International Journal of Research in Engineering and Science (IJRES)
ISSN (Online): 2320-9364, ISSN (Print): 2320-9356
www.ijres.org Volume 6 Issue 5 Ver. I ǁ 2018 ǁ PP. 53-61

Anapproach for Processing Fuzzy Data Based On Hedge Algebras

Conghao Nguyen
Information Technology Faculty, Hue University, Vietnam
Corresponding Author: Conghao Nguyen

Abstract: *In the paper we introduce a new model of fuzzy relational databases in which uncertain data are represented as just linguistic terms with the hedge algebra-based semantics. That is these terms can be considered as elements of hedge algebras and their quantitative semantics are defined by semantically quantifying mappings of hedges algebras that are defined by the fuzziness measure of terms. These mappings assign each linguistic term of an attribute a real value in the real domain associated with the attribute, called the semantic representative or semantic value of this term. Based on fuzziness measure of terms, a system of fuzziness intervals of an attribute domain can be determined, from which we can construct k-neighbourhoods of linguistic terms and k-partition of the attribute domain, where k is the length of strings which represent linguistic terms of hedge algebras. This k-partition allows define k-matching relations on the attribute domain under consideration. We show that under the equivalence classes of this k-partition, which are intervals, and the semantic representatives of linguistic data in the database, we can manipulate the data in a unified way, because the queries related to linguistic data can be converted into the classical ones.*

Keywords: *fuzzy relational databases, hedge algebra, linguistic term, fuzziness measure.*

Date of Submission: 13-06-2018 Date of acceptance: 28-06-2018

I. INTRODUCTION

We are often faced with fuzzy or uncertain information in almost fields of our life. Therefore, it is natural that there were many works dealing with relational database models with uncertain information, normally called fuzzy database models. There are several approaches to such fuzzy databases to solve the problem of representation and semantic treatment of fuzzy information. They depend of course on the viewpoint, upon which one models the fuzzy data.

In the simplest case, the fuzziness of database model lies in the fact that some of the relations of the database are viewed as fuzzy sets on the cross product of their attribute domains, that is each tuple of the relations is associated with a membership degree taking values in the interval [0,1]. The handling of the data in such a database is rather simple, because the attribute values are still crisp ones, i.e. the attributes of the database are not allowed to take fuzzy data. To compare two values of an attribute domain *D*, a fuzzy relation RES is utilized, where RES is a reflexive and symmetric relation defined on the cross product of *D*, $\mu_{\mathrm{RES}}: D \times D \rightarrow [0,1]$, and called resemblance or proximity relation.

The original approach to fuzzy database is the one in which some attributes allow taking fuzzy sets as their values, which represent the meaning of vague terms, i.e. the uncertain information of the available knowledge about the elements of the attribute domain. For an attribute *A*, *A*(*x*) denotes the value of the object or individual *x* at the attribute *A*. In the case *A*(*x*) is an uncertain data, *A*(*x*) is a fuzzy set defined on the real domain of *A*. To capture the semantics of this kind of data, there are various approaches to utilize the semantics of fuzzy sets for defining the similarity between two fuzzy sets. For example, to compare two fuzzy sets *A* and *B* one can introduce different equality indexes, which indicate their similarity, or different comparability measures *Com*(*A*,*B*), which are also fuzzy sets but defined on [0,1] to measure the extent to which the one fuzzy set is compatible with the other or to measure the compatibility of the fact that "*A* is *B*" [13]. There are also several other compatibility measures established based on certain intuitive semantic idea for particular categories of fuzzy sets such as the one of triangular and/or trapezoidal fuzzy sets [14].

It may be observed that although these approaches achieve many successes in both theory and applications, there are still some inconveniences in data manipulation for these databases:

i) Normally, fuzzy data in such fuzzy databases provided by the database user are linguistic data which are considered as the labels of fuzzy sets or possibility distributions. As it can be seen above, there are many computational approaches to the representation of the meaning of linguistic information.

ii) Secondly, in fuzzy databases there are attributes, whose values are crisp as well as imprecise data, the data types of these attributes are not unique, e.g. their values may be real numbers or fuzzy sets or possibility distributions. In comparison with the classical database it causes many changes in data manipulation. The difficulty lies in the fact that how we can define matching relations such as =, ≠, ≤, ≥, <and > between values of completely different types in a fuzzy environment. For example, how we can understand the expression $t[A]$ $\theta s[A]$, where t and s are tuples of the database and $t[A]$ and $s[A]$ denote the values of different types of these tuples of the attribute A and θ is one of the matching relations =, ≠, ≤, ≥, <and >. Therefore, this problem may become simpler if we can give another approach in which we can treat all uncertain data as real data in the data manipulation.

Hedge algebras form an algebraic approach to the natural qualitative semantics of linguistic terms [7] and establish a quite new methodology to solve effectively approximate reasoning problems [9]. So, it may allow us introduce a new approach to fuzzy databases. In the algebraic approach, each term-domain $X = LDom(A)$ of the linguistic variable [6], or an attribute A of a database can be considered as an algebra in the category of universal algebras: $(LDom(A), G, C, H, \leq)$, where C is a set of constants, G is a set of primary terms considered as generators, X is freely generated from G by means of one-argument operations in H, a set of linguistic *hedges* or *modifiers* in question, and ≤ is a *semantics-based ordering relation* on X. Intuitively, we can observe that term-domains can be ordered based on the natural meaning of their elements and two terms x and y satisfy the inequality $x \leq y$ if the meaning of x is less than the one of y.

In this paper, we will introduce a notion of neighbourhoods of the representative of a term x, called also neighbourhoods of the term x. They are subintervals of [0,1] that contain the representative of x considered as their topologically internal point and they can be defined based on the fuzziness measure of terms. It is interesting that this provides an ability to define fuzzy matching relations, called matching relations of degree k, where k is the length of a term, and to translate queries related with linguistic data into traditional ones. The paper is organized as follows: In the second section, some results of hedge algebras are introduced. In the section 3, a relational model of databases with linguistic data is proposed. Some conclusions will be given in the end of the paper.

II. FUNDAMENTAL CONCEPTS

In this section, we will shown that the meaning of linguistic terms can be expressed by utilizing the structure of hedge algebras or, in the other words, each terms-domain which can be ordered by the semantics of terms becomes a hedge algebra. Since in databases the attribute domains are linearly ordered sets, we shall restrict our presentation to linear hedge algebras.

