Current Development in Artificial Intelligence.
ISSN 0976-5832 Volume 5, Number 1 (2014), pp. 17-24

http://www.irphouse.com

Database Management System and Fuzzy Data

Prof. Rajeev G. Sapre and Ms. Sayali R. Sapre

Department of Mathematics , GogateJogalekar College, Ratnagiri.

Finolex Academy of Management and Technology, Ratnagiri.

Abstract:

When the available information is imperfect, it is often desirable to represent it in the database, so that it can be used to answer queries of interest as much as possible. We come across various types of data which is many times vague or imprecise. We have to extract some meaningful information from this data. The data as well as query in data sources are often vague or imprecise (fuzzy). Generally people use software likeSQL,Oracle, My-Sql to extract the information but all the queries that are fired are rigid and most of the times semantic intent is not reflected in the outcome of these queries. Statistical data are not always precise numbers, or vectors, or categories. Real data are frequently what is called fuzzy. Examples where this fuzziness is obvious, are environmental, ecological, biological, medical, Experimental, sociological , commercial and economics data. Also the results of measurements can be best described by using fuzzy numbers and fuzzy vectors respectively.In this paper, a comprehensive classification of fuzzy data is done. This classification will be used as framework for understanding how fuzzy data arise.

1 Introduction

Computers do not reason similar to human minds. They simply manipulate solid facts into ones and zeros whereas the human mind can process incomplete and vague data with uncertainties and unknowns and still function perfectly. Thus came the concept of Fuzzy Logic in order to create machines that can imitate the complex functions of the mind.

Fuzzy logic is a superset of conventional (Boolean) logic that has beenextended to handle the concept of partial truth -- truth values between"completely true" and "completely false". It was introduced by Dr.

LotfiZadeh of UC/Berkeley in the 1960's as a means to model the uncertaintyof natural language.

1.1 Fuzzy Database Evaluation

Fuzzy logic provides the solution by providing a mathematical approach to programming complex systems. It implements operator knowledge directly into the system. The thought process of how humans store and apply knowledge about controlling a system can take the form of

"if<situation> then <reaction>" rules.

For example, aprocess may require a rule such as

"if the temperature is decreasing fast than increase the power of heater".

"If the person is tall then he will be included in the game "

Since the rules are quite abstract this would be difficult to program into a traditional system.

One cannot define ideas like "intelligent" , "tall" etc precisely.

Fuzzy Logic is a method for grouping items into sets, which does not require that an item be either in or out of a set completely.

Fuzzy sets can be characterized by a*fuzzy membership function*, a function which takes a number of attributes of the itemsto be classified as input, and provides a fuzzy membership value between 0 and 1 asoutput.

The need for fuzziness in data modelingcan be viewed as providing the capabilityto directly represent imprecision. Clearly now that AI applications are becoming morecommon, the utilization of database technology becomes more critical if these applicationsare to reach their full potential and usefulness.

Below we give some of thereasons of using fuzzy logic in Database:

- Fuzzy logic provides a user-friendly data presentation for the report generationphase with linguistic variables and fuzzy values. It helps us to extract semantic intent of the query.
- Fuzzy Sets provides additional data security features due to the introduction of anadditional view-based data layer, which hides the numerical values from the users.
- The information is available for querying or evaluating the very largedatabases by using linguistic variables.
- Fuzzy logic provides optimizeddatabase performance by modeling imprecise, uncertainand vague data.

2 Classification of Fuzzy Data

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

information is properly generated from the analysis of he data then it helps in making decisions. Fuzzy theory allows us tomodel imprecise or vague data. The use of fuzzy theory also allows us to manage vagueknowledge. There have thus been several proposals for extending relational databasesystems 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 instandardizing fuzzy data in fuzzy relational databases (FRDBs).

To fill this gap, aclassification of fuzzy data is developed. This methodology contains extensions forrepresenting the imprecision of data in fuzzy relational database.

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

For many practical systems, important information comes from two sources: onesource is human experts who describe their knowledge about the system in naturallanguages (Knowledge base) the other is sensory measurements and mathematical models that are derivedaccording to physical laws. Doctors, lawyers, engineers can diagnose problems alot quicker if the expert system they use to diagnose the problem lists a few fuzzysolutions 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 thesame data. Imprecise, vague, uncertain, ambiguous, inconsistent, incomplete and nulldata 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 usto reject our null hypothesis. Negative data implies range ofpossible values as it denies one value but the possible value can be other than that. Asubset W of a set U can be represented by a function f_w: U $\rightarrow$ {0,1} than negativeinformation can be presented as $f_{w'}(x) = 1 - f_w(x)$. As shown in Table 2, the fuzzyset will have membership value of 0 for negative data but membership value will varyfor the range of all other possible values. Fig. 2 represents Negative Data.

