A new proposal to deal with hesitant linguistic expressions on preference assessments

Marie Durand, Isis Truck

A new proposal to deal with hesitant linguistic expressions on preference assessments

Marie Durand

UMR SFL 7023, Paris 8 University, 2 rue de la Liberte, F-93526 Saint-Denis
marie.durand@cnrs.fr

Isis Truck*

CHArt Laboratory – EA 4004, Paris 8 University, 2 rue de la Liberte, F-93526 Saint-Denis
isis.truck@univ-paris8.fr

Abstract

Information fusion and hesitant information fusion represent an important part of decision making processes. This paper focuses on hesitant expressions and the way to take them into account in the computations, using weights served by a simple but efficient process.

In a previous paper we have proposed to use an operator called the symbolic weighted median *to express hesitant linguistic assessments such as "I hesitate between this and that but I tend to lean toward that alternative". Now we go further in explaining in detail how to transform such expressions into our hesitant operators. Inspired by language science research, several hesitant linguistic expressions are discussed, including linguistic modifiers and qualifiers, then they are transformed into weight vectors before being aggregated to complete information fusion.*

I. Introduction

Hesitation between several alternatives is very common in decision making. Therefore there is a need for dealing with uncertainty during fusion information to make a decision. A lot of works and studies have been carried out and propose tools such as a hesitant fuzzy linguistic framework [Torra, 2010, Rodríguez et al., 2012, Liao et al., 2014, Rodríguez et al., 2014, Wang et al., 2015, Liang et al., 2016] and some of these tools use the 2-tuple linguistic model [Herrera and Martínez, 2000]. The basic concept is the following: people may hesitate between several alternatives while an alternative is expressed through a fuzzy set or a *linguistic fuzzy 2-tuple* such as a pair (s_i, α) where s_i is a linguistic term and α a number that represents a symbolic translation to avoid loss of information during the computations. There is a real need in practical cases, especially complex decision making problems [Rodríguez et al., 2016b]. Indeed, for many real-world decisions, the knowledge is either absent, or may only be known in some vague, hesitant, intuitive, way [Girle et al., 2003]. Of course, other models of 2-tuples can also be considered and we proposed in a previous paper a classification of several hesitant operators using several 2-tuple models [Truck and Abchir, 2014].

*Corresponding author

However, a recent paper pointed out the low quality of some proposals and discussed which direction new proposals on hesitant fuzzy sets should follow [Rodríguez et al., 2016a]. Indeed, one of the challenge is: "How can we represent human knowledge?" and the advice given is: "future extensions must be discussed in the context of representing uncertainty in a real world context, providing useful tools for those problems that require representing and managing the hesitancy in expert's knowledge." That is why our purpose here is to continue our work about the transformation of hesitant linguistic expressions into hesitant operators [Truck and Abchir, 2014]. This is not only a way to represent human knowledge, but a way to represent uncertainty in a real world context because linguistic expressions come from subjective assessments, judgements, feelings... from databases to study the mechanisms that underlie second language acquisition [Rast et al., 2011, Truck et al., 2014]. Another question arises: what can be learnt from the language science research?

In this recent paper, our proposal was to use an aggregation operator called the *symbolic weighted median* to express hesitant assessments such as "I hesitate between τ_2 and τ_3 but I tend to lean toward τ_3", where $\tau_i, i \in \{0, n-1\}$ is one among n alternatives. How to go from the linguistic expression to the mathematical modeling was future work. Now we are interested in this question and in the problem of the linguistic expressions themselves.

The present paper is organized as follows: Section 2 recalls the *Symbolic Weighted Median* and their underlying tools, the *Generalized Symbolic Modifiers*. The third section shows the importance of the weights according to the linguistic expression of hesitation that is divided into *qualifiers* and *modifiers*. Section 4 details the way to obtain those weights with a mapping function while Section 5 defines a formal method to aggregate such linguistic hesitant expressions. Finally Section 6 concludes this study.

II. Preliminaries

i. Previous works on Modifiers and Median

The *Generalized Symbolic Modifiers* (GSMs) have been proposed in [Truck and Akdag, 2006] and are used thereafter to express the result of information fusion through an operator called the *Symbolic Weighted Median* (SWM) that is entirely defined by GSMs. A GSM is associated to a semantic triplet of parameters: radius (denoted ρ — the more the radius, the more powerful the modifier), nature (i.e. dilated, eroded or conserved) and mode (i.e. reinforcing, weakening or centring). GSMs are defined through a totally ordered set of M alternatives $\mathcal{L}_M = \{\tau_0, \ldots, \tau_i, \ldots, \tau_{M-1}\}$ ($\forall i, j \in \{0, 1, \ldots, M-1\}$, $\tau_i \leq \tau_j \Leftrightarrow i \leq j$). Four basic operators are defined $\vee$ (max), $\wedge$ (min), $\neg$ (symbolic negation, with $\neg\tau_j = \tau_{M-j-1}$) and the Łukasiewicz implication $\rightarrow_L$: $\tau_i \rightarrow_L \tau_j = \min(\tau_{M-1}, \tau_{M-1-(i-j)})$

τ', the value after modification, is computed according to a GSM m with a radius ρ, denoted m_ρ. Actually m_ρ modifies the pair $(\tau_i, \mathcal{L}_M)$ into another pair $(\tau'_j, \mathcal{L}_{M'})$.

