A New Linguistic Modelling of the Symmetrical Threshold Semantics*

E. Herrera-Viedma
Dept. of Computer Science and Artificial Intelligence, Library Science Studies School, University of Granada, 18017 – Granada, Spain
viedma@decsai.ugr.es

A. G. López-Herrera
Dept. of Computer Science and Artificial Intelligence, Library Science Studies School, University of Granada, 18017 – Granada, Spain
agabriel@ugr.es

L. Hidalgo
Office of Transference of Investigation Results, University of Granada, E-18071 – Granada, Spain
lhidalgo@ugr.es

Abstract

In this contribution a new matching function for a symmetrical threshold semantics is presented. It can be interpreted as a tuning of that proposed in [10] which was defined using an ordinal fuzzy linguistic approach. This new matching function is defined using a 2-tuple fuzzy linguistic approach and, consequently, avoids the loss of precision and information in final results. On the other hand, it softens the behaviour of that proposed in [10] by processing in a more consistent way the query threshold weights. In such a way, this new linguistic matching function improves the results of linguistic information retrieval systems and therefore, can help to improve the users' satisfaction.

Keywords: Fuzzy Information Retrieval, Linguistic Modelling, Weighted Queries.

1. Introduction

The main activity of an Information Retrieval System (IRS) is the gathering of pertinent archived documents that better satisfy the user queries. IRSs present three components to carry out this activity [10, 11]: i) a *database*: which stores the documents and the representation of their information contents (index terms), ii) a *query subsystem*: which allows users to formulate their queries by means of a query language, iii) an *evaluation subsystem:* which evaluates the documents for a user query obtaining a Retrieval Status Value (RSV) form each document. The query subsystem supports the user-IRS interaction, and therefore, it should be able to account for the imprecision and vagueness typical of human communication. This aspect may be modelled by means of the introduction of weights in the query language. Many authors have proposed weighted IRS models using Fuzzy Set Theory [2, 3, 4, 7, 8, 9, 16, 19, 20, 21]. Usually, they assume numeric weights associated with the queries (values in [0, 1]). However, the use of query languages based on numeric weights forces the user to quantify qualitative concepts (such as "importance"), ignoring that many users are not able to provide their information needs precisely in a quantitative form but in a qualitative one. Some fuzzy linguistic IRS models [5, 6, 10, 11, 12, 17] have been proposed using a *fuzzy linguistic approach* [24] to model the query weights and document scores. A useful fuzzy linguistic approach which allows us to reduce the complexity of the design for the IRSs [10,11] is called the *ordinal fuzzy linguistic approach* [14, 15, 22]. In this approach, the query weights and document scores are ordered linguistic terms.
On the other hand, we have to establish the semantics associated with the query weights to formalize fuzzy linguistic weighted querying. There are four semantic possibilities [3, 10, 17]: i) weights as a measure of the importance of a specific element in representing the query, ii) as a threshold to aid in matching a specific document to the query, iii) as a description of an ideal or perfect document, and iv) as a limit on the amount of documents to be retrieved for a specific element.

In [10] a variant for a threshold semantics, called symmetrical threshold semantics, was proposed. This semantics has a symmetric behaviour in both sides of the mid threshold value. It assumes that a user may use presence weights or absence weights in the formulation of weighted queries. Then, it is symmetrical with respect to the mid threshold value, i.e., it presents the usual behaviour for the threshold values which are on the right of the mid linguistic value (presence weights), and the opposite behaviour for the values which are on the left (absence weights or presence weights with low value). To evaluate this semantics, in [10] was defined a parameterized symmetrical linguistic matching function. This function has like main limitation the loss of

* Este trabajo ha sido financiado por el proyecto: *GLIRS-II: Un Sistema de Recuperación de Información Documental Basado en Información Lingüística Difusa y Algoritmos Genéticos.* Ministerio de Ciencia y Tecnología. Ref. TIC-2003-07977.

information and precision in final results, i.e. in the computation of the linguistic RSVs of documents.

