A Novel Linguistic Cohesion Measure for Weighting Experts' Subgroups in Large-Scale Group Decision Making Methods

Álvaro Labella
Department of Computer Science
University of Jaén
Jaén, Spain
alabella@ujaen.es

Rosa M. Rodríguez
Department of Computer Science
University of Jaén
Jaén, Spain
rmrodrig@ujaen.es

Luis Martínez
Department of Computer Science
University of Jaén
Jaén, Spain
martin@ujaen.es

***Abstract*—Today, decision making problems are continually evolving due to the new needs of society, caused mainly by continuous technological advances. Many times, in order to face these new decision problems, it is no longer enough with the participation of only a few experts, but hundreds or thousands are necessary. The engagement of many experts implies, in turn, the appearance of new challenges such as the management of greater uncertainty, scalability, opinions' polarization etc. This contribution is focused on group decision making problems in a large-scale context in which uncertainty is modeled by linguistic information and how to deal with the scalability problem under these conditions through clustering methods. Clustering methods are used to manage the scalability problem by grouping the initial large group of experts into smaller subgroups according to the similarity of their opinions and assign different weights to such subgroups. The weights assignment is a key issue due to its influence in the final solution of the problem and classically, it has been carried out by taking into account exclusively the size of the subgroups, by ignoring other features such as the cohesion of the opinions in the subgroups, which may provoke a misassignment of the importance in experts' subgroups and thus, unfair solutions. Therefore, this paper introduces a new way to calculate and assign properly the relevance of experts' subgroups in clustering methods by taking into account both size and cohesion of such subgroup under uncertainty conditions in which experts use linguistic assessments.**

***Index Terms*—large-scale group decision making, hesitant fuzzy linguistic term set, fuzzy clustering, linguistic cohesion measure**

I. Introduction

Human beings are continually making decisions in their daily lives, choosing which clothes to wear, what to eat or which political party to vote for. In some problems the participation of a single person is enough to make the decision but, usually, due to the increasing complexity of decision problems and the need of considering several opinions, it is required the participation of several people in the decision process leading to *Group Decision Making* (GDM). In a GDM problem, a group of people with different knowledge and points of view work side by side in order to select one of several alternatives as solution of the decision problem [7].

978-1-7281-2348-6/19/$31.00 ©2019 IEEE

Nowadays, society undoubtedly has to face new and more complex decision making problems that have appeared mainly due to the dizzying advance of technology [8], [20], [21]. Finding a solution for this type of problems is not a simple task and requires the participation of a large number of experts, leading to *Large-Scale Group Decision Making* (LSGDM). So far, GDM problems have assumed a small number of decision makers, but in LSGDM problems may be from 20 to hundreds or thousands [3]. LSGDM implies new challenges [15], for instance, scalability, minority opinions, non-cooperative behaviors, stronger disagreement positions and so on.

This contribution is focused on the *scalability* problem, which makes reference to the difficulty to manage a huge amount of information at the same time. Specifically, for LSGDM problems, such amount of information is represented by the preferences of hundreds or thousands of experts. To reduce the scalability problem in LSGDM problems, clustering methods have been used successfully [9], [23]. Clustering methods classify the initial group of decision makers into smaller subgroups composed by those decision makers whose opinions are similar to each other. To obtain fair solutions in LSGDM problems, it is common to weight the experts' subgroups and, in this way, control the influence of each subgroup in the final decision. Classically, the relevance of the experts' subgroups has been exclusively based on the number of experts who compose the subgroups, in such a way the greatest number of experts are the most influential group. However, although experts belong to same subgroup, disagreements might appear to some extent by provoking a wrong assignment of the subgroups' weights and lastly, an unjust solution of the problem.

Additionally, in LSGDM the uncertainty and vagueness of the information inherent in the problems is often modeled by means of linguistic approaches [12], [13], [24]. Nevertheless, linguistic information is usually modeled by single linguistic terms, which may not enough due to experts may hesitate among several terms [19]. To manage these situations, Rodríguez et al. proposed the use of *Complex Linguistic Expressions* (CLEs) based on *Hesitant Fuzzy Linguistic Terms*

Set (HFLTS) which allow to model experts' hesitations [17], [18] and are similar to the expressions that experts use in real world decision problems.

In [16] was introduced a first attempt to deal with cohesion in LSGDM with fuzzy preference relations. However, it seems logical to extend this effort for linguistic LSGDM in which CLEs should be considered. Because, so far, there is no any proposal beyond the subgroup's size to improve the weights assignment to them.

