ISSN 2229-5518

Systematic Review of the State of the Art in Fuzzy Database Model

Otokwala, Uneneibotejit Job; Ofut, Ogar Tumenayu

Abstract— Most real-world data are vague, imprecise and imperfect and so classical relational databases are lacking in ability to integrate and manage them. The fuzzy database model uses the fuzzy set and fuzzy logic to extend the Classical relational database models as a functional way of supporting and managing imprecise, vague and imperfect data especially in Multiple Criteria Decision-Making. The Systematic review of the State of the Art in Fuzzy Database Model is therefore, a review and condensation of the various approaches by different authors to integrate the crispy and imperfect data.
.

Index Terms—Entity-Relationship, Fuzzy, Fuzzy Extended Entity-Relationship, imperfect, IFO, ExIFO, imprecised data, Non-Formal relation, vague.

---

1 INTRODUCTION

A database is an ordered collection of related data elements intended to meet the information needs of an organization and it is designed to be shared by multiple users [3]. The type of data and the values of attributes are not always known with sufficient precision [1]. This is because, most of the data are fuzzy, vague and complex either by their nature or by non-ideal measurement and the uncertainty arising from the fuzziness are always ambiguous [2,4]. The main motivation therefore for using fuzzy method lies in the need to resolve the fundamental problem of integration of crispy and imperfect data in the database [6]. By design, Relational Databases are based on Boolean logic with a bistable output of {1 or 0; true or false}. The fuzzy database approach measures information on the degree of truth and it has become the most convenient way to store and manage imprecise data.

The fundamental components of the fuzzy database model are the fuzzy logic and the fuzzy set. While the fuzzy logic uses a combination of various mathematical principles to represent knowledge depending on a gradual degree of membership, the fuzzy set theory on the other hand, provides a robust framework for systematically handling of uncertainty based on fuzziness [5,6].

- *Otokwala, Uneneibotejit Job is currently pursuing his PhD degree program in Computer Science at the University of Nigeria, Nsukka, Nigeria, PH-+2349035643372. E-mail: uneneibotejit.otokwala.pg81070@unn.edu.ng*
- *Ofut, Ogar Tumenayu is of the department of Computer Science, Cross River State University of Technology, Calarbar, Nigeria, PH-+2349099931484. E-mail: ofutogar@crutech.edu.ng*

2 LITERATURATURE REVIEW

Real world applications, data and information often times are imprecise, uncertain and vague and many sources contribute to it [16]. For example, we are increasingly faced with large volumes of data that are generated via both traditional and non-traditional means (*e.g.*, sensors, camera, genome, biological and geographical systems, etc). Because these data are imprecise, imperfect and uncertain, they pose a significant problem in terms of incorporation, representation and manipulation in the traditional Relational Database [14]. The introduction of fuzzy logic by Zadeh helped in the extension and integration of fuzzy data through different data models [5].

3 STATE OF THE ART IN FUZZY DATABASE MODEL

The uncertainty and incomplete data representation which are viewed as disadvantages in the ER model necessitated the use of fuzzy sets and fuzzy logic to extend existing relational database models [8]. The State of the Art in Fuzzy Database Model are the several approaches by different authors on the extension and implementation of the fuzzy database model and techniques. The approaches are:

3.1 Chen and Kerre Approach.

The Chen and Kerre's approach introduced the fuzzy extension where the superclass and subclass relationship concepts of the ER model are extended using fuzzy logic [10,11]. The basic idea is to define a set of members over a universal space such that, if E1 is a superclass of E2 and e ЄE2, then E1(e) ≤ E2(e), where E1(e) and E2(e) are the membership functions of e to E1 and E2, respectively [13]. Chen and Kerre further discussed three kinds of constraints with respect to fuzzy relationships. The constraints are:

ISSN 2229-5518

(1) The inheritance constraint. This constraint proposes that an instance of a subclass inherits all relationship instances in which it participated as a superclass entity.
(2) Total participation constraint. Defined when each entity, a*i* >0 in the entity set (E ∃ ⊚i) occurs in at least one relationship in that relationship set.
(3) The cardinality constraints. In an Entity–Relationship (ER) schemas, this constraint specifies the dependencies among the entities. A conventional simplified cardinality notation uses 1 for mini and maxi, and a letter (e.g., n) for mini ≥ 0, maxi = N [12,13]. There are kinds of cardinality constraints, and they could be represented thus: 1:1, 1: N, and N:M relationships.

