Linguistic Modifiers with Unbalanced Term Sets in Multi-valued Logic

Nouha Chaoued, Amel Borgi, Anne Laurent

Linguistic Modifiers with Unbalanced Term Sets in Multi-valued Logic

Abstract: Modeling human knowledge by machines should be as faithful as possible to reality. Therefore, it is imperative to take account of inaccuracies and uncertainties in this knowledge. This problem has been dealt with through different approaches. The most two common approaches are fuzzy logic and multi-valued logic. These two logics propose a linguistic term modeling. Generally, problems modeling qualitative aspect use linguistic variables assessed in linguistic terms that are uniformly distributed on the scale. However, in many cases, linguistic information needs to be defined by unbalanced term sets whose terms are not uniformly and/or not symmetrically distributed. In the literature, many researchers deal with these term sets in the context of fuzzy logic. Thereby, in our work, we introduce a new approach to represent and treat such term sets in the context of multi-valued logic. First, we propose an algorithm that allows representing terms within an unbalanced set. Then, we describe a second algorithm that permits the use of linguistic modifiers within unbalanced multi-sets.

1 INTRODUCTION

Knowledge handled by humans is often imperfect. These imperfections may be due to ambiguity, incompleteness, imprecision, uncertainty, inconsistency, etc. Several approaches were suggested in the literature for such knowledge representation and treatment. The most known are fuzzy logic (Zadeh, 1965) and multi-valued logic (De Glas, 1989, Akdag et al., 1992).

We notice that humans are able to perform reasoning without any exact measurements. They mostly use abstract terms of natural language (*young*, *old*, *mature*, etc.) and symbolic data rather than numerical values or qualitative ones. Terms can also be composed by using adverbs, such as *little*, *more or less* and *slightly*. Both fuzzy logic and multi-valued logic propose a linguistic term modeling to allow using words in reasoning process. They use linguistic variables that take values in a set of linguistic terms (Zadeh, 1975). These latter express the various nuances of the processed information using words.

Generally in Knowledge-Based Systems, experts use linguistic terms that are uniformly and symmetrically distributed on a scale. However, in some cases, we need to assess qualitative aspects by means of variables using linguistic term sets which are not uniformly distributed. For example, in the evaluation process, we often consider a single negative term, i.g. *Fail*, and many positive terms such as *Medium*, *Good*, *Excellent*, etc. The gap between these terms is unequal.

In this paper, we focus on multi-valued logic. It allows to symbolically represent imprecise knowledge using ordered adverbial expressions of natural language (De Glas, 1989). We have noticed that in the context of this logic, few studies have treated unbalanced linguistic term sets (Abchir, 2013, Chaoued and Borgi, 2015). The aim of this paper is to establish a methodology to represent and manage this kind of data. It is based on the work of (Chaoued and Borgi, 2015) that express unbalanced terms using a uniform multi-set. In the present work, we apply our proposal to an Information Retrieval System (IRS). The latter aims to retrieve a set of documents that satisfies a user query. This system is composed of three units (Herrera-Viedma and López-Herrera, 2007):

1. A **documentary archive** or a database including a set of documents. They are represented by means of index terms describing their subject content.

2. A **query subsystem** presenting user needs by means of weighted queries. It indicates the topics that he/she is asking for.

3. An **evaluation subsystem** allowing the evaluation of documents according to their relevance compared to the user query.

Usually, users are interested by documents whose contents are the most relevant to their queries. This implies the use of more precise labels in the left (positive) interval than in the right (negative) one (Herrera-Viedma and López-Herrera, 2007). Thus, in IRS the use of non-uniformly distributed term sets is recommended. Indeed, it is more appropriate to use such kind of multi-set to represent the relevance degrees of the documents or to express the weights of index terms in the queries. To achieve this, we will use the unbalanced set $S_{un} =$

$\{None, Low, Medium, High, QuiteHigh, VeryHigh, Total\}$ = $\{N, L, M, H, QH, VH, T\}$ (Herrera-Viedma and López-Herrera, 2007)(Fig. 1).

Figure 1 – Unbalanced linguistic term set

Two aspects are discussed in the present work. The first concerns the representation of terms within unbalanced multi-set. The second deals with the management of such kind of knowledge. Fig. 2 illustrates the different steps of this process. In the first one, we apply the single scale algorithm (Chaoued and Borgi, 2015) to express a term (L in the figure) using a uniform multi-set (L is represented by τ_2). Afterwards, an existing tool, as aggregation operator or modifiers, is applied to the obtained term (τ_2 is modified into τ_3). The last step aims to express the result of the computational phase with a term from the initial multi-set (Approximately L in the figure).
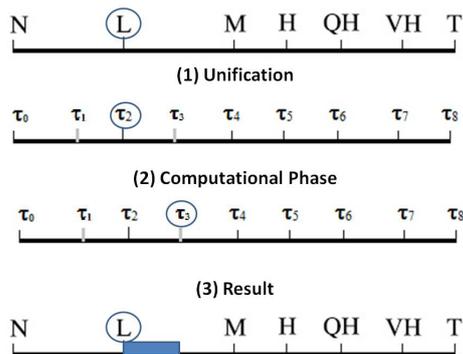

Figure 2 – Management of unbalanced linguistic term set

This article is organized as follows. We start, in section 2, with an explanation of what is unbalanced term sets. In section 3, we introduce the basic concepts of multi-valued logic and of linguistic modifiers. Existing works that concern the representation of unbalanced multi-sets are presented in section 4. Then, section 5 introduces our approach to express terms within an unbalanced set (Fig. 2 Step 3). Finally, in section 6, we propose a new way to use Generalized Symbolic Modifiers with unbalanced linguistic terms (Fig. 2 Step 2).