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 examplein Table 3, Fuzzy set of data will be developed based on membership function forgiven range. Membership value will be assigned to each data value. Fig. 3 representsRange Data.

2.3 Data with Error Margin

Data with error margin means a limit in a condition, beyond or below which datais 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 andits variation Ɛ. This can be expressed by W ϵMi ± Ɛwherei = 1,…..,n.For example, Table 4 gives information about which instrument allows how mucherror 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 bygiven margin of data. Membership function will be written on the basis of given marginand membership value will be assigned to each data value. Fig. 4 represents Datawith 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 isprecise when the set of possibilities is a singleton; a *null value* usually denotes that noinformation is available, yet could be regarded as imprecise information where the setof possible values encompasses the entire domain of legal values. A basic problemwith null values is that they have many plausible interpretations. Most authors agreethat the various expressions of nulls can be reduced to two basic interpretations. Theseare:*The unknown interpretation*: a value exists but it is not known. Additional informationmay 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 datacontains some vague predicate such as “tall” or “cloudy day” . When modelling the concept “tall” as afuzzy subset of [0,∞) with a membership function A: [0,∞) $\rightarrow$[0,1], is a description ofthe meaning “tall” in a mathematical way.For example, “Ramesh is tall”. The statement could be used to establish the range ofpossible 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 opinionabout the truth of a fact, which it does not know for certain. This lack of informationmakes it impossible to determine if certain statements about the world are true orfalse, all that can be done is to estimate the tendency of the statement to be true orfalse 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 maynot be measured accurately, due to some or other reasons, and will includesome unavoidable uncertainty. In such situations, the best that we can do is to try toestimate the tendency of the statement to be true (or to be false). This can be donewith the help of fuzzy set & by providing degree of membership to the statement to betrue or false.Such data occurs in social surveys.

2.9 Ambiguous Data

Ambiguous means doubtful, uncertain, or capable of being understood in

either of twoor 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 DinningRoom 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 ordata 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, itmay happen that A is derivable from the database DB1 and ¬A is derivable from thedatabase DB2. In this case uncertainty is about the database that contains the correctinformation. The mutual inconsistency of DB1 and DB2, that is, their lack of commonmodel, 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 differentdegrees of education while applying for different courses. Here Mr. AbhijitBhatye has shown differenteducational qualifications for applying for two different computer courses. Fuzzy function can bewritten and membership value will be assigned to each statement to assign the degreeof truth.

2.15 Inconsistent Data Across Multiple Files/Tables in Database

This type of data arises because integrity constraints that encompass all semanticallyrelated 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 databaseswould overlap. In multiple databases if one database have one or more elementsin 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 impreciseinformation where the set of possible values encompasses the entire domain of legalvalues. Incomplete processing of data occurs when the observation can't perform itsfunction on time. Incomplete Data occurs due to *dirty read, lost update,UnrepeatableRead* and *Missing Values* .

3 Conclusion

In this paper we have shown importance of fuzzy logic in database system. We presentgeneral-purpose classification of fuzzy data. Fuzzy data representation reflects how,where and to what extent fuzziness is incorporated into classical models. The classificationprovides framework for understanding the origins of fuzzy data and the impact offuzzy data in database management system. If you know the possibility that certainfuzzy data exist, you will be more prone to spot them and to plan your project to storeand manipulate fuzzy data in a best possible way. We expect that such classification willprovide a valuable guideline for further research in this direction.

References

1. Altrock, C.: Fuzzy Logic and NeuroFuzzy Applications Explained. Prentice Hall PTR Englewood Cliffs, New Jersey (1995)
2. Buckles, B.P.: An information retrieval perspective on fuzzy database systems. Proceedings of the ACM '82 conference, ISBN:0-89791-085-0, (1982) Pages: 186 – 187
3. DevalPopat, HemaSharda, and David Taniar :Classification of Fuzzy Data in Database ManagementSystem
4. Elmasri, R., Navathe, S.: Fundamentals of Database Systems, AddisionWisley, Third Edition, ISBN 9814053309 (2000)
5. Kwan, S., Olken, F., and Rotem, D.: Uncertain, Incomplete and Inconsistent Data in Scinetific and Statistical Databases. Uncertainty Management in Information Systems (1996) 127-154.