Desde los Datos hasta el Valor: Procesos de Agregación de Información basados en el Concepto de Mayoría

J.I. PELÁEZ
Dpto. Lenguajes y Ciencias de la Computación
Universidad de Málaga. España.
Investigador Proyecto Prometeo. Ecuador
jipeláez@uma.es

A.M. CASADO
Departamento de Economía y Finanzas
Universidad de Málaga
Málaga, España
acasado@uma.es

E.R. YANEZ
Grupo de Investigación SIBI
Universidad de Guayaquil
Guayaquil, Ecuador
Estela.yanezb@ug.edu.ec

RESUMEN

Las transformaciones tecnológicas y de información que está experimentando nuestra sociedad en la última década, está provocando un crecimiento exponencial de la información, no solamente para las empresas u organizaciones, sino también para todos los ciudadanos. Pero operar en el entorno de mayor capacidad de generación de datos de la historia conlleva la adaptación de herramientas y procesos. Así por ejemplo, hay que abordar problemas de compresión y disponibilidad de información, recuperación de datos faltantes en las BBDD, tratamiento de imágenes, o minería de datos para los procesos de toma de decisión. Entre todos estos problemas destacan aquellos que tratan de obtener información, dentro de este volumen de datos, para poder tomar las mejores decisiones. Los procesos de toma de decisión constan de dos pasos: un primer paso de agregación de la información, que trata de sintetizar un valor que represente al conjunto de datos; y un proceso de explotación. Para resolver el problema de agregación de datos se utilizan operadores como la media aritmética, pero todos estos operadores no obtienen resultados que representen a la mayoría. Para dar solución a este problema fueron propuestos los denominados operadores de mayoría, que tratan de obtener valores representativos de la mayoría representando también a las minorías. En este trabajo se realiza una revisión de los principales operadores de agregación de mayoría, y su aplicación a diferentes tipos de problemas.

Palabras Clave. Operadores de Agregación de Mayoría, Toma de Decisión Multicriterio, Operadores OWA.

1. INTRODUCCIÓN

Las transformaciones tecnológicas y de información que está experimentando la sociedad, especialmente en la última década, está produciendo un crecimiento exponencial de los datos en todos los ámbitos de la sociedad [14]. Sirva de ejemplo los volúmenes de datos que recientes estudios han mostrado de las principales redes sociales en el año 2014: Facebook: el número de usuarios activos por día es de 699 millones; el número total de páginas en Facebook es de 50 millones; el número promedio de fotos que se publican en Facebook por día es de 350 millones. Twitter: el número mensual de usuarios activos en esta red es de 215 millones; el número promedio de tweets enviados al día es de 500 millones. YouTube: el número total de videos que se han subido a YouTube es superior a 4 billones, 300 horas de video por minuto. LinkedIn: el número total de usuarios de LinkedIn es de más de 238 millones; cada segundo se unen dos usuarios nuevos a LinkedIn. Instagram: el número promedio de fotos que se suben a esta red por día es de 55 millones; el número total de fotos subidas a Instagram es de más de 16 billones. De manera que los volúmenes de información han pasado de manejarse en poco tiempo de Gigabytes a Zettabyte.

Los datos que se generan en los diferentes ámbitos, como muestra la pirámide de la información (figura 1) son la mínima unidad semántica, y se corresponden con elementos primarios de información que por sí solos son irrelevantes como apoyo a las tomas de decisiones. Para que estos datos puedan ser de utilidad en cualquier proceso de decisión, es preciso que se conviertan en información, es decir, en un conjunto de datos procesados con un significado, para ayudar a crear conocimiento.

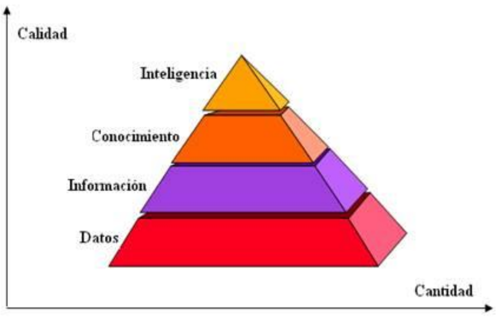

Figura 1. Pirámide de la Información.

