The impact of biased experts in the aggregation of fuzzy preference relations

Laura Smith
Dr. Simon James and Prof. Gleb Beliakov
Deakin University

February 2015
Abstract

The group decision making process usually contains a consensus model which focuses on achieving some level of agreement amongst experts before their preferences are aggregated and an alternative is chosen. These consensus models look at the distances between the experts’ preferences and use this to provide each expert with a consensus level. If an expert’s consensus level is below a predefined threshold, the expert will be given recommendations to update certain preferences until they receive a consensus level that is considered acceptable when compared to the group. This report considers situations where the current methods for calculating the consensus between experts can be exploited by a biased expert and then explores some of the ways to reduce the impact of such situations.

Keywords aggregation functions, consensus, group decision making, induced OWA
1 Introduction

The group decision making (GDM) problem consists of finding the best option over a set of alternatives, given a set of individuals that provide their preferences [1]. There are several methods that can be used to obtain the experts’ preferences [2], in our research we focussed on fuzzy preference relations.

A fuzzy preference relation is a matrix of values that shows the expert’s preference of the row alternative over the column alternative. In contrast to boolean logic or strict logic, fuzzy logic naturally lends itself to the decision making process [3]. Experts are able to easily express the degree of their preference of one alternative over another, rather than a partial order relation that only indicates whether or not an alternative is preferred to another.

Within GDM the notion of consensus has received a lot of attention, as it is desirable that experts have some level of agreement before aggregating their preferences [2]. Several consensus models have been presented [4, 5], that give an indication of the level of agreement between the experts. The typical consensus model looks at the distance between experts preferences and provide each expert with a consensus level. If the consensus level is below a predefined threshold, the expert will be given recommendations to update the preferences that differ significantly from the group opinion. Once all the experts have a consensus level above the threshold, the preferences are aggregated through a selection process.

The standard operator used in the aggregation of the individual preferences is the arithmetic mean, the use of which assumes that all experts are providing preferences with no ulterior motivation. A simple review of mass ratings provided for food outlets and movies indicates that this is not always the case. If a system has the potential to be exploited, chances are at some time it will be. The purpose of the consensus process is to draw attention to experts who are exploiting the process by providing preferences that differ significantly from the group. However the use of the arithmetic mean can lead to situations where the current algorithms fail to pick up on extreme preferences. In this report we consider the impact that a biased expert can exert in a group decision making context using the arithmetic mean and analyse the model to understand why this takes place. We then propose alternative weighting allocations and analyse whether these provide a more robust method against extreme preferences.

This report is set out as follows. In section 2, we provide an overview of the group decision making problem and basic definitions. In section 3, we outline a current consensus model. In section 4, we analyse the current consensus model and show the impact of biased experts. In section 5, we then propose alternative methods of weighting experts and analyse the robustness of these methods. Finally in section 6, we conclude and discuss further directions for research.

2 Preliminaries

In this section we present definitions necessary for the remainder of the report.

2.1 Group decision making

Group decision making considers the problem of combining preferences of multiple individuals to find the solution best acceptable to the group [1].

We consider a group of $m$ experts $V = v_1, v_2, ..., v_m$ who provide preferences over $n$ alternatives $U = u_1, u_2, ..., u_n$. The experts express their preferences in the form of a fuzzy preference relation.
**Definition 1 (Fuzzy preference relation.)** A fuzzy preference relation (FPR) is a $n \times n$ matrix of values $p_{ij} \epsilon [0, 1]$, which represent an expert’s preference of alternative $i$ over alternative $j$. Where an FPR satisfies additive reciprocity, i.e. that $p_{ij} + p_{ji} = 1$, it holds that if $p_{ij} > 0.5$, then the expert prefers alternative $i$ to alternative $j$. Similarly if $p_{ij} < 0.5$ then $j$ is preferred to $i$. $p_{ij} = 0.5$ indicates indifference between the two alternatives.

