Penalizing manipulation strategies in Consensus processes

F. Mata L. Martínez J.C. Martínez

Dept. of Computer Science. University of Jaén, 23071 - Jaén, Spain.
{fmata, martin}@ujaen.es, jcmc0006@estudiante.ujaen.es

Abstract

In this work, how to minimize the effects of manipulation strategies in group decision making (GDM) problems is addressed. In this kind of problems a consensus phase may be carried out in order to achieve a minimum level of agreement before making a decision on a set of alternatives. In this setting, we discuss the possibility that an expert tries to reach his personal goal rather than the group one. One way to do this is manipulating the opinions that the expert provides to group opinion and another one is not changing his opinions like a strategy to impose his opinion to the group. The goal of this work is to propose a consensus model that provides a penalization mechanism in order to prevent the effects of both kinds of malicious behaviors. To achieve it, the mechanism assigns to each expert an importance which may change according to expert' behavior during the consensus process.

Keywords: Consensus, group decision making, penalization, manipulation.

1 INTRODUCTION

GDM problems are decision situations where given a set of possible alternatives, a group of experts try to achieve a common solution taking into account their individual opinions. Usually, GDM problems have been solved carrying out Selection Processes where experts obtain the best set of alternatives to solve the problem from their opinions [6]. However, some experts might consider that their opinions have not been taken into account in order to obtain the solution, and therefore they might not agree with the proposed solution. To avoid this situation, it is suitable to carry out a consensus process consisting of several rounds (see Figure 1) where experts discuss and change their opinions in order to reach a minimum agreement before making a decision [2, 7, 8, 9].

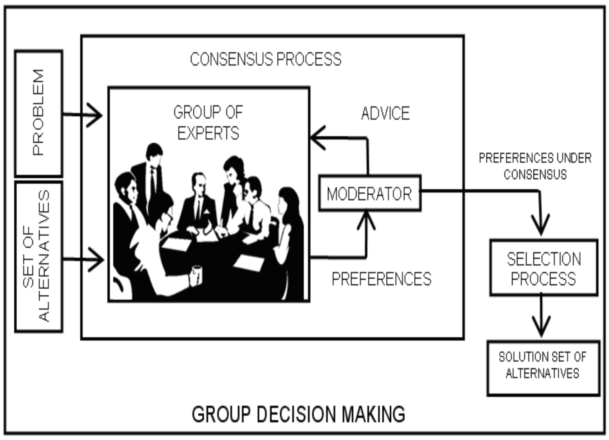

Figure 1: Resolution process of a GDM problem

A consensus process is usually coordinated by a moderator who evaluates the level of agreement of each round and helps experts to make their opinions closer to each other. However, the moderator may not be objective and to manage the consensus process toward his own goals. One way to avoid this situation is to automate the consensus reaching processes [7, 8, 10].

Most consensus processes are carried out into impartiality environments where all the experts' opinions are treated at the same level of importance [9]. However, the true goal of an expert could be the selection of his most preferred alternative and so he could be tempted to strategically manipulate his assessments in order to attain this goal [11, 13, 16]. In addition, an expert can decide not to follow the advice suggested by the moderator like a strategy to enhance his opinions, forcing the other group members to change their opinions towards the opinion of the manipulative expert.

To prevent the first type of manipulation (called *strategic manipulation*), some mechanisms have been proposed in the literature [5, 16, 17]. Also we can find an approach to combat the second type of malicious behavior in [12], but there is no any proposal to deal with both kinds of no desirable behaviors in consensus processes.

The aim of this contribution is to propose a consensus model that includes an initial approach of a penalization mechanism to minimize the effects of both types of manipulation attempts. To do so, in each consensus round, the model calculates the importance degree associated with each expert in basis to his predisposition for agreement. Then the collective opinion is obtained by means of an uninorm aggregation operator that takes into account the importance degrees. In this way, the model achieves to penalize manipulative experts. Finally, the model generates a set of changes for those experts whose assessments are more distant from the collective opinion.

This work is organized as follows. In Section 2, we introduce the GDM problems addressed in this work. In Section 3, the consensus model is briefly explained. In Section 4 we describe the penalization mechanism. In Section 5 the aggregation process based on uninorms is presented. Finally, some conclusions are drawn in Section 6.

