fuzzy class collects objects which have partial membership to the class.

T is a finite set of type names that could either be crisp or vague types denoting set of precise values or vague/imprecise values respectively.

A is a set of attribute names. Attributes could be simple (when domain is a type) or complex (when domain is a class). Single-valued and multi-valued attributes exists.

P is the property relation. P relates one class, Ci with its attribute names and both classes, Ci and Cj with the domain.

H is the inheritance relation that defines class hierarchies. If the inheritance relation is fuzzy then a label defines the extent to which instances of a subclass are instances of a super class.

6.1. Application of Fuzzy Object-Oriented Database (FOOD) model

In order to handle impreciseness in data, some modern data types such as text files, images, audio, video, medical and research data, spatial data, astronomical data etc. have developed the concept of Fuzzy Object-Oriented database in designing, storing and implementing of such data type [2].

Video content extraction, storage and retrieval, medical diagnosis system, Multimedia database, Geographical Information system are some applications that uses complex and fuzzy information and requires the implementation of Fuzzy Object-oriented database model.

Some applications and prototypes that adopts FOOD model are presented in [9], [10], and [11].

The authors in [9], presented a framework for the automatic extraction of semantic content of a raw video and its storage in a Fuzzy Object-oriented database Management System. The database is capable of storing fuzzy information contained in a video. Users' queries can be retrieved from the database via an interface. Similarly, [10] proposed a framework that integrates Fuzzy object-Oriented Database to a Fuzzy Knowledge base for video applications to handle data and semantic rules applicable to video database applications domains. In [11], a data cube representation for patient diagnosis system using Fuzzy Object-Oriented Database is presented in order to make information retrieval faster.

6.2. Limitations of Fuzzy Object-Oriented Database (FOOD) Model

FOOD model does not have a standard query language unlike the Fuzzy Relational Database model that has adopted SQL as its standard language [12]. The Fuzzy Relational database model seems to capture uncertainty in a more formal and better way.

In order to integrate the strengths of Fuzzy Relational and Object-oriented Database models, [27] proposed a Fuzzy Object-Relational Database model which is a fuzzy extension of the basic Object- Relational Database construct. It takes advantage of the Object-Relational Database systems in order to manipulate, represent, store and retrieve fuzzy data irrespective of its complexity.

7. Conclusion

Integration of fuzzy information in database models and approaches to handle them has been an ongoing research area in databases considering the complex and fuzzy nature of data. Previously, research was majorly on relational models. However, the need to develop a model that could represent complex data with fuzziness came up. Fuzzy Object-Oriented database has the capacity to represent complex data with imprecision. Fuzzy Object-Relational database has advantage over the precious models since it exhibits their attributes. With the rise of complex data from social media and other data sources, current research is being done in extending these models to handle more complex data.

This paper gives an overview on various approaches that have been adopted in handling information imprecision with practical illustrations.

8. References

i. Motro, A., "Management of uncertainty in database system" in Modern database system: The object model, interoperability and beyond W. Kim(Ed.), Addison-Wesley,1995, chapter 22, pages 457-476.[online]. Available at https://pdfs.semanticscholar.org/448d/267dd580609e725129ad6aa6815a35fdc65b.pdf.Accessed: December 20, 2019

ii. Puja Shrivastava, Prof. (Dr.) Laxman Sahoo, "An overview of Fuzzy Object oriented Databases(FOODs)" in International Journal of Scientific & Engineering Research, Volume 4, Issue 7, July 2013.[online]. Available: https://www.ijser.org/researchpaper/An-Overview-of-Fuzzy-Object-Oriented-Database-Systems-FOODBs.pdf. Accessed: November 29, 2019

iii. Codd E.F, "Extending the database relational model to capture more meaning", ACM transactions on Database systems, vol. 4, No.4, December 1979. Available: https: //gertjans.home.xs4all.nl/usenet/microsoft.public.sqlserver.programming/codd1979.pdf.[online].Accessed: December 26, 2019.

iv. "NULLSQL". Available: https://en.wikipedia.org/wiki/Null (SQL).Accessed: December 31, 2019.

v. Galindo, J, Urrutta A., and Piatini, M., "Fuzzy Database Approaches" in Fuzzy databases: Modelling, design and Implementation. Hershey, PA: Idea & Group Publishing, 2006, chapter 2, page 45-59. [online].Available: http://www.boente.eti.br/fuzzy/ebook/ebook-fuzzy-galindo.pdf. Accessed: October 29, 2019.

vi. Codd E.F, "Missing information". The Relational Model for Database Management, Version 2. Reading, MA: Addison-Wesley,1990.[online] Available at https://codeblab.com/wp-content/uploads/2009/12/rmdb-codd.pdf. Accessed: December 27, 2019.

