

Extensión del concepto de utopía para el problema de la agregación de rankings sin empates

Extension of the Concept of Utopia for Rank Aggregation Problem Without Ties

  Randy Reyna-Hernández¹;
 Alejandro Rosete²;

¹Universidad de Matanzas, Matanzas, Cuba,
randyrh91@gmail.com

²Universidad Tecnológica de La Habana “José Antonio Echeverría”,
La Habana, Cuba,
rosete@ceis.cujae.cu

Cómo citar / How to cite

R. Reyna-Hernández; A. Rosete, “Extensión del concepto de utopía para el problema de la agregación de rankings sin empates”, *Tecnológicas*, vol. 24, nro. 51, e1788, 2021. <https://doi.org/10.22430/22565337.1788>

ISSN-p: 0123-7799
ISSN-e: 2256-5337

Vol. 24, nro. 51, e1788, 2021

Recibido: 22 octubre 2020
Aceptado: 16 febrero 2021
Disponible: 19 marzo 2021

©Instituto Tecnológico Metropolitano
Este trabajo está licenciado bajo
una Licencia Internacional
Creative Commons Atribución
(CC BY-NC-SA)



Resumen

El uso de los rankings y la forma de agregarlos o resumirlos ha recibido una atención creciente en diversos campos: bibliometría, búsquedas web, minería de datos, estadística, calidad educativa y biología computacional. Para el Problema de Ordenamiento Óptimo con empates fue introducido recientemente el concepto de Matriz Utópica: una solución ideal y no necesariamente factible con una calidad insuperable para las soluciones factibles del problema. Este trabajo propone una extensión de la noción de Matriz Utópica para el Problema de Agregación de Rankings en que no se permiten empates entre elementos en el ranking de salida. Más allá de la extensión que es directa, el trabajo se centra en estudiar su valor como idealización o solución súper óptima. Como el Problema de Agregación de Rankings puede resolverse de forma exacta a partir de su definición como Problema de Programación Lineal Entera, se presenta un estudio experimental donde se analiza la relación que existe entre los valores utópicos (y anti utópicos) y la solución óptima en instancias resueltas con la ayuda del software de código abierto SCIP. Entre las 47 instancias analizadas, en 19 el Valor Utópico resultó ser igual al valor óptimo (40,43 % de factibilidad) y en 18 el Valor Anti Utópico también resultó ser factible (38,00 %). Este estudio experimental demuestra la utilidad de los valores utópicos y anti utópicos para ser considerados como valores extremos en el Problema de Agregación de Rankings, pudiendo así encontrarse muy rápidamente cotas superiores e inferiores para la optimización.

Palabras Claves

Agregación de rankings, programación lineal entera, minería de datos, algoritmos de optimización.

Abstract

The use of rankings and how to aggregate or summarize them has received increasing attention in various fields: bibliometrics, web search, data mining, statistics, educational quality, and computational biology. For the Optimal Bucket Order Problem, the concept of Utopian Matrix was recently introduced: an ideal and not necessarily feasible solution with an unsurpassed quality for the feasible solutions of the problem. This work proposes an extension of the notion of Utopian Matrix to the Rank Aggregation Problem in which ties are not allowed between elements in the output ranking. Beyond the extension that is direct, the work focuses on studying its usefulness as an idealization or super optimal solution. As the Rank Aggregation Problem can be solved exactly based on its definition as an Integer Linear Programming Problem, an experimental study is presented where it is analyzed the relationship that exists between utopian (and anti utopian) values and the optimal solution in several instances solved by using the open source software SCIP. Among the 47 instances analyzed, in 19 the Utopian Value turned out to be equal to the optimal value (40.43 % feasibility) and in 18 the Anti Utopian Value also turned out to be feasible (38.00 %). This experimental study demonstrates the usefulness of utopian and anti utopian values to be considered as extreme values in the Rank Aggregation Problem, thus being able to find higher and lower bounds for optimization very quickly.

Keywords

Rank aggregation, integer linear programming, data mining, approximation algorithms.

1. INTRODUCCIÓN

En la actualidad se generan grandes cantidades de datos a cada segundo. Las técnicas para el manejo de grandes volúmenes de información surgen con el nombre de minería de datos y se constituyen como la alternativa ideal en estos casos.

La minería de datos comprende un conjunto de técnicas tendientes a “realzar” o descubrir comportamientos y patrones presentes dentro de un conjunto de datos con la finalidad de generar conocimiento que proporcione soluciones a un problema determinado [1]. Como gran parte de estos datos son presentados siguiendo algún orden de preferencia, temas como la agregación de rankings han llamado la atención de numerosos investigadores y ya cuentan con diferentes aplicaciones en variados temas que van desde la calidad educativa [2] hasta la biología computacional [3]–[7].

Debido a su popularidad en la comunidad científica, y la amplia gama de aplicaciones que presenta el tema, se han desarrollado además numerosas herramientas para facilitar su uso [8],[9].

Dado un conjunto de elementos (u objetos), para ser ordenados por jueces, según su criterio de preferencias (por ejemplo, los mejores delante y los peores al final), los rankings proporcionados por cada juez, que ordenan todos los elementos, (o algunos de ellos), podrían ser combinados para obtener un ranking de consenso (agregación). El problema de encontrar dicho ranking es conocido en la literatura como Problema de Agregación de Rankings (RAP, por sus siglas en inglés) [10].

Existen diferentes tipos de RAP. Por ejemplo, cuando los rankings a agregar no tienen empates ni ausencias de algún elemento, entonces los rankings son permutaciones (rankings completos sin empates) y el RAP se reduce al Problema de Kemeny (KRP) [11].

Una variante más reciente del RAP es el Problema de Ordenamiento Óptimo de Rankings con Empates (OBOP, por sus siglas en inglés), cuya solución consiste en un ranking con empates, a diferencia del KRP [12]–[14].

En el RAP, para instancias de gran tamaño no es posible, en general, obtener la solución exacta. Sin embargo, se puede abordar mediante el uso de algoritmos heurísticos codiciosos (por ejemplo, Borda), los cuales encuentran una solución aproximada del problema. Estos algoritmos son rápidos, pero las soluciones así obtenidas están lejos de ser óptimas. Las metaheurísticas han mostrado un buen equilibrio entre la eficiencia y la precisión en problemas de este tipo [15],[16].

