Introducing Artificial Intelligence Query With fuzzy query over fuzzy database

Protap Mollick[a], Ashiqur Rahman[b], Md. Shrof Uddin Shibly[c], Shimul Kanti Bala[d]

[a] AIUB— American International University-Bangladesh,Banani, Dhaka, Bangladesh.
protap2010@gmail.com

[b] JU— Jahangirnagar University, Savar, Dhaka, Bangladesh
ashiqur128@yahoo.com

[c] AIUB—American International University-Bangladesh, Banani, Dhaka, Bangladesh.
shibly_shorof@yahoo.com

[d] AIUB—American International University-Bangladesh, Banani, Dhaka, Bangladesh.
shimulbala@gmail.com

Abstract. Information systems have revolutionized the way information can be stored and processed. As a result, the information volume has significantly increased leading to an information overload. It therefore becomes difficult to analyze the large amounts of available data and to generate appropriate management decisions. In practice, information systems mostly use relational databases in order to store these data collections. Another issue, using the relational model, is the restriction of having sharp, precise data and therefore a dichotomous querying process which is not well suited for decision making. This paper presents an artificial intelligent query system using fuzzy logic. The system is based on constructing a sample fuzzy database and is organized with fuzzy queries, employing vague or fuzzy terms in the database. The fuzzily defined data has been represented using *s*, *z*, and π shaped membership functions. This paper also makes a comparison between traditional database and fuzzy database by computing the time cost of classical query over classical database, fuzzy query over classical database and fuzzy query over fuzzy database. Experimental results demonstrate that the proposed intelligent fuzzy query is faster than the conventional query and it provides the user the flexibility to query the database using natural language.

Keywords: Artificial Intelligence, Emulation, Database, query, Fuzzy Set.

1 INTRODUCTION

The fuzzy set theory has been proposed in 1965 by Lofti A. Zadeh from the University of Berkeley. This theory is based on the intuitive reasoning by taking into account the human subjectivity and imprecision. It is not an imprecise theory but a rigorous mathematical theory which deals with subjectivity and/or uncertainty which are common in the natural language. The natural language is a very complicated structure which is fundamental; not only in the human communication, but also in the way human beings think and perceive the surrounding world. The main idea of the fuzzy logic is to capture the vagueness of the human thinking and to express it with appropriate mathematical tools. More precisely, "the fuzzy logic provides a mathematical power for the emulation of the higher order cognitive functions, the thought and perception" .Unlike computers, the human reasoning is not binary where everything is either yes (true) or no (false) but deals with imprecise concepts like 'a tall man', 'a moderate temperature' or 'a large profit'. These concepts are ambiguous in the sense that they cannot be sharply defined. For instance, the question whether a person is tall cannot be universally answered as some people will agree and others won't. Despite the fact that the definition of the word 'tall' is clear, it is not possible to sharply state if a person is tall because the answer may depend on the individual perception. Even for one person it may not be possible to give a

clear and precise answer as the belonging to a concept (e.g. tall person) is often not sharp but fuzzy, involving a partial matching expressed in the natural language by the expressions 'very', 'slightly', 'more or less', etc.

In order to illustrate this ambiguity, consider the concept 'middle-aged' for a person. This concept is clear in the people's mind however it is difficult to explicitly determine the precise beginning and ending years for a middle-aged person. Once again every individual might give a different definition of that concept. Let's assume that a survey states that a middle-aged person is between 35 and 55 years old. The concept 'middle-aged' can then be represented as a set illustrated in Figure 1. This set represents the truth function of the concept 'middle-aged' according to the survey where the X-axis represents the age and the Y-axis contains the truth value. A truth value of 1 means that the age corresponds to the concept and a value of 0 indicate the age does not belong to that concept. This definition of the concept 'middle-aged' implies that a person who is 34 years old would suddenly become middle-aged on his next birthday. Similarly, just after his 56th birthday this person would no longer be middle-aged. This definition of the concept 'middle-aged' seems therefore unnatural as it does not match the human perception due to the sharply fixed boundaries.

