



Softcomputing in neutrosophic linguistic modeling for the treatment of uncertainty in information retrieval

Dionisio Vitalio Ponce Ruiz¹, Juan Carlos Albarracín Matute², Edmundo José Jalón Arias³, Luis Orlando Albarracín Zambrano⁴, Luis Javier Molina Chalacán⁵, Ítalo Mecias Serrano Quevedo⁶, and Andrea Raquel Zuñiga Paredes⁷

¹ Professor, Universidad Regional Autónoma de los Andes – Extension Quevedo, Ecuador. E-mail: uq.dionisioponce@uniandes.edu.ec

² Professor, Universidad Regional Autónoma de los Andes - Extension Quevedo, Ecuador. E-mail: uq.juanalbarracin@uniandes.edu.ec

³ Professor, Universidad Regional Autónoma de los Andes – Extension Quevedo, Ecuador. E-mail: uq.edmundojalon@uniandes.edu.ec

⁴ Professor, Universidad Regional Autónoma de los Andes - Extension Quevedo, Ecuador. E-mail: uq.luisalbarracin@uniandes.edu.ec

⁵ Professor, Universidad Regional Autónoma de los Andes – Extension Quevedo, Ecuador. E-mail: uq.luismoreira@uniandes.edu.ec

⁶ Professor, Universidad Regional Autónoma de los Andes - Extension Quevedo, Ecuador. E-mail: uq.italoserrano@uniandes.edu.ec

⁷ Professor, Universidad Regional Autónoma de los Andes – Extension Quevedo, Ecuador. E-mail: uq.andreazuniga@uniandes.edu.ec

Abstract. Softcomputing is an extension of different concepts and techniques that try to overcome the difficulties presented by real problems that are imprecise, uncertain and difficult to categorize. The use of Softcomputing contributes to clarify that this term is not in itself a precise definition, but is a mixture of techniques, tools and methods that converge among themselves and each process has a specific objective. Within the group of Softcomputing techniques that are frequently used in linguistic modeling, for the treatment of uncertainty and the recovery of information, there is Neurocomputing, Probabilistic Reasoning, Fuzzy Logic and now neutrosophic Logic. In the present work, the combination of Softcomputing and neutrosophic logic is presented, for the neutrosophic linguistic modeling, in the treatment of uncertainty and information recovery, by means of a diffuse ordinal linguistic approximation. Modeling is performed by the linguistic information aggregation operator LOWA, which performs flexible modeling in the treatment of uncertainty and the evaluation of the proposed linguistic information retrieval. For this reason, the objective of this work is to deal with uncertainty in the process of recovering information.

Keywords: Softcomputing, neutrosophic linguistic modeling, uncertainty treatment, information retrieval.

1 Introduction

The interpretation of data, texts and contexts has become a great challenge for the achievement of human and professional development, emerging various informatic tools and philosophical theories that try to transform this reality.

Uncertainty and problem solving are a major issue in the effectiveness of the information management process. Incomplete information needs to be treated in order to reduce bias in all information management processes; therefore, it needs to be treated in order to obtain accurate results in all processes where incomplete and inaccurate information is present [1].

The treatment of uncertainty has become a process of different dimensions, a process that was not carried out until the terms related to the Theory of Probability were conceived. By recognizing from the Theory of Probability, the dimensions defined by [2] and [1], it was that uncertainty began to be dealt with.

According to the diversity that uncertainty presents, different types of uncertainty were established, where from that moment on, its classification began according to Bonnisone and Tong [3]. This uncertainty classification focused on knowledge imperfection, imprecision and incomplete information.

The imprecision and incomplete information, has been used frequently in different branches of science, specifically it has been used to indicate the absence of a value. Essentially imprecision indicates the existence of a value that cannot be measured in its entirety, and uncertainty reveals the fact of a subjective formulation about the veracity of a fact which is not known with certainty.

The need to consider uncertainty in the process of solving problems, bearing in mind that it may be present in the data of a given problem, as in the knowledge used to solve different types of problems, was that the development of various methods of solving uncertainty began.

