A Fuzzy Linguistic Representation Model based on a Symbolic Translation

F. Herrera[a], **Luis Martínez**[b]
[a]Dept. of Computer Science and A.I. University of Granada, 18071 - Granada, Spain. e-mail: herrera@decsai.ugr.es
[b] Dept. of Computer Science. University of Jaén, 23071 - Jaén, Spain. e-mail: martin@ujaen.es

Abstract

The fuzzy linguistic approach has been applied successfully to many problems. However, there is a limitation on this approach, the loss of information. It appears due to its information representation model (discrete terms) and the computational methods used when fusion and combination processes are performed on linguistic variables. In this contribution we propose a new fuzzy linguistic representation model based on the concept of "Symbolic Translation" for dealing with linguistic information in a continuous domain. Together with this representation model we shall develop a computational technique for fusing linguistic variables without loss of information.

Keywords: Linguistic variables, linguistic modeling, fusion of linguistic information.

1 Introduction

The problems depending on their aspects can deal with different types of information. Usually, the problems present quantitative aspects that can be assessed by means of precise numerical values, but in other cases the problems present qualitative aspects that are complex to assess by means of precise values. In the latter case, the use of the fuzzy linguistic approach [10] has provided good results. When a problem is solved using linguistic information, implies the need to use computational techniques that provide linguistic operators of fusion and comparison. Here an important limitation for this approach appears, "the loss of information". The computational models [1, 4, 3] used in the specialized literature for fusing linguistic variables present this drawback that implies a lack of precision in the final results. The loss of information appears because the linguistic representation model used by the fuzzy linguistic approach is discrete (symbolic linguistic terms) while the computational methods used by it are continuous. Therefore, it happens that the results usually do not exactly match any of the initial linguistic terms, then an approximation process must be developed to express the result in the source expression domain. This produces the consequent loss of information.

The aim of this contribution is to develop a new representation model for overcoming this limitation. We shall introduce the concept of "Symbolic Translation" that is the basic element to develop the new linguistic model. This representation model uses a 2-tuple composed by a linguistic term and a number that supports the value of the Symbolic Translation. The main advantage of this model is to be continuous in its domain, therefore it can express any counting of information in the universe of the discourse without need for any approximation process. Together with this representation model, we present a computational method to deal with the 2-tuples without loss of information. Finally, we shall apply this fuzzy linguistic representation model in a decision-making problem, in which we can see how this model is more precise than the previous ones.

In order to do that, this contribution is structured as follows: in Section 2 we shall present a brief review of the fuzzy linguistic approach; in Section 3 we analyze linguistic computational techniques; in Section 4 we develop the new linguistic representation model; in Section 5 we present an application over a decision process using the 2-tuple representation model and finally, some concluding remarks are included.

2 Fuzzy Linguistic Approach

Usually, we work in a quantitative setting, where the information is expressed by means of numerical values. However, many aspects of different activities in the real world cannot be assessed in a quantitative form, but rather in a qualitative one, i.e., with vague or imprecise knowledge. In that case a better approach may be to use linguistic assessments instead of numerical values. The fuzzy linguistic approach represents qualitative aspects as linguistic values by means of linguistic variables [10].

We have to choose the appropriate linguistic descriptors for the term set and their semantic. In the literature, several possibilities can be found (see [6] for a wide description). In order to accomplish this objective, an important aspect to analyze is the *"granularity of uncertainty"*, i.e., the level of discrimination among different counts of uncertainty. Typical values of cardinality used in the linguistic models are odd ones, such as 7 or 9, where the mid term represents an assessment of "approximately 0.5", and with the rest of the terms being placed symmetrically around it [1]. These classical cardinality values seem to fall into line with Miller's observation about that human beings can reasonably manage to bear in mind seven or so items [8].