2 GDM PROBLEMS AND PREFERENCES EXPRESSION

GDM problems are classically defined as decision situations in which a set of individuals (also called experts) $E = \{e_1, e_2, \ldots, e_m\}$ $(m \geq 2)$, try to choose the best alternative/s from a set of alternatives $X = \{x_1, x_2, \ldots, x_n\}$ $(n \geq 2)$. Utility vectors are usual preference structures used by the experts to give their opinions [3, 14],

$$U_{e_i} = \{u_i^1, u_i^2, \ldots, u_i^n\}$$

where each $u_i^k \in [0,1]$ represent the utility or value given by expert e_i for the alternative x_k. It is assumed that a greater value of u_i^k, implies a greater satisfaction of expert e_i on the alternative x_k. In this contribution, we deal with GDM problems where experts use utility vectors to provide their opinions.

3 MODEL DESCRIPTION

In this work we propose a penalization mechanism for a consensus reaching model in GDM problems. This model consists of three phases (see Figure 2):

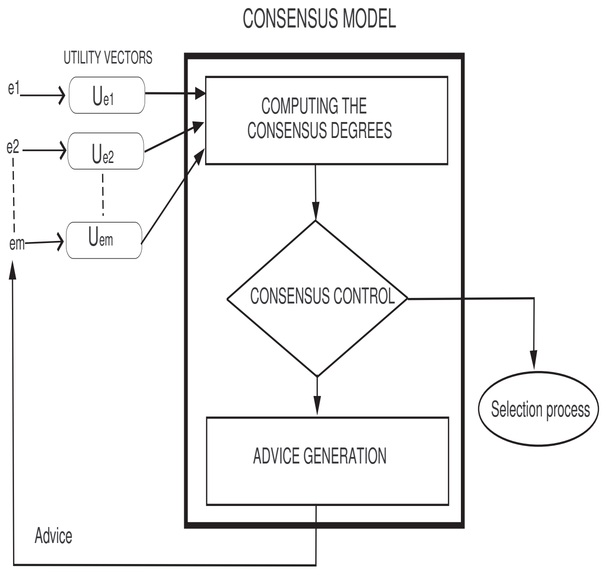

Figure 2: Consensus model

1. *Computing the consensus degree, cr.* The consensus degree cr measures the level agreement among the experts. Firstly, the model obtains a similarity vector for each pair of experts e_i, e_j with $(i < j)$

$$sv_{i,j} = \{s(u_i^1, u_j^1), s(u_i^2, u_j^2), \ldots, s(u_i^n, u_j^n)\}.$$

Each element of the vector is the similarity at level of alternatives between two experts' opinions. It is obtained from similarity function $s(\cdot)$:

$$s(u_i^k, u_j^k) = 1- \mid u_i^k - u_j^k \mid, \quad (1)$$

where u_i^k and u_j^k are the utility values given to alternative x_k by experts e_i and e_j respectively. Next, the model computes a consensus vector CV,

$$CV = \{cv^1, \ldots, cv^n\},$$

where each element of vector is obtained as the arithmetic mean of similarity vectors

$$cv^k = \theta(sv_{i,j}^k),\ i,j = 1, \ldots, m,\ i < j,\ k = 1, \ldots, n.$$

Finally, the consensus degree cr, which represents the overall agreement among all experts, is calculated as the average of values contained in CV,

$$cr = \frac{\sum_{k=1}^{n} cv^k}{n}.$$

2. *Consensus control.* In this phase the level of agreement cr is checked. If cr is greater or equal than a given consensus threshold γ then the consensus reaching process would stop. Otherwise,

the consensus process keeps going. The consensus threshold γ is the desired minimum consensus before starting the selection process and it should be fixed in advance.

3. *Advice generation.* If cr is not high enough, the model suggests a set of changes to the experts in order to make their preferences closer to each other. To do so, the model computes the collective utility vector by aggregating all experts' opinions

$$COLV = \{u_c^1, u_c^2, \ldots, u_c^n\},$$

where each

$$u_c^k = \delta(u_1^k, u_2^k, \ldots, u_m^k).$$

In this work, we propose to use an uninorm based operator δ as aggregation operator such as we explain in Section 5.