Additively consistent FPRs can be generated from a set of raw scores $(x_1, x_2, ..., x_n)$ provided over the unit interval using the following equation:

$$p_{ij} = \frac{x_i - x_j + 1}{2} \quad (1)$$

Using a preference relation allows experts to break the larger decision into smaller decisions where they are able to focus solely on two alternatives at a time. However, this can also lead to an expert submitting an FPR that is intransitive. Some models for GDM include a consistency module [6], the purpose of which is to ensure that individual FPRs are consistent so that if an expert prefers $A \succeq B$ and $B \succeq C$, then they also prefer $A \succeq C$. Some examples of a consistency modules can be found in [6, 7].

The basic structure of a group decision making model is outlined in Figure 1.

![Figure 1: Group Decision Making process.](image)

The next stage of the GDM process is the consensus model. The notion of consensus is similar to that of agreement. Consensus models seek to increase the agreement amongst experts before aggregating the individual preferences for the selection process. The basic framework of a consensus model is outlined in Figure 2:

![Figure 2: The basic framework of a consensus model.](image)

The individual consistent FPRs are input into the consensus model. Through some algorithm, consensus levels are calculated and assigned to each expert. If an expert’s consensus level is above a predefined level of agreement known as a consensus threshold, then their FPR is moved on to the selection stage. If however, the consensus level is below the predefined threshold, the expert will be given recommendations to update their preferences which were furthest from the group opinion.

After the predefined level of consensus has been met by all experts, the preferences are aggregated and an alternative selected.
2.2 Aggregation functions

Once we have ensured that all the individual preference relations are consistent and meet the necessary requirement of agreement, we are then able to aggregate the FPRs and select an alternative.

Aggregation functions are functions that take multiple inputs and produce an output that is reflective of the inputs [8]. In GDM we want to take the individual preferences and combine them so that we can determine which alternative is best acceptable to the group as a whole.

**Definition 2 (Aggregation functions.)** An aggregation function $A : [0, 1]^n \rightarrow [0, 1]$ is a function non-decreasing in each argument and satisfying $A(0, ..., 0) = 0$ and $A(1, ..., 1) = 1$.

The standard averaging aggregation function adopted in most GDM contexts is the weighted arithmetic mean (WAM).

**Definition 3 (Weighted arithmetic mean.)** Given a weighting vector $\mathbf{w}$, the weighted arithmetic mean is given by

$$WAM_{\mathbf{w}(x_1, ..., x_n)} = \sum_{i=1}^{n} w_i x_i$$

The weight $w_i$ indicates the importance of the $i$th input.

3 Current consensus model

The purpose of the consensus model in group decision making is to ensure that the final decision is acceptable to all experts. The aim of the process is to identify experts who provide preferences that are significantly different from the group, and to provide these experts with the opportunity to update these preferences to be closer to the majority. In this section we outline a current consensus model.

The consensus model provided in [5] is based upon a similarity degree and a proximity degree. The similarity degree considers the distance between an expert’s preferences with the rest of the group and assigns this as a weight for the expert in the aggregation of the FPRs into a collective preference relation (CPR). A high similarity degree results in a higher weight, while a low similarity degree is penalised with a lower weight. The proximity degree considers the distance between each expert’s preferences and the CPR. The overall consensus level for each expert is calculated through a combination of the similarity degree and proximity degree.

![Consensus model outlined in [5].](image)

Figure 3: Consensus model outlined in [5].
The similarity degree is calculated in three levels where the similarity between experts \( v_k \) and \( v_h \) on the pair of alternatives \( x_i \) and \( x_j \) is given as:

\[
SD_{ij}^{kh} = 1 - |p_{ij}^k - p_{ij}^h|
\]

Level 1. The similarity degree on the pair of alternatives \( x_i \) and \( x_j \) of expert \( v_h \) to the rest of the experts in the group:

\[
SPA_{ij}^h = \frac{\sum_{l=1, l \neq h}^m SD_{ij}^{hl}}{m-1}
\]

Level 2. The similarity degree on alternative \( x_i \) of an expert \( e_h \) to the rest of the group:

\[
SA_i^h = \frac{\sum_{j=1}^n SPA_{ij}^h}{n}
\]