Definition 1. *[Truck and Akdag, 2006]*
Given $\rho \in \mathbb{N}^*, i \in \{0, \ldots, M-1\}$, any $\tau'_j (j \in \{0, \ldots, M'-1\})$ can be computed through $m_\rho(\tau_i)$.

$$\begin{array}{rccc} m_\rho \colon & \mathcal{L}_M & \rightarrow & \mathcal{L}_{M'} \\ & \tau_i & \mapsto & \tau'_j \end{array}$$

Three GSM families have been defined: weakening, reinforcing (see Table 1) and central ones (see Definition 2 for an example of such a GSM , where DC' is a *d*ilated *c*entring modifier, *i.e.* granularity increases).

MODE / NATURE	Weakening		Reinforcing	
Erosion	$\tau'_j = \tau_{\max(0,i-\rho)}$ $\mathcal{L}_{M'} = \mathcal{L}_{\max(1,M-\rho)}$	EW(ρ)	$\tau'_j = \tau_i$ $\mathcal{L}_{M'} = \mathcal{L}_{\max(i+1,M-\rho)}$	ER(ρ)
			$\tau'_j = \tau_{\min(i+\rho,M-\rho-1)}$ $\mathcal{L}_{M'} = \mathcal{L}_{\max(1,M-\rho)}$	ER′(ρ)
Dilation	$\tau'_j = \tau_i$ $\mathcal{L}_{M'} = \mathcal{L}_{M+\rho}$	DW(ρ)	$\tau'_j = \tau_{i+\rho}$ $\mathcal{L}_{M'} = \mathcal{L}_{M+\rho}$	DR(ρ)
	$\tau'_j = \tau_{\max(0,i-\rho)}$ $\mathcal{L}_{M'} = \mathcal{L}_{M+\rho}$	DW′(ρ)		
Conservation	$\tau'_j = \tau_{\max(0,i-\rho)}$ $\mathcal{L}_{M'} = \mathcal{L}_M$	CW(ρ)	$\tau'_j = \tau_{\min(i+\rho,M-1)}$ $\mathcal{L}_{M'} = \mathcal{L}_M$	CR(ρ)

Table 1: *Summary of reinforcing and weakening GSMs, according their nature [Truck and Akdag, 2006].*

Definition 2. *[Truck and Akdag, 2006]*

$$DC'(\rho) = \begin{cases} \tau'_j = \begin{cases} \tau_{\frac{i*(M*\rho-1)}{M-1}} & \text{if } \tau_{\frac{i*(M*\rho-1)}{M-1}} \in \mathcal{L}_{M'} \\ \tau_{\left\lfloor \frac{i*(M*\rho-1)}{M-1} \right\rfloor} & \text{otherwise (pessimistic)} \\ \tau_{\left\lfloor \frac{i*(M*\rho-1)}{M-1} \right\rfloor+1} & \text{otherwise (optimistic)} \end{cases} \\ \mathcal{L}_{M'} = \mathcal{L}_{M*\rho} \end{cases}$$

Definition 3. *[Truck and Akdag, 2009] Let $\mathcal{L}_M = \{\tau_0, \tau_1, \ldots, \tau_{M-1}\}$ be a collection of M ordered elements. When the elements have weights, the collection is denoted $\langle \tau_0^{w_0}, \tau_1^{w_1}, \ldots, \tau_{M-1}^{w_{M-1}} \rangle \in \mathcal{B}^{\mathcal{L}_M}$ (set of collections) such that $\sum w_i = 1$, $i = \{0, \ldots, M-1\}$. The* Symbolic Weighted Median $\mathcal{M}$ *is defined as follows:*

$$\begin{array}{rcl} \mathcal{M} : \ \mathcal{B}^{\mathcal{L}_M} & \to & \mathcal{L}_{M'} \\ \langle \tau_0^{w_0}, \tau_1^{w_1}, \ldots, \tau_{M-1}^{w_{M-1}} \rangle & \mapsto & \mathcal{M}(\langle \tau_0^{w_0}, \tau_1^{w_1}, \ldots, \tau_{M-1}^{w_{M-1}} \rangle) \\ & & = \tau'^{w'_j}_j \ \textit{such that:} \ \left| \sum_{p=0}^{j-1} w'_p - \sum_{p=j+1}^{M'-1} w'_p \right| < \varepsilon \\ & & = m(\tau_i^{w_i}, \mathcal{L}_{M-1}) \ \textit{with} \ w_i = 1 \\ & & = m(\tau_i, \mathcal{L}_{M-1}) \end{array}$$

with $m(\tau_i, \mathcal{L}_{M-1})$ a GSM (or a composition of GSMs) applied to an element of the initial collection $\mathcal{L}_M$ and where $\sum_{p=0}^{j-1} w'_p$ ($\sum_{p=j+1}^{M'-1} w'_p$ respectively) is the sum $\mathscr{S}_1$ ($\mathscr{S}_2$ respectively) of the weights that are before (after respectively) element $\tau'^{w'_i}_j$.

ε has to be negligible ($\varepsilon \ll w_i$), so both sums $\mathscr{S}_1$ and $\mathscr{S}_2$ must be as close as possible. ε is a *negligible semantic gap*, *i.e.* a negligible difference between two linguistic descriptions of an object. The chosen method is to change scale, *i.e.* to subdivise the element into sub-elements. This way, a new collection is obtained and the sums $\mathscr{S}$ can be computed again. Thus the result of the aggregation is either an element of $\mathcal{L}_M$, or a *sub-element*. A sub-element is an element on which one or more GSMs have been applied.

ii. Expressing the doubt linguistically

In [Truck and Abchir, 2014], we focused on the various ways to express the hesitation or the doubt and we obtained two families of linguistic statements. The binary ones are: *between* τ_α *and* τ_β; τ_α *or* τ_β (there is no condition on τ_α nor τ_β, *i.e.* they don't need to be subsequent values); the unary ones are: *at most* τ_α; *at least* τ_α; *everything except* τ_α.