2 PRELIMINARIES

Most Knowledge-Based Systems use linguistic sets with terms that are uniformly and symmetrically distributed. However, there are other cases that need to assess qualitative aspect by means of variables using linguistic terms which are not uniformly and/or symmetrically distributed, named unbalanced linguistic sets. In many real-life situations, these latter are used as in project investment, negotiation process, evaluation process, etc. Asymmetric linguistic information can be a consequence of the nature of the linguistic variables involved in the problem such as personal examination or evaluation system (Fig. 3) (Martínez and Herrera, 2012). Terms are not equidistant, e.g. the distance between *Poor* and *Average* is greater than between *Average* and *Good*. This difference indicates the expert's interest in having more precise definition of a part of the domain, that leads to the use of more labels in this interval.
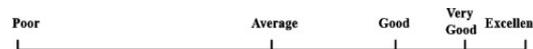

Figure 3 – Set of 5 linguistic terms not uniformly distributed

Moreover, in many problems, several decision makers are involved. In fact, it is more reliable to obtain a decision based on the opinion of several experts than on a single one. In a multi-experts decision making process, each expert may assess his knowledge with a particular scale having a specific granularity (Fig. 4) (Martínez and Herrera, 2012). It indicates their different knowledge backgrounds or judging abilities. In each used linguistic set, terms are uniformly distributed.

Figure 4 – Fuzzy multi-granular scales used by experts

Herrera et al. (Francisco Herrera and Martínez, 2008) considered that an unbalanced fuzzy set is a set with a minimum term, a maximum term and a central term s_c called also midpoint. The remaining terms are not uniformly distributed and asymmetrically on both side of central term :

$$S = S_L \cup S_C \cup S_R \qquad (1)$$

With :

- S_L: the subset including terms on the left of the central term s_c and $\#(S_L)$ its cardinality;

- S_C: the singleton containing the term s_c;
- S_R: the subset including terms on the right of the central term s_c and $\#(S_R)$ its cardinality.

Hence, S is described by:

$$S = \{(\#(S_L), density_L), 1, (\#(S_R), density_R)\} \quad (2)$$

where 1 is the midpoint. The density corresponds to a high granularity around the central term or the end of fuzzy partition. Therefore, the density value can be *middle* or *extreme*.

For example, the set represented by Fig. 3 will be described by $S = \{(1, extreme), 1, (3, extreme)\}$. In fact, the central term is $\{Average\}$. The left subset includes only the term $\{Poor\}$ which is the minimum term. While the right subset includes $\{Good, Very\ Good, Excellent\}$. These terms are closer to the maximum term *Excellent*. So the density corresponds to *extreme* for the two subset.

Xu (Xu, 2009) proposed the unbalanced set as a set of (2t-1) labels, with t a positive integer, defined by:

$$S^{(t)} = \{s_\beta^{(t)} | \beta = (1-t), \frac{2}{3}(2-t), \frac{2}{4}(3-t), ..., 0,$$
$$..., \frac{2}{4}(t-3), \frac{2}{3}(t-2), (t-1)\} \quad (3)$$

Thus, the central term has index 0 and other terms have positive or negative indices. The main idea of this definition is that the absolute value of the deviation between the indices of two successive terms increase regularly proceeding from the center term to the end of the scale. Fig. 5 (Xu, 2009) illustrates an unbalanced set with $t = 4$. We notice that terms are symmetrically distributed around the central term.
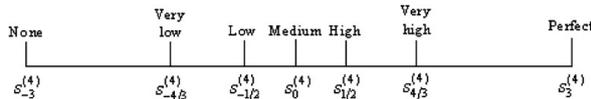

Figure 5 – A set of seven linguistic labels $S^{(4)}$

In this work, we propose to define an unbalanced linguistic term set S_{un} as a set of degrees that includes a minimum degree, *False*, a maximum degree, *True*, and the remaining ones are not necessarily uniformly distributed on the scale: the gaps between adjacent terms may be unequal. Set granularity, i.e. its cardinality, can be odd or even. Each term is defined by its position on the scale. They are supplied directly or by specifying the gap between each successive terms.

To illustrate our proposal, let us use the set represented in Fig. 3. This set will be described using the distance between terms :

- Distance Poor - Average: $d_1 = \frac{1}{2}$

- Distance Average - Good: $d_2 = \frac{1}{4}$
- Distance Good - Very Good: $d_3 = \frac{1}{8}$
- Distance Very Good - Excellent: $d_4 = \frac{1}{8}$

There are two important subjects to treat when dealing with imperfect knowledge: (1) their representation and (2) their management. Many researches had dealt with these term sets in the fuzzy logic context (Herrera-Viedma and López-Herrera, 2007, Martínez and Herrera, 2012, Herrera and Martínez, 2001, Wang et al., 2015, Xu, 2009, Marin et al., 2014, Bartczuk et al., 2012, Jiang et al., 2015). Researchers have proposed approaches for their representation (Martínez and Herrera, 2012, Herrera and Martínez, 2001, Bartczuk et al., 2012) and their management, such as operators for unbalanced aggregation (Marin et al., 2014, Jiang et al., 2015). In the context of multi-valued logic, only Abchir (Abchir, 2013) and Chaoued and Borgi (Chaoued and Borgi, 2015) proposed algorithms to represent unbalanced term sets within uniform set. Our work, developed in the multi-valued context, has two targets: the first one is to propose a new way to represent terms within unbalanced multi-sets; the second intends to define an approach to use linguistic modifiers with such knowledge.

3 MULTI-VALUED LOGIC AND GENERALIZED SYMBOLIC MODIFIERS

Multi-valued logic is based on De Glas's multi-set theory (De Glas, 1989). In this approach, each linguistic term is represented by a multi-set. Knowledge is expressed thanks to an ordered and finite scale of M symbols denoted by (De Glas, 1989, Akdag and Pacholczyk, 1989):

$$L_M = \{\tau_0, \tau_1, ..., \tau_{(M-1)}\}; M \geq 2 \quad (4)$$

τ_i is the membership degree to the multi-set (i $\in$ [0, M-1]).

