Deployment of Neutrosophic Technology to Retrieve Answer for Queries Posed in Natural Language

Meena Arora
CSE Department
JSS Academy of Technical Education
Noida, INDIA
meena23dec@gmail.com

Ranjit Biswas
CSE Department
Institute of Technology and Management
Gurgoan, INDIA
ranjitbiswas@yahoo.com

***Abstract*— In this paper, we have we have introduced a new intelligent soft-computing method of neutrosophic search with ranks and a new neutrosophic rank sets for neutrosophic relational data model (NRDM). Essentially the data and documents on the Web are heterogeneous; inconsistency is unavoidable in Web mining. Using the presentation and reasoning method of our data model, it is easier to capture imperfect information on the Web which will provide more potentially valued-added information. In Bio-informatics there is a proliferation of data sources. Each research group and each new experimental technique seems to generate yet another source of valuable data. But these data can be incomplete and imprecise, and even inconsistent. We could not simply throw away one data in favor of other data. So now we can represent and extract useful information from these data as a challenge. Thus it is a kind of an intelligent search for match in order to answer imprecise queries of the lay users. Our method, being an intelligent soft-computing method, will support the users to make and find the answers to their queries without iteratively refining them by trial and error. This important issue of closeness cannot be addressed with the crisp mathematics. That is why we have used the Neutrosophic tools. Neutrosophic-search method could be easily incorporated in the existing commercial query languages of DBMS to serve the lay users better. So in this Paper Authors are suggesting NRDM and Rank Sets to solve the imprecise query based on Rank Neutrosophic search which is a combination - Neutrosophic Proximity search and α-Neutrosophic-equality Search .**

Keywords-* *α-Neutrosophic-equality Search , proximity search, Rank neutrosophic sets, Rank Neutrosophic search , Neutrosophic relation, Neutrosophic relational data model.

I. 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. The root cause of the disparity between such 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. An item belongs to the database" is a probabilistic event and it can be extended to all data models; here we discuss probabilistic *relational* data.

With the rapid growth of the amount of data available in electronic libraries, through Internet and enterprise network mediums, advanced methods of search and information retrieval are in demand. Two Types of Probabilistic relational Data are there, *Database is deterministic and Query answers are probabilistic* or *Database is probabilistic and Query answers are probabilistic.* Deployment of neutrosophic technology allows stating flexible, smooth and vague search criteria and retrieving a rich set of relevance ranked documents aiming to supply the inquirer with more satisfactory answers.

Some previous approaches sidestepped complexity .Now our implementation includes Ranking query answers. To deal with uncertainties in searching match for such queries, neutrosophic ranking search will be the appropriate tool. In this paper we propose two things – neutrosophic search with ranks and the Rank sets.

Intuitionistic fuzzy sets can only handle incomplete information not the indeterminate information and inconsistent information which exists commonly in belief system.

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.

In neutrosophic set, indeterminacy is quantified explicitly and truth-membership, indeterminacy-membership and falsity-membership are independent. This assumption is very important in a lot of situations such as information fusion when we try to combine the data from different sensors.

And to deal with this imprecision we are using a Ranking method based on Neutrosophic Logic with rank neutrosophic sets.

II. Neutrosophic Logic

Definition 2.1

A logic in which each proposition is estimated to have the percentage of truth in a subset T, the percentage of indeterminacy in a subset I, and the percentage of falsity in a subset F, where T, I, F are defined below, is called *Neutrosophic Logic.* Constants: (T, I, F) truth-values, where

978-1-4244-5540-9/10/$26.00 ©2010 IEEE

T, I, F are standard or non-standard subsets of the nonstandard interval $]^-0, 1^+[$, where n_{inf} = inf T + inf I + inf F≥ $^-0$, and n_{sup} = sup T + sup I + sup F≤ 3^+. **Neutrosophic logic [6]** was created by Florentin Smarandache (1995)

III. NEUTROSOPHIC SETS

Definition 3.1 (neutrosophic Set) :Let X be a space of points (objects), with a generic element in X denoted by x. A neutrosophic set **A** in **X is** characterized by a truth-membership function $\mathbf{T_A}$, an indeterminacy-membership function $\mathbf{I_A}$ and a falsity-membership function $\mathbf{F_A}$. $\mathbf{T_A(x)}$, $\mathbf{I_A(x)}$ and $\mathbf{F_A(x)}$ are real standard or non-standard subsets of$]^-0, 1^+[$. That is

$$T_A : X \rightarrow]^-0, 1^+[\quad (1)$$

$$I_A : X \rightarrow]^-0, 1^+[\quad (2)$$

$$F_A : X \rightarrow]^-0, 1^+[\quad (3)$$

There is no restriction on the sum of T_A**(x)**, I_A**(x)** and FA(x) so -0 ≤sup TA(x)+sup IA(x) + sup FA(x) ≤3+.

