A Methodology for Generating the Semantics of Unbalanced Linguistic Term Sets

Francisco Herrera, Enrique Herrera-Viedma, Luis Martínez, Pedro J. Sánchez *†‡

Abstract

The use of linguistic information, in the resolution of any problem, implies the choice of the linguistic term set with its semantics. In this contribution we shall focus in the choice of the semantics of the linguistic terms belonging to unbalanced linguistic term sets, i.e., linguistic term sets whose linguistic terms are not equally informative or symmetrically distributed. We shall present a methodology to establish the semantics of these terms. This process will use the concept of linguistic hierarchy as representation basis of the unbalanced linguistic information.

Keywords: Linguistic variables, unbalanced term sets, semantics, computing with words

1 Introduction

In those problems dealing with aspects related to human beings, qualitative aspects, is difficult to assess them by means of precise numbers. Then, the use of the fuzzy linguistic approach [10] has obtained successful results in problems of different areas dealing with qualitative aspects [1, 2, 5, 6, 8, 9].

In any problem that uses linguistic information the first goal to satisfy is the choice of the linguistic term set with its semantics, for establishing the linguistic term set used in the problem. In the literature can be found different possibilities for choosing the linguistic descriptors and their semantics (see [2]).

For defining the semantics of the linguistic descriptors, can be assumed that all the terms of the term set are equally informative, i.e., symmetrically distributed as in [1, 8] or can be unbalanced term sets, i.e., the terms of the linguistic term set are not equally informative [3, 7] (e.g., Fig. 1).

In this contribution we shall focus in *Unbalanced Term Sets*. We shall present a process to define the semantics of the labels of these term sets.

*F. Herrera and E. Herrera-Viedma are with the Dept. of Computer Science and A.I., University of Granada, 18071 - Granada, Spain. E-mail: herrera,viedma@decsai.ugr.es

† L. Martínez and Pedro J. Sánchez are with the Dept. of Computer Science, University of Jaén, 23071 - Jaén. E-mail: martin,pedroj@ujaen.es

‡This work is supported by the Research Projects TIC2002-03348, TIC2002-03276 and FEDER Fonds.

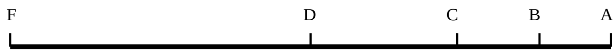

Figure 1: Unbalanced Term Set

To do so, we shall use the concept of linguistic hierarchy as representation basis of the unbalanced linguistic information.

In order to do that, this contribution is structured as follows: Section 2 reviews the linguistic hierarchical contexts that will be used in Section 3 that presents a mechanism for generating the semantics of linguistic terms of an unbalanced linguistic term set. And finally some concluding remarks are pointed out.

2 Linguistic Hierarchies

The hierarchical linguistic structure was used in [5] to improve the precision in the processes of computing with words, CW, in linguistic multi-granular contexts. In this paper, we use them to manage unbalanced linguistic term sets.

A *linguistic hierarchy* is a set of levels, where each level is a linguistic term set with different granularity from the remaining of levels of the hierarchy. Each level belonging to a linguistic hierarchy is denoted as $\mathbf{l}(\mathbf{t}, \mathbf{n}(\mathbf{t}))$, being:

1. t, indicates the level of the hierarchy,
2. $n(t)$, the granularity of the linguistic term set of the level t.

We assume levels containing linguistic terms whose membership functions are triangular-shaped, symmetrical and uniformly distributed in $[0, 1]$. In addition, the linguistic term sets have an odd number of elements.

The levels belonging to a linguistic hierarchy are ordered according to their granularity, i.e., for two consecutive levels t and $t + 1$, $n(t + 1) > n(t)$. This provides a linguistic refinement of the previous level.

From the above concepts, we define a linguistic hierarchy, LH, as the union of all levels t:

$$LH = \bigcup_t l(t, n(t))$$

Given a LH, we denote as $S^{n(t)}$, the linguistic term set of LH corresponding to the level t of LH with a granularity of uncertainty of $n(t)$:

$$S^{n(t)} = \{s_0^{n(t)}, ..., s_{n(t)-1}^{n(t)}\}$$

Generically, we can say that the linguistic term set of level $t+1$, $S^{n(t+1)}$, is obtained from its predecessor, $S^{n(t)}$, as:

$$l(t, n(t)) \rightarrow l(t+1, 2 \cdot n(t) - 1) \qquad (1)$$

A graphical example of a linguistic hierarchy is shown in Fig. 2:

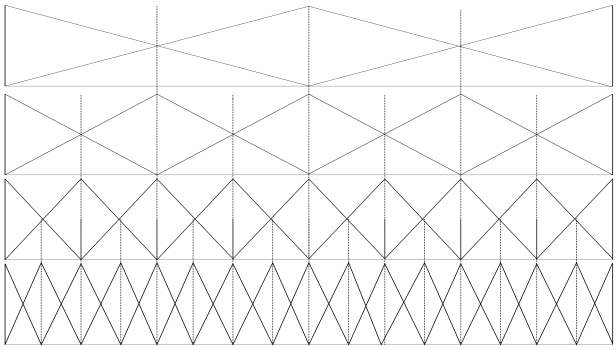

Figure 2: Linguistic Hierarchy of 3,5, 9 and 17 labels

3 Generating the Semantics of Unbalanced Linguistic Term Sets

The aim of this paper is to design a mechanism for generating the semantics for labels of an unbalanced linguistic term set, S, by means of linguistic hierarchies. This mechanism may be used whenever the unbalanced term set has a central label and the remaining ones are distributed around it, on the left and right sides.

3.1 Basic Ideas

First of all, we must choose a Linguistic Hierarchy, LH, to select over its levels the linguistic terms of the unbalanced term set, S. The linguistic terms of S will depend on the LH chosen.

We consider that the label set, S, is divided in three subsets,

$$S = S_L \cup S_C \cup S_R$$

- S_L is the subset with labels less than central label, and $\#S_L$ is its number of labels.
- S_C is a subset with the central label, i.e., $\#S_C = 1$. This label is builded joining the upside, $\overline{s}$, and the downside, $\underline{s}$, of two labels, $s_i = (a_i, b_i, c_i)$ and $s_j = (a_j, b_j, c_j)$, such that, $b_i = b_j$:

$$s_c = \overline{s_i} \cup \underline{s_j} = (a_i, b_i, c_j), \quad a_i \leq b_i \leq c_j.$$

- S_R is the subset with labels higher than central label, being $\#S_R$ the number of labels of S_R.

3.2 Algorithmic Process

Following, we present the algorithmic process for generating from a *Linguistic Hierarchy*, LH, the semantics of the terms of the unbalanced term set, S. We distinguish two possibilities:

1. The simple case that:

$$\exists l(t, n(t)) \in LH \;/\; \frac{n(t)-1}{2} = \begin{cases} \#S_R, & or \\ \#S_L \end{cases}$$

   We generate the labels for S_R as:

   (a) To assign the labels from $S_R^{n(t)}$ to S_R

$$S_R \leftarrow S_R^{n(t)}$$

   (b) To assign the downside of the central label

$$\underline{s_C} \leftarrow \underline{s_C^{n(t)}}$$

   And for S_L as:

   (a) To assign the labels from $S_L^{n(t)}$ to S_L

$$S_L \leftarrow S_L^{n(t)}$$

   (b) To assign the upside of the central label

$$\overline{s_C} \leftarrow \overline{s_C^{n(t)}}$$

2. In other cases, the generation of the labels depends upon the distribution of the unbalanced term set, S. The distribution of the label set is given by a set of five values:

$$\{(\#S_L, density_{S_L}), 1, (\#S_R, density_{S_R})\},$$

   - $density_{S_L}$ a label assessed in the term set $\{middle, extreme\}$, that indicates if the higher granularity of the left side is near of the central label or near of the left extreme.
   - $density_{S_R}$ equivalent to $density_{S_L}$ on the right side.

We present the process for the right subset, S_R, the process for the left subset, S_L, is symmetrically analogous to the right one. Here, we use graphics based on the generation method over S_R of the unbalanced term set shown in Fig. 1, for explaining the process.