En [14] se introducen dos nuevos conceptos, la Matriz Utópica y la Anti Utópica, que se asocian a una matriz de orden de pares que representa las precedencias en un conjunto de rankings. Siguiendo la investigación presentada en [14], en este trabajo se replantean los conceptos de Matriz Utópica y de Matriz Anti Utópica para RAP. De esta forma, estas idealizaciones constituyen soluciones súper-óptimas para el RAP, que se pueden utilizar como valores extremos. Además, se muestran los resultados experimentales de la relación entre el Valor Utópico y la solución óptima de 47 modelos de Programación Lineal Entera (PLE) resueltos con la ayuda del software de código abierto SCIP.

El resto del documento está estructurado de la siguiente manera. A continuación, se definen el RAP y OBOP, así como se presentan los conceptos de Matriz Utópica, Matriz Anti utópica, Valor Utópico y Valor Anti Utópico para OBOP planteados en [14]. Seguidamente, se introducen los conceptos nuevos relacionados con los Valores Utópicos y Anti Utópicos para RAP. Posteriormente se muestran los resultados de los experimentos realizados. Finalmente, se dan a conocer las conclusiones del presente trabajo.

2. ANTECEDENTES

2.1 Problema de Agregación de Rankings (RAP)

Informalmente, el Problema de Agregación de Rankings (RAP) consiste en combinar varios rankings (que ordenan el mismo conjunto de candidatos, o alternativas), para obtener un ordenamiento que exprese un consenso entre todos ellos. La agregación de rankings se ha estudiado en muchas disciplinas, principalmente en el contexto de la teoría de la elección social, donde existe una rica literatura que data de la segunda mitad del siglo XVIII [17].

Entonces, formalmente, dado un conjunto $[[n]] = \{1, \dots, n\}$ de elementos, un ranking π es un orden de preferencia de estos (o algunos de estos) elementos. Los rankings que ordenan todos los elementos de $[[n]]$ se denominan *completos*, mientras que los que ordenan los elementos de un subconjunto de $[[n]]$ se denominan *rankings parciales*. Por otro lado, los rankings que establecen una preferencia entre cada par de elementos ordenados se denominan sin empates, mientras que los que presentan empates son *rankings con empates*.

Conceptualmente, un empate puede entenderse como una falta de información de preferencia entre algunos elementos clasificados. Los elementos empatados forman un bucket. Entonces, un ranking también puede entenderse como un orden de preferencia (disjunto) entre sus buckets [18].

Para identificar el ranking que será el que mejor represente ese conjunto, es importante medir cuan diferentes son dos de ellos [4]. Las distancias son la forma común de medir la diferencia entre dos rankings.

A pesar de existir diferentes formas de calcular la distancia entre dos rankings, la más popular es la distancia de Kendall-Tau.

La distancia de Kendall [19], entre dos rankings π y σ , se define en (1) como el número total de pares de elementos en desacuerdo. Hay desacuerdo sobre un par de elementos $(i; j)$ si el orden relativo de i y j es diferente en π y σ . Más precisamente:

$$K(\sigma, \pi) = |\{(i, j): i < j, (\sigma(i) < \sigma(j) \wedge \pi(i) > \pi(j)) \vee (\sigma(i) > \sigma(j) \wedge \pi(i) < \pi(j))\}| \quad (1)$$

Donde $\sigma(i)$ y $\pi(i)$ indican la posición del elemento i dentro de las permutaciones σ y π , respectivamente. Además, debe notarse que los rankings se expresan como permutaciones y no hay empates entre los elementos.

Una propiedad interesante de la Distancia de Kendall es que su valor máximo entre dos rankings de n elementos es $\frac{n(n-1)}{2}$.

Entonces, según [20] dado un conjunto de N rankings, $\Sigma = \{\sigma_1, \sigma_2, \dots, \sigma_N\}$ que ordenan n elementos $1, 2, \dots, n$, con $\sigma_i \in \mathbb{S}_n$, donde \mathbb{S}_n es un grupo que contiene todas las permutaciones de los n elementos, el RAP consiste en encontrar la permutación π_0 que satisface (2):

$$\pi_0 = \operatorname{argmin}_{(\pi \in \mathbb{S}_n)} \frac{1}{N} \sum_{i=1}^N K(\sigma_i, \pi) \quad (2)$$

$K(\sigma_i, \pi)$ corresponde a la distancia Kendall-Tau entre los rankings σ_i y la permutación π .

2.2 Problema de Ordenamiento Óptimo de Rankings con Empates (OBOP)

Nuevamente, el objetivo del OBOP es encontrar un ranking de consenso que ordene cada uno de los rankings de entrada, pero en este caso la respuesta puede contener empates entre varios elementos del ranking.

Más formalmente, dado un conjunto de elementos $[[n]] = \{1, \dots, n\}$, un ranking completo con empates β es una partición ordenada de $[[n]]$ [14]. Más precisamente, se trata de una ordenación lineal de subconjuntos disjuntos (buckets) B_1, B_2, \dots, B_k de $[[n]]$, $1 \leq k \leq n$, con $\cup_{i=1}^k B_i = [[n]]$. Por lo tanto, teniendo en cuenta dos buckets B_i, B_j en β , se escribe $B_i <_{\beta} B_j$, para indicar que B_i precede a B_j de acuerdo con el orden de bucket β . Análogamente, dados dos objetos $u \in B_i, v \in B_j$, se escribe $u <_{\beta} v$ si $B_i <_{\beta} B_j$. Todos los elementos que pertenecen al mismo bucket se consideran empatados. Por lo tanto, si u y v están empatados con respecto a B , se escribe $u \sim_{\beta} v$ [14].

Se puede representar B como una matriz C $n \times n$, donde $C(u, v) = 1$ si $u <_B v$, $C(u, v) = 0$ si $v <_B u$ y $C(u, v) = 0,5$ si $u \sim_B v$. Debe notarse, además, que $C(u, v) + C(v, u) = 1$ [21].