If *E*, *R*, and *A are:* fuzzy entity type, fuzzy interrelation type, and fuzzy attribute set of the fuzzy ER model, and if μ_E, μ_R, and μ_A be their membership functions, then Chen and Kerre label types can be represented in the figures below.

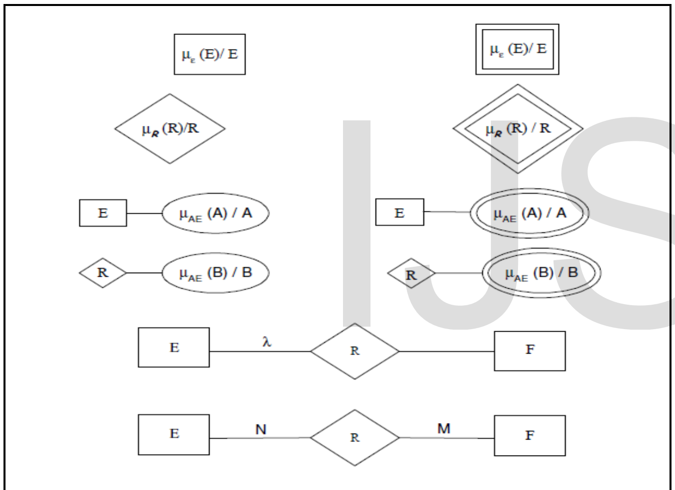

Figure 1 showing ER fuzzy notation proposed by Chen [12]

3.2 Zvieli and Chen's Approach

Imprecision at the modelling level was paramount in the minds of Zvieli and Chen and so, they offered the first approach to incorporate fuzzy logic into the extension of ER model. They adapted a design methodology for FRDBs, which contains extensions for representing the imprecision of data in the Entity-Relationship (ER) data model. They also proposed a set of steps for the derivation of a Fuzzy Relational Database (FRDB) from this extended ER model [9,12]. Zvieli and Chen allowed fuzzy attributes in entities and relationships and introduced three levels of fuzziness in the ER model. The three levels are: [9]

1. At the first level, entity set types, relationships and attributes may be fuzzy and thus have a membership degree. For example, in Figure 2, the fuzzy entity "Company" has a 0.9 membership degree, the relationship "Accepts" has a 0.7 membership degree, and the fuzzy attribute "EmailAddress" has a 0.8 membership degree.
2. The second level is related to the fuzzy occurrences of entities and relationships. Where instances belong to the entity or relationship with different membership degrees. For example, an entity "Young_Employees" must be fuzzy, because its instances, its employees, belong to the entity with different membership degrees.
3. The third level concerns the fuzzy values of attributes of special entities and relationships. For example, attribute "Quality" of a basketball player may be fuzzy (the possibilities include bad, good, very good, and so on). [11, 12, 14]

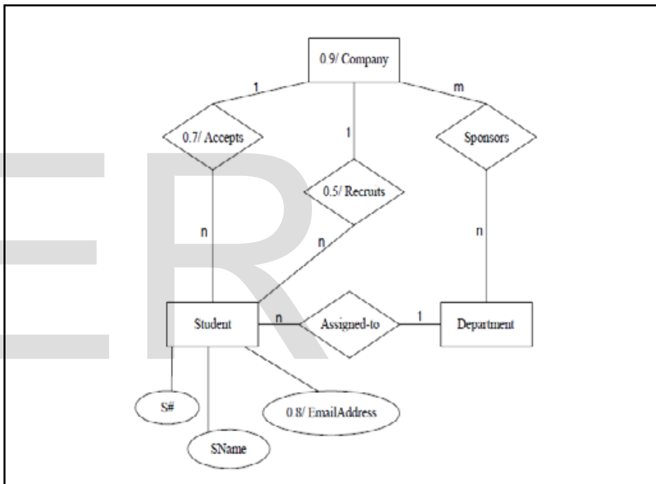

Figure 2, showing membership degrees to the model in some sets (entities, relationships, or attributes): The first level of the Zvieli and Chen approach. [12]
.