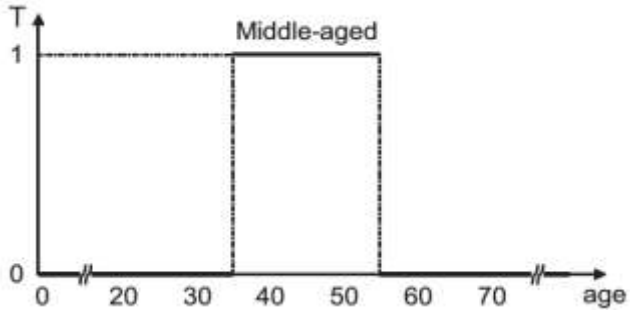

Figure 1: Sharply defined concept of a middle-aged person

A way of better modeling the imprecision of the human thinking is to introduce the notion of partial membership which allows a continuous transition between the different concepts. The notion of partial belonging can be represented by a fuzzy set (see Figure 2).. With this new definition a person enters the concept 'middle-aged' at the age of 20 with a continuous increment till the full belonging at the age of 45 and then progressively quits the concept. This way, there are no more extreme steps (sharp boundaries) such that just within a year somebody jumps into or out of the concept 'middle-aged'.

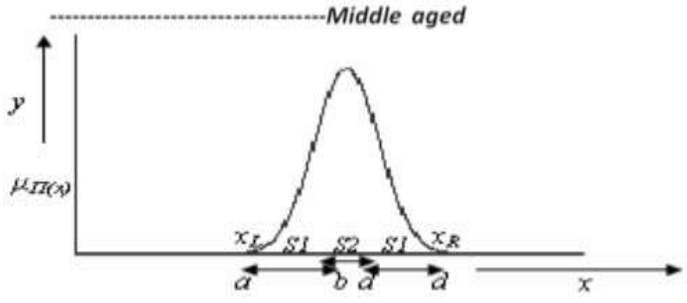

Figure 2: Concept of a middle-aged person defined with a fuzzy set

The ambiguity is part of the human thinking and is ubiquitous in the natural language. Different aspects of ambiguity can be distinguished.

2 DATABASE SYSTEM & APPLICATION

A database management system (DBMS) is a collection of interrelated data and a set of programs to access those data. The collection of data usually referred to as the database, contains information relevant to an enterprise. The primary goal of a DBMS is to provide a way to store and retrieve database information that is both convenient and efficient. Database systems are designed to manage large bodies of information. Management of data involves both defining strictures for storage of information and providing mechanisms for the manipulation of information. Computer scientists have developed a large body of concepts and techniques for managing data.

Database system are widely used. Here are some representative applications,

Banking: For customer information, accounts and loans and banking transactions.

Airlines: For reservations and schedule information. Airlines where among the first to use databases in a geographically distributed manner-terminals situated around the world accessed the central database system through phone lines and other data networks.

Universities: The student's information, course registration's and grades.

Credit card transactions: For purchases on credit cards and generation of monthly statements.

Telecommunication: For keeping records of calls made, generation monthly nulls, maintaining balance's on prepaid calling cards and storing information about the communication networks.

Finance: For storing information about holdings, sales and purchases of financial instruments such as stocks and bonds.

Sales: For customer, product and purchases information.

Manufacturing: For management of supply chain and for tracking production of items in factories, inventories of items in warehouses/stores and orders for items.

Human Resources: For information about employees, salaries, payroll taxes, benefits and for generation of paychecks.

As the list illustrates, databases for an essential part of almost all enterprises tide. The last four decades of the twentieth century, use of databases grew in all enterprises. In the early days, very few people interacted directly with database systems, although without realizing it they interacted with database indirectly though printed reports such as credit card statements or through agents such as bank tellers and airline reservation agents. Then automated teller machines came along and let users interact directly with databases. Phone interfaces to computer also allowed users to deal directly with databases a caller could dial a number and press phone keys to enter information or to select alternative options, to find flight arrival times, for example or to register for courses in a university. Although user interfaces hide details of access to a databases and most people are not even aware they are dealing with a database, accessing databases forms an essential part of almost everyone's life today.

The important database management system can be judged in another way today, database system vendors like Oracle are among the largest software companies in the world and database system form an important part of the product line of more diversified companies like Microsoft and IBM.

3 DATABASE & QUERY

An important point in the evolution of modern concept of uncertainty was the publication of a seminal paper by Lofti A Zadeh , in which 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. To create a toolkit for the analysis of customer relationships, a combination of relational databases and fuzzy logic is proposed. Fuzzy logic, unlike statistical data mining techniques such as cluster or regression analysis, enables the use of non-numerical values and introduces the notion of linguistic variables. Using linguistic terms and variables will result in a more human oriented querying process.. Fuzzy database consisted of incomplete, imprecise and vague data and fuzzy query use linguistic terms. There are two feasible ways to incorporate fuzziness in DBMS. One is making fuzzy queries to the classical databases and another is fuzzy queries to the fuzzy database. A system of logic developed for representing condition that deal with degree of membership and degree of truth. In this paper we use three membership function **S**-shaped membership function, **Z**-shaped membership function and **Pi**-shaped membership function that is based on database crisp balance information. The crisp attribute have precise data, such as Tk 8000 balance and fuzzy attribute consists of imprecise data.