Moreover, a proximity vector

$$pv_i = \{s(u_i^1, u_c^1), s(u_i^2, u_c^2), \ldots, s(u_i^n, u_c^n)\}$$

and a proximity value p_i,

$$p_i = \frac{\sum_{k=1}^{n} pv_i^k}{n}$$

are calculated for each expert e_i. These values represent the similarity between expert's opinion and the collective one.

Then, the model proceeds as follow:

i. It selects the set of experts with less proximity, called $EXPCH$. These experts are the furthest from collective opinion and therefore they should change their opinions in order to increase the level of agreement in the next consensus round.

ii. It identifies the set of alternatives where the level of agreement is lower,

$$ALT = \{x_l \in X \mid cv^l < cr\}.$$

iii. For each expert $e_i \in EXPCH$ and for each alternative $x_l \in ALT$, the model suggest the direction of the changes of opinion, i.e to increase or decrease the current utility value. To do this, the model takes into account the individual opinion as well as the collective one and it applies one of the following three "direction rules":

DR.1. If $(u_i^l - u_c^l) < 0$, the expert e_i should increase the utility given to alternative x_l.

DR.2. If $(u_i^l - u_c^l) > 0$, the expert e_i should decrease the utility given to alternative x_l.

DR.1. If $(u_i^l - u_c^l) = 0$, the expert e_i has not to modify the utility given to alternative x_l.

4 PENALIZATION MECHANISM

The purpose of a consensus reaching process is to achieve a high level of agreement before making a decision. We may assume that all experts want to reach the agreement and therefore they are willing to renounce their initial assessments in order to reach it. However, some experts may have other intentions and try to manipulate the consensus reaching process in order to increase their personal interests or benefits. The manipulation may be seem at least from two points of view:

i) *Strategic manipulation of the preferences.* This is the case of an expert who tries to impose his preferences by means of enhancing his most preferred alternative. For example, e_i can maximize the utility value of the alternative x_1 and minimize the utility values of the rest of alternatives, $U_{e_i} = \{1, 0, 0, \ldots, 0\}$, like a strategy to impose his opinion. In this case, e_i may be manipulating his opinions in order to make the collective opinion closer to his particular opinion.

ii) *Disobedient behavior.* On the other hand, let us suppose that an expert decides not to follow the advice suggested by the model forcing the other group members to change their opinions towards his individual opinion. This behavior may also be seem as a strategy to manipulate the consensus process and perhaps to obtain personal goals instead of group goals.

In this section a mechanism to penalize both types of manipulation strategies are proposed. This mechanism is based on computing an importance degree or weight for each expert. Experts' weights are taken into account by the aggregation operator when the collective preference is obtained. So, an expert who completely discounts all the alternatives except his favorite, or he manifests a disobedient behavior by ignoring the model's advice, he would have little influence on the construction of the group opinion. Both types of malicious behaviors are considered to compute the experts' importance such as we present in the following subsections.

4.1 STRATEGIC PREFERENCES MANIPULATION

In [17, 18] Yager proposed that the importance of an expert would be directly related to the total degree of satisfaction allocated, i.e., according to the assessments given on all the alternatives. The more satisfied expert is the most important, thus, "happy" experts get more importance. So, if an expert rejects all alternatives except his preferred one, the expert's importance should be very low or null in the group opinion. This could be detected by checking the "extreme" utility values and computed its difference.

Let $t_i = 1 - (Max_j[u_i^j] - Min_j[u_i^j])$. The importance of an expert e_i that deliberately manipulates his preferences, wsm_i, is obtained as:

$$wsm_i = \frac{t_i}{Max_{(1 \leq k \leq m)}[t_k]} \tag{2}$$

That is, the expert is more important when his most extreme utility values are not very different.

4.2 DISOBEDIENT BEHAVIOR

This may happen when an expert decides not to apply the advice suggested by the model as strategy to impose his opinion. The importance of a "disobedient" expert, wdb_i, is computed as:

$$wdb_i = Max(0, 1 - \frac{n\_rejections}{max\_rejections}) \tag{3}$$

where $n\_rejections$ is the number of times that expert e_i has refused to make the changes proposed by the model, and $max\_rejections$ is the maximum number of times that an expert might refuse them. The $max\_rejections$ value should be fixed before starting the consensus process.