vii. Mohammed Ali Ben Hassine, Amel Grissa-Touzi, Jose Galindo, "How to achieve Fuzzy Relational Databases managing Fuzzy data and meta data" in Handbook of Research on Fuzzy Information Processing in Databases, IGI global, 2008, chapter 12, page 351-380[online].Available: https://www.researchgate.net/publication/228095525_How_to_Achieve_Fuzzy_Relational_Databases_Managing_Fuzzy_Data_and_Metadata.Accessed:January 10, 2020

viii. Slawomir Zadronzny, Guy de Tre, Rita de Caluwe, Janusz Kacprzyk, "An overview of Fuzzy Approaches to Flexible Database Querying", in Handbook of Research on Fuzzy Information Processing in Databases, Galindo J.(Ed.), Volume I, IGI global, Chapter 2,pp.54[online]. .Available:https://www.ibspan.waw.pl/~kacprzyk/papers/Handbook-Galindo-2008.pdf.Accessed: December 12, 2019

ix. Koyuncu, M., Yilmaz, T., Yildirim Y., Yazici, A., "A Framework for Fuzzy Video Content Extraction, Storage and Retrieval" in IEEE Int. Conf. on Fuzzy Systems, 2010.

x. Nezihe Burcu Ozgura, Murat Koyuncub, Adnan Yazicia, "An intelligent fuzzy object-oriented database framework for video database applications" in Fuzzy Sets and Systems 160 (2009) 2253–2274, doi:10.1016/j.fss.2009.02.017.

xi. Shweta Dwivedi, Dr. Santosh Kumar, "Data Cube Representation for patient Diagnosis System Using Fuzzy Object-Oriented Database" in Transactions on Machine Learning and Artificial Intelligence, Volume 5, Issue 1, Feb 2017.

xii. Sonia, Subita Kumari, "Fuzzy Object-Oriented Database versus FRDB for Uncertainty Management" in International Journal of Computer Applications (0975 – 8887) Volume 74– No.17, July 2013, doi: 10.14738/tmlai.51.2778.

xiii. Umano, M., & Fukami, S., "Fuzzy relational algebra for possibility distribution- fuzzy-relation model of fuzzy data" in Journal of Intelligent Information Systems, 3, 7-27 (1994).

xiv. Buckles, B. P., & Petry, F. E., "Uncertainty models in information and database systems" in Information Sciences, 11, 1985, page 77-87.

xv. Prade, H., & Testemale, C., "Fuzzy relational databases: Representational issues and reduction using similarity measures" in Journal of the American Society for Information Science, 38(2), 1987.

xvi. Prade, H., "Lipski's approach to incomplete information databases restated and generalized in the setting of Zadeh's possibility theory" in Information Systems, 9(1), 27–42, 1987.

xvii. Prade, H., "Generalizing database relational algebra for the treatment of incomplete or uncertain information and vague queries" in Information Sciences, 34(2), 115–143, 1984.

xviii. Prade, H., & Testemale, C. (1985), "Representation of soft constraints and fuzzy attribute values by means of possibility distributions in databases". Available: https://dl.acm.org/doi/pdf/10.5555/3157680.3157710?download=true Accessed: January 14, 2020.

xix. Zemankova-Leech, M., & Kandel, A., "Implementing imprecision in information systems" in Information Sciences, 37, 107-14, 1985.

xx. Medina, J. M., Pons, O., & Vila, M. A., "GEFRED: A generalized model of fuzzy relational databases" in Information Sciences, 76(1/2), 87-109, 1994.

xxi. Bordogna, G., Leporati, A., Lucarella, D., & Pasi G. The fuzzy object-oriented database management system in Recent issues on fuzzy databases, G. Bordogna & G. Pasi (Eds.), Physica-Verlag, 2000, pp. 209-236.

xxii. Barbara, D., Garcia-Molina, H., & Porter, D, The management of probabilistic data in IEEE Transactions on Knowledge and Data Engineering, 4, pp. 487-502, 1992.

xxiii. Bellman, R. E., & Zadeh, L. A., Decision making in a fuzzy environment in Management Sciences, 17(4), pp. 141-175. [online] Available: http://ayouty.com/___Gam_Shorouq_final/5info_sys/FUZZY_DECISION.pdf

xxiv. Z. M. MA, Li Yan, "A Literature Overview of Fuzzy Database Models" in Journal of Information Science and Engineering, 24, PP.189-202, 2008.

xxv. Anupriya, PROF. Rahul Rishi, "Review of Fuzzy Logical Database Models" in IOSR Journal of Computer Engineering (IOSRJCE), Volume 8, Issue 4 (Jan. - Feb. 2013), PP 24-30.

xxvi. Priyanka J. Pursani, Prof A. B. Raut, "FI: Fuzzy Object Oriented Database (Food) Index" in International Journal of Advanced Research in Computer and Communication Engineering Vol. 3, Issue 3, March 2014.[online] Available: https://pdfs.semanticscholar.org/77b4/092f77ae14b7c124940dc49d69e78a1a8ea9.pdf. Accessed on November 30, 2019

xxvii. Carlos D Barranco, Jesus R. Campana, Juan M. Medina, "Towards a Fuzzy Object-relational Database Model" In Handbook of Research on Fuzzy Information Processing in Databases, IGI global, 2008, chapter 17, page 435 461.