In the statement "I hesitate between two alternatives but I tend to lean toward the second one", the words "tend to lean towards" add obviously a notion of a weight assigned on the second alternative (the second alternative is assigned a higher weight than the first one).

We have seen that the SWM permits to express an aggregation of a set of weighted alternatives. Considering M alternatives denoted τ_0 to τ_{M-1}, the above statement representing an expert's opinion can be expressed with the following collection of weighted alternatives: $\langle \tau_0^{w_0}, \ldots, \tau_i^{w_i}, \ldots, \tau_{M-1}^{w_{M-1}} \rangle$. The weights w_i permit to express the hesitation, with $\Sigma w_i = 1$.

The choice of the best weights to sum up the expert's opinion is considered below, according to the linguistic assessment.

So we assume in this paper that the SWM is able to express all kinds of hesitation, because they all are a question of weights on alternatives.

iii. Hesitant Fuzzy Linguistic Term Sets

Recently, several scientists focused on hesitant linguistic expressions and formalisms dedicated to them. Torra introduced the concept of Hesitant Fuzzy Sets as extensions and generalizations of fuzzy sets [Torra, 2010]. Hesitant fuzzy sets deal with quantitative settings.

Besides, Hesitant Fuzzy Linguistic Term Sets, as *qualitatives settings*, have been proposed to provide a linguistic and computational basis to increase the richness of linguistic elicitation based on the fuzzy linguistic approach [Rodríguez et al., 2012].

A *Hesitant Fuzzy Linguistic Term Set* (HFLTS), denoted H_S, is defined as an ordered finite subset of the consecutive linguistic terms of a linguistic term set.

Many basic operations have been defined such as upper bound and lower bound, the complement of an HFLTS, union and intersection between two HFLTS. The concept of *fuzzy envelope* of an HFLTS is also introduced as being a linguistic interval bounded by the minimal and the maximal elements of the HFLTS.

Several expressions are suggested to explicit the hesitation. These expressions are generated from a *context-free grammar* denoted G_H with primary and composite terms, unary and binary relations and conjunction. Three expressions obtained through a transformation function E_{G_H} are given: *at least*, *at most* and *between ... and ...* (see [Rodríguez et al., 2013].

III. Expression of a hesitant linguistic assessment

Sapir first introduced the term *scale* in language sciences in his work on the expression of gradation [Sapir, 1944]. A **scalar** is defined as a term relative to a graduated scale or in the Oxford English Dictionary "a set or series of graduations". Talking about expressions of hesitation on preference assessments refers to an hesitation about which linguistic term to choose in an ordered set. Quirk & al state that **degree adverbials** are concerned with "the assessment of gradable constituents in relation to an imaginary scale" [Quirk, 1985]. They distinguish **two families** of degree adverbials: those which express a measure (*more than*); adverbs which convey a degree of *intensity*, which are divided into *two kinds*: amplifiers and diminishers (respectively refered as "reinforcing" and "weakening" in this paper).

Besides, there are various ways of expressing the hesitation. Keeping in mind the scalars and the degree adverbials, we focus on the following expressions:

1. I think it is **really** τ_α;
2. I think it is **really more** (resp. **less**) **than** τ_α;
3. I think it is **rather** τ_α;
4. I think it is **rather more** (resp. **less**) **than** τ_α;
5. I think it is **a little bit more** (resp. **less**) **than** τ_α;
6. I think it is **between** τ_α **and** τ_β;
7. I think it is τ_α **or** τ_β;
8. I think it is **at most** τ_α. This is equivalent to " I think it is **less than** τ_α ";
9. I think it is **at least** τ_α. This is equivalent to " I think it is **more than** τ_α ";
10. I think it is **everything except** τ_α;
11. I think it is **between** τ_α **and** τ_β, but I **tend to lean toward** τ_β;

The seventh item (τ_α **or** τ_β) expresses a hesitation between two values that can be far apart from eachother. For example, when choosing a color, one may hesitate between yellow or blue. That doesn't mean that yellow and blue have to be close in the hue scale. Another example could be a scoring task. One may hesitate between a 5 (out of 20) and an 11 (out of 20). It depends on how to judge the homework, if one considers that it contains off-topic contents or not. In one case, it would be given a poor rating such as 5, in the other, it would score rather well.

A careful analysis of the interaction between the words which compose the expressions leads to the conclusion, according to Quirk *et al.*'s work, that there are **two kinds** of words: those for the **qualification of the expression of the doubt** and those for the **modification of the expression of the doubt**. It is important to notice that the modification here is not (directly) related to the GSMs.

Really, **rather**, **a little bit** and **tend to lean toward** are *modifiers* while **more than**, **less than**, **between ... and**, **or** and **everything except** are *qualifiers* of the doubt.

According to what is discussed above, a doubt can be expressed through a tuple of weighted alternatives, *i.e.* a (or several) GSM(s) applied to an alternative.

i. Qualifiers of the doubt

Each linguistic expression is composed of one — unary (resp. two — binary), *reference value(s)* τ_α (resp. τ_α and τ_β) $\in \mathcal{L}_M$ and an *influence area* A, defined as a subset of $\mathcal{L}_M$, such that $w_i \neq 0$, for each $\tau_i \in A$. k is the cardinality of A.