Definition 3.2 (Complement) The complement of a neutrosophic set A is denoted by c(A) and is defined by

$$T_{c(A)}(x) = \{1^+\} - T_A(x), \quad (4)$$

$$I_{c(A)}I(x) = \{1^+\} - I_A(x), \quad (5)$$

$$F_{c(A)}(x) = \{1^+\} - F_A(x), \quad (6)$$

for all x in X.

Definition 3.3 (Union) The union of two neutrosophic sets A and B is a neutrosophic set C, written as C = A ⋃ B, whose truth-membership, indeterminacy-membership and falsity-membership functions are related to those of A and B by

$$T_C(x) = T_A(x) + T_B(x) - T_A(x) \times T_B(x), \quad (7)$$

$$I_C(x) = I_A(x) + I_B(x) - I_A(x) \times I_B(x), \quad (8)$$

$$F_C(x) = F_A(x) + F_B(x) - F_A(x) \times F_B(x), \quad (9)$$

for all x in X.

Definition 3.4 (Intersection)

The intersection of two neutrosophic sets **A** and **B** is a neutrosophic **set C,** written **as C = A ∩ B,** whose truth-membership, indeterminacy-membership and falsity-membership functions are related to those of A and B by

$$T_C(x) = T_A(x) \times T_B(x), \quad (10)$$

$$I_C(x) = I_A(x) \times I_B(x), \quad (11)$$

$$F_C(x) = F_A(x) \times F_B(x), \quad (12)$$

for all x in X.

IV. RANKING

Definition 4.1

Ranking: Ranking is defined as Computing a similarity score between a tuple and the query,

Consider the query

Q = SELECT* From R

Where $A_1 = v_1$ and … and $A_m = v_m$

Query is a vector: Q = $(v_1, \ldots, v_m)$

Tuple is a vector: T = $(u_1, \ldots, u_m)$

Consider the applications: personalized search engines, shopping agents, logical user profiles, soft catalogs. To answer the queries related with the above application Two approaches are given:

- Qualitative→Pare to semantics (deterministic)
- Quantitative→alter the query ranking

With imprecise values specified this way, their probabilistic indexing weight can be derived easily. An excellent style manual for science writers is [7].

V. RANK NEUTROSOPHIC SETS

In this section, we will now present the notion of rank neutrosophic set (RNS). RNS is an instance of neutrosophic set which can be used in real scientific and engineering applications.

Consider parameters such as capability, trustworthiness and price of semantic Web services. These parameters are commonly used to define quality of service of semantic Web services. In this section, we will use the evaluation of quality of service of semantic Web services [7] as running example to illustrate every set-theoretic operation on rank neutrosophic sets.

Example 1 : Assume that X = $[x_1, x_2, x_3]$. x_1 is capability, x_2 is trustworthiness and x_3 is price. The values of x_1, x_2 and x_3 are in [0,1]. They are obtained from the questionnaire of some domain experts, their option could be a degree of "good service", a degree of indeterminacy and a degree of "poor service". A is a single valued neutrosophic set of X defined by

A = <0.3,0.4,0.5>/ x_1 + <0.5,0.2,0.3>/ x_2 + <0.7,0.2,0.2> / x_3. B is a single valued neutrosophic set of X defined by B = <0.6,0.1,0.2>/x_1 + <0.3,0.2,0.6>/x_2 + <0.4,0.1,0.5>/x_3.

Definition 5.1 (Complement) The complement of a single valued neutrosophic set A is denoted by c(A) and is defined by

$$T_{c(A)}(x) = F_A(x), \quad (13)$$

$$I_{c(A)}(x) = 1 - I_A(x), \quad (14)$$

$$F_{c(A)}(x) = T_A(x), \quad (15)$$

for all x in X.