Right Subset: $\mathbf{S_R}$

To find two levels t and $t+1$, such that,

$$\frac{n(t)-1}{2} < \#S_R < \frac{n(t+1)-1}{2}$$

The labels of the S_R are chosen from the assignable labels of the right subsets of levels t and $t+1$. These assignable labels subsets vary along the assignment process. Initially (Fig. 3):

(a) The assignable subset for level t is:

$$AS_R^{n(t)} = \{s^{n(t)}_{\frac{n(t)-1}{2}+1}, ..., s^{n(t)}_{n(t)-1}\},$$

(b) and the assignable subset for level $t+1$:

$$AS_R^{n(t+1)} = \{s^{n(t+1)}_{\frac{n(t+1)-1}{2}+1}, ..., s^{n(t+1)}_{n(t+1)-1}\}$$

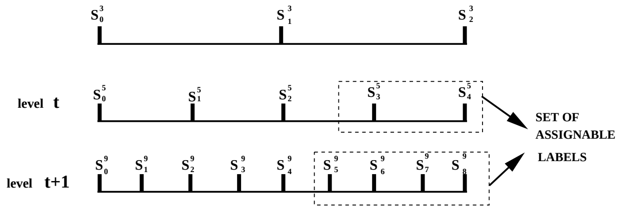

Figure 3: Ordered sets of assignable labels.

To assign the labels of S_R we distinguish two subsets:

$$S_R = S_{RC} \cup S_{RE}$$

- S_{RE} has assigned the labels close to the right extreme of S, and
- S_{RC} has assigned the labels close to the central label of S.

The generation method acts as follows:

IF $density_{S_R}$= extreme
THEN
$S_{RE} \subset AS_R^{n(t+1)}$ (max density: *extreme*)
$S_{RC} \subset AS_R^{n(t)}$
ELSE
$S_{RE} \subset AS_R^{n(t)}$
$S_{RC} \subset AS_R^{n(t+1)}$ (max density: *middle*)

We must remark that the method takes into account the rule for building LH shown in equation 1: *"Every label of level **t** has associated two labels of level **t+1**"*. Therefore, when the method assigns a label from $AS_R^{n(t+1)}$ eliminates its associated label from $AS_R^{n(t)}$ (see Fig. 4).

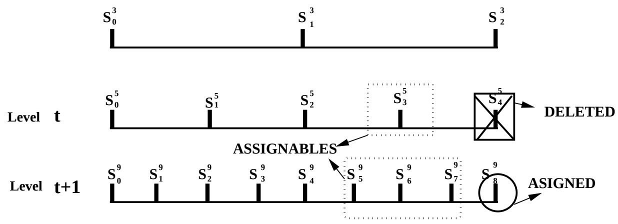

Figure 4: Generation Method. Right Side

To assign labels from $AS_R^{n(t+1)}$ and $AS_R^{n(t)}$ to S_{RE} and S_{RC} we define the functions:

1. $\mathbf{assign_{t+1}}(\mathbf{AS_R^{n(t+1)}}, \mathbf{density}, \mathbf{S_{subset}})$: deletes the first label, starting from *density*, from $AS_R^{n(t+1)}$ and assigns it into S_{subset}. Finally, if its associated label belongs to $AS_R^{n(t)}$ then is eliminated too. Where $density \in \{middle, extreme\}$ and $subset \in \{RC, RE\}$ (see Figs. 5 and 6).

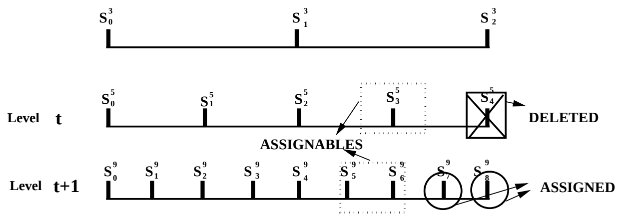

Figure 5: Assignment with $density_{S_R} = extreme$.

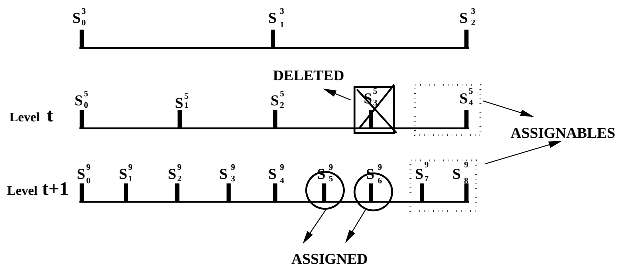

Figure 6: Assignment with $density_{S_R} = middle$.

2. $\mathbf{assign_t}(\mathbf{AS_R^{n(t)}}, \mathbf{S_{subset}})$: assigns the existing labels in $AS_R^{n(t)}$ into S_{subset}.