One possibility of generating the linguistic term set consists of directly supplying the term set by considering all terms distributed on a scale on which a total order is defined [9]. For example, a set of seven terms S, could be given as follows:

$$S = \{s_0 = none, s_1 = very\ low, s_2 = low, s_3 = medium, s_4 = high, s_5 = very\ high, s_6 = perfect\}$$

Usually, in these cases, it is required that in the linguistic term set there exist:

1) A negation operator: $\text{Neg}(s_i) = s_j$ such that $j = g$-i (g+1 is the cardinality).
2) A max operator: $\max(s_i, s_j) = s_i$ if $s_i \geq s_j$.
3) A min operator: $\min(s_i, s_j) = s_i$ if $s_i \leq s_j$.

The semantic of the linguistic terms is given by fuzzy numbers defined in the [0,1] interval. A computationally efficient way to characterize a fuzzy number is to use a representation based on parameters of its membership function [1]. The linguistic assessments given by the users are just approximate ones, some authors consider that linear trapezoidal membership functions are good enough to capture the vagueness of those linguistic assessments. The parametric representation is achieved by the 4-tuple (a, b, d, c), where b and d indicate the interval in which the membership value is 1, with a and c indicating the left and right limits of the definition domain of the trapezoidal membership function [1]. A particular case of this type of representation are the linguistic assessments whose membership functions are triangular, i.e., $b = d$, then we represent this type of membership functions by a 3-tuple (a, b, c). An example may be the following :

$P = Perfect = (.83, 1, 1)$ $VH = Very\_High = (.67, .83, 1)$
$H = High = (.5, .67, .83)$ $M = Medium = (.33, .5, .67)$
$L = Low = (.17, .33, .5)$ $VL = Very\_Low = (0, .17, .33)$
$N = None = (0, 0, .17)$.

Other authors use a non-trapezoidal representation, e.g., Gaussian functions [2].

3 Analysis of the Linguistic Computation Models

The linguistic variables are used in computational processes that imply their fusion, aggregation, comparison, etc. To perform these computations there are different models in the literature. Here we briefly review of them. Afterwards, we shall present a decision-making process to solve a linguistic decision problem using these models.

3.1 Linguistic Computational Model based on the Extension Principle

The Extension Principle is a basic concept in the fuzzy sets theory [5] which is used to generalize crisp mathematical concepts to fuzzy sets. The use of extended arithmetic based on the Extension Principle [5] increases the vagueness of the results. Therefore, the results (fuzzy sets) obtained by the linguistic aggregation operators based on the Extension Principle are counts of information that usually do not match any linguistic term in the initial term set, so a linguistic approximation process is needed to express the result in the original expression domain. The linguistic approximation process consists of finding a fuzzy set supporting the semantic of the linguistic term set whose meaning is the closest to the meaning of an unlabelled fuzzy set generated by a linguistic aggregation operation based on the Extension Principle. In the literature we can find different ways to make linguistic approximations [1, 4].

A linguistic aggregation operation based on the Extension Principle can be expressed formally as:

$$S^n \xrightarrow{\tilde{F}} F(R) \xrightarrow{app_1} S$$

where $\tilde{\mathrm{F}}$ is an aggregation operator based on the Extension Principle, $app_1(\cdot)$ is a linguistic approximation process and S is the initial term set.

3.2 Linguistic Computational Symbolic Model

On the other hand, a second approach used to operate on linguistic information is the symbolic one [3], that makes direct computations on labels. These methods use the order and properties of such linguistic assessments to perform the computations. Usually, they use the ordered structure of the linguistic term sets, $S = \{s_0, \ldots, s_g\}$ where $s_i < s_j$ iff $i < j$, to make the computations [3]. The intermediate results are numerical values, $\alpha \in [0, g]$, which must be approximated in each step of the process by means of an approximation operator $app_2(\cdot)$ to obtain a value, $app_2(\alpha) \in \{0, 1, \ldots, g\}$, that indicates the index of the associated linguistic term, $s_{app_2(\alpha)} \in S$.