Example 2 Let A be the single valued neutrosophic set defined in Example 1. Then, c(A) = <0.5,0.6,0.3>/ x_1 + <0.3,0.8,0.5>/x_2 + <0.2,0.8,0.7>/x_3.

Definition 5.2 (Union) The union of two single valued neutrosophic sets A and B is a single valued neutrosophic set C, written as C = A ⋃ B, whose truth-membership, indeterminacy-membership and falsity-membership functions are related to those of A and B by

$$T_C(x) = \max (T_A(x), T_B(x)), \quad (16)$$
$$I_C(x) = \max (I_A(x), I_B(x)), \quad (17)$$
$$F_C(x) = \min(F_A(x),F_B(x)), \quad (18)$$

for all x in X.

Example 3 Let A and B be the single valued neutrosophic sets defined in Example 1. Then, $A \cup B = <0.6, 0.4,0.2>/x_1 + <0.5, 0.2,0.3>/x_2 + <0.7, 0.2,0.2>/x_3$.

Definition 5.3 (Intersection) The intersection of two single valued neutrosophic sets A and B is a single valued neutrosophic set C, written as C = A ∩ B, whose truth-membership, indeterminacy-membership and falsity-membership functions are related to those of A and B by

$$T_C(x) = \min (T_A(x), T_B(x)), \quad (19)$$
$$I_C(x) = \min (I_A(x), I_B(x)), \quad (20)$$
$$F_C(x) = \max(F_A(x),F_B(x)), \quad (21)$$

for all x in X.

Example **4** Let A and B be the single valued neutrosophic sets defined in Example 1. Then, $A \cap B = <0.3,0.1,0.5>/ x_1 + <0.3,0.2,0.6>/x_2 + <0.4,0.1,0.5>/x_3$.

Definition 5.4 (Difference) The difference of two single valued neutrosophic set C, written as C = A \ B, whose truth-membership, indeterminacy-membership and falsity-membership functions are related to those of A and B by

$$T_C(x) = \min(T_A(x),F_B(x)), \quad (22)$$
$$I_C(x) = \min(I_A(x),1-I_B(x)), \quad (23)$$
$$F_C(x) = \max(F_A(x),T_B(x)), \quad (24)$$

for all x in X.

Example 5 Let A and B be the single valued neutrosophic sets defined in Example 1. Then $A \setminus B = <0.2,0.4,0.6>/ x_1 + <0.5,0.2,0.3>/x_2 + <0.5,0.2,0.4>/x_3$.

Now we will define two operators: truth-favorite (Δ) and falsity-favorite (∇) to remove the indeterminacy in the single valued neutrosophic sets and transform it into intuitionistic fuzzy sets or paraconsistent sets. These two operators are unique on single valued neutrosophic sets.

*Definition 5.6 (*Truth-*favorite)* The truth-favorite of a single valued neutrosophic set A is a single valued neutrosophic set B, written as B = ΔA, whose truth-membership and falsity-membership functions are related to those of A by

$$T_B(x) = \min (T_A(x)+I_A(x),1) \quad (25)$$
$$I_B(x) = 0, \quad (26)$$
$$F_B(x) = F_A(x), \quad (27)$$

for all x in X.

Example 7 Let A be the single valued neutrosophic set defined in Example 1. Then

$\Delta A = <0.7, 0, 0.5>/x_1 + <0.7, 0, 0.3>/x_2 + <0.9, 0, 0.2>/x_3$.

Definition 5.7 (Falsity-favorite) The falsity-favorite of a single valued neutrosophic set B, written as B = ∇A, whose truth-membership and falsity-membership functions are related to those of A by

$$T_B(x) = T_A(x), \quad (28)$$
$$I_B(x) = 0, \quad (29)$$
$$F_B(x) = \min(F_A(x)+I_A(x),1), \quad (30)$$

for all x in X.

Example 8 Let A be the single valued neutrosophic set defined in Example 1. Then

$\nabla A = <0.3, 0, 0.9>/x_1 + <0.5, 0, 0.5>/x_2 + <0.7, 0, 0.4>/x_3$.