Therefore, our proposal aims at defining a new way to weight subgroups in linguistic LSGDM that not only considers the size of the subgroups but also the togetherness of subgroups members' opinions, i.e, their cohesion. To achieve this aim, a new linguistic cohesion measure for assessing the togetherness in experts' subgroups is proposed that will be integrated in the linguistic LSGDM resolution process.

This paper is composed by different sections: Section II revises preliminary concepts in LSGDM, CLEs based on HFLTS and fuzzy clustering to understand easily the proposal. Section III presents a clustering method for linguistic LSGDM problems. Section IV introduces a new process to compute the cohesion in clusters of experts for LSGDM dealing with CLEs. Afterwards, Section V shows a case study to bring to light the utility of the proposal. To conclude, Section VI points out some findings and coming studies.

II. Preliminaries

In this section, contents about LSGDM, CLEs and fuzzy clustering are briefly.

A. Large-scale Group Decision Making

In a GDM process, several experts, who present different behaviors and opinions, attempt to achieve a common solution by choosing one or several alternatives among a group of them [11].

Formally, a GDM problem is characterized by a finite set of decision makers, $E = \{dm_1, \ldots, dm_m\}$ who evaluate a finite set of alternatives or solutions, $A = \{a_1, \ldots a_n\}$ [7]. Two main steps compose the classical resolution scheme for GDM problems (see Fig.1):

- *Aggregation*: the decision makers' opinions are aggregated by means of an aggregation operator for obtaining a global opinion for each alternative.
- *Exploitation*: the collective opinion for each alternative are ranked, by obtaining the solution of the problem.

In GDM problems is common the apparition of uncertainty and vagueness provoked by the complexity of the problems and the changing contexts in which they take place. Such uncertainty cannot be modeled in a quantitative way thus, classical decision theory models cannot be applied. On the other hand, linguistic information has been applied successfully to model such uncertainty. The use of linguistic preferences to evaluate the alternatives gives rise to *Linguistic Decision Making* (LDM). To solve this type of problems, a similar scheme to the one represented in Fig. 1 is used but with two additional phases [6] (see Fig. 2):

- *Define syntax and semantics*: first, a linguistic expression domain to model the experts' opinions is defined.
- *Select a linguistic aggregation operator*: second, it is essential to choose a proper linguistic aggregation operator to obtain the collective opinion from the decision makers' assessments.

On the other side, GDM problems is constantly envolved. Nowadays, the participation of a high number of decision makers to solve the increasingly complex GDM problems is becoming more and more necessary, which implies to manage a bigger amount of information. For this reason, LSGDM has attracted the attention of many scholars [4], [15], [22]. This concept is quite similar to GDM, but with two significant differences, the number of experts is much more bigger than in the latter and, in turn, the number of experts is much more bigger than the alternatives.

Formally, a LSGDM problem consists of a finite set of decision makers, $E = \{dm_1, \ldots, dm_m\}$ who express their opinions over a finite set of alternatives $A = \{a_1, \ldots, a_n\}$, $(n \geq 2)$ or possible solutions for the problem, being $(m >> n)$. The resolution of LSGDM problems can be carried out by following similar schemes to the ones represented in Figs. 1 and 2, these problems introduce novel challenges [15], focusing this contribution on the scalability problem related to the management of a great amount of information.

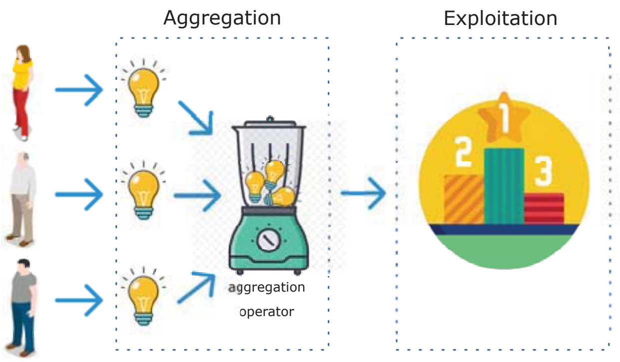

Fig. 1. GDM resolution scheme.

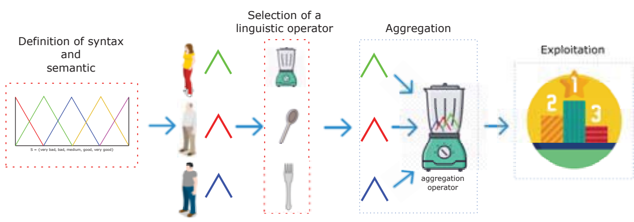

Fig. 2. LDM resolution scheme.