Table 2 shows the qualifiers with their corresponding reference values. It is to notice that $\emptyset$ is a limit case, particularly useful when using statements with **just a modifier**, *e.g. really* τ_α. So the reference value of $\emptyset$ is simply the alternative (denoted τ_α).

qualifier from linguistic statement	reference values	conditions	family
$\emptyset$	τ_α	no condition	unary
between ... *and*	τ_α, τ_β	$\alpha < \beta - 1$	binary
... *or* ...	τ_α, τ_β	$\alpha \neq \beta$	binary
at most ... or *less than* ...	τ_α	no condition	unary
at least ... or *more than* ...	τ_α	no condition	unary
everything except ...	τ_α	no condition	unary

Table 2: *The six qualifiers to express hesitation.*

modifier from linguistic statement	class	applies to	action
really[1]	reinforcing	unary qualifiers except $\emptyset$	modifies the influence area and decreases the weight assigned to the reference value in favour of the following or the preceding values
really	reinforcing	$\emptyset$ (limit case)	increases the weight assigned to the reference value at the expense of the following or the preceding values
rather	weakening	unary qualifiers	decreases the weight assigned to the reference value and extends the influence area according to the number of τ_i in the interval
a little bit	weakening	unary qualifiers, except $\emptyset$	concentrates the weights around the reference value according to a distribution function such as $f(x) = 1/x$
tend to lean toward	weakening	binary	increases the weight assigned to one of the reference values as well as the neighbouring values inside the influence area, at the expense of the other reference value and its neighbours

Table 3: *The four modifiers to express hesitation.*

ii. Modification of the expression of the doubt

We propose a simple classification of the modifiers into two classes: reinforcing and weakening ones. According to the linguistic expressions, the modifiers have a certain action on the weights and may extend or reduce the influence area, give more or less importance to a certain τ_i, compared with its neighbours, etc. Table 3 provides a summary of the various cases.

We now formalize these qualifiers and modifiers, through introduction of a context-free grammar with a set of prodution rules.

iii. Resulting context-free grammar

A context-free grammar permits to define rules in order to obtain all possible strings in a formal language [Chomsky, 1956]. Usually, a context-free grammar is defined by a 4-tuple where the first element is a finite set of nonterminal characters or variables; the second element is a finite set of terminals, *i.e.* the alphabet of the language; the third element is a finite relation defining

[1]Let us notice that "really" is a generic word. "Absolutely", "certainly", "truly"... could replace "really" in the linguistic expression.

rules or productions of the grammar; and the fourth element is the start variable (or symbol) [Bordogna and Pasi, 1993, Rodríguez et al., 2011].

Definition 4. *Let G_D be a context-free grammar to express the doubt, and $\mathcal{L}_M = \{\tau_0, \dots, \tau_i, \dots, \tau_{M-1}\}$ be a set of ordered elements. G_D is defined by the following 4-tuple: $G_D = (V, \sum, R, S)$ which syntax is described using the extended Backus-Naur form [Scowen, 1993]:*

$V = \{\langle \textit{reference value} \rangle, \langle S \rangle, \langle \textit{binary relation} \rangle, \langle \textit{unary relation} \rangle, \langle \textit{modifier} \rangle\}$

$\sum = \{$*at most, less than, at least, more than, everything except, between ... and, ... or ...,
tend to lean toward, rather, a little bit, really,*$\tau_0, \dots, \tau_i, \dots, \tau_{M-1}\}$

$R = \{S ::= \langle \textit{reference value} \rangle | \langle S \rangle \langle \textit{reference value} \rangle$
$S ::= \langle \textit{unary relation} \rangle | \langle \textit{binary relation} \rangle \langle \textit{reference value} \rangle | \langle \textit{modifier} \rangle$
$| \langle \textit{modifier} \rangle \langle \textit{unary relation} \rangle | \langle \textit{modifier} \rangle \langle \textit{binary relation} \rangle \langle \textit{reference value} \rangle$
$\langle \textit{reference value} \rangle ::= \tau_0 | \dots | \tau_i | \dots | \tau_{M-1}$
$\langle \textit{binary relation} \rangle ::=$ *between ... and* | *... or ...*
$\langle \textit{unary relation} \rangle ::=$ *at most* | *less than* | *at least* | *more than* | *everything except*
$\langle \textit{modifier} \rangle ::= \langle \textit{weakening} \rangle | \langle \textit{reinforcing} \rangle$
$\langle \textit{reinforcing} \rangle ::=$ *really*
$\langle \textit{weakening} \rangle ::=$ *tend to lean toward* | *rather* | *a little bit*$\}$

As Rodríguez *et al.*, we define a function E_{G_D} that transforms linguistic expressions obtained by G_D into $\mathcal{L}_M$, where $\mathcal{L}_M$ is the linguistic term set used by G_D:

$$E_{G_D} : e^A_{\tau_i} \longrightarrow \mathcal{L}_M$$

The next step is to define a formal way to obtain the weights that will be assigned to linguistic terms.

IV. Characterization of Linguistic Information based on the weights

i. A mapping function to compute the weights

The processing of linguistic expressions into a weighted tuple of τ_i is done by weight distribution (with normalized weights) on the set $\mathcal{L}_M$, according to their influence area A and their reference(s) value(s) τ_α (and τ_β).

To compute the weights, we propose a function which inputs a linguistic expression in a natural language that expresses hesitation or doubt, and which outputs a tuple of n weights denoted $\langle w_0, \dots, w_{n-1} \rangle$ summing to 1.

Definition 5. *Let $e^A_{\tau_\alpha}$ (respectively $e^A_{\tau_\alpha, \tau_\beta}$) be a unary (respectively binary) linguistic expression over $\mathcal{L}_M$.*

Let P be the set of following hesitant modifiers: $P = \{$really, rather, a little bit, tend to lean toward$\}$.