Galindo et al, [12] in analysing the proposal by Zvieli and Chen, observed that the first level may be useful, but that there is a need to decide whether such an entity, relationship, or attribute should or should not appear in the implementation phase. The second level is useful, but it is important to consider other varying degree of meanings (membership, importance, fulfilment degree, and so on). The third level is only useful to the extent that it is similar to the data type of some of the attributes.

3.3 Yazici and Merdan Approach

IFO data model is a mathematically defined data model that combines the fundamental principles of "semantic" database modeling using a graph-based formalism. It uses directed graph with various types of vertices and edges which represents atomic objects, structured objects, functional fragments and ISA relationships between them [15]. Yazici and Merdan studied this model and adopted the IFO model in order to incorporate imprecise attributes. They thereafter proposed an extension of the IFO model to ExIFO for the processing of imperfect data with special treatment of data where similarity exists in a label [12,15]. The implementation and validation of the representation of a fuzzy conceptual scheme is by looking at a representation of uncertain attributes. Also, they proposed 3 constructors in the conceptual ExIFO model such that the constructors will allow imprecision and uncertainty in database models, based on the IFO conceptual model. They use fuzzy values like: true attributes, incomplete-valued attributes, and null-valued attributes in their illustration [12]. In the first case for example, consider a Set of Real numbers, R= {1, 2, 3, 4, 5} and a subset, x= {1, 3, 5}; there exist a similarity relation between the domain of the real attributes and the subset, x϶R. The second valued-attribute is the incomplete attribute where the domain is non-specific but only provided a range of numbers (e.g. between 10 and 20) which is a classical incomplete attribute. In the third, the true data value is available but it is not expressly precise. An example of this attribute may be, whether a certain number exist. Note, the main contribution of this approach is the use of an extended Non-First Normal Form relation (NF2) which is aimed at transforming the conceptual design into a logical design [12].

3.4 Chaudhry, Moyne and Rundensteiner Approach

Chaudhry et' al are one of the many authors that proposed a method for the extension of the classical relational database. Their method proposes the extension of the ER model of Zvieli and Chen through a sequence of steps that maps the fuzzy EER model to the fuzzy relational database. The two types of imprecisions they considered are: (i) the imprecision in the degree of membership of a tuple in a relation, and (ii) the imprecision in a data value. According to them, "firstly, present the fuzzy relation construct that expresses the imprecision in the degree of membership of a tuple in a relation, and then the possibilistic relation construct that expresses the imprecision in a data value" [17]. Galindo et' al [12] defined *n* linguistic labels as *n* fuzzy sets over the universe of an attribute with each tuple in the crisp entity transformed up to the level of the *n value* of fuzzy. Each fuzzy tuple (or value) does not store the crisp value but a corresponding linguistic label and a degree of membership to which the corresponding crisp entity belongs in the new entity. The crisp entities and the new fuzzy entity are then mapped to separate tables. The design sequence for the extension of FRDBs are: refer to fig. 6

Step 1: Constructing an extended fuzzy ER data model.
Step 2: Transforming the ER model to relational tables.
Step 3: Normalization of the relations.
Step 4: Ensuring correct interpretation of the fuzzy relational operators.

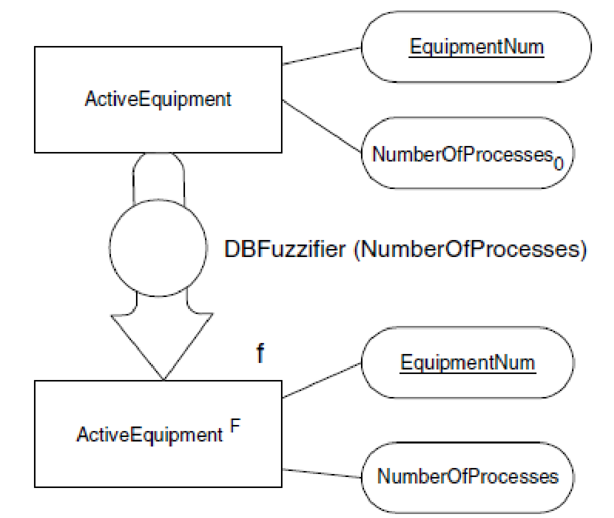

Figure 3. showing the model proposed by Chaudhry, Moyne, and Rundensteiner (1994): Example of DBFuzzifier transformation [12,17].