Los procesos de toma de decisión se componen de dos pasos: un primer paso de agregación de datos para obtener información, que supone el uso de uno o varios operadores de agregación capaces de proporcionar una relación de preferencia colectiva; y un segundo paso de explotación de la información, que permite transformar la información en conocimiento a través de técnicas de minería de datos.

El proceso de agregación de ha convertido en la principal tarea de los procesos de toma de decisión [18]. Para llevarla a cabo, tal como hemos indicado anteriormente, se hace uso de operadores de agregación que actúan entre el mínimo y el máximo a través de operadores de medias [1, 2, 4, 17, 16, 18, 19]. Pero estos operadores no obtienen valores que sean representativos de las mayoría de los datos a agregar [6, 7, 8, 12], ya que en la mayoría de los casos presentan los denominados problemas de reparto [8, 9, 11], beneficiando los datos minoritarios en contra de los mayoritarios. Para dar solución a estos problemas, se presentaron los operadores y modelos de mayoría [5, 6, 9].

El objetivo de este trabajo es mostrar los principales operadores de agregación de mayoría y su aplicación en diferentes áreas de investigación como la toma de decisión multicriterio, lógica fuzzy, imputación de información en BBDD, recuperación de imágenes, y minería de opinión. Para ello el trabajo ha sido organizado como sigue: en la segunda sección se presentan los operadores de mayoría como una familia de los operadores OWA, mostrando los diferentes procesos de cálculo de los pesos de ponderación así como su utilización con etiquetas lingüísticas; en la tercera sección se presentan las diferentes aplicaciones; y finalmente las conclusiones y referencias.

2. OPERADORES DE MAYORÍA

Los procesos de Mayoría son introducidos por Peláez & Doña [7] como una nueva forma de agregación cuyo objetivo principal es obtener un resultado que represente el valor/opinión de la mayoría, sin olvidar al mismo tiempo a las minorías. Posteriormente, trabajo como los realizados por Pasi & Yager [6] proponen un modelo de mayoría para las tomas de decisión en grupo.

Los principales operadores de agregación de mayoría son el: el operador MA-OWA que utiliza la cardinalidad de los datos a agregar para determinar su importancia [9]; el operador LAMA-OWA que es utilizado en entornos lingüísticos [8]; el operador WC-OWA que es utilizado en procesos sociales modelando los denominados grupos de trabajo en los procesos de decisión en grupo [5]; y el operador SMA-OWA que incorpora una variable que actúa como un potenciómetro, que permite flexibilizar el concepto de mayoría en la agregación de datos [3].

2.1. Operador MA-OWA

Se define el operador de pesos ponderados de mayoría aditiva, F_{MA}, como [9]:

$$F_{MA}(a_1, a_2, \ldots, a_n) = \sum_{j=1}^{n} w_j \cdot b_j =$$

$$= \sum_{j=1}^{n} f_j(b_1, b_2, \ldots, b_n) \cdot b_j$$

donde b_j es el *j-ésimo* mayor valor de las a_n y los w_j los pesos de ponderación, que cumpliendo las condiciones de normalización de los operadores OWA [18], son calculados en función de los elementos a agregar como:

$$w_j = f_j(b_1, \ldots, b_n) = \frac{1}{\prod_{k=g_j}^{n} h_k(b_1, \ldots, b_n)}$$

siendo g_j una función que indica el momento en que el elemento b_j es considerado en el proceso de agregación. Este momento o paso, viene dado por el número de elementos iguales a b_j que existen en el intervalo $[b_1, b_j]$; y finalmente, h_k es una función que indica el número de elementos que son agregados en cada paso del proceso de agregación:

$$h_k(b_1, b_2, \ldots, b_n) = \begin{cases} \sum_{j=1}^{n} p_{kj} & \text{si } k = 1 \\ \sum_{j=1}^{n-k+1} p_{kj} + 1 & \text{en otro caso} \end{cases}$$

$$p_{kj} = \begin{cases} 1 & \text{si } b_j = b_{j+k-1} \text{ y } b_j \neq b_{j-1} \\ 1 & \text{si } j = 1 \text{ y } k \geq 1 \text{ y } b_j = b_k \\ 0 & \text{en otro caso} \end{cases}$$