Let Q be the set of following hesitant qualifiers: $Q = \{$more than, less than, between...and, or, everything except$\}$.

The function I is a mapping which obtains the weights for each τ_i, $i = \{0, \dots, M-1\}$:

$$\begin{array}{rcl} I: \ P \times Q \times \mathcal{L}_M & \rightarrow & \mathcal{L}_M \\ e^A_{\tau_\alpha} (\text{resp. } e^A_{\tau_\alpha, \tau_\beta}) & \mapsto & \langle \tau_0^{w_0}, \tau_1^{w_1}, \dots, \tau_{M-1}^{w_{M-1}} \rangle \end{array}$$

ii. Characterization of the hesitant linguistic expressions through the mapping I

Regarding the qualifiers, we propose the following weights to take into account the linguistic expressions.

Let us consider the six qualifiers from Table 2. All $\tau_i^{w_i}$ which belong to A have their weight w_i different from 0 and τ_α (resp. τ_α and τ_β) belongs (resp. belong) to A.

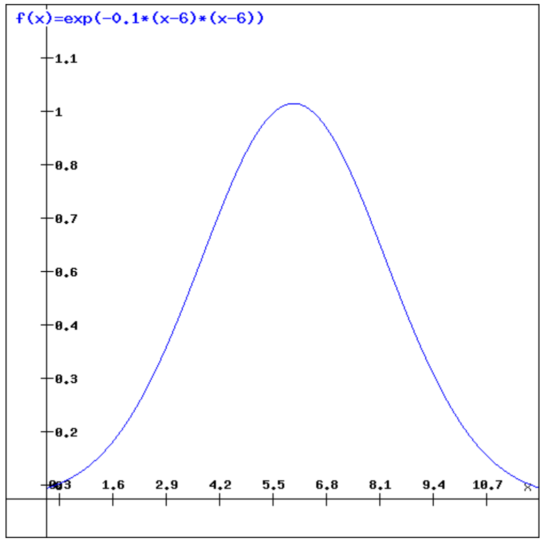
1. $\emptyset$ τ_α: $\{w_\alpha = 1\}$
2. between τ_α and τ_β: the rule to be applied is the following. Let k be the number of $\tau \in A$, *i.e.* $k = \beta - \alpha - 1$. w_α equals w_β and both are twice less important than $w_i, i \in]\alpha, \beta[$. It follows that $w_i = \dfrac{1}{k+1}$ for any $\alpha < i < \beta$; $w_\alpha = w_\beta = \dfrac{1}{2k+2}$ and $w_j = 0$ for any $j < \alpha$ or $j > \beta$
3. τ_α or τ_β: $\{w_\alpha = 0.5; w_\beta = 0.5$ and $w_j = 0$ for any $j \in \{0, \ldots, M-1\}$ and $j \neq \alpha, j \neq \beta\}$
4. at most τ_α: the rule to be applied is the following. $\{w_0 = 0, w_\alpha = 0.5,\ w_i = f(i)$ for $0 < i < \alpha\}$ where $f(x)$ is a linear increasing function. For example, $w_i = \dfrac{i}{2\alpha}$, then we normalize. So, $w_i = \dfrac{i}{2\alpha * \sum_i w_i}$
5. at least τ_α: the rule to be applied is the following. $\{w_\alpha = 0.5, w_{M-1} = 0$ and $w_i = f(i)$ for $\alpha < i < M-1\}$ where $f(x)$ is a linear decreasing function. For example, $w_i = \dfrac{M-1-i}{2(M-1-\alpha)}$, then we normalize. So, $w_i = \dfrac{M-1-i}{2(M-1-\alpha) * \sum_i w_i}$
6. everything except τ_α: the weights are evenly distributed, except for w_α whose value is zero. $\{w_\alpha = 0, w_i = \dfrac{1}{M-1}\}$, for $i = \{0, 1, \ldots, \alpha-1, \alpha+1, \ldots, M-1\}$

Regarding *at most* and *at least* qualifiers, in the common sense, when we say "at least 10 out of 20", for example, it means 10, 11, 12, ... but not 19 nor 20. Similarly, in the common sense, when we say "at most 10 out of 20", for example, it means 10, 9, 8, ... but not 0.

Regarding the modifiers, we have seen that basic linguistic expressions can be associated to other terms when used by evaluators, in order to insist or weaken a statement they made. Let us consider the four modifiers from Table 3.

1. Reinforcing: "really"
   - applied to unary qualifiers except $\emptyset$. The idea is to reinforce the weight of the reference value in order to free some weight in favour of the other values. The initial weight (denoted $w_{\alpha_{old}}$ and usually equals to 0.5) is multiplied by the inverse of a fuzzy index denoted $\zeta \in]0, 1[$ in order to increase the value. By default, $\zeta = 0.8$: $\{w_\alpha = w_{\alpha_{old}}/\zeta$ and the same linear function $f(x)$ is kept, then we normalize$\}$;
   - applied to $\emptyset$: $\{w_i = exp^{-\nu(i-\alpha)^2}$ for all i, with $\nu \in]0, +\infty[\ \}$, see Figure 1 where $\alpha = 6$ and $\nu = M/100 = 0.1$; then we normalize. The function for w_i is a Gaussian function because normal distribution may occur in natural phenomena. For example, Zadeh or Zimmermann used Gaussian membership functions to express vague linguistic terms and to model weight vectors [Zadeh, 1994, Zimmermann, 1996]. ν is an index of non-fuzziness. When ν decreases towards 0, the result is very fuzzy, when ν approaches infinity, the result is very crisp.

