

Evaluating Relevant Opinions within a Large Group

Ana Tapia-Rosero^{1,2} and Guy De Tré²

¹*FIEC, Escuela Superior Politécnica del Litoral, Campus Gustavo Galindo V., Guayaquil, Ecuador*

²*Department of Telecommunications and Information Processing, Ghent University,
St-Pietersnieuwstraat 41, B-9000 Ghent, Belgium*

Keywords: LSP, Evaluation, Relevant Opinions, Shape-similarity Method, Large-scale Group Decision-making.

Abstract: We propose to identify which opinions are relevant, from the decision-maker's point of view, within a large group of opinions that could be collected using social media. Our approach considers that each participating person expresses his/her preferences over a criterion specification as a matter of degree. First, using a shape-similarity method, we split a large group of opinions, where each opinion is represented through a membership function, into clusters —here, a cluster depicts a group of similar opinions over the criterion. Then, in order to evaluate the relevance of each cluster, we differentiate them based on some characteristics like the cohesion, the number of membership functions and the number of noticeable opinions. Within this paper, the cohesion of the cluster is a measure that takes into account the level of togetherness among its contained membership functions; and the representativeness of the cluster is obtained by combining the number of membership functions and the number of noticeable represented opinions (i.e., considered as more important or worthy of notice among other opinions). Moreover, relevant clusters result in the evaluation of combining their cohesion measure and their representativeness according to the decision-maker's point of view. Finally, as a part of the evaluation, this proposal includes the steps describing the process through an illustrative example.

1 INTRODUCTION

Nowadays, the use of social media makes it possible to involve a large group of people to express their opinions on criteria, e.g., opinion with respect to a feature like weight, length, or usefulness of a product. Thus, opinions from different points of view might be gathered and used in a decision-making context. Within this paper, each person that belongs to the aforementioned large group will be considered as an expert. But, how do we identify and evaluate relevant opinions in a large group that includes different points of view and some opinions are more representative than others? Here, our aim is to evaluate relevance through the wisdom of the crowd while avoiding to be overwhelmed with a huge amount of opinions.

Let us consider that a company wants to know the “usefulness level” (criterion) of a new feature in a product (e.g., a pressure sensor for an electric toothbrush, an augmented reality for a smart phone, a pedestrian detection in a car, a heart rate monitor in a cellphone, a pedometer in a waistband, among others) while the product is under design. Here, it is possible to gather this information using social media (e.g., a fan page) where opinions are given by different levels

of knowledge (students, non-experts and professionals), areas of expertise (engineering, medicine, journalism, among others) and personal profiles (single, married, parents, etc.). Hence, it is desired to differentiate noticeable opinions considering their importance from the decision-maker's point of view (e.g., the opinion of some specific professionals might be more important than the opinion of some regular users).

Using soft computing techniques, each expert will express his/her preferences with respect to a specific criterion (i.e., level of usefulness) through a membership function. In this way, experts using expressions like “the usefulness level is above 65%”, “it is below 50%” or “it is around 30%” could represent what he/she understands to be the level of usefulness through membership functions (Eshragh and Mamdani, 1979; Pedrycz, 2013). Within this paper, it is not required that each expert has preknowledge on soft computing techniques to represent his/her preferences $P(x)$ as a matter of degree, i.e., $0 \leq P(x) \leq 1$ where 0 denotes a complete disagreement on a criterion and 1 denotes the highest level of agreement, as long as they provide some values (Dujmović and De Tré, 2011). These values will be used to define the attribute criterion in a membership function.

Once all the membership functions have been gathered, we are able to group them using a shape-similarity method (Tapia-Rosero et al., 2014). The shape-similarity method obtains clusters of similar opinions, represented by symbolic notations, facilitating their further processing. Each cluster allows us to obtain the closest approximation to represent a group of expert opinions by means of its upper and lower bounds. These boundaries allow us to compute a *cohesion measure* among the contained membership functions, where a higher value denotes more togetherness and hence expresses a group of more confident opinions. The main advantage of the proposed cohesion measure is that it is possible to obtain a cluster where some of the contained membership functions do not overlap but are close enough to be considered similar (Tapia-Rosero and De Tré, 2013). Notwithstanding, any cluster with a single membership function will obtain the highest cohesion value. Therefore, we consider that besides the *cohesion measure*, additional attributes describe a relevant cluster. As well as one bright bulb could light up a room as good as a higher number of less brighter bulbs; we consider that the opinion of one expert might highlight among others. Based on this analogy, the *representativeness* of the cluster is obtained by combining the number of membership functions and the number of noticeable opinions.

Since any cluster might be categorized as *relevant* based on a combination of the aforementioned characteristics, in this paper we use the *logic scoring of preference (LSP) aggregation* (Dujmović, 2007; Dujmović et al., 2010) to obtain the overall evaluation value for each cluster. The LSP aggregation is based on the verbalized approach of the Generalized Conjunction/Disjunction (GCD) (Dujmović, 2012) and allows us to easily reflect aspects of human decision-making, i.e. relative importance given by weights and a combination of andness and orness. The overall evaluation values are used in the selection of the relevant clusters and it is made by the decision-maker. Here, it is possible to select one cluster with the best evaluation, i.e., the highest overall evaluation value, or to select a group of the top clusters.

The goal of this proposal is to identify and evaluate relevant opinions within a large group from the decision-maker's point of view. Within this respect, it is a challenge trying to accurately reflect someone's point of view. However, by using soft computing techniques it is possible to provide a method to model and handle importance among opinions including as a novelty the use of LSP aggregation which reflects aspects of human decision-making. In this paper, we studied how a large group of opinions is re-

duced to some of them considered to be relevant by the decision-maker.