3.5 Buckles and Petry Model

The Buckles-Petry Model was the first model that utilizes similarity relations in the relational model. It provides a structure for the representation of inexact information in the form of a relational database. This structure differs from ordinary relational databases in two aspects: (1) components of tuples need not be single values and (2) a similarity relation is required for each domain set of the database. In this model, assume a fuzzy relation is defined as a subset of the following Cartesian product: P(D1) × ... × P(Dm), where P(Di) represents the elements of the domain (Di), including all the subsets that could be considered within the Di domain [20]. The three data types that this proposal permits are: finite set of scalars, finite set of numbers and fuzzy number set [17].

3.6 Ma, Zhang, Ma, and Chen Approach

Ma, Zhang, Ma, and Chen reviewed the work of Zvieli and Chen especially the three levels which they then incorporated into the Fuzzy Extended Entity-Relationship model (FEER model). This approach tries to manage complex objects in the real world at conceptual level and associate their importance to the degree of each of the components (attributes, entities, etc.). So much restriction was however imposed because of their generalization of definitions for specialization, category, and aggregation [12]. Furthermore, in 2004, they introduced an extended object-oriented database model to handle imperfect, imprecise as well as complex objects. The modules of the EOODBM that they extended are: objects, classes, objects-classes relationships, subclass/superclass, and multiple inheritances. Here are **s**ome FEER notations they proposed:

a) fuzzy attributes, entities, and interrelations single-valued attribute type

b) specialization, aggregation, and fuzzy categories

4 POSSIBILISTIC MODELS.

Possibility theory is premised on the idea of how linguistic variables are related to fuzzy sets. In this way, it is then possible to evaluate the possibility of the variable X belonging to set Y just like the membership degree of X element in Y [18]. Below are examples of possibilistic model:

4.1 The Prade-Testemale Model.

This is a FRDB model that allows the integration of incomplete or uncertain data in the possibility theory and the relations corresponding to knowledgebase. After integration, the uncertain data are then stored in form of tables irrespective of the fact that differences may occur in the type of values in the columns. For example, an attribute A, having a X domain with e as a special element denotes a scenario where A is not applied to y [18]. Thus, the values of A for a y object can be represented by a possibility distribution $\pi A(y)$ about $X \cup \{e\}$ such that the PD, $\pi A(y)$ is an application that goes from $X \cup \{e\}$ to the [0, 1] interval. [18]

4.2 The Zemankova-Kaendel Model.

This model dates back to 1984 and 1985 and it is premised on three databases: value database, explanatory database and set of translating rules. Value database data are ordered in ways similar to the possibilistic models while on the other hand, explanatory databases have fuzzy subsets and fuzzy relations stored in them. The set of translating rules, are the various measures for handling of adjectives and modifiers [18]. The possibility measure, PA(S) is used to find the compatibility of the fuzzy subset, S of the condition, with an attribute A value for each tuple in the relation is given as $PA(S) = \sup_{x \in X} \{\mu F(x) \cdot \pi A(x)\}$.

4.3 The GEFRED Model.

The Generalized Fuzzy Relational Database (GEFRED) model was proposed in 1994 by Medina-Pons-Vila. In its development, the fuzzy domain was considered within the framework of possibilistic model. GEFRED is designed to contain unknown, undefined and null values hence its ability to handle various datatypes. It also redefines the relational algebraic operators like: union, intersection, difference, Cartesian product, projection, selection, join and division in the generalized fuzzy relational algebra. [18]

5 OBSERVATIONS & FUTURE RESEARCH

The approaches by the authors on the use of fuzzy model to integrate imprecise and imperfect data into the database have aligned with the benefits of the fuzzy model which are based on a generality of function estimators: clarity, modularity, ability to be explained, easy handling of uncertainty, and parallel processing of rules [6]. There are however, some very important drawbacks which portends major limitations to the fuzzy model that the authors did not factor in. The drawbacks are: the high computational costs, severe computing power restrictions, comprehensibility, and optimization [21]. Future research should therefore incorporate the computational complexities and the severe computing restriction. Typical references are the limitations inherent in bioinformatics settings such that the computational complexities have created hurdles for crispy data defuzzification [6].

6 CONCLUSION

Fuzzy set and fuzzy logic have become useful tools for accurate modeling and integration of real-world uncertain, imprecise and imperfect data. Condensing the several approaches of the extension of the traditional classical databases in this research work was with a view to presenting an up-to-date State of the Art in fuzzy database modeling, the limitations and the future research areas.

