International Journal of Information Technology and Knowledge Management
January-June 2011, Volume 4, No. 1, pp. 57-61

NEUTROSOPHIC RELATIONAL DATABASE SEARCH APPROACH TO CAPTURE INCONSISTENT INFORMATION

Meena Arora[1] & Ranjit Biswas[2]

Most of our traditional tools for formal modeling, reasoning and computing are crisp, deterministic and precise in nature. Complete description of a real system would require far more detailed data than a human being could ever recognize and process simultaneously. Hence, among the various paradigmatic changes in science and mathematics in last century, one such has been the concern of the concept of neutrosophy. There is a genuine necessity to develop such a system which should be able to answer the users queries posed in natural language, without giving much botheration to the users. Such type of queries are not crisp in nature, and involve predicates with fuzzy (or rather neutrosophic data, fuzzy/ neutrosophic hedges (with concentration or dilation) etc. I will propose a new type of searching techniques using neutrosophic theory to meet the predicates posed in natural language in order to answer imprecise queries of the users. Thus it is a kind of an intelligent search for match in order to answer imprecise queries of the lay users. We will propose a new type of searching techniques using neutrosophic theory to meet the predicates posed in natural language in order to answer imprecise queries of the users. Thus it is a kind of an intelligent search for match in order to answer imprecise queries of the lay users.

Keywords: Rank Neutrosophic Search, Neutrosophic Relation, Neutrosophic Relational Data Model

1. Introduction

In real-life problems, the data associated are often imprecise, or non-deterministic. All real data cannot be precise because of their fuzzy nature. Imprecision can be of many types: non-matching data values, imprecise queries, inconsistent data, misaligned schemas, etc. Ultimate cause for many research areas: data mining, semi structured data, schema matching, nearest neighbor. Consequently, there is a genuine necessity for the different large size organizations, especially for the industries, companies having world wide business, to develop such a system which should be able to answer the users queries posed in natural language, irrespective of their grammar, without giving much botheration to the users.

The root cause of the disparity between common-sense queries and the keyword approach of today's engines is this: a user's search queries are often an approximation and synopsis of his/her information needs, so purely matching against the terms in the search query is a woefully inadequate method for finding the correct or even correlated information.

For example, when we ask the opinion of an expert about certain statement, he or she may say that the possibility that the statement is true is 0.5 and the statement is false is 0.6 and the degree that he or she is not sure is 0.2.

To deal with uncertainties in searching match for such queries, neutrosophic ranking search will be the appropriate tool. In this paper we analyze how our new method of neutrosophic search will be different and rather an improved method from the existing methods ([2] [3] [4] [9] [11] [13] [14]). Our method, being an intelligent soft-computing method, will support the users to make and find the relevant answers without iteratively refining their queries.

2. Present Work

The various approaches used for Database search are discussed below:

2.1. The Classical Relational Model Approach

A classical relational database [7] consists of a collection of relations. A relation is a table of values where each row represents a collection of related data values. In a table, each row is called a tuple, a column header is called an attribute and the table as a whole is called the relation. A relation schema $R(A_1, A_2,....,A_n)$ consists of a relation name R and a list of attributes $A_1, A_2,....,A_n$.

There are various restrictions on data in the form of constraints. Domain constraints specify that each value of an attribute A_i must be an atomic value from the domain $dom(A_i)$. This includes restrictions on data types, on the range of values (if any), and on the format of data.

2.2. Fuzzy based Model Approach

Fuzziness can be defined as the vagueness concerning the semantic meaning of events, phenomenon or statements themselves. It is particularly frequent in all areas in which

[1]Department of CSE, JSS Academy of Technical Education, Noida, India
[2]Department of CSE, Manav Rachna University, Faridabad, India
Email: [1]meena23dec@gmail.com, [2]ranjitbiswas@yahoomail.com

human judgment, evaluation and decisions are important. As an example consider a student record database system. Supposing we want to find bright and young students in the whole batch. For a crisp system we would specify the query as

PROJECT (Student_Name) WHERE $19 \leq AGE \leq 23$ and $3 \leq GPA \geq 4$

But this system has a major flaw. Consider a student, Arjun whose age is 24 and has a good GPA of 4 out of 4. He should have been selected but is not. It is because of the rigid boundary conditions set by the normal crisp logic.