2.1. Qualitative semantics of linguistic terms

It has been seen above the meaning of terms of a linguistic variable X can be expressed through an ordering relation on its terms-domain. In other words, the meaning of a term is revealed through the algebraic structure related to this term. This viewpoint is similar to the observation that the meaning of the truth value "totally true" in the classical logics is interpreted not only by 1 itself, but by 1 in the context of 0 defined by the order-based structure of a two-element Boolean algebra. Therefore, the meaning of a term in a terms-domain X of a linguistic variable X can be represented by its ordering relationships with the remaining ones in this domain, i.e. by an algebraic structure of terms-domains. In [6-9], it is shown that each term-domain X can be regarded as an abstract algebra AX = $(X, G, C, H, \leq)$, where $C = \{0, W, 1\}$ is a set of constants, $X = H(G) \cup C$ with $H(G)$ being freely generated from the set of generators G by means of one-argument operations in H, a set of linguistic hedges or modifiers, and ≤ is a *semantics-based ordering relation* on X. Such algebras are called hedge algebras. In the case that the sets H^-, H^+ and G are linearly ordered, where H^- and H^+ are, respectively, the sets of negative and positive hedges and $H = H^- \cup H^+$, AX is called a linear hedge algebra. For instance, the domain of the linguistic variable *TRUTH*: T = {*0, W, 1, true, false, very true, more true, rather true, very false, rather false, less false,...*} can be considered as a linear hedge algebra AT = $(T, G, C, H, \leq)$ with G = {*true*, *false*}, H^+ = {*very*, *more*}, H^- = {*rather*, *less*} and ≤ is a relation induced by the natural meaning of terms in X, e.g. we have *1>very true>true*, *more true>true>rather true>less true, ... >W>less false>rather false>false>very false*, … *>0* …

2.2. Fuzziness measures and semantically quantifying mappings

Consider a complete linear HA $\underline{AX} = (\underline{X}, G, C, H, \Phi, \Sigma, \leq)$, which is a completion of AX = $(Dom(X), G, C, H, \leq)$, where $\underline{X} = H_e(G)$, $H_e = H \cup \{\Phi, \Sigma\}$ and Σ, Φ are two artificial hedges the meaning of which is defined, respectively, by the supremum (denoted by sup, for short) and infimum (denoted by inf) of the set $H(x)$ in the poset $(\underline{X}, \leq)$. The word "completion" means that it is necessary to complete certain elements in the original hedge algebra AX so that the operations Σ and Φ will be defined for all $x \in \underline{X}$. Set $Lim(X) = \underline{X} \setminus H(G)$, elements of which are called *limit elements* of $\underline{AX}$.

Definition 2.1[8]. A Comp-HAs$\underline{AX}$ = ($\underline{X}$, $G, C, H, \Sigma, \Phi, \leq$)is said to be a linear hedge algebra(Lin-HA, for short) if the sets $G \cup C = \{0, c^-, W, c^+, 1\}$, $H^+ = \{h_1,..., h_p\}$and $H^- = \{h_{-1}, ..., h_{-q}\}$are linearly ordered with $h_1<...<h_p$and $h_{-1}< ...< h_{-q}$,where $p, q >1$. Note that $H = H^- \cup H^+$.

Let Xbe a linguistic variable and $\underline{AX}$ = ($\underline{X}$, $G, C, H, \Sigma, \Phi, \leq$) be a ComHA which models a linguistic domain *Dom*(X) of the variable X. Assume that this algebra is *free*, that is $hx \neq x$, for every $x \in H(G)$ and $h \in H$. Semantically, for each term x, the set $H(x)$ consists of all terms whose meaning originates from a definitive essential meaning of the term x. For example, consider two terms x = *RatherTrue* and y = *ApproximatelyTrue*. The term *VeryRatherTrue* reflects a certain meaning of *RatherTrue* but not of *AppTrue*, while the term *VeryAppTrue* reflects a definitive meaning of *AppTrue*, but not of *RatherTrue*. In addition, it can be observed that a term x is vague if and only if its meaning is still changed by using hedges. For example, *True* is vague, but *Absolutely True* is not and hence H(*Absolutely True*) = {*Absolutely True*}. Therefore, we may consider the set $H(x)$ as an expression of certain essential characteristics of fuzziness of the term x and we may use it to model the fuzziness of the linguistic term, qualitatively.

In connection with measuring the fuzziness of linguistic terms in the algebraic approach, we need study the problem of quantification of hedge algebras. Similarly, as the fuzzy defuzzification which assigns a real number of the reference domain to each fuzzy set, we introduce a notion of semantically quantifying mappings (SQMs) of hedge algebras. Fuzziness measure of terms and hedges are very difficult to define in the framework of the fuzzy sets theory. But hedge algebras provide a good intuition and good mathematical basic to define these notions in a reasonable way. Moreover, as it can be seen below, there is a close relation between SQMs and fuzziness measure of hedge algebras.

Definition 2.2[8].Given a Lin-HA $\underline{AX}$ = ($\underline{X}$, $G, C, H, \Sigma, \Phi, \leq$). A mapping $f: \underline{X} \to [0,1]$ is called a semantically quantifying mapping (SQM) of $\underline{AX}$ if it satisfies the following conditions:

1)f is a one-to-one mapping;

2)$f$preserves the ordering relation on $\underline{X}$, i.e. $x < y \Rightarrow f(x) < f(y)$, and $f(0) = 0$, $f(1) = 1$;

3)f is continuous in the sense that for $\forall x \in \underline{X}, f(\Phi x) = inf\ f(H(x))$ and $f(\Sigma x) = sup\ f(H(x))$.

Note that Σx = *supremum* $H(x)$ and Φx = *infimum* $H(x)$, where the supremum and infimum are taken in the linearly ordered set $\underline{X}$.

Based on the structure of *free*Lin-HAs, it is observed that the "size" of $H(x)$ can model the fuzziness degree of the term x, since it consists of terms which have a meaning originating from that of x. If x is an exact concept, then it is a fixed point and hence $H(x) = \{x\}$, a single element – a minimal set among the sets $H(x)$, $\forall x \in X$. And, if $x = hu$ and $y = kv$, $\forall h, k \in H$, $h \neq k$,we always have $H(hu) \subseteq H(x)$ and $H(hu) \cap H(ku) = \emptyset$, that is that if x and y originate from two terms having essentially different meaning then their fuzziness models can be consider as independent events. In addition, the equality $H(x) = \bigcup_{k \in H} H(kx)$ indicates that the fuzziness model of x equals the union of the fuzziness models of the more specific and "independent" terms originated from x. So, as a consequence, for a given SQM f, the diameter of the image $f(H(x)) \subseteq [0,1]$ can be interpreted as fuzziness measure of the term x. Therefore, we can introduce the following definition which can be adopted as an axiomatization of the fuzziness measure:

Definition 2.3 [8]. Given a *free*Lin-HA $\underline{AX}$ = ($\underline{X}$, $G, C, H, \Sigma, \Phi, \leq$). An fm: $\underline{X} \to [0,1]$ is said to be a fuzziness measure (FM, for short) of terms in $\underline{X}$ if:

1)$fm(c^-) + fm(c^+) = 1$ and for $\forall u \in \underline{X}$, $\sum_{h \in H} fm(hu) = fm(u)$; In this case fm is called complete;

2) If $H(x) = \{x\}$, then $fm(x) = 0$. Especially, $fm(0) = fm(W) = fm(1) = 0$;

3)$\forall x, y \in \underline{X}$, $\forall h \in H$, $\frac{fm(hx)}{fm(x)} = \frac{fm(hy)}{fm(y)}$, that is this proportion does not depend on specific elements and, hence, it is called the fuzziness measure of the hedge h and denoted by $\mu(h)$.It is easy to show the following proposition:

Proposition 2.1 [8].For fm and $\mu(h)$ defined in Definition. 2.3, the following statements hold:

1) $fm(hx) = \mu(h) fm(x)$, $\forall x \in \underline{X}$ *and* $fm(x) = 0$, *for all* $x \in Lim(\underline{X})$;

2) $fm(c^-) + fm(c^+) = 1$;

3) $\sum_{-q \leq i \leq p, i \neq 0} fm(h_i c) = fm(c)$, *where* $c \in \{c^-, c^+\}$;

4) $\sum_{-q \leq i \leq p, i \neq 0} fm(h_i x) = fm(x), x \in X$;

5) $\sum\{\mu(h_i) : -q \leq i \leq -1\} = \alpha$ *and* $\sum\{\mu(h_i) : 1 \leq i \leq p\} = \beta$, *where* $\alpha, \beta > 0$ *and* $\alpha + \beta = 1$.