After the right side has been assigned it is assigned the downside of the central label of S, S_C, from the central label of the term set whose level was assigned the most left label on the right side:

if density = *extreme* then $\underline{s_c} \leftarrow \underline{s_c^{n(t)}}$
if density = *middle* then $\underline{s_c} \leftarrow \underline{s_c^{n(t+1)}}$

The algorithm that implements the generation method is shown in Table 1.

IF $density_{S_R}$= extreme
THEN
  REPEAT
    $assign_{t+1}(AS_R^{n(t+1)}, extreme, S_{RE})$
  UNTIL $(\#S_R - \#S_{RE}) = \#AS_R^{n(t)}$
  $assign_t(AS_R^{n(t)}, S_{RC})$
  $\underline{s_C} \leftarrow \underline{s_C^{n(t)}}$
ELSE
  REPEAT
    $assign_{t+1}(AS_R^{n(t+1)}, middle, S_{RC})$
  UNTIL $(\#S_R - \#S_{RC}) = \#AS_R^{n(t)}$
  $assign_t(AS_R^{n(t)}, S_{RE})$
  $\underline{s_C} \leftarrow \underline{s_C^{n(t+1)}}$

ENDIF

Table 1: Generation Method. Algorithm

According to this method the semantics assigned to the labels of Fig. 1 are the following ones:

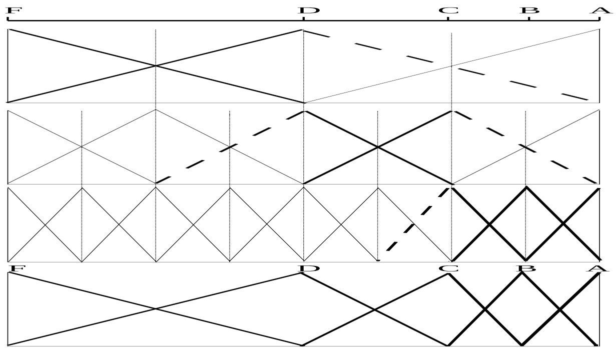

Figure 7: Semantics of the Grading System

4 Concluding Remarks

We have presented a method based on the use of linguistic hierarchies for establishing the semantics of the labels belong to unbalanced term sets. The semantics obtained with this method allows to use linguistic unbalanced term sets in processes of Computing with Words without loss of information, by means of the transformation functions between labels from different levels of a linguistic hierarchy presented in [5] and the linguistic 2-tuple computational model [4]. This computational model over unbalanced linguistic term sets will be explored and discussed in detail during the presentation.

References

[1] G. Bordogna, M. Fedrizzi, and G. Pasi. A linguistic modeling of consensus in group decision making based on OWA operators. *IEEE Trans. on Systems, Man and Cybernetics, Part A: Systems and Humans*, 27:126–132, 1997.

[2] F.Herrera and E. Herrera-Viedma. Linguistic decision analysis: Steps for solving decision problems under lingusitic information. *Fuzzy Sets and Systems*, 115:67–82, 2000.

[3] F. Herrera, E. Herrera-Viedma, and L. Martínez. A hierarchical ordinal model for managing unbalanced linguistic term sets based on the linguistic 2-tuple model. In *Proceedings of the Eurofuse Workshop on Preference Modelling and Applications*, pages 201–206, Granada (Spain), April 2001.

[4] F. Herrera and L. Martínez. Computing with words using the 2-tuple linguistic representation model: Analysis of consistency and description. In *Proceedings of the Eighth International Conference on Information Processing and Management of Uncertainty in Knowledge-Bases Systems. IPMU'2000*, volume Vol II, pages 765–772, Madrid(Spain), 2000.

[5] F. Herrera and L. Martínez. A model based on linguistic 2-tuples for dealing with multigranularity hierarchical linguistic contexts in multi-expert decision-making. *IEEE Trans. on Systems, Man and Cybernetics. Part B: Cybernetics*, 31(2):227–234, 2001.

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

[7] V. Torra. Negation function based semantics for ordered linguistic labels. *Int. Journal of Intelligent Systems*, 11:975–988, 1996.

[8] R.R. Yager. Non-numeric multi.criteria multi-person decision making. *Group Decision and Negotation*, 2:81–93, 1993.

[9] R.R. Yager. An approach to ordinal decision making. *International Journal of Approximate Reasoning*, 12:237–261, 1995.

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