Information Fusion with the Power Average Operator

Ronald R. Yager

Abstract. The power average provides an aggregation operator that allows argument values to support each other in the aggregation process. The properties of this operator are described. We see this mixes some of the properties of the mode with mean. Some formulations for the support function used in the power average are described. We extend this facility of empowerment to a wider class of mean operators such as the OWA and generalized mean.

Keywords: information fusion, aggregation operator, averaging, data mining.

1 Introduction

Aggregating information using techniques such as the average is a task common in many information fusion processes. Here we provide a tool to aid and provide more versatility in this process. In this work we introduce the concept of the power average [1]. With the aid of the power average we are able to allow values being aggregate to support each other. The power average is provides a kind of empowerment as it allows groups of values close to each other to reinforce each other. This operator is particularly useful in group decision making [2].

2 Power Average

In the following we describe an aggregation type operator called the **P**ower **A**verage (P–A), this operator takes a collection of values and provides a single value [1]. We define this operator as follows:

$$\text{P-A}(a_1, ..., a_n) = \frac{\sum_{i=1}^{n} (1 + T(a_i))\, a_i}{\sum_{i=1}^{n} (1 + T(a_i))}$$

Ronald R. Yager
Machine Intelligence Institute
Iona College
New Rochelle, NY 10801

S. Greco et al. (Eds.): Preferences and Decisions, STUDFUZZ 257, pp. 397–414.
springerlink.com

where $T(a_i) = \sum_{\substack{j=1 \\ j \neq i}}^{n} \text{Sup}(a_i, a_j)$ and is denoted the support for a from b.

Typically we assume that Sup(a, b) satisfies the following three properties:

1. $\text{Sup}(a, b) \in [0, 1]$
2. $\text{Sup}(a, b) = \text{Sup}(b, a)$
3. $\text{Sup}(a, b) \geq \text{Sup}(x, y)$ if $|a - b| \leq |x - y|$

In condition three we see the more similar, closer, two values the more they support each other.

We shall find it convenient to denote $V_i = 1 + T(a_i)$ and $w_i = \frac{V_i}{\sum_{i=1}^{n} V_i}$. Here the w_i are a proper set of weights, $w_i \geq 0$ and $\Sigma_i w_i = 1$. Using this notation we have

$$\text{P-A}(a_1, ..., a_n) = \Sigma_i w_i a_i,$$

it is a weighted average of the a_i. However, this is a non-linear weighted average as the w_i depend upon the arguments.

Let us look at some properties of the power average aggregation operator. First we see that this operator provides a generalization of the simple average, if $\text{Sup}(a_i, a_j) = k$ for all a_i and a_j then $T(a_i) = k(n - 1)$ for all i and hence $\text{P-A}(a_1, ..., a_n) = \frac{1}{n} \Sigma_i a_i$. Thus when all the supports are the same the power average reduces to the simple average.

We see that the power average is commutative, it doesn't depend on the indexing of the arguments. Any permutation of the arguments has the same power average.

The fact that $\text{P-A}(a_1, ..., a_n) = \Sigma_i w_i a_i$ where $w_i \geq 0$ and $\Sigma_i w_i = 1$ implies that the operator is bounded, $\text{Min}[a_i] \leq \text{P-A}(a_1, a_2, ..., a_n) \leq \text{Max}_i[a_i]$. This in turn implies that it is idempotent, if $a_i = a$ for all i then $\text{P-A}(a_1, ..., a_n) = a$.

As a result of the fact that the w_i depend upon the arguments, one property typically associated with averaging operator that is not generally satisfied by the power average is monotonicity. We recall that monotonicity requires that if $a_i \geq b_i$ for all i then $\text{P-A}(a_1, ..., a_n) \geq \text{P-A}(b_1, ..., b_n)$. As the following example illustrates, the increase in one of the arguments can result in a decrease in the power average.

Example: Assume the support function Sup is such that

Sup(2, 4) = 0.5	Sup(2, 10) = 0.3	Sup(2, 11) = 0
	Sup(4, 10) = 0.4	Sup(4, 11) = 0

the required symmetry means S(a, b) = S(b, a) for these values.

Consider first P-A(2, 4, 10), in this case

$$T(2) = Sup(2, 4) + Sup(2, 10) = 0.8$$
$$T(4) = Sup(4, 2) + Sup(4, 10) = 0.9$$
$$T(10) = Sup(10, 2) + Sup(10, 4) = 0.7$$

and therefore $\text{P-A}(2, 4, 10) = \dfrac{(1 + 0.8)\,2 + (1 + 0.9)\,4 + (1 + 0.7)\,10}{(1 + 0.8) + (1 + 0.9) + (1 + 0.7)} =$ 5.22.

Consider now P-A(2, 4, 11), in this case

$$T(2) = Sup(2, 4) + Sup(2, 11) = 0.5$$
$$T(4) = Sup(4, 2) + Sup(4, 11) = 0.5$$
$$T(11) = Sup(11, 2) + Sup(11, 2) = 0$$

and therefore

$$\text{P-A}(2, 4, 11) = \frac{(1.5)(2) + (1.5)\,4 + (1)(1.1)}{1.5 + 1.5 + 1} = 5$$

Thus we see that P-A(2, 4, 10) > P(2, 4, 11).