2.3. A family of fuzziness intervals and its structure

The following notion will be useful for studying the structure of a family of fuzziness intervals and for defining a specific family of SQMs induced by the fuzziness measure *fm*. It is interesting that we can define the *sign* of elements of hedge algebras, based on the positiveness and negativeness of primary terms and PN-property of hedges:

Definition 2.4 (*Sign function*) $Sign$:$\underline{\boldsymbol{X}} \to \{-1, 0, 1\}$is a function which is defined recursively as follows, where h, $h' \in \boldsymbol{H}$, $c \in \{c^-, c^+\}$:
1) $Sign(c^-) = -1$, $Sign(c^+) = +1$,
2) $Sign(h'hx) = Sign(h'hx)$, if $h'hx = hx$ and otherwise,
$Sign(h'hx) = -Sign(hx)$, if $h'hx \neq hx$ and h' is negative w.r.t. h (or w.r.t. c, when $h = I$ and $x = c$);
$Sign(h'hx) = +Sign(hx)$, if $h'hx \neq hx$ and h' is positive w.r.t. h (or w.r.t. c, when $h = I$ and $x = c$).
3) $Sign(x) = 0$, for all limit elements $x \in \underline{X} \setminus X$, where $X = H(G)$.
Based on the fuzziness measure above, we define a notion of *fuzziness intervals* of terms that model an aspect of quantitative term meaning. For every term x, the fuzziness interval of x is a subinterval of [0,1] of length $fm(x)$, denoted by $\mathfrak{I}_{fm}(x)$, which is defined by induction on the length of x as follows:
i) For terms x of length 1, i.e. $x \in \{c^+, c^-\}$, $\mathfrak{I}_{fm}(c^-)$ and $\mathfrak{I}_{fm}(c^+)$ are defined so that they constitute a partition of [0,1] and satisfy the conditions that $\mathfrak{I}_{fm}(c^-) \leq \mathfrak{I}_{fm}(c^+)$, i.e. their order is induced by that between c^+ and c^-, $|\mathfrak{I}_{fm}(c^-)| = fm(c^-)$ and $|\mathfrak{I}_{fm}(c^+)| = fm(c^+)$, where $|\mathfrak{I}(x)|$ denotes the length of $\mathfrak{I}(x)$. Here, the notation $U \leq V$ means that, for $\forall x \in U$, $\forall y \in V$, we have $x \leq y$.
ii) Suppose that $\mathfrak{I}_{fm}(x)$ has been defined and $|\mathfrak{I}_{fm}(x)| = fm(x)$, for all x of length k. Then, $\{\mathfrak{I}_{fm}(h_ix): i \in [-q \wedge p]\}$ is constructed so that it is a partition of $\mathfrak{I}_{fm}(x)$ and satisfies the conditions that $|\mathfrak{I}_{fm}(h_ix)| = fm(h_ix)$ and $\{\mathfrak{I}_{fm}(h_ix): i \in [-q \wedge p]\}$ is a linearly ordered set, whose order is induced by that of $\{h_{-q}x, h_{-q+1}x, ..., h_px\}$, i.e. if, for example, $h_{-q}x > h_{-q+1}x > ... > h_px$, then $\mathfrak{I}_{fm}(h_{-q}x) \geq \mathfrak{I}_{fm}(h_{-q+1}x) \geq ... \geq \mathfrak{I}_{fm}(h_px)$ (see Figure 2.2).
Assume from now on that the fuzziness intervals of linguistic terms always contain their right end-point, e.g. by this convention we have $\mathfrak{I}_{fm}(c^-) = [0, fm(c^-)]$ and $\mathfrak{I}_{fm}(c^+) = (fm(c^+), 1]$.
If $l(x) = k$, then the fuzziness interval $\mathfrak{I}_{fm}(x)$ of x is said to be of depth k and, if necessary and the index fm is understood, we use the notation $\mathfrak{I}_k(x)$ to indicate explicitly that this fuzziness interval is of depth k.
Set $X_k = \{x \in X : l(x) = k\}$. Obviously, $X = \bigcup_{1 \leq k < \infty} X_k$.
Put $\mathrm{I}_k = \{\mathfrak{I}_k(x): x \in X_k\}$, the set of all fuzziness intervals of depth k (or fuzziness k-intervals for short), for a given positive integer k, and put $\mathrm{I} = \{\mathfrak{I}(x): x \in X\} = \bigcup_{1 \leq k < \infty} \mathrm{I}_k$, the set of all fuzziness intervals of a hedge algebra $\underline{AX}$.
The following proposition describes the structure of the family of the fuzziness intervals which serves as a basis for studying the similarity of data in an attribute domain (refer to Figure 2.1).

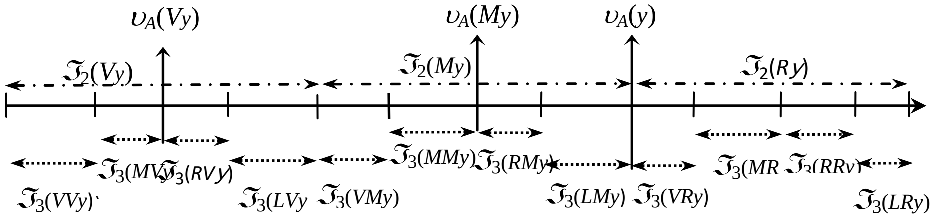

Figure 2.1

Definition 2.5 [7]. Let $\underline{AX} = (X, G, C, H, \Sigma, \Phi, \leq)$ be a *free* complete Lin-HA and $fm(c^-)$, $fm(c^+)$ and $\mu(h)$ be fuzziness measures of the primary terms c^-, c^+ and hedge $h \in H$, respectively, which satisfy 2) and 5) of Proposition 2.1. Then, the mapping υ induced by the fuzziness measure fm is defined recursively as follows:
1) $\upsilon(W) = \kappa = fm(c^-)$, $\upsilon(c^-) = \kappa - \alpha fm(c^-) = \beta fm(c^-)$, $\upsilon(c^+) = \kappa + \alpha fm(c^+)$;
2) $\upsilon(h_jx) = \upsilon(x) + Sgn(h_jx)\{\sum_{i=Sgn(j)}^{j} \mu(h_i)\, fm(x) - \omega(h_jx)\mu(h_j)\, fm(x)\}$, where

$\omega(h_jx) = \frac{1}{2}[1 + Sgn(h_jx)Sgn(h_ph_jx)(\beta - \alpha)] \in \{\alpha, \beta\}$, *for all* $j \in [-q \wedge p]$;

3) $\upsilon(\Phi c^-) = 0$, $\upsilon(\Sigma c^-) = \kappa = \upsilon(\Phi c^+)$, $\upsilon(\Sigma c^+) = 1$, and for all $j \in [-q \wedge p]$, we have:

$\upsilon(\Phi h_jx) = \upsilon(x) + Sgn(h_jx)\{\sum_{i=sign(j)}^{j-sign(j)\frac{1+Sign(h_jx)}{2}} \mu(h_i)\, fm(x)\}$ and

$\upsilon(\Sigma h_jx) = \upsilon(x) + Sgn(h_jx)\{\sum_{i=sign(j)}^{j-sign(j)\frac{1-Sign(h_jx)}{2}} \mu(h_i)\, fm(x)\}$.