B. Use of Complex Linguistic Expressions

Uncertainty in LSGDM problems has been modeled successfully by linguistic information but most of the linguistic proposals only consider that decision makers assess the alternatives by using single just one linguistic term, insufficient to reflect their opinions when they consider that several linguistic terms might be used to evaluate the different alternatives. Taking into account this premise, Rodríguez et al. [17] proposed the use of richer linguistic expressions generated by a context-free grammar (see [18] for further details). Some examples of CLEs may be *between good and bad*, *at most poor* or *at least medium*.

These CLEs are easy to understand and can represent experts' opinions in a comprehensive way. Furthermore, with the aim of facilitating the experts' assessments elicitation by using such expressions, they are usually modeled by means of a *Hesitant Fuzzy Linguistic Preference Relation* (HFLPR) [25], that is a matrix $A \times A \rightarrow S_{ll}$, where S_{ll} is the set of CLEs generated by the linguistic terms belonging to the linguistic term set S. Let $S = \{\textit{very expensive}, \textit{expensive}, \textit{fair}, \textit{cheap}, \textit{very cheap}\}$, be a linguistic term set, an example of HFLPR provided by the expert dm_i could be:

$$P^i = \begin{pmatrix} - & \textit{fair} & \textit{cheap} \\ \textit{fair} & - & \textit{at most expensive} \\ \textit{expensive} & \textit{at least cheap} & - \end{pmatrix} \quad (1)$$

The linguistic computations with CLEs are performed by using a transformation function was defined. This function transforms the CLEs into HFLTS that is a new representation of linguistic information.

Definition 1: [17] Let $S = \{s_0, \ldots, s_g\}$ be a linguistic term set, a HFLTS, H_S, is an ordered finite subset of consecutive linguistic terms of S.

$$H_S = \{s_i, s_{i+1}, \ldots, s_j\}, s_k \in S, k \in \{i, \ldots, j\}$$

Definition 2: [18] Let E_{CG_H} be a function that transforms the CLEs, $ll \in S_{ll}$, obtained by the context-free grammar CG_H, into HFLTSs, H_S. Being S the linguistic term set used by CG_H and S_{ll} the expression domain generated by CG_H.

$$E_{CG_H} : S_{ll} \rightarrow H_S$$

The transformations of the CLEs generated by the context-free grammar CG_H are the following ones:

$E_{CG_H}(s_i) = \{s_i | s_i \in S\}$
$E_{CG_H}(\textit{at most } s_i) = \{s_j | s_j \leq s_i \textit{ and } s_j \in S\}$
$E_{CG_H}(\textit{at least } s_i) = \{s_j | s_j \geq s_i \textit{ and } s_j \in S\}$
$E_{CG_H}(\textit{between } s_i \textit{ and } s_j) = \{s_k | s_i \leq s_k \leq s_j \textit{and } s_k \in S\}$

Different computational models have been introduced to deal with HFLTS [10], [18]. In this proposal we will use the fuzzy envelope, a fuzzy representation of the CLEs that models the uncertainty and vagueness expressed by them. This fuzzy representation allows to carry out the computations by fuzzy sets and obtain more precise results than using intervals [18].

Definition 3: [10] The *fuzzy envelope*, $env_F(H_S)$, is defined as a trapezoidal fuzzy membership function as follows:

$$env_F(H_S) = \tilde{T}(a_1, a_2, a_3, a_4)$$

where H_S is a HFLTS and $\tilde{T}(a_1, a_2, a_3, a_4)$ is a fuzzy trapezoidal membership function (see [10] for further detail).

C. Fuzzy C-means

One of the biggest drawbacks in LSGDM problems is scalability problem. To tackle this problematic, clustering techniques aim at dividing a big group of data objects into different subgroups or clusters and manage them independently, by reducing the problem of scalability. There are several clustering algorithms proposed in the literature, in this contribution we will use the fuzzy c-means algorithm [2], because it has provided very good results in different problems and it is quite simple. In a brief summary, this algorithm computes iteratively centroids and classify each data object to a specific cluster through the computation of distances among it and the centroid (see [2] for further details). One of the most important step in fuzzy c-means is to initialize the clusters and centroids. For this contribution, the number of clusters, N, is fixed as the number of alternatives, $N = n$, and their respective centroids are initialized with HFLPRs (see (12)) in which each alternative predominates over the others. In this way, the formed subgroups are composed by experts with similar opinions respect to the alternatives. For each iteration, t, the centroids C^k are computed together with the membership degree of each expert' preference P^i to each centroid C^k. As in this contribution, experts elicit their preferences by using HFLPRs, the algorithm has been adapted as follows:

Algorithm 1 Fuzzy c-means for HFLPRs

Inputs: decision makers provide their opinions by HFLPRs, P^i

Start :

Fix N, $(N \geq 2)$ and b.

