Consensus-based clustering under hesitant qualitative assessments

José Luis García-Lapresta

PRESAD Research Group, IMUVA, Dept. de Economía Aplicada, Universidad de Valladolid, Spain

David Pérez-Román

PRESAD Research Group, Dept. de Organización de Empresas y Comercialización e Investigación de Mercados, Universidad de Valladolid, Spain

Abstract

In this paper, we consider that agents judge the feasible alternatives through linguistic terms –when they are confident in their opinions– or linguistic expressions formed by several consecutive linguistic terms –when they hesitate. In this context, we propose an agglomerative hierarchical clustering process where the clusters of agents are generated by using a distance-based consensus measure.

1. Introduction

In different decision-making problems, agents show their opinions about a set of alternatives and then an aggregation procedure generates an outcome (a winning alternative, several winning alternatives, a ranking on the set of alternatives, etc.). The opinions given by the agents may be provided in different ways: the favorite alternative, a subset of acceptable alternatives, a ranking on the set of alternatives, an assessment for each alternative, etc.

In the case of agents assess independently each alternative, the corresponding assessments can be of different nature depending of the context: numerical values, intervals of real numbers, fuzzy numbers, linguistic terms, etc.

Usually people prefer to handle the imprecision through linguistic terms rather than with exact numerical values. Since opinions are imprecise, trying to represent them by using a precise number is meaningless (see Zimmer [38]). Wallsten *et al.* [34] have shown empirically how most people are more comfortable using words rather than numbers to describe probabilities. On the other hand, for evaluating qualitative aspects, as the comfort of a car, is more

---

Email addresses: `lapresta@eco.uva.es` (José Luis García-Lapresta), `david@emp.uva.es` (David Pérez-Román)

appropriate to use linguistic terms than quantitative values (see Levrat *et al.* [23]).

In the context of voting theory, it is worth mentioning that Balinski and Laraki [4, 5] have proposed a voting system called *Majority Judgment* where voters assess the alternatives through linguistic terms ('excellent', 'very good', 'good', 'acceptable', 'poor' and 'to reject'). See Balinski and Laraki [6] for an experimental analysis of their proposal.

Clearly, the use of linguistic assessments allows the agents to show their imprecise opinions in a suitable way (see Zadeh [36, 37]). However, in some situations agents, even experts assessing alternatives within their field of expertise may have doubts about which feasible linguistic term would assign to an alternative (see Agell *et al.* [2] and García-Lapresta *et al.* [16]). For this reason, it is interesting to allow the agents to judge in a more imprecise way, giving them the option of assigning several linguistic terms[1].

Our approach concerning the imprecision is based on an adaptation of the *absolute order of magnitude spaces* introduced by Travé-Massuyès and Dague [31] and Travé-Massuyès and Piera [32]; more specifically in the extensions devised by Roselló *et al.* [25, 26, 27] (see also Agell *et al.* [1] and Falcó *et al.* [13, 14]).

In many contexts and disciplines, observations or objects are grouped in clusters in such a way that elements within each cluster are similar to one another with respect to an attribute. Then, objects are classified in homogeneous clusters and the objects in a cluster are more similar to each other than they are to an object belonging to a different cluster (see Jain *et al.* [21] and Everitt *et al.* [12], among others).

In this paper, we have devised a clustering procedure in the context of hesitant qualitative assessments[2]. More specifically, we have analyzed how agents can be grouped into clusters when such agents rate alternatives by means of linguistic terms from a predetermined linguistic scale[3], if the agents are confident in their opinions. If they are not confident about which term to use, they are allowed to use a linguistic expression generated by several consecutive linguistic terms.

In the cluster formation, we have considered that the similarity between two groups of agents with respect to an alternative is the degree of consensus[4]

---

[1]In this way, Torra [29] introduced the notion of *hesitant fuzzy set* as a generalization of the intuitionistic fuzzy sets and fuzzy multisets, by allowing the agents to assign several values for the membership function to each alternative.

[2]Other clustering procedures have been used in decision-making. For instance, Valls and Torra [33] proposed one for grouping alternatives taking into account the assessments obtained in different criteria; García-Lapresta and Pérez-Román [18] proposed another one for grouping individuals who rank alternatives through weak orders.

[3]We have assumed that the linguistic scale is uniformly and symmetrically distributed. Thus, the distance between consecutive linguistic terms is assumed to be constant.

[4]The notion of *consensus measure* was introduced by Bosch [8] in the context of linear orders. Additionally, Bosch [8] and Alcalde-Unzu and Vorsatz [3] provided axiomatic characterizations of several consensus measures in the context of linear orders. García-Lapresta

in the merged group. That consensus is measured taking into account the distances between all the pairs of individual assessments over the alternative that is been evaluated (according to a previously fixed metric on the set of linguistic expressions).

Given an alternative, the agglomerative hierarchical clustering process starts by joining the two agents that are more similar with respect to that alternative (ties are broken in a lexicographic manner). The next cluster is created by merging a new agent to the previous cluster or by joining two agents in such a way that the similarity is maximized (again ties are broken in a lexicographic manner). The process continues until the last cluster join all the agents.

The overall agglomerative hierarchical clustering process follows the same pattern that the previous one, but now taking all the alternatives into account. The (overall) degree of consensus in a group of agents is defined as the outcome given by an aggregation function to the degrees of consensus in that group of agents with respect to all the alternatives. The overall similarity between two groups of agents is defined as the (overall) degree of consensus in the merged group.

We have proven some properties that both consensus measures satisfy. These properties ensure that the similarities that generate the clustering processes work in a suitable way.

The information provided by the clustering procedure is interesting for knowing what is the agreement of the agents over each alternative and also what is the overall agreement. It is also useful for having a picture on what agents are more similar regarding a specific alternative as well as from an overall perspective. In this way, possible outliers are easily detected.

A clear application of our proposal is in consensus reaching processes, where a human or virtual moderator invites the agents to modify their opinions for increasing the agreement before applying a decision-making procedure (see, for instance, Saint and Lawson [28] and Eklund *et al.* [9]).

In order to show how our proposal works, we illustrate the procedure from the data of the field experiment carried out by Agell *et al.* [2].