Level 3. The similarity degree on the preference relation of expert \( e_h \) to the rest of the experts in the group:

\[
SD^h = \frac{\sum_{i=1}^n SA_i^h}{n}
\]

The similarity degree is used to assign a weight to each expert relative to the others which is given by

\[
RSD^h = \frac{SD^h}{\sum_{k=1}^m SD^k}
\]

The purpose of this is clear, we want to give higher weights to experts that give preferences closer to the rest of the group and lower weights to those experts who provide preferences that aren’t reflected by the group opinion. These weights are then used to combine the individual FPRs, \( P^k \) into a CPR which is calculated using the following equation:

\[
CPR = \sum_{k} w_k P^k
\]

The proximity degree is then calculated similarly in three steps beginning with

\[
PD_{ij}^k = 1 - |p_{ij}^{CPR} - p_{ij}^k|
\]

The similarity degree and proximity degree are then used to calculate a consensus level for each expert. If the consensus level is not above the predefined threshold, usually around 0.8, experts will be given recommendations to update those preferences which are considered too different to the rest of the group.

4 Analysis of existing consensus methods

In this section we analyse the method outlined in section 3 and investigate the properties of the model that allow minorities to go unidentified while significantly affecting the decision.

The model shown above has four main steps. Firstly we calculate a similarity degree, then using this to weight the experts we calculate a CPR. We then calculate the proximity degree of
each expert from the CPR and finally assign each expert with a consensus level based upon the similarity degree and proximity degree.

We consider the calculation of the similarity degree and assigning of weights to the experts as all experts are able to exert some impact in this stage. Once the CPR has been calculated, it becomes the standard group opinion for the rest of the calculations.

4.1 Assigning weights to the experts

The purpose behind weighting the experts is to ensure that those experts whose preferences differ significantly from the group will have less impact on the overall ranking of the alternatives. To gain a further insight into the weight allocation and subsequent impact of each expert in the formation of the CPR, we ran experiments where we randomly generated 4 sets of additively consistent fuzzy preference relations via Equation (1). We then calculated the relative weight that would be assigned to each expert, and recorded the maximum and minimum weight each time. We ran this experiment 10000 times using the R software package [9]. With four experts and four alternatives, the minimum weighting given to an expert over these trials was 0.1926 while the maximum weighting was 0.2818. The small range of weights obtained during this experiment shows that even when an expert provides an FPR that is significantly different from the other experts, they are still able to have a notable influence on the CPR. The weights are calculated using the similarity degree that is assigned to each expert. We analyse the process of calculating the similarity degree below.

4.2 Calculation of the similarity degree

The calculation of the similarity degree consists of a series of arithmetic means that aggregate the distances between an expert’s preferences to the rest of the group. We provide an example where this generates an undesirable result and discuss the repercussions of using the arithmetic mean to aggregate the similarity distances.

Example 1 A widely circulated newspaper employs four food critics to provide their preferences over four high end restaurants A,B,C,D in order to rank the restaurants for an upcoming feature article. Three of the critics agree that the ranking order of the restaurants should be $A \succ B \succ C \succ D$. However, the owner of B is keen for their restaurant to gain the extra publicity, and being aware that A is the favourite, they bribe the fourth critic to provide a poor score for A with hopes to bring down the overall score for A enough for them to be ranked number 1. So the fourth expert, accepting the bribe, provides a ranking order of $B \succ C \succ D \succ A$.

