Biased experts and similarity based weights in preferences aggregation

Gleb Beliakov, Simon James, Laura Smith, Tim Wilkin

School of Information Technology, Deakin University

Abstract

In a group decision making setting, we consider the potential impact an expert can have on the overall ranking by providing a biased assessment of the alternatives that differs substantially from the majority opinion. In the framework of similarity based averaging functions, we show that some alternative approaches to weighting the experts' inputs during the aggregation process can minimize the influence the biased expert is able to exert.

Keywords: aggregation functions, non-monotonic averaging, consensus, pairwise preferences, group decision making, induced OWA

1. Introduction

Fundamental to the group decision making problem is the aggregation of expert preferences into an overall preference relation that can be used as a basis for ranking the alternatives. While the arithmetic mean represents the standard operator of choice, the focus on consensus guided aggregation in recently proposed systems has called for the consistency, expertise and agreement between experts to be incorporated into the decision making process. In the field of soft, or fuzzy computing, the term *consensual processes* usually refers to multi-stage models where expert preferences are iteratively revised based on proximity to the overall opinion until a final decision can be reached (see, for example [1, 2, 3, 4, 5]).

Underlying the use of the arithmetic mean (or its weighted versions) in consensual processes are a number of assumptions, including those regarding the independence of input sources, the distribution and accuracy of assessments, and so on. A particularly significant assumption is that all assessments are provided in good faith. We need only look to community-based ratings online or news reports of corruption to note that this will not always be the case (see [6] for an example of some recent research in this area). If a decision system has the potential to be exploited, then chances are that at some time it will be.

We use the following example to illustrate this and motivate the investigations of this paper.

Example 1 *A government department evaluates tenders from 4 software companies $\{A,B,C,D\}$ to implement an automated fare system on a public transportation network. There are 4 experts in the government department who are each to express their preferences over the set of tender applications. Three of the experts are unanimous in their rankings, believing the order should be $A \succ B \succ C \succ D$. Expert 4, however has been bribed to give preference to company B. Knowing the reasonable ranking for the companies, expert 4 attempts to exploit the aggregation procedure by providing a ranking of $B \succ C \succ D \succ A$, hoping to bring down the ranking of company A enough so that B wins the tender.*

In a group decision making context, one method to try to prevent situations where expert 4 can influence the final result is to allocate importance to each of the experts based on how similar they are to the rest of the group. The rationale here is that experts with extreme scores (whether deliberate or due to lack of knowledge, input error etc.) should not have their votes counted as significantly as those with opinions closer to the majority. In the study of aggregation functions, this problem can be considered in the framework of outlier detection and noise reduction, with a number of functions available (such as the density-based averages in [7, 8]) which can limit how much an extreme value (or set of extreme values) can influence the overall evaluation.

In this contribution we pay attention to consensus guided approaches to weights allocation and investigate how susceptible some existing approaches are to biased manipulation such as the situation in Example 1. We focus on additively consistent pairwise preferences (which can be obtained from an underlying vector of scores) and the combined preference relation that would be used as a basis for making recommendations to the experts in the first consensus round. We find that if weights are determined from the entire preference relation, biased experts may be able to manipulate the consensus process. We therefore make mention of some alternative weight allocation methods and investigate whether altering the structure of the aggregation process could be useful in limiting the effect of outliers.

The paper will be set out as follows. In Section 2 we give an overview of the group decision making problem with fuzzy preference relations and the basic definitions of aggregation functions that will be necessary for the rest of the paper. In Section 3 we provide an example of how some existing consensus systems can be influenced by a corrupt expert and for specific assumptions provide theoretical bounds on the extent to which he or she can alter the results. In Section 4, we consider the problem of how to weight the experts in a way that discourages extreme preferences but still takes into account any 'reasonable' preferences provided by the experts. In Section 5 we explore whether aggregating the preference pairs

individually can alleviate the problems associated with allocating importance based on an overall measure of similarity. We note the ramifications that this may have on the consistency of the combined preference relation. Finally, in Section 6 we conclude and discuss further directions for research.

2. Preliminaries

In this section we will present the concepts, definitions and notation relevant to the rest of the paper.

2.1. Group decision making

We consider a group of m experts $V = v_1, v_2, \ldots, v_m$ expressing their preferences over a set of n alternatives $U = u_1, u_2, \ldots, u_n$. An expert expresses his or her preference for alternative i over j with the value $p_{ij} \in [0,1]$ with $p_{ij} > 0.5$ indicating that u_i is preferred to u_j. A value of 0.5 is interpreted as the two alternatives being considered equal by the expert, while a value less than 0.5 indicates that u_j is preferred to u_i.