An advantage within the scope of this proposal is that it handles a large group of opinions gathered through social media, where the initially given preferences are not modified. Furthermore, it evaluates different points of view separately (i.e., previously clustered) unlike it occurs in some consensual processes and it permits the decision-maker to select the group of opinions that best suits his/her choice based on the combination of some cluster profile characteristics (i.e., cohesion, number of membership functions and number of noticeable opinions).

The remainder of this paper is structured as follows. Section 2 gives some preliminary concepts for clustering similar opinions and computing a cohesion measure. Section 3 describes the LSP method based on the verbalized approach of the generalized conjunction/disjunction aggregators. Section 4 describes how to identify and quantify relevant opinions within a large group using an illustrative example that demonstrates its applicability in a decision-making context. Section 5 concludes the paper and presents some opportunities for future work.

2 PRELIMINARIES

This section defines preliminary concepts to properly understand the remaining sections. These include concepts on fuzzy sets for representing expert opinions, some definitions to cluster similarly shaped membership functions, and how to compute a cohesion measure denoting the level of togetherness for each of the clusters.

2.1 Representing Expert Opinions

A membership function μ_A , from the preference point of view, represents a set of more or less preferred values of a decision variable x in a fuzzy set A . Hereby, $\mu_A(x)$ represents the intensity of preference or preference level in favor of value x (Dubois and Prade, 1997).

In this paper, trapezoidal membership functions are used considering that they are widely known (Klir and Yuan, 1995) and they could be built with a few input values through parameters a , b , c , and d (Equation 1) to represent the expert preferences over criteria (Dubois, 2000). These dividing points between the segments, denoted by the aforementioned parameters, hold the relation $a \leq b \leq c \leq d$ among them.

$$\mu_A(x) = \begin{cases} 0 & , x < a \\ \frac{x-a}{b-a} & , a < x < b \\ 1 & , b \leq x \leq c \\ \frac{d-x}{d-c} & , c < x < d \\ 0 & , x \geq d \end{cases} \quad (1)$$

If we return to the introductory example, trapezoidal membership functions allow experts to express the *usefulness level* using percentages to denote their preferences (Eshragh and Mamdani, 1979; Pedrycz, 2013). In this way, experts might use expressions like “the usefulness level is above 65%” (Figure 1a) hereby $b = 65\%$, “it is below 40%” (Figure 1b) hereby $c = 40\%$ or “it is between 25% and 50%” (Figure 1c) hereby $b = 25\%$ and $c = 50\%$. These are cases where $P(x) = 1$ denote the highest level of preference. Analogously, other expressions given by the experts will lead us to denote the lowest level of preference agreement on the criterion where $P(x) = 0$.

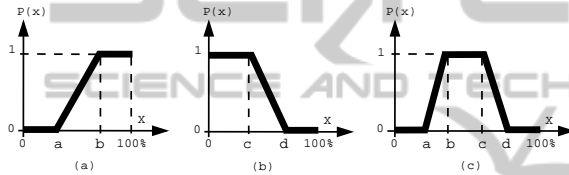


Figure 1: Trapezoidal membership functions expressing expert preferences through percentages.

Triangular membership functions, a particular case of trapezium, could be treated considering that b and c have equal values. This particular membership function could be used by experts through expressions “around x ” where x denotes the highest level of agreement on the criteria, here $x = b = c$, and the spread of less preferred values (i.e., between a and d) might vary among experts.

2.2 Clustering Similar Opinions: A Shape Based Approach

Considering that we use a shape-similarity method proposed in (Tapia-Rosero et al., 2014), within this subsection some definitions borrowed from (Tapia-Rosero et al., 2014) are included to make this paper self-contained. The shape-similarity method assumes that similar opinions are reflected by similarly shaped membership functions. It uses as inputs several membership functions, representing the opinion of experts over a specific criterion, and builds clusters of similar opinions (Figure 2).

The shape-similarity method has three phases that could be summarized as follows:

1. A *shape-symbolic notation* for each normalized membership function is built, which depicts

a membership function through a sequence of symbolic-characters (See Definition 1).

2. A *similarity measure* in the unit interval among shape-symbolic notations is obtained, where 0 denotes no similarity and 1 denotes full similarity between them.
3. A *clustering step* is performed based on the aforementioned similarity measure between notations. The clustering stops when the highest similarity is considered too low according to a previously determined threshold τ , where $0 \leq \tau \leq 1$. The selected threshold τ allows us to adjust the clusters starting from the highest similarity with the smallest number of membership functions (i.e., where each cluster contains a single membership function) to a lower similarity with the highest number of membership functions (i.e., one cluster containing all the membership functions).

Let $S^{category} = \{+, -, 0, 1, L, I, H\}$ be the set that is used to represent the category of a segment in a membership function, and S^{length} a linguistic term set used to represent its relative length on the X-axis compared to the sum of all segments. Using the aforementioned sets, a *symbolic-character* is defined as follows:

Definition 1. A *symbolic-character* is a representation of a segment in a membership function as a pair $\langle t, r \rangle$ with $t \in S^{category}$ and $r \in S^{length}$; where t represents the category of the segment and r depicts its relative length by means of a linguistic term.

In this way, each segment of the membership function uses a sign $\{+, -\}$ to represent its slope, a value $\{0, 1\}$ to represent its preference level on the criterion (i.e., the lowest level or the highest level of agreement respectively) and a letter $\{L, I, H\}$ to denote a *low*, *intermediate* or *high* point (e.g., a peak in a triangular membership function corresponds to a high point annotated as H). The linguistic term set S^{length} , depicted in Figure 3, expresses the relative length of the segment on the X-axis by means of labels (e.g., the label ES corresponds to an “extremely short” segment while label EL corresponds to an “extremely long” segment).