The rest of the paper is organized as follows. Section 2 is devoted to introduce some notation and basic notions. Section 3 includes the clustering procedure. Section 4 contains the illustrative example. Finally, Section 5 concludes with some remarks.

2. Notation and basic notions

We now show some concepts and notation we need for introducing our proposal (for more details, see Falcó *et al.* [14]).

Let $A = \{1, \ldots, m\}$, with $m \geq 2$, be a set of agents and let $X = \{x_1, \ldots, x_n\}$, with $n \geq 2$, be the set of alternatives which have to be evaluated. Under total

_____

and Pérez-Román [17] extended that notion to the context of weak orders and they analyzed a class of consensus measures generated by distances.

3

certainty, each agent assigns a linguistic term to every alternative within a linguistic ordered scale $L = \{l_1, \ldots, l_g\}$, where $l_1 < l_2 < \cdots < l_g$. The linguistic scale is balanced and equispaced between consecutive terms. The elements of L can be linguistic terms as 'very good', 'good', etc.

Based on the *absolute order of magnitude spaces* introduced by Travé-Massuyès and Piera [32], we define the *set of linguistic expressions* as follows

$$\mathbb{L} = \{[l_h, l_k] \mid l_h, l_k \in L, \ 1 \leq h \leq k \leq g\},$$

where $[l_h, l_k] = \{l_h, l_{h+1}, \ldots, l_k\}$. Since $[l_h, l_h] = \{l_h\}$, this linguistic expression can be replaced by the linguistic term l_h. In this way, $L \subset \mathbb{L}$.

Example 1. Consider the set of linguistic terms $L = \{l_1, l_2, l_3, l_4, l_5\}$ with the meanings given in Table 1.

l_1	l_2	l_3	l_4	l_5
very bad	bad	acceptable	good	very good

Table 1: Meaning of the linguistic terms.

Linguistic expressions have clear meanings. For instance, $[l_1, l_3]$ means 'between very bad and acceptable', $[l_4, l_5]$ means 'between good and very good', or 'at least good', etc.

Taking into account the approach introduced in Roselló *et al.* [27], the set of all the linguistic expressions can be represented by a graph $G_{\mathbb{L}}$. In the graph, the lowest layer represents the linguistic terms $l_h \in L \subset \mathbb{L}$, the second layer represents the linguistic expressions created by two consecutive linguistic terms $[l_h, l_{h+1}]$, the third layer represents the linguistic expressions generated by three consecutive linguistic terms $[l_h, l_{h+2}]$, and so on up to last layer where we represent the linguistic expression $[l_1, l_g]$. As a result, the higher an element is, the more imprecise it becomes.

The vertices in $G_{\mathbb{L}}$ are the elements of $\mathbb{L}$ and the edges $\mathcal{E} - \mathcal{F}$, where $\mathcal{E} = [l_h, l_k]$ and $\mathcal{F} = [l_h, l_{k+1}]$, or $\mathcal{E} = [l_h, l_k]$ and $\mathcal{F} = [l_{h+1}, l_k]$.

Fig. 1 shows the graph representation of Example 1.

When an agent is confident about his opinion on an alternative, he can assign a linguistic term $l_h \in L$ to this alternative. However, if he is unconfident about his opinion, he might use a linguistic expression $[l_h, l_k] \in \mathbb{L}$, with $h < k$. For more details, see Roselló *et al.* [25, 26, 27].

In Falcó *et al.* [13, 14] the mentioned approach has been used for introducing and analyzing some voting systems where agents show their linguistic assessments over the alternatives under uncertainty.

A *profile* V is a matrix (v_i^a) consisting of m rows and n columns of linguistic expressions, where the element $v_i^a \in \mathbb{L}$ represents the linguistic assessment given
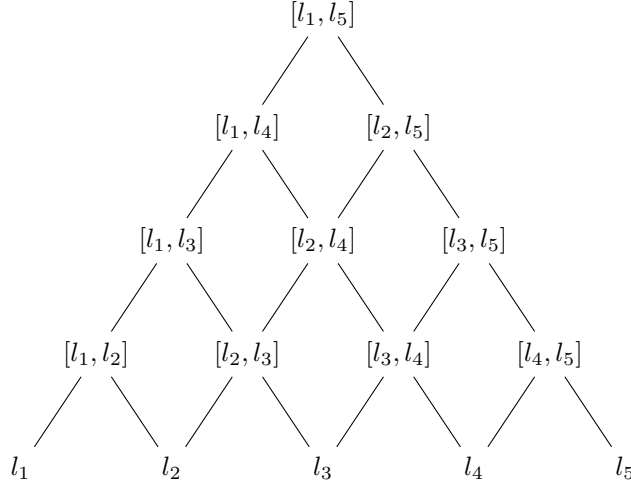
4

Figure 1: Graph representation of the linguistic expressions for $g = 5$.

by the agent $a \in A$ to the alternative $x_i \in X$. Then,

$$
V = \begin{pmatrix}
v_1^1 & \cdots & v_i^1 & \cdots & v_n^1 \\
\cdots & \cdots & \cdots & \cdots & \cdots \\
v_1^a & \cdots & v_i^a & \cdots & v_n^a \\
\cdots & \cdots & \cdots & \cdots & \cdots \\
v_1^m & \cdots & v_i^m & \cdots & v_n^m
\end{pmatrix} = (v_i^a).
$$

The distance between two linguistic expressions $\mathcal{E}, \mathcal{F} \in \mathbb{L}$ is defined as the geodesic distance in the graph $G_\mathbb{L}$ between their associated vertices and it is denoted by $d(\mathcal{E}, \mathcal{F})$. The geodesic distance between two vertices in a graph is the number of edges in one of the shortest paths connecting them[5]. These geodesic distances can be analytically expressed through

$$
d\big([l_h, l_k], [l_{h'}, l_{k'}]\big) = |h - h'| + |k - k'|.
$$

Notice that the geodesic distance between consecutive linguistic terms is always 2: $d(l_h, l_{h+1}) = d\big([l_h, l_h], [l_{h+1}, l_{h+1}]\big) = 2$, for every $h \in \{1, \ldots, g - 1\}$. This fact is consistent with the assumption that L is uniformly and symmetrically distributed.