In this contribution we present a new modelling of the symmetrical threshold semantics defined in [10] which overcomes its difficulties. We present a new and alternative definition of the symmetrical matching function that synthesizes the symmetrical threshold semantics, softens the behaviour of that defined in [10], and allows to achieve more precise RSVs, improving the results of the retrieval, and consequently, allowing to increase the users' satisfaction. This new symmetrical matching function is defined in a 2-tuple fuzzy linguistic approach [13]. By using the 2-tuple fuzzy linguistic representation model we improve the precision in the representation of linguistic information and by using the 2-tuple computational model we avoid the loss of information in the combination operations of linguistic information.

The paper is structured as follows. Section 2 presents the 2-tuple fuzzy linguistic approach. Section 3 defines the new linguistic symmetrical matching function and accomplishes a study of its performance. And finally, in Section 4, some concluding remarks are pointed out.

2. A 2-tuple fuzzy linguistic approach

The *ordinal fuzzy linguistic approach* is an approximate technique appropriate to deal with qualitative aspects of problems [15]. An ordinal fuzzy linguistic approach is defined by considering a finite and totally ordered label set $S = \{s_0, ..., s_T\}$, $T+1$ is the cardinality of S in the usual sense, and with odd cardinality (7 or 9 labels). The mid term representing an assessment of "approximately 0.5" and the rest of the terms being placed symmetrically around it [1]. The semantics of the linguistic terms set is established from the ordered structure of the terms set by considering that each linguistic term for the pair (s_i, s_{T-i}) is equally informative. For each label s_i is given a fuzzy number defined on the [0,1] interval, which is described by a membership function. The computational model to combine ordinal linguistic information is based on the following operators:

1. Negation operator: $Neg(s_i) = s_j, j = T - i.$
2. Maximization operator: $MAX(s_i, s_j) = s_i$ *if* $s_i >= s_j$.
3. Minimization operator: $MIN(s_i, s_j) = s_i$ *if* $s_i <= s_j$
4. Aggregation operators: Usually to combine ordinal linguistic information we use aggregation operators based on symbolic computation, e.g. the LOWA operator [15] or the LWA operator [14].

Let S be a linguistic term set, if a symbolic method aggregating linguistic information obtains a value $\beta \in [0, T]$, and $\beta \notin \{0, \dots, T\}$ then an approximation function (app(.)) is used to express the index of the result in S [13]. For example, in the LOWA, app(.) is the simple function *round.*

Definition 1. [13] Let $\beta \in [0, T]$ be the result of an aggregation of the indexes of a set of labels assessed in a linguistic term set S, i.e., the result of a symbolic aggregation operation. Let $i = round(\beta)$ and $\alpha_i = \beta - i$ be two values, such that, $i \in \{0, 1, .., T\}$ and $\alpha_i \in [-.5, .5)$ then α_i is called a *Symbolic Translation*.

From this concept, F. Herrera and L. Martínez developed a linguistic representation model which represents the linguistic information by means of 2-tuples (s_i, α_i), $s_i \in S$ and $\alpha_i \in [-.5, .5)$ [13]:
- s_i represents the linguistic label of the information, and
- α_i is a numerical value expressing the value of the translation from the original result β to the closest index label i in S.

This model defines a set of transformation functions between numeric values and linguistic 2-tuples.

Definition 2. [13] Let S be a linguistic term set and $\beta \in [0, T]$, then the 2-tuple that expresses the equivalent information to β is obtained with the following function:

$$\Delta: [0, T] \rightarrow S x [-.5, .5);$$

$$\Delta(\beta) = (s_i, \alpha_i), \text{ with } \begin{cases} s_i & i = round(\beta) \\ \alpha_i = \beta - i & \alpha_i \in [-.5, .5) \end{cases} \quad (1)$$

where s_i has the closest index label to "β" and "α_i" is the value of the symbolic translation.

Proposition 1. [13] Let (s_i, α_i), $s_i \in S$ be a linguistic 2-tuple. There is always a Δ^{-1} function, such that, from a 2-tuple it returns its equivalent numerical value $\beta \in [0, T] \subset \Re$.