La entrada del problema es un conjunto de rankings (completos o parciales) que se representan en una matriz de precedencias P de dimensiones $n \times n$ con valores en el intervalo $[0,1]$ tal que $P(u, v) + P(v, u) = 1$ para todo $u, v \in [[n]]$, $u \neq v$, y $P(u, u) = 0,5$ para todo $u \in [[n]]$. Usualmente $P(u, v)$ es interpretado como la fracción de permutaciones de la muestra en las que el elemento u ha sido ordenado antes del elemento v [14].

Entonces, el objetivo del OBOP consiste encontrar el orden de buckets tal que la distancia entre las matrices C y P sea mínima. En otras palabras, el valor óptimo del OBOP para la matriz P consiste en encontrar una Matriz de Bucket C $n \times n$ que minimice la función (3):

$$f_c(C) = D(C, P) = \sum_{u,v} |C(u, v) - P(u, v)| \quad (3)$$

Esto quiere decir, que la distancia $D(C, P)$ es el valor modular o absoluto de las diferencias sumadas entre cada par de elementos de la matriz de entrada (matriz de precedencias) y la forma matricial de representar un ranking usada en el OBOP.

Es importante notar que la matriz de precedencia puede representar tanto rankings completos, incompletos como con empates. En [22] se demuestra que la distancia $D(C, P)$ es en realidad una extensión de la distancia de Kendall, formulada en (1) con anterioridad, con la diferencia que permite la comparación entre cualquier par de rankings.

Por lo tanto, el OBOP tiene una matriz de entrada P y una salida formada por la permutación de los conjuntos de partición. Así, se clasifican los elementos del ranking inicial en buckets que contienen subconjuntos con propiedades homogéneas, y proporciona la permutación de buckets que mejor los represente [21].

2.3 La Matriz Utópica (U_p) y el Valor Utópico (u_p) para OBOP

Según [14], dada una Matriz de Precedencia P , la Matriz Utópica asociada a P es la matriz $n \times n$ definido en (4) como:

$$U_p(u, v) = Y(P(v, u)) \quad (4)$$

Donde:

$$Y(x) = \begin{cases} 1 & \text{si } x > 0,75 \\ 0,5 & \text{si } 0,25 \leq x \leq 0,75 \\ 0 & \text{si } 0 < x < 0,25 \end{cases}$$

Debe notarse que cada celda de la Matriz Utópica toma uno de los valores posibles de precedencia cuando se permiten empates (0,0,5 o 1) escogiendo el que está más cerca de la matriz de precedencia de entrada.

Siendo así, no puede existir ninguna representación matricial de un ranking que tenga una celda con distancia menor a la matriz de precedencia que el valor presente en la Matriz Utópica. Entonces el Valor Utópico U_P asociado con P es $u_P = D(U_P, P)$.

Con base en lo anterior, el Valor Utópico U_P es una cota del valor óptimo del problema OBOP asociada a la matriz de precedencia P . Debe notarse que no siempre la Matriz Utópica representa una solución factible para el OBOP [14].

Debe tenerse en cuenta además que:

-Para cualquier Matriz de Precedencia P , la distancia máxima entre una salida en particular y la correspondiente en la Matriz Utópica es 0,25, y sucede cuando el valor en la matriz de precedencia de entrada es 0,25 o 0,75.

$$-P(u, u) = U_P(u, u) = 0,5$$

-Para una matriz de dimensión n , el mayor Valor Utópico es $u(n) = 0,25n(n - 1)$. Este valor de utopía corresponde a una matriz P con valores en $\{0,25, 0,75\}$ en todas las celdas, excepto en las de la diagonal principal.

-Si $P(u, v) = U_P(u, v) \in \{0, 0,5, 1\}$ para todo $u, v \in M$ entonces el Valor Utópico u_P asociado a P es 0. Es los demás casos, no es posible encontrar una solución con distancia 0.

Por otro lado, la utopicidad $U(P)$ de una matriz P se puede definir como se muestra en (5):

$$U(P) = \frac{u(n) - u_P}{u(n)} \tag{5}$$

La utopicidad $U(P)$ puede considerarse como una normalización en el intervalo $[0,1]$ de la similitud entre P y su Matriz Utópica U_P . En particular, si $U(P) = 1$ entonces $U_P = P$. Es decir, P es utópica.

2.4 La Matriz Anti Utópica (A_P) y el Valor Anti Utópico (A_P) para OBOP

Según [14], dada una Matriz de Precedencia P , la Matriz Anti Utópica asociada a P es la matriz $n \times n$ definido en (6):

$$A_P(u, v) = \Gamma(P(v, u)) \tag{6}$$

Donde:

$$\Gamma(x) = \begin{cases} 0 & \text{si } x > 0,5 \\ 1 & \text{si } x \leq 0,5 \end{cases}$$

Entonces el Valor Anti Utópico a_P asociado con P es $a_P = D(A_P, P)$.

Definido de esta manera, es una idealización a la peor solución posible al OBOP, y podría ser útil como límite superior para $D(\bullet, P)$. En otras palabras, dado una Matriz de Precedencia

arbitraria P , $D(B, P)$ está en el intervalo $[u_p, a_p]$ para cualquier orden de bucket B . Esto quiere decir que con los conceptos anteriores es posible determinar cotas superiores e inferiores para el problema OBOP. En la sección siguiente se mostrará cómo pueden extenderse estos conceptos para el problema RAP donde no es posible que haya empates en el ranking de consenso que brinda como salida.

3. EXTENSIÓN DEL CONCEPTO DE UTOPIA PARA RAP

3.1 La Matriz Utópica (UR_p) y el Valor Utópico (ur_p) para RAP

Dada una Matriz de Precedencia P , la Matriz Utópica asociada a P es la matriz $n \times n$ definida como (7):

$$UR_p(u, v) = \Psi(P(v, u)) \forall u \neq v \quad (7)$$

Donde:

$$\Psi(x) = \begin{cases} 0 & \text{si } x < 0,5 \\ 1 & \text{si } x \geq 0,5 \end{cases}$$

Entonces el Valor Utópico ur_p asociado con P es $ur_p = D(UR_p, P)$.

Debe notarse que, para una matriz E , no se puede obtener una solución que tenga valor de función objetivo menor que ur_p , pues es una cota inferior que solo se alcanza en algunos casos.