Figure 4 shows a trapezoidal membership function with five segments, each of them represented by a shape-symbolic character. Thus, the *shape-symbolic notation* for this figure could be expressed as:

$$\langle 0, S \rangle \langle +, VS \rangle \langle 1, S \rangle \langle -, ES \rangle \langle 0, EL \rangle$$

Hereafter we will consider that using different thresholds, different clusters containing similarly shaped membership functions were obtained. Thus, for each threshold τ a set of k clusters $C_\tau =$

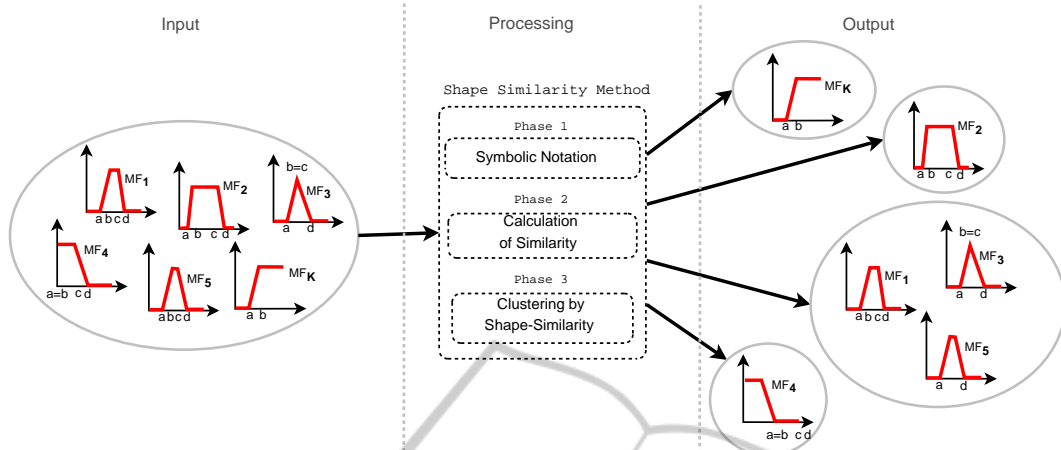


Figure 2: General architecture of the shape-similarity method.

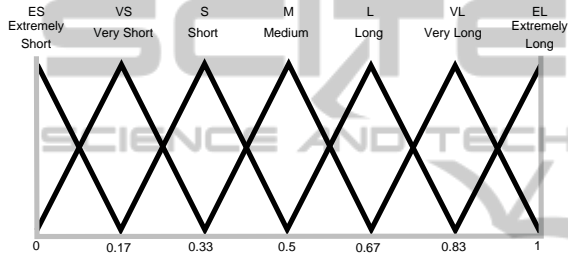


Figure 3: Linguistic term-set S^{length} represented by triangular membership functions denoting the relative length of segments on the X-axis.

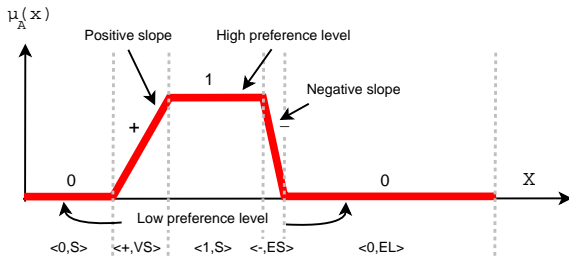


Figure 4: Segments of a trapezium and their corresponding shape-symbolic characters.

$\{C_1, \dots, C_k\}$, where each cluster C_j will be represented through an array of n characteristics or attributes (a_{1j}, \dots, a_{nj}) is used.

2.3 A Cohesion Measure for Expert Preferences

When several clusters of membership functions, representing similar expert opinions, are present it is possible to establish a way to compare them for further processing. This paper proposes to use a *cohesion measure*, which computes the level of togetherness among the membership functions contained

in the cluster. We assume that clusters with a high level of cohesion are more confident than those with a lower cohesion level, since they are closer and do not necessary overlap. On the one hand, we could think about a cluster that contains one hundred membership functions representing the same opinion (i.e., each of them has the same membership function representation) where we graphically expect a group of membership functions with the highest *cohesion* (i.e., these membership functions will be drawn one over the other). On the other hand, we could think about a cluster with the same amount of membership functions where some of them overlap and others are close enough to be considered similar. The latter scenario might lead us to graphically identify the boundaries where all the membership functions are contained, however this cluster will have a lower cohesion than the first one. In a decision-making context, the first scenario is the “ideal case” that might be considered unrealistic, while the second scenario could guide us to think about a group of similar opinions where the degree of similarity might be given by a *cohesion measure*. For example, Figure 5 shows two clusters with different levels of cohesion.

There are several strategies to compute the level of togetherness or cohesion among the membership

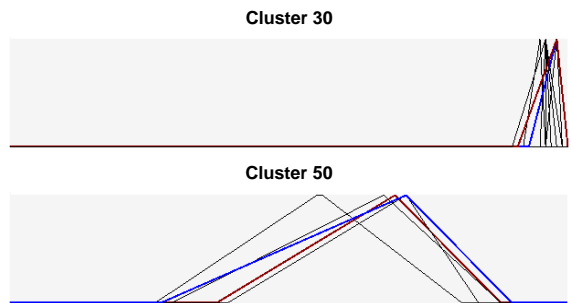


Figure 5: C_{30} denotes a higher cohesion level than C_{50} .

functions contained in a cluster that might be considered. In (Tapia-Rosero and De Tré, 2013) two of them have been proposed, however in this paper we will only consider the geometrical approach. The geometrical approach takes into account the area contained between the upper and lower boundaries compared to the total available area (Figure 6).