Remark 1: [13] From Definition 2 and Proposition 1, it is obvious that the conversion of a linguistic term into a linguistic 2-tuple consists of adding a value 0 as symbolic translation: $s_i \in S \rightarrow (s_i, 0)$.

The 2-tuple linguistic computational model operates with the 2-tuples without loss of information and is based on the following operations [13]:

1. Negation operator of a 2-tuple:

$$Neg(s_i, \alpha_i) = \Delta(T - \Delta^{-1}(s_i, \alpha_i)).$$

2. Comparison of 2-tuples: The comparison of linguistic information represented by 2-tuples is carried out according to an ordinary lexicographic order.

3. Aggregation of 2-tuples: Using the functions Δ and Δ^{-1} any numerical aggregation operator can be easily extended for dealing with linguistic 2-tuples.

Definition 3. [23] Let $A = \{a_1, \ldots, a_m\}$, $a_k \in [0,1]$ be a set of assessments to aggregated, then the OWA operator, ϕ, is defined as $\phi(a_1, \ldots, a_m) = W \cdot B^T$, where $W = [w_1, \ldots, w_m]$, is a weighting vector, such that $w_i \in [0,1]$ and $\Sigma_i w_i = 1$, and $B = \{b_1, \ldots, b_m\}$ is a vector associated to A, such that, $B = \sigma(A) = \{a_{\sigma(1)}, \ldots, a_{\sigma(m)}\}$, with σ being a permutation over the set of assessments A, such that $a_{\sigma(j)} \leq a_{\sigma(i)}\ \forall i \leq j$.

A 2-tuple linguistic extended definition of ϕ would be as follows:

Definition 4. Let $A = \{(a_1, \alpha_1), \ldots, (a_m, \alpha_m)\}$ be a set of assessments in the linguistic 2-tuple domain, then the 2-tuple linguistic OWA operator, ϕ_{2t} is defined as $\phi_{2t}((a_1,\alpha_1),\ldots,(a_m,\alpha_m)) = \Delta(W \cdot B^T)$,
$B = \sigma(A) = \{(\Delta^{-1}(a_1,\alpha_1))_{\sigma(1)},\ldots,(\Delta^{-1}(a_m,\alpha_m))_{\sigma(m)}\}$.

3. A new linguistic matching function for a symmetrical threshold semantics

In this section we present a new proposal to model the symmetrical threshold semantics defined in [10] in order to improve its performance. Before presenting it, we show the linguistic IRS assumed.

3.1. An ordinal linguistic weighted IRS based on a symmetrical threshold semantics

The ordinal linguistic weighted IRS that we use in our research presents the following elements to carry out its activity:

1. *Database:* we assume a database of a traditional fuzzy IRS as in [8, 19, 18]. The database stores the finite set of documents D = $\{d_1, \ldots, d_m\}$ represented by a finite set of index terms T = $\{t_1, \ldots, t_l\}$, which describe the subject content of the documents. The representation of a document is a fuzzy set of terms characterized by a numeric indexing function F: D×T → [0, 1], which is called index term weight [19]: $d_j = F(d_j, t_1)/t_1 + F(d_j, t_2)/t_2 + \ldots + F(d_j, t_l)/t_l$. F weighs index terms according to their significance in describing the content of a document. Thus $F(d_j, t_i)$ is a numerical weight that represents the degree of significance of t_i in d_j.

