IJournals: International Journal of Software & Hardware Research in Engineering
ISSN-2347-4890
Volume 3 Issue 9 September, 2015

On Fuzzy Logic to handle Vague and Imprecise Data

By Dr. Rajeev GajananSapre,
Head, Department of Mathematics, GogateJogalekar College, Ratnagiri.
rgsapre@rediffmail.com

Abstract

The reasoning and decision-making process of human beings is often uncertain and vague in nature. Traditional databases with the help of SQL , Oracle manage data, which is non fuzzy and concrete, deterministic and precise in nature. The data retrieval process is done with the help of machine having no thinking capacity. If we want the computer system to work according to semantic intent of human being , then we need to introduce Fuzziness into our databases based upon the concept of fuzzy logic, which is an extension of normal Boolean logic ,representing something between absolute truth and absolute false.

The concept of fuzzy database and the ways of handling the fuzzy queries to databases/ fuzzy databases are explained in this paper followed by a comparison between the two different set theories on which the traditional and the fuzzy databases are based.

Key words : Fuzzy database, Fuzzy logic, Fuzzy sets.
Subject Classification : 03B52

1. Introduction

Most of our traditional tools for reasoning and computing are concrete or non fuzzy, deterministic and precise in nature. But it gives birth to two major complications. In real life we come across many situations where we have to deal with Vague data , imprecise data, The query languages like SQL etc allows us to fire commands in its DML which are rigid and these commands can not reflect semantic intent of our query. In real life we describe height of a person by saying he is very short, very tall, tall , not much tall etc rather by telling height of a person in number. Secondly, complete description of a real system would require far more detailed data than a human being could ever recognize and process simultaneously. In medical diagnosis when doctor makes diagnosis he has to rely on the data which is too imprecise. Since most conventional databases in use today are based on the relational model, so such kind of description becomes difficult to represent under the conventional relational model. If we want the computer system to work according to human speech, then we need to introduce Fuzziness into our databases .

2.Traditional Database Approach: An Example

Consider a student record database system. Supposing we want to find all the **bright** (percentage above 90) and **young**students (age between 15 and 20) in the whole batch. For a crisp system we would specify the query as

*select * from student where percentage>90 and age between 15 and 20*

Now this system has a major flaw. Consider a student, A whose percentage is 92% but whose age is 20 years and 15 days. He should have been selected as our requirement is to select bright students but because of the rigid boundary conditions set by the normal crisp logic, he has been rejected. The query contains two conditions. First is marks > 90 it is true in this case but age has crossed 20. So rigidity of the query will not treat A as bright and young students. Here we know that A is bright and young but this semantic intent is not reflected in the retrieval.

2. The Concept of Fuzziness

A **fuzzy concept** is a concept of which the boundaries of application can vary considerably according to context or conditions, instead of being fixed once and for all. This means the concept is *vague* in some way, lacking a fixed, precise meaning, without however being unclear or meaningless altogether. It has a definite meaning, which can become more precise only through further elaboration and specification, including a closer definition of the context in which the concept is used.

A fuzzy concept is understood by scientists as a concept which is "to an extent applicable" in a

situation, and it therefore implies gradations of meaning.

The concept of fuzziness as described by Zadeh [1] includes imprecision, uncertainty, and degrees of truthfulness of values. Zadeh introduced a theory whose objects fuzzy sets are sets with boundaries that are not precise and the membership in this fuzzy set is not a matter of true or false, but rather a matter of degree. This concept was called Fuzziness and the theory was called Fuzzy Set Theory. It is particularly frequent in all areas in which human judgement , evaluation and decisions are important.

3. Definition of a Fuzzy Set

fuzzy set *: Let X be a nonempty set. A fuzzy set A in X is characterized by its membership*

functionμA: $X \rightarrow [0, 1]$ *and* $\mu A(x)$ *is interpreted as the degree of membership of element x in fuzzy set A for each* $x \in X$. It is clear that *A* is completely determined by the set of tuples $A = \{(u, \mu A(u)) | u \in X\}$.