One of the techniques that has become an effective way to deal with uncertainty, especially when represented

by inconsistency in the data is Neutrosophy. Neutrosophy is a new branch of philosophy [4] that studies the origin, nature and scope of neutrality, as well as its interactions with different ideational spectra.

Neutrosophy has formed the basis for a set of mathematical theories, which generalize the use of neutrosophic logic [4], neutrosophic sets, neutrosophic probability, neutrosophic statistics and multiple practical applications. It constitutes one of the most far-reaching applications at present, to deal with uncertainty and in particular neutrosophic logic [5].

Neutrosophic logic is a non-probabilistic approximate reasoning method [6], which is defined as an extension of fuzzy logic and facilitates the modeling of qualitative information in an approximate manner. Its success is mainly due to the possibility of solving complex and poorly defined problems which, by means of traditional methods, are difficult to solve; among the problems to be solved is the treatment of uncertainty and the retrieval of information.

Information retrieval is dedicated to the storage and retrieval of documents that are not quantitative in nature, i.e. documents with vague and imprecise information. Therefore, in view of this problem, it is recommended to make use of Softcomputing techniques, tools, tools and methods [7].

It is in this context of information retrieval that the development of Softcomputing begins. Softcomputing aims to take advantage of the knowledge of a problem that leads to imprecision and uncertainty, to achieve greater profit, strength and solution.

The main disciplines that converge with Softcomputing are Neurocomputation and Probabilistic Ratio, Fuzzy Logic and now Neutrosophic Logic. Neutrosophic Logic according to [4] is a logic in which each proposition is T% true, I% indeterminate, and F% false.

Based on the above, in the present work, the use of neutrosophic logic is proposed, for the linguistic modeling in the treatment of uncertainty for the recovery of qualitative information. Linguistic modeling has frequently been used in problems where phenomena related to human perceptions and relationships (design, taste, fun, etc.) are evaluated.

2 Preliminaries

2.1 Neutrosophic Linguistic Modeling

In human perceptions and relationships, words from natural language (beautiful, ugly, sweet, salty, sympathetic, many, few, others...) are used instead of numerical values to emit refer valuations. For this reason the use of neutrosophic linguistic modeling is proposed, which has as its theoretical basis the theory of neutrosophic sets and the theory of diffuse sets, techniques that have proved effective in evaluating aspects of qualitative nature [8].

Linguistic variables are used to represent the qualitative aspects whose domain of expression are sets of words or linguistic terms according to [9]. A linguistic variable according to [10] and Herrera and Herrera-Viedma, is characterized by a syntactic value or label and by a semantic value or meaning.

The label is a word or phrase belonging to a set of linguistic terms and the meaning of that label is given by a diffuse subset in a universe of discourse. Since words are less precise than numbers, the concept of linguistic variable is a good proposal to characterize those phenomena that are not suitable to be evaluated through numerical values.

The methodology to be used in neutrosophic linguistic modeling for the treatment of uncertainty in information retrieval follows a diffuse ordinal linguistic approach, an approach that constitutes a technique designed to model the qualitative aspects of a given problem. In a diffuse ordinal linguistic approach, the set of finite labels $S = \{S_i, i \in H = 0, K, T\}$ is assumed, totally ordered in the normal sense and with an odd cardinal (7 or 9 labels) [11].

The central label represents an approximate value of 0.5 and the rest is located symmetrically around it. The semantics of the labels are defined based on the ordered structure of the set of labels when considering the pair par (S_i, S_{T-i}) which means; which is also informative.

Each label has associated a neutrosophic number in the interval $[0,1]$, this neutrosophic number is defined by a trapezoidal membership function represented by a 4- tuple $a_i, b_i, \alpha_i, \beta_i$. The first two parameters indicate the interval in which the value of belonging is 1; the third and fourth indicate the amplitude to the left and to the right of the distribution), with the following properties:

1. Order: $S_i \geq S_j$ if $i \geq j$
2. Negation: $Neg(S_i) = S_j$, with $j = T - i$
3. Maximum: $MAX(S_i, S_j) = S_i$ if $S_i \geq S_j$
4. Minimum: $MIN(S_i, S_j) = S_i$ if $S_j \leq S_i$