2. *Query subsystem:* we use a query subsystem with a fuzzy linguistic weighted Boolean query language to express user information needs. With this language each query is expressed as a combination of the weighted index terms that are connected by logical operators AND ($\wedge$), OR ($\vee$), and NOT($\neg$). The weights are ordinal linguistic values taken from a label set *S,* and they are associated with a symmetrical threshold semantics [10, 11]. As in [5], our atomic components are pairs, $\langle t_i, c_i \rangle$, where t_i is an index term, but defining the linguistic variable *Importance* with the ordinal linguistic approach and associating c_i with a symmetrical threshold semantics. Accordingly, the set $\boldsymbol{Q}$ of the legitimate queries is defined by the following syntactic rules:

1. $\forall q = \langle t_i, c_i \rangle \in \boldsymbol{T} \times S \rightarrow q \in \boldsymbol{Q}$.
2. $\forall q, p \in \boldsymbol{Q} \rightarrow q \wedge p \in \boldsymbol{Q}$.
3. $\forall q, p \in \boldsymbol{Q} \rightarrow q \vee p \in \boldsymbol{Q}$.
4. $\forall q \in \boldsymbol{Q} \rightarrow \neg q \in \boldsymbol{Q}$.
5. *All legitimate queries* $q \in \boldsymbol{Q}$ *are only those obtained by applying rules 1-4, inclusive.*

3. *Evaluation subsystem:* The evaluation subsystem for weighted Boolean queries acts by means of a constructive bottom-up process based on the *criterion of separability* [9, 19]. The RSVs of the documents are ordinal linguistic values whose linguistic components are taken from the linguistic variable *Importance* but representing the concept of *relevance*. Therefore, the set of linguistic terms *S* is also assumed to represent the relevance values. The evaluation subsystem acts in two steps:

1. Firstly, the documents are evaluated according to their relevance only to atoms of the query. In this step, the symmetrical threshold semantics is applied in the evaluation of atoms by means of a parameterized linguistic matching function $g : \boldsymbol{D} \times \boldsymbol{T} \times S \rightarrow S$, which is defined as [10]:

$$g(d_j, t_i, c_i) = \begin{cases} s_{Min\{a+\beta,T\}} & s_{T/2} \leq s_b \leq s_a \\ s_{Max\{0,a-\beta\}} & s_{T/2} \leq s_b \wedge s_a < s_b \\ Neg(s_{Max\{0,a-\beta\}}) & s_a \leq s_b < s_{T/2} \\ Neg(s_{Min\{a+\beta,T\}}) & s_b < s_{T/2} \wedge s_b < s_a \end{cases} \quad (2)$$

such that, (i) $s_b = c_i$; (ii) s_a is the linguistic index weight obtained as $s_a = Label(F(d_j,t_i))$, being $Label$:[0,1]→S a function that assigns a label in S to a numeric value $r \in [0,1]$; and iii) β is a bonus value that rewards/penalizes the relevance degrees of documents for the satisfaction /dissatisfaction of request <t_i,c_i>, which can be defined depending on the closeness between $Label(F(d_j,t_i))$ and c_i, for example as $\beta = round(2|b-a|/T)$. We should point out that whereas the traditional threshold matching function are always non-decreasing [17], g is non-decreasing on the right of the mid term and decreasing on the left of the mid term in order to be consistent with the meaning of the symmetrical threshold semantics.

2. Secondly, the documents are evaluated according to their relevance to Boolean combinations of atomic components, and so on, working in a bottom-up fashion until the whole query is processed. In this step, the logical connectives AND and OR are modelled by means of LOWA [15] operators with $orness(W)$<0.5 and $orness(W) \geq 0.5$ respectively, being $orness(W)$ a orness measure introduced by Yager in [23] to classify the aggregation of the OWA operators: $orness(W) = (1/m-1)(\sum_{i=1}^{m} (m-i)\, w_i)$.

Remark 2: We should point out that if we have a negated query, or a negated subexpression, or a negated atom, their evaluation is obtained from the negation of the relevance results computed for the query, or the subexpression, or atom in a no-negated situation.