It should be noted that symbolic degrees are connected only by the total order relation $\leq$, defined by (Adkag, 1992)

$$\tau_\alpha \leq \tau_\beta \iff \alpha \leq \beta; \forall \alpha \ and \ \beta \in [0, M-1] \quad (5)$$

In most existing works on multi-valued logic, these degrees are assumed to be uniformly distributed on the scale. The membership relation in multi-valued logic is partial:

$$x \in_\alpha A \iff x\ belongs\ to\ A\ at\ a\ degree\ \alpha \quad (6)$$

To express the imprecision of a predicate, a qualifier ϑ_α is associated with each degree:

$$x \; is \; \vartheta_\alpha \; A \iff (x \; is \; \vartheta_\alpha \; A) \; is \; true$$
$$\iff "x \; is \; A" \; is \; \tau_\alpha \; true$$

The work of Zadeh (Zadeh, 1972), in fuzzy logic, and more recently those of Akdag et al. (Akdag et al., 2000, Akdag et al., 2001) and Truck (Truck, 2002), in multi-valued logic context, consider that any fuzzy subset or multi-valued symbol can be considered as a modifier of another fuzzy subset or multi-valued symbol respectively. A modifier allows modeling knowledge and gradual reasoning to build new terms from the initial ones, or to compare two values by finding the modifier that allows the transformation from one to another.

In the multi-valued logic, a membership in a multi-set is characterized by a symbolic membership degree τ_i defined on a scale of ordered degrees L_M. Thus, the data modification is a transformation of a degree and/or the scale of the multi-set. Indeed, some linguistic modifiers preserve the same multi-set, but modify the membership degree. Others transform a multi-set towards another one. Hence, it leads to an expansion or erosion of the original scale.

Symbolic linguistic modifiers have been proposed by Akdag et al. (Akdag et al., 2000, Akdag et al., 2001) and they were generalized and formalized by Truck (Truck, 2002, Truck et al., 2002). They were named Generalized Symbolic Modifiers (GSM). According to Truck (Truck, 2002), a GSM (Definition 1) is a triplet of parameters: radius, nature (i.e. dilated, eroded or preserved) and mode (i.e. reinforcing, weakening or central). The radius is denoted by ρ with $\rho \in \mathbb{N}^*$. The higher ρ, the more powerful the modifier.

Definition 1. *(Truck, 2002) Let τ_i be a symbolic degree, such that $i \in \mathbb{N}$, in a scale L_M of M terms $(M \in \mathbb{N}^* \backslash \{1\})$ with $i < M$. Let m be a GSM with radius ρ denoted m_ρ. The modifier m_ρ is a function that performs a linear transformation of τ_i to a new degree $\tau_{i'} \in L_{M'}$ (where $L_{M'}$ is the linear transformation of L_M) according to a radius $\rho : m_\rho(i) = i'$; $m_\rho(M) = M'$*

For each linguistic degree, Akdag et al. (Akdag et al., 2000) associate a numerical rate, i.e. an intensity level. In fact, an item from the multi-set can be considered as a precision degree of a proposition. This latter is the quotient $prop(\tau_i) = \frac{p(\tau_i)}{M-1}$ associated with each τ_i such that $p(\tau_i)$ is its position in the scale. Thus, $Prop(\tau_i)$ is the weight of the degree τ_i relatively

---

$^1 \lfloor . \rfloor$ is the flour function.

to the linguistic set granularity, i.e. its intensity compared to the truth degree $\tau_{(M-1)}$ (True).

Symbolic modifiers are classified as: weakening when the proportion decreases, i.e. $prop(\tau_{i'}) < prop(\tau_i)$ (the four weakening modifiers defined in (Truck, 2002) are $EW_\rho, DW_\rho, DW'_\rho$ and CW_ρ); reinforcing when it increases, i.e. $prop(\tau_{i'}) > prop(\tau_i)$ (the four reinforcing modifiers proposed in (Truck, 2002) are $ER_\rho, DR_\rho, DR'_\rho$ and CR_ρ) or central if $prop(\tau_i)$ does not change, such modifiers may act as a zoom on the base (the four central modifiers presented in (Truck, 2002) are $EC_\rho, EC'_\rho, DC_\rho$ and DC'_ρ). Some examples of GSM are presented in Table 1.

4 REPRESENTING UNBALANCED MULTI-SETS

In the context of multi-valued logic, few works have treated the non-uniformly distributed knowledge. Abchir proposed in the discussion of his thesis (Abchir, 2013) a first approach based on the use of Generalized Symbolic Modifiers (GSM). Chaoued and Borgi also proposed a modified version of the Abchir's algorithm to represent unbalanced degrees on a single uniform scale (Chaoued and Borgi, 2015).

4.1 Abchir's Algorithm

This approach aims to represent numerical input values $v_1, v_2, ...,$ representing the position of terms on a scale, as symbolic degrees of a uniform multi-set L_M. The algorithm starts by calculating a γ coefficient multiplier to transform input values (v_i) to integer ones, denoted val_i, if they are not. Then, each obtained integer value is transformed into a couple (τ_i, L_M) according to the following rule:

$$M = max(2, val_i + 1); \tau_i = \tau_{val_i} \tag{7}$$

The procedure is iterative. It stops when all input values have been treated. The author denotes the first M by M_1. If the next value to treat cannot be integrated in this set, i.e., $val_2 \geq M_1$, a new value M, denoted by M_2, is calculated such that: $M_2 = val_2 + 1$.

Thus, the couple (val_2, M_2) leads to (τ_{val_2}, L_{M_2}). When M changes, previously calculated couples must be represented within the new scale L_{M_2}. The weakening expanding modifier $DW(M_2 - M_1)$ (Table 1) is used to do this.