3.2 La Matriz Anti Utópica (AR_p) y el Valor Anti Utópico (ar_p) para RAP

Análogamente a la Matriz Anti Utópica y al Valor Anti Utópico del OBOP, se pueden definir estos conceptos para RAP, entonces, dada una Matriz de Precedencia P , la Matriz Anti Utópica asociada a P es la matriz $n \times n$ definida como (8):

$$AR_p(u, v) = \Phi(P(v, u)) \forall u \neq v \quad (8)$$

Donde:

$$\Phi(x) = \begin{cases} 0 & \text{si } x > 0,5 \\ 1 & \text{si } x \leq 0,5 \end{cases}$$

Entonces el Valor Anti Utópico ar_p asociado con P es $ar_p = D(AR_p, P)$.

Definido de esta manera, es una idealización a la peor solución posible al RAP, y podría ser útil como límite superior para $D(\bullet, P)$. En otras palabras, dado una Matriz de Precedencia arbitraria P , $D(B, P)$ está en el intervalo $[ur_p, ar_p]$ para cualquier ranking sin empates.

Es interesante notar que el espacio de solución del problema RAP es un subconjunto del espacio de soluciones del problema OBOP. Por esta razón, no es posible encontrar una solución del RAP que no sea solución del OBOP. De esta manera, el intervalo $[ur_p, ar_p]$ de los posibles valores de las soluciones del RAP está contenido en el intervalo $[u_p, a_p]$ de posibles soluciones para el OBOP. Así, las cotas definidas por el intervalo $[ur_p, ar_p]$ para el RAP están

más ajustadas. Esto resalta la importancia la extensión del concepto de Utopía para el RAP, ya que permite un mejor ajuste que si se usara el intervalo $[u_p, a_p]$.

4. RESULTADOS Y DISCUSIÓN

En esta sección se presenta un estudio experimental de la relación entre el Valor Utópico del RAP y los valores óptimos reales para esas instancias. Esto es posible, debido a la existencia de una formulación de PLE para el problema RAP que fue presentada en [10].

Por esta razón, dada una matriz de precedencia es posible encontrar su solución óptima usando una herramienta de solución de problemas de PLE. En este caso, se empleará la herramienta libre SCIP.

SCIP es un *framework* para la Programación Entera con Restricciones (CIP), un novedoso paradigma que integra la Programación con Restricciones (CP), la Programación Mixta-Entera (MIP) y las técnicas de modelado y solución de satisfacción (SAT) [23]. SCIP está disponible gratuitamente en código fuente para uso comercial y se puede descargar desde [24].

Es importante estudiar la relación que existe entre la solución óptima y la solución utópica en problemas en los cuales es posible encontrar su solución óptima en un tiempo prudencial, ya que puede servir de referencia para estimar lo que ocurriría en instancias mayores cuya solución óptima se desconozca, aprovechando que la obtención de las cotas utópicas es computable en muy poco tiempo.

Todos los experimentos fueron realizados en un ordenador personal con un procesador Intel i7 - 4790, 3,60 GHz, 4 núcleos y 4GB de memoria RAM.

Para los experimentos se usaron 47 conjuntos de datos (*dataset*) de rankings reales disponibles en [25]. En particular, se descargaron ficheros *pwg* asociados a los siguientes conjuntos de “Datos de Elección”: ED-00006-Skate Data(3–4, 11–12, 18, 28, 46, 48), ED-00011-Web Search(1), ED-00014-Sushi Data(1) and ED-00015-Clean Web Search (1–2, 7, 9,12, 14, 16–20, 23–25, 27, 29–30, 32, 34, 40–42, 44, 46, 48, 50, 54, 55,57, 59, 65–66, 67, 69, 73, 74, 77). Estos ficheros han sido utilizados previamente en experimentos de OBOP publicados en [14].

Todos los ficheros *pws* usados para los experimentos, así como los modelos de PLE y las herramientas con las que se ejecutaron los modelos [26] para poder ser analizados y replicados. Una descripción general de los 47 conjuntos de datos utilizados se muestra en la Tabla 1.

Tabla 1. Descripción de los conjuntos de rankings usados en los experimentos. Fuente: elaboración propia.

	Pro	Med.	Min.	Max.	Des Est.
n	78,87	64,0	10,00	240,0	59,55
v	111,19	4,0	4,00	5000,0	728,61
u_p	850,85	377,5	2,33	5277,5	1207,29
ur_p	930,85	487,0	2,67	6142,4	1438,67

Por cada conjunto de datos, se muestra el promedio (Pro), la mediana (Med), los valores mínimos (Min) y máximos (Max) y la desviación estándar (Des Est.) del número de elementos del conjunto de rankings a agregar (n), cantidad de votantes (v), Valor Utópico para OBOP (u_p) y el Valor Utópico para RAP (ur_p).

4.1 Relación entre el Valor Utópico y el valor óptimo en el RAP

La Tabla 2 muestra una comparación entre los valores utópicos para RAP (ur) para cada instancia y su valor óptimo real (vor_min). Además, se muestra la diferencia entre ambos valores. A partir de la Tabla 2 se puede apreciar que entre el valor óptimo real y su Valor Utópico correspondiente, normalizado teniendo en cuenta el tamaño de cada instancia, existe un desfase promedio de aproximadamente 0,083.

Como se muestra en la Figura 1, fue factible el Valor Utópico en 19 de los 47 modelos, resultando un 40,43 % de factibilidad aproximadamente.

Además, se puede apreciar que, en varios modelos, ambos valores coinciden porque la solución utópica es factible y por tanto la mínima diferencia proporcional es 1, sin embargo, es más interesante notar que el promedio es de 1,01 y el mayor valor es 1,10. Esto quiere decir que nunca en los modelos la solución óptima fue más de 10 % peor que el Valor Utópico.

La Figura 2 muestra un gráfico con la diferencia proporcional. Es interesante notar que en estos resultados la matriz utópica fue factible (se correspondió con la solución óptima) en una cantidad de instancias mucho mayor que las reportadas en [14]. Aunque este resultado merece un estudio a fondo, creemos que hay aspectos que favorecen esta diferencia.

Por una parte, aquí se ha usado un algoritmo exacto que garantiza el óptimo, mientras que en [14] se reporta el resultado promedio de algoritmos aproximados.