2.2. Operador LAMA-OWA

Sea $p_1, p_2, \ldots, p_n, \in P$ un conjunto de etiquetas, tal que $t > 0$ y sea $\delta_1, \delta_2, \ldots, \delta_n \in N$, la frecuencia o cardinalidad de las etiquetas, donde $\delta_i \leq \delta_{i+1}$ para todo $1 \leq i \leq n-1$. El operador LAMA es la etiqueta p_m definida como [8]:

$$p_m = LAMA((p_1, \delta_1), (p_2, \delta_2), \ldots, (p_n, \delta_n)) =$$

$$= p_1 \otimes \lambda_1 \oplus p_2 \otimes \lambda_2 \oplus \ldots \oplus p_n \otimes \lambda_n$$

donde:

$$\lambda_i = \begin{cases} \frac{1}{d_1} & \text{if } i=1 \\ \frac{1}{d_1} \cdot \frac{1-n^{\delta_2}}{1-n} & \text{if } i=2 \\ \lambda_{i-1} + \frac{1}{d_{i-1}} \cdot \frac{1-(n-i+2)^{\delta_i - \delta_{i-1}}}{1-(n-i+2)} & \text{if } i > 2 \end{cases}$$

con

$$d_i = \begin{cases} 1 & \text{if } i = 1, n = 1 \\ n^{\delta_2} & \text{if } i = 1, n = 2 \\ n^{\delta_2} \cdot \prod_{j=1}^{n-2} (n-j)^{\delta_{j+2} - \delta_{j+1}} & \text{if } i = 1, n > 2 \\ \prod_{j=i-1}^{n-2} (n-j)^{\delta_{j+2} - \delta_{j+1}} & \text{if } i > 1 \end{cases}$$

y donde $\oplus$ es la suma de etiquetas y $\otimes$ es el producto de una etiqueta por un real positive defiido en [4].

2.3. Operador WKC-OWA

El operador Work Committee-OWA es definido para agregar información en problemas de decisión democrática, usando el concepto de comités de trabajo. El operador es definido como [5]:

$$F_{WKC}(a_1, a_2, \ldots, a_n) = \sum_{j=1}^{n} w_j \cdot a_j =$$

$$= \sum_{j=1}^{n} f_j(a_1, a_2, \ldots, a_n) \cdot a_j$$

donde $w_j \in [0,1]$ y $\sum_{j=1}^{n} w_j = 1$.

w_j es igual a:

$$F_{WKC}(a_1, a_2, \ldots, a_n) = \sum_{j=1}^{n} w_j \cdot a_j$$

$$w_j = f_j(b_1,\ldots,b_n) = \frac{1}{\max|\{b_1,\ldots,b_n\}| \cdot h_j(b_1,\ldots,b_n)}$$

h_j es una función que indica el número de elementos con cardinalidad ≤ a_j para todo $a_k \in \{a_j, \ldots, a_n\}$, por ejemplo:

$$h_k(b_1, b_2, \ldots, b_n) = 1 + \sum_{i=1}^{n} L_{ki}$$

donde

$$L_{ki} = \begin{cases} 1 & if \ |b_i| \geq C_k \quad with \quad b_i \neq b_k \wedge b_i \neq b_{i-1} \\ 0 & otherwise \end{cases}$$

$$C_k = \sum_{j=1}^{k} p_{kj}$$

$$p_{kj} = \begin{cases} 1 & \text{if } |b_j - b_k| \leq \alpha \\ 0 & otro\_caso \end{cases}$$

El valor α modela el tamaño final de cada comité. Socialmente este grado es determinado por la flexibilidad de los ciudadanos para agruparse en grupos y reforzar sus opiniones.