3.2. Limitations of the symmetrical threshold semantics modelled by *g*

According to the symmetrical threshold semantics the evaluation subsystem assumes that a user may search for documents with a minimally acceptable presence of one term in their representations (as in the classical interpretation happens [17]) or documents with a maximally acceptable presence of one term in their representations. Then, when a user asks for documents in which the concept(s) represented by a term t_i is (are) with the value *High Importance*, the user would not reject a document with a *F* value greater than *High*; on the contrary, when a user asks for documents in which the concept(s) represented by a term t_i is (are) with the value *Low Importance*, the user would not reject a document with a *F* value less than *Low*. Given a request $<t_i, c_i> \in \mathbf{T}\times S$; this means that the query weights that imply the presence of a term in a document $c_i \geq s_{T/2}$ (e.g. *High, Very High*) they must be treated differently to the query weights that imply the absence of one term in a document $c_i < s_{T/2}$ (e.g. *Low, Very Low*). Then, if $c_i > s_{T/2}$ the request $<t_i, c_i>$, is synonymous with the request $<t_i$, *at least* $c_i>$, which expresses the fact that the desired documents are those having *F* values as high as possible; and if $c_i < s_{T/2}$ is synonymous with the request $<t_i$, *at most* $c_i>$, which expresses the fact that the desired documents are those having *F* values as low as possible. The linguistic matching function *g* defined in [10] represents a possible modelling of the meaning of the symmetrical threshold semantics. However, such modelling or interpretation presents the following limitations:

1. *The loss of precision*: This problem is a consequence of ordinal linguistic framework which works with discrete linguistic expression domains and this implies to assume limitations in the representation domain of RSVs. Therefore, as linguistic term sets (*S*) assumed have a limited cardinality (5,7 or 9 labels) to assess the linguistic RSVs, in consequence, it is difficult to distinguish or specify what documents really satisfy better the atomic weighted request $<t_i, c_i>$. Although the system retrieves many documents the possible relevance assessments are limited by the cardinality of the label set considered.

2. *The loss of information*: This problem also is a consequence of the ordinal linguistic approach because it forces us to apply approximation operations in the definition of *g*, in particular, the *rounding* operation used to calculate the parameter β, and as it is known [13], in such a case almost always there exists a loss of information

Example 1: Let $S= \{s_0 =Null\ (N),\ s_1=Extremely\_Low\ (EL),\ s_2=Very\_Low\ (VL),\ s_3=Low\ (L),\ s_4=Medium\ (M),\ s_5=High\ (H),\ s_6=Very\_High\ (VH),\ s_7=Extremely\_High\ (EH),\ s_8=Total\ (TO)\}$ be a label set used to assess the linguistic information in a IRS and consider two documents d_1 and d_2, such that, $Label(F(d_1,t_i))=EH$ and $Label(F(d_2,t_i))=TO$, respectively, then if the atomic request is $<t_i, M>$ we obtain the same relevance degree for both documents as a consequence of the loss of information, $g(d_1, t_i, M)= TO$ and $g(d_2, t_i, M)= TO$.

3. *g tends to overvalue the satisfaction/dissatisfaction of the requests*: This problem is a consequence of the own definition of *g*. For example, if we analyze its definition we can observe that relevance degrees generated when the threshold value is satisfied, i.e. $s_{Min\{a+\beta ,T\}}$, always are limited by the index term weight, s_a. This shows a too optimistic evaluation of the satisfaction of threshold value and reduces the possibilities of discrimination among the documents that satisfy the threshold value. Similarly, it happens in the dissatisfaction case.

In the following subsection, we try to overcome these problems by defining a new threshold matching function.

3.3. A 2-tuple linguistic matching function to model the symmetrical threshold semantics

In this section, we present a new symmetrical matching function to model the symmetrical threshold semantics that overcomes the problems of the matching function *g* [10] aforementioned. We design it by using as base the 2-tuple fuzzy linguistic representation model [13] and we call it like 2-tuple linguistic matching function g_{2t}.
Firstly, we should point out that the simple fact to define the new matching function g_{2t} in a 2-tuple linguistic approach allows us to solve the first problem of *g*, given that using the 2-tuple linguistic representation model in its definition g_{2t} inherits its properties, and one of the main properties of the 2-tuple linguistic representation model is to eliminate the loss of precision of the ordinal linguistic model [13]. On the other hand, to overcome the second problem we have to avoid to include approximation operations in the definition of g_{2t}, and to overcome the third problem we have to soften the relevance degrees generared by g_{2t} when threshold value is minimally satisfied by the index term weight.