Adicionalmente, el problema OBOP tiene un espacio de búsqueda mucho mayor que el del RAP y con una mayor complejidad en las restricciones a cumplir. Por ejemplo, el primer ejemplo mostrado en [14] cuya solución utópica para el OBOP no es factible, sí tiene una solución utópica que se corresponde con el ranking 1 | 2 | 3 que es factible para el RAP. Estos aspectos pueden explicar las diferencias.

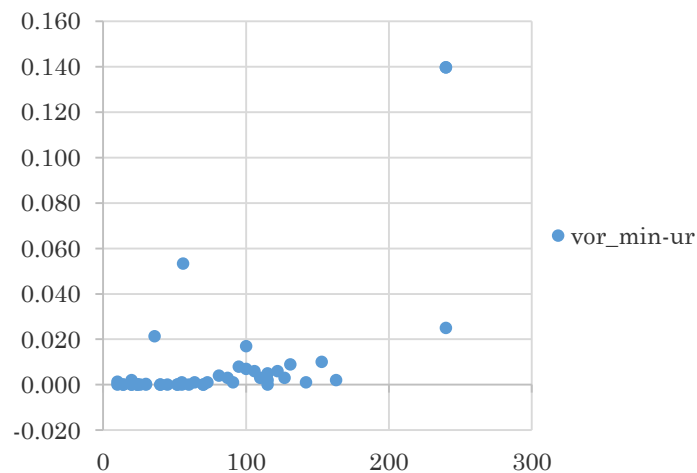


Figura 1. Comportamiento del desfase entre el Valor Utópico y el valor óptimo real según crecen los elementos a ordenar. Fuente: elaboración propia.

Tabla 2. Relación Valor Utópico – valor óptimo real. Fuente: elaboración propia.

dataset	n	ur	vor_min	vor_min - ur	vor_min / ur
06_03	14	7,111	7,111	0	1
06_04	14	2,667	2,667	0	1
06_11	20	19,111	19,111	0	1
06_12	20	9,778	9,778	0	1
06_18	24	12,444	12,444	0	1
06_28	24	42,222	42,444	0,222	1,005
06_46	30	28,857	29,143	0,286	1,01
06_48	24	18,667	18,667	0	1
11_01	240	6142,4	6282	139,6	1,023
14_01	10	30,917	30,917	0	1
15_01	240	7204,5	7229,5	25	1,003
15_02	240	6.142	6282	139,6	1,023
15_07	110	1908	1911	3	1,002
15_09	115	2257	2262	5	1,002
15_12	100	1652	1659	7	1,004
15_14	163	4458,5	4460,5	2	1
15_16	70	868	868	0	1
15_17	127	2955	2958	3	1,001
15_18	115	2235,5	2235,5	0	1
15_19	87	1254,5	1257,5	3	1,002
15_20	122	3135,5	3141,5	6	1,002
15_23	142	3612	3613	1	1
15_24	91	1375,5	1376,5	1	1,001
15_25	115	2469,5	2471,5	2	1,001
15_27	95	1639,5	1647,5	8	1,005
15_29	106	1848	1854	6	1,003
15_30	64	777,5	778,5	1	1,001
15_32	153	3899,5	3909,5	10	1,003
15_34	55	564,5	564,5	0	1
15_40	131	3498	3507	9	1,003
15_41	70	1025	1025	0	1
15_42	100	1994	2011,	17	1,009
15_44	45	331	331	0	1
15_46	40	247	247	0	1
15_48	10	18,667	20	1,333	1,071
15_50	26	148,5	148,5	0	1
15_54	60	599,5	599,5	0	1
15_55	52	382	382	0	1
15_57	73	1133,5	1134,5	1	1,001
15_59	55	490	491	1	1,002
15_65	40	325,5	325,5	0	1
15_66	52	364,5	364,5	0	1
15_67	30	148	148	0	1
15_69	81	1013	1017	4,000	1,004
15_73	36	224	245,333	21,333	1,095
15_74	20	52,667	54,667	2	1,038
15_77	56	594	647,333	53,333	1,09

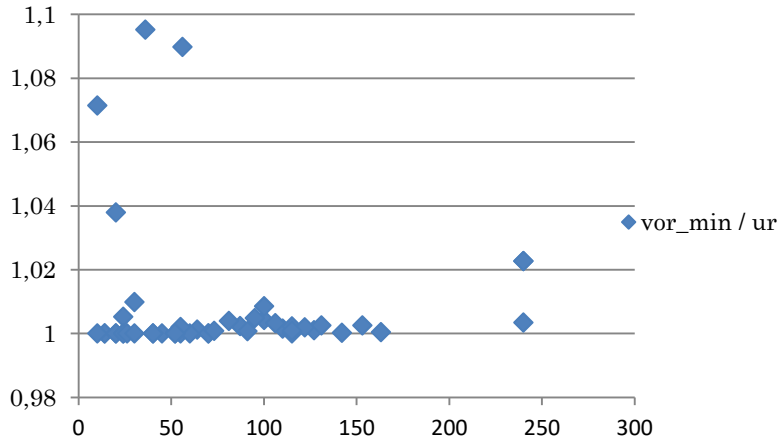


Figura 2. Diferencia proporcional entre el Valor Utópico y el valor óptimo real
Fuente: elaboración propia.

4.2 Relación entre el Valor Anti Utópico y el valor óptimo en el RAP

Para obtener la relación entre los valores anti utópicos y el valor óptimo real en el RAP se maximizó la función objetivo. La Tabla 3 muestra una comparación entre los valores anti utópicos para RAP (aur) para cada instancia y su valor óptimo real (vor_max). Además, se muestra la diferencia entre ambos valores.

En la Tabla 3 se pueden apreciar los resultados obtenidos para el valor óptimo real, en general, se encuentran muy próximos a su Valor Anti Utópico correspondiente, existiendo un desfase promedio aproximado de tan solo 0,37, normalizado teniendo en cuenta el tamaño de cada instancia.

Para una mejor comprensión se muestra la Figura 3. Además, fue factible el Valor Anti Utópico en 18 de los 47 modelos, resultando un 38 % de factibilidad aproximadamente.