The process continues for each value, ensuring that the multiplier coefficient γ always generates integer values. If this is not the case, the multiplier γ should be changed in a new γ that allows transforming

Table 1 – Some examples of Generalized Symbolic Modifiers (Truck, 2002)

Mode	Nature	Modifier	Effect
Weakening	Erosion	EW_ρ	$m(i) = \max(0, i - \rho)$ $m(M) = \max(2, M - \rho)$
	Dilatation	DW_ρ	$m(i) = i$ $m(M) = M + \rho$
		DW'_ρ	$m(i) = \max(0, i - \rho)$ $m(M) = M + \rho$
	Conservation	CW_ρ	$m(i) = \max(0, i - \rho)$ $m(M) = M$
Reinforcing	Erosion	ER_ρ	$m(i) = i$ $m(M) = \max(i + 1, M - \rho)$
		ER'_ρ	$m(i) = \min(i + \rho, M - \rho - 1)$ $m(M) = \max(1, M - \rho)$
	Dilatation	DR_ρ	$m(i) = i + \rho$ $m(M) = M - \rho$
	Conservation	CR_ρ	$m(i) = \min(i + \rho, M - 1)$ $m(M) = M$
Central	Erosion	EC_ρ	$m(i) = \max(\lfloor \frac{i}{\rho} \rfloor, 1)$ [1] $m(M) = \max(\lfloor \frac{M}{\rho} \rfloor + 1, 2)$ [1]
	Dilatation	DC_ρ	$m(i) = i\,\rho$ $m(M) = M\,\rho - \rho + 1$

all considered values into integers, its old value will be denoted by γ_{old}. Then, all previous values must be recalculated using the central expanding modifier DC(c) (Table 1). The multiplier c allows to transform γ_{old} into γ. It is obtained as follows:

$$\gamma = c * \gamma_{old} \qquad (8)$$

For this algorithm, neither considered values to treat nor their number is known in advance. Partitioning will be done as the value is specified. The final result remains the same if a change is made in input terms order. The main criticism made to this approach is its iterative part for recalculating pairs representing terms already treated. This is done each time that the used multi-set or the value of the multiplier γ is modified.

4.2 Single Scale Algorithm

The difference between Abchir's algorithm (Abchir, 2013) and Chaoued and Borgi approach (Chaoued and Borgi, 2015) is in the input data as well as in their treatment. In this latter, input data can be the positions of the values to partition or the distances between them. This last case, not allowed with Abchir's approach, offers more flexibility to the users.

The gap between each pair of successive terms reflect a difference in their meaning. In this proposal, inputs are expressed by numerical values. These latter are supplied directly (d_i) (Fig. 6-a) or specified by the position of each term on the scale (v_i) (Fig. 6-b). In the latter case, the distance between the terms (more precisely their positions) v_{i+1} and v_i are calculated as follows:

$$d_i = v_{i+1} - v_i \qquad (9)$$

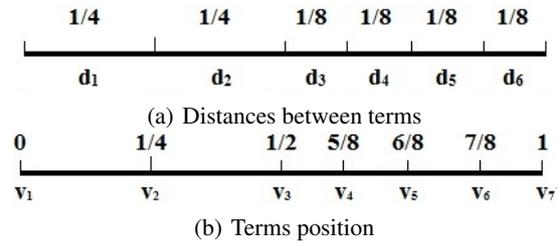(a) Distances between terms

(b) Terms position

Figure 6 – Algorithm inputs

This proposal is based on Espinilla et al. work (Espinilla et al., 2011). The authors proposed a new definition of linguistic hierarchies: extended linguistic hierarchies (ELH), within fuzzy logic. To build an ELH, they use a finite number of levels $l(t, n(t))$ with:

- t: the number indicating the hierarchy level with t=1,..,m; such m is the number of experts ;

- $n(t)$: the granularity of the linguistic term sets corresponding to the level t. Each linguistic term has a triangular membership function, uniformly and symmetrically distributed on [0,1]. The granularity of each level is always odd.

To express his knowledge, each expert can use a specific level, i.e. a particular granularity (Espinilla

et al., 2011). Espinilla et al. proposed to add a new level $l(t^*, n(t^*))$ (Definition 2) to the hierarchy with $t^* = m + 1$. This level retains all the modal points of all the previous levels. The modal points have a membership degree equal to one.

Definition 2. *(Espinilla et al., 2011) Let $\{S^{n(1)}, ..., S^{n(m)}\}$ be the set of linguistic scales with any odd value of granularity. A new level, $l(t^*, n(t^*))$ with $t^* = m + 1$, that keeps the former modal points of the previous m levels can have the following granularity:*

$$n(t^*) = 1 + LCM[(n(1) - 1), (n(2) - 1), ...,$$
$$(n(m+1) - 1)]$$

With m the number of experts

Fig. 7 (Espinilla et al., 2011) shows an ELH with three levels of 3, 5 and 7 terms. The fourth level (t^*) has as granularity the LCM of the previous ones: $1 + LCM(3 - 1, 5 - 1, 7 - 1) = 13$.
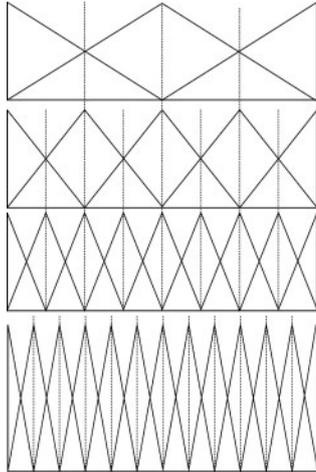

Figure 7 – Extended Linguistic Hierarchy with 3, 5, 7 and 13 terms

In a similar way, Chaoued and Borgi (Chaoued and Borgi, 2015) proposed to express unbalanced terms within a new uniform multi-set. In fact, the granularity of the uniform multi-set equals the LCM of the sets representing initial terms granularity. The input data are unbalanced linguistic terms set S_{un} (L_M, M the number of terms) and the gap between terms. These are provided directly or by specifying the position of each term.