Although in practice an expert's thinking can sometimes be contradictory or inconsistent, we can impose a number of constraints on the set of preferences that ensure logical behavior. Since a preference relation is assumed to infer an ordering over the set of alternatives, these constraints correspond with the properties of *reflexivity*, i.e. that $p_{ii} = 0.5$; *asymmetry*, that if $p_{ij} > 0.5$ then $p_{ji} < 0.5$ and vice-versa; and *transitivity*, i.e. that if $p_{ij} > 0.5$ and $p_{jk} > 0.5$ then $p_{ik} > 0.5$. More strictly, we have the notion of *additively consistent* preference relations, which require $p_{ik} = p_{ij} + p_{jk} - 0.5$. There are a number of existing methods for converting an expert's supplied preferences into an additively consistent preference relation and we also note that there exists the notion of *multiplicative transitivity* or *multiplicative consistency* [9]. However herein we will assume preference relations that are additively consistent.

Given a set of raw scores $x_1, x_2, \ldots, x_n \in [0,1]$ provided by each expert over the set of alternatives, we can obtain an additively consistent preference relation using

$$p_{ij} = \frac{x_i - x_j + 1}{2}. \tag{1}$$

2.2. Aggregation functions

The usual approach in group decision making is to combine the preference relations of multiple experts into an overall preference relation. For this, we use aggregation functions [10, 11, 12].

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

In particular, we are usually interested in averaging functions, i.e. those bounded by their inputs such that for a set of inputs $\mathbf{x} \in [0,1]^n$, $\min(\mathbf{x}) \leq A(\mathbf{x}) \leq \max(\mathbf{x})$.

The standard averaging aggregation function adopted in most practical contexts is the weighted arithmetic mean (WAM), which is defined with respect to a weighting vector, $\mathbf{w}$.

Definition 2 *A vector $\mathbf{w} = (w_1, \ldots, w_n)$ is called a weighting vector if $w_i \in [0,1]$ and $\sum_{i=1}^{n} w_i = 1$.*

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

$$WAM_{\mathbf{w}}(x_1, \ldots, x_n) = \sum_{i=1}^{n} w_i x_i. \tag{2}$$

The weight w_i is indicative of the importance of the i-th contributing input. In the case of group decision making, this could be the importance or reliability of an expert, or the proportional representation of a particular profile. For example, if we have 37% of shareholders supporting a particular opinion, then a weight of 0.37 can be allocated to that opinion.

On the other hand, another important class of weighted averaging functions used in soft-computing and decision making are the ordered weighted averaging (OWA) operators [13]. The OWA weights are assigned based on the magnitude of the inputs.

Definition 4 *Given a weighting vector $\mathbf{w}$, the Ordered Weighted Averaging (OWA) function is*

$$OWA_{\mathbf{w}}(x_1, \ldots, x_n) = \sum_{i=1}^{n} w_i x_{\sigma(i)}, \tag{3}$$

where the $\sigma(.)$ notation denotes the components of $\mathbf{x}$ being arranged in non-increasing order $x_{\sigma(1)} \geq x_{\sigma(2)} \geq \ldots \geq x_{\sigma(n)}$.

The OWA is capable of expressing a number of order statistics such as the maximum function, with $\mathbf{w} = (1,0,\ldots,0)$, and the minimum with $\mathbf{w} = (0,\ldots,0,1)$. It is also convenient for expressing the median, $\mathbf{w}_k = 1$, for $n = 2k+1$ (n is odd) or $w_k = w_{k+1} = 0.5$ for $n = 2k$ (n is even) and $w_i = 0$ otherwise.

A further development of the OWA operator considered re-ordering weights based on an auxiliary variable [14, 15]. The definition of the induced OWA (IOWA) is presented here as given by Yager and Filev in [14], in particular, their convention for ties is used.

Definition 5 *Given a weighting vector $\mathbf{w}$ and an inducing variable $\mathbf{z}$ the Induced Ordered Weighted Averaging (IOWA) function is*

$$IOWA_{\mathbf{w}}(\langle x_1, z_1 \rangle, \ldots, \langle x_n, z_n \rangle) = \sum_{i=1}^{n} w_i x_{\eta(i)}, \tag{4}$$

where the $\eta(.)$ notation denotes the inputs $\langle x_i, z_i \rangle$ reordered such that $z_{\eta(1)} \geq z_{\eta(2)} \geq \ldots \geq z_{\eta(n)}$ and the convention that if q of the $z_{\eta(i)}$ are tied, i.e. $z_{\eta(i)} = z_{\eta(i+1)} = \ldots = z_{\eta(i+q-1)}$,

$$x_{\eta(i)} = \frac{1}{q} \sum_{j=\eta(i)}^{\eta(i+q-1)} x_j. \tag{5}$$

An inducing variable can be based on any notion that associates a variable with each input x_i. Where x_i provides information to be aggregated, z_i provides some information about x_i, e.g. the importance, distance from the source, time displacement of the reading etc. The input pairs $\langle x_i, z_i \rangle$ may be two independent features of the same input, or can be related by some function, i.e. $z_i = f_i(x_i)$.

As many authors have shown, the IOWA provides a useful framework for modeling an aggregation based on consensus or agreement between the inputs. Examples include the consensus-based operator of Pasi and Yager [16]. Approaches to defining and characterizing weights associated with OWA operators have recently been discussed in [17].