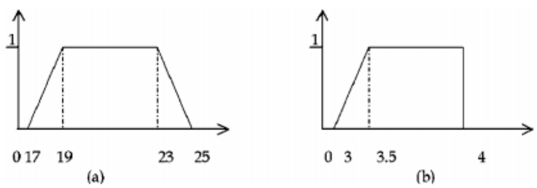

Fig. 2.1: a) AGE, b) GPA

In Set to be considered as Arjun also satisfies the query to some extent, which is represented by its membership grade.

Once fuzzy relations are defined, it is possible to develop fuzzy relational databases.

2.3. The Similarity Based Fuzzy Relation Approach

In similarity based fuzzy relation, the imprecision of attribute values is primarily in their meaning e.g. the values of the attribute EXPERTISE in the relation shown often partially overlap. The overlapping degree of these areas of expertise is specified by similarity matrix. Consider a PROFILE relation as shown in table 2.1 and let its similarity matrix be as shown in Table 2.2.

Table 2.1
Profile Relation

FName	Job Type	Expertise
Bob	Academic	A.I
Rob	Industry	Expert System
Nancy	Government	Statistics
John	Government	Robotics

A Similarity Matrix:

Table 2.2
Similarity Matrix Table

	Robotics	Expert System	AI	Statistics
Robotics	1.0	0.6	0.6	0.2
Expert System	0.6	1.0	0.9	0.2
AI	0.6	0.9	1.0	0.2
Statistics	0.2	0.2	0.2	1.0

In addition the imprecision due to semantics, a similarity based fuzzy relation could also describe imprecision due to incomplete information. To do this, an attribute in a tuple is allowed to have set of values which describe all possible values for the attribute in the tuple. A similarity based fuzzy relation involving Domain D1, D2,….Dk is thus a subset of Cartesian product.

$$2^{D1} * 2^{D2} * \ldots\ldots 2^{Dk}$$

Furthermore, the values in domains D1, D2, ….Dk are related by similarity relation S1, S2,….Sk, that map each pair of values in a domain to the interval [0, 1]. Where 1 means “identical” and 0 means “Totally different”.

2.4. The Possibility based Logic Approach

This is based on possibility distribution theory. The possibility / necessity approach is more general than the similarity approach in the sense that it handles all types of information.

Possibilitic logic is a weighted logic which aims to enable reasoning with uncertain knowledge introduced by L. Zadeh in [15].

A possibility based fuzzy relation generalizes a relation by allowing the value of an attribute A to be a possibility distribution $\Pi_{A(t)}$ of the attributes domain. Let D1, D2,….Dk be the attributes of a fuzzy relation R. A tuple in a possibility based fuzzy relation is denoted by (v1,v2,….vk) where (v1, v2,….vk) are fuzzy subsets of D1, D2,…..Dk. The interest of such an approach is its ability to represent in a unified manner, precise values (singletons) NULL values, as well as fuzzy values. It should be noted that since data is imprecise, then the result of query will also be imprecise. In theory both possibility measures and necessity measures can be used for processing queries in fuzzy database. In practice, however, the necessity measures are rarely used.

2.5. The Probabilistic Logic based Approach

Probabilistic logic corresponds to “probability, likelihood”; probabilities range between 0 and 1 and hence may seem similar at first. For example, let a 100 ml glass contain 30 ml of water. Then we may consider two concepts: Empty and Full. The meaning of each of them can be represented by a certain fuzzy set. Then one might define the glass as being 0.7 empty and 0.3 full. Note that the concept of emptiness would be subjective and thus would depend on the observer or designer.

3. RANK NEUTROSOPHIC SEARCH APPROACH

Compared with all other types of sets, in the neutrosophic set each element has three components which are subsets (not numbers as in fuzzy set) and considers a subset, similarly to intuitionistic fuzzy set, of “indeterminacy” - due to unexpected parameters hidden in some sets, and let the superior limits of the components to even boil over 1 (overflooded) and the inferior limits of the components to even freeze under 0 (under dried).