Frequently we will write $A(x)$ instead of $\mu A(x)$. If $X = \{x1, \ldots, xn\}$ is a finite set and *A* is a fuzzy set in *X* then we often use the notation A = μ1/x1 + ------ + μn/xnwhere the term μi/xi, i= 1, . . . , n signifies that μiis the grade of membership of xi in A and the plus sign represents the union.

Assume someone wants to buy a cheap car. Cheap can be represented as a fuzzy set on a universe of prices, and depends on his purse. For instance, 'cheap' is roughly interpreted as follows:

• Below Rs 2,00,000 cars are considered as cheap, and prices make no real difference to buyer's eyes.

• Between Rs 2,00,000and 4,50,000, a variation in the price induces a weak preference in favor of the cheapest car.

• Between 4,50,000and 6,00,000, a small variation in the price induces a clear preference in favor of the cheapest car.

• Beyond 6,00,000the costs are too high (out of consideration).

This consideration is useful to decide membership function of a concept 'cheap'. This function may not be unique.

So Fuzzy set is a function from universal set X to [0,1].

4. Comparison of Fuzzy and Boolean Logic

- Fuzzy logic is a concept of `certain degree'.
  Fuzzy logic differs to Boolean logic in a sense that something can be true to a certain extent and does not have to be just true or false.
  Japan is currently the most active users of fuzzy logic.
  When it was founded in the 60s the Americans and the rest of the world totally ignored the idea. Instead, it was adopted by the Japanese followed by Korea and other parts of the East. Currently 70% of Japanese products use fuzzy logic.
  Boolean logic is a subset of fuzzy logic. Since Boolean logic only holds for values of 1 or 0 and fuzzy logic holds for a range of values *from* 1 *to* 0.
- Boolean algebra brings logic operators AND, OR, NOT. But by using these operators when searching with imprecise or incomplete information, it is not ensured that we get the requested data.
- But imagine if there is no information meeting the query conditions, still the user gets information which meet conditions to some extent/degree.
- And this isn't possible using classic approach. Only fuzzy logic can handle such queries
- Hence Fuzzy logic is an extension of normal boolean logic representing something between absolute truth and absolute false.
- Advantage of fuzzy logic is the mathematical ability to catch up the words described by words. This gives us the possibility to work with ambiguous terms like "small", "near", "far" ,"about", "very" and with number of other words used in human language. If 20 m is close, then what about 21 m. Will it be categorized as far?
- Next difference is direct in query where fuzzy statement can be used i.e. statement which allows us to use the words like big, far, near, average etc.
- Hence the basic difference between fuzzy approach and boolean one is the ability to get "rated result".

5. Classification of Fuzzy Data

1 Classification of Fuzzy Data

Many real world systems and applications require information management components that provide support for managing imprecise data. If

required information is properly generated from the analysis of the data then it helps in making decisions. Fuzzy theory allows us to model imprecise or vague data. The use of fuzzy theory also allows us to manage vague knowledge. There have thus been several proposals for extending relational database systems in order to represent as well as query such imprecise data. Little work, however, has been done in modeling uncertainty at the conceptual schema level and in standardizing fuzzy data in fuzzy relational databases (FRDBs).

To fill this gap, a classification of fuzzy data is developed. This methodology contains extensions for representing the imprecision of data in fuzzy relational database.

We limit the scope of the paper with the following assumption that the types of data considered in this paper are only numerical and string data.

For many practical systems, important information comes from two sources: one source is human experts who describe their knowledge about the system in natural languages (Knowledge base) the other is sensory measurements and mathematical models that are derived according to physical laws. Doctors, lawyers, engineers can diagnose problems a lot quicker if the expert system they use to diagnose the problem lists a few fuzzy solutions that they can use to augment their own findings.