As aforementioned, symmetrical threshold semantics has a symmetric behaviour in both sides of the mid threshold value because it is defined to distinguish two situations in the threshold interpretation: i) when the threshold value is on the left of the mid term and ii) when it is on the right. It assumes that a user may use presence weights or absence weights in the formulation of weighted queries. Then, it is symmetrical with respect to the mid threshold value. Therefore, analyzing the case of presence weights, i.e. threshold values which are on the right of the mid threshold value, we rapidly derive the case of absence weights.

When the linguistic threshold weight s_b given by a user is higher, in the usual sense, than middle label of the term linguistic set, $s_{T/2}$, the matching function g is non-decreasing. As aforesaid, in this case the problem of g is that it rewards excessively to those documents whose F values overcomes to the threshold weight s_b and penalizes excessively to those documents whose F values do not overcome s_b. We look for a non-decreasing matching function g_{2t} that softens the behaviour of g. Concretely, to achieve this goal g_{2t} should work as follows: the more the F values exceed the threshold values and the closer they are to the maximum RSV s_T, the greater the RSVs of the documents. However, when the F values are below the threshold values and closer to s_0, the lower the RSVs of the documents and the closer to s_0 they are. These two circumstances are called in the literature oversatisfaction and undersatisfaction [17]. Assuming a continuous numeric domain [0, T], in Figure 1 we represent graphically the desired behaviour of g_{2t} for three possible threshold values *T/2*, *u* and *u´*, being values *0*, *T/2*, and *T* the indexes of the following terms of *S:* bottom term, middle term and top term, respectively.

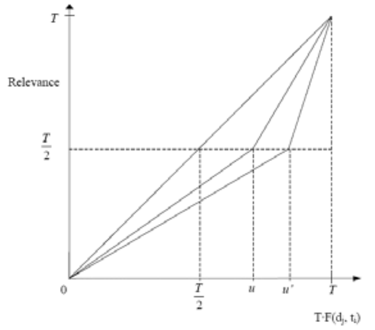

Figure 1. Desired behaviour of the matching function g_{2t}.

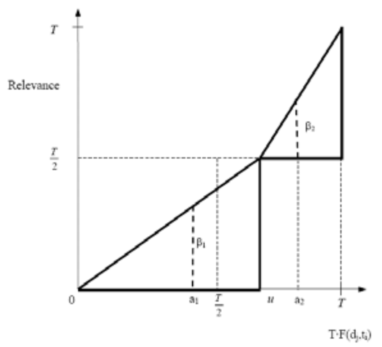

Figure 2. Desired behaviour of g_{2t} for a threshold value on the right of the mid term.

If we focus on the case of threshold value u (see Figure 2), then given two possible values of index term weight $a_1<u$ and $a_2>u$, the relevance degrees obtained by a desired matching function should be β_1 and $(T/2) + \beta_2$. Assuming this hypothesis the definition of the 2-tuple linguistic matching function g_{2t} on the right of the mid term would be as follows:

$g_{2t} : \boldsymbol{D}\times\boldsymbol{T}\times(S\times[-.5, .5)) \rightarrow S\times[-.5, .5))$

$$g_{2t}(d_j,t_i,(s_b,0)) = \begin{cases} \Delta(\beta_2 + \frac{T}{2}) & (s_a,\alpha_a) \geq (s_b,0) \wedge (s_b,0) \geq (s_{T/2},0) \\ \Delta(\beta_1) & (s_a,\alpha_a) < (s_b,0) \wedge (s_b,0) \geq (s_{T/2},0) \end{cases} \quad (3)$$