As we shall subsequently see, this ability to display non-monotonic behavior provides one of the useful features of this operator that distinguishes it from the usual average. For example the behavior displayed in the example is a manifestation of the ability of this operator to discount outliers. For as we shall see in the subsequent discussion, as an argument moves away from the main body of arguments it will be accommodated, by having the average move in its direction, this will happen up to point then when it gets too far away it is discounted by having its effective weighting factor diminished.

To some degree this power average can be seen to have some of the characteristics of the mode operator. We recall that the mode of a collection of arguments is equal to the value that appears most in the argument. We note that the mode is bounded by the arguments and commutative, however as the following example illustrates it is not monotonic.

Example: Mode(1, 1, 3, 3, 3) = 3. Consider now Mode(1, 1, 4, 7, 8) = 1, here we increased all the threes and obtain a value less than the original.

As we shall subsequently see, while both the power average and mode in some sense are trying to find the most supported value, a fundamental difference exists between these operators. We note that in the case of the mode we are not aggregating, blending, the values we are counting how many of each, the mode must be one of the arguments. In the case of power average we are allowing blending of values.

It is interesting, however, to note a formal relationship between the mode and the power average. To understand this we introduce an operator we call a **P**ower **M**ode. In the case of the power mode we define a support function $Sup_m(a, b)$, indicating the support for a from b, such that

1) $Sup_m(a, b) \in [0, 1]$
2) $Sup_m(a, b) = Sup_m(b, a)$

3) $Sup_m(a, b) \geq Sup_m(x, y)$ if $|a - b| \leq |x - y|$
4). $Sup_m(a, a) = 1$.

We then calculate $\text{Vote}(i) = \sum_{j=1}^{n} Sup_m(a_i, a_j)$ and define

$$\text{Power Mode}(a_1, ..., a_n) = a_{i^*}$$

where i* is such that $\text{Vote}(i^*) = \text{Max}_i[\text{Vote}(i)]$, it is the argument with the largest vote.

If $Sup_m(a, b) = 0$ for $b \neq a$ ($Sup_m(a, a) = 1$ by definition) then we get the usual mode. Here we are allowing some support for a value by neighboring values). It is also interesting to note the close relationship to the mountain clustering method introduced by Yager and Filev [3] and particularly with the special case of mountain clustering called the subtractive method suggested by Chu [4]. Some connection also seems to exist between the power mode and the idea of fuzzy typical value introduced in [5].

3 Power Average with Binary Support Functions

In order to obtain some intuition for the power average aggregation operator we shall consider first a binary support function. Here we assume

$$Sup(a, b) = K \quad \text{if } |a - b| \leq d$$
$$Sup(a, b) = 0 \quad \text{if } |a - b| > d.$$

Thus two values support each if they are less than or equal d away, otherwise they supply no support. Here K is the value of support. In the following discussion we say a and b are neighbors if $|a - b| \leq d$. The set of points that are neighbors of x will be denoted N_x. We shall call a set of points such that all points are neighbors and no other points are neighbors to those points a cluster. We note if x and y are in the same cluster then the subset $\{x\} \cup N_x = \{y\} \cup N_y$ defines the cluster.

Let us first assume that we have two disjointed clusters of values $A = \{a_1, ..., a_{n_1}\}$ and $B = \{b_1, ..., b_{n_2}\}$. Here all points in A support each other but support none in B while the opposite holds for B. In this case for all i and j, $|a_i - a_j| \leq d$, $|b_i - b_j| \leq d$ and $|a_i - b_j| > d$. Here for each a_i in A, $T(a_i) = K(n_1 - 1)$ and for each b_j in B, $T(b_j) = K(n_2 - 1)$. From this we get $1 + T(a_i) = (1 - K) + n_1 K$ and $1 + T(b_j) = (1 - K) + n_2K$. Using this we have

$$\text{P-A}(a_1, ..., a_{n_1}, b_1, ..., b_{n_2}) = \frac{\sum_{i=1}^{n_1} ((1 - K) + n_1K)a_i + \sum_{j=1}^{n_2} ((1 - K) + n_2K))b_j}{n_1(1 - K + n_1K) + n_2(1 - K + n_2K)}$$

Letting $\bar{a} = \frac{1}{n_1} \sum_{i=1}^{n_1} a_i$ and $\bar{b} = \frac{1}{n_2} \sum_{j=1}^{n_2} b_j$ we have

$$PA(a_1, ..., a_{n_1}, b_1, ..., b_{n_2}) = \frac{((1 - K) + n_1K)n_1\bar{a} + ((1 - K) + n_2K))n_2\bar{b}}{n_1(1 - K + n_1K) + n_2(1 - K + n_2K)}$$

We get a weighted average of the cluster averages. If we let

$$w_a = \frac{(1 - K + n_1K)n_1}{n_1(1 - K + n_1) + n_2(1 - K + n_2K)} \text{ and } w_b = \frac{(1 - K + n_2K)n_2}{n_1(1 - K + n_1) + n_2(1 - K + n_2K)}$$

then $PA(a_1, ..., a_{n_1}, b_1, ..., b_{n_2}) = w_a\,\bar{a} + w_b\,\bar{b}$. We note $w_a + w_b = 1$ and $\frac{w_a}{w_b} = \frac{(1 - K + n_1K)\,n_1}{(1 - K + n_2K)\,n_2}$