To store expert opinions, fuzzy database is necessary which stores fuzzy data (linguistic terms). Fuzzy data means imprecise, vague data or non-standard representation of the same data. Imprecise, vague, uncertain, ambiguous, inconsistent, incomplete and null data are fuzzy data. Here we discuss each type of fuzzy data in detail.

2.1 Negative Data

Negative data means denying or contradicting data.Negative data are data that do not enable us to reject our null hypothesis. Negative data implies range of possible values as it denies one value but the possible value can be other than that. A subset W of a set U can be represented by a function f_w: U $\rightarrow$ {0,1} than negative information can be presented as $f_{w'}(x) = 1 - f_w(x)$. As shown in Table 2, the fuzzy set will have membership value of 0 for negative data but membership value will vary for the range of all other possible values.

Table 1. Route Table

Destination	Bus number
Kalina	313 or 312 or 357
Fort	165 or 161 or 31

Table 2. Population Table

Surname	Number of Children	Number of girls
Desai	3	1
Gavade	2	Nil
Soni	2	1

2.2 Range Data

Range data means data that vary within or over specified limits. If students of a particular college come from various towns say T1, T2,……,Tn, then when a particular student from that college is selected then that student will be from one of those cities. This is expressed as student belongs to { T1, T2, ….,Tn}. For example in Table 3, Fuzzy set of data will be developed based on membership function for given range. Membership value will be assigned to each data value.

2.3 Data with Error Margin

Data with error margin means a limit in a condition, beyond or below which data is no longer possible or acceptable. If the theory DB has several distinct modelsM1, M2, ….,Mn, and we know that W is equal to one of the value from M1 to Mn and its variation Є. This can be expressed by W ∈□ Mi ± Є□ wherei = 1,…..,n. For example, Table 4 gives information about which instrument allows how much error margin. In lot of scientific experiments error margin is assumed for the readings. In Fuzzy database, upper limit and lower limit of the fuzzy set is set by given margin of data. Membership function will be written on the basis of given margin and membership value will be assigned to each data value. Fig. 4 represents Data with Error Margin.

Table 3 Student Table

Name of Student	Age
Pranesh	20-25
Ramesh	21-28
Rakesh	20-26
Haresh	15-20

Table 4. Instrument Table

Instrument Name	Level of Sugar	Allowed error margin
Sugar measurer	150	+/- 10

2.4 Null Data

The two extreme kinds of imprecision are precise values and null values: a value is precise

when the set of possibilities is a singleton; a *null value* usually denotes that no information is available, yet could be regarded as imprecise information where the set of possible values encompasses the entire domain of legal values. A basic problem with null values is that they have many plausible interpretations. Most authors agree that the various expressions of nulls can be reduced to two basic interpretations. These are: *The unknown interpretation*: a value exists but it is not known. Additional information may be available on a specific instance of a null .*The nonexistent interpretation*: a value does not exist. A value is undefined. This meaning of null, however, is not related to uncertainty or fuzziness.

2.5 Vague Data

Vague means not clearly expressed or not clear in meaning or application. Vague data contains some vague predicate such as "tall" or "cloudy day" . When modelling the concept "tall" as a fuzzy subset of [0,∞) with a membership function A: [0,∞) →□ [0,1], is a description of the meaning "tall" in a mathematical way. For example, "Ramesh is tall". The statement could be used to establish the range of possible values for Tom's height, based on fuzzy set interpretation of term "tall". Different Individual can interpret the word "tall" differently and such concepts are subjective.

2.6 Uncertain Data

Uncertainty arises from the fact that an agent has constructed a subjective opinion about the truth of a fact, which it does not know for certain. This lack of information makes it impossible to determine if certain statements about the world are true or false, all that can be done is to estimate the tendency of the statement to be true or false by using some numerical measure of the degree to which one may be sure

2.7 Uncertain Data Due to Statistical Analysis

Some data is recorded statistically and so is inherently uncertain. This type of data generally arises when readings are taken in case of scientific experiments.

