Procesamiento del Lenguaje Natural, Revista nº 65, septiembre de 2020, pp. 21-28 recibido 06-03-2020 revisado 14-06-2020 aceptado 14-06-2020

Un estudio de los métodos de reducción del frente de Pareto a una única solución aplicado al problema de resumen extractivo multi-documento

A study of methods for reducing the Pareto front to a single solution applied to the extractive multi-document summarization problem

Jesus M. Sanchez-Gomez[1], **Miguel A. Vega-Rodríguez**[1], **Carlos J. Pérez**[2]
[1]Dpto. de Tecnología de Computadores y Comunicaciones, Universidad de Extremadura
[2]Dpto. de Matemáticas, Universidad de Extremadura
jmsanchezgomez@unex.es, mavega@unex.es, carper@unex.es

Resumen: Los métodos de resumen automático son actualmente necesarios en muchos contextos diferentes. El problema de resumen extractivo multi-documento intenta cubrir el contenido principal de una colección de documentos y reducir la información redundante. La mejor manera de abordar esta tarea es mediante un enfoque de optimización multi-objetivo. El resultado de este enfoque es un conjunto de soluciones no dominadas o conjunto de Pareto. Sin embargo, dado que solo se necesita un resumen, se debe reducir el frente de Pareto a una única solución. Para ello, se han considerado varios métodos, como el mayor hipervolumen, la solución consenso, la distancia más corta al punto ideal y la distancia más corta a todos los puntos. Los métodos han sido probados utilizando conjuntos de datos de DUC, y han sido evaluados con las métricas ROUGE. Los resultados revelan que la solución consenso obtiene los mejores valores promedio
Palabras clave: Resumen extractivo multi-documento, Optimización multi-objetivo, Frente de Pareto, Solución única

Abstract: Automatic summarization methods are currently needed in many different contexts. The extractive multi-document summarization problem tries to cover the main content of a document collection and to reduce the redundant information. The best way to address this task is through a multi-objective optimization approach. The result of this approach is a set of non-dominated solutions or Pareto set. However, since only one summary is needed, the Pareto front must be reduced to a single solution. For this, several methods have been considered, such as the largest hypervolume, the consensus solution, the shortest distance to the ideal point, and the shortest distance to all points. The methods have been tested using datasets from DUC, and they have been evaluated with ROUGE metrics. The results show that consensus solution achieves the best average values
Keywords: Extractive multi-document summarization, Multi-objective optimization, Pareto front, Single solution

1 Introducción

Hoy en día, la información en Internet crece exponencialmente, y obtener la más importante sobre un tema concreto es de interés en muchas áreas. Extraer la información más relevante es posible mediante las herramientas de minería de texto, las cuales son capaces de generar un resumen automático a partir de información textual (Hashimi, Hafez, y Mathkour, 2015).

Existen varios tipos de resumen. Dependiendo de dónde se obtiene la información, un resumen puede ser mono-documento o multi-documento (Zajic, Dorr, y Lin, 2008): los mono-documento reducen la información de un único documento, y los multi-documento seleccionan la información de una colección. Un resumen también puede ser abstractivo, donde las palabras y oraciones que lo forman pueden no existir en el texto original, o extractivo, donde sus oraciones sí existen en la fuente original (Wan, 2008).

El problema del resumen extractivo multi-documento puede ser formulado como un pro-

ISSN 1135-5948. DOI 10.26342/2020-65-2

blema de optimización tanto mono-objetivo como multi-objetivo. En la optimización mono-objetivo solo se optimiza una única función objetivo, la cual incluye todos los criterios de forma ponderada (Alguliev, Aliguliyev, y Mehdiyev, 2011). Por otro lado, en la optimización multi-objetivo todas las funciones objetivo se optimizan de manera simultánea. Además, la optimización multi-objetivo ha logrado mejores resultados que la mono-objetivo (Saleh, Kadhim, y Attea, 2015). Las funciones objetivo utilizadas en estos trabajos fueron la cobertura del contenido y la reducción de la redundancia.

En la optimización multi-objetivo la solución generada no es única, sino que es un conjunto de soluciones no dominadas denominado conjunto de Pareto (Sudeng y Wattanapongsakorn, 2015). Los métodos de reducción del frente de Pareto pueden clasificarse en tres grupos. Primero, los métodos basados en preferencias, que necesitan que el usuario las asigne de forma previa (Antipova et al., 2015). Segundo, los métodos de *clustering*, que seleccionan uno o varios subconjuntos de soluciones del frente mediante el uso de técnicas basadas en similitud, como en Taboada y Coit (2007). Y tercero, los métodos basados en distancias seleccionan una única solución del frente basándose en la distancia al punto ideal (Padhye y Deb, 2011).