The critics each provide a set of evaluations from which an additive pairwise preference relation is constructed using Equation (1). For experts 1-3, the scores are given as (1,0.8,0.5,0.2) corresponding to companies A to D respectively. Expert 4 allocates the scores (0,0.8,0.5,0.2), giving A an extremely low score, but ensuring a reasonably high similarity value with respect to companies B to D. The two corresponding preference relations are:

$$P_1 = P_2 = P_3 = \begin{bmatrix}
0.5 & 0.6 & 0.75 & 0.9 \\
0.4 & 0.5 & 0.65 & 0.8 \\
0.25 & 0.35 & 0.5 & 0.65 \\
0.1 & 0.2 & 0.35 & 0.5
\end{bmatrix}$$
and \( P^4 = \begin{bmatrix}
0.5 & 0.1 & 0.25 & 0.4 \\
0.9 & 0.5 & 0.65 & 0.8 \\
0.75 & 0.35 & 0.5 & 0.65 \\
0.6 & 0.2 & 0.35 & 0.5
\end{bmatrix} \)

Using the equations for the similarity degree the consensus model presented above, the weight allocated to experts 1-3 would be 0.2586 while the weight for expert 4 would be 0.2241. This results in the following collective preference relation.

\[
CPR = \begin{bmatrix}
0.5 & 0.4879 & 0.6379 & 0.7879 \\
0.5121 & 0.5 & 0.65 & 0.8 \\
0.3621 & 0.35 & 0.5 & 0.65 \\
0.2121 & 0.2 & 0.35 & 0.5
\end{bmatrix}
\]

which corresponds with an overall ranking of \( B \succ A \succ C \succ D \) for the alternatives.

The experts 1-3 receive a consensus level of 0.9477 and expert 4 receives a consensus level of 0.8335. Therefore we see that expert 4 was able to change the overall ranking of the restaurants without being identified by the consensus model as providing extreme preferences.

The overall ranking for the alternatives in Example 1 implies that B is the favourite alternative. However this is clearly not the alternative that is best acceptable to the group as a whole, so we look at how Expert 4 went unidentified by the model even though their preference was able to change the overall outcome.

In level 2 of the similarity degree calculation, the similarity of an expert over each alternative is calculated through the arithmetic mean. For alternative A, expert 4 has provided extreme preferences and so the similarity degrees for each preference will be low, and result in the SA for alternative A being low. However, for alternatives B, C and D, three of the four similarity degrees for the preferences are identical to the group, with only alternative A having a low similarity degree. This results in the extreme distance being averaged over four preferences, resulting in a SA for each of the alternatives B, C and D that are considered close to the group. In level 3 of the similarity degree, we calculate the arithmetic mean of the similarity degrees for each alternative. Alternative A has a low SA degree, however alternatives B, C and D have relatively high SA degrees. Again the low score is averaged over the four alternatives resulting in an overall similarity degree that is high enough to get expert 4’s FPR through the consensus model without being identified as containing an extreme preference.

4.3 Conclusions on existing consensus models

Our analysis of the consensus model outlined in [5] shows that it is possible for a biased expert to change the overall ranking of the alternatives which is an undesirable result given that the goal of the process is to find the alternative best acceptable to the group. We sum up our observations in the following points: The small range of weights calculated in the first stage of the consensus process mean that a biased expert still has a noticeable impact on the CPR and consequently the consensus levels assigned to each expert. Using the arithmetic mean can cause extreme preferences to go unnoticed as the extreme distance of one preference is averaged over four preferences.
5 Alternative weighting processes and their robustness

In this section we present several possible alternative methods to weight the experts in the consensus process. We provide an overview of each method and look at how each method deals with the biased expert in the example given above, however we note that because the method may give desirable results for one example does not mean that it is robust for all situations. We present each method as a direction for further research.

5.1 Increasing the range of weights assigned using transformation functions

The similarity degree assigned in the current model is a direct reflection of an individual’s preferences compared with the others in the group. This is a natural result, though we determined in our analysis that the range of weights that can be assigned directly from the similarity degree is quite small. As shown, this means that even if an expert is to give a preference that is opposite to the group, the minimum weight that can be allocated to that expert still allows their preferences to have a noticeable effect on the collective preference relation. One way to increase the range of weights is to use transformation functions on the original similarity degrees which will increase the distance between experts before normalising the relative similarity degrees and aggregating the collective preference relation. In Table 1, we show how raising the similarity degree to different powers affects the weight given to each expert from Example 2 and in turn how these weights affect the consensus levels calculated for each expert. The original similarity degrees for each expert were (0.9375, 0.9375, 0.9375, 0.8125), taking the square of these we obtain similarities of (0.8789063, 0.8789063, 0.8789063, 0.6601562). We can see here that the range in weights is increased significantly, which in turn has the desirable effect of decreasing the consensus level for the fourth expert.