It will be seen that there is a close relationship between this SQM and the fuzziness intervals of a hedge algebra which is shown in the following theorem:

Theorem 2.1 [7].*Let* $\underline{\mathrm{AX}} = (X, G, H, \Sigma, \Phi, \leq)$ *be a free complete Lin-HA and* υ *be defined as in Definition 2.5. Then,* υ *is an SQM and* $\upsilon(H(x))$ *is dense in the interval* $[\upsilon(\Phi x), \upsilon(\Sigma x)], \forall x \in X$. *Moreover,* $\upsilon(\Phi x) = \inf \upsilon(H(x)), \upsilon(\Sigma x) = \sup \upsilon(H(x))$ *and* $fm(x) = \upsilon(\Sigma x) - \upsilon(\Phi x)$, *and hence* $fm(x) = d(\upsilon(H(x)))$, *whered*$(A)$ *denotes the diameter of* $A \subseteq [0,1]$. *As a consequence,*$\upsilon[H(G)]$ *is dense in* $[0,1]$.

Example 2.1. Let us consider a linear hedge algebra of *AGE*, $\underline{\mathrm{AX}} = (X, G, C, H, \Sigma, \Phi, \leq)$, where $G = \{young, old\}$, $H^- = \{P, L\}$ and $H^+ = \{M, V\}$, where P, L, M and V stand for *Rather*, *Less*, *More* and *Very*, respectively. Suppose that $D_A = [0, 120]$, $fm(old) = 0.55$, $fm(young) = 0.45$, $\mu(R) = 0.32$, $\mu(L) = 0.20$, $\mu(M) = 0.30$ and $\mu(V) = 0.18$. So, $\alpha = 0.52$ and $\beta = 0.48$. Since, in applications instead of the SQM taking values in [0,1], it should takes values in [0, 120], we set $\upsilon_{A,r}(x) = \upsilon_A(x) \times 120$, where the index r means that it is a value of the “real” domain in question. So, by Definition 2.5, we have:

$\upsilon_{A,r}(young) = (0.45 - 0.45 \times 0.52) \times 120 = 0.234 \times 120 = 25.92$

$\upsilon_{A,r}(Ryoung) = 25.92 + 0.48 \times 0.32 \times 0.45 \times 120 = 34.2144$

$\upsilon_{A,r}(MRyoung) = 34.2144 + (-1) \times [1 - \beta] \times 0.30 \times 0.32 \times 0.45 \times 120 = 31.51872$

$\upsilon_{A,r}(VRyoung) = 34.2144 + (-1) \times [0.30 + 0.18 - 0.18 \times \beta] \times 0.32 \times 0.45 \times 120 = 27.412992$

$\upsilon_{A,r}(Vyoung) = 25.92 + (-1)[0.30 + 0.18 - 0.48 \times 0.18] \times 0.45 \times 120 = 4.6656$

$\upsilon_{A,r}(\Sigma Vyoung) = 25.92 + (-1) \times 0.30 \times 0.45 \times 120 = 9.72$

$\upsilon_{A,r}(\Phi Lyoung) = 25.92 + (+1) \times 0.32 \times 0.45 \times 120 = 42.2$

III. A RELATIONAL MODEL OF DATABASES WITH LINGUISTIC DATA

As usual, a relational database model is a set $DB = \{U, R_1, R_2, \ldots, R_m; Const\}$, where $U = \{A_1, A_2, \ldots, A_n\}$ is the universe of attributes, R_i is a relation scheme, which is a subset of U, *Const* is a set of constraints on data in database. Each A_j will be associated with a set D_{Aj}, called the domain of A_j.

In a relational database with linguistic data, if an attribute A_i is allowed to take linguistic values, it can be regarded as a linguistic variable, denoted also by A_i, whose reference set is the real domain D_{Ai} of A_i. So, the domain of such an attribute A_i consists of two parts, a real domain D_{Ai} and a terms-set $X = LDom(A_i)$, a linguistic domain of the linguistic variable A_i. Such attributes are called *linguistic attributes*. Set $\mathrm{D}(A_i) = D_{Ai} \cup LDom(A_i)$. Here, assume that for A_i not being linguistic attribute, $LDom(A_i) = \emptyset$. Elements of D_{Ai} will be denoted by $a, b, c, \ldots$ and elements of $LDom(A_i)$ will be denoted by $x, y, z, u, v \ldots$

As usual, a tuple t on$U$ is a mapping t: $U \to \mathrm{D}(A_1) \cup \ldots \cup \mathrm{D}(A_n)$ such that $t(A_i) \in \mathrm{D}(A_i)$, for $1 \leq i \leq n$. Tuples will be denoted by t, s with indexes, if necessary. By $t[A_i]$ we mean the value of a tuple t at the attributeA_i. For any subset X of U, $t[X]$ denotes the restriction of the mapping t on$X$ and it is called a tuple on X.

Consider a relation scheme, i.e. a subset R of U. An instance of R is a set of distinct tuples on R and called a relation. Relations of R are denoted by $r[R]$, $s[R]$ ... If R is understood and there will be no confusion, R may be omitted in these notations. Because t may take real data as well as linguistic ones, it is necessary to construct a method for handling the data in database based on their semantics. If we interpret linguistic data as labels of fuzzy sets then we can manipulate data based on fuzzy sets theory and we have a concept of fuzzy databases.

In this paper, linguistic domain $LDom(A_i)$ is assumed to be a subset of the underlying set of a complete linear hedge algebra $\underline{\mathrm{AX}}_{Ai} = (X, G, H, \Sigma, \Phi, \leq)$ of the linguistic variable A_i. It raises a question: is there a method to manage the semantics of data based on hedge algebras? If the answer is positive, we may consider linguistic values in databases as elements of hedge algebras, whose underlying set is a term-set. From this viewpoint, the above database model is called a *relational database model with linguistic data*.

3.1. HA-based semantics of linguistic data and semantics-based topology

Let us consider a linguistic attribute A and suppose as above that $\mathrm{D}(A) = LDom(A) \cup D_A$ is a mixed domain of A. The question is that how we can define a similarity of data in the mixed domain $\mathrm{D}(A)$?

In traditional way, each linguistic value is interpreted as the label of a fuzzy set on D_A. It is well known that under such fuzzy data representation, manipulation of data is much more difficult than in the case of crisp data. Especially, it is not easy to define data similarity in a unified way, since its data types are not unique and their semantics is completely different.