3. Assigning weights by similarity in consensual processes

A typical framework for a consensual process is to first obtain a set of pairwise preference relations over a set of options or alternatives from each expert (which may involve pre-processing or some iterative process to ensure consistency). Following this, the preference matrices are aggregated into a combined preference relation (CPR) and the level of consensus amongst the experts is established. From this information, suggestions are made to the experts to revise their preferences so that they are closer, overall, to the CPR and their new preferences are then used to construct an updated CPR and so on until a satisfactory level of consensus is achieved. Recommendations could be in the form of suggested preference values, e.g. "consensus will increase if you assign the preference $p_{34} = 0.57$"; or statements identifying large disparities with the rest of the group, e.g. "you gave the least preference to alternative A, however this was the most preferred alternative by the majority of experts".

Given the strong dependence of revised individual preference relations on the initial CPR, herein we focus on the impact an expert may have during the first iteration, from which the first round of suggestions are made and an initial level of consensus is determined.

3.1. Summary of consensual process

A typical consensus process using the similarity between experts to guide the weights allocation is described in [18]. Similarity is calculated in three stages: the similarity on a pair of alternatives, the similarity on an alternative, and the similarity of the preference relation of an expert to the rest of the group. For the k-th expert, these are given respectively by[1],

$$sim_{ij}^k = \frac{1}{m-1} \sum_{h=1, h \neq k}^{m} 1 - |p_{ij}^k - p_{ij}^h|, \qquad (6)$$

$$sim_i^k = \frac{1}{n-1} \sum_{j=1, j \neq i}^{n} sim_{ij}^k, \qquad (7)$$

$$sim^k = \frac{1}{n} \sum_{i=1}^{n} sim_i^k. \qquad (8)$$

Overall, the similarity is essentially the average difference between each of expert k's preferences p_{ij}^k and those supplied by the rest of the group. It can be calculated succinctly using the triple sum,

$$sim^k = \frac{1}{n(n-1)(m-1)} \sum_{i=1}^{n} \sum_{j=1, j \neq i}^{n} \sum_{h=1, h \neq k}^{m} 1 - |p_{ij}^k - p_{ij}^h|. \qquad (9)$$

These similarity degrees are used to calculate the relative weight w_k of an expert using,

$$w_k = \frac{sim^k}{\sum\limits_{h=1}^{m} sim^h}. \qquad (10)$$

These weights are incorporated into the aggregation of the individual fuzzy preference relations P^k resulting in a collective fuzzy preference relation given by,

$$CPR = \sum_{l=1}^{m} w_k \cdot P^k, \qquad (11)$$

where '$\cdot$' denotes the standard multiplication of a matrix by a scalar.

The final stages of this first iteration are to calculate the proximity degree of each expert to the CPR by taking the average of $1 - |p_{ij}^k - p_{ij}^{CPR}|$ for all entries in the preference matrix, and then to merge together the proximity degrees and similarity degrees to give an overall consensus between each expert and the group. If an expert has a consensus level lower than the threshold value (often around 0.8), recommendations will be given to update the preferences that are most different from the collective fuzzy preference relation.

The rationale for using similarity to guide the weight process is clear: we want to downgrade the weights of any expert who differs from the majority, while those who have supplied preferences similar to the rest of the group should be given a higher weight. This weighting is similar to that used in nearest-neighbor type aggregation.

3.2. Numerical example

A potential problem, however is that since similarity is calculated across the entire preference matrix, the weight allocated to an 'extreme' expert may not be sufficiently low to minimize the impact on the CPR that is obtained in the first iteration, and furthermore identify when an expert is providing 'unreasonable' scores and should be required to think more carefully when updating their preferences. We can illustrate this idea with the following example.

Example 2 *(continued from Example 1) The experts 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* $(0, 0.8, 0.5, 0.2)$*, ensuring a reasonably high similarity value due to companies B to D receiving the same scores. The two corresponding preference relations are as follows*

$$P^1 = P^2 = P^3 = \begin{bmatrix} 0.5 & 0.6 & 0.75 & 0.9 \\ 0.4 & 0.5 & 0.65 & 0.8 \\ 0.25 & 0.35 & 0.5 & 0.65 \\ 0.1 & 0.2 & 0.35 & 0.5 \end{bmatrix},$$

$$\text{and } P^4 = \begin{bmatrix} 0.5 & 0.1 & 0.25 & 0.4 \\ 0.9 & 0.5 & 0.65 & 0.8 \\ 0.75 & 0.35 & 0.5 & 0.65 \\ 0.6 & 0.2 & 0.35 & 0.5 \end{bmatrix}.$$

Using Equations (9) and (10), the weight allocated to experts 1-3 would be 0.2619 while the weight for expert 4 would be 0.2143. This results in the following aggregated matrix.

$$CPR = \begin{bmatrix} 0.5 & 0.4929 & 0.6429 & 0.7929 \\ 0.5071 & 0.5 & 0.65 & 0.8 \\ 0.3571 & 0.35 & 0.5 & 0.65 \\ 0.2071 & 0.2 & 0.35 & 0.5 \end{bmatrix},$$

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

Furthermore, the resulting proximity level for the expert would be 0.8036.