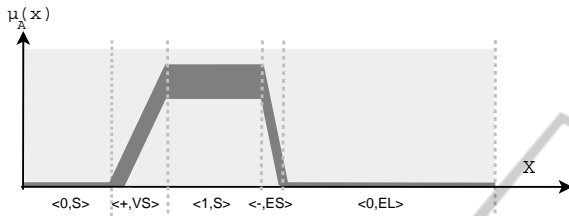


Figure 6: Area contained between boundaries (dark gray) compared to the available area (light gray).

Equation 2 sets a general form to obtain the cohesion measure in cluster C_j with threshold τ based on these area comparisons.

$$cohesion(C_j, \tau) = 1 - \frac{A^U - A^L}{A^T}. \quad (2)$$

Hereby, A^U denotes the area under the upper bound, A^L denotes the area under the lower bound and A^T corresponds to the total present area. For illustration purposes, the computed cohesion for clusters (using Equation 2) C_{30} and C_{50} with $\tau = 0.95$ are 0.9350 and 0.7547 respectively.

3 LOGIC SCORING OF PREFERENCE METHOD

Within this paper, the *logic scoring of preference (LSP) method* is used to evaluate relevant opinions considering, as mentioned in Section 2.2, that a set of k clusters $C_\tau = \{C_1, \dots, C_k\}$ were previously obtained. Here, each cluster C_j is represented through an array of n attributes $a_{i,j}$ where i is the identifier of the attribute and j is the identifier of the cluster.

The LSP method consists of a set of input attributes and elementary criteria reflecting the decision-maker's point of view on these attributes. The main advantage lies in that it is possible to build a precise model of logic aggregation of preferences, by combining the proper aggregation operators reflecting the user's needs (Dujmović and Nagashima, 2006). In order to proceed with the evaluation, the LSP method has the following steps:

- 1) Considering the decision-maker's point of view, it is necessary to define his/her evaluation attributes.

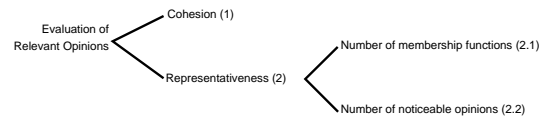


Figure 7: Example of a system attribute tree for evaluating relevant opinions.

Several attributes can be considered, thus the first step allows us to *create a system attribute tree*. In this step, different characteristics or attributes for relevant clusters of opinions are stated and hierarchically structured. For example, if we want to evaluate relevant opinions the decision-maker can use the structure shown in Figure 7.

The leaves of the tree represent the elementary attributes $(a_{1,j}, \dots, a_{n,j})$ of cluster C_j . These are not further decomposed, they have been previously measured and they are ready to be evaluated. Notice that an intermediate node (e.g. *Representativeness*), depicts that the attribute has been decomposed in more elementary attributes (i.e., number of membership functions and number of noticeable opinions).

For the sake of readability, the elementary attributes of Figure 7 include their identifiers in parenthesis, i.e. the identifiers for cohesion, number of membership functions and number of noticeable opinions are 1, 2.1 and 2.2 respectively.

- 2) The evaluation of the elementary attributes is based on their level of satisfaction or preference. Thus, the second step is to *define the elementary criteria*, through functions G_i that determine the elementary preference score reflecting the acceptable and unacceptable values of attribute i . In this step, a fuzzy set for each elementary attribute is used to represent the decision-maker's preference. For example, Figure 8 shows a membership function representing that the decision-maker accepts clusters with $cohesion \geq 0.5$ but he/she prefers $cohesion \geq 0.6$. Furthermore, the decision-maker considers that lower values, i.e. $cohesion < 0.5$, are not acceptable.

Thus, elementary criteria might be expressed using piecewise linear approximations of functions

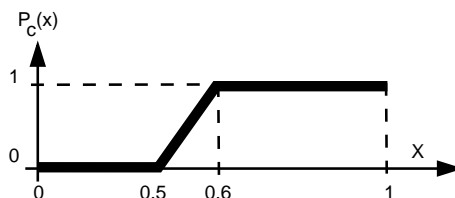


Figure 8: Example of the decision-maker's preference for elementary attribute "cohesion".

where after defining certain dividing points between segments (i.e., parameters a, b, c and d), we could use linear interpolation between them. As it has been mentioned in (Dujmović et al., 2010) “this approach yields a good combination of simplicity and accuracy”.

Once all the elementary criteria G have been defined, it is possible to evaluate all the attributes in each cluster. Thus, $e_{i,j} = g_i(a_{i,j})$ corresponds to the evaluation of attribute i in cluster C_j . For example, given $a_{1,30} = 0.9350$ which corresponds to the cohesion of cluster C_{30} we obtain $e_{1,30} = 1$ according to the decision-maker’s preference (Figure 8).

- 3) In order to satisfy all the decision-maker’s preferences we need to *create an aggregation structure*, which establishes the proper aggregation operators based on the generalized conjunction/disjunction (GCD) principle (Dujmović, 2007) while being consistent with the previously created system attribute tree.