We shall solve this question in such a way that we can manipulate the data in databases simply and in a unified way. Firstly, to unify the data types, we embed $LDom(A)$ into the underlying set of a Lin-HA $\underline{\mathrm{AX}}_A = (X, G, H, \Sigma, \Phi, \leq)$ and utilize a quantitative semantic mapping $\upsilon_{A,r}$ associated with the variable A to transforms linguistic data to real data, i.e. $\upsilon_{A,r} : X \to D_A$, where the index r indicates as above that $\upsilon_{A,r}$ takes values in the real domain D_A. So, each linguistic datum x of A can be regarded as a label of a real value defined by $\upsilon_{A,r}$. Since $\upsilon_{A,r}(x) \in D_A$, we may establish a method for manipulating the data of real type or linguistic type in a unified way.

However, the semantics of $\upsilon_{A,r}(x)$ is completely different from the usual real data, because the real value $\upsilon_{A,r}(x)$ is only a representative of the term x that depends on the fuzziness parameters. So, we can not handle by simply using $\upsilon_{A,r}(x)$ instead of x in comparison with real values or linguistic values of A, but we should handle it in another way, utilizing the fuzziness intervals of x.

Let a terms-domain X of a linguistic attribute A be given and X_k be the set of all term x of the length k. Let us denote by x_{0k} the least element in X_k. Then, $\upsilon_A(\Phi x_{0k}) = 0$ and according to Theorem 2.1 and the definition of the fuzziness intervals, we have $\mathfrak{I}_k(x_{0k}) = [\upsilon_A(\Phi x_{0k}), \upsilon_A(\Sigma x_{0k})]$ and $\mathfrak{I}_k(x) = (\upsilon_A(\Phi x), \upsilon_A(\Sigma x)]$, for all $x \in X_k$ and $x \neq x_{0k}$ with a note, by our convention, that the fuzziness intervals are closed at their right end-point. If we denote by λ_k the greatest length of intervals in $I_k = \{\mathfrak{I}_k(x): x \in X_k)$ and by η the greatest fuzziness measure of hedges in H, then by (1), Proposition 2.1 we have $\lambda_{k+1} \leq \eta\lambda_k \leq \eta^k\lambda_1$. Since $\eta < 1$, it follows that we can find a fuzziness interval of x as small as we desire.

Let $\underline{AX}$ be a Lin-HA with $H^+ = \{h_1,...,h_p\}$and $H^- = \{h_{-1}, ..., h_{-q}\}$ being linearly ordered sets with $h_1<...<h_p$and $h_{-1}< ...< h_{-q}$,where p, q >1. In applications, we usually assume $p, q \leq 3$. Denote by $H_1 = \{h_i, h_{-j} : 1 \leq i \leq [p/2]$ & $1 \leq j \leq [q/2]\}$ the set of hedges that generate terms which have a meaning near the meaning of the original x and by $H_2 = \{h_i, h_{-j} : [p/2] < i \leq p$ and $[q/2] < j \leq q\}$ the set of hedges that generate terms which have meaning far from x. This partition of hedges has only a technical value to serve our definition of basic open sets of certain depth k. First of all, we need some notions and notations.

Let INT_k be a set of fuzziness intervals of depth k. Two fuzziness intervals of depth k, $\mathfrak{I}_k(x)$ and $\mathfrak{I}_k(y)$, is said to be connected in INT_k if there is a chain of consecutive fuzziness intervals in INT_k such that it connects $\mathfrak{I}_k(x)$ and $\mathfrak{I}_k(y)$. In this case $\mathfrak{I}_k(x)$ is also said to be connected in INT_k to a point $a$in $\mathfrak{I}_k(y)$. The set INT_k is said to be connected if any two of its intervals are connected and, otherwise, it is said to be disconnected.

This notion of connectivity defines the so-called *connectivity relation* on INT_k and it is an equivalence relation. Set $INT_k(H_n) = \{\mathfrak{I}_k(h_iy): y \in X_{k-1}, h_i \in H_n\}$ and $INT_k(H_n, x) = \{\mathfrak{I}_k(h_ix): l(x) = k - 1, h_i \in H_n\}$, where $n = 1, 2$. By the definition of H_1, it can easily be verified that $INT_k(H_1, x)$ is *connected*. Note that the point $\upsilon_A(x)$ is the unique common end-point of the intervals $\mathfrak{I}_k(h_1x)$ and $\mathfrak{I}_k(h_{-1}x)$. Hence, every fuzziness interval in $INT_k(H_1, x)$ is connected to $\upsilon_A(x)$, that is they lie around the point $\upsilon_A(x)$. From this observation, for every term $x \in H(G)$, $l(x) = k - 1$, we put $O_{k,H1}(x) = \bigcup\{\mathfrak{I}_k(y) : l(y) = k, \mathfrak{I}_k(y)$ is connected to $\upsilon_A(x)$ in $INT_k(H_1, x)\} = \bigcup_{hi \in H1}\mathfrak{I}_k(h_ix)$.

For illustration, it can be seen in Figure 2.1 that $O_{2,H1}(y) = \mathfrak{I}_2(My) \cup \mathfrak{I}_2(Py)$ and $O_{3,H1}(My) = \mathfrak{I}_3(MMy) \cup \mathfrak{I}_3(PMy)$, where $H_1 = \{M, R\}$ and y is an abbreviation of "*young*".

In the contrast, $INT_k(H_2, x)$, where $H_2 = \{L, V\}$, is disconnected and consists of the fuzziness intervals which can be considered as lying far from the point $\upsilon_A(x)$. For example, in Figure 2.1, $\mathfrak{I}_3(VMy)$ and $\mathfrak{I}_3(LMy)$ belonging to $INT_3(H_2, My)$ are disconnected in $INT_3(H_2)$ and lie far from $\upsilon_A(My)$, but $\mathfrak{I}_3(LMy)$ and $\mathfrak{I}_3(VPy)$ are connected in $INT_3(H_2)$ and lie around the value $\upsilon_A(y)$, since they are connected to $\upsilon_A(y)$ in $INT_3(H_2)$. Note that $l(y) = 1 < 3$.

Now, we are ready to define a system of neighbourhoods of a linguistic term. Remember that for $x = h_{k-1}... h_1c$, $x|_i = h_{i-1}... h_1c$.

Definition 3.1.For each term $x = h_{k-1}... h_1c$ of length k, where $c \in G$, a basic semantic neighbourhoods system of x up to a depth l, $l \geq k$, under the mapping υ_A, denoted by $Neig^l_{\upsilon}(x)$, is the set of the following intervals $\theta_i(x)$:

1) $\theta_i(x) = \mathfrak{I}_i(x|_i) = (\upsilon_A(\Phi x|_i), \upsilon_A(\Sigma x|_i)]$, $k \geq i \geq 1$;
2) $\theta_{k+1}(x) = O_{(k+1),H1}(x)$, $i = k + 1$;
3) $\theta_j(x)$, $l \geq j > k+1$, is defined as follows: $\theta_j(x) = \bigcup\{\mathfrak{I}_j(y) : l(y) = j, \mathfrak{I}_j(y)$ is connected to $\upsilon_A(x)$ in $INT_j(H_2)\}$.