2.4. Operador SMA-OWA

El operador SMA-OWA es una función $F_{SMA}: \mathbb{R}^n \times \mathbb{N}^N \to \mathbb{R}$ definida como [3]:

$$\sum_{i=1}^{n} w_{i,N} v_{\sigma(j)}$$

donde: $N = \max_{1 \leq i \leq n} m_i$,
$\sigma \in S_n$ es una permutación ordenada, por ejemplo, tal que $v_{\sigma(i)} \geq v_{\sigma(i+1)}$, y los pesos son definidos por la relación de recurrencia

$$w_{i,1} = \frac{1}{u_1} = \frac{1}{n}; \ u_1 = n \qquad (1)$$

$$w_{i,k} = \frac{\gamma_{i,k} + w_{i,k-1}}{u_k}$$

$$u_k = 1 + \sum_{j=1}^{n} \gamma_{j,k}, \quad 2 \leq k \leq N$$

donde:

$$\gamma_{ik} = \begin{cases} \delta & m_{\sigma(i)} \geq k \\ 1-\delta & otro\_caso \end{cases}$$

δ es el factor de relevancia de la cardinalidad factor (CRF) y 0 ≤ δ ≤ 1.

3. APLICACIÓN DE LOS OPERADORES DE MAYORÍA

La necesidad de obtener valores en los procesos de agregación de datos que represente a la mayoría, no es exclusivo de los procesos de toma de decisión multicriterio. Se puede encontrar en muchos ámbitos de profesionales o de investigación.

A continuación se presentan algunos problemas donde el concepto de mayoría ha sido utilizado con buenos resultados.

3.1. Datos Faltantes en las BBDD

Los datos faltantes en las BBDD es motivo de gran preocupación en la comunidad científica o empresarial, especialmente en el campo de las ciencias sociales o médicas.

Hacer frente a los datos faltantes suele ser un asunto difícil, por lo que se hace necesario aplicar métodos de recuperación para inferir datos que sean válidos.

Una estratégica simple y común es ignorar los valores perdidos, lo que reduce el tamaño de datos útiles, y puede introducir sesgos importantes en los estudios, especialmente cuando los datos que faltan se distribuyen de una manera no aleatoria.

Otra forma es utilizar métodos para dar respuesta a este problema como: (1) los métodos simples de imputación, se basan en calcular un valor medio de la variable donde faltan los datos, o una media de clase o de otra variable. Métodos basados en regresiones; métodos Hot Deck; métodos del vecino más cercano; los Predictive Mean Matching; los Repeated Imputation; o los métodos basados en mayoría difusa (Fuzzy majority imputation), los cuales están basados en el concepto de mayoría a través de los diferentes operadores expuestos anteriormente. Estos métodos han sido aplicados en diferentes escenarios de recuperación de datos, obteniendo buenos resultados [13].

3.2. Tratamiento de Imágenes

La eliminación de ruido en imágenes digitales es un aspecto de vital importancia, ya que cualquier sistema de visión artificial comienza con el procesamiento de los valores de intensidades. Una manera de eliminar ese ruido es mediante los operadores lineales de convolución. Estos operadores o filtros, eliminan el ruido en las imágenes, calculando un nuevo valor de intensidad para cada pixel de la imagen mediante una convolución de los píxeles adyacentes, por ejemplo, mediante la media aritmética.

En este tipo de problemas la aplicación del concepto de mayoría, es decir, reemplazar un pixel por un valor que represente a la mayoría de la ventana de recuperación, a través de los operadores de agregación ha mostrado buenos resultados [10]. Por ejemplo, tiene mejores resultados que la mediana, media o gauss, cuando es aplicado en imágenes con ruido alto y medio, y buenos resultados con imágenes con ruido bajo. La figura 2, muestra un ejemplo de imágenes con ruido de Poisson y diferentes filtros.