For example, in order to obtain the *representativeness* of cluster C_j it is necessary to take into account its components (i.e., attributes $a_{2.1,j}$ and $a_{2.2,j}$) and the level of simultaneity or replaceability among them. Figure 9 shows the aforementioned *representativeness* annotated as $e_{2,j}$. In a similar way, we will obtain the evaluation of *relevant opinions* for cluster C_j given by e_j , where we need to aggregate its components $e_{1,j}$ and $e_{2,j}$. The level of simultaneity or replaceability will be given by the proper selection of the aggregation operators represented as A in this Figure.

In this paper, the aggregation structure allows us to obtain the evaluation of relevant opinions e_j for cluster j . However, to create this structure with the appropriate selection of aggregation operators it is necessary to introduce the generalized conjunction/disjunction principle in the next section.

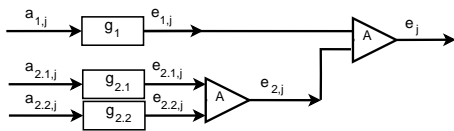


Figure 9: Example of the aggregation structure for elementary attributes 1, 2.1 and 2.2 of cluster C_j .

3.1 Generalized Conjunction/Disjunction

The generalized conjunction/disjunction (GCD) operator is a continuous logic function that integrates conjunctive and disjunctive properties in a single func-

tion (Dujmović and Larsen, 2007), denoted as $y = x_1 \diamond \dots \diamond x_n$, $x_i \in I = [0, 1]$, $i = 1, \dots, n$, and $y \in I$. GCD includes two parameters: the *andness* and the *orness*. The *andness*, $\alpha \in I$, expresses the conjunction degree used to denote simultaneity while the *orness*, $\omega \in I$, expresses the disjunction degree (Dujmović and Nagashima, 2006) used to denote replaceability. These parameters are complementary, i.e., $\alpha + \omega = 1$. The location of GCD with respect to conjunction and disjunction is defined in (Dujmović and Nagashima, 2006) as follows:

$$x_1 \diamond \dots \diamond x_n = \omega(x_1 \vee \dots \vee x_n) + \alpha(x_1 \wedge \dots \wedge x_n)$$

If $\alpha > 0.5 > \omega$, the expression $x_1 \diamond \dots \diamond x_n$ is called *partial conjunction* and is denoted by $x_1 \Delta \dots \Delta x_n$. If $\alpha < 0.5 < \omega$, the expression $x_1 \diamond \dots \diamond x_n$ is called *partial disjunction* and is denoted by $x_1 \nabla \dots \nabla x_n$. If $\alpha = \omega = 0.5$, the expression $x_1 \diamond \dots \diamond x_n$ is called the *neutrality function*, which is implemented as the arithmetic mean and is denoted by $x_1 \ominus \dots \ominus x_n$.

Although GCD can be implemented in several ways (Dujmović, 2008), within this paper we will only consider an implementation based on the weighted power means (WPM) as follows:

$$x_1 \diamond \dots \diamond x_n = (W_1 x_1^r + \dots + W_n x_n^r)^{\frac{1}{r}}, \quad (3)$$

hereby W_i denotes the weight assigned to the parameter x_i and the parameter r can be computed as a function of andness α using a suitable numerical approximation (Dujmović, 2007).

Table 1 includes the corresponding orness, andness and exponent r for 17 levels of GCD implemented using WPM as a reference. Notice that symbols D and C correspond to full disjunction ($\omega = 1$), and full conjunction ($\alpha = 1$) respectively.

3.2 GCD Verbalized Approach

The GCD verbalized approach presented in (Dujmović, 2012) facilitates the use of the LSP method. Within this approach the decision-maker specifies the overall degree of importance using a multi-level overall importance scale (Table 2) for each attribute.

The multi-level overall importance scale has L levels from “lowest” to “highest”, denoted S for simultaneity and R for replaceability. Thus, the decision-maker should provide the overall importance and the selection of simultaneity or replaceability. This information will allow us to obtain the appropriate aggregator.

Table 1: Aggregation operators for 17 levels of GCD implemented by WPM.

Symbol	Orness(ω)	Andness(α)	Exponent r
D	1	0	$+\infty$
D++	0.9375	0.0625	20.63
D+	0.8750	0.1250	9.521
D+-	0.8125	0.1875	5.802
DA	0.7500	0.2500	3.929
D+	0.6875	0.3125	2.792
D-	0.6250	0.3750	2.018
D-	0.5625	0.4375	1.449
A	0.5	0.5	1
C-	0.4375	0.5625	0.619
C-	0.3750	0.6250	0.261
C+	0.3125	0.6875	-0.148
CA	0.2500	0.7500	-0.72
C+	0.1875	0.8125	-1.655
C+	0.1250	0.8750	-3.510
C++	0.0625	0.9375	-9.06
C	0	1	$-\infty$

Table 2: Overall importance scale with $L = 16$ levels.

Level	Overall importance
16	Highest
15	<i>Slightly below highest</i>
14	Very high
13	<i>Slightly above high</i>
12	High
11	<i>Slightly below high</i>
10	Medium-high
9	<i>Slightly above medium</i>
8	Medium
...	...
0	Lowest

In the case of n attributes of overall importance, (S_1, \dots, S_n) for simultaneity, the andness α is defined as the mean normalized overall importance:

$$\alpha = \frac{(S_1 + \dots + S_n)}{nL}, S_i \in [0, L] \quad (4)$$

In a similar way in the case of replaceability (R_1, \dots, R_n) , the orness ω is defined as:

$$\omega = \frac{(R_1 + \dots + R_n)}{nL}, R_i \in [0, L] \quad (5)$$

Within this paper, clusters with a large number of membership functions are considered to be important, but a cluster with a single membership function given by a noticeable expert might also be relevant. In this case, the representativeness of the cluster indicates *replaceability* among the number of membership functions and the number of noticeable represented opinions. For example, if the decision-maker considers that the “representativeness” of a cluster is given by

the number of membership functions (R_1 considered as *very high*) and the number of noticeable opinions (R_2 considered as *high*), the computation of the *orness* level is as follows:

$$\omega = \frac{(R_1 + R_2)}{nL} = \frac{14 + 12}{2(16)} = 0.8125. \quad (6)$$

Once the level of andness/orness have been given by the decision-maker, we need to map these into normalized weights $W_1 + \dots + W_n = 1$. Although it is possible to use the GCD verbalized approach to compute the weights with ease (Dujmović, 2012), within this paper we consider that these will be given by the decision-maker as well. For example, in the case that the decision-maker considers the representativeness components equally important, then both weights are 0.5.