Compute centroids $C^k, k \in \{1, \ldots, n\}$.

while condition **do**

Compute the membership degree of each decision maker's preference P^i, $\mu_{C^k}(P^i)$ to each cluster represented by its centroid, C^k

$$\mu_{C^{k,t}}(P^i) = \frac{(1/d_H(P^i, c^{k,t}))^{2/(b-1)}}{\sum_{u=1}^{n}(1/d_H(P^i, c^{u,t}))^{2/(b-1)}} \quad (2)$$

Update cluster centers, C^k

$$C^k_{lj} = \frac{\sum_{l=1}^{n}\sum_{j=1}^{n}\sum_{i=1}^{m} env(E_{CG_H}(P^i_{lj}))}{\#C^k} \quad (3)$$

end while

where $d_H(\cdot)$ is a distance measure between two HFLPRs [25], t is the current iteration, $\#C^k$ the number of experts in the cluster C^k and b the fuzziness degree [2].

Remark 1: Note that the elements that compose the centroids are represented by trapezoidal fuzzy numbers, $T^k_{lj}(a, b, c, d)$, obtained from the aggregation of the fuzzy envelopes of the assessments of all the experts that compose the cluster C^k for each pair of alternatives (a_l, a_j).

III. A Large-scale Method based on Fuzzy C-means with a Cohesion Measure for Complex Linguistic Expressions

This contribution presents a linguistic cohesion measure for computing the weights of experts' clusters when clustering techniques are applied and in which experts express their preferences by using CLEs modeled by HFLPRs taking into consideration both the size and cohesion of the subgroups. Therefore, this section introduces a large-scale method based on CLEs and fuzzy clustering that includes the proposed linguistic cohesion measure introduced in the following section.

The method consists of several steps described in further detail below and represented in Fig. 3:

1) *Inputs*: decision makers express their assessments by means of CLEs that are modeled by HFLPRs (see the example introduced in Section II-B). Afterwards, the fuzzy envelope of each CLE is obtained.
2) *Clustering process*: all the experts are divided into several subgroups, $G = \{G_1, \dots G_N\}$, by using the clustering algorithm introduced in Section II-C.
3) *Weighting subgroups*: the relevance of each subgroup, $W = \{w_1, \dots w_N\}$, is computed by taking into account both its cohesion and size (see Section IV).
4) *Aggregation*: to represent the global opinion of all the decision makers in the decision problem, the centroids C^k of each subgroup G^k are aggregated by obtaining a matrix. Several aggregation operators could be applied but, without losing of generality, in this contribution we have used, a weighted average operator, although any other could be applied.

   Definition 4: Let $C^k = (c^k_{lj})_{n\times n}$ be the centroids of the clusters G^k and $W = \{w_1, \dots w_N\}$ the weights assigned to each subgroup, the collective matrix $P^c = (p^c_{lj})_{n\times n}$ is obtained as follows:

$$C^k_{lj} = \sum_{l=1}^{n} \sum_{j=2}^{n-1} \sum_{k=1}^{N} c^k_{lj} w_k \quad (4)$$

5) *Exploitation*: finally the alternatives are ranked by computing the non-dominance degree [14] among them in P^c.

Fig. 3. LSGDM clustering method scheme.

IV. Linguistic Cohesion Measure

As it was aforementioned, once a clustering process is applied in the linguistic LSGDM and the experts' subgroups are obtained, the next step is to compute the importance of each subgroup. Classically, such importance has been determined just by the clusters size, the greater the number of decision makers in a cluster, the greater its importance. However, the fact that a large number of experts belong to the same subgroup does not imply implicitly that their opinions match each other. Therefore, from our view, a cluster composed by a large numbers of experts but whose opinions are not cohesive should not be considered so relevant as a cluster with a similar size and greater cohesion. For this reason, our proposal consists of a novel linguistic cohesion measure able to determine the degree of closeness in the decision makers' opinions. Therefore, to calculate more appropriately the importance in experts' subgroups, we take into consideration both size and cohesion. In this way, better solutions for LSGDM problems will be obtained.