where $(s_a,\alpha_a) = \Delta(T\cdot F(d_j,t_i))$, $(s_b,0)$ is the representation in the linguistic 2-tuple model of the linguistic threshold weight given by a user, and β_1 and β_2 are numerical values obtained as follows. In Figure 2, two triangles are showing the behaviour of the desired matching function. The triangle on the right of the mid value *T/2* shows the way in which documents that have an index term weight a_2 higher than a threshold value u are rewarded, and the triangle on the left of the mid value shows the way in which documents that have an index term weight a_1 lower than u are penalized. Analysing both triangles we can calculate the following expressions for β_2 and β_1:

$$\frac{T-(\frac{T}{2})}{T-u} = \frac{\beta_2}{a_2-u} \Rightarrow \beta_2 = \frac{T\cdot(a_2-u)}{2\cdot(T-u)}$$

$$\frac{\frac{T}{2}}{u} = \frac{\beta_1}{a_1} \Rightarrow \beta_1 = \frac{a_1\cdot\frac{T}{2}}{u} = \frac{a_1\cdot T}{2\cdot u}.$$

To apply these expressions in the 2-tuple linguistic matching function g_{2t} we must know that:

- $u = \Delta^{-1}(s_b,0)$, being s_b the linguistic threshold value provided by a user,
- a_2 would be the numeric weight of some index term t_i representing the content of a document d_j, i.e., $a_2 = T\cdot F(d_j, t_i)$, and similarly
- a_1 would be the numeric weight of some index term t_i representing the content of a document d_k, i.e., $a_1 = T\cdot F(d_k, t_i)$.

Summarizing, given that g_{2t}, like g, must present a symmetric behaviour in both sides of the mid threshold value, then the complete definition of g_{2t} is easily obtained as follows:

$$g_{2t}(d_j,t_i,(s_b,0)) = \begin{cases} \Delta(\beta_2 + \frac{T}{2}) & if\,(s_a,\alpha_a) \geq (s_b,0) \wedge (s_b,0) \geq (s_{T/2},0) \\ \Delta(\beta_1) & if\,(s_a,\alpha_a) < (s_b,0) \wedge (s_b,0) \geq (s_{T/2},0) \\ \Delta(\beta_2^* + \frac{T}{2}) & if\,(s_a,\alpha_a) \leq (s_b,0) \wedge (s_b,0) < (s_{T/2},0) \\ \Delta(\beta_1^*) & if\,(s_a,\alpha_a) > (s_b,0) \wedge (s_b,0) < (s_{T/2},0) \end{cases}$$

(4)

with $\beta_2 = \frac{T\cdot(a_2-u)}{2\cdot(T-u)}$, $\beta_1 = \frac{a_1\cdot T}{2\cdot u}$, $\beta_2^* = \frac{T\cdot(u-a_1)}{2\cdot u}$, $\beta_1^* = \frac{T\cdot(T-a_2)}{2\cdot(T-u)}$, $u = \Delta^{-1}(s_b,0)$, $a_1 = T\cdot F(d_k, t_i)$ and

$a_2 = T \cdot F(d_j, t_i)$.

4. CONCLUDING REMARKS

In this paper we have described a new linguistic modelling of the symmetrical threshold semantics [10] in a linguistic framework. We have defined a new symmetrical linguistic matching function to model the meaning of the symmetrical threshold semantics that overcomes the problems found in the linguistic matching function defined in [10]. We have defined this new linguistic matching function in a 2-tuple fuzzy linguistic context [13] to take advantage of the usefulness of the 2-tuple fuzzy linguistic representation model with respect to avoid the problems of loss of precision and information in the results.

In the future, we shall research the impact of the different threshold matching functions existing in the literature in order to define a general application framework that facilitates us their design and use in the IRSs.

References

[1] P.P. Bonissone and K.S. Decker, Selecting uncertainty calculi and granularity: An experiment in trading-off precision and complexity, in: L.H. Kanal and J.F. Lemmer, Eds., *Uncertainty in Artificial Intelligence* (North-Holland, 1986) 217-247.

[2] A. Bookstein, Fuzzy request: An approach to weighted Boolean searches, *Journal of the American Society for Information Science* 31 (1980) 240-247.