Figure 1: *Distribution function of the weights for the expression "really τ_α" ($\alpha = 6$ and $\nu = 0.1$).*



2. Weakening

   - rather: the idea is to weaken the weight of the reference value in order to free some weight in favour of the other values. The initial weight (denoted $w_{\alpha_{old}}$ and usually equals to 0.5) is multiplied by $\zeta \in]0,1[$ in order to decrease the value. By default, $\zeta = 0.8$: $\{w_\alpha = w_{\alpha_{old}} * \zeta$ and the same linear function $f(x)$ is kept, then we normalize $\}$
   - a little bit: the idea is to concentrate the weights around the reference value, increasing or decreasing the weights before or after the reference value: $\{w_\alpha = w_{\alpha_{old}} * \zeta; w_{\alpha+1} = w_{\alpha+1} + (1-\zeta) * w_{\alpha_{old}}$ (or $w_{\alpha-1} = w_{\alpha-1} + (1-\zeta) * w_{\alpha_{old}}$, depending on whether the qualifier is "more than" or "less than") and the same linear function $f(x)$ is kept, then we normalize$\}$
   - tend to lean toward: depends on the reference value (τ_α or τ_β). Increases the weight assigned to τ_α (or τ_β) and decreases the weight of the other reference value. $\{w_\alpha = w_{\alpha_{old}}/\zeta, \zeta = 0.9$ by default, $\Delta = w_{\alpha_{old}} - w_\alpha$ (resp. w_β) and $w_\beta = w_{\beta_{old}} - \Delta$ (resp. w_α)$\}$.

We now propose an example to show the feasibility of our approach.

iii. A practical example

Example. *Let $C = \{c_1, c_2, c_3, c_4\}$ be a set of criteria on a same object O, $\mathcal{L}_M = \{\tau_0, \tau_1, \ldots, \tau_{10}\} = \{$nothing, absolutely low, very low, low, medium low, fair, medium high, high, very high, absolutely high, perfect$\}$, with $M = 11$ a linguistic term set used to described O. The context-free grammar G_D from Definition 4 is applied on these two linguistic term sets. The assessments provided for O are presented in Table 4.*

- *Criteria 1*
  For the first criteria the value $e^A_{\tau_\alpha}$ ="everything except high", where $\alpha = 7$ and $A = \mathcal{L}_M = \{$nothing, absolutely low, very low, low, medium low, fair, medium high, high, very high, absolutely high, perfect$\}$ because the whole set is taken as influence area, according to the above definition. We then

		qualifier	modifier	reference value(s)
O	c_1	everything except	$\emptyset$	high
	c_2	between ... and ...	tend to lean toward	very low, medium high
	c_3	at least	$\emptyset$	medium low
	c_4	$\emptyset$	really	fair

Table 4: *Assessments provided for the description of object O*

*apply the mapping function I to compute weights for each element in $\mathcal{L}_M$: the weights are evenly distributed, except for w_7 whose value is zero (*cf. *subsection ii). $\{w_7 = 0, w_i = \frac{1}{11-1} = 0.1\}$, for $i = \{0, 1, \ldots, 6, 8, \ldots, 10\}$*
To sum up, the result that has been produced is the following:

$$\begin{array}{rcl} I : P \times Q \times \mathcal{L}_M & \longrightarrow & \mathcal{L}_M \\ e^A_{\tau_7} & \mapsto & \langle \tau_0^{0.1}, \tau_1^{0.1}, \tau_2^{0.1}, \tau_3^{0.1}, \tau_4^{0.1}, \tau_5^{0.1}, \tau_6^{0.1}, \tau_7^{0}, \tau_8^{0.1}, \tau_9^{0.1}, \tau_{10}^{0.1} \rangle \end{array}$$

- *Criteria 2*
  *For the second criteria the doubt is expressed through a binary qualifier and a weakening modifier. $e^A_{\tau_\alpha, \tau_\beta}$ ="between very low and medium high, but tend to lean toward medium high", where $\alpha = 2, \beta = 6$ and $A = \{\tau_2, \tau_3, \ldots, \tau_6\}$. We then apply the mapping function I to compute weights for each element in $\mathcal{L}_M$. For the binary qualifier, $k = 6 - 2 - 1 = 3$, $w_2 = w_6$ and both are twice less important than w_3, w_4 and w_5 (*cf. *subsection ii). $w_2 = w_6 = \frac{1}{2*3+2} = 0.125$ and $w_3 = w_4 = w_5 = \frac{1}{3+1} = 0.25$*
  With the modifier, $w_6 = \frac{w_{\beta_{old}}}{\zeta} = 0.125/0.9 = 0.139$, $\Delta = 0.139 - 0.125 = 0.014$, $w_2 = w_{\alpha_{old}} - \Delta = 0.125 - 0.014 = 0.111$
  To sum up, the result that has been produced is the following:

$$\begin{array}{rcl} I : P \times Q \times \mathcal{L}_M & \longrightarrow & \mathcal{L}_M \\ e^A_{\tau_2, \tau_6} & \mapsto & \langle \tau_0^{0}, \tau_1^{0}, \tau_2^{0.111}, \tau_3^{0.25}, \tau_4^{0.25}, \tau_5^{0.25}, \tau_6^{0.139}, \tau_7^{0}, \tau_8^{0}, \tau_9^{0}, \tau_{10}^{0} \rangle \end{array}$$