First, the granularity M_k of each uniformly distributed linguistic set L_{M_K} used to represent each input term v_k is calculated. In this set, the distance between each two successive terms is denoted by d_k.

$$M_k = \frac{(1 + d_k)}{d_k}; k = 1, ..., (M - 1) \quad (10)$$

Once the sets L_{M_K} including all the terms of the initial set are determined, the granularity of the uniform set $L_{M'}$ is deduced as:

$$M' = 1 + LCM_{k=1}^{M-1}(M_k - 1) \quad (11)$$

Each value will be expressed on the new uniform scale $L_{M'}$ as a couple $(\tau_{\gamma_k}, L_{M'})$ according to the following rule:

$$d_k = \gamma_k * d'; k = 1, ..., (M - 1) \quad (12)$$

Where d' is the distance between any pair of successive degrees in $L_{M'}$. For a uniform linguistic set, this distance is:

$$d' = \frac{1}{(M' - 1)} \quad (13)$$

To express the distances d_k according to the distance d', the multiplier γ_k is calculated using the rule (12).

Algorithm 1: Single Scale algorithm

Input:
M unbalanced linguistic terms
Normalized distances d_k with k=1, ..., (M-1)
begin
    Calculate granularities M_k; $M_k \in \mathbb{N}^* \setminus \{1\}$
    with k=1, ..., (M-1)
    Calculate granularity:
    $M' = 1 + LCM_{k=1}^{M-1}(M_k - 1)$
    Deduce distance $d' = \frac{1}{(M'-1)}$
    $s_0 \leftarrow (\tau'_0, L_{M'})$
    Calculate $\gamma_k = \frac{d_k}{d'}$ with k=1, ..., (M-1)
    $s_k \leftarrow (\tau'_{\gamma_k}, L_{M'})$
Output: $S' = \{(\tau'_0, L_{M'}), (\tau'_{\gamma_1}, L_{M'}), ...,$
    $(\tau'_{\gamma_{p-1}}, L_{M'})\}$

The minimum term is represented by $(\tau'_0, L_{M'})$ and denoted by s_0. The process, presented in Algorithm 1, continues for other values. The obtained couples $(\tau'_{\gamma_k}, L_{M'})$ are denoted by s_k. To ensure the succession of terms, the calculated multiplier γ_k is added to the one calculated in the previous iteration named γ_{old}. Thus, the comparison is always done regarding the first term of the set $L_{M'}$, i.e., τ'_0.

This algorithm allows to bring us back to the uniform case which will give us the capacity to use different existing tools as linguistic modifiers (Kacem et al., 2015), aggregation operators, etc.

Considering the complexity, this proposal is less complex (O(p)) than Abchir's (O(p^2)), with p the number of input terms (Chaoued and Borgi, 2015). Indeed, in this approach the treatment is done after

introducing all data. Hence, the couples representing terms are calculated once.

To illustrate this algorithm, let us consider an Information Retrieval System (IRS) and the linguistic term set $S_{un} = L_7 = \{N, L, M, H, QH, VH, T\}$ (Herrera-Viedma and López-Herrera, 2007) (Fig. 1). This latter is used to express the relevance of documents in the retrieval process. We aim to represent each term of L_7 using a term from a uniform linguistic set. As input, we consider the normalized distances in L_7 as:

- Distance N - L: $d_1 = \frac{1}{4} = 0.25$

- Distance L - M: $d_2 = \frac{1}{4} = 0.25$

- Distance M - H: $d_3 = \frac{1}{8} = 0.125$

- Distance H - QH: $d_4 = \frac{1}{8} = 0.125$

- Distance QH - VH: $d_5 = \frac{1}{8} = 0.125$

- Distance VH - T: $d_6 = \frac{1}{8} = 0.125$

First, we calculate the values M_k:
$M_1 = \frac{1+d_1}{d_1} = 5 = M_2$; $M_3 = \frac{1+d_3}{d_3} = 9 = M_4 = M_5 = M_6$

Then, we determine the granularity M' of the uniform set and the distance between any of its successive terms d': $M' = 1 + LCM(5-1, 9-1) = 9$;
$d' = \frac{1}{M'-1} = 0.125$

We associate the couple (τ'_0, L_9) to the term N, and we treat the other values:

- For the term L (k = 1)
  $\gamma_{old} \leftarrow 0; \gamma_1 \leftarrow \gamma_{old} + \frac{d_1}{d'} = 2; L \leftarrow (\tau_2, L_9)$

- For the term M (k = 2)
  $\gamma_{old} \leftarrow 2; \gamma_2 \leftarrow \gamma_{old} + \frac{d_2}{d'} = 4; M \leftarrow (\tau_4, L_9)$

- For the term H (k = 3)
  $\gamma_{old} \leftarrow 4; \gamma_3 \leftarrow \gamma_{old} + \frac{d_3}{d'} = 5; H \leftarrow (\tau_5, L_9)$

- For the term QH (k = 4)
  $\gamma_{old} \leftarrow 5; \gamma_4 \leftarrow \gamma_{old} + \frac{d_4}{d'} = 6; QH \leftarrow (\tau_6, L_9)$

- For the term VH (k = 5)
  $\gamma_{old} \leftarrow 6; \gamma_5 \leftarrow \gamma_{old} + \frac{d_5}{d'} = 7; VH \leftarrow (\tau_7, L_9)$

- For the term T (k = 6)
  $\gamma_{old} \leftarrow 7; \gamma_6 \leftarrow \gamma_{old} + \frac{d_6}{d'} = 8; T \leftarrow (\tau_8, L_9)$

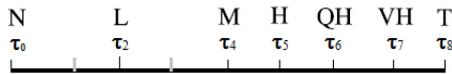The algorithm output is illustrated in Fig. 8.

