

Review of Fuzzy Logical Database Models

Anupriya¹, Prof. Rahul Rishi²

¹(Department of Information Technology, AIM &ACT, Banasthali University, Rajasthan, India)

²(Department of Computer Science & Engg, UIET (Maharshi Dayanand University), Haryana, India)

Abstract - Fuzzy set theory has been extensively applied to extend various database models (conceptual and logical) and resulted in numerous contributions, mainly with respect to the popular relational model or to some related form of it. So this paper reviews fuzzy logical database models, in which fuzzy relational databases are discussed.

Keywords: Fuzzy databases, fuzzy set, possibility distribution, database models.

I. Introduction

Information given in real-world applications is often vague, imprecise and uncertain. Ignoring the imperfect nature of real-world will introduce some deformation of human perception of real-world. It may eliminate several substantial information. That information may be very useful in several data-intensive applications. At present time, the Relational Data Model is worldwide used tool for building database systems and applications. Relational Database Management Systems have developed a powerful technology for data storage and retrieval. Nevertheless, these systems suffer the problem of rigidity. So, fuzzy databases have been introduced. Some user requirements may not be solved directly by a classical querying system like as “Find all middle age male patients whose blood pressure is high.” So for these types of queries, various data models have been introduced.

A significant body of research in the area of fuzzy database modeling has been developed over the past thirty years and tremendous gain is hereby accomplished in this area. Various fuzzy database models have been proposed. Some major issues related to these models have also been investigated.

The remainder of this paper is organized as follows. Section 2 gives the basic knowledge about imperfect information and fuzzy sets theory. Issues about fuzzy relational database models are described in section 3. The last section concludes this paper.

II. Imperfect Information and Fuzzy Sets Theory

2.1. Imperfect Information

Fuzzy databases are used basically to handle imperfect information. Imperfect information can be inconsistent, imprecise, ambiguous, uncertain or vague. According to Z.M. Ma, imperfect information is [1]

1. Inconsistent when some real world aspect is having more than one value. (Example – the age of a patient is stored as 25 and 27.)
2. Imprecise when attribute's value choice has to be made from any given interval or range. (Example – the age of a patient is the set [23, 24, 25, 26] or the Heartbeat rate is in interval [50-65].)
3. Ambiguous when some elements of information lead to various possible interpretations.
4. Uncertain when there exists degree of truth in attribute value. It occurs due to lack of information. (Example- the possibility that age of a person is 30 is 80 %.)
5. Vague when attribute value is represented by linguistic variables. (Example – the age of patient is ‘Young’.)

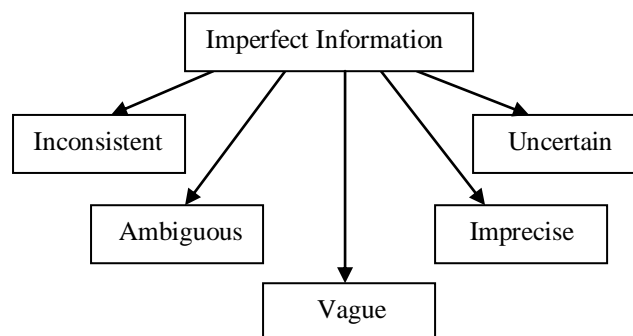


Fig.1 Types of imperfect information

2.2. Basic Preliminaries on Fuzzy Set Theory

Let $X = \{x_1, x_2, \dots, x_n\}$ be universe of discourse.

2.2.1. Fuzzy Sets

Fuzzy sets are extension and generalization of the basic concepts of crisp sets. It allows partial membership. A fuzzy set F over a universe of discourse X can be defined as set of ordered pairs[2].

$$F = \{\mu_F(x) / x : x \in X, \mu_F(x) \in [0,1]\}$$

Here, $\mu_F(x)$ is membership function of the element x to the fuzzy set F . So, a fuzzy set is thus defined by a function that maps objects in a domain of concern to their membership value in a set. Such a function is known as membership functions (μ) whose value lies in range $[0,1]$ i.e.

$$\mu_F(x): X \rightarrow [0,1]$$

$\mu_F(x) = 0$ indicates that x does not belong to the fuzzy set F .