3. Clustering

There are many clustering algorithms because the notion of cluster cannot be precisely defined (see Ward [35], Jain *et al.* [21], Everitt *et al.* [12] and

---

[5]In Falcó *et al.* [13] these distances are modified through two parameters for penalizing the increase of imprecision.

Estivill-Castro [11], among others).

Hierarchical clustering deals with how objects should be grouped into clusters. Given a distance matrix, linkage objects can be computed through a criterion to compute distances between groups. There are common and basic criteria such as *single-link*, *complete-link*, *average-link* and *centroid distance criterion*. These criteria are mostly based on a measure of (dis)similarity between sets of observations. Mostly, this is achieved through a metric that measures the distance between pairs of observations as well as a linkage criterion which measures the (dis)similarity between sets, defined as a function of the pairwise distances of observations in the sets.

In our proposal, we will use consensus measures for defining the (dis)similarities and also for defining the linkage criterion.

In single-link and complete-link, the (dis)similarity of two clusters is defined as the (dis)similarity of their most (dis)similar elements. These criteria do not consider what happens to the other elements of the clusters.

Average-link and centroid distance criterion evaluate cluster quality based on all (dis)similarities between members, thus avoiding the pitfalls of the single-link and complete-link criteria. However, these criteria increase the overall time complexity because it is necessary to obtain a representative element of each new cluster. It is important to mention that we do not represent each cluster by a single representative element, but we use all the information included in the cluster.

For the above four criteria, its time complexity is $\Theta(N^2 \log N)$ (see Manning *et al.* [24, Chapter 17]). The consensus criterion in the alternatives level has the same time complexity. The structure of the algorithm matches the above criteria, but the similarity function for potential merge pairs considers what happens to the other members in the clusters. This procedure does not increase time complexity.

3.1. Measuring consensus

For measuring the consensus over an alternative in a group of agents (Definition 1), we propose a method based on the Gini coefficient [19] (a classical dispersion measure over income distributions that is based on the sum of absolute differences between individual incomes) and the consensus measures introduced by García-Lapresta and Pérez-Román [17] in the context of weak orders (see also Erdamar *et al.* [10] for an extension of this notion to the preference-approval setting).

With $\mathcal{P}_2(A) = \{I \subseteq A \mid \#I \geq 2\}$ we denote the family of subsets of at least two agents. The inverse of a linguistic expression $\mathcal{E} = [l_h, l_k] \in \mathbb{L}$ is defined as $\mathcal{E}^{-1} = [l_{g+1-k}, l_{g+1-h}]$. Given a profile $V = (v_i^a)$, its inverse is defined as $V^{-1} = (u_i^a)$ with $u_i^a = (v_i^a)^{-1}$. Given a permutation π on A, a profile $V = (v_i^a)$ and a subset of agents $I \in \mathcal{P}_2(A)$, the profile $V^\pi = (u_i^a)$ is defined as $u_i^a = v_i^{\pi(a)}$; and $I^\pi = \{\pi^{-1}(a) \mid a \in I\}$, i.e., $a \in I^\pi \iff \pi(a) \in I$.

Following the pattern of the consensus measures introduced by García-Lapresta and Pérez-Román [17], we now define the degree of consensus over

an alternative through distances between linguistic expressions (instead of weak orders) and the corresponding normalization.

Definition 1. *Given a profile* $V = (v_i^a)$*, the* degree of consensus *in a subset of agents* $I \in \mathcal{P}_2(A)$ *over the alternative* $x_i \in X$ *is defined as*

$$
C(V, I, x_i) = 1 - \frac{\displaystyle\sum_{\substack{a,b \in I \\ a < b}} d\left(v_i^a, v_i^b\right)}{\displaystyle\binom{|I|}{2} \cdot 2 \cdot (g-1)}.
$$

The following result includes some interesting properties of the proposed consensus measure. Normalization means that consensus is always between 0 and 1. Anonymity requires symmetry with respect to agents. Unanimity means that the maximum consensus is only achieved when all opinions are the same. Positiveness means that consensus is always positive whenever more than two individuals are involved. Maximum dissension means that in each subset of two agents, the minimum consensus is only reached whenever the assessments of agents are extreme and each one is the inverse of the other. Reciprocity means that if all individual assessments are reversed, then the consensus does not change.

Proposition 1. *For every profile* $V = (v_i^a)$*, every subset of agents* $I \in \mathcal{P}_2(A)$ *and every alternative* $x_i \in X$*, the following statements hold:*

1. *Normalization:* $C(V, I, x_i) \in [0, 1]$.
2. *Anonymity:* $C(V^\pi, I^\pi, x_i) = C(V, I, x_i)$ *for every permutation* π *on* A.
3. *Unanimity:* $C(V, I, x_i) = 1 \Leftrightarrow \left(v_i^a = v_i^b \text{ for all } a, b \in I\right)$.
4. *Positiveness: If* $\#I > 2$*, then* $C(V, I, x_i) > 0$.
5. *Maximum dissension:* $C(V, \{a, b\}, x_i) = 0 \Leftrightarrow \left(\left(v_i^a = l_1 \text{ and } v_i^b = l_g\right) \text{ or } \left(v_i^a = l_g \text{ and } v_i^b = l_1\right)\right)$.
6. *Reciprocity:* $C(V^{-1}, I, x_i) = C(V, I, x_i)$.

PROOF: Let $V = (v_i^a)$ be a profile.

1. Since $2 \cdot (g-1)$ is the maximum distance between linguistic expressions and $\binom{|I|}{2}$ is the number of terms in the numerator of the quotient appearing in $C(V, I, x_i)$, we have that $C(V, I, x_i) \in [0, 1]$ for all $I \in \mathcal{P}_2(A)$ and $x_i \in X$.