2.2 LOWA operator

For the aggregation of linguistic values the operator LOWA is used, which is an ordinate-do operator based

on symbolic computation [11]. It operates on labels, and considers only the order of them and not their associated semantics. The LOWA operator is defined as:

Let $A = \{a_1, \dots, a_m\}$, the set of labels to add, then the LOWA operator, Φ is defined as $\Phi(a_1, \dots, a_m) = W \cdot B^T = C^M\{w_k, b_k, k = 1, \dots, m\} = w_1 \ominus b_1 \oplus (1 - w_1) \ominus C_m - 1 \{\beta_h, b_h, h = 2, \dots, m\}$.

(1)

Where; $W = [w_1, \dots, w_m]$ is a vector of weights, such that, $w_i \in [0, 1]$ y $\sum_i w_i = 1$. $\beta_h = w_h / (\sum_2^m w_k)$, $h = 2, \dots, m$, y $B = \{b_1, \dots, b_m\}$ is an ordered vector associated with A , such that, $B = \sigma(A) = \{a_{\sigma(1)}, \dots, a_{\sigma(m)}\}$, where $a_{\sigma(j)} \leq a_{\sigma(i)} \forall i \leq j$, σ is a permutation about A . C_m is the convex combination operator of m labels and if $m=2$, then $C_2\{w_i, b_i, i = 1, 2\} = w_1 \ominus s_j \oplus (1 - w_1) \ominus s_i = s_k$, such that $k = \min\{T, i + \text{round}(w_1 \cdot (j - i))\}$ } $s_j, s_i \in S, (j \geq i)$, Where; round is the rounding operator, y $b_1 = s_j$, $b_2 = s_i$. $S_i w_j = 1$ y $w_i = 0$ with $i \neq j \forall i$, that $C_m\{w_i, b_i, i = 1, \dots, m\} = b_j$.

The LOWA operator is an or-and operator [12], operating between the MIN and MAX operators. In the neutrosophic linguistic modeling proposed in this paper, this operator is used to evaluate Boolean queries and to classify OWA operators according to how they operate between the and and the or, as proposed by Yager [15], who defined a measure of orness, associated with vector W , as shown in equation 1.

$$\text{orness}(W) = \frac{1}{m-1} \sum_{k=1}^m (m-k)w_k \quad (2)$$

Set to a W , the closer the OWA operator's behavior is to or closer will be its orness measure to 1; the closer it is to an and, the closer it will be to 0. An OWA operator with a lot of non-zero weights in the first positions, behaves like an or (orness ≥ 0.5), however the one presented in the last positions behaves like an and [13, 14].

3. Neutrosophic linguistic modeling framework

Based on the above, Figure 1 shows the components for neutrosophic linguistic modeling framework useful for the treatment of uncertainty in information retrieval.



Figure 1: Components for linguistic modeling in the treatment of uncertainty and information retrieval. Source: Own preparation.

- The Database for the neutrosophic linguistic modeling in the treatment of the uncertainty and the recovery of the information is necessary to store the documents and the representation of their informative contents whose components are the index terms and their weights. These weights indicate the concepts induced by the terms regarding the index of the documents.

- Consultations make it easier for users to specify their information needs by formulating queries.

- Evaluations are performed to evaluate documents for each query using an equality function, where a Recovery status value is assigned to each document.

According to the components described above, it is possible to carry out neutrosophic linguistic modeling to deal with uncertainty and recover the information. This modeling supports weighted linguistic Boolean queries, which solves the inconvenience of AND and OR operators, inconveniences that are given because they operate very restrictively and very inclusively, respectively.

⁷⁶ For the proposed neutrosophic linguistic modeling, in the first place, a diffuse ordinal linguistic approximation is defined as in [15]. The terms of the queries are weighted by linguistic labels interpreted as significant limits constituting linguistic weights.

Linguistic weights are interpreted as limits to be satisfied in the equalization between documents and consultations. The Boolean operators AND and OR are modeled using the linguistic information aggregation operator LOWA [16], which facilitates the flexibility of the consultation evaluation process. Retrieved documents are classified into notability classes, identified by language tags.