$\mu_F(x) = 1$ indicates that x completely belongs to the fuzzy set F .

The universe of discourse X can be classified as

1. Finite or discrete universe of discourse $X = \{x_1, x_2, \dots, x_n\}$, where a fuzzy set F can be represented by:

$$F = \mu_1 / x_1 + \mu_2 / x_2 + \dots + \mu_n / x_n$$

Here, μ_i with $i = 1, 2, \dots, n$ represents the membership degree of the element x_i .

2. Infinite universe of discourse, where a fuzzy set F over X can be represented by:

$$F = \int \mu_F(x) / x$$

2.2.2. Fuzzy Union or t-conorms

If A and B are two fuzzy sets of the universe X , then the fuzzy union of A and B is denoted by $A \cup B$ and is defined as

$$A \cup B = \{(x, \max \{ \mu_A(x), \mu_B(x) \}) : x \in X\}$$

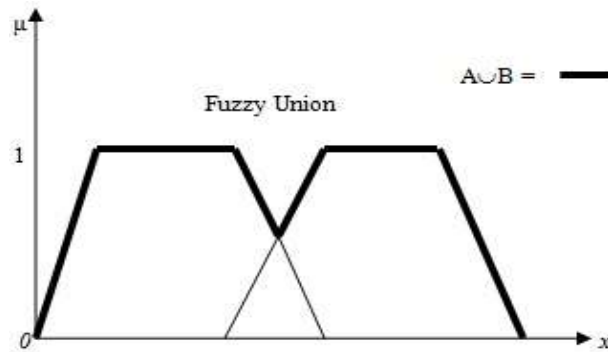


Fig. 2 Representation of a Fuzzy Union

2.2.3. Fuzzy Intersection or t-norms

If A and B are two fuzzy sets of the universe X , then the fuzzy intersection of A and B is denoted by $A \cap B$ and is

defined as

$$A \cap B = \{(x, \min \{ \mu_A(x), \mu_B(x) \}) : x \in X\}$$

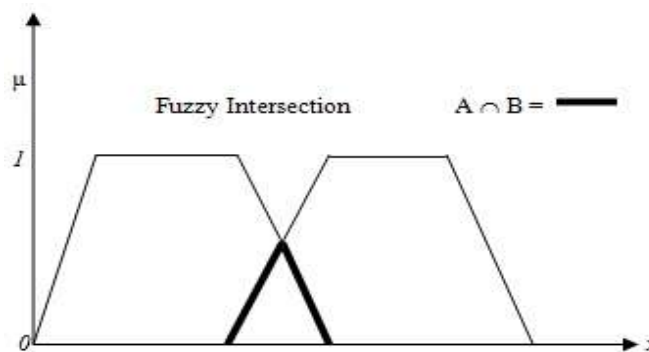


Fig. 3 Representation of a Fuzzy Intersection

2.2.4. Fuzzy Difference

The difference of set A with respect to set B is the collection of all elements in the universe that belong to A but do not belong to B.

$$A - B = A - (A \cap B)$$

$$B - A = B - (B \cap A)$$

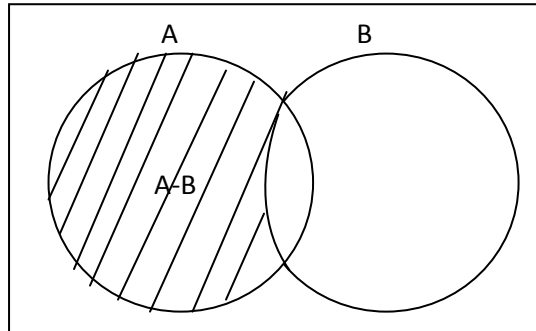


Fig. 4 Representation of a Fuzzy Difference

2.2.5. Fuzzy Relation

Let A and B be two sets. A fuzzy relation R from A to B is a fuzzy set on $A \times B$ and is denoted by $R(A \rightarrow B)$.