Formally, the symbolic aggregation is:

$$S^n \xrightarrow{C} [0, g] \xrightarrow{app_2} \{0, ..., g\} \longrightarrow S$$

where C is a symbolic aggregation operator, $app_2(\cdot)$ is an approximation operation.

3.3 Example

Here, we propose a simple decision-making process for solving a linguistic decision problem using the two models we have just reviewed.

A. **Linguistic Decision Problem**

A distribution company needs to renew its computing system, so it contracts a consulting company to carry out a survey of the different possibilities existing on the market, to decide which is the best option for its needs. The alternatives are the following:

(a) x_1 is a UNIX based system,
(b) x_2 is a Windows-NT based system,
(c) x_3 is an AS/400 based system,
(d) x_4 is a VMS based system.

The consulting company has a group of four consultancy departments

1. p_1 is the cost analysis department,
2. p_2 is the systems analysis department,
3. p_3 is the risk analysis department,
4. p_4 is the technology analysis department.

Each deparment provides a performance vector expressing its preferences for each alternative. The performance values are assessed in the linguistic term set $S = \{N, VL, L, M, H, VH, P\}$ (defined in section 2). In this case all the deparments have the same degree of importance in the decision process. The performance vectors provided by the departments are:

		alternatives			
	L_{ij}	x_1	x_2	x_3	x_4
experts	p_1	VL	M	M	L
	p_2	M	L	VL	H
	p_3	H	VL	M	M
	p_4	H	H	L	L

where "L_{ij}" is the performance value provided by the expert i to the alternative j, whose membership functions are of the triangular type $C_{ij} = (a_{ij}, b_{ij}, c_{ij})$.

B. **Selection Model**

The selection model we shall use to solve the above problem has the following steps:

1. The performance values provided by experts are aggregated to obtain a *collective performance value* on each alternative, i.e., a collective performance vector.

2. A selection process over the collective performance vector is applied. We select the best alternatives that will be those with the maximum value in the collective performance vector.

We shall implement this selection model with the two above computational techniques.

C. **Solution based on the Extension Principle**

Collective performance vector. We shall use the arithmetic mean as aggregation operator, assuming equal importance for each expert ($w_i = 0.25 \quad i = 1,...,4$). Afterwards, a collective performance value for each alternative "x_j" is obtained using:

$$C_j = (\sum_{i=1}^{m} w_i a_{ij}, \sum_{i=1}^{m} w_i b_{ij}, \sum_{i=1}^{m} w_i c_{ij})$$

$$C_1 = ((.25 * 1.33), (.25 * 2.01), (.25 * 2.66)) = (.33, .5, .66)$$
$$C_2 = ((.25 * 1), (.25 * 1.67), (.25 * 2.33)) = (.25, .42, ..58)$$
$$C_3 = ((.25 * .83), (.25 * 1.5), (.25 * 2.17)) = (.21, .38, .54)$$
$$C_4 = ((.25 * 1.17), (.25 * 1.83), (.25 * 2.5)) = (.3, .45, .625)$$

These fuzzy sets do not exactly match with any linguistic term in S, therefore we must apply to them a linguistic approximation process to obtain the results in the initial term set S. For selecting a term for the fuzzy set C_j, we shall use a linguistic approximation process ($app_1(\cdot)$) based on the euclidean distance:

$$d(s_l, C_j) = \sqrt{P_1(a_l - a_j)^2 + P_2(b_l - b_j)^2 + P_3(c_l - c_j)^2}$$

$$s_l = (a_l, b_l, c_l) \in S \qquad C_j = (a_j, b_j, c_j)$$

where P_1, P_2, P_3 are weights representing the importance of a, b, c. So $app_1(C_j)$ chooses s_l^*, such that,

$$d(s_l^*, C_j) \leq d(s_l, C_j) \quad \forall s_l \in S$$

This linguistic approximation process is applied to the above fuzzy sets, with $P_1 = 0.2,\ P_2 = 0.6,\ P_3 = 0.2$:

$$app_1(C_1) = d(s_l^*, C_1) = M \quad app_1(C_2) = d(s_l^*, C_2) = M$$
$$app_1(C_3) = d(s_l^*, C_3) = L \quad app_1(C_4) = d(s_l^*, C_4) = M$$

The main problem of the methods based on the Extension Principle is the loss of information. Therefore the solution sets obtained are not very precise.