En este trabajo se han estudiado y comparado varios métodos automáticos de reducción del frente de Pareto a una única solución aplicados al problema de resumen extractivo multi-documento. Estos métodos se han basado en los conceptos del hipervolumen, de la solución consenso y de varios tipos de distancias. La experimentación ha sido realizada con los conjuntos de datos de DUC (*Document Understanding Conferences*), y los resultados se han evaluado con las métricas ROUGE (*Recall-Oriented Understudy for Gisting Evaluation*). Además, también se ha realizado un análisis estadístico de los resultados obtenidos.

2 Trabajo relacionado

A continuación se analizan los métodos más usados para reducir el frente de Pareto.

En primer lugar, los métodos basados en preferencias de usuario necesitan que los usuarios asignen sus preferencias a priori para obtener un conjunto reducido de soluciones. Estas preferencias suelen estar relacionadas con los pesos de las funciones objetivo. El método de función de estrés ponderado propuesto por Ferreira, Fonseca, y Gaspar-Cunha (2007) integra las preferencias del usuario para encontrar la mejor región del frente de Pareto de acuerdo con estas preferencias. De la misma manera, el método de utilidades aditivas discriminatorias (Soylu y Ulusoy, 2011) requiere que el usuario asigne algunas referencias que reflejen sus preferencias. Otro método es el basado en el filtro de Pareto, desarrollado por Antipova et al. (2015), usado para reducir y facilitar el análisis post-óptimo con la idea de clasificar las soluciones de acuerdo a una eficiencia global y seleccionar aquellas que muestren un mejor equilibrio entre los objetivos.

En segundo lugar, los métodos de *clustering* dividen el frente de Pareto en varios grupos. Estos métodos comienzan con un número concreto de grupos, calculando el centroide para cada uno de ellos de forma iterativa y agrupando los puntos del frente de Pareto con el centroide más similar. El algoritmo de partición de grupos k-means (Taboada y Coit, 2007) proporciona conjuntos más pequeños de soluciones intermedias óptimas, calculando el centroide de cada grupo y asignando cada solución al grupo con el centroide más cercano. Aguirre y Taboada (2011) usaron el enfoque de árbol de crecimiento dinámico auto-organizado para reducir de forma inteligente el tamaño del conjunto, obteniendo así soluciones representativas. Además, la técnica del agrupamiento jerárquico consiste en la formación de subconjuntos de acuerdo a las decisiones de diseño (Veerappa y Letier, 2011).

Por último, los métodos basados en distancias seleccionan el punto del frente de Pareto que tiene la distancia más corta al punto ideal, que representa una solución en la que los valores de las funciones objetivo son óptimos. Padhye y Deb (2011) utilizaron el método de la métrica L_2 para seleccionar la solución con la distancia Euclídea más corta al punto ideal. Del mismo modo, Siwale (2013) presentó la solución de compromiso Chebyshev, que utiliza el criterio de la cercanía a un punto ideal basándose en esta distancia.

Los métodos basados en preferencias de usuario y los basados en *clustering* no seleccionan una única solución. Además, los métodos basados en distancias que se han revisado solo tienen en cuenta dos tipos de distancias.

En este trabajo se han evaluado otros tipos de distancias, como la Manhattan, la Mahalanobis y la Levenshtein, además de otros métodos, como el del mayor hipervolumen y la solución consenso.

3 Definición del problema

El problema de resumen extractivo multi-documento se formula a continuación como un problema de optimización multi-objetivo. Los métodos más usados en este campo son los basados en vectores de palabras. En ellos, una oración se representa como un vector de palabras, y para medir la similitud entre oraciones se utiliza un criterio particular, siendo la similitud coseno el más utilizado.

3.1 La similitud coseno

Dado el conjunto $T = \{t_1, t_2, ..., t_m\}$ que contiene los m términos distintos existentes en la colección de documentos D. Cada oración $s_i \in D$ se representa como $s_i = (w_{i1}, w_{i2}, ..., w_{im}), i = 1, 2, ..., n$, siendo n el número de oraciones en D. Cada componente w_{ik} representa el peso del término t_k en la oración s_i, que se calcula mediante el esquema tf_isf (*term-frequency inverse-sentence-frequency*) de la siguiente forma:

$$w_{ik} = tf_{ik} \cdot log(n/n_k) \tag{1}$$

donde tf_{ik} cuenta cuántas veces está el término t_k en la oración s_i, y $log(n/n_k)$ es el factor isf, siendo n_k el número de oraciones que contienen el término t_k.