It is obvious that in the first consensus round the system has not suggested any advice yet, and therefore $wdb_i = 1$ for all $e_i \in E$.

About the disobedient behavior there are several open questions that we are studying how to solve. For example, what happens if an expert follows the advice by just changing his assessments in a minimal fraction? Let us suppose expert e_i gives $U_{e_i} = \{0.4, 0.5, 0.9\}$ and the model suggests decreasing the first and third value and increasing the second one. Then the expert provides $U_{e_i} = \{0.3999, 0.5001, 0.8999\}$. He has applied the advice, but clearly he is not being honest. On the other hand, what happens if the expert does not apply all the advice? Let us suppose e_i decides to change only the first two values, ignoring the third one. The discussion about when an expert shows a disobedient behavior is not trivial. We can try to identify disobedient experts by defining two parameters: a minimum change tax and a minimum percentage of changes. For example, we can establish that the minimum variation of the utility values should be ± 0.25 and the minimum percentage of changes carried out by the expert bigger than 85%. So an expert which does not observe these conditions, he might be considered "disobedient".

4.3 COMPUTING THE IMPORTANCE DEGREE OF AN EXPERT

When both weights have been calculated, then the importance degree for an expert w_i can be obtained. This value is computed by aggregating both weights using the Min operator,

$$w_i = Min(wsm_i, wdb_i). \tag{4}$$

In this way, the model always chooses the lowest weigh and therefore the penalization is higher.

Afterwards the model recalculates the utility values given by each expert taking into account the w_i value. The new values obtained are aggregated in order to build the collective utility vector $COLV$. This aggregation process is described in the next section.

5 USING UNINORMS AS AGGREGATION OPERATOR

Several aggregation operators may be used to obtain the collective opinion. In [8], we used the mean operator to aggregate the experts' preferences. However this operator has a disadvantage, an expert cannot know whether his assessments will increase or decrease the "support" for an alternative because it depends on the other experts' assessments.

In this work we propose to use an aggregation operator based on uninorms. Uninorms are an unification and generalization of the t-norm and t-conorm operators [1, 4, 15]. A uninorm is a mapping $R : I \times I \rightarrow I$, having the following properties:

1. $R(x, y) = R(y, x)$: Commutativity.
2. $R(x, y) \geq R(u, v)$ for $x \geq u, y \geq v$: Monotonicity.
3. $R(x, R(y, z)) = R(R(x, y), z)$: Associativity.
4. There exists some $g \in [0, 1]$ such that for all $x, R(x, g) = x$: Identity element.

Using uninorms, we can see that additional arguments have an effect on the aggregation depending on their relationship to a boundary point, the identity g, but this boundary point doesn't vary with the prior value aggregated, i.e., g is fixed. The identity provides an "a priori" knowledge of the effect of the values provided by the expert, so by providing a greater value than the identity, an expert knows "a priori" that he is helping to increase support for an alternative. On the other hand, providing a smaller value than the identity g, an expert knows he is helping to decrease support for an alternative, and providing a value equal to g the expert expresses his indifference with respect an alternative [18]. The behavior of a uninorm is very important in our case since according to property 4,

$$R(a_1, a_2, \ldots, a_n, g) = R(a_1, a_2, \ldots, a_n) \qquad (5)$$

then we can see that an utility value $u_i^k = g$ can be seen as an indifferent assessment with respect to the alternative x_k.

In our model, the result of the aggregating process is a collective utility vector

$$COLV = \{u_c^1, u_c^2, \ldots, u_c^n, \}$$

where each element is

$$u_c^k = R_g(u_1^k, u_2^k \ldots, u_m^k)$$

being R a uninorm operator with identity g.