We see that if $k = 1$, then $\frac{w_a}{w_b} = (\frac{n_1}{n_2})^2$, the weights proportional to the square of the number of elements in the clusters. Thus in this case $w_a = \frac{n_1^2}{n_1^2 + n_2^2}$ and $w_b = \frac{n_2^2}{n_1^2 + n_2^2}$. On the other hand if we allow no support, $K = 0$, then $\frac{w_a}{w_b} = \frac{n_1}{n_2}$, the weights are just proportional to the number of elements in each cluster. In this case $w_a = \frac{n_1}{n_1 + n_2}$ and $w_b = \frac{n_2}{n_1 + n_2}$. Thus we see as we move from $K = 0$ to $K = 1$ we move from being proportional to number of elements in each cluster to being proportional to the square of the number of elements in each cluster. We now begin to see the effect of this power average. If we allow support then elements that are close gain power. This becomes a reflection of the adage that there is power in sticking together. We also observe that if n_1K and $n_2K >> (1 - K)$, there are a large number of arguments, then again $\frac{w_a}{w_b} = (\frac{n_1}{n_2})^2$. Furthermore we note if $n_1 = n_2$ then we always have $\frac{w_a}{w_b} = 1$, here we take the simple average.

Consider now the case when we have q disjoint clusters, each only supporting elements in its neighborhood. Let a_{ji} for $i = 1$ to n_j be the elements in the j^{th} cluster. In this case

$$P\text{-}A = \frac{\sum_{j=1}^{q} \left(\sum_{i=1}^{n_j} (1 - K + n_jK)a_{ji} \right)}{\sum_{j=1}^{q} n_j(1 - K + n_jK)}$$

Letting $\frac{1}{n_j}\sum_{i=1}^{n_j} a_{ji} = \bar{a}_j$, the individual cluster averages, we can express this power average as

$$\text{P-A} = = \frac{\sum_{j=1}^{q} ((1 - K + n_j K)\, n_j\, \bar{a}_j)}{\sum_{j=1}^{q} (1 - K + n_j K)\, n_j}$$

Again we get a weighted average of the individual cluster averages, $\text{P-A} = \sum_{j=1}^{q} w_j\ \bar{a}_j$. In this case $w_i = \frac{(1 - K + n_i K)\, n_i}{\sum_{j=1}^{q} (1 - K + n_j K)\, n_j}$ and $\frac{w_i}{w_j} = \frac{(1 - K + n_i K)\, n_i}{(1 - K + n_j K)\, n_j}$.

Again we see if K= 1, then $\frac{w_i}{w_j} = \frac{n_i^2}{n_j^2}$, the proportionality factor is the square of the number of elements. Here then $w_i = \frac{n_i^2}{\sum_{j=1}^{q} n_j^2}$. If we allow no support, K = 0, then $\frac{w_i}{w_j} = \frac{n_j}{n_i}$, here we get the usual average. We note that K is the value of support.

Consider a case with small value of support, 1 - K ≈1. Furthermore assume n_i is a considerable number of elements while n_j is a very small number. Here |(1 - K) + n_j K ≈ 1 while (1 - K) + n_i K ≈ n_1K then $\frac{w_i}{w_j} = \frac{n_i^2 K}{(1 - K) n_j} \approx \frac{n_i^2 K}{n_j}$.

On the other hand if n_i and n_j are large, $n_i K$ and $n_j K >>> 1$ then $\frac{w_i}{w_j} = \frac{n_i^2}{n_j^2}$.

We that if $(1 - K) << n_j K$ for all j then $\text{P-A} = \frac{\sum_{j=1}^{q} n_j^2\, \bar{a}_j}{\sum_{j=1}^{q} n_j^2}$, the weights in proportion to the square of the number of elements.

Let us observe another interesting property of this P-A. To most clearly illustrate the property we shall assign K = 1. Assume we have two clusters then with K = 1 we have

$$\text{P-A} = \frac{n_1^2\,\bar{a}_1 + n_2^2\,\bar{a}_2}{n_1^2 + n_2^2}$$

If $n_1 \approx n_2 = \frac{1}{2}n$, they have the same number of elements then P-A = $\frac{1}{2}\bar{a}_1 + \frac{1}{2}\bar{a}_2$. Assume now that the second cluster is broken into two equal disjoint clusters. Then $P(A) = \frac{n_1^2\,\bar{a}_1 + n_2^2\,\bar{a}_2 + n_3^2\,\bar{a}_3}{n_1^2 + n_2^2 + n_3^2}$ with $n_1 = \frac{1}{2}$ n, $n_2 = \frac{1}{4}$ n and $n_3 = \frac{1}{4}$ n. From this we see that

$$P(A) = \frac{\frac{1}{4}\bar{a}_1 + \frac{1}{16}\bar{a}_2 + \frac{1}{16}\bar{a}_3}{\frac{1}{4} + \frac{1}{16} + \frac{1}{16}} = \frac{4\,\bar{a}_1 + \bar{a}_2 + \bar{a}_3}{6}$$

We see cluster one's influence (power) has greatly increased because of the fragmentation of cluster two..