N	L	M	H	QH	VH	T
τ_0	τ_2	τ_4	τ_5	τ_6	τ_7	τ_8

Figure 8 – Result of the single scale algorithm

5 REPRESENTATION OF TERMS WITHIN UNBALANCED SET

In the previous section, we have presented Chaoued and Borgi's work (Chaoued and Borgi, 2015) to represent unbalanced term set S_{un} within a uniform term set $L_{M'}$. Thereby, existing linguistic modifiers or aggregation operators can be applied to its terms. The obtained values are also expressed using the uniform set $L_{M'}$. However, it is more understandable for the user to represent these results using terms from the initial set S_{un}.

In this section, we focus on the representation of a term τ'_i from the uniform set $L_{M'}$ with a term from S_{un} denoted by L_M. We suppose that the used distances on both sets are normalized. For more clarity, we denote the degrees of the uniform set $L_{M'}$ by $\tau'_0, ..., \tau'_i...$ and those of S_{un} by $\tau_0, ..., \tau_i...$ (knowing that τ_0 and τ'_0 correspond to the same position 0 and τ_M and $\tau'_{M'}$ to 1). We propose a way to determine the nearest position, in S_{un}, to an initial term τ'_i, represented by its position v_i in $L_{M'}$. For that purpose, first, we need to define the proportion of an unbalanced degree from S_{un}.

Definition 3. *Let τ_i be a symbolic degree of an unbalanced multi-set $S_{un}(L_M)$. Its proportion is defined as:*

$$Prop(\tau_i) = \sum_{j=1}^{i} d_j$$

With d_j the normalized distance between each pair of successive degrees in the unbalanced multi-set S_{un}.

We aim to identify the term τ_{pos} in S_{un} which distance between it and the first term (τ'_0) of L'_M is the closest to v_i, the position of τ'_i on L'_M. Thus, we will compare v_i with the sum l_k of the distances separating successive terms in L_M: $l_k = \sum_{j=1}^{k} d_j$.

The process will stop when the value of l_k is higher or equal to v_i. If the values v_i and l_k are equal, the position of τ'_i in $L_{M'}$ is that of τ_{k-1}. If $v_i < l_k$, we check the closest term to v_i between τ_{k-1} and τ_k. In this case, a proportion error, denoted by α, exists and its value is the difference between the position of a term τ'_i and that of the closest term, i.e. τ_{k-1} or τ_k. If these terms are equidistant to τ'_i, we choose the term with the smallest index. This approach is described in Algorithm 2.

The inputs for this function are the position v_i of the term τ'_i, the M terms of the unbalanced multi-set S_{un} and the normalized distances d_k between its successive terms. As output, we obtain the couple (τ_{pos}, α).

To illustrate our approach, let us continue with the

IRS already described. Let us suppose that a document D is described in the database with a set of 5 index $t_i (i \in \{1,...,5\})$:

D = $0.7/t_1 + 0.5/t_2 + 0.9/t_3 + 0.6/t_4 + 0.4/t_5$.

In fact, the system indicates for each index term a numeric value corresponding to its importance in describing the subject discussed in the document. If the index term t_i is not linked to the document subject content, its value is equal to 0. However, index value equals to 1 means that t_i is too important in the document subject. We aim to represent each index value, i.e. v_i, by means of a couple (τ_{pos}, α); such τ_{pos} is a membership degree from L_7 (S_{un}).

The counter k is initialized to 1 and the distance l_k to 0. For the first index value v_1:

- k = 1: $l_1 = 0.25 < 0.7 = v_1$
- k = 2: $l_2 = 0.25 + 0.25 = 0.5 < 0.7$
- k = 3: $l_3 = 0.5 + 0.125 = 0.625 < 0.7$
- k = 4: $l_4 = 0.625 + 0.125 = 0.75 > 0.7$

In this case v_1 is lower than l_4 (= 0.75). We can say that the index value is between τ_3, *High* (H), and τ_4, *Quite High* (QH). So, we check the closest term between τ_3 and τ_4 by comparing values of $(l_4 - v_1)$ and $(v_1 - (l_4 - d_4))$: (0.75 - 0.7) < (0.7 - (0.75 - 0.125)).

Algorithm 2: Representation of terms within unbalanced multi-set

Input:
The position v_i of the term τ'_i, $\tau_i \in L'_M$
M linguistic terms of $L_M(S_{un})$
Normalized distances d_k in L_M; k=1, ..., (M-1)
begin
    k = 1
    $l_1 = 0$
    while $l_k < v_i$ **do**
        $l_k = \sum_{j=1}^{k} d_j$
        k ← k + 1
    if $l_k = v_i$ **then**
        pos ← k-1
        $\alpha \leftarrow 0$
    else
        if $l_k - v_i < v_i - (l_k - d_k)$ **then**
            pos ← k
            $\alpha \leftarrow - (l_k - v_i)$
        else
            pos ← k-1
            $\alpha \leftarrow v_i - (l_k - d_k)$
Output: The couple (τ_{pos}, α)

Hence, τ_4 is closer than τ_3 and the proportion error is α = -0.05. Thus, v_1 is represented by (QH, -0.05)

(Fig. 9). We can say that the degree τ_4 is weakened of 0.05. We proceed using the same approach for the
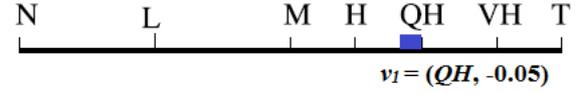

Figure 9 – Representation of the index v_1 within an unbalanced set

other index values. Thus, we obtain:

D = (QH, -0.05)/t_1 + (M, 0)/t_2 + (VH, 0.025)/t_3 + (H, -0.025)/t_4 + (M, -0.1)/t_5.