El contenido principal de D puede representarse como la media de los pesos de los m términos en T mediante un vector de medias $O = (o_1, o_2, ..., o_m)$, cuyas componentes se calculan como:

$$o_k = \frac{1}{n} \sum_{i=1}^{n} w_{ik} \tag{2}$$

La similitud coseno se basa en los pesos definidos previamente. Concretamente, mide la semejanza entre el par de oraciones $s_i = (w_{i1}, w_{i2}, ..., w_{im})$ y $s_j = (w_{j1}, w_{j2}, ..., w_{jm})$ de la siguiente forma:

$$sim(s_i, s_j) = \frac{\sum_{k=1}^{m} w_{ik} w_{jk}}{\sqrt{\sum_{k=1}^{m} w_{ik}^2 \cdot \sum_{k=1}^{m} w_{jk}^2}} \tag{3}$$

3.2 La optimización multi-objetivo

La colección de documentos $D = \{d_1, d_2, ..., d_N\}$, que contiene N documentos, también puede representarse como un conjunto de n oraciones: $D = \{s_1, s_2, ..., s_n\}$. El fin es generar un resumen $R \subset D$ que tenga en cuenta los siguientes tres aspectos:

- Cobertura del contenido. El resumen debe cubrir el contenido principal de la colección de documentos.
- Reducción de la redundancia. El resumen no debe contener oraciones de la colección de documentos similares entre sí.
- Longitud. El resumen debe tener una longitud prefijada L.

El problema de resumen extractivo multi-documento implica la optimización simultánea de la cobertura del contenido y de la reducción de la redundancia. Sin embargo, estos criterios son contradictorios entre sí. Por lo tanto, la mejor forma de resolver este problema es mediante un enfoque de optimización multi-objetivo.

Antes es necesario definir la representación de una solución, $X = (x_1, x_2, ..., x_n)$, donde x_i es una variable binaria que tiene en cuenta la presencia o ausencia ($x_i = 1$ o $x_i = 0$) de la oración s_i en el resumen R.

El primer objetivo a optimizar, $\Phi_{CC}(X)$, es el relativo al criterio de la cobertura del contenido. Dadas las oraciones $s_i \in R$, este objetivo se representa como la similitud entre las oraciones s_i y todas las oraciones en la colección D, representada por el vector medio O. Por lo tanto, la siguiente función debe ser maximizada:

$$\Phi_{CC}(X) = \sum_{i=1}^{n} sim(s_i, O) \cdot x_i \tag{4}$$

El segundo objetivo a optimizar, $\Phi_{RR}(X)$, se refiere a la reducción de la redundancia. En este caso, se necesita otra variable binaria y_{ij} que relacione la presencia o ausencia simultánea ($y_{ij} = 1$ o $y_{ij} = 0$) del par de oraciones s_i y s_j en el resumen R. Para cada par de oraciones $s_i, s_j \in R$, su similitud $sim(s_i, s_j)$ debe ser minimizada, lo que es equivalente a maximizar la siguiente función:

$$\Phi_{RR}(X) = \frac{1}{\left(\sum_{i=1}^{n-1} \sum_{j=i+1}^{n} sim(s_i, s_j) \cdot y_{ij}\right) \cdot \sum_{i=1}^{n} x_i} \tag{5}$$

Tras la definición de los objetivos, se puede formular el problema de optimización multi-objetivo:

$$\text{máx } \Phi(X) = \{\Phi_{CC}(X), \Phi_{RR}(X)\} \tag{6}$$

$$\text{sujeto a } L - \varepsilon \leq \sum_{i=1}^{n} l_i \cdot x_i \leq L + \varepsilon \qquad (7)$$

donde l_i es la longitud de la oración s_i y ε es la tolerancia de la longitud del resumen:

$$\varepsilon = \max_{i=1,2,\ldots,n} l_i - \min_{i=1,2,\ldots,n} l_i \qquad (8)$$

4 Métodos de reducción del frente de Pareto a una única solución

En esta sección se presentan los diferentes métodos estudiados para reducir el frente de Pareto a una única solución.

4.1 Mayor hipervolumen

El hipervolumen mide el espacio cubierto por los valores de las funciones objetivo correspondientes a cada solución del frente de Pareto (Beume et al., 2009). Al tratarse de un problema bidimensional, el hipervolumen es el área cubierta por cada punto (solución). Este método selecciona la solución asociada al punto con mayor hipervolumen (MH) (ver Figura 1).

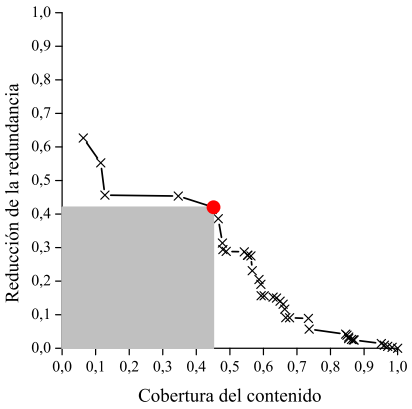

Figura 1: Mayor hipervolumen

4.2 Solución consenso

La solución consenso (SC) es aquella que se genera a partir de todas las soluciones del frente de Pareto (Pérez et al., 2017). En este problema se forma con las oraciones de los resúmenes asociados, seleccionando las oraciones más utilizadas hasta alcanzar la restricción de longitud $L \pm \epsilon$ indicada en la Ecuación (7). Al tratarse de un resumen generado a partir de otros, su solución asociada puede no existir en el frente (ver Figura 2).