The generalized conjunction/disjunction aggregation adequately reflects the reasoning and preferences of the decision-maker and it is possible to extend it with ease, i.e. through changes in the model. In this paper GDC aggregation principle is used to *create the aggregation structure* of the LSP method (step 3) where the attributes have been evaluated and aggregated taking into account the decision-maker’s preferences.

4 EVALUATING RELEVANT OPINIONS

The aim of this section is to describe the steps that allows us to distinguish clusters that are relevant to represent expert opinions in a group decision-making context, where these have been gathered through social media. Here, the main contribution is to provide a method to handle and model the importance of opinions from the decision maker’s point of view, including as a novelty the application of LSP reflecting aspects of human decision-making.

On the assumption that similar opinions have been clustered by a shape based approach and that certain attributes of each cluster are available, we ask the decision-maker for his/her preferences to reflect his/her point of view in the selection of relevant opinions. Thus, the steps to evaluate relevant opinions using LSP are described as follows (Figure 10):

1. *Creation of a system attribute tree* made by the decision-maker based on the available attributes of the cluster. The attributes are hierarchically organized, where the leaves of the tree correspond to the elementary attributes selected from the cluster

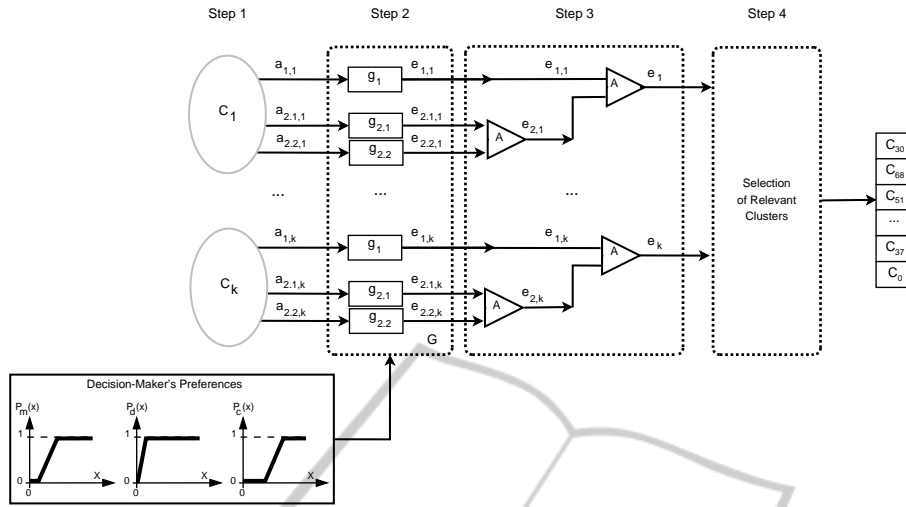


Figure 10: Evaluation of relevant opinions from clusters grouped by shape-similarity.

and the root will lead us to the overall evaluation value.

2. *Definition of elementary criteria* to reflect the acceptable and unacceptable values for each elementary attribute. These will be given by the decision-maker as membership functions.
3. *Creation of an aggregation structure* using the GCD verbalized approach. This step allows us to obtain the aggregation operators and weights reflecting the decision-maker's point of view with ease.
4. *Selection of relevant clusters* based on the top overall evaluation values.

For illustration purposes, the aforementioned steps will be described using the following example. A company wants to know the perceived “level of usefulness” of adding a digital lock (new feature) in a previously well positioned suitcase model (product), from a not uniform crowd (i.e., a large group of opinions with different points of view). Therefore, the company gathered this information using a fan page of the original product. In this case, the non uniformity of opinions is given by different levels of knowledge, areas of expertise and personal profiles. In this example, the head of the design department acts as the decision-maker.

The decision-maker considers that all the opinions are important, but those given by a specific profile (i.e., frequent flyers) will be considered noticeable. Additionally, he considers opinions within a large group relevant.

The opinions about the “level of usefulness” were clustered by the shape-similarity method presented in Section 2.2 under the assumption that sim-

ilarly shaped membership functions represent similar opinions. Hereby, a set of $k = 50$ clusters $C = \{C_1, \dots, C_{50}\}$ have been obtained representing a total of $t = 120$ opinions. For each cluster C_j the cohesion measure $a_{1,j}$ (cf. Equation 2), the number of membership functions $a_{2.1,j}$, and the number of noticeable opinions $a_{2.2,j}$ are computed.

Step 1. The decision-maker's point of view, is reflected in the system attribute tree shown in Figure 11.



Figure 11: System attribute tree for evaluating relevant opinions within a large group.

This system attribute tree establishes that the *representativeness* of the cluster is given by a combination of the number of membership functions and the number of noticeable opinions. In a similar way, it is indicated that the *evaluation of relevant opinions* is given by the cohesion and the representativeness of the cluster.