[3] G. Bordogna, C. Carrara and G. Pasi, Query term weights as constraints in fuzzy information retrieval, *Information Processing & Management* 27 (1991) 15-26.

[4] G. Bordogna and G. Pasi, Linguistic aggregation operators of selection criteria in fuzzy information retrieval, *International Journal of Intelligent Systems* 10 (1995) 233-248.

[5] G. Bordogna and G. Pasi, A fuzzy linguistic approach generalizing Boolean Information retrieval: A model and its evaluation, *Journal of the American Society for Information Science* 44 (1993) 70-82.

[6] G. Bordogna and G. Pasi. An ordinal information retrieval model. International Journal of Uncertain, Fuzziness and Knowledge System, 9 (2001) 63-76.

[7] D. Buell and D.H. Kraft, Threshold values and boolean retrieval systems, *Information Processing & Management* 17 (1981) 127-136.

[8] D. Buell and D.H. Kraft, A model for a weighted retrieval system, *Journal of the American Society for Information Science* 32 (1981) 211-216.

[9] C.S. Cater and D.H. Kraft, A generalization and clarification of the Waller-Kraft wish list, *Information Processing & Management* 25 (1989) 15-25.

[10] E. Herrera-Viedma, Modelling the retrieval process for an information retrieval system using an ordinal fuzzy linguistic approach, *Journal of the American Society for Information Science and Technology* 52:6 (2001) 460-475.

[11] E. Herrera-Viedma, An information retrieval system with ordinal linguistic weighted queries based on two weighting elements. *Int. J. of Uncertainty, Fuzziness and Knowledge-Based Systems* 9 (2001) 77-88.

[12] E. Herrera-Viedma, O. Cordón, M. Luque, A.G. Lopez, A.M. Muñoz, A model of fuzzy linguistic IRS based on multi-granular linguistic information, *International Journal of Approximate Reasoning* 34 (2003) 221-239.

[13] F. Herrera and L. Martínez, A 2-tuple fuzzy linguistic representation model for computing with words, *IEEE Transactions on Fuzzy Systems* 8:6 (2000) 746-752.

[14] F. Herrera and E. Herrera-Viedma, Aggregation operators for linguistic weighted information, *IEEE Transactions on Systems, Man, and Cybernetics, Part A: Systems and Humans,* 27 (1997) 646-656.

[15] Herrera, F., Herrera-Viedma, E., and Verdegay, J. L. (1996). Direct approach processes in group decision making using linguistic OWA operators. *Fuzzy Sets and Systems*, 79: 175-190.

[16] D.H. Kraft and D.A. Buell, Fuzzy sets and generalized Boolean retrieval systems, *International Journal of Man-Machine Studies* 19 (1983) 45-56.

[17] D.H. Kraft, G. Bordogna and G. Pasi, An extended fuzzy linguistic approach to generalize Boolean information retrieval, *Information Sciences* 2 (1994) 119-134.

[18] S. Miyamoto, *Fuzzy Sets in Information Retrieval and Cluster Analysis* (Kluwer Academic Publishers, 1990).

[19] W.G. Waller and D.H. Kraft, A mathematical model of a weighted Boolean retrieval system, *Information Processing & Management* 15 (1979) 235-245.

[20] R.R. Yager, A Hierarchical Document Retrieval Language, *Information Retrieval* 3 (2000) 357-377.

[21] R.R. Yager, A note on weighted queries in information retrieval system, *Journal of American Society of Information Sciences* 38 (1987) 23-24.

[22] R.R. Yager, An approach to ordinal decision making, *International Journal of Approximate Reasoning* 12 (1995) 237-261.

[23] R.R. Yager, On ordered weighted averaging aggregation operators in multicriteria decision making, *IEEE Transactions on Systems, Man, and Cybernetics* 18 (1988) 183-190.

[24] L.A. Zadeh, The concept of a linguistic variable and its applications to approximate reasoning. Part I, *Information Sciences* 8 (1975) 199-249, Part II, *Information Sciences* 8 (1975) 301-357, Part III, *Information Sciences* 9 (1975) 43-80.