3 Case Study

According to the components defined in Figure 1, the Database for neutrosophic linguistic modeling in the treatment of uncertainty and information retrieval stores the documents $D = \{d_1, \dots, d_m\}$ and their respective representations $R = \{R_1, \dots, R_d\}$. The documents are represented by index terms $T = \{t_1, \dots, t_n\}$. Each term has an

associated weight describes the content of the documents and is represented through the numerical indexation function that assigns to each document d_j and to each index term t_i a numerical weight between 0 and 1, this function is shown through the expression 2.

$$F : D \times T \rightarrow [0,1] \quad (2)$$

Where;

$F(d_j, t_i)$ is a weight that represents the degree to which t_i in d_j is significant.

$F(d_j, t_i) = 0$ implies that the document d_j , is not represented by you.

$F(d_j, t_i) = 1$ implies that the document is perfectly represented by you

$F(d_j, \dots, t_i) \in (0,1)$ represents the different intermediate significant degrees.

The Database in the present work contains seven documents $D=\{d_1, \dots, d_7\}$. The documents are indexed by the indexation function F which assigns the following weights to each term in the documents:

$$\begin{aligned} Rd1 &= 0.7/t_5 + 0.4/t_6 + 1/t_7; Rd2 = 1/t_4 + 0.6/t_5 + 0.8/t_6 + 0.9/t_7; Rd3 \\ &= 0.5/t_2 + 1/t_3 + 0.8/t_4; Rd4 = 0.9/t_4 + 0.6/t_6 + 1/t_7; Rd5 \\ &= 0.7/t_3 + 1/t_4 + 0.4/t_5 + 0.8/t_9 + 0.6/t_{10}; Rd6 = 1/t_5 + 0.99/t_6 + 0.8/t_7; Rd7 \\ &= 0.8/t_5 + 0.02/t_6 + 0.8/t_7 + 0.9/t_8 \end{aligned}$$

The queries, as component 2, defined in Figure 1, are performed from the linguistic labels stored in the Database. Each query is expressed as a combination of weighted index terms, connected by the Boolean operators *AND* ($\wedge$), *OR* ($\vee$), *y NOT* ($\neg$), and weighted by language tags. The linguistic variable "Relevance" is then used to model the linguistic weights, defined by the neutrosophic linguistic approach. For this purpose, a set of ordered labels is considered that are denoted as S to express the linguistic weights.[17] The linguistic variable "Relevance" is used to model linguistic weights, defined by the neutrosophic linguistic approach.

The query itself is considered as a Boolean expression whose components are 2-tuples $\langle t_i, c_i \rangle$ belonging to the set, $T \times S$; $t_i \in T$, where c_i is a value of the linguistic variable "Relevance", associated to a limit semantics [18].

The Q set of reliable queries is defined by the following syntax rules:

1. $\forall q = \langle t_i, c_i \rangle \in T \times S \rightarrow q \in Q$
2. $\forall q, p \in Q \rightarrow q \wedge p \in Q$
3. $\forall q, p \in Q \rightarrow q \vee p \in Q$
4. $\forall q \in Q \rightarrow \neg(q) \in Q$
5. All credible queries $q \in Q$ are only those obtained by applying the above rules.

Therefore, from the linguistic variables stored in the Database, the query shown in expression 3 is obtained.

$$q = ((t_5, VH) \wedge (t_6, L)) \vee (t_7, H) \quad (3)$$

Through the consultation produced, the representation of the documents is obtained expressed in linguistic form.

$$\begin{aligned} \text{Label: } Rd1 &= H/t_5 + M/t_6 + T/t_7; Rd2 = T/t_4 + M/t_5 + H/t_6 + VH/t_7; Rd3 \\ &= M/t_2 + T/t_3 + H/t_4; Rd4 = VH/t_4 + VL/t_6 + T/t_7; Rd5 \\ &= H/t_3 + T/t_4 + M/t_5 + H/t_9 + M/t_{10}; Rd6 = T/t_5 + EH/t_6 + H/t_7; Rd7 \\ &= H/t_5 + EL/t_6 + H/t_7 + VH/t_8 \end{aligned}$$

Once the query is obtained, component 3 is executed, relative to the evaluation, for this purpose the sensitivity parameter $K=2$ is assumed, which means the distance between the linguistic values. Therefore, the evaluations are carried out according to the symmetric limit semantics expressed by g^1 , which is nothing more than an equalization function different from the classical functions associated with limit semantics, which are increasing monotonous.