4.3 Distancia más corta al punto ideal

El punto ideal es aquel en el que los valores de los objetivos son los mejores posibles.

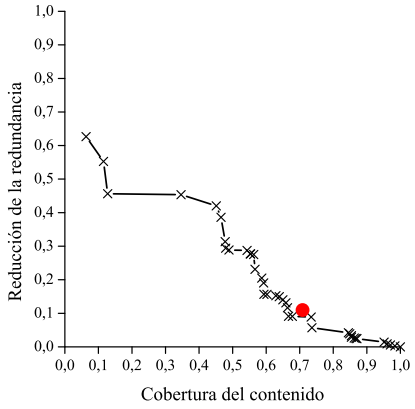

Figura 2: Solución consenso

En un espacio objetivo normalizado, con rango [0,1], donde las funciones objetivo deben ser maximizadas, el punto ideal está situado en (1,1). Por lo tanto, este método (DCPI) mide la distancia entre cada punto del frente y el punto ideal, y selecciona el que tiene la distancia más corta (ver la Figura 3, que se basa en la distancia Euclídea). Dados dos puntos del espacio $P_1 = (p_{11}, p_{12})$ y $P_2 = (p_{21}, p_{22})$, la distancia d entre P_1 y P_2 se define como $d(P_1, P_2)$. Además de las distancias Euclídea (DCPI_E) y Chebyshev (DCPI_C), también se han estudiado las distancias Manhattan (DCPI_M) y Mahalanobis (DCPI_B).

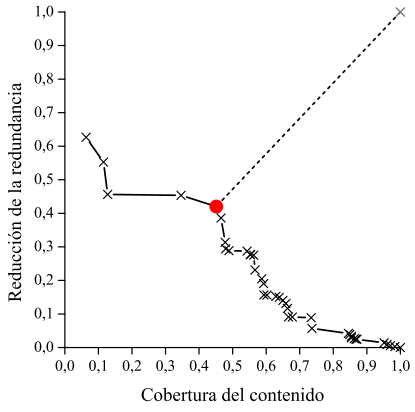

Figura 3: Distancia más corta al punto ideal

En primer lugar, la distancia Euclídea es la distancia ordinaria entre dos puntos en un espacio Euclídeo (Padhye y Deb, 2011):

$$d_E(P_1, P_2) = \sqrt{\sum_{i=1}^{2} (p_{1i} - p_{2i})^2} \qquad (9)$$

En segundo lugar, la distancia Manhattan está basada en la geografía callejera cuadri-

culada del distrito de Manhattan (Wu et al., 2015). Se define como la suma de las longitudes de las proyecciones en los ejes de coordenadas del segmento de línea entre dos puntos:

$$d_M(P_1, P_2) = \|P_1, P_2\| = \sum_{i=1}^{2} |p_{1i} - p_{2i}| \quad (10)$$

En tercer lugar, la distancia Chebyshev es la mayor de las longitudes de las proyecciones en los ejes de coordenadas del segmento de línea entre dos puntos (Siwale, 2013):

$$d_C(P_1, P_2) = \max_{i=1,2} |p_{1i} - p_{2i}| \quad (11)$$

Y en cuarto lugar, la distancia Mahalanobis determina la similitud entre dos variables aleatorias multi-dimensionales, teniendo en cuenta sus correlaciones (Zhao et al., 2017). En este contexto se define como:

$$d_B(P_1, P_2) = \sqrt{\sum_{i=1}^{2} \left(\frac{p_{1i} - p_{2i}}{\sigma_i} \right)^2} \quad (12)$$

donde σ_i es la desviación estándar no corregida, que se calcula midiendo la dispersión de los valores de la función objetivo i a través de los G puntos del frente de Pareto:

$$\sigma_i = \frac{1}{G} \sum_{g=1}^{G} (p_{gi} - \bar{p_i})^2 \quad (13)$$

siendo $\bar{p_i} = \frac{1}{G} \sum_{g=1}^{G} p_{gi}$.