2.2.6. Fuzzy Tolerance Relation

A fuzzy relation $R(A \rightarrow A)$ is said to be

- i) reflexive : iff $\forall x \in X, \mu_R(x, x) = 1$
- ii) symmetric : iff $\forall x_1, x_2 \in X, \mu_R(x_1, x_2) = \mu_R(x_2, x_1)$

So, the fuzzy relation is said to be fuzzy tolerance relation if it is reflexive and symmetric.

2.2.7. Fuzzy Equality

If A and B are two fuzzy sets of universe X, then A and B are equal if $A \supseteq B$ and $A \subseteq B$.

$$A = B \Leftrightarrow \forall x \in X, \mu_A(x) = \mu_B(x)$$

2.2.8. Fuzzy Quantifiers

Fuzzy or linguistic quantifiers [3] allow us to express approximate idea of number of a subset's elements fulfilling a certain condition about fuzzy quantities. Linguistic quantifiers or labels express fuzziness because the definition of these labels varies from one person to another. It also varies with the context in which it is applied.

Fuzzy quantifiers can be *absolute* or *relative*:

- *Absolute quantifiers* express quantities over the total number of elements of a particular set. Example, "much more than 20", "close to 90", "a small number of", "approx 22" etc.
- *Relative quantifiers* express measurements over the total number of elements, which fulfill a certain condition depending on the total number of possible elements (the proportion of elements). Consequently, the truth of the quantifier depends on two quantities. This type of quantifier is used in expressions such as "the majority" or "most," "the minority," "little of", "about half of," and so forth. In this case, in order to evaluate the truth of the quantifier, we need to find the total number of elements fulfilling the condition and to consider this value with respect to the total number of elements that could fulfill it (including those that do fulfill it and those that do not).

2.2.9. Possibility and Necessity Measures

It measures the extent to which A and B superpose each other, denoted as $Poss(A, B)$ and defined as [4]:

$$Poss(A, B) = \sup_{x \in X} [\min (A(x), B(x))]$$

The necessity measure describes the degree to which B is included in A and is denoted by $Nec(A, B)$ [4]:

$$Nec(A, B) = \inf_{x \in X} [\max (A(x), 1 - B(x))]$$

III. Fuzzy Relational Databases

Fuzzy relational databases models are based on either similarity relations or possibility theory. The Possibility theory is based on the idea of linguistic variables and how they are related to fuzzy sets So, the various models based on these relations are as follows [4]:

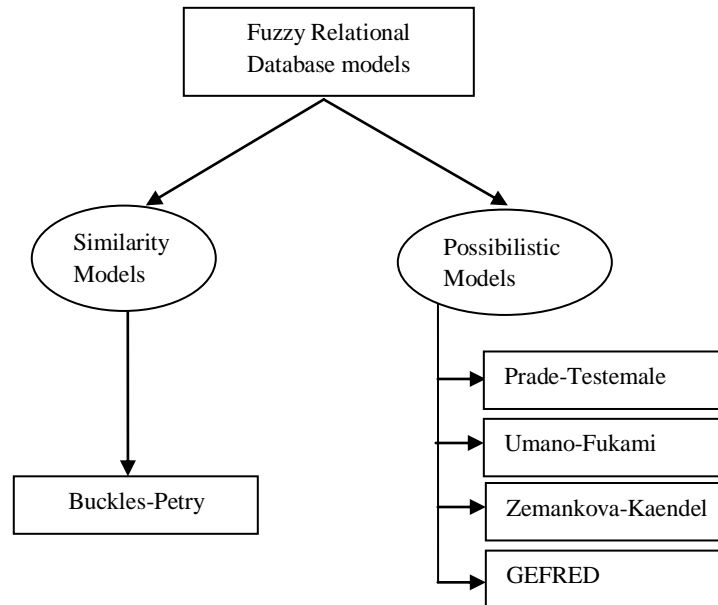


Fig. 5 Fuzzy relational databases models

3.1. Similarity Models

3.1.1 The Buckles-Petry Model

The first model that utilizes similarity relations [5] in the relational model was Buckle-Petry Model. It was proposed by Buckles and Petry [6]. In this model, a fuzzy relation is defined as a subset of the following Cartesian product: $P(D_1) \times \dots \times P(D_m)$, where $P(D_i)$ represents the parts set of a D_i domain, including all the subsets that could be considered within the D_i domain (having any number of elements).