3.3. Minería de Opinión

En los últimos años, la adquisición de productos los servicios a través de Internet ha crecido de manera exponencial. Hoy en día, una persona pasa a través de diferentes sitios web en busca de mejores alternativas y precios. En este proceso de búsqueda, él / ella utilizan aplicaciones de medios sociales para examinar las opiniones de las demás personas que han adquirido este tipo de servicios o productos. Entre la información que ofrecen las web, está la satisfacción de los compradores con dichos productos o servicios, la cual se hace mediante una escala predefinida, expresada comúnmente con un sistema de estrellas o etiquetas lingüísticas. En la mayoría de los casos, el sistema determina el valor de satisfacción mediante una operación de media aritmética, y en muchos casos, este cálculo no refleja de manera adecuada las opiniones de todos los usuarios.

Figura 2. Imágenes de Lápices con Ruido de Poisson y Filtros.

Para resolver este problema, los operadores de agregación basados en el concepto de mayoría, como por ejemplo el operador SMA-OWA [14], se han mostrado muy adecuados tanto con sistemas de estrellas como lingüísticos. Los resultados obtenidos han resultado más precisos. Así mismo, el proceso ha sido más dinámico y automático, ya que los pesos se calculan cuando se detecta una nueva votación.

También este tipo de operadores, aborda de manera más eficiente opiniones desesperadas/extremas y el resultado refleja mayor estabilidad que la media aritmética.

3.4. Coalición de Criterios

Finalmente, el concepto de mayoría está trascendiendo el área de los operadores de agregación de datos, y se está generalizando a otras situaciones más complejas en los procesos de toma de decisión multicriterio, como es el caso de la coalición de criterios en problemas de toma de decisión [15].

Modelar las interacciones entre criterios en los procesos de toma de decisión multicriterio es una tarea difícil. Esta complejidad surge cuando hay redundancias y sinergias visibles entre criterios, que los métodos tradicionales no pueden hacer frente. Para dar solución a este problema, Bernal et all en [15] proponen haciendo uso de los procesos de mayoría y de la integral de Choquet un método para determinar el valor de los pesos de cada criterio.

En la figura 3, se muestra el diagrama de flujo para el modelo propuesto para el problema de Coalición de Criterios.

4. CONCLUSIONES

La sociedad actual junto con las transformación tecnológicas y los volúmenes de información que se están generando, precisa de métodos que permitan obtener a partir de los datos, información que refleje el sentimiento o el sentir de la mayoría de dichos datos.

Tradicionalmente, estos procesos de conversión de datos en información se han llevado a cabo mediante operadores de agregación que no reflejan un valor de la mayoría, todo lo contrario, potencian el valor de las minorías. Este hecho es de gran importancia, porque estamos utilizando información que no refleja el sentir de la mayoría.

Para dar respuesta a estos problemas, los cuales han sido denominados como problemas de reparto, han surgido los procesos de mayoría, los cuales obtienen información que representa a la mayoría pero sin olvidar al mismo tiempo a las minorías.

Estos procesos de mayoría han sido modelados a través de diferentes operadores o de modelos, destacando entre estos el operador MA-OWA, como origen del resto de operadores o de modelos.

Dichos operadores se han mostrado muy adecuados en diferentes tipos de problemas, como por ejemplo en el tratamiento de imágenes, recuperación de datos faltantes en las BBDD, problemas de toma de decisión multicriterio, minería de opinión, o en el modelado de las sinergias en las coaliciones de criterios.

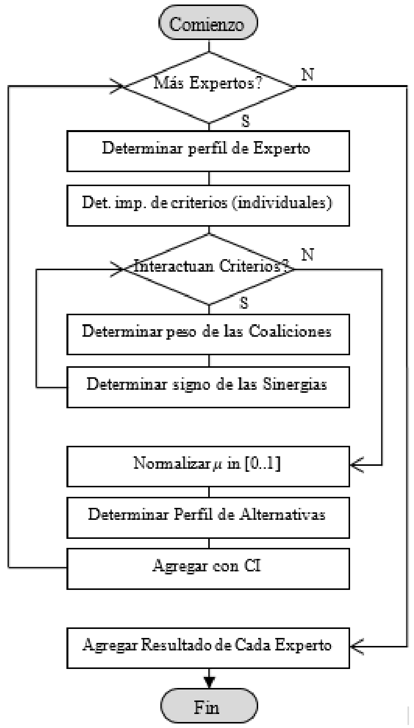

Figura 3. Diagrama de Flujo del Modelo.