- *Criteria 3*
  For the third criteria the doubt is expressed through a unary qualifier. $e^A_{\tau_\alpha}$ ="at least medium low", where $\alpha = 4$ and $A = \{\tau_4, \tau_5, \ldots, \tau_9\}$. We then apply the mapping function I to compute weights for each element in $\mathcal{L}_M$: $w_4 = 0.5, w_{10} = 0, w_5 = \frac{11-1-5}{2(11-1-14)} = 5/12, w_6 = 4/12$, $w_7 = 3/12, w_8 = 2/12, w_9 = 1/12$, then we normalize.
  To sum up, the result that has been produced is the following:

$$\begin{array}{rcl} I : P \times Q \times \mathcal{L}_M & \longrightarrow & \mathcal{L}_M \\ e^A_{\tau_4} & \mapsto & \langle \tau_0^{0}, \tau_1^{0}, \tau_2^{0}, \tau_3^{0}, \tau_4^{0.286}, \tau_5^{0.238}, \tau_6^{0.19}, \tau_7^{0.143}, \tau_8^{0.095}, \tau_9^{0.048}, \tau_{10}^{0} \rangle \end{array}$$

- *Criteria 4*
  For the fourth criteria the doubt is simply expressed through a modifier. $e^A_{\tau_\alpha}$ ="really fair", where $\alpha = 5$ and $A = \{\tau_0, \tau_1, \ldots, \tau_{10}\}$. We then apply the mapping function I to compute weights for each element in $\mathcal{L}_M$: $w_0 = exp^{-0.1*(0-5)^2} = 0.082, w_2 = exp^{-0.1*(1-5)^2} = 0.202, \ldots, w_5 = 1, \ldots, w_{10} = exp^{-0.1*(10-5)^2} = 0.082$, then we normalize.

To sum up, the result that has been produced is the following:

$$\begin{array}{rcl} I : P \times Q \times \mathcal{L}_M & \longrightarrow & \mathcal{L}_M \\ e^A_{\tau_5} & \mapsto & \langle \tau_0^{0.015}, \tau_1^{0.036}, \tau_2^{0.074}, \tau_3^{0.121}, \tau_4^{0.164}, \tau_5^{0.181}, \tau_6^{0.164}, \tau_7^{0.121}, \tau_8^{0.074}, \tau_9^{0.036}, \tau_{10}^{0.015} \rangle \end{array}$$

We have entirely defined the mapping I which assigns the weights on the alternatives τ_i depending on the linguistic statement and hesitation expressions. Now we are interested in the way to aggregate them.

V. Information Aggregation

A hesitant linguistic expression reflecting assessments of an expert based on various criteria on a same object can be expressed through a matrix $C \times T$ where T belongs to a linguistic term set S of n terms and C represents one of the p criteria.

The global representation is a weight matrix where each line (each vector) represents a linguistic assessment on a criteria: w_i^j is the weight of linguistic term τ_i for criteria C_j.

$$\begin{pmatrix} w_0^0 & w_1^0 & \dots & w_{n-1}^0 \\ w_0^1 & w_1^1 & \dots & w_{n-1}^1 \\ \dots & \dots & \dots & \dots \\ w_0^{p-1} & w_1^{p-1} & \dots & w_{n-1}^{p-1} \end{pmatrix}$$

Example. *Following Example 1, the resulting matrix of weights for the object O is :*

$$\begin{pmatrix} 0.1 & 0.1 & 0.1 & 0.1 & 0.1 & 0.1 & 0.1 & 0 & 0.1 & 0.1 & 0.1 \\ 0 & 0 & 0.111 & 0.25 & 0.25 & 0.25 & 0.139 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0.286 & 0.238 & 0.19 & 0.143 & 0.95 & 0.48 & 0 \\ 0.015 & 0.036 & 0.074 & 0.121 & 0.164 & 0.181 & 0.164 & 0.121 & 0.074 & 0.036 & 0.015 \end{pmatrix}$$

To perform the information fusion, there are three distinct steps:

- for each matrix, each row vector is given as input to the SWM algorithm (see [Truck and Abchir, 2014]) which returns a $m(\tau_i, \mathcal{L}_{M-1})$, *i.e.* a GSM (or a composition of GSMs) applied to an element of the initial collection $\mathcal{L}_M$. This part of the work won't be described here since it has already been explained in [Truck and Abchir, 2014], Section 4.2. So the fusion is done by criteria;
- we thus obtain a column vector

$$\begin{pmatrix} m(\tau_i, \mathcal{L}_{M-1})_{C_0} \\ m(\tau_i, \mathcal{L}_{M-1})_{C_1} \\ \dots \\ m(\tau_i, \mathcal{L}_{M-1})_{C_{p-1}} \end{pmatrix}$$

- this new vector will be aggregated into a single value, that will also be a GSM or a composition of GSMs (see [Truck and Abchir, 2014]), knowing that those $m(\tau_i, \mathcal{L}_{M-1})$ may have various granularity levels (M has no reason to be the same for each $\tau_i^{w_i}$). The reduction to a common denominator goes beyond the scope of this paper which focused on the way to assign the weights before information aggregation.

VI. Conclusions

We have proposed a study of various expressions of the doubt, focusing on the fact that there are two parts in such expressions: qualifiers and modifiers. Moreover, we assume in this paper that a linguistic hesitant expression is a set of linguistic terms assigned with weights.

Qualifiers are divided into two families (unary and binary) focusing on 6 different behaviours. Modifiers are divided into two classes (reinforcing and weakening) focusing on five different behaviours. Modifiers act on qualifiers in changing the weights assigned on the linguistic terms. In a nutshell, a hesitant linguistic expression is a kind of composition of two functions, one related to the modifiers, and the other one related to the qualifiers. It is the composition of functions that makes the wholeness of the mapping I described above.

Thus the various cases are studied giving each time a (generic) function to obtain the weights.

Future works will be to take into account the various levels of granularity obtained after the computations (the aggregation) of several hesitant linguistic expressions.