Three data types are permitted by this model are finite set of scalars (labels), finite set of numbers and fuzzy number set. The similarity values given by users are standardized in the $[0,1]$ interval, where 0 corresponds to “totally different” and 1 to “totally similar.”

A value between 0 and 1, can be taken as a *similarity threshold* to get the values whose similarity is greater than the threshold or to consider those values undistinguishable.

3.2. Possibilistic Models

Possibility theory is based on idea how linguistic variables are related to fuzzy sets. In this way, we can evaluate the possibility of variable X belonging to set Y. It is just like the membership degree of X element in Y. Possibilistic models are

1. Prade-Testemale Model
2. Umano-Fukami Model
3. Zemankova-Kaendel Model
4. GEFRED Model

3.2.1. The Prade-Testemale Model

Prade and Testemale published an FRDB model that allows the integration of incomplete or uncertain data in the possibility theory [7]. The relations corresponding to knowledgebase are stored in form of tables as in Codd's relational model. The difference may occur in the type of values in columns. In Codd's classic relational model [8], atomic values are stored in attributes of a tuple. While, in Prade-Testemale model, attribute value of a tuple can be a possibility distribution. An attribute A, having a X domain, is considered. e is a special element denoting the case in which A is not applied to y. All the available information about the values of A for an y object can be represented by a possibility distribution $\pi_{A(y)}$ about $X \cup \{e\}$. In other words, $\pi_{A(y)}$ is an application that goes from $X \cup \{e\}$ to the $[0, 1]$ interval. So, all value types adopted by this model can be represented.

If $\pi_{A(y)}(x) = 1 \forall x \in X$ then x value is completely possible for A(y).

1. If precise data is known and this is crisp(c) then

$$\pi_{A(y)}(e) = 0$$

$$\pi_{A(y)}(c) = 1$$

$$\pi_{A(y)}(x) = 0, \forall x \in X, x \neq c$$
2. If information is unknown but applicable then

$$\pi_{A(y)}(e) = 0$$

$$\pi_{A(y)}(x) = 1, \forall x \in X$$
3. If information is not applicable then

$$\pi_{A(y)}(e) = 1$$

$$\pi_{A(y)}(x) = 0, \forall x \in X$$
4. If there is lack of information (no idea that values are applicable or not applicable) then

$$\pi_{A(y)}(x) = 1, \forall x \in X \cup \{e\}$$
5. If given information is in range [a , b] then

$$\pi_{A(y)}(e) = 0$$

$$\pi_{A(y)}(x) = 1 \text{ if } x \in [a, b] \subseteq X$$

$$\pi_{A(y)}(x) = 0 \text{ in other case}$$
6. If the information available is a possibility distribution μ_a , then

$$\pi_{A(y)}(e) = 0$$

$$\pi_{A(y)}(x) = \mu_a(x) \forall x \in X$$

3.2.2. The Umano- Fukami Model

Through this model [9] , non-applicable information may be modeled by possibility distribution about a given domain .Here each domain value has a possible value equal to 0 .Suppose universe of discourse of A(y) is X and possibility that A(y) takes value $x \in X$ is $\pi_{A(y)}(x)$.