Step 2. The decision-maker's preferences were given through trapezoidal membership functions $P_c(x)$ for elementary attribute cohesion, $P_m(x)$ representing the number of membership functions and $P_d(x)$ for the number of noticeable opinions. These membership functions are shown in Figure 12.

The aforementioned membership functions reflect his acceptable and unacceptable values for each elementary attribute. As mentioned in Section 3, we could obtain his preference for values that lie in the slopes using a linear approximation.

For example, let us consider cluster C_{29} and its at-

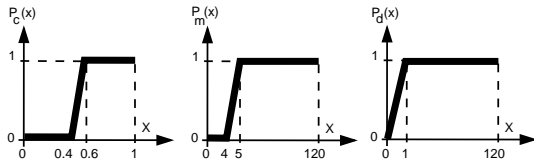


Figure 12: Decision-maker's preferences for elementary attributes cohesion $P_c(x)$, number of membership functions $P_m(x)$ and number of noticeable opinions $P_d(x)$.

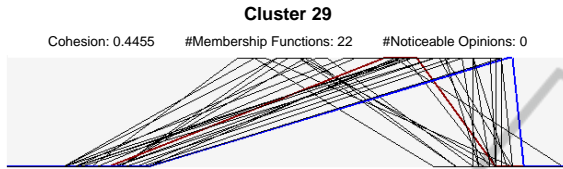


Figure 13: Cluster C_{29} and its attributes.

tributes $a_{1,29} = 0.4455$, $a_{2,1,29} = 22$ and $a_{2,2,29} = 0$ shown in Figure 13. In order to evaluate the cohesion $e_{1,29}$ from the decision-maker's point of view, we need to interpolate its value using $g_1(x) = \frac{x-0.4}{0.6-0.4}$, as follows:

$$e_{1,29} = g_1(0.4455) = \frac{0.4455 - 0.4}{0.6 - 0.4} = 0.2275$$

Within this step, all the attributes of each cluster will be evaluated using their corresponding function to reflect the decision-maker's preferences. Thus, e.g., $e_{2,1,29} = 1$ and $e_{2,2,29} = 0$.

Step 3. In order to build the aggregation structure it is necessary to select the aggregation operators properly. In this paper, these will be selected using the GCD verbalized approach which allows the decision-maker to use the overall importance scale in Table 2. In this example, the decision-maker considers that the cohesion and the representativeness in a cluster should be simultaneously satisfied. Here, the importance of each attribute has been established as follows: The cohesion is "high" ($S_1 = 12$) and the representativeness is "medium high" ($S_2 = 10$). With this approach the level of andness α is given by

$$\alpha = \frac{S_1 + S_2}{nL} = \frac{12 + 10}{2(16)} = 0.6875$$

The obtained α value allows us to note that even though the cohesion and the representativeness should be simultaneously satisfied, the level of andness is not too high. Thus the minimal partial conjunction where both parameters are mandatory is used. In a similar way the aggregation operator for the representativeness is obtained as shown in Equation (6) where the level of orness ω is 0.8125. These aggregators are annotated by symbols $C-+$ and $D+-$, and using Table 1 we obtained the r exponents -0.148 and 5.802 respectively.

Next, the decision-maker has to select the weight of each attribute denoting its importance. For instance, if the *cohesion* is two times more important than the *representativeness* then the weights for these attributes are $W_1 = 0.67$ and $W_2 = 0.33$ respectively. In a similar way, if the components for the *representativeness* are equally important then their weights are 0.5 (i.e., $W_{2,1} = W_{2,2} = 0.5$). Hence, the aggregation structure to evaluate relevant opinions within a large group, including the weight of each attribute, is shown in Figure 14.

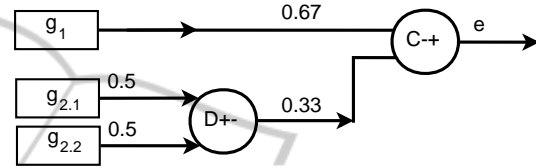


Figure 14: Aggregation structure based on the decision-maker's point of view.

Using the previously obtained aggregation structure we obtain a single value representing the overall evaluation of relevant opinions for each cluster (Equation 3). For illustration purposes, let us compute the overall evaluation of relevant opinions for cluster C_{29} . First, let us evaluate its *representativeness* given the selected aggregator $D+-$.

$$e_{2,29} = (0.5(e_{2,1,29})^r + 0.5(e_{2,2,29})^r)^{\frac{1}{r}}$$

$$e_{2,29} = (0.5(1)^{5.802} + 0.5(0)^{5.802})^{\frac{1}{5.802}}$$

$$e_{2,29} = 0.887393$$

Then, in a similar way, using aggregator $C-+$ we will compute the overall evaluation of relevant opinions as follows:

$$e_{29} = (0.67(e_{1,29})^r + 0.3(e_{2,29})^r)^{\frac{1}{r}}$$

$$e_{29} = (0.67(0.2275)^{-0.148} + 0.3(0.88739)^{-0.148})^{\frac{1}{-0.148}}$$

$$e_{29} = 0.34610$$

Thus, the evaluation of cluster C_{29} is given by the previously obtained value.

Notice, that using this approach it is possible to easily change the input parameters, given by the decision-maker, in order to accurately represent his/her point of view. For example, if the decision-maker would have changed the given weights (i.e., 33% for cohesion and 67% for representativeness) in the aggregation structure, the overall evaluation value would have been 0.5490246.