We now consider a situation in which we have three sets of elements, $A = \{a_1, ..., a_{n_1}\}$, $B = \{b_1, ..., b_{n_2}\}$ and $C = \{c_1, ..., c_{n_3}\}$. We assume all the elements in A are a neighbors with each other as well as with those in B. Those in B are neighbors with each other and also with those in both A and C. The elements in C are neighbors with themselves and B. Thus B is seen to be between A and C. Here we see that for all a_i we have $T(a_i) = K(n_1 + n_2 - 1)$, for all b_i $T(b_i) = K(n_1 + n_2 + n_3 - 1)$ and for all $c_i T(c_i) = K(n_2 + n_3 - 1)$. Let $\bar{a} = \frac{1}{n_j}\Sigma\, a_j$, $\bar{b} = \frac{1}{n_2}\Sigma\, b_j$ and $\bar{c} = \frac{1}{n_3}\Sigma\, b_j$. Using this we have

$$\text{P-A} = \frac{(1 - K + K(n_1 + n_2))n_1\bar{a} + (1 - K + K(n_1 + n_2 + n_3)n_2\bar{a} + (1 - K + K(n_2 + n_3)n_3\bar{a}}{(1 - K + K(n_1 + n_2))n_1 + (1 - K + K(n_1 + n_2 + n_3)n_2 + (1 - K + K(n_2 + n_3)n_3}$$

Again for illustrative purposes we assume K =1 hence

$$\text{P-A} = \frac{(n_1 + n_2)n_1\bar{a} + n\,n_2\bar{b} + (n_2 + n_3)n_3\bar{c}}{(n_1 + n_2)n_1 + n\,n_2 + (n_2 + n_3)n_3}$$

$$\text{P-A} = \frac{(n - n_2)n_1\bar{a} + n\,n_2\bar{b} + (n - n_1)n_3\bar{c}}{n^2 - 2n_1n_3}$$

We see that relationship between the weights associated A and C is

$$\frac{w_a}{w_c} = \frac{(n - n_3)n_1}{(n - n_1)n_3} = \frac{(n_2 + n_1)n_1}{(n_2 + n_3)n_3}$$

If n_2 is large compared with both n_1 and n_3 then $\frac{w_a}{w_c} = \frac{n_1}{n_3}$, their relationship is proportion to the number of elements in A and C. If n_2 is small compared with both n_1 and n_3 then $\frac{w_a}{w_c} = \frac{n_1{}^2}{n_3{}^2}$. Consider the relationship between A and B, which is analogous to B and C, $\frac{w_a}{w_b} = \frac{n_1(n_1 + n_2)}{(n)(n_2)}$. If n_2 is large compared with n_1 and n_3 then $\frac{w_a}{w_b} \approx \frac{n_1\, n_2}{(n)(n_2)} \approx \frac{n_1}{n}$

We consider now another situation that exemplifies the possibility for non-monotonicity. Let $\{a_1, ..., a_n, a_{n+1}\}$ be a collection of points in the same cluster, for all a_i and a_j, $|a_i - a_j| \le d$. In this case $\text{P-A}\{a, ..., a_{n+1}\} = \frac{1}{n+1} \sum_{j=1}^{n+1} a_i = \bar{a}$.

Assume now that we replace a_{n+1} by $\hat{a}_{n+1}$ where $\hat{a}_{n+1} \ge a_{n+1}$ and $|a_{n+1} - a_j| > d$ for all other a_j. That is we have moved the n+1th observation all the way to the right. In this case we can view the situation having two disjoint clusters one being $\{a_1, ..., a_n\}$ and the other $\{\hat{a}_{n+1}\}$. As we already established the power average of this situation is

$$\text{P-A}(a_1, a_2, ..., a_n, \hat{a}_{n+1}\} = w_1\tilde{a} + w_2\hat{a}_{n+1}$$

here $\tilde{a} = \frac{1}{n} \sum_{i=1}^{n} a_i$ and $\hat{a}_{n+1} = a_{n+1} + \Delta$. We also note that $\bar{a} = \frac{1}{n+1} a_{n+1} + \frac{n}{n+1}\tilde{a}$ hence

$$\tilde{a} = \frac{(n + 1)\ \bar{a} - a_{n+1}}{n}$$

In the situation where K = 1 we have $\frac{w_1}{w_2} = \frac{n_1{}^2}{n_2{}^2} = \frac{n^2}{1}$. This gives us $w_1 = \frac{n^2}{n^2 + 1}$ and $w_1 = \frac{1}{n^2 + 1}$ and hence

$$\text{P-A}(a_1, ..., \hat{a}_{n+1}) = \frac{n^2}{n^2 + 1}\tilde{a} + \frac{1}{n^2 + 1}\hat{a}) = a + \frac{\Delta - (n - 1)\ (a_{n+1} - \bar{a})}{n^2 + 1}$$

Thus we see that if a_{n+1} was the right most element then we get a non-monotonicity as long as Δ is not too big.

4 Forms for the Support Function

The support function is a crucial part of the power average method. The form of the support function is context dependent. Here we describe some useful parameterized formulations for expressing the Sup function. The determination of the values of the parameters may require the use of some learning techniques. We recall if **R** is the range of the values to be aggregated then Sup:$\mathbf{R} \times \mathbf{R} \to [0, 1]$ such that Sup(a, b) = S(b, a), and Sup(a, b) ≥ Sup(x, y) if |a - b| ≤ |x - y|.

In the preceding we assumed a binary Sup function, Sup(a, b) = K if |a - b| ≤ d and Sup(a, b) = 0 if |a - b| > d. A natural extension of this is to consider a partitioned type support function. Let K_i for i = 1 to p be a collection of values such that $K_i \in [0, 1]$ and where $K_i > K_j$ if i < j. Let d_i be a collection of values such that $d_i \geq 0$ and where $d_i < d_j$ if i < j. We now can define a support function as

$$\begin{array}{ll} \text{If } |a - b| \leq d_1 \text{ then Sup}(a, b) = K_1 & \\ \text{If } d_{j-1} < |a - b| \leq d_j \text{ then Sup}(a, b) = K_j & \text{for } j = 2 \text{ to } p - 1 \\ \text{If } d_{p-1} < |a - b| \text{ then Sup}(a, b) = K_p & \end{array}$$

Inherent in the above type of support function is a discontinuity as we move between the different ranges.