As an example, from the structure of fuzziness intervals given in Figure 2.1, we have

$$\theta_2(y) = O_{2,H1}(y) = \mathfrak{I}_2(My) \cup \mathfrak{I}_2(Py)$$

$$\theta_3(y) = \bigcup\{\mathfrak{I}_3(y) : l(y) = 3, \mathfrak{I}_3(y) \text{ is connected to } \upsilon_A(y) \text{ in } INT_3(H_2, y)\}$$
$$= \mathfrak{I}_3(LMy) \cup \mathfrak{I}_3(VPy)$$

$$Neig^3_{\upsilon}(y) = \{\mathfrak{I}_1(y), \mathfrak{I}_2(My) \cup \mathfrak{I}_2(Py), \mathfrak{I}_3(LMy) \cup \mathfrak{I}_3(VPy)\},$$

$$Neig^3_{\upsilon}(Vy) = \{\mathfrak{I}_1(y), \mathfrak{I}_2(Vy), \mathfrak{I}_3(MVy) \cup \mathfrak{I}_3(PVy)\}.$$

It is worth emphasizing that $\upsilon_A(x)$ is always an internal point of every neighbourhood in $Neig^l_{\upsilon}(x)$.

Example 3.1.Let us consider a linear hedge algebra of AGE, $\underline{AX} = (X, G, C, H, \Sigma, \Phi, \leq)$, where $G = \{young, old\}$, $H^- = \{P, L\}$ and $H^+ = \{M, V\}$, where R, L, M and V stand for *Rather*, *Little*, *More* and *Very*, respectively. Set$D_A = [0, 120]$, $fm(old) = 0.55$, $fm(young) = 0.45$, $\mu(R) = 0.32$, $\mu(L) = 0.20$, $\mu(M) = 0.30$ and $\mu(V) = 0.18$. So, $\alpha = 0.52$. Note that $H_1 = \{R, M\}$ and $H_2 = \{L, V\}$.

1) Basic semantic neighbourhoods of *young*: By the definition, the basic semantic neighbourhood system of depth 1 of *young*, $Neig^1_{\upsilon}(young)$, consists of a unique interval $\mathfrak{I}_1(young) = [\upsilon_{A,r}(\Phi young), \upsilon_{A,r}(\Sigma young)] = [0, fm(young) \times 120] = [0, 54.00]$, where the subscript r indicates, as previously, that the mapping $\upsilon_{A,r}$ is defined in the *real* domain D_A.

The basic semantic neighbourhoods system of depth 2 of *young*, $Neig^2_{\upsilon}(young)$, consists of the intervals $\mathfrak{I}_1(young) = [0,54.00]$ and $\theta_{2,r}(young) = \Omega_{2,H1,r}(young) = (\upsilon_{A,r}(\Phi My), \upsilon_{A,r}(\Sigma My)] \cup (\upsilon_{A,r}(\Phi Ry), \upsilon_A(\Sigma Ry)] =$

$(\upsilon_{A,r}(\Phi My), \upsilon_{A,r}(\Sigma Ry)] = (\upsilon_{A,r}(y) - 120 \times fm(My),\ \upsilon_{A,r}(y) + 120 \times fm(Ry)] = (25.92 - 0.30 \times 0.45 \times 120,\ 25.92 + 0.32 \times 0.45 \times 120) = (9.72, 43.20]$, where *y* stands for *young*, for short.
The neighbourhood of depth 3 of *young* under $\upsilon_{A,r}$ is the interval $\theta_{3,r}(young) = \mathfrak{I}_3(LMy) \cup \mathfrak{I}_3(VPy) = (\upsilon_{A,r}(\Phi LMy), \upsilon_{A,r}(\Sigma LMy)] \cup (\upsilon_{A,r}(\Phi VRy), \upsilon_{A,r}(\Sigma VRy)] = (\upsilon_{A,r}(\Phi LMy), \upsilon_{A,r}(\Sigma Vy)] = (\upsilon_{A,r}(y) - 120 \times fm(LMy),\ \upsilon_{A,r}(y) + 120 \times fm(VRy)] = (25.92 - 0.20 \times 0.30 \times 0.45 \times 120,\ 25.92 + 0.18 \times 0.32 \times 0.45 \times 120] = (22.68, 29.0304]$, since $L, V \in H_2$.
2) Semantic neighbourhoods of *Rather young* (*Ry*):we have $\upsilon_{A,r}(Ry) = 25.92 + 0.48 \times 0.32 \times 0.45 \times 120 = 34.2144$. Then, by Definition 3.1, we observe that$Neig^2{}_{\upsilon}(Ry)$ consists of a unique interval $[\upsilon_{A,r}(\Phi Ry), \upsilon_{A,r}(\Sigma Ry)] = (25.92,\ 25.92 + 0.32 \times 0.45 \times 120] = (25.92, 43.2]$;
$Neig^3{}_{\upsilon}(Ry)$ consists of the neighbourhoods in $Neig^2{}_{\upsilon}(Ry)$, $\theta_{2,r}(Ry)$, and the following ones with a notice that the length of *Ry* is 2: $\theta_{3,r}(Ry) = O_{3,H1}(Ry) = (\upsilon_{A,r}(Ry) - fm(MRy) \times 120,\ \upsilon_{A,r}(Ry) + fm(RRy) \times 120] = (34.2144 - 0.30 \times 0.32 \times 0.45 \times 120,\ 34.2144 + 0.32 \times 0.32 \times 0.45 \times 120] = (29.0304, 39.744]$.

3.2. Manipulation of linguistic data semantics

Normally, the mathematical foundation for data manipulation is the relational algebra, whose important operations such as "select", "join", … , are defined based on the evaluation of matching relations $=, \neq, \leq, \geq, <$ and $>$ on mixed domains of the linguistic attributes. We shall try to solve this question in such a way that we can manage the data in databases in a most advantageous way which utilizes the new data semantics.
Let *t* and *s* be two tuples defined on *U*. Each linguistic attribute A_i will be equipped with a quantitative semantic mapping $\upsilon_{Ai} : LDom(A_i) \rightarrow D_{Ai}$. The first question is how can we define matching relations on $D(A_i) = D_{Ai} \cup LDom(A_i)$, which contains linguistic data as well as crisp data? First off all, we will define the notion of "equality" on $D(A_i)$. In a fuzzy environment, there is a degree of equality, called a similarity between data. In the algebraic approach, we will introduce a notion of similarity of degree *k*, where *k* is the length of terms, based on the notion of neighbourhoods of depth *k* examined above.
Let *X* be a set of linguistic terms of an attribute *A*, which is considered as a subset of *H*(*G*) of a Lin-HA. We will define a *k*-partition of the unit interval [0,1], based on the neighbourhoods of depth *k* of terms in *X* defined by Definition 3.1.