For sake of simplicity, let us suppose a cluster composed by two experts, $G^k = \{dm_1, dm_2\}$ who take part of a LSGDM with three alternatives $A = \{a_1, a_2, a_3\}$. The fuzzy linguistic term set used to model the assessments is S = {*No Importance (NI), Very Unimportant (VU), Unimportant (U), Fair (F), Important (I), Very Important (VI), Essential (E)*} and modeled by the following HFLPRs:

$$P^1 = \begin{pmatrix} - & I & VI \\ U & - & At\ most\ U \\ VU & At\ least\ I & - \end{pmatrix} \quad (5)$$

$$P^2 = \begin{pmatrix} - & VI & VI \\ VU & - & U \\ VU & I & - \end{pmatrix} \quad (6)$$

Then, the experts' linguistic assessments are transformed into HFLTS:

$$P^1 = \begin{pmatrix} - & \{I\} & \{VI\} \\ \{L\} & - & \{NI, VU, U\} \\ \{VU\} & \{I, VI, E\} & - \end{pmatrix} \quad (7)$$

$$P^2 = \begin{pmatrix} - & \{VI\} & \{VI\} \\ \{VU\} & - & \{U\} \\ \{VU\} & \{I\} & - \end{pmatrix} \quad (8)$$

Finally, the fuzzy envelope of each HFLTS [10] is computed:

$$P^1 = \begin{pmatrix} - & T^1_{12}(0.5, 0.667, 0.833) & T^1_{13}(0.667, 0.833, 1) \\ T^1_{21}(0.167, 0.333, 0.5) & - & T^1_{23}(0, 0, 0.15, 0.5) \\ T^1_{31}(0, 0.167, 0.333) & T^1_{32}(0.5, 0.86, 1, 1) & - \end{pmatrix} \quad (9)$$

$$P^2 = \begin{pmatrix} - & T^2_{12}(0.667, 0.833, 1) & T^2_{13}(0.667, 0.833, 1) \\ T^2_{21}(0, 0.167, 0.333) & - & T^2_{23}(0.167, 0.333, 0.5) \\ T^2_{31}(0, 0.167, 0.333) & T^2_{32}(0.5, 0.667, 0.833) & - \end{pmatrix} \quad (10)$$

Afterwards, a matrix Q^k, in which each element $[q^-_{lj}, q^+_{lj}]$ is an interval that represents the smallest fuzzy envelope q^-_{lj} and the largest fuzzy envelope q^+_{lj} of the experts' preferences belonging to G^k for each pair of alternatives (a_l, a_j) is obtained.

Remark 2: To compare the fuzzy envelopes, the concept of *magnitude* of a fuzzy number [1] is used.

$$Mag(\tilde{T}(a_1, a_2, a_3, a_4)) = \frac{1}{2}\left(a_2 + a_3 - \frac{a_1 - a_2}{6} + \frac{a_3 - a_4}{6}\right) \tag{11}$$

So that:

A) $Mag(T_1) > Mag(T_2) \iff T_1 \succ T_2$
B) $Mag(T_1) < Mag(T_2) \iff T_1 \prec T_2$
C) $Mag(T_1) = Mag(T_2) \iff T_1 \sim T_2$

where T_1 and T_2 are two fuzzy numbers.

The resulting Q^k for our example is shown as follows:

$$\tilde{Q}^k = \begin{pmatrix} - & [T^1_{12}, T^2_{12}] & [T^1_{13}, T^2_{13}] \\ [T^1_{21}, T^2_{21}] & - & [T^1_{23}, T^2_{23}] \\ [T^1_{31}, T^2_{31}] & [T^1_{32}, T^2_{32}] & - \end{pmatrix} \tag{12}$$

The pairs of fuzzy envelopes in Q^k on (a_l, a_j) are graphically shown in Fig. 4 according to the geometrical fuzzy distance [5] among each pair. The items represented in the figure are:

- *X-axis*: this axis is formed by the set of all the pair of alternatives over A, noted as Z, where $z_t = (a_l, a_j), l, j \in (1, 2, 3),\ l \neq j$.
- q^-_{lj}: the fuzzy envelope with less magnitude for the pair of alternatives (a_l, a_j) in $\tilde{Q}^k$.
- q^+_{lj}: the fuzzy envelope with greater magnitude for the pair of alternatives (a_l, a_j) in $\tilde{Q}^k$.
- T^k: the area enclosed by the points a^T, b^T, c^T and d^T, where $T^k = g^T \times n^T$ where g^T is the maximal distance between two fuzzy envelopes, which is 1, and n^T is equal to the number of pair of alternatives.
- A^k: the shaded area. This are represents the cohesion of G^k, the larger the area, the less cohesion.

Remark 3: Notice that the pairs z_t are located by the minimum assessments q^-_{lj} in increasing order.