In other words, by providing a biased opinion for one alternative, expert 4 has been able to change the ranking (despite being outnumbered 3 to 1). If the experts are asked to adjust their scores toward increasing the level of consensus, each of the experts 1-3 will receive the feedback that consensus can be increased if they allocate less preference to company A.

We note that a single example does not negate the usefulness of the consensus method. Indeed, expert 4 still has the lowest overall consensus to the group. However this example highlights two stages of this process at which the monotonic behavior of the arithmetic mean can be exploited. The first is that one extreme preference has an equivalent influence on the similarity degree calculation as would be exerted by a number of small differences in preferences. The biased expert can provide reasonable preferences for companies B, C, D in order to hide the effect of providing an extremely low value for company A. Conversely, when the number of experts is small, the extreme preference will have a significant impact on the score, in this case bringing down the ranking of alternative 1 to second place[2].

To gain a further insight into this characteristic of the proposed consensus process, we ran experiments where

we randomly generated 4 sets of additively consistent pairwise preference relations[3]. We calculated the relative weight that would be assigned to each expert, and recorded the maximum and minimum weight each time. We ran this experiment 1000 times using the R software package [19]. 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. It is clear that in general, the influence of any expert is very likely to remain quite significant when generating the initial CPR. We note that this also has an impact on the resulting proximity allocated to this expert, since their preference relation made a relatively large contribution to generating the CPR from which it is measured.

3.3. Analysis of similarity calculations in a 4 expert, 4 option model

Here we investigate the conditions under which a biased expert could affect the overall ranking in the scenario provided in Examples 1-2. Clearly if experts 1-3 all expressed their preference for company A over company B as $p_{ij} = 1$, which is equivalent to providing raw scores $x_1 = 1$ and $x_2 = 0$, then expert 4 will not be influential enough to change the ranking between these two alternatives. We consider potential strategies that could be used by expert 4:

1. provide the same raw scores to each of the companies B, C and D to maximize similarity to the rest of the group and a score of 0 to company A in order to downgrade its overall preference;
2. provide the same raw scores to companies C and D, then allocate 0 to company A and 1 to company B in order to maximize the preference for B over A.

In the following we will use the indices $1 - 4$ corresponding to options $A - D$ accordingly, so the preference p_{12} indicates the allocated preference for option A over option B.

3.3.1. All experts have the same preferences over B, C, D

In this case, the preferences involving $i, j \in \{2, 3, 4\}$ will all be equal and we need only focus on differences in the preferences p_{12}, p_{13}, p_{14} and their symmetric cases. Furthermore, for sim^k when $k = 1, 2, 3$, we need only take account of differences to expert 4. We will use k to denote the three experts in the consensus group and l to refer to the biased expert. Using Equation (9) we have $n(n-1)(m-1) = 36$ and hence the similarity level simplifies to,

$$sim^k = 1 - \frac{1}{18} \left(|p_{12}^k - p_{12}^l| + |p_{13}^k - p_{13}^l| + |p_{14}^k - p_{14}^l| \right).$$

[2]On the other hand, we have an equivalent problem for larger pools of experts if there is a coalition comprising 25% of the total population that wishes to unfairly manipulate the aggregation process.

[3]We first created vectors of raw scores and then converted these to fuzzy pairwise preference relations using Equation (1).

For the final expert, we will need to multiply the differences by 3, so we have

$$sim^l = 1 - \frac{1}{6}\left(|p^k_{12} - p^l_{12}| + |p^k_{13} - p^l_{13}| + |p^k_{14} - p^l_{14}|\right).$$

We can assume that in each case $p^k_{ij} > p^l_{ij}$ since expert l is trying to bring down the ranking of option 1. We can represent these absolute differences in terms of the raw scores using Equation (1), giving

$$|p^k_{12} - p^l_{12}| + |p^k_{13} - p^l_{13}| + |p^k_{14} - p^l_{14}|$$

$$= \frac{x^k_1 - x^l_1 + x^l_2 - x^k_2}{2} + \frac{x^k_1 - x^l_1 + x^l_3 - x^k_3}{2} + \frac{x^k_1 - x^l_1 + x^l_4 - x^k_4}{2}$$

$$= \frac{3x^k_1 - 3x^l_1 + x^l_2 - x^k_2 + x^l_3 - x^k_3 + x^l_4 - x^k_4}{2}.$$

Since we assume that the raw scores (and resulting degrees of preference) for x_2, x_3, x_4 are the same, we therefore simply have