<table>
<thead>
<tr>
<th>Transformation</th>
<th>Experts 1-3</th>
<th>Consensus Level</th>
<th>Expert 4</th>
<th>Consensus Level</th>
</tr>
</thead>
<tbody>
<tr>
<td>$SD$</td>
<td>0.2586</td>
<td>0.9477</td>
<td>0.2241</td>
<td>0.8335</td>
</tr>
<tr>
<td>$SD^2$</td>
<td>0.2666</td>
<td>0.9207</td>
<td>0.2002</td>
<td>0.7551</td>
</tr>
<tr>
<td>$SD^5$</td>
<td>0.2866</td>
<td>0.8490</td>
<td>0.1401</td>
<td>0.5964</td>
</tr>
</tbody>
</table>

Using these transformation functions, we look at the affect on the CPR. When using the transformation $SD^2$, we obtain the following CPR:

$$CPR = \begin{bmatrix} 0.5 & 0.4999 & 0.6499 & 0.7999 \\ 0.5001 & 0.5 & 0.65 & 0.8 \\ 0.3501 & 0.35 & 0.5 & 0.65 \\ 0.2001 & 0.2 & 0.35 & 0.5 \end{bmatrix}$$

When using the transformation $SD^5$, we obtain the following CPR:

$$CPR = \begin{bmatrix} 0.5 & 0.5299 & 0.6799 & 0.8299 \\ 0.4701 & 0.5 & 0.65 & 0.8 \\ 0.3201 & 0.35 & 0.5 & 0.65 \\ 0.1701 & 0.2 & 0.35 & 0.5 \end{bmatrix}$$
We can observe that the CPR is more reflective of the majority when the range in weights is increased by the transformation functions. This can be explored in greater depth in future research.

5.2 Induced ordered weighting averaging operators

Another weighting method could be developed using an induced ordered weighting averaging function (IOWA). An IOWA is an averaging function that includes a reordering of the inputs before assigning weights [10]. This method uses predefined weights which are assigned to experts according to their relative similarity degrees. So in each case the expert with the greatest agreement to the group has the biggest impact on the collective preference relation, while an expert who is furthest from the group opinion has little to no impact on the CPR. In Table 2 we show the effects of two different weighting vectors on the consensus levels calculated for each expert in Example 2. As this method uses predefined weights, we can ensure that we are only allowing those experts with preferences closest to the group to impact the collective preference relation. However this also presents problems when experts work together in an effort to bias a particular alternative. This can be explored in greater depth in future research.

<table>
<thead>
<tr>
<th>Weight vector</th>
<th>Experts 1-3 Consensus Level</th>
<th>Expert 4 Consensus Level</th>
</tr>
</thead>
<tbody>
<tr>
<td>(0.4,0.4,0.10,0.1)</td>
<td>0.9594</td>
<td>0.8219</td>
</tr>
<tr>
<td>(0.6,0.3,0.1,0)</td>
<td>0.9663</td>
<td>0.8114</td>
</tr>
</tbody>
</table>

6 Conclusion

In this report we have shown that current consensus models are susceptible to attacks by biased experts to the extent that one extreme preference by one expert is able to affect the overall ranking order of the alternatives. We showed that the small range of weights that can be assigned to experts in this model allows experts that provide preferences that are significantly different to the group to still have a noticeable impact the CPR and consequently all further levels of the GDM process. We also showed that the use of the arithmetic mean in the calculation of the similarity degrees provides a buffer for extreme preferences, allowing an extreme preference to go unidentified due to the large distance being spread over four preferences. We have then provided several methods that have potential to increase the robustness of the consensus process. These will provide the direction for further research for this problem.

References