To compute the cohesion for G^k is necessary to compute the values q^-_{lj} and q^+_{lj}.

$$q^-_{lj} = min\left\{q^1_{lj}, q^2_{lj}, \dots q^s_{lj}\right\}, \forall\ (l, j) \in I \tag{13}$$

where I is the n^T pairs over the set of alternatives $A = \{a_1, \dots, a_n\}$.

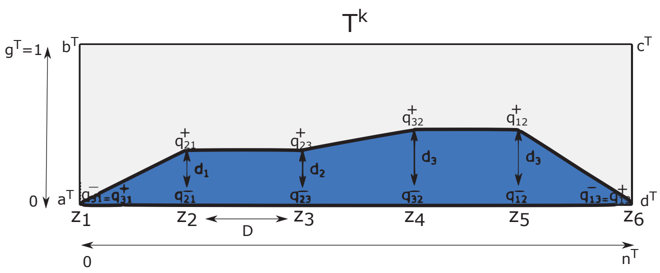

Fig. 4. Graphical visualization for cohesion computation.

$$q^+_{lj} = max\left\{q^1_{lj}, q^2_{lj}, \dots q^s_{lj}\right\}, \forall\ (l, j) \in I \tag{14}$$

and,

$$q^-_{ab} = min_{l,j \in I}\left\{q^-_{lj}\right\}, (a, b) \in I \tag{15}$$

$$q^+_{cd} = max_{l,j \in I}\left\{q^-_{lj}\right\}, (c, d) \in I \tag{16}$$

Therefore, A^k is obtained as:

$$A^k = \left[\sum_{l,j \in I}\left(d_T(q^+_{lj}, q^-_{lj})\right) - \frac{d_T(q^+_{ab}, q^-_{ab}) + d_T(q^+_{cd}, q^-_{cd})}{2}\right] \cdot D \tag{17}$$

where $d_T(\cdot, \cdot)$ represents the geometric distance for fuzzy numbers introduced in [5] as follows:

$$d(T^1, T^2) = \begin{cases} \frac{1}{4}(|a_1 - a_2|^\lambda + |b_1 - b_2|^\lambda + |c_1 - c_2|^\lambda + \\ |d_1 - d_2|^\lambda)^{\frac{1}{\lambda}}, if\ 1 \leq \lambda \leq \infty \\ max(|a_1 - a_2|, |b_1 - b_2|, |c_1 - c_2|, |d_1 - d_2|), \\ if\ \lambda = \infty \end{cases} \tag{18}$$

where T^1 and T^2 are two parametric fuzzy numbers.

To compute the cohesion for the cluster G^k:

$$cohesion(G^k) = 1 - \frac{A^k}{T^k} \in [0, 1] \tag{19}$$

Finally, to obtain the weight for each subgroup G^k:

Definition 5: Let $Y_{G^k} = \{c, s\}$ be cohesion and size of the cluster G^k, the cluster's weight is obtained as follows

$$\delta(Y_{G^k}) = (1 + s)^{c\gamma} \tag{20}$$

where $\gamma > 0$ allows to measure the influence of c in the resulting weights.

Finally, the values $\delta(Y_{G^k})$ are normalized:

$$w_k = \frac{\delta(Y_{G^k})}{\sum_{z=1}^{n} \delta(Y_{G^z})} \tag{21}$$

V. Case Study

Once the new measure to obtain the weights for experts' subgroups has been introduced, this section evaluates its performance through a case study. First, a LSGDM problem is defined and by means of the LSGDM model introduced in Section III is solved. Such a model will be applied with different values of the parameter γ, to see how the cohesion affects to the solution of the problem. Finally, we study and analyse the results obtained.

Let $E = \{dm_1, \dots dm_{50}\}$ be a group of 50 business people, who decide to build a sustainable hospital in China. Three cities are candidates for the building of the hospital, $A = \{a_1 : Shanghai, a_2 : Pekin, a_3 : Canton\}$.

Following the different steps introduced in Section III.

1) *Inputs*: decision makers assess the alternatives by CLEs which are modeled by HFLPRs. Afterwards, the fuzzy envelope of each CLE is computed.