$$\frac{3x^k_1 - 3x^l_1}{2}.$$

So

$$sim^k = 1 - \frac{3(x^k_1 - x^l_1)}{36} = 1 - \frac{x^k_1 - x^l_1}{12},$$

and

$$sim^l = 1 - \frac{9(x^k_1 - x^l_1)}{36} = 1 - \frac{x^k_1 - x^l_1}{4}.$$

Focusing on the preference for option 1 over option 2, we have

$$p^{CPR}_{12} = 3w^k p^k_{12} + w^l p^l_{12}$$

$$= \frac{3 \cdot sim^k}{3 \cdot sim^k + sim^l} p^k_{12} + \frac{sim^l}{3 \cdot sim^k + sim^l} p^l_{12}$$

$$= \frac{3 \cdot sim^k p^k_{12} + sim^l p^l_{12}}{3 \cdot sim^k + sim^l},$$

which in terms of x^k_1, x^l_1 will be

$$p^{CPR}_{12} = \frac{3(1 - \frac{x^k_1 - x^l_1}{12})p^k_{12} + (1 - \frac{x^k_1 - x^l_1}{4})p^l_{12}}{3(1 - \frac{x^k_1 - x^l_1}{12}) + (1 - \frac{x^k_1 - x^l_1}{4})}.$$

The biased expert can exert the largest influence on the overall degree of preference using $x^l_1 = 0$. To find the threshold at which preference can be given to option 2, we set $x^k_1 = 1$ and $p^{CPR}_{12} = 0.5$ then solve for x^k_2. We omit the steps for brevity, noting that the solution yields $x^k_2 = \frac{11}{14} \approx 0.7858$.

Experts 1-3 would hence have to provide a raw score for company B of 0.7858 or lower to ensure that the biased expert could not affect the overall ranking. This is equivalent to a preference[4] of 0.6071 or higher for option 1 over option 2.

---

[4]We note that a translation of the raw scores, i.e. $x^k_1 = 0.8, x^k_2 = 0.5858$ would lead to the same preferences and the same overall result.

3.3.2. The biased expert also increases the score given to option 2 (as well as decreasing option 1)

In this case, we will have differences in $p_{12}, p_{13}, p_{14}, p_{23}, p_{24}$ and the symmetric cases. We further assume that $p^k_{23} \leq p^l_{23}$ and $p^k_{24} \leq p^l_{24}$. Taking this into account, similar to the previous case, Equation (9) will reduce to

$$1 - \frac{x^k_1 - x^l_1 + x^l_2 - x^k_2}{12}.$$

This leads to an overall preference expressed as

$$p^{CPR}_{12} = \frac{3(1 - \frac{x^k_1 - x^l_1 + x^l_2 - x^k_2}{12})p^k_{12} + (1 - \frac{x^k_1 - x^l_1 + x^l_2 - x^k_2}{4})p^l_{12}}{3(1 - \frac{x^k_1 - x^l_1 + x^l_2 - x^k_2}{12}) + (1 - \frac{x^k_1 - x^l_1 + x^l_2 - x^k_2}{4})}.$$

To test the threshold at which the fourth expert can affect the decision, we assume $x^k_1 = 1, x^l_1 = 0$, and lastly that $x^l_2 = 1$.

Solving for $p^{CPR}_{12} = 0.5$ and x^k_2 in this case leads to the quadratic,

$$0 = -(x^k_2)^2 - 10x^k_2 + 12,$$

which has real solutions

$$\frac{10 \pm \sqrt{132}}{-2}.$$

The obtained value is approximately 0.7446. This means that if the three experts give a score of 1 to option 1 and option 2 has a score below 0.7446 (or equivalently that the preference for option 1 over 2 is above 0.6277), then the 4th expert will not be able to affect the order, however if the results are closer than this, then it will be possible to supply preferences that can affect the ranking.

A question we might ask is whether three experts providing a preference of $p_{ij} = 0.6$ should be significant enough to result in an overall preference of $p^{CPR}_{ij} > 0.5$. In the following two sections, we propose some alternative approaches to weighting the experts that attempt to minimize the influence of an unreasonable and extreme score provided by a single expert.

4. Alternative ways to weight the experts

In this section we investigate two approaches to weighting the experts' preferences in order to avoid situations where unreasonable scores can upset the ranking, however which still utilize the similarity calculations. In the first case, we consider predefining the weights to the experts according to their degree of similarity to the rest of the group, while in the second case we consider transformation functions of the similarity evaluations.

4.1. IOWA operator with predefined weights

We consider aggregation of the experts' preference matrices using the IOWA operator, i.e.

$$CPR = \sum_{k=1}^{n} w_k P^{\eta(k)}, \qquad (12)$$

where $\eta(k)$ denotes the expert with the k-th highest similarity rating. In other words, we use the auxiliary variable $z_k = sim^k$ so that the expert with the highest overall similarity to the group is associated with w_1, the second highest with w_2 and so on.