Let, for each *k*, a set of *k*-intervals associated with *x*, $\forall x \in X$, where *k* indicates that they are constructed by utilizing fuzziness intervals of depth *k*, is defined as follows:
Let us consider the sets H_1, H_2 and the sets $INT_k(H_n) = \{\mathfrak{I}_k(h_i y): y \in X_{k-1}, h_i \in H_n\}$, $n = 1, 2$, as mentioned in Section 3.1. It is clear that $INT_k(H_1) \cap INT_k(H_2) = \varnothing$ and $INT_k(H_1) \cup INT_k(H_2) = J_k$, the set of all *k*-fuzziness intervals. The connectivity relation on $INT_k(H_n)$ partitions the set $INT_k(H_n)$ into the connectivity components. Intuitively, each component defines a similarity of degree *k* between its real values. This suggests us to introduce the following

Definition 3.2.Each connectivity component *C* of $INT_k(H_n)$, for $n = 1, 2$, determines an interval $S_k = \bigcup\{\mathfrak{I}_k: \mathfrak{I}_k \in C\}$.It is called a *k-similarity interval* of *x*, for any $x \in X$ such that$\upsilon_{Ai}(x) \in S_k$, and denoted by $S_k(x)$.
This definition is correct, since the *k*-fuzziness intervals in *C* are consecutive and, therefore, they constitute a subinterval of [0,1].
To exemplify this notion, consider the structure of the fuzziness intervals given in Figure 2.1. For $k = 3$, it can be seen for example that $\{\mathfrak{I}_3(LMy), \mathfrak{I}_3(VRy)\}$, $\{\mathfrak{I}_3(MMy), \mathfrak{I}_3(RMy)\}$, $\{\mathfrak{I}_3(LVy), \mathfrak{I}_3(VMy)\}$, $\{\mathfrak{I}_3(VVy)\}$ are connectivity components of $INT_3(H_1)$ and $INT_3(H_2)$.

Definition 3.3.Let a Lin-HA $\underline{AX}$ and its fuzziness measure *fm* be given. Suppose that υ_{Ai} is the SQM induced by the fuzziness measure *fm*. Then, for any two tuples *t* and *s* on*U*, we shall write $t[A_i] =_{\upsilon,k} s[A_i]$ and call it an equality of degree *k*, or *k*-equality, if the following conditions hold:
1) If $t[A_i], s[A_i] \in D_{Ai}$ then $t[A_i] = s[A_i]$;
2) If only one of $t[A_i]$, $s[A_i]$ is a linguistic datum, say $t[A_i]$, then $s[A_i] \in S_k(t[A_i])$;
3) If both $t[A_i]$, $s[A_i]$ are linguistic data, then $S_k(t[A_i]) = S_k(s[A_i])$.

Definition 3.4.Let us assume the same assumptions as in Definition 3.3. Then,
1) We shall write $t[A_i] \leq_{\upsilon,k} s[A_i]$, if either $t[A_i] =_{\upsilon,k} s[A_i]$ or $S_k(t[A_i]) < S_k(s[A_i])$;
2) We shall write $t[A_i] <_{\upsilon,k} s[A_i]$, if $S_k(t[A_i]) < S_k(s[A_i])$;
3) We shall write $t[A_i] >_{\upsilon,k} s[A_i]$, if $S_k(t[A_i]) > S_k(s[A_i])$.
For illustration, let us consider the following example.

Table 3.1.An instance r of the relation scheme R

Name	Age	Title	NumOS	NumOP
Tim	35	PhD	*Small*	8
Kerry	55	Ass. Professor	59	46
William	41	Ass. Professor	68	*RRlarge*
Johnson	65	Professor	*Vlarge*	40
Mary	29	Ass. Professor	*MRlarge*	32
Robert	59	Professor	63	49
Martin	54	Ass. Professor	55	39

Example 3.2. Let us consider a relation scheme R = {NAME, AGE, TITLE, NumOS, NumOP}, where *NAME*, *AGE*, *TITLE*, *NumOS* and *NumOP* stand for Name, Age, Academic Title, Number of Scientific Works and Number of post graduate students which have defended their Master or Doctoral thesis successfully. This relation scheme may be called TEACHER-ABILITY. Assume an instance r of R given in Table 3.1.

For manipulating the linguistic data in such a database, the database system will calculate the quantitative semantics of the linguistic data. Let be given that D_{AGE} = [0, 120], D_{NumOS} = [0, 75] and D_{NumOP} = [0, 50] and that the fuzziness parameters of SQM assigned to *AGE* are the same as in the Example3.1, while the ones for *NumOS* and *NumOP* are the same and given as follows: $\kappa = fm(c^-)$ = 0.40, *Less*) = 0.25, $\mu(Rather)$ = 0.30, $\mu(More)$ = 25 $\mu(Very)$ = 0.20. Then, the system will calculate the pre-established data for the relation scheme R. To exemplify, we calculate some of them:

1) For the attribute *AGE*, we will calculate some neighbourhoods of *young* and the quantitative semantic values of *y*, *MRy* and *VRy*, which stand for *young*, *More Rather young* and *Very Rather young*, as follows. Firstly, as calculated in Example 3.1, we have:

$\upsilon_{A,\mathbf{r}}(young)$ = 25.92; $\upsilon_{A,\mathbf{r}}(Ryoung)$ = 34.2144 ; $\upsilon_{A,\mathbf{r}}(MRyoung)$ = 31.51872; $\upsilon_{A,\mathbf{r}}(VRyoung)$ = 27.412992 and $\upsilon_{A,\mathbf{r}}(Vyoung)$ = 4.6656.

Now, we calculate some similarity intervals of 2- and 3-partition of the domain of the attribute AGE:

$\theta_{2,AGE}(y) = \Im_{\boldsymbol{AGE}}(My) \cup \Im_{\boldsymbol{AGE}}(Ry) = (\upsilon_A(\Phi My), \upsilon_A(\Sigma Ry)] = (25.92 - 0.30\times0.45\times120,\ 25.92 + 0.32\times0.45\times120]$= (9.72, 43.2]. As computed in Example 3.1, we have $\theta_{3,AGE}(y) = \Im_3(LMy) \cup \Im_3(VPy)$ = (22.68, 29.0304].

2) Now, we compute some equivalence classes of 2- and 3-partition for the attributes *NumOS*. Firstly, we compute the representatives of *l* and *s*, where *l* and *s* stand for *large* and *small*, respectively:

$\upsilon_{W,r}(l) = [\kappa+\alpha fm(l)]\times75 = [0.40 + (0.30 + 0.25)\times0.60]\times75 = 54.75$;

$\upsilon_{W,r}(Rl) = \upsilon_{W,r}(l) + (-1)\times[\mu(R)\times fm(l) - \alpha\times\mu(R)\times fm(l)]\times75 = 54.75 - 0.45\times0.30\times0.60\times75 = 48.675$;

$\upsilon_{W,r}(MRl) = \upsilon_{W,r}(Rl) + (+1)\times[\mu(M)\times\mu(R)\times fm(l) - \beta\times\mu(M)\times\mu(R)\times fm(l))]\times75 = 48.675 + 0.55\times0.25\times0.30\times0.60\times75 = 50.53125$;

$\upsilon_{W,r}(Vl) = \upsilon_{W,r}(l) + (+1)\times[(\mu(M) + \mu(V))\times fm(l) - \beta\times\mu(V)\times fm(l))]\times75 = 54.75 + [0.45\times0.60 - 0.45\times0.20\times0.60]\times75 = 70.95$.