One form of the Sup function with a continuous transition is Sup(a, b) = $K e^{-\alpha(a - b)^2}$ where $K \in [0,1]$ and $\alpha \geq 0$. We easily see that this function is symmetric and lies in the unit interval. We see K is the maximal allowable support and α is acting as a attenuator of the distance. The larger the α the more meaningful differences in distance. We note here that a = b gives us Sup(a, b) = K and as the distance between a and b gets larger, Sup(a, b) →0.

Using this form for support function we have

$$\text{P-A}(a_1, ..., a_n) = \frac{\sum_{i=1}^{n} (1 + T(a_i))a_i}{\sum_{i=1}^{n} (1 + T(a_i))}$$

where $T(a_i) = \sum_{\substack{j=1 \\ j \neq i}}^{n} K e^{-\alpha(a_i - a_j)^2}$. Denoting $V_i = 1 + T(a_i)$ we express

$\text{P-A}(a_1, ..., a_n) = \Sigma_i w_i a_i$ where $w_i = \dfrac{V_i}{\sum_{j=1}^{n} V_j}$. Since $e^{-\alpha(a_i - a_i)^2} = 1$ we can express $V_i = 1 - K + K M_i$ where $M_i = \sum_{j=1}^{n} e^{-\alpha(a_i - a_j)^2}$. Noting the similarly of M_i to the mountain function used in mountain clustering [3] we call M_i the support mountain at i. It's clear that if $a_p = a_q$ then $M_q = M_p$ and hence $V_q = V_p$. It is also noted that $M_i \geq 1$ for all i.

We see here that

$$\text{P-A}(a, ..., a_n) = \frac{\sum_{i=1}^{n} (1 - K)\, a_i + K \sum_{i=1}^{n} M_i a_i}{n(1 - K) + K \sum_{i=1}^{n} M_i}$$

In the special case where K = 1 then $V_i = M_i$ and hence

$$\text{P-A}(a_1, ..., a_n) = \frac{\sum_{i=1}^{n} M_i a_i}{\sum_{i=1}^{n} M_i}$$

A simple algorithm approach somewhat is in spirit of the mountain method is as follows:

1. For each argument value a_i, i = 1 to n, initialize $M_i = 0$
2. For each data point a_j j = 1 to n augment M_i, $M_i = M_i + e^{-\alpha(a_i - a_j)^2}$

This builds the support mountain.

3. Calculate $V_i = (1 - K) + K\, M_i$ - linear transformation of mountain values
4. Calculate $w_i = \dfrac{V_i}{\sum_{j=1}^{n} V_j}$
5. $\text{P-A} = \Sigma_i\, w_i\, a_i$

As we have noted an important characteristic of this power average is its possibility for displaying non-monotonicity, a feature that can provide one of the benefits of this method. The following example illustrates the occurrence of non-monotonicity.

Example: Consider the Power average of twenty elements, 10 of which are ten's and 10 of which are five's. In this case the ordinary average evaluates to 7.5 and for any choice of K and α the power average also evaluates to 7.5. The following table shows what happens as we change one of the values originally equal to 10. For illustrative purposes we used K = 1 and $\alpha = 0.3$

Value	**AVE**	**P-A**
10	7.5	7.5
9	7.45	7.398
8	7.4	7.278
7	7.35	7.193
6	7.3	7.083

5	7.25	6.982
11	7.55	7.4727
12	7.6	7.346
13	7.65	7.259
14	7.7	7.232
15	7.75	7.22856
16	7.8	7.22828
17	7.85	7.22829
18	7.9	7.22831
19	7.9	7.22832
20	8	7.22834

We see that as we decrease the value and move it towards the cluster of fives our P-A decrease, although more dramatically than the average. Essentially the variable value is beginning to join the cluster of fives and increase its power. In the case of increasing the value, initially the power average instead of increasing as does the average begins to decrease, exhibiting non–monotonicity. This decrease is a reflection of the fragmentation of the cluster at 10, it is losing its power because it lost a member and the cluster at five has gained in power more than compensating for the increase in value. This decreasing in the P-A continues as we increase the element until it reaches eighteen at which time we see a reversal and now the P-A starts increasing At this point the increase in value begins overcoming the loss of power. But still we are favoring the cluster of fives.

We describe another approach to obtaining the support function that combines the partitioning of the first method with the continuity displayed by the exponential function. This approach motivated by Zadeh's idea of computing with words [6] makes use of fuzzy systems modeling technology [7]. We shall briefly describe the possibilities for this approach. Using this approach we can express our support function by a description of its performance in terms of a set of rules using linguistic values. For example.

If difference is *very small* then support is K_1

If difference is *small* then support is K_2

If difference is *moderate* the support is K_3

If difference is *large* the support is K_4

If difference is *very large* the support is K_5

Representing the italic terms as fuzzy sets, VS, S, M, L, and VL respectively and denoting the difference between a and b as Δ than we have a collection of fuzzy if-then rules, a fuzzy systems model:

If Δ is VS then $S(a, b) = K_1$

If Δ is S then $S(a, b) = K_2$

If Δ is M then $S(a, b) = K_3$

If Δ is L then $S(a, b) = K_4$

If Δ is VL then $S(a, b) = K_4$

here $K_i < K_j$ if $i > j$.