2) *Clustering process*: the fuzzy c-means algorithm is applied to group the decision makers into clusters with similar preferences. Table I shows the resulting experts' clusters, $\{G_1, G_2, G_3\}$, one for each alternative.
3) *Weighting subgroups*: the importance of the experts' clusters are calculated by Eqs. 20 and 21. Additionally, such a weight has been also computed by ignoring the cohesion to compare and analyse the results. Table II shows the results obtained for the cohesion and size and Table III display the weights for each group taking into account different values of γ.
4) *Aggregation*: the aggregation of the the centroids, $\{C^1, C^2, C^3\}$, is carried out by using Eg 4.
5) *Exploitation*: the alternatives are sorted using the collective opinion P^c for the different values of γ (see Table IV).
According to Table IV, different values of $gamma$ provide different solutions for the decision making problem. When cohesion is not taken into account in the weights computations, the most important experts' subgroups are those that have a greater number of experts, in this case G_1 and G_3 (see Table III), and their weights are the same, 0.36. In addition, the ranking of alternatives when the cohesion is not applied is $a_1 \succ a_2 \succ a_3$ being a_1 the best solution (see Table IV). Nevertheless, when the cohesion is considered to compute the weights, the subgroups G_1 and G_3 have a different weight (see Table III), although their size is the same, being the biggest weight assigned to the subgroup G_3. Regarding the ranking of the cities obtained, in both experiments $gamma = 1.5$ and $\gamma = 3.0$, the ranking is the same, although the variation of the value γ might cause different solutions. Therefore, it depends on the value for γ, the effect of the cohesion can increase or decrease the computation of the subgroups' weights. In this case, both experiments present a different solution to the one provided without considering the cohesion. The ranking of the cities is $a_2 \succ a_1 \succ a_3$ being the best solution, a_2. Therefore, the inclusion of cohesion can provoke different solutions, consequently, it is a measure that should be studied and analyzed.

TABLE I
EXPERTS' CLUSTERS

Subgroup	Experts
G_1	$dm_1, dm_3, dm_5, dm_6, dm_7, dm_{13}, dm_{18}, dm_{19}, dm_{20}, dm_{23}, dm_{33}, dm_{35}, dm_{38}, dm_{42}, dm_{45}, dm_{46}$ dm_{47}, dm_{50}
G_2	$dm_2, dm_9, dm_{10}, dm_{12}, dm_{16}, dm_{17}, dm_{25}, dm_{27}, dm_{32}, dm_{34}, dm_{37}, dm_{43}, dm_{44}, dm_{49}$
G_3	$dm_4, dm_8, dm_{11}, dm_{14}, dm_{15}, dm_{21}, dm_{22}, dm_{24}$ $dm_{26}, dm_{28}, dm_{29}, dm_{30}, dm_{31}, dm_{36}, dm_{39}, dm_{40}, dm_{41}, dm_{48}$

TABLE II
EXPERTS' CLUSTERS FEATURES

Cluster	Size	Cohesion
G_1	18	0.3367
G_2	14	0.3398
G_3	18	0.3565

VI. CONCLUSIONS

One of the pivotal key when clustering algorithms are used for LSGDM problems is the computation of the weights of the experts' subgroups. Classically, such weights have been computed only considering the size of the subgroups and ignoring other features as the cohesion degree in the experts' opinions, which might lead to unfair solutions. Furthermore, LSGDM problems usually present uncertainty related to their complexity, which have been modeled in a successful way by means of linguistic information. However, such complexity can provoke that experts hesitate when they have to elicit their opinions by using single linguistic terms. Under these conditions, CLEs are useful to model the experts' hesitancy and similar to the expressions that experts use in decision processes. This work has presented a linguistic cohesion measure to obtain the relevance of decision makers' clusters in linguistic LSGDM by considering not only their size, but also their cohesion. Additionally, a novel fuzzy clustering LSGDM method has been introduced to evaluate the performance of the proposed measure.

The performance of the proposed linguistic cohesion measure in a consensus reaching process for LSGDM problems may be considered for future research.

ACKNOWLEDGMENTS

This work is partially supported by the Spanish Ministry of Economy and Competitiveness through the Spanish National Research Projects TIN2015-66524-P and PGC2018-099402-B-I00 and the Postdoctoral fellow Ramón y Cajal (RYC-2017-21978).

REFERENCES

[1] S. Abbasbandy and T. Hajjari. A new approach for ranking of trapezoidal fuzzy numbers. *Computers & Mathematics with Applications*, 57(3):413–419, 2009.

[2] J. Bezdek. *Pattern Recognition with Fuzzy Objective Function Algorithms.* Springer, 1981.

[3] X.H. Chen and Rong Liu. Improved clustering algorithm and its application in complex huge group decision-making. *Systems Engineering and Electronics*, 28(11):1695–1699, 2006.

[4] X. Gou, Z.S. Xu, and F. Herrera. Consensus reaching process for large-scale group decision making with double hierarchy hesitant fuzzy linguistic preference relations. *Knowledge-based Systems*, 157:20–33, OCT 1 2018.