Selection process. The alternatives with the highest collective performance value are "$\{x_1, x_2, x_4\}$". This is not a good solution, because due to the lack of precision presented by this method, we are not able to choose only one alternative.

D. **Solution based on Symbolic Methods**

Collective performance vector. We shall use the Convex Combination [3] as aggregation operator, the weighting vector will be $\{.25, .25, .25, .25\}$, so the collective performance values are:

$$C^4(\{VL, M, H, H\}, \{.25, .25, .25, .25\}) = M$$
$$C^4(\{M, L, VL, H\}, \{.25, .25, .25, .25\}) = M$$
$$C^4(\{M, VL, M, L\}, \{.25, .25, .25, .25\}) = L$$
$$C^4(\{L, H, M, L\}, \{.25, .25, .25, .25\}) = M$$

Selection process. The alternatives with the highest collective performance are "$\{x_1, x_2, x_4\}$". Here again we find a multiple alternative solution set, that coincides with the above solution. Symbolic methods present a loss of information as well, in this case it is caused by the "*round*" operator used by the Convex Combination.

Both computational models present loss of information, caused by the need to express the results in the initial expression domain that is discrete. In the following section we propose a continuous linguistic representation model that can express any counting of information although it does not exactly match any linguistic term.

4 A 2-tuple Linguistic Representation Model based on the Symbolic Translation

To develop this model we shall define the concept of Symbolic Translation and use it to represent the linguistic information by means of 2-tuples, (s, α), where s is a linguistic term and α is a number supporting the value of the Symbolic Translation. Together with this representation model we shall present a computational technique to deal with linguistic 2-tuples without loss of information.

4.1 The Symbolic Translation. A 2-tuple Linguistic Representation

Let $S = \{s_0, ..., s_g\}$ be a linguistic term set, a symbolic method aggregating linguistic information obtains a value $\beta \in [0, g]$, if $\beta \notin \{0, ..., g\}$ then an approximation process ($app_2(\cdot)$) is used to express the index of the result in S.

Definition 1. *The symbolic translation is a numerical value assessed in* $[-.5, .5)$ *that supports the "difference of information" between a counting of information* β *assessed in* $[0, g]$ *and the closest value in* $\{0, ..., g\}$ *that indicates the index of the closest linguistic term in* S.

From this concept we shall develop a linguistic representation model which represents the linguistic information by means of 2-tuples (r_i, α_i), $r_i \in S$ and $\alpha_i \in [-.5, .5)$. r_i represents the linguistic label center of the information and α_i is a number that supports the value of the Symbolic Translation.

We have to define how to convert a classical linguistic term into an equivalent 2-tuple, and how to build a 2-tuple from a value, β, that does not exactly express the information of a linguistic term.

Definition 2. *Let $s_i \in S$ be a linguistic term, then its equivalent 2-tuple representation is obtained by means of the function θ as:*

$$\theta : S \longrightarrow (S \ x \ [-0.5, 0.5))$$
$$\theta(s_i) = (s_i, 0) \ / \ s_i \in S$$

Definition 3. *Let $S = \{s_0, ..., s_g\}$ be a linguistic term set and $\beta \in [0, g]$ a value supporting the result of a symbolic aggregation, then the 2-tuple that expresses the equivalent information to β is obtained with the following function:*

$$\Delta : [0, g] \longrightarrow Sx[-0.5, 0.5)$$

$$\Delta(\beta) = \begin{cases} s_i & i = round(\beta) \\ \alpha = \beta - i & \alpha \in [-.5, .5) \end{cases}$$

where round is the usual round *operation, s_i has the closest index label to "β" and "α" is the value of the Symbolic Translation.*
Example. Given a symbolic aggregation operation over labels assessed in $S = \{s_0, s_1, s_2, s_3, s_4, s_5, s_6\}$ that obtains as its result $\beta = 2.8$, then its equivalent 2-tuple will be: $\Delta(2.8) = (s_3, -0.2)$.