In this paper S-shaped membership function use three linguistic terms very low, low, not so low that depends on the range of the balance of the database.

So we can say that, the whole crisp balance in the database can be classified nine linguistic terms. The schema of used relation in this paper is as follows: Account schema = (account no, account name, branch name, balance), where all attribute are crisp for classical database. For fuzzy database, the balance is only fuzzy attribute and others are crisp. In both cases, account no is primary key. An example, that shows in the table1 about account relation.

A_no	A_name	B_name	Balance
1001	Jamal	Dhaka	7890
1002	Kamal	Dhaka	4500
1003	Saiful	Comilla	5150
1004	Ashiquie	Narial	9450
1005	Sabuj	Narial	8100
1006	Kasham	Kushtia	2500
1007	Sajjad	Bagura	3250
1008	Sumon	Chittagong	15000

Table 1: Account relation

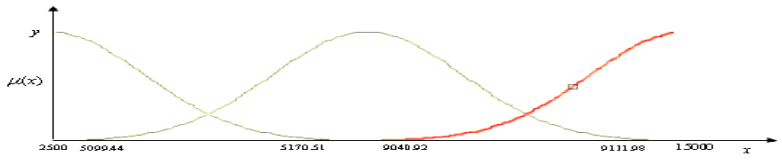

Figure 4: Query sets characterizing balance

This paper, automatically change the range of the membership function of the database. We know the lowest value and height value of the database. If new balance input the database then check the balance it's less then lowest value or greater then height value. If (lowest balance< balance <height balance) then put the balance in appropriate place. Otherwise we follow the steps,

1. First find the arithmetic mean of the whole balance of the database.

$$\mathrm{AM} = \frac{1}{2} \sum_{\mathrm{i\text{-}1}}^{\mathrm{n}} \mathrm{x_i}$$

2. Second find the Standard Deviation.

$$\mathrm{SD} = \sqrt{\frac{1}{n}\sum_{i-1}^{n}(x_i - \mathrm{AM})^2}$$

3. Finally calculate the range of the three membership function for nine linguistic terms.

Higher value of the Zmf is [36],

$$H_l = mean - \frac{sd}{2} + mean \times overlap$$

Lower value of the Pimf is,

$$L_m = mean - \frac{sd}{2} - mean \times overlap$$

Higher value of the Pimf is,

$$H_m = mean + \frac{sd}{2} + mean \times overlap$$

Lower value of the Smf is,

$$L_h = mean + \frac{sd}{2} - mean \times overlap$$

To build a fuzzy database, we construct an individual index file for each linguistic term of fuzzy attribute of that database. This research uses nine index files. These are very low index file, low index file, not so low index file, not so moderate index file, moderate index file, very moderate index file, not so high index file, high index file and very high index file. When insert a new balance in the database then put the balance in appropriate index file according to the range with create fuzzy balance and linguistic balance. For example, the current records is given below,

A_no	A_name	B_name	Balance
1002	Kamal	Dhaka	4500

Table 2: Classical database table

This balance is classified according to the range. We see that, the balance is converting according to the *Z*–shaped membership function and we find the fuzzy balance and linguistic balance. This are given below,

A_no	A_name	B_name	Balance	F_Balance	L_Balance
1002	Kamal	Dhaka	4500	0.999573617	***Not so low***

Table3: Fuzzy database table for not so low linguistic balance

Again we insert the current records

A_no	A_name	B_name	Balance
1003	Saiful	Comilla	5150

Table.4: classical database table for overlapping position balance

Balance is converting in *Z*-shaped membership function and creates fuzzy balance and linguistic balance. Second the balance is converting in*Pi*-shaped membership function and creates fuzzy balance and linguistic balance. That show in the table,

A_no	**A_name**	**B_name**	**Balance**	**F_Balance**	**L_Balance**
1003	Saiful	Comilla	5150	0.999573617	***Not so low***
1003	Saiful	Comilla	5150	0.000198467	***medium***

Table 5: Fuzzy database table for overlapping position balance

The query statement, use to retrieve data from database, which involves imprecise information is called fuzzy query.