[5] S. Heilpern. Representation and application of fuzzy numbers. *Fuzzy sets and Systems*, 91(2):259–268, 1997.

[6] F. Herrera and E. Herrera-Viedma. Linguistic decision analysis: Steps for solving decision problems under linguistic information. *Fuzzy Sets and Systems*, 115(1):67–82, 2000.

[7] J. Kacprzyk. Group decision making with a fuzzy linguistic majority. *Fuzzy Sets and Systems*, 18(2):105–118, 1986.

TABLE III
EXPERTS' CLUSTERS WEIGHTS

γ	**Cluster**	**Weight**
No cohesion	G_1	0.36
	G_2	0.28
	G_3	0.36
$\gamma = 1.5$	G_1	0.33
	G_2	0.30
	G_3	0.37
$\gamma = 3.0$	G_1	0.33
	G_2	0.27
	G_3	0.40

TABLE IV
RANKING OF THE CITIES

City	$\gamma = 0$	$\gamma = 1.5$	$\gamma = 3.0$
a_1	1	2	2
a_2	2	1	1
a_3	3	3	3

[8] J. Kim. A model and case for supporting participatory public decision making in e-democracy. *Group Decision and Negotiation*, 17(3):179–192, 2008.

[9] B. Liu, Y. Shen, X. Chen, Y. Chen, and X. Wang. A partial binary tree DEA-DA cyclic classification model for decision makers in complex multi-attribute large-group interval-valued intuitionistic fuzzy decision-making problems. *Information Fusion*, 18:119 – 130, 2014.

[10] H. Liu and R. M. Rodríguez. A fuzzy envelope for hesitant fuzzy linguistic term set and its application to multicriteria decision making. *Information Sciences*, 258:220–238, 2014.

[11] J. Lu and D. Ruan. *Multi-objective group decision making: methods, software and applications with fuzzy set techniques*, volume 6. Imperial College Press, 2007.

[12] L. Martínez. Sensory evaluation based on linguistic decision analysis. *International Journal of Aproximated Reasoning*, 44(2):148–164, 2007.

[13] L. Martínez, D. Ruan, and F. Herrera. Computing with words in decision support systems: an overview on models and applications. *International Journal of Computational Intelligence Systems*, 3(4):382–395, 2010.

[14] S.A. Orlovsky. Decision-making with a fuzzy preference relation. *Fuzzy Sets and Systems*, 1(3):155–167, 1978.

[15] I. Palomares, L. Martínez, and F. Herrera. A consensus model to detect and manage noncooperative behaviors in large-scale group decision making. *IEEE Transactions on Fuzzy Systems*, 22(3):516–530, 2014.

[16] R. M. Rodríguez, A. Labella, G. De Tré, and L. Martínez. A large scale consensus reaching process managing group hesitation. *Knowledge-Based Systems*, 2018.

[17] R. M. Rodríguez, L. Martínez, and F. Herrera. Hesitant fuzzy linguistic term sets for decision making. *IEEE Transactions on Fuzzy Systems*, 20(1):1109–119, 2012.

[18] R. M. Rodríguez, L. Martínez, and F. Herrera. A group decision making model dealing with comparative linguistic expressions based on hesitant fuzzy linguistic term sets. *Information Sciences*, 241(1):28–42, 2013.

[19] R.M. Rodríguez, A. Labella, and L. Martínez. An overview on fuzzy modelling of complex linguistic preferences in decision making. *International Journal of Computational Intelligence Systems*, 9:81–94, 2016.

[20] C. Sueur, J.L. Deneubourg, and O. Petit. From social network (centralized vs. decentralized) to collective decision-making (unshared vs. shared consensus). *PLoS one*, 7(2):e32566, 2012.

[21] S.J. Walker. Big data: A revolution that will transform how we live, work, and think. *International Journal of Advertising*, 33(1):181–183, 2014.

[22] Z. Wu and J. Xu. A consensus model for large-scale group decision making with hesitant fuzzy information and changeable clusters. *Information Fusion*, 41:217 – 231, 2018.

[23] S. Zahir. Clusters in a group: Decision making in the vector space formulation of the analytic hierarchy process. *European Journal of Operational Research*, 112(3):620 – 634, 1999.

[24] X. Zhang. A novel probabilistic linguistic approach for large-scale group decision making with incomplete weight information. *International Journal of Fuzzy Systems*, 20(7):2245–2256, 2018.

[25] B. Zhu and Z. Xu. Consistency measures for hesitant fuzzy linguistic preference relations. *IEEE Transactions on Fuzzy Systems*, 22(1):35–45, Feb 2014.