Graphically it is represented in Figure 1.

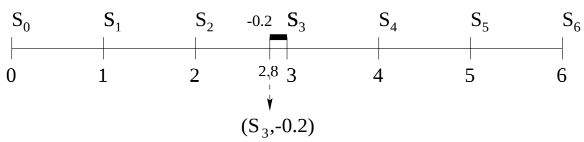

Figure 1: Symbolic Translation Computation

Definition 4.*Let $S = \{s_0, ..., s_g\}$ be a linguistic term set and (s_i, α) be a 2-tuple. There is always a Δ^{-1} function, such that, from a 2-tuple it returns its equivalent numerical value $\beta \in [0, g]$.*

$$\Delta^{-1} : Sx[-.5, .5) \longrightarrow [0, g]$$
$$\Delta^{-1}(s_i, \alpha) = i + \alpha = \beta$$

4.2 Linguistic Computational Model Based on the Symbolic Translation

In this subsection, we present a computational technique to operate with linguistic 2-tuples without loss of information.

1. Comparison of 2-tuples

We shall introduce how to compare linguistic information represented by 2-tuples.

Let (s_k, α_1) and (s_l, α_2) be two 2-tuples, each one representing a counting of information, then

- if $k < l$ then (s_k, α_1) is smaller than (s_l, α_2)
- if $k = l$ then

  1 if $\alpha_1 = \alpha_2$ then (s_k, α_1), (s_l, α_2) represents the same information

  2 if $\alpha_1 < \alpha_2$ then (s_k, α_1) is smaller than (s_l, α_2)

  3 if $\alpha_1 > \alpha_2$ then (s_k, α_1) is bigger than (s_l, α_2)

2. Aggregation of 2-tuples

The aggregation of information consists of obtaining a value that summarizes a set of values.

Let $\{(s_i, \alpha_1), ..., (s_k, \alpha_n)\}$ be a set of 2-tuples to be aggregated, then the aggregation of linguistic 2-tuples can be performed as:
a) All the 2-tuples are transformed into their equivalent numerical values by means of Δ^{-1}, then a set of numerical values $\{\beta_1, ..., \beta_n\}$ to be aggregated are obtained.
b) A numerical aggregation operator (the mean, the weighting average, ...) is used to aggregate $\{\beta_1, ..., \beta_n\}$, obtaining an aggregated value β.
c) $\Delta(\beta)$ obtains its correspondent 2-tuple.

3. Negation operator of a 2-tuple

We define the negation operator on 2-tuples as:

$$Neg((s_i, \alpha)) = \Delta(g - (\Delta^{-1}(s_i, \alpha)))$$

where $g + 1$ is the cardinality of S, $S = \{s_0, ..., s_g\}$.

In [7] can be found a set of aggregation operators based on the 2-tuple fuzzy linguistic representation and these computational techniques.

5 Example

We use the 2-tuple linguistic representation model to solve the decision-making problem presented in the subsection 3.3. The performance vectors of the experts are transformed into 2-tuples using the θ function:

		alternatives			
		x_1	x_2	x_3	x_4
experts	p_1	$(VL, 0)$	$(M, 0)$	$(M, 0)$	$(L, 0)$
	p_2	$(M, 0)$	$(L, 0)$	$(VL, 0)$	$(H, 0)$
	p_3	$(H, 0)$	$(VL, 0)$	$(M, 0)$	$(M, 0)$
	p_4	$(H, 0)$	$(H, 0)$	$(L, 0)$	$(L, 0)$