By Definition 3.2 of *k*-partition, we have :

$S_{W,2,r}(Rl) = S_{W,2}(l) = \Im_r(Rl) \cup \Im_r(Ml) = (\upsilon_{W,r}(\Phi Rl), \upsilon_{W,r}(\Sigma Ml)] = (54.75 - 0.30\times0.60\times75,\ 54.75 + 0.25\times0.60\times75]$ = (41.25, 66.00];

$S_{W,3,r}(Rl) = \Im_r(RRl) \cup \Im_r(MRl) = (\upsilon_{W,r}(\Phi RRl), \upsilon_{W,r}(\Sigma MRl)] = (48.675 - 0.30\times0.30\times0.60\times75,\ 48.675 + 0.25\times0.30\times0.60\times75]$= (44.625, 52.05]; $S_{W,2,r}(MRl) = S_{W,2,r}(Rl)$; $S_{W,3,r}(MRl) = S_{W,3,r}(Rl)$;

$S_{W,2,r}(Vl) = \Im_{2,r}(Vl) = (\upsilon_{W,r}(\Phi Vl), \upsilon_{W,r}(\Sigma Vl)] = (54.75 + (+1)\times0.25\times0.6\times75,\ 54.75 + (+1)\times[0.25 + 0.20]\times0.60\times75$= (66, 75];

$S_{W,3,r}(Vl) = \Im_r(RVl) \cup \Im_r(MVl) = (\upsilon_{W,r}(\Phi RVl), \upsilon_{W,r}(\Sigma MVl)] = (70.95 - 0.30\times0.20\times0.60\times75,\ 70.95 + 0.25\times0.20\times0.60\times75]$= (68.25, 73.20].

IV. CONCLUSION

In this paper, a relational model of databases with linguistic data is introduced in which the semantics of the linguistic data will be defined and manipulated based on the semantic-order-based structure of term-domains, called hedge algebras.

The way the semantics of uncertain data will be represented in databases is very important for handling the data in databases. Hedge algebras seem to be a useful tool to represent the qualitative meaning of term which can be formulated in term of an ordering relation on term-domains associated with certain attributes of a database. It is shown that various intuitive essential properties of the meaning of linguistic terms can be expressed in the structure of hedge algebras, while almost of them can not formulated in the framework of the fuzzy sets theory. Since in this algebraic approach, the ordering relation of hedge algebras can be regarded as being induced by the meaning of vague terms in natural language, the structure of hedge algebras can be considered as a direct mathematical model of term-domains, i.e. the underlying sets of hedge algebras can be understood to be just the mathematical structure of term-domains. On this research viewpoint, hedge-algebra-based semantics of linguistic data may reflect faithfully the natural meaning of terms in natural languages. As a consequence, this nature may bring many advantages in representing and handling data semantics in databases with linguistic data.

Taking these advantages, the semantics of linguistic data can be represented by a structure of fuzziness intervals of real domains of linguistic attributes of a database and, in data manipulation, linguistic data can be represented by their representatives (semantic values) in the respective real domains. Both kinds of data, intervals and values, are two components for representing the semantics of the linguistic data. Outward they are classical real data, however, different from the classical data they may be changed, because they depend on the fuzziness parameters which are subjective and in principle can be changed in the life cycle of the database. In spite of this, we have a basis to construct a method for manipulating the real data as well as the linguistic data in a unified way. Therefore, it is shown for example that queries related with linguistic data can be converted into certain classical ones. This of course is one considerable reason for simplifying the data manipulation tasks in management of the linguistic databases in comparison with that of fuzzy databases. The second reason for this is that on'ce given the fuzziness parameters, the data set for manipulating the linguistic data in databases can be pre-established. This speeds up the data manipulation process in the linguistic databases.

REFERENCES

[1]. L.V.T.Lan, N.M.Han, N.C.Hao, Algorithm to build fuzzy decision tree for data classification problem based on fuzziness intervals matching, *Journal of Computer sciences and Cybernetics*, Vol 32, No 4 (2016), 367-380.

[2]. L.V.T.Lan, Han.N.M, Hao.N.C, A novel method to build a fuzzy decision tree based on hedge algebras, *International Journal of Research in Engineering and Science*, Vol 4, No 4 (2016), 16-24.

[3]. Nguyen Cong Hao, Truong Thi My Le, Fuzzy object-oriented database model based on hedge algebra, *Journal of Computer Science & Cybernetics*, Vol 20, No 3 (2012), 129-140.

[4]. Nguyen Cong Hao, Truong Thi My Le, Data dependencies in fuzzy object oriented databases model based on hedge algebras, *Annales Univ. Sci Budapest., Sect. Comp*. 44 (2015), 165-182.

[5]. Nguyen Cong Hao, Fuzzy functional dependencies based on hedge algebras, *International Journal of Computer Technology and Applications*, Vol 6, No 6 (2015), 1052-1059.

[6]. Nguyen Cat Ho, Le Xuan Vinh, Nguyen Cong Hao, Unify data and establish similarity relation in linguistic databases base on hedge algebra, *Journal of Computer Science & Cybernetics*, Vol 25, No 4 (2009), 314-332.

[7]. N.C. Ho, A Topological Completion of Refined Hedge Algebras and a Model of Fuzziness of Linguistic terms, *Fuzzy Sets and Systems* 158(4) (2007) 436-451.

[8]. N.C. Ho, N.V. Long, Fuzziness Measure on Complete Hedge Algebras and Quantitative Semantics of Terms in Linear Hedge Algebras, *Fuzzy Sets and Systems* 158 (4) (2007) 452-471.

[9]. N.C. Ho, Quantifying Hedge Algebras and Interpolation Methods in Approximate Reasoning, *Proc. of the 5th Inter. Conf. on Fuzzy Information Processing, Beijing, March 1-4* (*2003*), 105-112.

[10]. Jyothi S, Babu M.S, Multivalued dependencies in fuzzy relational database and lossless join decomposition, *Fuzzy sets and systems* 88 (1997), 315-332.

[11]. Mustafa LLKerSozat, Adnan Yazici, A complete axiomatization for fuzzy functional and multivalued dependencies in fuzzy database relations, *Fuzzy Set and Systems* 117 (2001), 161-181.

[12]. Raju K.V.S.V.N, Majumdar A.K 1998, Fuzzy functional dependencies and lossless join decomposition of fuzzy relational database systems, *ACM TODS* 13 (4), 129-166.

[13]. Umano, M., &Fukami, S. (1994). Fuzzy relational algebra for possibility distribution-fuzzy-relation model of fuzzy data.*Journal of Intelligent Information Systems*, *3*, 7-28.

[14]. Wei-Yi Liu, Fuzzy data dependencies and implication of fuzzy data dependencies, *Fuzzy sets and systems*, Vol 92 (1997), 341-348.

Conghao Nguyen " Anapproach for Processing Fuzzy Data Based On Hedge Algebras." International Journal of Research in Engineering and Science (IJRES), vol. 06, no. 05, 2018, pp. 53–61.