2. Let $V = (v_i^a)$ and $V^\pi = (u_i^a)$ with $u_i^a = v_i^{\pi(a)}$. Then, we have

$$C(V^\pi, I^\pi, x_i) = 1 - \frac{\displaystyle\sum_{\substack{a,b \in I^\pi \\ a<b}} d\left(u_i^a, u_i^b\right)}{\dbinom{|I^\pi|}{2} \cdot 2 \cdot (g-1)} =$$

$$1 - \frac{\displaystyle\sum_{\substack{\pi(a),\pi(b) \in I \\ \pi(a)<\pi(b)}} d\left(v_i^{\pi(a)}, v_i^{\pi(b)}\right)}{\dbinom{|I|}{2} \cdot 2 \cdot (g-1)} = C(V, I, x_i).$$

3. It follows from the fact that d is a distance.
4. If $C(V, I, x_i) = 0$, then $d(v_i^a, v_i^b)$ is maximum for all $a, b \in I$. Since the maximum distance between linguistic expressions is only reached when comparing l_1 and l_g, we have $v_i^a \in \{l_1, l_g\}$ for every $a \in I$. Consequently, $\#I = 2$.
5. It follows from the fact that the maximum distance between linguistic expressions is only reached when comparing l_1 and l_g.
6. It is due to the fact that $d\left(\mathcal{E}^{-1}, \mathcal{F}^{-1}\right) = d(\mathcal{E}, \mathcal{F})$ for all $\mathcal{E}, \mathcal{F} \in \mathbb{L}$. ∎

Once we know the degree of consensus of a subset of agents over each alternative, it would be interesting to determine the overall consensus of this group of agents. For this matter, we will use aggregation functions (see Fodor and Roubens [15], Beliakov *et al.* [7], Torra and Narukawa [30] and Grabisch *et al.* [20]).

Definition 2. *An* aggregation function *is a mapping* $F : [0,1]^n \longrightarrow [0,1]$ *that is monotonic in each component and satisfies the boundary conditions* $F(0, \ldots, 0) = 0$ *and* $F(1, \ldots, 1) = 1$. *Moreover,*

1. F *is* symmetric *if for all* $(t_1, \ldots, t_n) \in [0,1]^n$ *and permutation* σ *on* $\{1, \ldots, n\}$ *it holds* $F(t_{\sigma(1)}, \ldots, t_{\sigma(n)}) = F(t_1, \ldots, t_n)$.
2. F *is* idempotent *if for every* $t \in [0,1]$ *it holds* $F(t, \ldots, t) = t$.
3. F *is* compensative *if for every* $(t_1, \ldots, t_n) \in [0,1]^n$ *it holds*

$$\min(t_1, \ldots, t_n) \leq F(t_1, \ldots, t_n) \leq \max(t_1, \ldots, t_n).$$

It is easy to see that every idempotent aggregation function is compensative, and viceversa.

Typical symmetric, idempotent (hence compensative) aggregation functions are the family of OWA operators (including the arithmetic mean, the median, the maximum, the minimum, the mid-range and the family of trimmed means, among others) and the family of quasiarithmetic means.

Definition 3. *Given an aggregation function* $F : [0,1]^n \longrightarrow [0,1]$, *the* degree of consensus relative to F *of a set of agents* $I \in \mathcal{P}_2(A)$ *in a profile* V *is defined as*

$$\mathcal{C}_F(V, I) = F(C(V, I, x_1), \ldots, C(V, I, x_n)).$$

Although the properties of $\mathcal{C}_F(V, I)$ depend on the chosen aggregation function F, in the following result we show some properties that are always satisfied. The new property neutrality means symmetry with respect to alternatives.

Given a permutation σ on $\{1, \ldots, n\}$ and a profile $V = (v_i^a)$, the profile $V_\sigma = (u_i^a)$ is defined as $u_i^a = v_{\sigma(i)}^a$.

Proposition 2. *For every aggregation function* $F : [0,1]^n \longrightarrow [0,1]$, *every profile* V *and every subset of agents* $I \in \mathcal{P}_2(A)$, *the following statements hold:*

1. *Normalization:* $\mathcal{C}_F(V, I) \in [0, 1]$.
2. *Anonymity:* $\mathcal{C}_F(V^\pi, I^\pi) = \mathcal{C}_F(V, I)$ *for every permutation* π *on* A.
3. *Unanimity: If* $v_i^a = v_i^b$ *for all* $x_i \in X$ *and* $a, b \in I$, *then* $\mathcal{C}_F(V, I) = 1$.
4. *Positiveness: If* $\#I > 2$ *and* F *is idempotent, then* $\mathcal{C}_F(V, I) > 0$.
5. *Maximum dissension: If* $\left(\left(v_i^a = l_1 \text{ and } v_i^b = l_g \right) \text{ or } \left(v_i^a = l_g \text{ and } v_i^b = l_1 \right) \right)$ *for all* $x_i \in X$, *then* $\mathcal{C}_F(V, \{a, b\}) = 0$.
6. *Neutrality: If* F *is symmetric, then* $\mathcal{C}_F(V_\sigma, I) = \mathcal{C}_F(V, I)$ *for every permutation* σ *on* $\{1, \ldots, n\}$.
7. *Reciprocity:* $\mathcal{C}_F(V^{-1}, I) = \mathcal{C}_F(V, I)$.
8. *Compensativeness: If* F *is idempotent, then*

$$\min(C(V, I, x_1), \ldots, C(V, I, x_n)) \leq \mathcal{C}_F(V, I) \leq \max(C(V, I, x_1), \ldots, C(V, I, x_n)).$$

PROOF: By Proposition 1 and the properties of F. ■

3.2. Clustering in the alternatives level

We now introduce a similarity function for each alternative that is constructed through the distance-based degrees of consensus $C(V, I, x_i)$. With these functions we will formulate our proposal of clustering in the alternatives level. With $\mathcal{P}(A)$ we denote the power set of A, i.e., the set of all the subsets of agents in A.

Definition 4. *Given a profile* V *and an alternative* $x_i \in X$, *the* similarity function relative to x_i

$$S^i : \left(\mathcal{P}(A) \setminus \{\emptyset\} \right)^2 \longrightarrow [0, 1]$$

is defined as

$$S^i(I, J) = \begin{cases} C(V, I \cup J, x_i), & \text{if } \#(I \cup J) \geq 2, \\ 1, & \text{if } \#(I \cup J) = 1. \end{cases}$$

9

Obviously, $S^i(I,J) = S^i(J,I)$ for all $I,J \in \left(\mathcal{P}(A) \setminus \{\emptyset\}\right)$. Moreover, the mapping $\delta^i : A^2 \longrightarrow [0,1]$ defined as $\delta^i(a,b) = 1 - S^i(\{a\},\{b\})$ is a metric:

$$\delta^i(a,b) = \frac{d\left(v_i^a, v_i^b\right)}{2 \cdot (g-1)}.$$

Given a profile $V = (v_i^a)$, the *distance matrix* $D_i = \left(d\left(v_i^a, v_i^b\right)\right)$ contains all the distances between the assessments given by the agents to the alternative x_i, $i = 1,\ldots,n$. Taking into account these matrices, all the consensus and similarities computations can be made.