In this case the semantics is symmetrical with respect to the central limit values ($S_{T/2}$) which means that g^1 is increasing in $Label(F)$ for limit values greater than ($S_{T/2}$) and decreasing in $Label(F)$ for limit values that are less than ($S_{T/2}$). K is a sensitivity parameter defined to control the importance of the proximity or distance between $Label(F)$ and sbexpression in the final result. The higher the value of K , the smaller the importance of the distance. If $K=1$ means that symmetrical limit semantics are not applied. Then, the evaluations according to the symmetrical limit semantics expressed by g^1 are:

1. $E * (\langle t_5, VH \rangle, d1) = VH, E * (\langle t_5, VH \rangle, d2) = H, E * (\langle t_5, VH \rangle, d5) = H, E * (\langle t_5, VH \rangle, d5) = H, E * (\langle t_5, VH \rangle, d6) = T, E * (\langle t_5, VH \rangle, d7) = VH$
2. $E * (\langle t_6, L \rangle, d1) = M, E * (\langle t_6, L \rangle, d2) = M, E * (\langle t_6, L \rangle, d4) = VH, E * (\langle t_6, L \rangle, d6) = L, E * (\langle t_6, L \rangle, d7) = VH$

$$3. E * (< t7, H >, d1) = T, E * (< t7, H >, d2) = VH, E * (< t7, H >, d4) = T, E * (< t7, H >, d6) = H, E * (< t7, H >, d7) = H$$

The evaluation of sb expression $(t5, VH) \wedge (t6, L)$ through Φ with $W = [0.3, 0.7]$ is:

$$\begin{aligned} 1. E * (< t55, VH > \wedge < t6, L >, d1) &= \Phi(VH, M) = H; \quad 2. E * (< t5, VH > \wedge < t6, L >, d2) = \Phi(H, M) = M \\ 2. E * (< t5, VH > \wedge < t6, L >, d4) &= \Phi(N, VH) = VL; \quad 4. E * (< t5, VH > \wedge < t6, L >, d5) = \Phi(H, N) = VL \\ 3. E * (< t5, VH > \wedge < t6, L >, d6) &= \Phi(H, L) = M; \quad 6. E * (< t5, VH > \wedge < t6, L >, d7) = \Phi(H, VH) = H \end{aligned}$$

Finally, the complete query is evaluated by using the operator Φ with $W = [0.7, 0.3]$:

$$\begin{aligned} 1. E * (q, d1) &= \Phi(H, T) = EH; \quad 2. E * (q, d2) = \Phi(M, VH) = H; \quad 3. E * (q, d4) = \Phi(VL, T) = VH \\ 2. E * (q, d5) &= \Phi(VL, N) = EL; \quad 5. E * (q, d6) = \Phi(M, H) = H; \quad 6. E * (q, d7) = \Phi(H, H) = H \end{aligned}$$

It is demonstrated with the result obtained, when making use of the operators Max and Min are less strict, with which the following result is obtained.

$$[RSV1 = T, RSV2 = VH, RSV4 = T, RSV5 = N, RSV6 = H, RSV7 = H]$$

The operator LOWA, when applied to model the evaluation of the neutrosophic connectives AND and OR in the evaluation process of the consultations, indicates that it is in front of the use of the Soft-computing techniques, which are introduced in the recovery of the information and contribute to eliminate the too strict evaluation of the operators that model the neutrosophic connectives, which is useful in the treatment of the uncertainty.