4.4 Distancia más corta a todos los puntos

Este método (DCTP) evalúa la suma de las distancias entre un punto y todos los restantes, y selecciona el punto del frente con la menor distancia (ver Figura 4). Para este método, además de las cuatro distancias definidas anteriormente (Euclídea DCTP_E, Chebyshev DCTP_C, Manhattan DCTP_M y Mahalanobis DCTP_B), también se ha considerado la distancia Levenshtein (DCTP_L) (Ristad y Yianilos, 1998). En este contexto se define como el número de inserciones y eliminaciones de oraciones necesario para transformar un resumen en otro. Dados dos resúmenes (o conjuntos de oraciones) $\mathcal{R}$ y $\mathcal{S}$, la distancia de Levenshtein entre ellos se calcula como:

$$d_L(\mathcal{R}, \mathcal{S}) = |\mathcal{R}| + |\mathcal{S}| - 2 \cdot |\mathcal{R} \cap \mathcal{S}| \quad (14)$$

donde $|\cdot|$ es la cardinalidad o número de elementos en un conjunto.

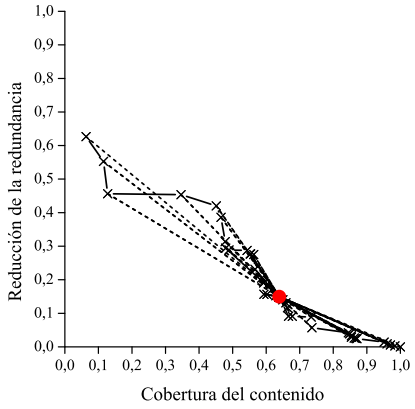

Figura 4: Distancia más corta a todos los puntos

5 Resultados

5.1 Conjuntos de datos

Para la ejecución de los experimentos se han utilizado los conjuntos de datos DUC, que es un banco de pruebas para la evaluación de resúmenes automáticos. En este estudio se han utilizado 10 temas (del d061j al d070f) del conjunto de datos DUC2002 (NIST, 2014). La longitud de resumen establecida en estos conjuntos es de 200 palabras.

5.2 Métricas de evaluación

El rendimiento de los métodos ha sido evaluado con las métricas ROUGE. Estas métricas son las que se utilizan en DUC (Lin, 2004). ROUGE mide la similitud entre un resumen generado automáticamente y un resumen generado por un experto. Las puntuaciones ROUGE empleadas han sido el ROUGE-2 (R-2) y el ROUGE-L (R-L), y se ha usado la media aritmética para medir la tendencia central en cada tema.

Además, se ha realizado un análisis estadístico de las puntuaciones ROUGE mediante pruebas de hipótesis adecuadas. Los p-valores menores que 0,05 han sido considerados como estadísticamente significativos.

5.3 Configuración

Sanchez-Gomez, Vega-Rodríguez, y Pérez (2018) propusieron y aplicaron un enfoque de optimización multi-objetivo basado en el algoritmo de colonia de abejas artificiales aplicado al problema de resumen extractivo multi-documento. Los siguientes parámetros produjeron buenos resultados: tamaño de población = 50; número de ciclos = 1000; y probabilidad de mutación = 0,1.

Método	d061j	d062j	d063j	d064j	d065j	d066j	d067f	d068f	d069f	d070f	Media
MH	**0,297**	0,186	0,212	0,201	0,112	0,189	0,273	0,242	0,131	0,099	0,194
SC	0,268	**0,400**	0,199	**0,203**	0,167	0,195	0,273	0,277	**0,254**	0,172	**0,241**
DCPI$_{\mathbf{E}}$	**0,297**	0,191	0,205	0,168	0,115	0,059	0,231	0,242	0,233	0,099	0,184
DCPI$_{\mathbf{M}}$	0,248	0,295	**0,256**	0,187	0,168	0,063	0,172	0,208	0,195	**0,258**	0,205
DCPI$_{\mathbf{C}}$	0,222	0,140	0,183	0,197	0,124	0,189	**0,353**	0,260	0,145	0,101	0,191
DCPI$_{\mathbf{B}}$	**0,297**	0,190	0,153	0,170	0,143	0,059	0,231	**0,318**	0,111	0,098	0,177
DCTP$_{\mathbf{E}}$	0,219	0,245	0,192	0,175	0,163	0,203	0,272	0,291	0,119	0,111	0,199
DCTP$_{\mathbf{M}}$	0,232	0,270	0,185	0,169	0,164	**0,221**	0,268	0,286	0,122	0,112	0,203
DCTP$_{\mathbf{C}}$	0,222	0,242	0,196	0,169	0,164	0,202	0,284	0,282	0,118	0,117	0,200
DCTP$_{\mathbf{B}}$	0,219	0,245	0,191	0,175	0,166	0,203	0,274	0,292	0,117	0,111	0,199
DCTP$_{\mathbf{L}}$	0,272	0,232	0,175	0,167	**0,177**	0,196	0,268	0,261	0,148	0,130	0,203