For each alternative $x_i \in X$, the agglomerative hierarchical clustering procedure we propose consists of a sequential process addressed by the following stages:

1. The initial clustering is $\mathcal{A}_0^i = \{\{1\},\ldots,\{m\}\}$.
2. Calculate the similarities between all the pairs of agents, $S^i(\{a\},\{b\})$ for all $a,b \in A$.
3. Select the two agents $a,b \in A$ that maximize these similarities in a lexicographic manner and construct the first cluster $A_1^i = \{a,b\}$.
4. The new clustering is $\mathcal{A}_1^i = \left(\mathcal{A}_0^i \setminus \{\{a\},\{b\}\}\right) \cup \{A_1^i\}$.
5. Calculate the similarities $S^i(A_1^i, \{c\})$ and take into account the previously computed similarities $S^i(\{c\},\{d\})$, for all $\{c\},\{d\} \in \mathcal{A}_1^i$.
6. Select the two elements of $\mathcal{A}_1^i$ with maximum value of S^i in a lexicographic manner and construct the second cluster A_2^i.
7. Proceed as in previous items until obtaining the next clustering $\mathcal{A}_2^i$.

The process continues in the same way until obtaining the last cluster, $\mathcal{A}_{m-1}^i = \{A\}$.

3.3. Overall clustering

Once the clustering in the alternatives level has been formulated, we now introduce an overall clustering procedure that is based on the consensus degrees $\mathcal{C}_F(V,I)$ for a fixed aggregation function F.

Definition 5. *Given a profile V and aggregation function $F : [0,1]^n \longrightarrow [0,1]$, the* overall similarity function associated with F,

$$S_F : \left(\mathcal{P}(A) \setminus \{\emptyset\}\right)^2 \longrightarrow [0,1],$$

is defined as

$$S_F(I,J) = \begin{cases} \mathcal{C}_F(V, I \cup J), & \text{if } \#(I \cup J) \geq 2, \\ 1, & \text{if } \#(I \cup J) = 1. \end{cases}$$

Obviously, $S_F(I, J) = S_F(J, I)$ for all $I, J \in \left(\mathcal{P}(A) \setminus \{\emptyset\}\right)$.

The overall clustering procedure follows the same pattern that the one introduced in the alternatives level, but now using the overall similarity function S_F instead of S^i, starting from $\mathcal{A}_0 = \{\{1\}, \dots, \{m\}\}$ and following until $\mathcal{A}_{m-1} = \{A\}$.

Dendrograms are a useful tool for viewing the hierarchical dynamic process. Taking into account the visual information provided by a dendrogram, it is possible to decide when to stop the clustering formation. But if a fixed number of clusters has to be constructed, the process should stop whenever this number of clusters is reached.

4. An illustrative example

In order to illustrate our proposal of clustering, we consider the field experiment carried out by Agell *et al.* [2] where five experts assessed five fruits with respect to their suitability to combine with dark chocolate.

The five experts of $A = \{1, 2, 3, 4, 5\}$ evaluated the five fruits of $X = \{x_1, x_2, x_3, x_4, x_5\}$ (see Table 2) through the linguistic expressions of $\mathbb{L}$ generated by the set of five linguistic terms $L = \{l_1, l_2, l_3, l_4, l_5\}$ whose meanings are in Table 3.

x_1	x_2	x_3	x_4	x_5
red-bilberry	gooseberry	west Indian cherry	cashew fruit	guava

Table 2: Fruits.

l_1	l_2	l_3	l_4	l_5
it does not combine at all	it does not combine well	combines well	combines very well	it is an excellent combination

Table 3: Meaning of the linguistic terms.

Table 4 contains the assessments given by the experts.

	x_1	x_2	x_3	x_4	x_5
1	$[l_4, l_5]$	$[l_3, l_4]$	l_1	$[l_4, l_5]$	l_1
2	l_5	$[l_1, l_2]$	$[l_2, l_3]$	l_3	$[l_1, l_2]$
3	l_5	l_3	l_4	l_5	l_2
4	l_5	$[l_2, l_3]$	l_4	l_5	$[l_1, l_2]$
5	l_5	$[l_1, l_2]$	l_2	$[l_1, l_5]$	l_3

Table 4: Assessments given by the experts.

Notice that the experts provided 25 assessments: 15 of them were single linguistic terms, 9 linguistic expressions with two consecutive linguistic terms

11

and one linguistic expression with all the linguistic terms. It is important emphasizing that even experts may hesitate when they assess alternatives (in this experiment 40% of the cases).

Taking into account the individual assessments included in Table 4, the distance matrices $D_i = \left(d\left(v_i^a, v_i^b \right) \right)$, $i = 1, \ldots, 5$, will be used for all the needed computations in the clustering process.

$$
D_1 = \begin{pmatrix} 0 & 1 & 1 & 1 & 1 \\ 1 & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 \end{pmatrix}, \quad
D_2 = \begin{pmatrix} 0 & 4 & 1 & 2 & 4 \\ 4 & 0 & 3 & 2 & 0 \\ 1 & 3 & 0 & 1 & 3 \\ 2 & 2 & 1 & 0 & 2 \\ 4 & 0 & 3 & 2 & 0 \end{pmatrix},
$$

$$
D_3 = \begin{pmatrix} 0 & 3 & 6 & 6 & 2 \\ 3 & 0 & 3 & 3 & 1 \\ 6 & 3 & 0 & 0 & 4 \\ 6 & 3 & 0 & 0 & 4 \\ 2 & 1 & 4 & 4 & 0 \end{pmatrix}, \quad
D_4 = \begin{pmatrix} 0 & 3 & 1 & 1 & 3 \\ 3 & 0 & 4 & 4 & 4 \\ 1 & 4 & 0 & 0 & 4 \\ 1 & 4 & 0 & 0 & 4 \\ 3 & 4 & 4 & 4 & 0 \end{pmatrix},
$$

$$
D_5 = \begin{pmatrix} 0 & 1 & 2 & 1 & 4 \\ 1 & 0 & 1 & 0 & 3 \\ 2 & 1 & 0 & 1 & 2 \\ 1 & 0 & 1 & 0 & 3 \\ 4 & 3 & 2 & 3 & 0 \end{pmatrix}.
$$