Assuming that expert 4 provides an evaluation of 0 for the preference p_{12}, the aggregated value will hence be

$$p_{ij}^{CPR} = (1 - w_4)p_{12}^k.$$

So if $w_4 = 0.1$ then any preference above $0.5/0.9 \approx 0.5556$ given by experts 1 to 3 (or a difference of $x_1 - x_2 = 0.1111$ in the raw scores) would be enough to ensure that the last expert had no influence. The weighting vector $\mathbf{w} = (1/3, 1/3, 1/3, 0)$ would mean that expert 4 in Example 1 would have no influence on the initial *CPR* at all, while other alternatives that could enforce this would be decreasing vectors such as $\mathbf{w} = (0.6, 0.3, 0.1, 0)$.

There is still the potential problem that two experts could collude and exert an unfair influence on the *CPR*. In fact, allocating too high a weight to any expert still leaves the consensual process vulnerable to manipulation. Consider the following example.

Example 3 *A consensual process utilizes the preference relation of the expert with the highest degree of consensus as its initial combined preference relation. In evaluating the same 4 software companies from Example 1, the following information is provided by experts 1-3,*

	1	2	3
A	1	1	1
B	0.6	0.9	0.7
C	0.5	0.3	0.6
D	0	0.2	0.4

Even though all experts provide a preference ordering of $A \succ B \succ C \succ D$, expert 4 is still able to exploit the differences in their pairwise preferences by providing a median score for companies C and D then keeping A and B close to the existing evaluations however with $B \succ A$. If the expert provides the vector $(0.9, 1, 0.5, 0.2)$, the overall similarity will be 0.8833 while experts 1-3 will have a similarity of 0.8722 and so expert 4's preferences will be used as the representative CPR and company B will be leading the tender application in the first round.

While we wish to minimize the effect of outliers, it may still be desirable that everyone contributes to the final evaluation, i.e. we discount outliers in the average rather then omit them from the data[5]. In the following we consider using transformations of the sim^k evaluations.

4.2. Transforming the similarity functions rather than using raw scores

Here our aim is to utilize the similarity calculations but to ensure that the weights are more sensitive to large differences amongst the experts. In this case, as an expert's

---

[5]This is a preferred approach since we may not easily know which data are outliers and which are not.

position becomes more extreme, their contribution becomes less significant at an increasing rate. We consider the following three transformations:

$$\psi_1(sim^k) = (sim^k)^2, \quad (13)$$

$$\psi_2(sim^k) = (sim^k)^5, \quad (14)$$

$$\psi_3(sim^k) = \frac{3}{2}(sim^k - 1/3). \quad (15)$$

The rational behind ψ_1 and ψ_2 is to make the weight disparity higher as the similarity levels decrease, while ψ_3 scales the similarity calculations to the unit interval, since the absolute minimum similarity between two consistent experts is $1/3$.

Table 1 gives an idea of how substituting these similarity equations can affect the initial aggregation of the consensus process. We focus on the aggregated preference for company A over company B (i.e., p_{12}^{CPR}) and the weight that is allocated to the biased expert (w_4).

Table 1: Aggregated values of p_{12}^{CPR} and allocated weight w_4 where 3 experts express their preferences as in Example 2.

	p_{12}^{CPR}		w_4	
p_{12}^4	0.1	0	0.1	0
sim^k	0.4929	0.4765	0.2143	0.2059
$(sim^k)^2$	0.5088	0.4993	0.1824	0.1678
$(sim^k)^5$	0.5455	0.5480	0.1089	0.0867
$\frac{3}{2}(sim^k - 1/3)$	0.5038	0.4935	0.1923	0.1774

In all cases, if expert 4 provides preferences that correspond with the raw scores $\mathbf{x} = (0, 0.8, 0.5, 0.2)$, i.e. trying to maximize the similarity score by providing identical scores for companies B, C, D, the overall preference still has company A as preferred to B. However for ψ_1 and ψ_3, the expert can still push the preference higher by providing a higher score for B, even though it decreases the overall similarity. On the other hand, ψ_2 has a desirable effect of decreasing the weight allocated to expert 4 to the extent that the decrease to the preference p_{12}^4 actually results in a small increase to the overall preference p_{12}^{CPR}.

5. Aggregating across individual cells

The calculation of similarity by using all pairwise preferences across the entire preference relation may miss the identification of a single or a few small extreme values. While using ψ_2 provided the desired behavior in the given example, we note that an honest expert providing reasonable scores could be allocated a small weight even though their scores are not extreme. For example, suppose expert 4's preferences correspond with the raw scores $\mathbf{x} = (0.8, 1, 0.3, 0.4)$. Admittedly, the expert's preferences differ to the rest of the group, however the differences could be consistent with a natural difference in opinion. The corresponding weight using ψ_2 would

be 0.1698 (compared with 0.2767 to the remaining experts), which could be considered unfairly low given that no individual score is extremely different. Here we consider the aggregated value of p_{ij}^{CPR} to instead reflect the consensus/majority opinion for that particular pair, aggregating across each of the preference relations.

We note that there will be ramifications for consistency, however if the purpose of this aggregated matrix is only to guide experts toward a more reasonable evaluation, we can still use an overall aggregation that preserves consistency in the final stages.