1. If information available is unknown and applicable then

$$\text{Unknown} = \pi_{A(y)}(x) = 1 \forall x \in X$$
2. If information is non-applicable then

$$\text{Undefined} = \pi_{A(y)}(x) = 0 \forall x \in X$$
3. If there is lack of information (no idea that values are applicable or not applicable) then

$$\text{Null} = \{1 / \text{Unknown}, 1 / \text{Undefined}\}$$
4. If given information is in range [a , b] then

$$\pi_{A(y)}(x) = 1 \text{ if } x \in [a, b] \subseteq X$$

$$\pi_{A(y)}(x) = 0 \text{ in other case}$$
5. If the information available is a possibility distribution μ_a , then

$$\pi_{A(y)}(x) = \mu_a(x) \forall x \in X$$

A fuzzy relation \mathfrak{R} , with n attributes, is defined as the following membership function:

$$\mu_{\mathfrak{R}}: \pi(X_1) \times \pi(X_2) \times \dots \times \pi(X_n) \rightarrow \pi([0, 1])$$

where the \times symbol denotes the Cartesian product, $\pi(X_i)$ with $i = 1, 2, \dots, n$ is the set of all the possibility distributions in the discourse universe X_i of the i-th \mathfrak{R} attribute.

During query session, the model divides the set of instances involved in the relation into three subsets:

1. Instances completely satisfying the query are in first subset.
2. Instances that may satisfy the query are in second subset.
3. Instances that do not satisfy the query are in third subset.

3.2.3. The Zemankova-Kaendel Model

Zemankova-Kandel Model dates back to 1984 and 1985. It consists of three databases [10]: value database, explanatory database and set of translating rules. In value database, data are organized in a similar way as in the possibilistic models. In explanatory database, the fuzzy subsets and fuzzy relations are stored. In a set of translating rules, various translating rules for the handling of adjectives and modifiers are used. In this model, query is posed in same manner as in Prade- Testemale model except the possibility measure is

$$P_A(S) = \sup_{x \in X} \{\mu_F(x) \cdot \pi_A(x)\}$$

Where $P_A(S)$ is the possibility measure used to find the compatibility of the fuzzy subset S of the condition, with an attribute A value for each tuple in the relation.

The certainty measure is used in this model given by this equation:

$$C_A(S) = \max_{x \in X} \{0, \inf \{\mu_F(x) \cdot \pi_A(x)\}\}$$

Query result is represented as a relation having two fields in which possibility value and certainty values, both are considered for each tuple. Minimum threshold values can also be considered.

In this model, possibility and certainty measures are being used. Interpretation of certainty measure is unclear. It is because there does not occur any relationship between possibility and certainty measure as occur in possibility (P) and necessity measures (N) as

$$N(S) = 1 - P(\neg S)$$

So, this model has various limitations which make it incomplete.

3.2.4. The GEFRED Model

The GEFRED (Generalized Fuzzy Relational Database) model was published in 1994 by Medina-Pons-Vila[11]. It is developed in possibilistic framework, so fuzzy domains are considered. It also includes the case where the underlying domain is not fuzzy i.e. numeric. Various datatypes given in GEFRED are:

1. A single scalar (e.g., Age = Young, represented by the possibility distribution 1/Young).
2. A single number (e.g., Height = 160, represented by the possibility distribution 1/160).
3. A set of mutually exclusive possible scalars (e.g., Age = {Young, Old}, represented by {1/Young, 1/Old}).
4. A set of mutually exclusive possible numbers (e.g., Age = {14, 50}, represented by {1/14, 1/50}).
5. A possibility distribution in a scalar domain (e.g., Age = {0.6/Young, 1.0/Middle}).
6. A possibility distribution in a numeric domain (e.g., Age = {0.5/23, 1.0/26, 0.8/24}, fuzzy numbers or linguistic labels).

It includes Umano-Fukami models's datatypes UNKNOWN, UNDEFINED and NULL also.