Step 4. In this step, the purpose of selecting relevant clusters is based on the selection made by the decision-maker from the previously evaluated clusters. It is possible that some decision-makers select

only the cluster with the best evaluation, i.e. the highest value, while other decision-makers prefer to select a group of the top clusters. Within this example, the decision-maker had selected the “top 5” clusters representing relevant opinions from his point of view.

It is worth to mention that within this example, the decision-maker started with 120 opinions gathered from social-media that were grouped into 50 clusters. Processing these clusters based on this proposal, allowed the decision-maker to select the top 5 clusters representing relevant opinions taking into account the cohesion and the representativeness of the clusters.

One of the advantages of the presented approach is that the flexibility in the LSP method allows the decision-maker, changing the definition of elementary criteria, in order to select relevant opinions that best suits his/her point of view. One remark within this respect is that the number of noticeable opinions, considered as a component of the representativeness, might be extended in order to represent different levels of importance (e.g., low, intermediate, high, etc.) among experts. In this case, it is possible to assign different weights to each expert opinion and its normalization will become part of the criteria definition. In a similar way the number of membership functions could be replaced by the *relative* number of membership functions, considering the total number of present opinions.

In order to validate the results of this proposal, five experts of the soft-computing area, were asked to rank a small selection of clusters (i.e., eight clusters from the original group of 50) based on the cohesion and the number of membership functions. Here, all the experts had the same selection for the top 4 and the cluster with the lowest overall evaluation, but the order of the other intermediate clusters were slightly different. Based on the computation described in this proposal, those intermediate clusters had a slightly different value in the overall evaluation, which justifies the small differences among the experts. However, more elaborated experiments should be performed, and are subject of further study.

5 CONCLUSIONS

This paper proposed to evaluate relevant opinions within a large number of expert opinions, expressed as membership functions, that might be gathered through social media. A shape-similarity method is used to cluster similar preferences in order to reduce the number of evaluations for different points of view. The evaluation results from selecting the best combination of *cohesion* and *representativeness* in the avail-

able clusters from the decision-maker’s point of view. The cohesion is a measure obtained from computing the area among the upper and lower bounds of the cluster compared to the total available area, while its representativeness is given by aggregating the number of membership functions and the number of noticeable opinions. In order to properly reflect the decision-maker’s point of view this proposal uses the LSP method that builds a precise representative model of logic aggregation of preferences.

The main advantage of this proposal is that it can handle a large group of opinions gathered through social media, where the preferences initially given are not modified. Furthermore, it evaluates different points of view separately (i.e., previously clustered) and it permits the decision-maker to select the group of opinions that best suits his/her choice (i.e., given as preferences in the LSP method) based on the combination of some cluster profile characteristics. Within this paper the cluster characteristics or attributes are a cohesion measure, the relative number of membership functions and the number of noticeable opinions.

We consider exploring some crowdsourcing applications as opportunities for future work, and evaluating clusters of opinions with different strategies in order to compare them with the presented approach are subject to further study as well.

ACKNOWLEDGEMENTS

This research is supported by Escuela Superior Politécnica del Litoral (ESPOL) and it is financed by SENESCYT under Ph.D. studies 2014.

REFERENCES

- Cabrerizo, F. J., Ureña, R., Pedrycz, W., and Herrera-Viedma, E. (2014). Building consensus in group decision making with an allocation of information granularity. *Fuzzy Sets and Systems*.
- Dubois, D. and Prade, H. (1997). The three semantics of fuzzy sets. *Fuzzy Sets and Systems*, 90:141–150.
- Dubois, D., P. H. (2000). *Fundamentals of Fuzzy Sets (THE HANDBOOKS OF FUZZY SETS Volume 7)*. Springer.
- Dujmović, J. (2007). Continuous Preference Logic for System Evaluation. *IEEE Transactions on Fuzzy Systems*, 15(6):1082–1099.
- Dujmović, J. (2008). Characteristic forms of generalized conjunction/disjunction. *2008 IEEE International Conference on Fuzzy Systems (IEEE World Congress on Computational Intelligence)*, pages 1075–1080.

- Dujmović, J. (2012). Andness and orness as a mean of overall importance. In *Fuzzy Systems (FUZZ-IEEE), 2012 IEEE International Conference on*, pages 1–6. IEEE.
- Dujmović, J. and De Tré, G. (2011). Multicriteria methods and logic aggregation in suitability maps. *International Journal of Intelligent Systems*, 26(10):971–1001.
- Dujmović, J., De Tré, G., and Van De Weghe, N. (2010). LSP suitability maps. *Soft Computing*, 14(5):421–434.
- Dujmović, J. and Larsen, H. L. (2007). Generalized conjunction/disjunction. *International Journal of Approximate Reasoning*, 46(3):423–446.
- Dujmović, J. and Nagashima, H. (2006). LSP method and its use for evaluation of Java IDEs. *International Journal of Approximate Reasoning*, 41(1):3–22.
- Eshragh, F. t. and Mamdani, E. (1979). A general approach to linguistic approximation. *International Journal of Man-Machine Studies*, 11(4):501–519.
- Klir, G. J. and Yuan, B. (1995). *Fuzzy Sets and Fuzzy Logic: Theory and Applications*. Prentice Hall.
- Pedrycz, W. (2013). *Granular Computing: Analysis and Design of Intelligent Systems*. CRC Press.
- Tapia-Rosero, A., Bronselaer, A., and De Tré, G. (2014). A method based on shape-similarity for detecting similar opinions in group decision-making. *Information Sciences*, 258:291–311.
- Tapia-Rosero, A. and De Tré, G. (2013). *A Cohesion Measure for Expert Preferences in Group Decision-Making*. Systems Research Institute Polish Academy of Sciences.