7 REFERENCES

[1] M. Hudec, *Fuzzy Data in Traditional Relational Databases, IEEE conference of Neurel, Belgrade (2014), 195-199*.

[2] Ankita Srivastava*and Awadhesh Kumar Sharma. *A Study and Comparison of Methods for Fuzzy Data Equivalence Volume 3, No. 3, May-June 2012 International Journal of Advanced Research in Computer Science ISSN No. 0976-5697.*

[3] Amit Garg & Dr. Rahul Rishi. *Global Journal of Computer Science and Technology Volume 12 Issue 6 Version 1.0 March 2012*

[4] F. Pavese. Why should correction values be better known than the measured true value? *Journal of Physics*: Conference series 459, (2013).

[5] Miroslav Hudec. *Collecting, storing and managing fuzzy data in statistical relational databases*

[6] Tuqyah Abdullah Al Qazlan, Aboubekeur Hamdi-Cherif, and Chafia Kara-Mohamed. State of the Art of Fuzzy Methods for Gene *Regulatory Networks Inference Scientific World Journal, Volume 2015 (2015), Article ID,11 pages* http://dx.doi.org/10.1155/2015/148010

[7] https://books.google.com.ng/books?isbn=3319159704

[8] B.P. Buckles and F.E. Petry, "A fuzzy representation of data for relational databases" *Fuzzy Sets and Systems*, 7, pp. 213–226, 2011.

[9] A. Zvieli and P. Chen, "ER modeling and fuzzy databases" *Proc. of the Second International Conference on Data Engineering*, 2008, pp. 320-327.

[10] G.Q. Chen and E.E. Kerre "Extending ER/EER concepts towards fuzzy conceptual data modeling" *Proc. of the IEEE International Conference on Fuzzy Systems*, 1998, pp. 1350-1325.

[11] E.E. Kerre and G.Q. Chen "Fuzzy data modeling at a conceptual level: Extending ER/EER concepts" *Knowledge management in fuzzy databases*, 2001, pp. 3-11

[12] J. Galindo, A. Urrutia, M. Piattini, *Fuzzy Databases: Modeling Design and Implementation*, Hershey, USA: IDEA Group, 2006

[13] Mira Balaban et tal, *Enforcing Cardinality Constraints in the ER Model with Integrity Methods1*. Ben-Gurion University of the Negev, Israel Journal of Database Management, 11(4), 2002, pp. 14-23

[14] Z. M. Ma+ And Li Yan. A Literature Overview of Fuzzy Conceptual Data Modeling Journal *of Information Science and Engineering 26, 427-441 (2010) 427*

[15] Adnan YAZICI and Ali ~INAR. Conceptual Modeling for the Design of Fuzzy Object Oriented Databases. Springer-Verlag Berlin Heidelberg 2000 https://link.springer.com/chapter/10.1007/978-3-7908-1865-9_2

[16] S. Parsons, "Current approaches to handling imperfect information in data and knowledge bases," *IEEE Transactions on Knowledge and Data Engineering*, Vol. 8, 2005, pp. 353-372

[17] Nauman A. Chaudhry, James R. Moyne, Elke A. Rundensteiner. A Design Methodology for Databases with Uncertain Data. https://www.computer.org/csdl/proceedings/ssdm/1994/6610/00/00336964.pdf

[18] Anupriya, Prof. Rahul Rishi, Mira Balaban et tal. Review of Fuzzy Logical Database Models. *IOSR Journal of Computer Engineering (IOSRJCE) ISSN: 2278-0661, ISBN: 2278-8727 Volume 8, Issue 4 (Jan. - Feb. 2013), PP 24-30.* www.iosrjournals.org www.iosrjournals.org

[19] Ms. Poonam, Banasthali Vidyapith, Rajasthan. Fuzzy Database Data Models and Its Framework. *International Journal of Computer Science Trends and Technology – Volume 3 Issue 2, Mar-Apr 2015.*

[20] The Entity-Relationship and Enhanced Entity-Relationship Model. *http://www.peterlo.com/Teaching/IT354/L02.pdf*

[21] N.M. de Reus. Assessment of benefits and drawbacks of' using fuzzy logic. NWO especially in fire control systems II *https://apps.dtic.mil/dtic/tr/fulltext/u2/a285428.pdf*