First, we obtain a collective perfomance vector aggregating the above 2-tuples, using the arithmetic mean $\overline{x}$ (all the experts have the same degree of importance):

$$\Delta(\overline{x}(1,3,4,4)) = (M, 0) \quad \Delta(\overline{x}(3,2,1,4)) = (M, -.5)$$
$$\Delta(\overline{x}(3,1,3,2)) = (L, .25) \quad \Delta(\overline{x}(2,4,3,2)) = (M, -.25)$$

The collective performance vector is:

$$\{(M,0),(M,-.5),(L,.25),(M,-.25)\}$$

Now we obtain the solution set of alternatives. In this case is "$\{x_1\}$" that is the collective value with the maximum 2-tuple. Therefore the distribution company will receive a survey where the best computing system for their needs is the "UNIX based system".

5.1 Comparative Analysis

Throughout this contribution we have solved a decision problem using three different methods, obtaining the following results:

	Dominance Degree				*Sol. Set*
	x_1	x_2	x_3	x_4	
EP	M	M	L	M	$\{x_1,x_2,x_4\}$
SM	M	M	L	M	$\{x_1,x_2,x_4\}$
2-t	(M,0)	(M,-.5)	(L,-.25)	(M,-.25)	$\{x_1\}$

Table I. Results using the three methods

From *Table I* we notice that all the methods, even the method using the 2-tuple representation, obtain similar linguistic values for the dominance degrees which indicates the correctness of the results and hence of the methods. However, an important difference appears in the *"Solution Set"* column in *Table I*, the solution set obtained by the 2-tuple method is more precise (is a subset) than the sets obtained by the other ones. This is because collective values obtained by the methods of the first two rows of *Table I* are discrete, then when different alternatives have the same linguistic value as the dominance degree we cannot discern which is better than the others. Using the 2-tuple representation the collective values are continuous, therefore if several alternatives have the same linguistic term but a different value for the Symbolic Translation we can choose the best among them.

6 Concluding Remarks

We have seen that the fuzzy linguistic approach has an important limitation due to its model for representing linguistic terms as discrete values. In this contribution we have presented a 2-tuple linguistic representation model based on the Symbolic Translation. It represents the information by means of 2-tuples, which are composed by a linguistic term and a numerical value that supports the value of the Symbolic Translation. In this way, the linguistic information is presented as continuous instead of discrete minimizing the loss of information.

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] G. Bordogna and G. Passi, A Fuzzy Linguistic Approach Generalizing Boolean Information Retrieval: A Model and Its Evaluation, *J. of the American Society for Information Science* **44** (1993) 70-82.

[3] M. Delgado, J.L. Verdegay and M.A. Vila, On Aggregation Operations of Linguistic Labels, *Int. J. of Intelligent Systems* **8** (1993) 351-370.

[4] R. Degani and G. Bortolan, The problem of Linguistic Approximation in Clinical Decision Making, *Int. Journal of Approximate Reasoning* **2** (1988) 143-162

[5] D. Dubois and H. Prade, Fuzzy Sets and Systems: Theory and Applications, *Academic*, New York, 1980.

[6] F. Herrera and E. Herrera-Viedma, Linguistic Decision Analisys:Steps for Solving Decision Problems under Linguistic Information, *Fuzzy Sets and Systems* (1998, To appear)

[7] F. Herrera and L. Martínez, A 2-tuple Fuzzy Linguistic Representation Model for Computing with Words, Dept. Computer Sciences and A.I., Granada University, Technical Report#DECSAI-990102, (1999)

[8] G.A. Miller, The Magical Number Seven or Minus Two: Some limits On Our Capacity of Processing Information, *Psychological Rev.* **63** (1956) 81-97.

[9] R.R. Yager, An Approach to Ordinal Decision Making, *Int. J. of Approximate Reasoning*, **12** (1995) 237-261.

[10] 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.