Tabla 1: Resultados para ROUGE-2 (en negrita los mejores valores)

Método	d061j	d062j	d063j	d064j	d065j	d066j	d067f	d068f	d069f	d070f	Media
MH	**0,541**	0,342	0,469	**0,404**	0,366	0,428	0,508	0,450	0,321	0,402	0,423
SC	0,509	**0,571**	0,458	0,390	**0,407**	0,461	0,517	0,506	0,517	0,480	**0,482**
DCPI$_{\mathbf{E}}$	0,540	0,347	0,459	0,347	0,375	0,278	0,417	0,450	**0,536**	0,402	0,414
DCPI$_{\mathbf{M}}$	0,508	0,483	**0,489**	0,373	0,393	0,281	0,413	0,450	0,461	**0,481**	0,433
DCPI$_{\mathbf{C}}$	0,494	0,291	0,414	0,400	0,375	0,425	**0,558**	0,465	0,380	0,406	0,421
DCPI$_{\mathbf{B}}$	0,540	0,345	0,357	0,312	0,341	0,279	0,417	**0,531**	0,256	0,402	0,378
DCTP$_{\mathbf{E}}$	0,481	0,463	0,440	0,368	0,390	0,458	0,525	0,518	0,441	0,435	0,452
DCTP$_{\mathbf{M}}$	0,492	0,477	0,430	0,365	0,390	**0,468**	0,518	0,512	0,449	0,440	0,454
DCTP$_{\mathbf{C}}$	0,438	0,462	0,445	0,364	0,391	0,460	0,532	0,507	0,443	0,440	0,453
DCTP$_{\mathbf{B}}$	0,481	0,463	0,438	0,368	0,392	0,458	0,525	0,518	0,443	0,435	0,452
DCTP$_{\mathbf{L}}$	0,518	0,406	0,384	0,349	0,395	0,456	0,488	0,485	0,322	0,431	0,423

Tabla 2: Resultados para ROUGE-L (en negrita los mejores valores)

5.4 Resultados

Los métodos han sido evaluados con los frentes de Pareto resultantes de las 20 repeticiones ejecutadas para cada tema. Las Tablas 1 y 2 contienen los resultados para R-2 y R-L respectivamente, mostrando que la solución consenso (SC) ha conseguido el mejor valor medio en ambos ROUGE. También es el método que ha obtenido el mejor valor en mayor número de temas (3 en R-2 y 2 en R-L). Los métodos de la distancia más corta al punto ideal y a todos los puntos con la distancia Manhattan (DCPI$_{\mathrm{M}}$ y DCTP$_{\mathrm{M}}$) han logrado el segundo mejor valor medio (DCPI$_{\mathrm{M}}$ en R-2 y DCTP$_{\mathrm{M}}$ en R-L). Además, DCPI$_{\mathrm{M}}$ y DCTP$_{\mathrm{M}}$ han obtenido el mejor valor en 2 temas y en 1 tema en ambos ROUGE, respectivamente. El método del mayor hipervolumen (MH) ha obtenido el mejor resultado en 1 tema en R-2 y en 2 temas en R-L. En cuanto a los métodos basados en distancias, y agrupando todas ellas, DCPI ha obtenido el mejor resultado en 5 temas en ambos ROUGE, mientras que DCTP los ha obtenido en 2 temas para R-2 y en 1 tema para R-L.

Una vez analizados los resultados de las puntuaciones ROUGE, se muestra el análisis estadístico. Se han aplicado pruebas de hipótesis estadísticas para analizar si existen diferencias estadísticamente significativas entre los diferentes métodos. Las condiciones de normalidad y homokedasticidad para el ANOVA uni-factorial no se pueden asumir para ningún tema, por lo que se ha utilizado el test de Kruskal-Wallis. Los p-valores son inferiores a 0,001 para todos los temas, por lo que al menos un par de métodos son significativamente diferentes. Las comparaciones por pares se han realizado entre el método de la solución consenso y el resto de los métodos mediante el test de Conover con la corrección de Bonferroni. Las Tablas 3 y 4 muestran las diferencias estadísticamente significativas, y de ellas se concluye que el método de la solución consenso (SC) aporta diferencias estadísticamente significativas con respecto al resto. Concretamente, SC ha obtenido diferencias estadísticamente significativas con el 80 % de los métodos en R-2 en al menos la mitad de los temas. En cuanto a R-L, SC ha obtenido diferencias estadísticamente significativas con todos los métodos en al menos la mitad de los temas.