References

[Bordogna and Pasi, 1993] Bordogna, G. and Pasi, G. (1993). A fuzzy linguistic approach generalizing boolean information retrieval: A model and its evaluation. *Journal of the American Society for Information Science*, 44(2):70.

[Chomsky, 1956] Chomsky, N. (1956). Three models for the description of language. *IEEE TRansactions on Information Theory*, 2(3):113–124.

[Girle et al., 2003] Girle, R., Hitchcock, D., McBurney, P., and Verheij, B. (2003). *Decision Support for Practical Reasoning: a theoretical and computational perspective*, pages 55–84. Kluwer Academic Publishers, Netherlands.

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

[Liang et al., 2016] Liang, X., Wei, C., and Cheng, X. (2016). *Entropy Measures for Extended Hesitant Fuzzy Linguistic Term Sets and Their Applications*, pages 151–164. Springer Singapore, Singapore.

[Liao et al., 2014] Liao, H., Xu, Z., and Zeng, X.-J. (2014). Distance and similarity measures for hesitant fuzzy linguistic term sets and their application in multi-criteria decision making. *Information Sciences*, 271:125–142.

[Quirk, 1985] Quirk, R. (1985). *A Comprehensive grammar of the English language*. General Grammar Series. Longman.

[Rast et al., 2011] Rast, R., Dimroth, C., and Watorek, M. (2011). Language teaching and acquisition: What can we learn from ab initio learners? *Discours, acquisition et didactique des langues, les termes d'un dialogue*, Collection Universités, série Sciences du Langage:71–85.

[Rodríguez et al., 2013] Rodríguez, R., Liu, H., and Martínez, L. (2013). A Fuzzy Representation for the Semantics of Hesitant Fuzzy Linguistic Term Sets. In *8th International Conference on Intelligent Systems and Knowledge Engineering (ISKE)*, pages 745–757.

[Rodríguez et al., 2016a] Rodríguez, R. M., Bedregal, B., Bustince, H., Dong, Y. C., Farhadinia, B., Kahraman, C., Martínez, L., Torra, V., Xu, Y., Xu, Z., et al. (2016a). A position and perspective analysis of hesitant fuzzy sets on information fusion in decision making. Towards high quality progress. *Information Fusion*, 29:89–97.

[Rodríguez et al., 2016b] Rodríguez, R. M., Labella, Á., and Martínez, L. (2016b). An overview on fuzzy modelling of complex linguistic preferences in decision making. *International Journal of Computational Intelligence Systems*, 9(suppl. 1):81–94.

[Rodríguez et al., 2011] Rodríguez, R. M., Martínez, L., and Herrera, F. (2011). Hesitant fuzzy linguistic term sets. In Wang, Y. and Li, T., editors, *Foundations of Intelligent Systems: Proceedings of the Sixth International Conference on Intelligent Systems and Knowledge Engineering, Shanghai, China, Dec 2011 (ISKE2011)*, pages 287–295, Berlin, Heidelberg. Springer Berlin Heidelberg.

[Rodríguez et al., 2012] Rodríguez, R. M., Martínez, L., and Herrera, F. (2012). Hesitant Fuzzy Linguistic Term Sets for Decision Making. *Fuzzy Systems, IEEE Transactions on*, 20(1):109–119.

[Rodríguez et al., 2014] Rodríguez, R. M., Martínez, L., Torra, V., Xu, Z. S., and Herrera, F. (2014). Hesitant Fuzzy Sets: State of the Art and Future Directions. *International Journal of Intelligent Systems*, 29(6):495–524.

[Sapir, 1944] Sapir, E. (1944). Grading, a study in semantics. *Philosophy of science*, 11(2):93–116.

[Scowen, 1993] Scowen, R. S. (1993). Extended BNF-a generic base standard. Technical report, Technical report, ISO/IEC 14977.

[Torra, 2010] Torra, V. (2010). Hesitant fuzzy sets. *International Journal of Intelligent Systems*, 25(6):529–539.

[Truck and Abchir, 2014] Truck, I. and Abchir, M.-A. (2014). Towards a classification of hesitant operators in the 2-tuple linguistic model. *International Journal of Intelligent Systems*, 29(6):560–578.

[Truck and Akdag, 2006] Truck, I. and Akdag, H. (2006). Manipulation of qualitative degrees to handle uncertainty: Formal models and applications. *International Journal of Knowledge and Information Systems*, 9(4):385–411.

[Truck and Akdag, 2009] Truck, I. and Akdag, H. (2009). A tool for aggregation with words. *International Journal of Information Sciences, Special Issue: Linguistic Decision Making: Tools and Applications*, 179(14):2317–2324.

[Truck et al., 2014] Truck, I., Durand, M. E., and Watorek, M. (2014). Toward a Polish Intelligent Virtual Tutor : An Overview Of Existing Work. In *The 11th International FLINS Conference on Computational Intelligence in Decision and Control*, pages 472–477, Joao Pessoa, Brazil.

[Wang et al., 2015] Wang, J., Wang, J.-Q., Zhang, H.-Y., and Chen, X.-H. (2015). Multi-criteria decision-making based on hesitant fuzzy linguistic term sets: An outranking approach. *Knowledge-Based Systems*, 86:224–236.

[Zadeh, 1994] Zadeh, L. A. (1994). Fuzzy logic, neural networks, and soft computing. *Communications of the ACM*, 37(3):77–84.

[Zimmermann, 1996] Zimmermann, H.-J. (1996). *Fuzzy Set Theory&Mdash;and Its Applications (3rd Ed.)*. Kluwer Academic Publishers, Norwell, MA, USA.