6 LINGUISTIC MODIFIERS WITH UNBALANCED MULTI-SETS

In multi-valued logic, any symbol can be considered as a modifier of another one. Thereby, modifiers allow to represent small variations of imprecise characterizations of a linguistic variables. Symbolic modifiers may transform simultaneously the membership degrees and the scale of the multi-set. We propose in this section to illustrate how using linguistic modifiers with unbalanced imperfect knowledge. We present an approach to apply Generalized Symbolic Modifiers (Truck, 2002) (Table 1), initially designed for balanced sets, to unbalanced term sets.

To perform this, first, we express the unbalanced term to modify with a term from a balanced multi-set. Afterwards, we apply the GSM modifier on the obtained term within a balanced set. Then, we should find the closest matching term back in the original unbalanced multi-set. Our proposal has as input M linguistic terms of the unbalanced multi-set S_{un} (represented on a scale L_M), a term τ_i from this set, the normalized distances d_k between its successive terms and the GSM m to apply. The treatment will be as follows (Algorithm 3):

1. Express the term τ_i within a uniform multi-set $L_{M'}$ using the single uniform scale approach (Subsection 4.2). The new term is denoted by τ'_i.

2. Apply the GSM m to the term τ'_i. Thus, we get a new uniform linguistic set L''_M, such that $M'' = m(M')$, and a new term $\tau''_i = m(\tau'_i)$.

3. Represent the term τ''_i within the initial unbalanced linguistic set L_M (S_{un}) using the approximation function presented in Algorithm 2.

Thus, for this last step, we have to determine the nearest position to τ''_i in S_{un} (L_M). First, we calculate the position v''_i, i.e. the distance between the term τ''_i

and τ_0''. Afterwards, we determine the term τ_{pos} by comparing v_i'' to the sum l_k of the distances between each successive terms in L_M (d_k).

Algorithm 3: Using GSM with unbalanced multi-set

Input:
M unbalanced linguistic terms of $L_M(S_{un})$
The term $\tau_i \in L_M$
Normalized distances d_k of L_M; k=1, ..., (M-1)
The GSM m
begin
> $L_{M'} \leftarrow$ Single scale approach (M unbalanced terms, d_k) (Algo.1)
> Identify the couple ($\tau_{i'}$, $L_{M'}$) corresponding to (τ_i, L_M) in $L_{M'}$
> (v_i'', M'') $\leftarrow$ m ($\tau_{i'}$, M')
> $v_i'' \leftarrow \frac{i''}{(M''-1)}$
> Calculate the closest term τ_{pos} to v_i'' in S_{un} and the proportion error α (Algo.2)

Output: The couple (τ_{pos}, α)

To illustrate our approach, we use the same example of IRS previously presented. We consider the unbalanced set L_7 of relevance degrees (Fig. 10). We apply the reinforcing modifier CR(1) to the relevance degree Medium, i.e. τ_2. The process is represented in Fig. 11.
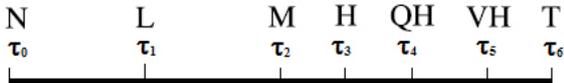
Figure 10 – Unbalanced set of relevance degrees

As mentioned before, the distances d_k are:
$d_1 = d_2 = 0.25$; $d_3 = d_4 = d_5 = d_6 = 0.125$.
The proportion of the term τ_2 in the unbalanced term set L_7 (Definition 3) is:
Prop (τ_2) = $d_1 + d_2 = 0.25 + 0.25 = 0.5$
The granularity of the uniform set, previously calculated in Subsection 4.2, is 9 and the term τ_2 from S_{un} is represented by τ_4' in L_9. We notice that the proportion is preserved Prop(τ_4') = $\frac{4}{8}$ = 0.5.
For this example, we apply the GSM CR (1) to the couple (τ_4', L_9) (Table 1):
$i'' = m(4) = min(4+1, 9-1) = min(5,8) = 5$;
$M'' = m(9) = 9$
The obtained couple is (τ_5'', L_9). We remind that the modifier CR is a reinforcing one, i.e. the proportion is increased: Prop (τ_5'') = $\frac{5}{8}$ > Prop(τ_4') = $\frac{4}{8}$.
Finally, we must express the term τ_5'' within the initial set L_7. We first calculate the distance: $v_5'' = $

$\frac{i''}{(M''-1)} = \frac{5}{8} = 0.625$.
The counter k is initialized to 1 and the distance l_1 to 0.

- For k = 1; $l_1 \leftarrow 0.25 < v_5''$

- For k = 2 ; $l_2 \leftarrow 0.25 + 0.25 = 0.5 < v_5''$

- For k = 3 ; $l_3 \leftarrow 0.5 + 0.125 = 0.625 = v_5''$

In this case, the value of l_3 is equal to v_5''. Thus, τ_3 is the correspondent term in the unbalanced set L_7. The proportion of this term is Prop (τ_3) = 0.625. It corresponds to the proportion of the term τ_5'' in the balanced multi-set L_9.

We can say that the result of applying CR (1) to the relevance degree *Medium* corresponds to the relevance degree *High*.
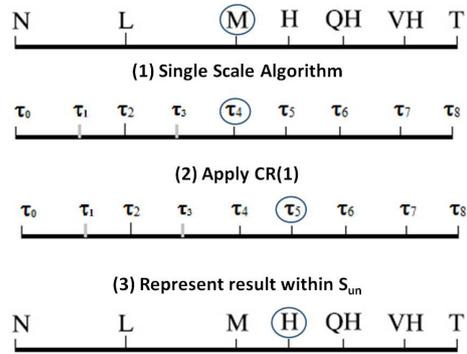
Figure 11 – The process of applying CR (1) to relevance degree M

7 CONCLUSIONS

This work deals with imperfect knowledge management in the multi-valued logic context. Specifically, it deals with the representation and management of unbalanced multi-sets. In fact, the available literature that involves unbalanced term sets are only concerned with the fuzzy logic. They also do not propose a formal definition of unbalanced term sets, which could be an important improvement in imperfect knowledge management.