VI. Neutrosophic Relational Data Model

A. *Neutrosophic relation*

In this section, we will define the Neutrosophic relation. A tuple in a neutrosophic relation is assigned a measure that will be referred to as the *truth* factor and also as the *false* factor. The interpretation of this measure is that we believe with confidence and doubt with confidence that the tuple is in the relation. The truth and false confidence factors for a tuple need not add to exactly 1. This allows for incompleteness and inconsistency to be represented.

B. 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.

VII. Rank Neutrosophic Search

Most of the type of queries are not crisp in nature, and involve predicates with fuzzy (or rather vague) data, fuzzy/rank edges. Thus, these types of queries are not strictly confined within the domains always. The corresponding predicates are not hard as in crisp predicates. Some predicates are soft because of neutrosophic/ fuzzy nature and thus to answer a query a hard match is not always found from the databases by search, although the query is nice and very real, and should not be ignored or replaced according to the business policy of the industry. To deal with uncertainties in searching match for such queries, neutrosophic rank logic will be the appropriate tool.

A rank neutrosophic search of predicates is basically composed of two types of search which are:-

(i) α-neutrosophic equality search, and

(ii) neutrosophic-proximity search

Consider a STUDENTS database which is STUDENTS (STUDENT_NAME, ROLL_NO, AGE, COMPLEXION, PHONE_NO, GPA). Suppose that the all required integrity constraints imposed on this database are on the domains of some attributes, given by dom(AGE) = [20,25], dom(COMPLEXION) = {black, white, fair, tan},

dom(SEX) = {M, F}, dom(GPA) = [0,5] Now consider a crisp query in a QL made by a System-Manager like below: PROJECT (STUDENT_NAME), WHERE $20 \leq AGE \leq 25$ and $3.5 \leq GPA \leq 4.5$. The answer will be immediately available. But if there is a query posed in natural language (by a lay user) like below:

Query 1.

PROJECT (STUDENT_NAME) WHO ARE "bright" AND "young", then the existing standard query languages will fail to answer it.

This type of query can be solved by rank neutrosophic search.

A. *α-NEUTROSOPHIC equality search*

Consider one half of above query like

Query 2.

PROJECT (STUDENT_NAME)

WHERE AGE = "approximately 22".

Now in both the above cases 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 22" or "young" and "bright" 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 neutrosophic numbers (NN) .

Denote any one of them, say the neutrosophic number "approximately x" by the notation I(x). We know that a Neutrosophic number is a Neutrosophic Set of the real numbers. Clearly for every member a Є dom(AGE), there is a membership value $t_{I(x)}(a)$ proposing the degree of equality of this crisp number a with the quantity "approximately x", and a non membership value $f_{I(x)}(a)$ proposing the degree of non equality . Thus, in neutrosophic philosophy of samarandache, every element of dom (AGE) satisfies the predicate AGE = "approximately 22" up to certain extent and does not satisfy too, up to certain extent. But we will restrict ourselves to those members of dom(AGE) which are α-neutrosophic-equal, the concept of which we will define below. Any imprecise predicate of type AGE = "approximately 22", or of type AGE = "young" (where the attribute value "young" is not a member of the dom(AGE)), is to be called by Neutrosophic-predicate, and a query involving Neutrosophic-predicate is called to be a Neutrosophic query.

Definition 7.1 Consider a choice –parameter α Є [0, 1]. A member of a of dom (AGE) is said to be -rank-equal to the quantity "approximate x" if a Є I_{α} (x) where a is the α -cut of the 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-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.

B. *NEUTROSOPHIC-Proximity Search*

The notion of α-neutrosophic-equality search and β – value of an interval as explained above is appropriate while there is a Neutrosophic-predicate in the query involving neutrosophic numbers. But there could be a variety of neutrosophic predicates existing in a neutrosophic query, many of them may involve neutrosophic fuzzy hedges (including concentration/dilation) like "good", "very good", "excellent", "too much tall", "young", "not old", etc. In this section we present another type of search for finding out a suitable match to answer imprecise queries. In this search we will use the theory of neutrosophic-proximity relation [1]. We know that a neutrosophic-proximity relation on a universe U is a neutrosophic relation on U which is both neutrosophic reflexive and neutrosophic-symmetric.

Consider the STUDENT database as described below and a query like

PROJECT (STUDENT_NAME)

WHERE COMPLEXION = "very-fair".