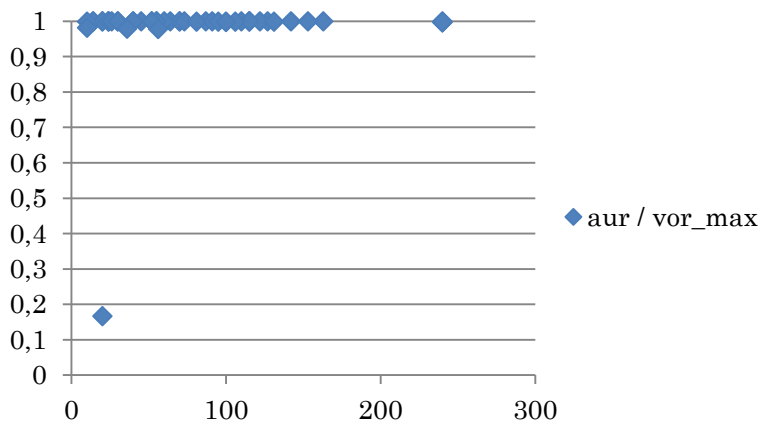


Figura 3. Comportamiento del desfase entre Valor Anti Utópico – valor óptimo real, maximizando, según crecen los elementos a ordenar. Fuente: elaboración propia.

Tabla 3. Relación Valor Anti Utópico – valor óptimo real. Fuente: elaboración propia.

dataset	n	aur	vor_max	aur - vor_max	aur / vor_max
06_03	14	174,89	174,89	0	1
06_04	14	179,33	179,33	0	1
06_11	20	360,890	360,890	0	1
06_12	20	370,22	370,22	0	1
06_18	24	539,56	539,56	0	1
06_28	24	509,78	509,56	0,22	0,999
06_46	30	841,14	840,86	0,28	0,999
06_48	24	533,330	533,330	0	1
11_01	240	51217,6	51078	139,6	0,997
14_01	10	59,16	59,08	0,08	0,998
15_01	240	50155,5	50130,5	25	0,999
15_02	240	51217,6	51078	139,6	0,997
15_07	110	10082	10079	3	0,999
15_09	115	10853	10848	5	0,999
15_12	100	8248	8241	7	0,999
15_14	163	21947,5	21945,5	2	0,999
15_16	70	3962	3962	0	1
15_17	127	13047	13044	3	0,999
15_18	115	10874,5	10874,5	0	1
15_19	87	6227,5	6224,5	3	0,999
15_20	122	11626,5	11620,5	6	0,999
15_23	142	16410	16409	1	0,999
15_24	91	6814,5	6813,5	1	0,999
15_25	115	10640,5	10638,5	2	0,999
15_27	95	7290,5	7282,5	8	0,998
15_29	106	9282	9276	6	0,999
15_30	64	3254,5	3253,5	1	0,999
15_32	153	19356,5	19346,5	10	0,999
15_34	55	2405,5	2405,5	0	1
15_40	131	13532	13523	9	0,999
15_41	70	3805	3805	0	1
15_42	100	7906	7889	17	0,997
15_44	45	1649	1649	0	1
15_46	40	1313	1313	0	1
15_48	10	71,33	70	1,33	0,981
15_50	26	501,5	501,5	0	1
15_54	60	2940,5	2940,5	0	1
15_55	52	2270	2270	0	1
15_57	73	4122,5	4121,5	1	0,999
15_59	55	2480	2479	1	0,999
15_65	40	1234,5	1234,5	0	1
15_66	52	2287,5	2287,5	0	1
15_67	30	722	722	0	1
15_69	81	5467	5463	4	0,999
15_73	36	1036	1014,67	21,33	0,979
15_74	20	327,330	54,67	272,7	0,167
15_77	56	2486	2432,67	53,33	0,978

Además, se puede apreciar que, como promedio, el óptimo es 0,98 del anti utópico por tanto se logró una muy buena estimación del óptimo. La Figura 4 muestra estos resultados.

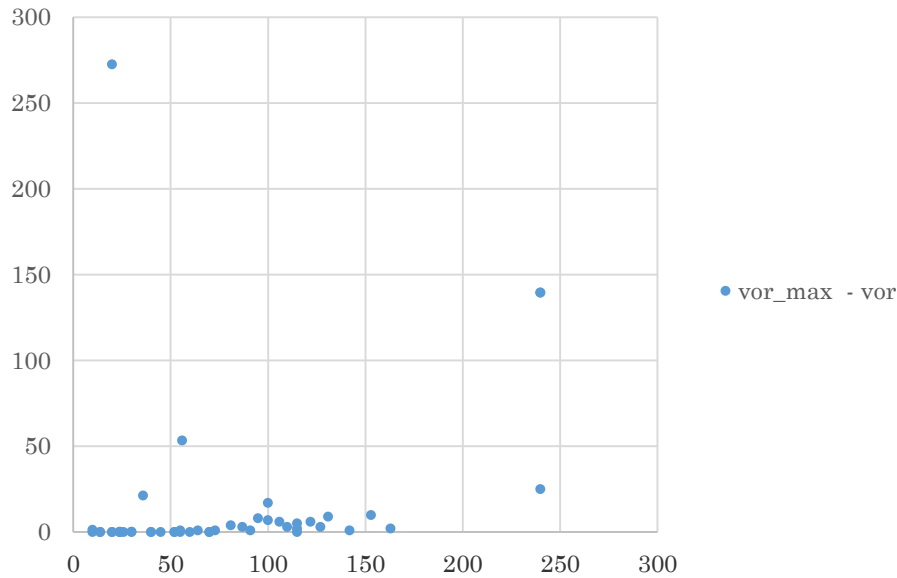


Figura 4. Diferencia proporcional entre el Valor Anti Utópico y el valor óptimo real. Fuente: elaboración propia.

La Figura 5 representa la relación entre los valores utópicos, anti utópico y óptimos reales.