In order to incorporate the experts' importance degrees in the aggregation process, we associate each u_i^k with its importance degree $w_i \in [0, 1]$ calculated in (4),

$$u_c^k = R_g((w_1, u_1^k), (w_2, u_2^k) \ldots, (w_m, u_m^k)).$$

Then, $COLV$ can be obtained as weighted uninorm aggregation. Let h_g be the importance transformation function that penalizes the experts' opinion and let $\hat{u}_i^k = h_g(w_i, u_i^k)$ be the penalized value, then we have that

$$u_c^k = R_g(\hat{u}_1^k, \hat{u}_2^k \ldots, \hat{u}_m^k).$$

The question now becomes: what is the form of function h_g. Several properties can be assigned to that kind of functions [17]:

i) It is required that $h_g(1, a) = a$, and $h_g(0, a) = g$, the identity of the uninorm. That is, the opinion from an expert that has 0 importance will not contribute to the aggregation.

ii) h_g should be monotonic with respect the value. Specifically, if a and $\hat{a}$ are two arguments such that $a \geq \hat{a}$ then any modification by importance will not interchange their order, and $\forall w$ $h_g(w, a) \geq h_g(w, \hat{a})$.

iii) $h_g(w, a)$ is bounded by its values for $w = 0$ and $w = 1$.

iv) As fourth condition we require that $h_g(w, a)$ transitions from $w = 0$ to $w = 1$ in a consistent monotonic manner. Let w and $\tilde{w}$ such that $\tilde{w} > w$. Then

- if $h_g(1, a) > h_g(0, a)$ $(a > g)$, then for all $\tilde{w} > w$ we have $h_g(\tilde{w}, a) \geq h_g(w, a)$;
- if $h_g(1, a) < h_g(0, a)$ $(a < g)$, then for all $\tilde{w} > w$ we have $h_g(\tilde{w}, a) \leq h_g(w, a)$;
- if $h_g(1, a) = h_g(0, a)$ $(a = g)$, then for all $\tilde{w} > w$ we have $h_g(\tilde{w}, a) \leq h_g(w, a)$.

In [17] we can find some interesting examples of such functions. Here there are two of them that are valid for any identity element g:
$h_g(w, a) = wa + \overline{w}g$,
$h_g(w, a) = (w \wedge a) \vee (\overline{w} \wedge g) \vee (a \wedge g)$,
with $\overline{w} = 1 - w$.

It can be seen that using a function h_g to transform the pairs (w_i, u_k^i) into a single numeric value $\hat{u}_i^k$ together with an uninorm to aggregate these values, we obtain an aggregation operator able to deal with experts with different importance degree and to penalize manipulative behaviors.

Example

Let us consider three experts and three alternatives. We use the function $h_g(w, a) = wa + \overline{w}g$ to obtain the penalized utility for any expert. As aggregation operator we use the called three Π operator proposed by Yager in [15],

$$R(x, y) = \frac{xy}{\overline{xy} + xy} \qquad (6)$$

where $\overline{x} = 1 - x, \overline{y} = 1 - y$. In this uninorm, the identity $g = 0.5$. Let us suppose the three experts' utility vectors are:

$U_{e_1} = \{1, 0, 0\}, U_{e_2} = \{0.4, 0.5, 0.9\}$ and $U_{e_3} = \{0.8, 0.7, 0\}$.

From those opinions, we might think e_1 is trying to strategically manipulate the consensus process. If we use (2) to compute the experts' importance associated with strategic manipulation behavior, we obtain the following results:

$$wsm_1 = 0, \quad wsm_2 = 1 \; and \; wsm_3 = 0.4.$$

Let us suppose we are in the first consensus round, and therefore $wdb_i = 1$ for all e_i (there is not disobedient behavior because the system has not generated any advice yet). Therefore, according to (4), the importance associated with each expert is

$$w_1 = 0, \quad w_2 = 1, \quad and \; w_3 = 0.4$$

We can see that the most important expert is e_2, and the less important expert is e_1. Taking into account these weights and using the function h_g described above, we obtain the penalized utility vector for each expert:

$\hat{U}_{e_1} = \{0.5, 0.5, 0.5\}$, $\hat{U}_{e_2} = \{0.4, 0.5, 0.9\}$ and $\hat{U}_{e_3} = \{0.62, 0.58, 0.3\}$.

Note that penalized preferences for expert e_2 (whose importance is $w_2 = 1$) have not changed, i.e, he has not been really penalized. However, all the penalized utilities of e_1 (whose importance degree is $w_1 = 0$) are 0.5, i.e. the value of g, the identity of the uninorm. This implies that according to (5), the opinions of expert e_1 will not affect to the collective vector, $COLV = \{0.52, 0.58, 0.79\}$, and therefore the other experts will not have to modify their opinions in order to be close to expert e_1.