When considering the whole set of experts, the consensus degrees over the five alternatives are as follows:

$$
C(V, A, x_1) = 0.95 > C(V, A, x_5) = 0.787 > C(V, A, x_4) = 0.75 >
$$
$$
C(V, A, x_2) = 0.725 > C(V, A, x_3) = 0.6 \,.
$$

According to the decision procedure considered by Agell *et al.* [2], the social ranking of the five alternatives is $x_1 \succ x_4 \succ x_3 \succ x_2 \succ x_5$. Thus, the maximum consensus is reached over the best alternative.

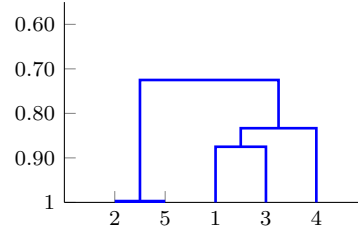After some computations, we obtain the following clusterings in the alternatives level (see also Fig. 2):

$$
\mathcal{A}_0^1 = \{\{1\}, \{2\}, \{3\}, \{4\}, \{5\}\} \qquad \mathcal{A}_0^2 = \{\{1\}, \{2\}, \{3\}, \{4\}, \{5\}\}
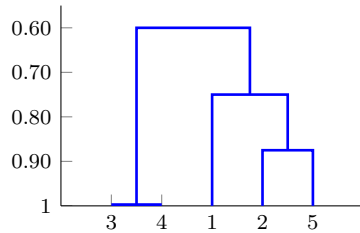$$
$$
\mathcal{A}_1^1 = \{\{2, 3\}, \{1\}, \{4\}, \{5\}\} \qquad \mathcal{A}_1^2 = \{\{2, 5\}, \{1\}, \{3\}, \{4\}\}
$$
$$
\mathcal{A}_2^1 = \{\{2, 3, 4\}, \{1\}, \{5\}\} \qquad \mathcal{A}_2^2 = \{\{2, 5\}, \{1, 3\}, \{4\}\}
$$
$$
\mathcal{A}_3^1 = \{\{2, 3, 4, 5\}, \{1\}\} \qquad \mathcal{A}_3^2 = \{\{2, 5\}, \{1, 3, 4\}\}
$$
$$
\mathcal{A}_4^1 = \{\{1, 2, 3, 4, 5\}\} \qquad \mathcal{A}_4^2 = \{\{1, 2, 3, 4, 5\}\}
$$

$$\mathcal{A}_0^3 = \{\{1\}, \{2\}, \{3\}, \{4\}, \{5\}\} \qquad \mathcal{A}_0^4 = \{\{1\}, \{2\}, \{3\}, \{4\}, \{5\}\}$$
$$\mathcal{A}_1^3 = \{\{3,4\}, \{1\}, \{2\}, \{5\}\} \qquad \mathcal{A}_1^4 = \{\{3,4\}, \{1\}, \{2\}, \{5\}\}$$
$$\mathcal{A}_2^3 = \{\{3,4\}, \{2,5\}, \{1\}\} \qquad \mathcal{A}_2^4 = \{\{1,3,4\}, \{2\}, \{5\}\}$$
$$\mathcal{A}_3^3 = \{\{3,4\}, \{1,2,5\}\} \qquad \mathcal{A}_3^4 = \{\{1,2,3,4\}, \{5\}\}$$
$$\mathcal{A}_4^3 = \{\{1,2,3,4,5\}\} \qquad \mathcal{A}_4^4 = \{\{1,2,3,4,5\}\}$$

$$\mathcal{A}_0^5 = \{\{1\}, \{2\}, \{3\}, \{4\}, \{5\}\}$$
$$\mathcal{A}_1^5 = \{\{2,4\}, \{1\}, \{3\}, \{5\}\}$$
$$\mathcal{A}_2^5 = \{\{1,2,4\}, \{3\}, \{5\}\}$$
$$\mathcal{A}_3^5 = \{\{1,2,3,4\}, \{5\}\}$$
$$\mathcal{A}_4^5 = \{\{1,2,3,4,5\}\}.$$
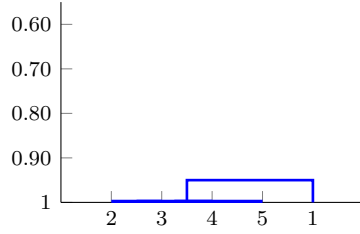
(a) Alternative x_1.

(b) Alternative x_2.

(c) Alternative x_3.
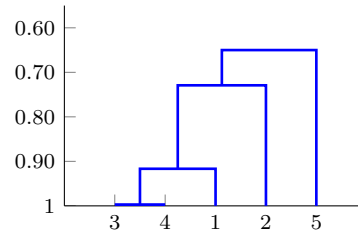
(d) Alternative x_4.
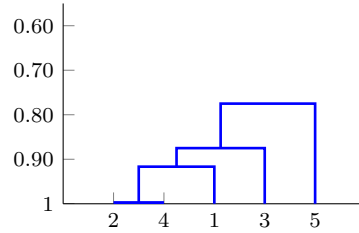
(e) Alternative x_5.

Figure 2: Dendrograms in the alternatives level.

Notice that the clustering processes are quite different depending on the alternatives. Thus, it makes sense to take into account the overall clustering by using different aggregation functions. In this example we have considered the two most important OWA operators, the arithmetic mean and the median, and additionally the 1-trimmed mean[6], a compromise between the arithmetic mean and the median.