2.8 Uncertain Data Due to Security Reasons

Other data is deliberately made uncertain for security reasons. Other data may not be measured accurately, due to some or other reasons, and will include some unavoidable uncertainty. In such situations, the best that we can do is to try to estimate the tendency of the statement to be true (or to be false). This can be done with the help of fuzzy set & by providing degree of membership to the statement to be true or false. Such data occurs in social surveys.

2.9 Ambiguous Data

Ambiguous means doubtful, uncertain, or capable of being understood in either of two or more possible senses.

Table 5. Building Table

Type of Room	Length	Breadth
DR	30	20
DR	20	20
DR	40	30

2.10 Ambiguous Data Due to Use of Abbreviation

For example, Building Table. Here confusion is that DR should be interpreted as Dinning Room or DR should be interpreted as Drawing Room, Discussion Room.

2.11 Ambiguous Data Due to Incomplete Context

For example, suppose length and breadth of a rectangle is given to be 30 and 20 then no one will understand whether it is in meters or centimeters. We come across such type of data many times where context is incomplete.

2.12 Ambiguous Data Due to Different Orderings and different abbreviations

:Ambiguous data occur when different values for the same item are stored by the system. For example, Mr.GogateVaibhav at one place and Mr.VaibhavGogate at other place. This creates confusion. Also people use different abbreviations for the same thing. For Computer Science they use CS at one place and C.S. at other place. Human mind can easily understand that these two things are the same but machine can not understand.

2.13 Inconsistent Data

Inconsistent data means, data that doesn't agree with itself or which is not reliable or data that does not have one common solution, as of two or more equations. For example, in the context of distributed databases, if each database is considered an agent, it may happen that A is derivable from the database DB1 and ¬A is derivable from the database DB2. In this case uncertainty is about the

database that contains the correct information. The mutual inconsistency of DB1 and DB2, that is, their lack of common model, can be expressed by W ∈ M (DB1) or W ∈ M (DB2).

2.14 Inconsistent Data Due to Unreliable Sources

These types of data exist because the sources of data are unreliable.

Table 6 .Qualification Table

Name of Student	Course Name	Qualification
BhatyeAbhijit	Web designing	H.Sc
BhatyeAbhijit	MCA	B.Sc.

For example, in the above table, there is an inconsistency in values. People may show different degrees of education while applying for different courses. Here Mr.AbhijitBhatye has shown different educational qualifications for applying for two different computer courses. Fuzzy function can be written and membership value will be assigned to each statement to assign the degree of truth.

2.15 Inconsistent Data across Multiple Files/Tables in Database

This type of data arises because integrity constraints that encompass all semantically related tables are not specified and enforced.

2.16 Inconsistency Due to Database Overlap

In any environment of multiple databases it is practically unavoidable that the databases would overlap. In multiple databases if one database have one or more elements in common with another database than we can say that databases are overlapped.

2.17 Incomplete Data

Incompleteness usually denotes absence of a value, yet could be regarded as imprecise information where the set of possible values encompasses the entire domain of legal values. Incomplete processing of data occurs when the observation can't perform its function on time. Incomplete Data occurs due to *dirty read, lost update*, *Unrepeatable Read* and *Missing Values.*

Fuzzy data can be further divided into two types:
a) **Approximate Value**: The information data is not totally vague and there is some approximate value, which is known and the data lies near that value
b) **Linguistic Variable**: This variable is used to represent a fuzzy number.