Para terminar, la Tabla 5 muestra los por-

Método	d061j	d062j	d063j	d064j	d065j	d066j	d067f	d068f	d069f	d070f	Total
MH	**<0,001**	**<0,001**	0,056	0,077	**<0,001**	**<0,001**	1,000	**<0,001**	**<0,001**	**<0,001**	7
DCPI$_E$	**<0,001**	**<0,001**	1,000	**<0,001**	**<0,001**	**<0,001**	**<0,001**	**<0,001**	1,000	**<0,001**	8
DCPI$_M$	1,000	**<0,001**	**0,002**	1,000	1,000	**<0,001**	**<0,001**	**<0,001**	**0,002**	1,000	6
DCPI$_C$	0,057	**<0,001**	**0,005**	0,429	**<0,001**	**0,006**	**<0,001**	**0,008**	**0,025**	**<0,001**	8
DCPI$_B$	**<0,001**	**<0,001**	**<0,001**	**<0,001**	0,683	**<0,001**	**<0,001**	1,000	**<0,001**	**<0,001**	8
DCTP$_E$	**<0,001**	**<0,001**	1,000	0,202	1,000	0,415	1,000	1,000	**<0,001**	**<0,001**	4
DCTP$_M$	**0,007**	**<0,001**	**0,017**	**<0,001**	1,000	**<0,001**	1,000	1,000	**<0,001**	**<0,001**	7
DCTP$_C$	**0,002**	**<0,001**	1,000	**<0,001**	1,000	0,197	1,000	1,000	**<0,001**	**<0,001**	5
DCTP$_B$	**<0,001**	**<0,001**	1,000	0,202	1,000	0,619	1,000	1,000	**<0,001**	**<0,001**	4
DCTP$_L$	1,000	**<0,001**	**<0,001**	**<0,001**	1,000	1,000	1,000	0,442	**<0,001**	**0,003**	5

Tabla 3: p-valores obtenidos de la comparación por pares entre el método de la solución consenso y el resto de métodos para ROUGE-2 (en negrita las diferencias estadísticamente significativas)

Método	d061j	d062j	d063j	d064j	d065j	d066j	d067f	d068f	d069f	d070f	Total
MH	**<0,001**	**<0,001**	1,000	0,112	**<0,001**	**<0,001**	1,000	**<0,001**	**<0,001**	**<0,001**	7
DCPI$_E$	**<0,001**	**<0,001**	1,000	**<0,001**	**<0,001**	**<0,001**	**<0,001**	**<0,001**	1,000	**<0,001**	8
DCPI$_M$	0,457	**<0,001**	**0,024**	1,000	**0,027**	**<0,001**	**<0,001**	**<0,001**	1,000	1,000	6
DCPI$_C$	1,000	**<0,001**	**<0,001**	0,900	**<0,001**	**<0,001**	**<0,001**	**<0,001**	**<0,001**	**<0,001**	8
DCPI$_B$	**<0,001**	**<0,001**	**<0,001**	**<0,001**	**<0,001**	**<0,001**	**<0,001**	1,000	**<0,001**	**<0,001**	9
DCTP$_E$	**<0,001**	**<0,001**	0,419	**0,011**	0,275	1,000	1,000	1,000	**<0,001**	**<0,001**	5
DCTP$_M$	0,134	**<0,001**	**0,003**	**0,004**	0,205	1,000	1,000	1,000	**<0,001**	**<0,001**	5
DCTP$_C$	**0,005**	**<0,001**	1,000	**<0,001**	0,303	1,000	0,635	1,000	**<0,001**	**<0,001**	5
DCTP$_B$	**<0,001**	**<0,001**	0,206	**0,027**	0,599	1,000	1,000	1,000	**<0,001**	**<0,001**	5
DCTP$_L$	1,000	**<0,001**	**<0,001**	**<0,001**	1,000	1,000	1,000	**0,048**	**<0,001**	**<0,001**	6

Tabla 4: p-valores obtenidos de la comparación por pares entre el método de la solución consenso y el resto de métodos para ROUGE-L (en negrita las diferencias estadísticamente significativas)

Método	MH	DCPI$_E$	DCPI$_M$	DCPI$_C$	DCPI$_B$	DCTP$_E$	DCTP$_M$	DCTP$_C$	DCTP$_B$	DCTP$_L$	Media
ROUGE-2	24,23	30,98	17,56	26,18	36,16	21,11	18,72	20,50	21,11	18,72	23,53
ROUGE-L	13,95	16,43	11,32	14,49	27,51	6,64	6,17	6,40	6,64	13,95	12,35

Tabla 5: Porcentajes de mejora obtenidos por el método de la solución consenso

centajes de mejora obtenidos por la solución consenso con respecto al resto de métodos. Los porcentajes de mejora obtenidos por la solución consenso varían del 17,56 % al 36,16 % para R-2 y del 6,17 % al 27,51 % para R-L. Además, los porcentajes de mejora global obtenidos son del 23,53 % y del 12,35 % en R-2 y R-L, respectivamente.