When using the arithmetic mean, the overall clustering is as follows (see also Fig. 3):

$$\mathcal{A}_0 = \{\{1\}, \{2\}, \{3\}, \{4\}, \{5\}\}$$
$$\mathcal{A}_1 = \{\{3, 4\}, \{1\}, \{2\}, \{5\}\}$$
$$\mathcal{A}_2 = \{\{2, 3, 4\}, \{1\}, \{5\}\}$$
$$\mathcal{A}_3 = \{\{1, 2, 3, 4\}, \{5\}\}$$
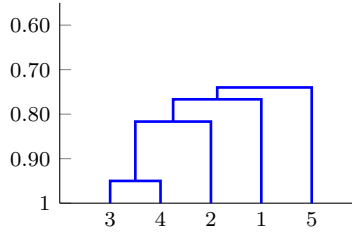$$\mathcal{A}_4 = \{\{1, 2, 3, 4, 5\}\}.$$

Figure 3: Overall dendrogram with the arithmetic mean.

When using the 1-trimmed mean, the overall clustering is as follows (see also Fig. 4):

$$\mathcal{A}_0 = \{\{1\}, \{2\}, \{3\}, \{4\}, \{5\}\}$$
$$\mathcal{A}_1 = \{\{3, 4\}, \{1\}, \{2\}, \{5\}\}$$
$$\mathcal{A}_2 = \{\{1, 3, 4\}, \{2\}, \{5\}\}$$
$$\mathcal{A}_3 = \{\{1, 3, 4\}, \{2, 5\}\}$$
$$\mathcal{A}_4 = \{\{1, 2, 3, 4, 5\}\}.$$

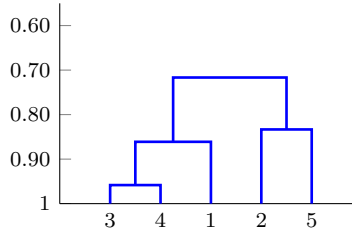[6]The arithmetic mean after discarding the highest and lowest values.

14

Figure 4: Overall dendrogram with the 1-trimmed mean.

When using the median, the overall clustering is as follows (see also Fig. 5):

$$\mathcal{A}_0 = \{\{1\}, \{2\}, \{3\}, \{4\}, \{5\}\}$$
$$\mathcal{A}_1 = \{\{3, 4\}, \{1\}, \{2\}, \{5\}\}$$
$$\mathcal{A}_2 = \{\{3, 4\}, \{2, 5\}, \{1\}\}$$
$$\mathcal{A}_3 = \{\{1, 3, 4\}, \{2, 5\}\}$$
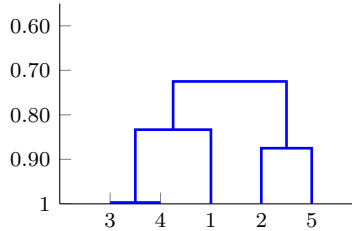$$\mathcal{A}_4 = \{\{1, 2, 3, 4, 5\}\}.$$

Figure 5: Overall dendrogram with the median.

Clearly, the aggregation function used in the overall clustering process has been determinant in the cluster formation. In this example, for obtaining the corresponding overall similarities, five values have been aggregated: the consensus degrees over each of the five alternatives in different subsets of agents. In each case, the arithmetic mean takes into account all the five values and the median only one, while the 1-trimmed mean considers the three central values. It is worth mentioning that the only coincidence among the three cases is that $\{3, 4\}$ is the first cluster. The clustering processes generated by the 1-trimmed mean and the median are similar (see Figures 4 and 5), but the one generated by the arithmetic mean is totally different (see Figure 3). The reason is that the arithmetic mean is very sensitive to extreme values. In our opinion, in this example the overall clustering generated by the 1-trimmed mean could be considered the most suitable of the three cases. But, in general, an important and controversial task of the decision maker is to fix an appropriate aggregation function for conducting the overall clustering process.

15

5. Concluding remarks

In this paper, we have proposed a clustering algorithm in a hesitant linguistic evaluation setting where agents assign linguistic terms to the alternatives or, when they are unconfident in their opinions, linguistic expressions formed by several consecutive linguistic terms.

The clustering procedure is based on a notion of consensus generated by geodesic distances between individual assessments. The main feature of this process is that agents are joined in a cluster when they maximize the consensus.

Clearly, other metrics can be used, in particular the parameterized family introduced by Falcó *et al.* [14]. The geodesic metric used in this paper, for simplicity reasons, is the degenerate case of the mentioned family, just when the parameters that penalize the imprecision are zero.

A possible application to the devised clustering procedure is to consensus reaching processes in the hesitant linguistic evaluation setting. They are dynamic and sequential procedures where a moderator tries to convince agents to change their opinions in several rounds for increasing the agreement in the group (see Kacprzyk *et al.* [22], Saint and Lawson [28] and Eklund *et al.* [9], among others).

Due to the monotonicity of aggregation functions, the overall consensus (Definition 3) is sensitive to the increase of the consensus degrees in the alternatives level (Definition 1). Thus, the moderator only needs to help the agents increase the agreement in each alternative. In this way, the corresponding dendrograms show who are the outliers. Then, the moderator may focus on these agents. Once a consensus threshold is reached, a group decision-making procedure can be carried out in order to rank the alternatives (within our framework those given by Falcó *et al.* [13, 14], for instance).

References

[1] Agell, N., Sánchez, M., Prats, F., Roselló, L., Ranking multi-attribute alternatives on the basis of linguistic labels in group decisions. Information Sciences 209, pp. 49-60, 2012.

[2] Agell, N., Sánchez, G., Sánchez, M., Ruiz, F.J., Selecting the best taste: a group decision-making application to chocolates design. Proceedings of the 2013 IFSA-NAFIPS Joint Congress, Edmonton, pp. 939-943, 2013.

[3] Alcalde-Unzu, J., Vorsatz, M., Measuring the cohesiveness of preferences: an axiomatic analysis. Social Choice and Welfare 41, pp. 965-988, 2013.

[4] Balinski, M., Laraki, R., A theory of measuring, electing and ranking. Proceedings of the National Academy of Sciences of the United States of America 104, pp. 8720-8725, 2007.

[5] Balinski, M., Laraki, R., Majority Judgment. Measuring, Ranking, and Electing. The MIT Press, Cambridge MA, 2011.

[6] Balinski, M., Laraki, R., Election by Majority Judgement: Experimental evidence, in: B. Dolez, B. Grofman, A. Laurent (Eds.), Situ and Laboratory Experiments on Electoral Law Reform: French Presidential Elections, Studies in Public Choice 25, Springer, Berlin, pp. 13-54, 2011.