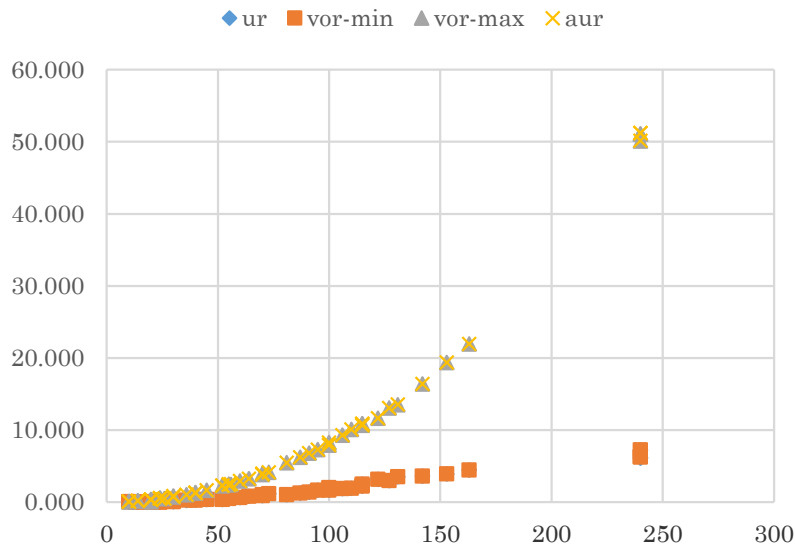


Figura 5. Relación entre Valor Utópico, Valor Anti Utópico, valor óptimo real minimizando y valor óptimo real maximizando. Fuente: elaboración propia.

En la Figura 6 se realiza una proporción llevando los valores utópicos, anti utópicos y reales a una escala [0,1], donde 0 es el Valor Utópico, 1 el anti utópico y ubica según la proporción el valor óptimo, lo cual permite ver con claridad que están acotados bien con respecto a los valores utópicos.

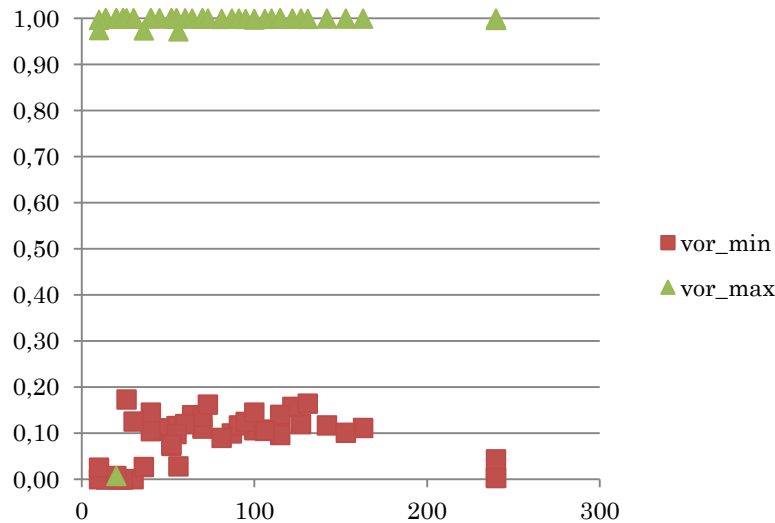


Figura 6. Relación entre Valor Utópico, Valor Anti Utópico y valor óptimo real a escala [0,1]
Fuente: elaboración propia.

5. CONCLUSIONES

En este trabajo, se extiende la investigación realizada en [14] y se realizó una nueva reformulación de los conceptos de Matriz Utópica y Anti Utópicas para el problema RAP.

En los experimentos se compararon los valores obtenidos de la matrices utópicas y anti utópicas con los valores óptimos del problema, donde se pudo demostrar la efectividad de los valores utópicos y anti utópicos como valores óptimos extremos (mínimos y máximos), y en ocasiones factibles, para medir el valor de calidad de la solución (ranking de consenso) del RAP sin empates.

La Matriz Utópica y la Matriz Anti Utópica son conceptos que se puede utilizar para evaluar las matrices de precedencia, que son la entrada de los RAP, por tanto, obtener una solución factible a partir de una matriz utópica es un problema interesante y abierto.

Además, es necesario conocer los casos en los que la Matriz Utópica y la Matriz Anti Utópica no son factibles, ya que ambos factores se vinculan directamente a la toma de decisiones, con el objetivo de ayudar a seleccionar los algoritmos correctos de acuerdo con las instancias del problema a resolver.

Además, encontrar rápidamente, con solo recorrer la matriz de entrada, una buena cota de los valores óptimos es muy interesante cuando se quiere resolver este tipo de problemas.

El hecho de que en algunos casos la Matriz Utópica refleje una matriz factible tiene varias implicaciones, y una importante es que el proceso de construir la Matriz Utópica se convierte en algunos casos en un algoritmo de solución del RAP, lo cual debe analizarse a profundidad en trabajos futuros.

6. AGRADECIMIENTOS

El trabajo no es producto de un proyecto económicamente soportado. Los autores agradecen a los revisores por todos los comentarios y sugerencias que han permitido mejorar el trabajo.

CONFLICTOS DE INTERÉS DE LOS AUTORES

Los autores reconocen no tener algún conflicto de intereses respecto al trabajo.

CONTRIBUCIÓN DE LOS AUTORES

Randy Reyna-Hernández: Conceptualización, curación de datos, análisis formal, investigación, metodología, administración del proyecto, software, recursos, validación, visualización y redacción.

Alejandro Rosete: Conceptualización, análisis formal, investigación, metodología, administración del proyecto, recursos, supervisión, validación y redacción.