4 RESULT

This paper we use fuzzy query in classical database. The necessary steps require to retrieve a record from classical database by suing fuzzy queries, this are given below,

Procedure find (value V)
Set C=root node
While C is not a leaf node begin
Let Ki=smallest search key value, in any greater than V
If there is no such value then begin
Let m= the number of pointers in the node
Set C= node pointed to by Pm
End
Else set C=the node pointed to by Pi
End
If there is a key value Ki in C such that Ki=V
Then pointer Pi directs us to the desired record
Else no record with key value k exists
End procedure.

Q.2 “find the all account numbers whose balance is very high”
The output of this query statement is shown in Table:

A_no	A_name	B_name	Balance	F_Balance	L_Balance
1008	Sumon	Chittagong	15000	1	Very high

Table 6: Output of query statement Q.2

In our example, k=8, n=4, m=1.

$$\text{The query cost} = \left(\log\left\lceil\frac{4}{2}\right\rceil 8+1\right)\times(4+0.1)$$
$$= (3+1)\times(4+0.1)$$
$$= 4\times 4.1$$
$$= 16.4 \text{ ms}$$

5CONCLUSION

This paper presents an intelligent query employing fuzzy logic with an objective to reduce the time complexities of large database. By defining suitable membership functions on the different attributes, we can find the appropriate degree of acceptance / unacceptable for a given value of an attribute. It has been shown that a query costs less time in fuzzy database

and speed up the system. So our proposed system is more useful than traditional queries on classical database. Though it requires some more space because of increasing index file as increasing the linguistic terms in fuzzy database but nevertheless, it is not a matter of considerable problem due to advance in memory technology. On the contrary, the time complexity is drastically reduced. Our proposed intelligent query employing fuzzy logic reduce time complexities of large database then the traditional database system, but yet it suffers some problems that may be possible to overcome. The main drawback is it requires some more space because of increasing index file as increasing the linguistic terms in fuzzy database. But we think that this is not a matter of considerable problem due to advance in memory technology.

Acknowledgments

We are earnestly grateful to one our group member, Protap Mollick, Graduated, Department of CSE, American International University-Bangladesh. For providing us with his special advice and guidance for this project. Finally, we express our heartiest gratefulness to the Almighty and our parents who have courageous throughout our work of the project.

References

S. Yasunobu, S. Miyamoto (1985). *Automatic Train Operation System by Predictive Fuzzy Control*. In M. Sugeno, editor, Industrial Applications of Fuzzy Control (pp. 1-18). North Holland.

V. Novák, R. Yager. (1992). *an Introduction to Fuzzy Logic Applications in Intelligent Systems* (2nd Ed.). New York.

J. Godjevac, I. Nettessur. (1999). *Practical Fuzzy concepts*. Lausanne.

M. M. Gupta. (1995). *Fuzzy Logic and Neural Systems*. In D. Ruan, editor, International Series in Intelligent Technologies, (pp.225-244). Kluwer Academic.

M. Mukaidono. Fuzzy Logic for Beginners (2001). World Sciatic Publishing, London.

L. A. Zadeh (1975). *The Concept of a Linguistic Variable and its Application to Approximate Reasoning Part I*. Information Sciences, (pp199-249).

L. A. Zadeh (1975). *The Concept of a Linguistic Variable and its Application to Approximate Reasoning _ Part II*. Information Sciences, (pp.301-357).

S. Rajashkaran, G.A.Vijayalakshmi (2003).*Pi.Neural Network, Fuzzy Logic, and Genetic Algorithms*, Synthesis and applications.(pp169-173),Hall of India.

L. A. Zadeh (1965). *Fuzzy Sets & Information and Control* (pp.338-353).

Protap Mollick Completed BSc in computer science & engineering from American International University-Bangladesh. Now he is working in a software company as a Software Engineer. His main research area is artificial intelligence, fuzzy logic, Microwave antenna, Human computer interaction. He has a good knowledge about microcontroller embedded system.

Ashiqur Rahman Completed MSc from Jahangirnagar University& BSc from Northern University of Bangladesh. Now he is working as a Lecturer of Royal University of Dhaka.

Md. Shrof Uddin Shibly is a software engineer with extensive experience and management skills and works for a software company. As a software engineer he is responsible for the development of management software products produced by the company. He received his B.Sc. degree in Computer Science and Software Engineering from the American International University-Bangladesh in 2014.

Shimul Kanti Bala. Completed his BSc graduate on software engineering in AIUB which is the top leading university in my country. Now he is working in a software company named by Data edge as an ASP.net programmer.