Some potential approaches to aggregating the preferences for each pair of alternatives are:

(a) Calculate similarity in an analogous way across the individual cells of each of the preference matrices, i.e.

$$p_{ij}^{CPR} = \sum_{k=1}^{m} w_{ij}^k p_{ij}^k, \qquad (16)$$

where

$$w_{ij}^k = \frac{\sum_{h=1,h\neq k}^{n} 1 - |p_{ij}^k - p_{ij}^h|}{\sum_{l=1}^{n} \sum_{h=1,h\neq l}^{n} 1 - |p_{ij}^h - p_{ij}^l|}.$$

(b) Find a 'majority' cluster or central value and only include p_{ij}^k within α of the central tendency of the cluster[6]. For example with 4 experts we can use the median of the *closest* 3 experts as our central value and set α to 0.3,

$$p_{ij}^{CPR} = \frac{1}{Card(K)} \sum_{k\in K} p_{ij}^k, \qquad (17)$$

where $K = \{k : |p_{ij}^k - Med_{UL}| < 0.3\}$ and $Med_{UL} = p_{ij}^{\sigma(2)}$ if $|p_{ij}^{\sigma(1)} - p_{ij}^{\sigma(2)}| < |p_{ij}^{\sigma(3)} - p_{ij}^{\sigma(4)}|$ and $p_{ij}^{\sigma(3)}$ otherwise. We remind that the $\sigma(\cdot)$ notation refers to the arguments being arranged in non-increasing order.

(c) Use the IOWA to aggregate the scores where the auxiliary variable is equivalent to the calculated similarity between the scores. As mentioned in the previous section, typical choices of weights that represent a majority type aggregation are $\mathbf{w} = (1/3, 1/3, 1/3, 0)$ or $\mathbf{w} = (0.6, 0.3, 0.1, 0)$, however in this case the experts cannot use their similarity standing across the irrelevant preferences in order to have an extreme preference included.

In approach (a), assuming expert 4 allocates $p_{12}^l = 0$ the similarity between this and the pairs of the other experts will simply be p_{12}^k and we have,

$$p_{12}^{CPR} = \frac{3(1 - \frac{p_{12}^k}{3})}{3(1 - \frac{p_{12}^k}{3}) + 1 - p_{12}^k} p_{12}^k$$

$$= \frac{3 - p_{12}^k}{4 - 2p_{12}^k} p_{12}^k.$$

Again to find the threshold, we set $p_{12}^{CPR} = 0.5$, leading to the quadratic equation,

$$0 = -(p_{12}^k)^2 + 4p_{12}^k - 2$$

which has solutions

$$\frac{-4 \pm \sqrt{8}}{-2}.$$

This results in a score of 0.5858 or higher required by experts 1-3 or a score of 0.8284 or lower required to be allocated to company 2. The weight allocated to expert 4 in this case would be 0.1464.

We note that the form of this aggregation is similar to the density-based averages [7], however whereas the weights in this case are based on the relative similarities to other experts in the group, $sim_{ij}^k = \sum_{h=1}^{m} 1 - |p_{ij}^k - p_{ij}^h|$, the density-based weights would be

$$w_{ij}^k = \frac{K_C \left(\frac{1}{m-1} \sum_{h=1,h\neq k}^{m} |p_{ij}^k - p_{ij}^h| \right)}{\sum_{l=1}^{m} K_C \left(\frac{1}{m-1} \sum_{h=1,h\neq l}^{m} |p_{ij}^l - p_{ij}^h| \right)},$$

where K_C is the Cauchy kernel given by $K_C(t) = (1 + t)^{-1}$.

Use of the Cauchy kernel with the preference values means that the aggregation will always allocate a reasonably high weight to expert 4. For example, if experts 1-3 have the preference $p_{ij}^k = 0.6$ and expert 4 has $p_{ij}^4 = 0$, then we still have $w_4 = 0.2$ and the expert exerts a relatively high influence on the overall preference.

In approach (b), the threshold of $\alpha = 0.3$ means that expert 4's preference will only be counted if it differs by no more than 0.3 from the majority. At most the expert can bring down the score by a maximum of $0.25 \times 0.3 = 0.075$, so that if the other experts provided any preference of 0.575 or above to p_{ij}^k, then the preference cannot be affected. This threshold can be adjusted to strike a balance between including all scores provided by the experts as long as they are not significantly different from the others. A similar aim is achieved by the robust estimators of Huber [20]. Whereas the weighted arithmetic mean can be considered as the value that minimizes the sum of squares, Huber estimators set a threshold at which the penalty becomes linear, i.e. in our situation, as our overall value we choose y that minimizes

$$\sum_{k=1}^{m} Hub(|p_{ij}^k - y|),$$

where $Hub(t) = \begin{cases} \frac{1}{2}t^2, & t < \alpha, \\ \alpha(t - \frac{\alpha}{2}), & \text{otherwise.} \end{cases}$

This function has no closed-form solution, however it can be solved numerically. When experts 1-3 provide the preference $p_{ij}^k = 0.6$ and expert 4 has a preference of 0, using $\alpha = 0.3$ would lead to an aggregated value of $p_{ij}^{CPR} = 0.5$.