[7] Beliakov, G., Pradera, A., Calvo, T., Aggregation Functions: A Guide for Practitioners. Springer, Heidelberg, 2007.

[8] Bosch, R., Characterizations of Voting Rules and Consensus Measures. Ph. D. Dissertation, Tilburg University, 2005.

[9] Eklund, P., Rusinowska, A., de Swart, H., Consensus reaching in committees. European Journal of Operational Research 178, pp. 185-193, 2007.

[10] Erdamar, B., García-Lapresta, J.L., Pérez-Román, D., Sanver, M.R., Measuring consensus in a preference-approval context. Information Fusion 17, pp. 14-21, 2014.

[11] Estivill-Castro, V., Why so many clustering algorithms: a position paper. ACM SIGKDD Explorations Newsletter 4, pp. 65-75, 2002.

[12] Everitt, B.S., Landau, S., Leese, M., Cluster Analysis, 4th Edition. Oxford University Press, New York, 2001.

[13] Falcó, E., García-Lapresta, J.L., Roselló, L., Aggregating imprecise linguistic expressions, in: P. Guo and W. Pedrycz (Eds.), Human-Centric Decision-Making Models for Social Sciences. Springer-Verlag, Berlin, 2014, pp. 97-113.

[14] Falcó, E., García-Lapresta, J.L., Roselló, L., Allowing agents to be imprecise: A proposal using multiple linguistic terms. Information Sciences 258, pp. 249-265, 2014.

[15] Fodor, J., Roubens, M., Fuzzy Preference Modelling and Multicriteria Decision Support. Kluwer Academic Publishers, Dordrecht, 1994.

[16] García-Lapresta, J.L., Aldavero, C., de Castro, S., A linguistic approach to multi-criteria and multi-expert sensory analysis. Proceedings of the 15th International Conference on Information Processing and Management of Uncertainty in Knowledge-Based Systems – IPMU 2014, Montpellier, 2014.

[17] García-Lapresta, J.L., Pérez-Román, D., Measuring consensus in weak orders, in: E. Herrera-Viedma, J.L. García-Lapresta, J. Kacprzyk, H. Nurmi, M. Fedrizzi, S. Zadròzny (Eds.), Consensual Processes, STUDFUZZ, vol. 267. Springer-Verlag, Berlin, 2011, pp. 213-234.

[18] García-Lapresta, J.L., Pérez-Román, D., Consensus-based hierarchical agglomerative clustering in the context of weak orders. Proceedings of the 2013 IFSA-NAFIPS Joint Congress, Edmonton, pp. 1010-1015, 2013.

[19] Gini, C., Variabilità e Mutabilità. Tipografia di Paolo Cuppini, Bologna, 1912.

[20] Grabisch, M., Marichal, J.L., Mesiar, R., Pap, E., Aggregation Functions. Cambridge University Press, Cambridge, 2009.

[21] Jain, A.K., Murty, M.N., Flynn, P.J., Data clustering: A review. ACM Computing Surveys 31 (3), pp. 264-323, 1999.

[22] Kacprzyk, J., Fedrizzi, M., Nurmi, H., Group decision making and consensus under fuzzy preferences and fuzzy majority. Fuzzy Sets and Systems 49, pp. 21-31, 1992.

[23] Levrat, E., Voisin, A., Bombardier, S., Bremont, J., Subjective evaluation of car seat comfort with fuzzy set techniques. International Journal of Intelligent Systems 12, pp. 891-913, 1997.

[24] Manning, C.D., Raghavan, P., Schütze, H., Introduction to Information Retrieval. Cambridge University Press, 2008.

[25] Roselló, L., Prats, F., Agell, N., Sánchez, M., Measuring consensus in group decisions by means of qualitative reasoning. International Journal of Approximate Reasoning 51, pp. 441-452, 2010.

[26] Roselló, L., Prats, F., Agell, N., Sánchez, M., A qualitative reasoning approach to measure consensus, in: E. Herrera-Viedma, J.L. García-Lapresta, J. Kacprzyk, H. Nurmi, M. Fedrizzi, S. Zadròzny (Eds.), Consensual Processes, STUDFUZZ, vol. 267. Springer-Verlag, Berlin, 2011, pp. 235-261.

[27] Roselló, L., Sánchez, M., Agell, N., Prats, F., Mazaira, F.A., Using consensus and distances between generalized multi-attribute linguistic assessments for group decision-making. Information Fusion 17, pp. 83-92, 2014.

[28] Saint, S., Lawson, J.R., Rules for Reaching Consensus: A Modern Approach to Decision Making. Jossey-Bass, San Francisco, 1994.

[29] Torra, V., Hesitant fuzzy sets. International Journal of Intelligent Systems 25, pp. 529-539, 2010.

[30] Torra, V., Narukawa, Y., Modeling Decisions: Information Fusion and Aggregation Operators. Springer, Berlin, 2007.

[31] Travé-Massuyès, L., Dague, P. (Eds.), Modèles et Raisonnements Qualitatifs. Hermes Science, Paris, 2003.

[32] Travé-Massuyès, L., Piera, N., The orders of magnitude models as qualitative algebras. Proceedings of the 11th International Joint Conference on Artificial Intelligence, Detroit, 1989.

[33] Valls, A. Torra, V., Using classification as an aggregation tool in MCDM. Fuzzy Sets and Systems 115, pp. 159-168, 2000.

[34] Wallsten, T.S., Budescu, D.V., Rapoport, A., Zwick, R., Forsyth, B., Measuring the vague meanings of probability terms. Journal of Experimental Psychology 115, pp. 348-365, 1986.

[35] Ward Jr., J.H., Hierarchical grouping to optimize an objective function. Journal of the American Statistical Association 58, pp. 236-244, 1963.

[36] Zadeh, L.A., Fuzzy logic = computing with words. IEEE Transactions on Fuzzy Systems 4, pp. 103-111, 1996.

[37] Zadeh, L.A., Computing with Words. Springer, Berlin, 2012.

[38] Zimmer, A.C., Verbal vs. numerical processing of subjective probabilities. In: R.W. Scholz (Ed.), Decision Making under Uncertainty. North-Holland, Amsterdam, pp. 159-182, 1983.