The value/data "very-fair" is not in the set dom (COMPLEXION). Therefore 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 (COMPLEXION) which may closely match the complexion of "white" or "fair".

Consider a new universe given by

W = dom (COMPLEXION) U {very-fair}.

Now propose a neutrosophic-proximity relation R over W. Choose a decision-parameter α Є [0,1]. We propose that search is to be made for the match e Є dom(COMPLEXION) such that t_R(very-fair, e) ≥ α. (It may be mentioned here that the condition tR(very-fair, e) ≥ α does also imply the condition f_R(very-fair, e) ≤ 1- α). We say that e is a close match with "very-fair" with the degree or amount of closeness being the interval $m_{very-fair}(e)$ given by

$m_{very-fair}(e) = [\ t_R(\text{very-fair},e),\ 1- f_R(\text{very-fair},e)\ ]$.

At β level of choice, the truth-value *t*(p1,p2) of the matching of the predicate p1: given by COMPLEXION = "very-fair" with the predicate p2: AGE = e is equal to the β-value of the interval $m_{very-fair}(e)$.

Definition 7.2 Consider a choice value β Є [0,1]. At β level of choice, for every element a of AGE_a (x), the truth – value t (p_1,p_2) of the matching of the predicate p_1: given by AGE="approximately x" with the predicate p_2: AGE = a is equal to the β –value of the interval $m_{I(x)}$ (a).

C. *Neutrosophic Search with ranks*

In this section we will now present the most generalized method of search called by neutrosophic rank –search. The neutrosophic rank search of matching is actually a combined concept of α –equality search, proximity search and crisp search. For example, consider a query like PROJECT (STUDENT_NAME) WHERE (SEX="M", COMPLEXION ="very-fair", AGE="approximately 22"). This is a 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: COMPLEXION="very-fair" and P3: AGE="approximately 22", Where P1 is crisp and P2, P3 are neutrosophic. Clearly, to answer this query the proposed neutrosophic rank search method is to be applied, because in addition to crisp search, both of α-neutrosophic equality search and neutrosophic proximity search will be used to answer this query. The truth value of the matching of the conjunction P of P1, P2 and P3 will be the product of the individual truth-values, (where it is needless to mention that for crisp match the truth –value will be exactly 1).There could be a multiple number of answers to this query, and the system will display all the results ordered or ranked according to the truth –values of p. It is obvious that the neutrosophic rank –search technique for predicate-matching reduces to a new type of fuzzy-search technique as a special case. neutrosophic sets.

VIII. Conclusion And Future Work

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 ([1], [2], [3], [5], [6], [8]) 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. When neutrosophic data are processed, their indeterminacies are processed as well and the consequent results are more meaningful. 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. This is a complete new Method of Answering Queries based on Neutrosophic logic.

As future work, I want to extend this paper to study Neutrosophic functional dependencies which constitute an important part of a good NRDM

References

[1] Bosc P.,Lietard I and Pivert O,"Evaluation of flexible queries: The quantified statement case".Technologies for constructing Intellegent Systems I, physica-Verlag Heidelberg New Yourk,pp 337-350 (2002).

[2] Chiang D., Chow L.R. and Heien N"fuzzy information in extended Fuzzy relational databases", fuzzy sets and systems 92, pp.1-10. (1997)

[3] Cohen,W., Integration of heterogeneous databases without common domains using queries based on textual similarity,proc. of SIGMOND,June 1998,pages 201-212.

[4] Haveliwala,T.,Gionis,A, Klein,D,and Indyk,P,Evaluatingstrategies for similarity search on the web , proceeding of WWW, Hawai, USA, may 2002.

[5] Kaushik S,and Nanda,H., Web based access of relational databases using fuzzy Natural Language Queries, International conference on cognitive Systems,Delhi,India, (1999). .

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

[7] Wang, H., Zhang, Y., Sunderraman, R. (2004). "Soft semantic web services agent", The Proc. Of NAFIPS 2004, pp. 126-129

[8] Yang Q.,Zhang W., Liu C., Wu J., Nakajima H., and Rishe N.D.EffecientProcessing of nested Fuzzy Database, IEEE Trans. On Knowledge and Data Eng., Vol.13,No.6,pp 884-901,Nov./Dec 2001