3.1. The Neutrosophic Logic

In the Neutrosophic Logic (which is a generalization of fuzzy logic, especially of intuitionistic fuzzy logic) every logical variable x is described by an ordered triple x = (T; I; F), where T is the degree of truth, F is the degree of falsehood, and I the degree of indeterminacy (or neutrality, i.e. neither true nor false, but vague, unknown, imprecise), with T; I; F standard or non-standard subsets of the non-standard unit interval $]^{-}0; 1+[$. In addition, these values may vary over time, space, hidden parameters, etc.

There is a genuine necessity to develop such a system which should be able to answer the users queries posed in natural language, without giving much botheration to the users.

Let A and B be two neutrosophic sets. One can say, by language abuse, that any element neutrosophically belongs to any set, due to the percentages of truth/indeterminacy/falsity involved, which varies between 0 and 1 or even less than 0 or greater than 1.

Thus: x(50,20,30) belongs to A (which means, with a probability of 50% x is in A, with a probability of 30% x is not in A, and the rest is undecidable); or y(0,0,100) belongs to A (which normally means y is not for sure in A); or z(0,100,0) belongs to A (which means one does know absolutely nothing about z's affiliation with A).

More general, x ((20-30), (40-45) $\cup$ [50-51], {20, 24, and 28}) belongs to the set A, which means:

- With a probability in between 20-30% x is in A (one cannot find an exact approximate because of various sources used);
- with a probability of 20% or 24% or 28% x is not in A;
- the indeterminacy related to the appurtenance of x to A is in between 40-45% or between 50-51% (limits included);

The subsets representing the appurtenance, indeterminacy, and falsity may overlap, and n_sup = 30+51+28 > 100 in this case.

3.2. Neutrosophic Relational Data Model (NRDM)

It is based on the neutrosophic set theory which is an extension of intuitionistic fuzzy set theory and is capable of manipulating incomplete as well as inconsistent information. We use both truth-membership function grade α and falsity-membership function grade β to denote the status of a tuple of a certain relation with $\alpha, \beta \in [0, 1]$ and $\alpha + \beta \leq 2$. NRDM is the generalization of fuzzy relational data model (FRDM) i.e. when $\alpha + \beta = 1$, neutrosophic relational model is the ordinary fuzzy relation. neutrosophic sets. (0.6, 0.42) (0.8, 0.01).

3.3. Interval Ranking

Intervals are not ordered. Owing to this major weakness, there is no universal method of ranking a finite (or infinite) number of intervals. But in real life problems dealing with intervals we need to have some tactic to rank them in order to arrive at some conclusion. We will now present a method of ranking of intervals, which we shall use in our work here in subsequent section.

A Rank Neutrosophic search of predicates is basically composed of two types of search which are:

1) α_i Rank Neutrosophic-equality search,
2) Rank Neutrosophic-proximity search.

Therefore, first of all we will introduce the methods of these above two searches, and then finally we will introduce the method of Rank Neutrosophic-search.

3.4. α Rank Neutrosophic Equality Search

Consider a STUDENTS database which is STUDENTS (STUDENT NAME, ROLL NO, SEX, AGE, EYE COLOUR, PHONE NO, and GPA).

If there is a query posed in natural language (by a lay user) like below:

PROJECT (STUDENT NAME) WHO ARE "bright" AND "young";

Then the existing standard query languages will fail to answer it.

Now consider another normal type of query like

PROJECT (STUDENT NAME) WHERE AGE = "approximately 20":

The standard SQL is unable to provide any answer to this query as the search for an exact match for the predicate will fail. The value "approximately 20" is not a precise data. Any data of type "approximately x", "little more than x", "slightly less than x", "much greater than x" etc. are not precise or crisp, but they are Rank Neutrosophic numbers(RNN).

Definition 3.4.1 α parameter

Consider a choice-parameter $\alpha \in [0, 1]$. A member of a of Dom (AGE) is said to be α-Rank Neutrosophic-equal to the quantity "approximate x" if a ∈ $I_{\alpha(x)}$ where is the α -cut of the Rank Neutrosophic number I(x). The degree or amount of this quality is measured by the interval $m_{I(x)}(a) = [t_{I(x)}(a), 1 - f_{I(x)}(a)]$.