Using an IOWA in approach (c) will have similar ramifications as was the case in Section 4.2, however with similarity only being measured across the preference pairs at hand. The weight w_4 then directly controls the maximum effect that a single extreme score can have on the aggregation.

6. Conclusion and future work

Although consensual processes usually involve many rounds of mediation and the potential for experts to change their scores, it seems important that these experts are being asked that their scores converge toward an aggregated value that reflects a majority viewpoint rather than one which has been biased by a minority of the experts. In this paper, we have investigated the problem of experts providing an unreasonable or biased evaluation in providing their initial preferences. In a 4-expert setting, we have shown the extent to which a single expert can affect the ranking of alternatives, even when 3 out of the 4 are unanimous in their evaluations. It is clear that even with more decision makers, there will always be cases when one or more biased experts can exert undue influence on the initial combined preference relation. We have suggested a number of approaches that could be employed in consensual process to alleviate this potential problem. Of course, the best method depends on the structure of the consensus process and interpretations surrounding it: how preference scores are interpreted, how the consensus threshold is used, whether the strength of preference is more important than the ranking of alternatives and so on. The problem of potential manipulation or influential noise in aggregated scores, however is not one that is unique to group decision making, and so the study of similarity-guided aggregation functions has been a growing area with many practical applications. In future work, we will continue to look at the usefulness of these approaches in other scenarios, in particular towards a characterization of functions in terms of the extent to which collusion attacks or extreme values can influence the output.

References

[1] 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*, 32:394–402, 2002.

[2] E. Herrera-Viedma, S. Alonso, F. Chiclana, and F. Herrera. A consensus model for group decision making with incomplete fuzzy preference relations. *IEEE Transactions on Fuzzy Systems*, 15(5):863–877, 2007.

[3] Z. Xu and X. Cai. Group consensus algorithms based on preference relations. *Information Sciences*, 181:150–162, 2011.

[4] I. J. Pérez, F. J. Cabrerizo, S. Alonso, and E. Herrera-Viedma. A new consensus model for group decision making problems with non-homogeneous experts. *IEEE Transactions on Systems, Man, and Cybernetics: Systems*, 44(4):494 – 498, 2014.

[5] E. Herrera-Viedma, F. J. Cabrerizo, J. Kacprzyk, and W. Pedrycz. A review of soft consensus models in a fuzzy environment. *Information Fusion*, 17:4 – 13, 2014.

[6] M. Allahbakhsh and A. Ignjatovic. An iterative method for calculating robust rating scores. *IEEE Tansactions on Parallel and Distributed Systems*, 26:340–350, 2015.

[7] P. Angelov and R. Yager. Density-based averaging - a new operator for data fusion. *Information Sciences*, 222:163 – 174, 2013.

[8] G. Beliakov and T. Wilkin. On some properties of weighted averaging with variable weights. *Information Sciences*, 281:1–7, 2014.

[9] J. Wu and F. Chiclana. Multiplicative consistency of intuitionistic reciprocal preference relations and its application to missing values estimation and consensus building. *Knowledge-Based Systems*, 71:187 – 200, 2014.

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

[11] M. Grabisch, J.-L. Marichal, R. Mesiar, and E. Pap. *Aggregation Functions*. Cambridge University press, Cambridge, 2009.

[12] Y. Torra, V. Narukawa. *Modeling Decisions. Information Fusion and Aggregation Operators*. Springer, Berlin, Heidelberg, 2007.

[13] R. Yager. On ordered weighted averaging aggregation operators in multicriteria decision making. *IEEE Transactions on Systems, Man and Cybernetics*, 18:183 – 190, 1988.

[14] R. Yager and D. Filev. Induced ordered weighted averaging operators. *IEEE Transactions on Systems, Man and Cybernetics - Part B: Cybernetics*, 20:141 – 150, 1999.

[15] R. Yager. Induced aggregation operators. *Fuzzy Sets and Systems*, 137:59 – 69, 2003.

[16] G. Pasi and R. R. Yager. Modeling the concept of majority opinion in group decision making. *Information Sciences*, 176:390–414, 2006.

[17] R. Yager and N. Alajlan. On characterizing features of OWA aggregation operators. *Fuzzy Optimization and Decision Making*, 13:1 – 32, 2013.

[18] J. Wu and F. Chiclana. Visual information feedback mechanism and attitudinal prioritisation method for group decision making with triangular fuzzy complementary preference relations. *Information Sciences*, 279:716 – 734, 2014.

[19] R Development Core Team. R: A language and environment for statistical computing. *R Foundation for Statistical Computing*, 2011, http://www.R-project.org.

[20] P. Huber. *Robust Statistics*. John Wiley and Sons, Hoboken, NJ, 2009.