Agradecimientos. *Este trabajo está financiado por el proyecto PROMETEO del Gobierno de Ecuador.*

5. REFERENCIAS

[1]. Calvo, T. & Mesiar, R. (2003). "Aggregation operators: ordering and bounds". Fuzzy Sets and Systems,

[2]. Fodor, J.C.; Marichal, J.L. & Roubens, M. (1995). "Characterization of some aggregation functions arising from MCDM problems" en b. Bouchonmeunier, R. R. Yager & L.A. Zadeh (eds.). Fuzzy logic and soft computing. Series: Advances in fuzzy systems-Applications and theory, 4. World Scientific Singapore, pp. 194-201.

[3] Karanik M, Peláez J.I. Bernal R, (2015 aceptado). Selective Majority Additive Ordered Weighting Averaging Operator. European Journal of Operational Research.

[4]. Llamazares B. (2007). Choosing OWA operator weights in the field of Social

Choice. Information Sciences: an International Journal. 177, 21, 4745-4756.

[5]. La Red D.L, Doña J.M, Peláez J.I, Fernández E.B. (2011). WKC-OWA. A new neat-owa operator to aggregate information in democratic decision problems. International Journal of Uncertainty fuzziness and knowledge based systems.

[6]. Pasi G. & Yager R. (2006). Modeling the Concept of Majority Opinion in Group Decision Making. Recent Advancements of Fuzzy Sets: Theory and Practice. Information Sciences, 176, 4, 390-414.

[7] Peláez J.I. Doña J.M. (2001). LAMA: A Linguistic Aggregation of Majority Additive Operator. Proceeding of the congress: Eurofuse. Granada. Spain.

[8] Peláez J.I. Doña J.M. (2003). LAMA: A Linguistic Aggregation of Majority Additive Operator. International Journal of Intelligent Systems.

[9] Peláez J.I. Doña J.M. (2003). Majority Additive-Ordered Weighting Averaging: A New Neat Ordered Weighting Averaging Operators Based on the Majority Process. International Journal of Intelligent Systems, 18, 4, 469-481.

[10] Peláez J.I. Doña J.M, Sánchez P, Mesas A. (2005). Aplicación y uso del operador MA-OWA en el tratamiento de imágenes. Proceeding Conferencia Caepia.

[11] Peláez J.I. Doña J.M. (2006). A Majority Model in Group Decision Making Using QMA-OWA Operators. International Journal of Intelligent Systems, 193-208.

[12] Peláez J.I. Doña J.M, Gómez-Ruiz J.A. (2007). Analysis of OWA operators in decision making for modelling the majority concept. Applied mathematics and Computation.

[13] Peláez J.I, Doña J.M, La Red D.L. (2008) Fuzzy Imputation Method for Database Systems. Hardbook of Research on Fuzzy Information Processing in Databases.

[14] Peláez J.I. (2014). Bernal R. Karanik M. Majority OWA Operator for opinion ratting in social media. Soft Computing.

[15] Bernal R, Karanik M, Peláez J.I. (2015) Fuzzy measure identification for criteria coalitions using linguistic information. Soft Computing.

[16] Scott, M.J. Antonsson, E.K (1998). “Aggregation functions for engineering design trade-offs”. Fuzzy Sets and Systems

[17] Smolíkova, R. Wachowiak, M.P. (2002). “Aggregation operators for selection problems”. Fuzzy Sets and Systems,

[18] Yager R. (1998). On ordered weighted averaging aggregation operators in multi-criteria decision making. IEEE Trans. On Systems, Man and Cybernetics 18. 183-190.

[19] Zimmermann, H.J. (1991). Fuzzy sets theory and its application. Kluwer Academia Publishers. Boston/ Dordrecht/ London.