To obtain the Sup(a, b) we use the inference mechanism of fuzzy systems modeling. Letting $\Delta = |a - b|$ then the analytic formulation of our support function is

$$\mathrm{Sup}(a, b) = \frac{K_1 VS(\Delta) + K_2 S(\Delta) + K_3 M(\Delta) + K_4 L(\Delta) + VL(\Delta)}{VS(\Delta) + S(\Delta) + M(\Delta) + L(\Delta) + VL(\Delta)}$$

here VS(Δ) indicates the membership of Δ in the fuzzy subset VS.

We now look at the power average in the special situation in which the arguments that are being aggregated, the a_i, always be in the unit interval [0, 1]. This is a situation that occurs in many environments when the arguments are degrees of belief. We note a particular important situation is in the aggregation of fuzzy subsets.

In the case when the arguments lie in the unit interval a very natural definition for the Sup function is

$$\mathrm{Sup}(a, b) = K(1 - |a - b|^{\alpha})$$

for $\alpha \geq 0$. Here we see that the term $|a - b|$ is a measure of distance between the arguments. We note since a and b are assumed to lie in the unit interval then $|a - b|$ must also lie in the unit interval as well as $|a - b|^{\alpha}$. We see $|a - b| \to 0$ indicates the elements are close and $|a - b| \to 1$ indicates the elements are far. We see that is Sup is related to the negation of the distance.

We notice that because a and b always lie in the unit interval, $|a - b| = 1$ if and only if one of the arguments equal zero and the other equals one. Furthermore we note that α modifies the effects of distance. Since $(a - b) < 1$ then $\alpha > 1$ reduces the effect of distance while $\alpha < 1$ increase the effects of distance. We note Sup(a, b) = K when a = b.

As in the preceding $\mathrm{P\text{-}A}(a_1, ..., a_n) = \dfrac{\sum_{i=1}^{n} V_i a_i}{\sum_{i=1}^{n} V_i}$. Let us consider the case when $\alpha = 2$, $\mathrm{Sup}(a, b) = K(1 - (a - b)^2)$. Here $V_i = 1 + T(a_i)$ with $T(a_i) = K \sum_{\substack{j=1 \\ i \neq j}}^{n} (1 - (a_i - a_j)^2)$. Realizing $1 - (a_i - a_i)^2 = 1$ then $V_i = (1 - K) + K \sum_{j=1}^{n} (1 - (a_i - a_j)^2)$. Letting $Q_i = \sum_{j=1}^{n} (a_i - a_j)^2$ we have

$$V_i = 1 - K + Kn - Kq_i$$

Let us carefully look at the term Q_i. We shall denote $\bar{a} = \frac{1}{n}\sum_{j=1}^{n} a_j$, it is the average, and denote $\text{Var}(a) = \frac{1}{n}\sum_{j=1}^{n} (a_j - \bar{a})^2$. Using these notations we can express

$$Q_i = \sum_{j=1}^{n} (a_i - a_j)^2 = \sum_{j=1}^{n} \left[(a_i - \bar{a}) - (a_j - \bar{a})\right]^2$$

$$Q_i = \sum_{j=1}^{n} (a_i - \bar{a})^2 + \sum_{j=1}^{n} (a_j - \bar{a})^2 - 2\sum_{j=1}^{n} (a_i - \bar{a})(a_j - \bar{a})$$

Realizing that $\sum_{j=1}^{n} (a_i - \bar{a})(a_j - \bar{a}) = (a_i - \bar{a})\sum_{j=1}^{n} (a_j - \bar{a}) = 0$ we have

$$Q_i = \sum_{j=1}^{n} (a_i - \bar{a})^2 + \sum_{j=1}^{n} (a_j - \bar{a})^2$$

Letting $\Delta_i = |a_i - \bar{a}|$. we have $Q_i = n\ \Delta_i^2 + n\ \text{Var}(a)$.

From this we have $V_i = (1 - K) + Kn - nK(\Delta_i^2 + \text{Var}(a))$. Using this we get that

$$\sum_{i=1}^{n} V_i = n(1 - K) + Kn^2 - n^2\ K\ \text{Var}(a) - n\ K \sum_{i=1}^{n} \Delta_i^2$$

Since $\frac{1}{n}\sum_{i-} \Delta_i^2 = \text{Var}(a)$ then $\sum_{i=1}^{n} V_i = n(1 - K) + Kn^2 - 2n^2K\text{Var}(a)$

Let us consider the special case where K = 1, here $V_i = n\ (1 - \text{Var}(a) - \Delta_i^2)$ and $\sum_{i=1}^{n} V_i = n^2\ (1 - 2\ \text{Var}(a))$. Using this

$$\text{P-A}(a_1, ..., a_n) = \frac{\sum_{i=1}^{n} V_i\ a_i}{n^2(1 - 2\text{Var}(a))} = \bar{a} + \frac{\bar{a}\sum_{i=1}^{n} \Delta_i^2 - \sum_{i=1}^{n} \Delta_i^2 a_i}{n(1 - 2\text{Var}(a))}$$

We see that if the arguments are such that there are a few large values far away from the the rest of the values mean then the power average tends to pull $\bar{a}$ downwards.