7. An Unknown value with possibility distribution:
Unknown = {1/x : x ∈ D}
8. An Undefined value with possibility distribution:
Undefined = {0/x : x ∈ D}
9. A NULL value given by:
NULL = {1/Unknown, 1/Undefined}

As GEFRED suggests, Generalized fuzzy domain (D) and generalized fuzzy relations (R) concepts are being included. Generalized fuzzy domain (D) includes classic domain, with the possibility distributions defined for this domain and the NULL value also. If X is the universe of discourse and P(X) is set of all possibility distributions including unknown, undefined types and Null type. Then, generalized fuzzy domain (D) can be represented as $D \subseteq P(X) \cup \text{NULL}$.

Generalized fuzzy relations are relations whose attributes are having generalized fuzzy domain. Fuzzy attributes may have associated with compatibility attribute where compatible degree can be stored. Generalized fuzzy relations are given by two sets: Head H and Body B. The head includes the name of each one of the n attributes, their domains, and their compatibility attributes (which are optional). The body includes the values of the m tuples:

$$\mathfrak{R} = \begin{cases} H = \{(A_1:D_1[C_1], \dots, A_n:D_n[C_n])\} \\ B = \{(A_1:d_{i1}[c_{i1}], \dots, A_n:d_{in}[c_{in}])\} \quad i=1, \dots, m \end{cases}$$

GEFRED model also defines fuzzy comparators which are based on the classic comparators (<, >, = etc). When these fuzzy comparators are used on non-fuzzy values, meanings of comparators does not change i.e. classical output should come during crisp values.

GEFRED redefines the relational algebraic operators: union, intersection, difference, Cartesian product, projection, selection, join [11] and division [12] in generalized fuzzy relational algebra.

GEFRED model is based on previously discussed various models. So, this model can be used to represent them.

IV. Conclusion

This paper outlined the brief introduction of fuzzy relational models. This paper also discusses imperfect information. Fuzzy set theory basics are also being reviewed in brief. Fuzzy information incorporated in database models has been an important topic of database research because such information extensively exists in data and knowledge intensive applications. Research has been conducted into various approaches to represent and handle fuzzy data in the context of databases. The previous work which has been performed in field of fuzzy relational database models is discussed in brief.

References

- [1] Z.M. Ma, A conceptual design methodology for fuzzy relational databases .Journal of Database Management , Vol. 16, No. 2 , 2005 ,66-83.
- [2] Zadeh, L. A., Fuzzy sets. *Information and Control*, 8, 1965, 338-353.
- [3] Zadeh, L. A., A computational approach to fuzzy quantifiers in natural languages. *Computer Mathematics with Applications*, 9,1983, 149-183.
- [4] Galindo, J., Urrutia, A., Piattini, M.: *Fuzzy Databases: Modeling Design and Implementation*. IDEA Group Publishing, Hershey, USA. (2006)
- [5] Zadeh, L. A. , Similarity relations and fuzzy orderings. *Information Sciences*, 3,1971, 177-200
- [6] Buckles, B. P., & Petry, F. E. , Uncertainty models in information and database systems. *Information Sciences*, 11,1985, 77-87.
- [7] Prade, H., & Testemale, Fuzzy relational databases: Representational issues and reduction using similarity measures. *J. Am. Soc. Information Sciences*, 38(2),1987, 118-126.
- [8] Codd, E. F. , The relational model for database management, Version 2,1990, Reading, MA: Addison-Wesley.
- [9] Umamo, M., & Fukami, S. , Fuzzy relational algebra for possibilitydistribution- fuzzy-relation model of fuzzy data. *Journal of Intelligent Information Systems*,1994, 3, 7-28.
- [10] Zemankova-Leech, M., & Kandel, A., Implementing imprecision in information systems. *Information Sciences*, 37, 1985,107-141.
- [11] Medina, J. M., Pons, O., & Vila, M. A., GEFRED: A generalized model of fuzzy relational databases. *Information Sciences*, 76(1/2), 1994,87-109.
- [12] Galindo, J., Medina, J. M., Cubero, J. C., & Garcia, M. T. Relaxing the universal quantifier of the division in fuzzy relational databases. *International Journal of Intelligent Systems*, 16(6),2001, 713-742.