This paper proposes an algorithm that allows to represent uniformly distributed terms within an unbalanced term set. This is performed by an approximation function that provides the closest term in the unbalanced multi-set to the desired value. In some cases, a proportion error may exist. Besides that, based on Chaoued and Borgi's work (Chaoued and Borgi, 2015) that allows the representation of terms

on a single uniform scale, we introduce a second algorithm that applies linguistic modifiers, initially designed for balanced terms, to unbalanced ones.

As future work, we notice that, in both Abchir's and Nouha and Borgi's algorithms, the input terms for the representation are numerical. It would be interesting to propose a way to treat symbolic cases where the distance between terms are defined with words instead of numbers. It would also be more understandable by humans to express the proportion error using adverbs like *little*, *more or less*, *slightly*,... Another aspect in the management of imperfect knowledge that should be treated is the approximate reasoning (Zadeh, 1975). It is based on the Generalized Modus Ponens. Its principle is to deduce a fact similar to the conclusion rule from an observation approximately equal to the premise rule. It would be important to propose a new Generalized Modus Ponens rules dealing with unbalanced multi-sets.

REFERENCES

Abchir, M.-A. (2013). *Towards fuzzy semantics for geolocation applications*. Phd thesis, University of Paris VIII Vincennes-Saint Denis.

Adkag, H. (1992). *Une approche logique du raisonnement incertain*. Phd thesis, University of Paris 6.

Akdag, H., De Glas, M., and Pacholczyk, D. (1992). A qualitative theory of uncertainty. *Fundamenta Informaticae*, 17(4):333–362.

Akdag, H., Mellouli, N., and Borgi, A. (2000). A symbolic approach of linguistic modifiers. *Information Processing and Management of Uncertainty in Knowledge-Based Systems, Madrid*, pages 1713–1719.

Akdag, H. and Pacholczyk, D. (1989). Incertitude et logique multivalente, première partie : Etude théorique. *BUSEFAL*, 38:122–139.

Akdag, H., Truck, I., Borgi, A., and Mellouli, N. (2001). Linguistic modifiers in a symbolic framework. *Int. J. Uncertain. Fuzziness Knowl.-Based Syst.*, 9(Supplement):49–61.

Bartczuk, L., Dziwinski, P., Starczewski, J., Rutkowski, L., Korytkowski, M., Scherer, R., Tadeusiewicz, R., Zadeh, L., and Zurada, J. (2012). A new method for dealing with unbalanced linguistic term set. volume 1, Czestochowa University of Technology, Department of Computer Engineering, al. Armii Krajowej 36, Czestochowa, 42-200, Poland.

Chaoued, N. and Borgi, A. (2015). Representation of unbalanced terms in multi-valued logic. In *The 12th IEEE International Multi-Conference on Systems, Signals & Devices*.

De Glas, M. (1989). Knowledge representation in a fuzzy setting. *Rapport interne*, 89:48.

Espinilla, M., Liu, J., and Martínez, L. (2011). An extended hierarchical linguistic model for decision-making problems. *Computational Intelligence*, 27(3):489–512.

Francisco Herrera, E. H. and Martínez, L. (2008). A fuzzy linguistic methodology to deal with unbalanced linguistic term sets. *IEEE T. Fuzzy Systems*, 16(2):354–370.

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

Herrera-Viedma, E. and López-Herrera, A. G. (2007). A model of an information retrieval system with unbalanced fuzzy linguistic information. *Int. J. Intell. Syst.*, 22(11):1197–1214.

Jiang, L., Liu, H., and Cai, J. (2015). The power average operator for unbalanced linguistic term sets. *Information Fusion*, 22(0):85 – 94.

Kacem, S. B. H., Borgi, A., and Tagina, M. (2015). Extended symbolic approximate reasoning based on linguistic modifiers. *Knowl. Inf. Syst.*, 42(3):633–661.

Marin, L., Valls, A., Isern, D., Moreno, A., and Merigó, J. (2014). Induced unbalanced linguistic ordered weighted average and its application in multiperson decision making. *Scientific World Journal*, page 642165.

Martínez, L. and Herrera, F. (2012). An overview on the 2-tuple linguistic model for computing with words in decision making: Extensions, applications and challenges. *Inf. Sci.*, 207:1–18.

Truck, I. (2002). *Approches symbolique et floue des modificateurs linguistiques et leur lien avec l'agrégation: Application: le logiciel flous*. Phd thesis, University of Reims Champagne-Ardenne.

Truck, I., Borgi, A., and Akdag, H. (2002). Generalized modifiers as an interval scale: Towards adaptive colorimetric alterations. In *Advances in Artificial Intelligence - IBERAMIA 2002, 8th Ibero-American Conference on AI, Seville, Spain, November 12-15, 2002, Proceedings*, pages 111–120.

Wang, B., Liang, J., Qian, Y., and Dang, C. (2015). A normalized numerical scaling method for the unbalanced multi-granular linguistic sets. *International Journal of Uncertainty, Fuzziness & Knowledge-Based Systems*, 23(2):221 – 243.

Xu, Z. (2009). An interactive approach to multiple attribute group decision making with multigranular uncertain linguistic information. *Group Decision and Negotiation*, 18(2):119 – 145.

Zadeh, L. A. (1965). Fuzzy sets. *Information and Control*, 8(3):338 – 353.

Zadeh, L. A. (1972). A fuzzy-set-theoretic interpretation of linguistic hedges. *Journal of Cybernetics*, 2(3):4–34.

Zadeh, L. A. (1975). The concept of a linguistic variable and its application to approximate reasoning - I. *Inf. Sci.*, 8(3):199–249.