Another interesting case of $Sup(a, b) = K(1 - |a - b|^{\alpha})$ occurs when $\alpha = 1$, here $Sup(a\ b) = K(1 - |a - b|)$. We note that $|a - b| = Max(a, b) - Min(a, b) = (a \vee b) - (a \wedge b)$. Here again $P\text{-}A(a_1, ..., a_n) = \dfrac{\sum_{i=1}^{n} V_i a_i}{\sum_{i=1}^{n} V_i}$. In this case

$$V_i = 1 + (\nu - 1) K - K \sum_{j=1}^{n} [(a_i \vee a_j) - (a_j \wedge a_i)]$$

Without loss of generality let us assume that the a_i have been indexed in descending order, thus a_i is the i^{th} largest of the arguments. In this case

$a_i = Min[a_i, a_j]$ and $a_j = Max[a_i, a_j]$ for $j = 1$ to $i - 1$
$a_j = Min[a_i, a_j]$ and $a_i = Max[a_i, a_j]$ for $j = i + 1$ to n
$a_i = Min[a_i, a_j] = Max[a_i, a_j]$ for $j = 1$

If we denote $Q_i = \sum_{j=1}^{n} |a_i - a_j|$ then

$$Q_i = \sum_{j=1}^{n} (a_i \vee a_j) - (a_i \wedge a_j) = \sum_{j=1}^{i-1} a_j + \sum_{j=i+1}^{n} a_i - (\sum_{j=1}^{i-1} a_i + \sum_{j=i+1}^{n} a_j)$$

$$Q_i = \sum_{j=1}^{i-1} a_j - \sum_{j=i+1}^{n} a_j - (\sum_{j=1}^{i-1} a_i - \sum_{j=i+1}^{n} a_i) = \sum_{j=1}^{i-1} a_j - \sum_{j=i+1}^{n} a_j + (n - 2i)a_i$$

Denoting $SL(i) = \sum_{j=1}^{i} a_j$ and $SU(i) = \sum_{j=i+1}^{n} a_j$ then $Q_i = SL(i) - SU(i) + (n - 2i)\, a_i$ and

$$V_i = 1 + (n - 1) K - K\, (SL(i) - SU(i) + (n - 2i)\, a_i)$$

and $\sum_{i=1}^{n} V_i = n + n\,(n - 1) K - K \sum_{j=1}^{n} (SL(i) - SU(i) + (n - 2i)\, a_i)$.

Let us consider the special case where $K = 1$, hence

$$V_i = n - (SL(i) - SU(i) + (n - i)a_i)$$

$$\sum_{i=1}^{n} V_i = n^2 - \sum_{i=1}^{n} SL(i) - S(u)i + (n - 2i)a_i$$

Since a_i appears in n - i + 1 of the SL and in i of SU, $\sum_{i=1}^{n} SL(i) - SU(i) = \sum_{i=1}^{n} (n - 2i + 1)a_i$ then $\sum_{i=1}^{n} V_i = n^2 - \sum_{i=1}^{n} (2n - 4i + 1)\, a_i = n^2 - n(2n - 1)\bar{a} + 4\sum_{i=1}^{n} i\, a_i$

5 Empowering Alternative Mean Operators

The average operator, $\frac{1}{n}\sum_{i=1}^{n} a_i$, provides one example of mean type aggregation operators [8]. We recall that mean type operators are characterized by boundedness, commutativity and monotonicity. Other examples of mean type operators are the Max, Min, and Median. In the preceding with the power average we extended the average operator by introducing the idea of support. That is with the P-A operator we allowed arguments in the aggregation to support each. This effectively result is a weights associated with the different arguments depending upon the support they obtained from other elements being aggregated. In this section we want to generalize the idea of supported aggregation to a wider class of mean operators.

We first look at the OWA operator [9] and introduce the **Power-OWA** operator. An OWA operator can be defined in terms of function $g:[0, 1] \rightarrow [0, 1]$, called a BUM function, having the properties: **1**. $g(0) = 0$, **2**. $g(1) = 1$ and **3**. $g(x) \geq g(y)$ if $x > y$. Using this BUM function the OWA aggregation $OWA_g(a_1, ..., a_n)$ can be expressed as $OWA_g(a_1, ..., a_n) = \sum_{i=1}^{n} w_i b_i$ where b_i is the i^{th} largest of a_j and the w_i are a collection of weights such that $w_i = g(\frac{i}{n}) - g(\frac{i-1}{n})$. It can be easily shown these weights are proper, $w_i \in [0, 1]$ and $\sum_{i=1}^{n} w_i = 1$.

By appropriately selecting g we can implement different types of aggregation imperative. For example if $g(x) = x$ then the OWA operator becomes the ordinary average with $w_j = \frac{1}{n}$ for all j. If g is such that $g(x) = 1$ for all $x > 0$ then we get the maximal aggregation, $OWA_g(a_1, ..., a_n) = Max_i[a_i]$. If g is such that $g(x) = 0$ for all $x < 1$ the we get the minimal aggregation, $OWA_g(a_1, ..., a_n) = Min_i[a_i]$. A median type operator can be implemented if $g(x) = 0$ for $x < 0.5$ and $g(x) = 1$ for $x \geq 0.5$. A class of OWA operators can be obtained if $g(x) = x^{\alpha}$ with $\alpha \geq 0$.