6 CONCLUSIONS

In this contribution we have considered the possibility when experts try to manipulate a consensus process in order to impose their preferences on the group opinion. This manipulation may come from preferences that have been deliberately manipulated by the experts or from disobedient experts which do not change their preferences in order to attain their personal goals. A mechanism for a consensus model is suggested in order to minimize the negative effects of these attempts of manipulation. This mechanism is based on associating an importance degree with each expert and to apply a uninorm aggregation operator to obtain the collective preference.

Acknowledges

This work has been supported by the Research Project TIN2006-02121.

Referencias

[1] B. De Baets. *Fuzzy Logic and Intelligent Technologies for Nuclear Science and Industry.* World Scientific, Singapore, 1998.

[2] C. Carlsson, D. Ehrenberg, P. Eklund, M. Fedrizzi, P.Gustafsson, P. Lindholm, G. Merkuryeva, T. Riissanen, and A.G.S.Ventre. Consensus in distributed soft environments. *European Journal of Operational Research*, 61:165–185, 1992.

[3] L. Dombi. *A General Framework for the Utility-Based and Outranking Methods*, pages 202–208. in: Fuzzy Logic and Soft Computing. World Scientific, 1995.

[4] J.C. Fodor, R.R. Yager, and A. Rybalov. Structure of uni-norms. *International Journal of Uncertainty, Fuzziness and Knowledge-Based Systems*, 5:411–427, 1997.

[5] J.L. Garcia-Lapresta. Favoring consensus and penalizing disagreement in group decision making. *Journal of Advanced Computational Intelligence and Intelligent Informatics*, Forthcoming.

[6] F. Herrera, E. Herrera-Viedma, and J.L. Verdegay. A sequential selection process in group decision making with linguistic assessment. *Information Sciences*, 85:223–239, 1995.

[7] E. Herrera-Viedma, F. Herrera, and F. Chiclana. A consensus model for multiperson decision making with different preference structures. *IEEE Transactions on Systems, Man and Cybernetics-Part A: Systems and Humans*, 32(3):394–402, 2002.

[8] E. Herrera-Viedma, L. Martínez, F. Mata, and F. Chiclana. A consensus support system model for group decision-making problems with multigranular linguistic preference relations. *IEEE Transactions on Fuzzy Systems*, 13(5):644–658, 2005.

[9] J. Kacprzyk, M. Fedrizzi, and H.Nurmi. *Consensus Under Fuzziness.* Kluwer Academic Publishers, 1997.

[10] N.I. Karacapilidis and C.P. Pappis. A framework for group decision making support systems: Combining AI tools and OR techniques. *European Journal of Operational Research*, 103:373–388, 1997.

[11] J.S. Kelly. Almost all social choice functions are highly manipulable, but a few aren't. *Social Choice and Welfare*, 10:161–175, 1993.

[12] F. Mata, L. Martínez, E. Herrera-Viedma, and F. Herrera. A penalized consensus model for group decision making problems with multigranular linguistic information. In *Proceedings of EUROFUSE Workshop on New Trends in Preference Modeling (EUROFUSE'07)*, pages 23–29, Jaen (Spain), April 2007.

[13] D.A. Smith. Manipulability of commond social choice functions. *Social Choice and Welfare*, 16:639–661, 1999.

[14] T. Tanino. *On Group Decision Making Under Fuzzy Preferences*, pages 172–185. in: Multiperson Decision Making Using Fuzzy Sets and Possibility Theory. Kluwer Academic Publishers, 1990.

[15] R.R. Yager. Uninorm aggregation operators. *Fuzzy Sets and System*, 80:111–120, 1996.

[16] R.R. Yager. Penalizing strategic preference manipulation in multi-agent decision making. *IEEE Transactions of Fuzzy Systems*, 9(3):393–403, 2001.

[17] R.R. Yager. Defending against strategic manipulation in uninorm-based multi-agent decision making. *European Journal of Operational Research*, 141:217–232, 2002.

[18] R.R. Yager. *Using Importances in Group Preference Aggregation to Block Strategic Manipulation*, pages 177–191. in: Studies in Fuzziness and Soft Computing, Aggregation operators. Physica-Verlag Heidelberg, 2002.