Denote the collection of all such a-Rank Neutrosophic-equal members from dom(AGE) by the notation $AGE_\alpha(x)$, which is a subset of dom(AGE). If $AGE_\alpha(x)$ is not a null-set or singleton, then the members can be ranked by ranking their corresponding degrees of equality.

3.5. Rank Neutrosophic Proximity Search

The notion of α-Rank Neutrosophic-equality search as explained above is appropriate while there is an Rank Neutrosophic-predicate in the query involving rank neutrosophic numbers. But there could be a variety of Rank Neutrosophic predicates existing in a Rank Neutrosophic query, many of them may involve Rank Neutrosophic fuzzy hedges (including concentration/dilation) like "good", "very good", "excellent", "too much tall", "young", "not old", etc. In this section I present another type of search for finding out a suitable match to answer imprecise queries. In this search I will use the theory of Rank Neutrosophic-proximity relation [1][5].

Consider the STUDENTS database as described earlier and a query like PROJECT(STUDENT NAME) WHERE EYE-COLOR="dark-brown":

The value/data "dark-brown" is not in the set dom(EYECOLOR). There fore a crisp search will fail to answer this. The objective of this research work is to overcome this type of drawbacks of the classical SQL. For this we notice that there may be one or more members of the set dom (EYE-COLOR) which may closely match the eye-color of "brown" or "dark brown".

Consider a new universe given by

W = dom (EYE-COLOR) ∪ {dark-brown}.

Consider a Rank Neutrosophic-proximity relation R over W. Choose a decision-parameter α ∈ [0, 1]. Search is to be made for the match e ∈ dom(EYE-COLOR) such that t_R(dark-brown,e) ≥ α. (It may be mentioned here that the condition t_R(dark-brown, e) ≥ α does also imply the condition f_R(dark-brown, e) ≥ 1 – α). We can say that e is a close match with "dark-brown" with the degree or amount of closeness being the interval mdark¡brown(e) given by

$_{\text{mdark¡brown}}(e) = [t_R(\text{dark –brown}, e), 1 - f_R(\text{dark –brown}, e)]$:

At β level of choice, the truth-value $t(p_1; p_2)$ of the matching of the predicate p_1: given by EYE COLOR ="dark-brown" with predicate p_2: AGE=e is equal to the β-value of the interval $_{\text{mdark-brown}}(e)$.

Definition 3.5.1 β-value of an Interval

Let J = [a; b] be an interval. The β -value of the interval J is a non-negative real number J, given by $J_\beta = (1 - \beta).a + \beta.\ b$. Clearly, $0 \le J_\beta \le 1$, and β = 0, J_β = a which signifies that the decision-maker is pessimistic, and also for β = 1, J_β = b which signifies that the decision maker is optimistic. For β = :5 it is the arithmetic-mean to be chosen usually for a moderate decision.

3.6. Rank Neutrosophic Search Technique

In this section we will now present the most generalized method of search called by Rank Neutrosophic-search. The Rank Neutrosophic search of matching is actually a combined concept of á-Rank Neutrosophic-equality search, Rank Neutrosophic-proximity search and crisp search.

For example, consider a query like

PROJECT (STUDENT NAME) WHERE (SEX="M", EYE-COLOR="dark-brown", AGE="approximately 20"):

This is a Rank Neutrosophic query. To answer such a query, matching is to be searched for the three predicates P1, P2 and P3 given by

P1: SEX="M" ;P2: EYE COLOR="dark-brown" and P3: AGE="approximately 20",

where P1 is crisp and P2, P3 are Rank Neutrosophic.

Clearly, to answer this query the proposed Rank Neutrosophic search method is to be applied, because in addition to crisp search, both of a-Rank Neutrosophic-equality search and Rank Neutrosophic proximity search will be used to answer this query. It is obvious that the Rank Neutrosophic-search technique for predicate matching reduces to a new type of fuzzy-search technique as a special case.