Conclusion

In this paper, the neutrosophic linguistic modeling was presented for the treatment of uncertainty in the recovery of information based on an ordinal neutrosophic linguistic approach that accepts linguistic weighted Boolean consultations. This neutrosophic linguistic modeling takes advantage of the property of being an operator of the LOWA operator, used to aggregate ordinal linguistic information. The LOWA operator is applied to model the evaluation of the logical AND and OR connectives in the query evaluation process. In this way, Softcomputing becomes effective in the retrieval of information and it is possible to eliminate the too strict evaluations of the operators that model the neutrosophic connectives, which is advantageous in the treatment of uncertainty.

References

- [1] Gómez Vieites, Á. and C. Suárez Rey, *Sistemas de información: herramientas prácticas para la gestión empresarial*. Madrid: Ra-Ma Editorial, 2005.
- [2] Salazar, A.A.P., *Modelo de implantación de Gestión del Conocimiento y Tecnologías de Información para la Generación de Ventajas Competitivas*. Universidad Tecnología Federico Santa María. Valparaíso, 2000.
- [3] Tong, R.M. and P.P. Bonissone, *A linguistic approach to decisionmaking with fuzzy sets*. IEEE Transactions on Systems, Man, and Cybernetics, 1980. **10**(11): p. 716-723.
- [4] Smarandache, F. and M. Leyva-Vázquez, *Fundamentos de la lógica y los conjuntos neutrosóficos y su papel en la inteligencia artificial*. 2018: Infinite Study.
- [5] Smarandache, F., *Neutrosophic set-a generalization of the intuitionistic fuzzy set*. International journal of pure and applied mathematics, 2005. **24**(3): p. 287.
- [6] Abdel-Basset, M. and M. Mohamed, *The role of single valued neutrosophic sets and rough sets in smart city: Imperfect and incomplete information systems*. Measurement, 2018. **124**: p. 47-55.
- [7] Zadeh, L.A., *Soft computing and fuzzy logic*, in *Fuzzy Sets, Fuzzy Logic, and Fuzzy Systems: Selected Papers by Lotfi a Zadeh*. 1996, World Scientific. p. 796-804.
- [8] Liang, R., J. Wang, and H. Zhang, *Evaluation of e-commerce websites: An integrated approach under a single-valued trapezoidal neutrosophic environment*. Knowledge-Based Systems, 2017. **135**: p. 44-59.
- [9] Zadeh, L.A., *Computing with words in Information/Intelligent systems 1: Foundations*. Vol. 33. 2013: Physica.
- [10] Delgado, M., et al., *Fuzzy quantification: a state of the art*. Fuzzy Sets and Systems, 2014. **242**: p. 1-30.
- [11] Haenni, R. *Shedding new light on Zadeh's criticism of Dempster's rule of combination*. in *2005 7th International conference on information fusion*. 2005. IEEE.

- [12] Herrera, F., E. Herrera-Viedma, and J.L. Verdegay, *Direct approach processes in group decision making using linguistic OWA operators*. Fuzzy Sets and systems, 1996. **79**(2): p. 175-190.
- [13] Yager, R.R., *Applications and extensions of OWA aggregations*. International Journal of Man-Machine Studies, 1992. **37**(1): p. 103-122.
- [14] Salido, J.F. and S. Murakami, *Extending Yager's orness concept for the OWA aggregators to other mean operators*. Fuzzy Sets and Systems, 2003. **139**(3): p. 515-542.
- [15] Hernandez, N.B. and J.E. Ricardo, *Gestion empresarial y posmodernidad*. 2018: Infinite Study Pons Publishing House, Bruxelles Belgium.
- [16] LEYVA, M., et al., *A framework for PEST analysis based on fuzzy decision maps*. Revista ESPACIOS, 2018. **39**(16).
- [17] Hernandez, N.B., et al., *LA TOMA DE DECISIONES EN LA INFORMATICA JURIDICA BASADO EN EL USO DE LOS SISTEMAS EXPERTOS*. Investigación Operacional, 2019. **40**(1): p. 131-140.
- [18] Herrera, F. and L. Martínez, *A model based on linguistic 2-tuples for dealing with multigranular hierarchical linguistic contexts in multi-expert decision-making*. IEEE Transactions on Systems, Man, and Cybernetics, Part B (Cybernetics), 2001. **31**(2): p. 227-234.

Received: January 11, 2019.

Accepted: May 15, 2019