7. REFERENCIAS

- [1] H. Ramírez-Murillo; C. A. Torres-Pinzón; E. F. Forero-García, “Photovoltaic Potential Estimation by Means of Data Mining in Four Colombian Cities,” *Tecnológicas*, vol. 22, no. 46, pp. 65–85, Sep. 2019. <https://doi.org/10.22430/22565337.1345>
- [2] F. Ganga-Contreras; J. López-Nunez; W. Sáez, “Portal de ranking de universidades iberoamericanas: una propuesta para facilitar procesos decisionales,” *Rev. Ibérica Sist. e Tecnol. Informação*, no. E25, pp. 472–488, Jan. 2020. [URL](https://doi.org/10.1145/371920.372165)
- [3] C. Dwork; R. Kumar; M. Naor; D. Sivakumar, “Rank Aggregation Methods for the Web,” in *Proceedings of the 10th International Conference on World Wide Web*, New York, 2001, pp. 613–622. <https://doi.org/10.1145/371920.372165>
- [4] L. J. Pérez Lugo, “Método para la agregación de rankings a partir de dos grupos con intereses contrapuestos,” (Tesis Doctorado), Facultad de Matemática, Física y Computación. Departamento de Ciencias de la Computación, Universidad Central “Marta Abreu” de Las Villas, 2015. [URL](https://doi.org/10.1137/1.9781611972771.66)
- [5] D. Sculley, “Rank Aggregation for Similar Items,” in *Proceedings of the Seventh SIAM International Conference on Data Mining, April 2007, Minneapolis, Minnesota, USA*, 2007, pp. 587–592. <https://doi.org/10.1137/1.9781611972771.66>
- [6] S. Chaudhuri; G. Das; V. Hristidis; G. Weikum, “Probabilistic Ranking of Database Query Results,” in *Proceedings of the Thirtieth International Conference on Very Large Data Bases, VLDB 2004*, Toronto, Canada 2004, pp. 888–899. <https://doi.org/10.1016/B978-012088469-8.50078-4>
- [7] G. Dahl; H. Minken, “A note on permutations and rank aggregation,” *Math. Comput. Model.*, vol. 52, no. 1–2, pp. 380–385, Jul. 2010. <https://doi.org/10.1016/j.mcm.2010.02.052>
- [8] H. L. Turner; J. van Etten; D. Firth; I. Kosmidis, “Modelling rankings in R: the PlackettLuce package,” *Comput. Stat.*, vol. 35, no. 3, pp. 1027–1057, Feb. 2020. <https://doi.org/10.1007/s00180-020-00959-3>
- [9] V. Pihur; S. Datta; S. Datta, “RankAggreg, an R package for weighted rank aggregation,” *BMC Bioinform.*, vol. 10, no. 62, Feb. 2009. <https://doi.org/10.1186/1471-2105-10-62>
- [10] A. Ali; M. Meila, “Experiments with Kemeny ranking: What works when?,” *Math. Soc. Sci.*, vol. 64, no. 1, pp. 28–40, Jul. 2012. <https://doi.org/10.1016/j.mathsocsci.2011.08.008>
- [11] A. Rosete, “Reformulación eficiente del problema de programación lineal de agregación de rankings,” *Ing. Ind.*, vol. 39, no. 3, Dic. 2018. [URL](https://doi.org/10.1145/1376616.1376625)
- [12] J. Feng; Q. Fang; W. Ng, “Discovering bucket orders from full rankings,” in *Proceedings of the ACM SIGMOD International Conference on Management of Data, SIGMOD 2008*, Vancouver 2008, pp. 55–66. <https://doi.org/10.1145/1376616.1376625>
- [13] A. Gionis; H. Mannila; K. Puolamäki; A. Ukkonen, “Algorithms for discovering bucket orders from data,” in *Proceedings of the Twelfth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, Philadelphia, PA, USA, August, 2006, pp. 561–566. <https://doi.org/10.1145/1150402.1150468>
- [14] J. A. Aledo; J. A. Gámez; A. Rosete, “Utopia in the solution of the Bucket Order Problem,” *Decis. Support Syst.*, vol. 97, pp. 69–80, May. 2017. <https://doi.org/10.1016/j.dss.2017.03.006>
- [15] J. A. Aledo; J. A. Gámez; A. Rosete, “Approaching rank aggregation problems by using evolution strategies: the case of the optimal bucket order problem,” *Eur. J. Oper. Res.*, vol. 270, no. 3, pp. 982–998, Nov. 2018. <https://doi.org/10.1016/j.ejor.2018.04.031>
- [16] J. A. Aledo; J. A. Gámez; D. Molina, “Approaching the rank aggregation problem by local search-based

- metaheuristics,” *J. Comput. Appl. Math.*, vol. 354, pp. 445–456, Jul. 2019. <https://doi.org/10.1016/j.cam.2018.03.014>
- [17] C. Dwork; R. Kumar; M. Naor; D. Sivakumar, “Rank aggregation revisited.” Manuscript, 2001. [URL](#)
- [18] J. A. Aledo; J. A. Gámez; D. Molina; A. Rosete, “Consensus-based journal rankings: A complementary tool for bibliometric evaluation,” *J. Assoc. Inf. Sci. Technol.*, vol. 69, no. 7, pp. 936–948, 2018. <https://doi.org/10.1002/asi.24040>
- [19] J. A. Aledo; J. A. Gámez; D. Molina, “Tackling the rank aggregation problem with evolutionary algorithms,” *Appl. Math. Comput.*, vol. 222, pp. 632–644, Oct. 2013. <https://doi.org/10.1016/j.amc.2013.07.081>
- [20] D. Molina García, “Contribuciones al problema de agregación de rankings. Aplicaciones al aprendizaje automático.” (Tesis Doctorales), Departamento de Matemáticas, Universidad de Castilla-La Mancha, 2015. [URL](#)
- [21] E. M. García Nové, “Nuevos problemas de agregación de rankings: Modelos y algoritmos,” (Tesis Doctorales), Departamento de Estadística, Matemáticas e Informática, Universidad Miguel Hernández de Elche, 2018. [URL](#)
- [22] W. D. Cook; M. Kress; L. M. Seiford, “An axiomatic approach to distance on partial orderings,” *RAIRO-Operations Res.*, vol. 20, no. 2, pp. 115–122, 1986. [URL](#)
- [23] T. Achterberg, “SCIP: solving constraint integer programs,” *Math. Program. Comput.*, vol. 1, no. 1, pp. 1–41, 2009. <https://doi.org/10.1007/s12532-008-0001-1>
- [24] Z. I. Berlin, “SCIP: solving constraint integer programs,” 2017. <https://www.scipopt.org/>
- [25] N. Mattei; T. Walsh, “PrefLib: A Library for Preferences <http://www.preflib.org>,” in *Algorithmic Decision Theory - Third International Conference, ADT 2013, Bruxelles, 2013*, vol. 8176, pp. 259–270. https://doi.org/10.1007/978-3-642-41575-3_20
- [26] R. Reyna-Hernández, “Herramientas y ficheros para replicar y analizar los experimentos.” 2021. [URL](#)