4. CONCLUSION

In this paper, we have introduced a new method of neutrosophic search with rank and rank neutrosophic sets to answer imprecise queries of the lay users from the databases which will be a great help to bioinformatics groups, consisting of computational biologists and bioinformatics computer scientists in unraveling the mass of information generated by large scale sequencing efforts underway in laboratories around the world. The search used to answer different queries suggested in ([4], [6], [8], [10], [12], [15]) are not the same to our proposed method. In this paper we have introduced a new paradigm that offers for greater resources for managing complexity. Consequently it can effectively deal with broader class of problems. In addition our search method as explained will also help in evaluating information gathered through Web mining. Also this will help decision makers to compile useful information from a combination of raw data, documents or business models to identify and solve problems and make decisions. As future work, I want to extend this paper to study Rank Neutrosophic Armstrongs Axioms which constitute an important part of a good NRDM.

REFERENCES

[1] Atanassov, K., Intuitionistic Fuzzy Sets : Theory and Applications, Physica-Verlag (2000), New-York.

[2] Bosc P., Liétard I and Pivert O, "Evaluation of Flexible Queries : The Quantified Statement Case", Technologies for Constructing Intelligent Systems I, Physica-Verlag Heidelberg, New York, pp 337-350, 2002.

[3] Bosc P. & Pivert O., "Some Approaches for Relational Databases Flexible Querying", Journal of Intelligent Information Systems, 1, 1992, pp 323-354.

[4] Buckles, B.P., Petry, F.E., 1985. "Query Languages for Fuzzy Databases", In: Kacprzyk, J., Yager, R. (Eds.), Management Decision Support Systems Using Fuzzy Sets and Possibility Theory, TUV Rheiland, Verlag, Berlin, pp. 241-251.

[5] Chen, Shyi-Ming, "Analyzing Fuzzy System Reliability using Rank Neutrosophic Set Theory", Int. Jou. of Applied Sc. & Engg, 1, 2003, 82-88.

[6] Chiang D., Chow L.R. and Heien N, "Fuzzy Information in Extended Fuzzy Relational Databases", Fuzzy Sets and Systems 92, pp.1-10., 1997.

[7] Codd, E.F., "A Relational Model of Data for Large Shared Data Banks", Communications of the ACM, 13(6), p.377-387, June 1970.

[8] Cohen,W., "Integration of Heterogeneous Databases without Common Domains using Queries based on Textual Similarity", Proc.OF SIGMOND, June 1998,pages 201-212. the Web, Proceeding of WWW, Hawai, USA, May 2002.

[9] Didier Dubois and Henri Prade, "Using Fuzzy Sets in Flexible Querying : why and how?", Flexible Query Answering Systems, Kluwer Academic Publishers, Norwell, MA, 1997.

[10] Kaushik S,and Nanda, H., "Web based Access of Relational Databases using Fuzzy Natural Language Queries", International Conference on Cognitive Systems, Delhi, India, (1999).

[11] Rolly Intan and Masao Mukaidono, "Approximate Data Querying in Fuzzy Relational Database", Journal of Advanced Computational Intelligence and Intelligent Informatics, 6, No.1, pp. 33-40, 2002.

[12] Smarandache, F. (1999), "A Unifying Field in Logics. Neutrosophy: Neutrosophic Probability", Set and Logic, Rehoboth: American Research Press.

[13] Shyi-Ming Chen and Woei-Tzy Jong, "Fuzzy Query Translation for Relational Database Systems", IEEE Trans. Systems, Man and Cybernetics, (Part B), Aug 1997, 27(4), pp 714-721.

[14] Yang Q., Zhang W., Liu C., Wu J., Nakajima H. and Rishe N.D., "Efficient Processing of Nested Fuzzy SQL Queries in a Fuzzy Database", IEEE Trans. On Knowledge and Data Eng., 13, No. 6, pp. 884-901, Nov/Dec 2001.

[15] Zadeh, L.A., "Fuzzy Sets," Information and Control, 8, pp338-353, (1965).

[16] Ullman, J.D., Principles of Database Systems, Galgotia, India, 1984.

[17] Joseph Amadee Goguen, " The Logic of Inexact Concepts", In : Synthese 19.3-4(1969), Pages 325-373.