Before preceding we shall find it convenient to use a slightly different notation for the OWA operator. We shall let *index* be an indexing function such that index(i) is the index of the i^{th} largest of the a_j. Thus we order the argument in descending order and then index(i) is the index of i^{th} element in this list. Since b_i is the i^{th} largest of the a_j using this index function we see that $b_i = a_{index(i)}$. Using this we can express the OWA aggregation as

$$OWA_g(a_i, ..., a_n) = \sum_{i=1}^{n} w_i \, a_{index(i)},$$

where the w_i as before are $w_i = g(\frac{i}{n}) - g(\frac{i-1}{n})$.

As in the preceding we shall let Sup(a, b) indicate the support for a from b. We note that using the index operator $Sup(a_{index(i)}, a_{index(j)})$ still represents the support of the second argument for the first. Because of the nature of the Sup function, Sup(a, b) ≥ Sup(x, y) when |a – b| < |x - y|, and the ordering captured by the index function we note that if i < j < k then $Sup(a_{index(i)}, a_{index(j)}) \geq Sup(a_{index(i)}, a_{index(k)})$ and $Sup(a_{index(j)}, a_{index(k)}) \geq Sup(a_{index(i)}, a_{index(k)})$. We let $T(a_{index(i)})$ denote the support of the i^{th} largest argument by all the other arguments, hence

$$T(a_{index(i)}) = \sum_{\substack{j=1 \\ j \neq i}}^{n} Sup(a_{index(i)} \; a_{index(j)}).$$

In addition we shall let $V_{index(i)} = 1 + T(a_{index(i)})$ and denote $TV = \sum_{i=1}^{n} V_{index(i)}$. We now can define the Power OWA operator as

$$POWA_g(a_1, ..., a_n) = \sum_{i=1}^{n} u_i \, a_{index(i)}$$

where $u_i = g(\frac{R_i}{TV}) - g(\frac{R_{i-1}}{TV})$ with $R_i = \sum_{j=1}^{i} V_{index(j)}$, by definition $R_{i-1} = 0$. We note that $TV = R_n$. We also observe that $R_i = R_{i-1} + V_{index(i)}$.

We can show in the special case where g(x) = x that this reduces to the Power Average. In this case

Another class of mean operators, called generalized means [8], are defined by

$$GM_\alpha(a_1, a_2, ..., a_n) = \left(\frac{1}{n}\sum_{j=1}^{n} a_j^\alpha\right)^{1/\alpha}$$

where $\alpha \in [-\infty, \infty]$. It is required when using these operators that $a_j \geq 0$. The inclusion of support in this class of mean operators can be accomplished in the following manner. Again let $T(a_i) = \sum_{j=1, j \neq i}^{n} Sup(a_i, a_j)$, $V_i = 1 + T(a_i)$ and $TV = \sum_{i=1}^{n} V_i$. the power generalized mean is defined as.

$$PGM_\alpha(a_1, ..., a_n) = \left(\frac{1}{TV} \sum_{i=1}^{n} V_i \, a_i^{\alpha}\right)^{\frac{1}{\alpha}}$$

We shall not further look at the properties of the Power OWA or the power generalized mean only to indicate that they act with respect to their mother operations in a manner similar to the way the power average acts with respect to the average.

In the preceding we assumed that all of the objects being aggregated were of equal importance. Here we shall consider the effect on the power operations of having differing importances associated with the objects being aggregated. We assume that each being aggregated has a weight $\omega_i \in [0, 1]$ indicating its importance. The procedure for including this importance involves a simple modification of the value V_i which we recall is defined as $V_i = 1 + T(a_i)$ where $T(a_i) = \sum_{j=1, j \neq i}^{n} Sup(a_i, a_j)$. In order to include the weights we suggest redefining V_i as

$$V_i = \omega_i \left(1 + \sum_{j=1, j \neq i}^{n} \omega_j \, Sup(a_i, a_j)\right)$$

and then continuing as described in preceding.

6 Conclusion

We introduced the power average operator to provide an aggregation operator which allows argument values to support each other in the aggregation process. The properties of this operator were described. We discussed the idea of a power median. We introduced some formulations for the support function used in the power average. We extended the idea of empowerment, supported aggregation, to a wider class of mean operators such as the OWA and generalized mean. Interesting applications of this approach to aggregation can be seen in data mining, group decision making and information fusion.

References

[1] Yager, R.R.: The power average operator. IEEE Transaction on Systems, Man and Cybernetics Part A 31, 724–730 (2001)
[2] Xu, Z., Yager, R.R.: Power geometric operators and their use in group decision making. IEEE Transactions on Fuzzy Sets and Systems (to appear)
[3] Yager, R.R., Filev, D.P.: Approximate clustering via the mountain method. IEEE Transactions on Systems, Man and Cybernetics 24, 1279–1284 (1994)
[4] Chiu, S.L.: Fuzzy model identification based on cluster estimation. Journal of Fuzzy and Intelligent Systems 2, 267–278 (1994)
[5] Yager, R.R.: A note on a fuzzy measure of typicality. International Journal of Intelligent Systems 12, 233–249 (1997)
[6] Zadeh, L.A.: Generalized theory of uncertainty (GTU)-principal concepts and ideas. Computational Statistics and Data Analysis 51, 15–46 (2006)
[7] Yager, R.R., Filev, D.P.: Essentials of Fuzzy Modeling and Control. John Wiley, New York (1994)
[8] Beliakov, G., Pradera, A., Calvo, T.: Aggregation Functions: A Guide for Practitioners. Springer, Heidelberg (2007)
[9] Yager, R.R.: On ordered weighted averaging aggregation operators in multi-criteria decision making. IEEE Transactions on Systems, Man and Cybernetics 18, 183–190 (1988)