A **linguistic term** is a name given to the fuzzy set.[3][4].
Consider a linguistic variable AGE which may have a Linguistic term YOUNG associated with it. Then the possibility distribution of this linguistic term YOUNG for the linguistic variable AGE will be as follows:

Now recall the same example of A whose percentage is 92% but whose age is 20 years and 15 days.
• In fuzzy logic, we would do the same by specifying two fuzzy sets YOUNG and BRIGHT and each student will have some membership grade associated with the two sets.
• So according to our definition, the student A will have a non-zero membership grade although it will be less than other students in the age group of 15 to 20
• Hence, using fuzzy logic even A will be included in the result set to be considered as "A also satisfies the query to some extent, which is represented by its membership grade."

6. Fuzzy Querying to Relational Databases

To incorporate fuzziness we introduce fuzzy sets/ linguistic terms on the attributes/linguistic variables[5] e.g. on the attribute HEIGHT, we may define sets as TALL,MEDIUM and SHORT. Similarly on the attribute AGE we may define sets as YOUNG, MIDDLE and OLD.
If it is a medical field then there can be different parameters related to symptoms of particular disease and in this case also we can assign linguistic terms like EXTREME , MODERATE, NILL etc.

When we plot the graphs of these fuzzy sets there arise triangular numbers, Trapezoidal numbers , bell shaped numbers and these functions allow us to decide membership grades to the linguistic terms.

The SQL query that can be fired can be of the type :

Select <attributes> from <relation> where <fuzzy condition>

e.g. *Select * from students where (age=middle) and (percentage =high)*

Today most query languages represent SQL which is based on Boolean Logic. FuzzySQL(SQLf) is based upon SQL with additional properties of impreciseness and fuzziness. The objective is to

introduce some fuzziness in the base block of SQL as shown above[6]

7. Memory and Time Consumption of Fuzzy Databases

A database that can handle imprecise information shall store not only the raw data but also the related information that shall allow us to interpret the data in a much deeper context. Hence fuzzy databases require more storage spaces as compared to traditional databases than handle crisp information. It takes more I/O time to transfer ill-known data between main memory and secondary memory than crisp data. Furthermore, since fuzzy queries require nonboolean degrees of satisfaction to be computed, it takes more CPU time to evaluate a fuzzy query condition than does a crisp query condition.

But the main advantage is using fuzzy logic and fuzzy mathematics ideas in Retrieving meaningful information from database is that it gives us information which reflects our semantic intent.

8. Conclusion

Today exists huge applications of fuzzy sets and their properties in areas such as Medical Diagnosis, Employment etc. One of those are databases. Fuzzy sets represent basement of fuzzy database systems , which brings opportunities to query data by language close to human speech. But fuzzy databases are still not very much in use because people are reluctant to replace their crisp data by fuzzy data before they are convinced that it is worthwhile and necessary to do so

REFERENCES

[1] Zadeh, L. A. Fuzzy sets. Information and Control, 8 (1965), 338-353.
[2] Klir G. J. and Yuan B. [2001], "Fuzzy Sets and Fuzzy Logic : Theory and Applications", Prentice Hall, Inc. Englewood Cliffs, N. J., U.S.A.
[3] W. Zhang, C. Yu, G. Wang, T. Pham, and H. Nakajima, ªA Relational Model for Imprecise Queries,º Proc. Int'l Symp.Methodologies in Intelligent Systems, 1993.
[4] H. Nakajima, T. Sogoh, and M. Arao, ªDevelopment of an Efficient Fuzzy SQL for Large Scale Fuzzy Relational Database,ºProc. Fifth Int'l Fuzzy Systems Assoc. World Congress '93, 1993.
[5] H. Prade and C. Testemale, ªFuzzy Relational Databases: Representational Issues and Reduction Using Similarity Measures,º J. Am. Soc. Information Science, vol. 38, no. 20, pp. 118-126,1988.
[6] P. Bosc, M. Galibourg, and G. Hamon, ªFuzzy Querying with SQL: Extensions and Implementation Aspects,º Fuzzy Sets and Systems, 1988.
[7] Navnit Kaur Johal, Amrit Kaur "Fuzzy Databases: Approach to handle vague information" ISSN: 0975-5462 361