6 Conclusiones

El problema del resumen extractivo multi-documento se ha resuelto aplicando optimización multi-objetivo. Como el resultado es un conjunto de soluciones no dominadas o frente de Pareto, el objetivo de este trabajo ha sido llevar a cabo un análisis para seleccionar un único resumen del frente. Para ello, se han implementado, evaluado y comparado diferentes métodos basados en el mayor hipervolumen, en la solución consenso, en la distancia más corta al punto ideal y en la distancia más corta a todos los puntos, evaluando para estos dos últimos varios tipos de distancias. Los resultados indican que la solución consenso es el método que mejores resultados ha obtenido. Los porcentajes de mejora global alcanzados son del 23,53 % para ROUGE-2 y del 12,35 % para ROUGE-L.

Agradecimientos

Esta investigación ha sido apoyada por la Agencia Estatal de Investigación (proyectos PID2019-107299GB-I00 y MTM2017-86875-C3-2-R), por la Junta de Extremadura (proyectos GR18090 y GR18108) y por la Unión Europea (Fondo Europeo de Desarrollo Regional). Jesus M. Sanchez-Gomez está apoyado por un Contrato Predoctoral financiado por la Junta de Extremadura (contrato PD18057) y la Unión Europea (Fondo Social Europeo).

Bibliografía

Aguirre, O. y H. Taboada. 2011. A Clustering Method Based on Dynamic Self Organizing Trees for Post-Pareto Optima-

lity Analysis. *Procedia Computer Science*, 6:195–200.

Alguliev, R. M., R. M. Aliguliyev, y C. A. Mehdiyev. 2011. Sentence selection for generic document summarization using an adaptive differential evolution algorithm. *Swarm Evol. Comput.*, 1(4):213–222.

Antipova, E., C. Pozo, G. Guillén-Gosálbez, D. Boer, L. F. Cabeza, y L. Jiménez. 2015. On the use of filters to facilitate the post-optimal analysis of the Pareto solutions in multi-objective optimization. *Comput. Chem. Eng.*, 74:48–58.

Beume, N., C. M. Fonseca, M. López-Ibáñez, L. Paquete, y J. Vahrenhold. 2009. On the Complexity of Computing the Hypervolume Indicator. *IEEE Trans. Evol. Comput.*, 13(5):1075–1082.

Ferreira, J. C., C. M. Fonseca, y A. Gaspar-Cunha. 2007. Methodology to Select Solutions from the Pareto-Optimal Set: A Comparative Study. En *Proceedings of the 9th Annual Conference on Genetic and Evolutionary Computation*, páginas 789–796. ACM.

Hashimi, H., A. Hafez, y H. Mathkour. 2015. Selection criteria for text mining approaches. *Comput. Hum. Behav.*, 51:729–733.

Lin, C.-Y. 2004. ROUGE: A package for automatic evaluation of summaries. En *Proceedings of the ACL-04 Workshop*, volumen 8, páginas 74–81. ACL.

NIST. 2014. Document Understanding Conferences. http://duc.nist.gov. Último acceso: 11 de agosto de 2020.

Padhye, N. y K. Deb. 2011. Multi-objective optimisation and multi-criteria decision making in SLS using evolutionary approaches. *Rapid Prototyping Journal*, 17(6):458–478.

Pérez, C. J., M. A. Vega-Rodríguez, K. Reder, y M. Flörke. 2017. A Multi-Objective Artificial Bee Colony-based optimization approach to design water quality monitoring networks in river basins. *Journal of Cleaner Production*, 166:579–589.

Ristad, E. S. y P. N. Yianilos. 1998. Learning string-edit distance. *IEEE Trans. Pattern Anal. Mach. Intell.*, 20(5):522–532.

Saleh, H. H., N. J. Kadhim, y B. A. Attea. 2015. A genetic based optimization model for extractive multi-document text summarization. *Iraqi Journal of Science*, 56(2):1489–1498.

Sanchez-Gomez, J. M., M. A. Vega-Rodríguez, y C. J. Pérez. 2018. Extractive multi-document text summarization using a multi-objective artificial bee colony optimization approach. *Knowledge-Based Syst.*, 159:1–8.

Siwale, I. 2013. Practical Multi-Objective Programming. Informe técnico, Technical Report RD-14-2013. APEX Research.

Soylu, B. y S. K. Ulusoy. 2011. A preference ordered classification for a multi-objective max–min redundancy allocation problem. *Comput. Oper. Res.*, 38(12):1855–1866.

Sudeng, S. y N. Wattanapongsakorn. 2015. Post Pareto-optimal pruning algorithm for multiple objective optimization using specific extended angle dominance. *Eng. Appl. Artif. Intell.*, 38:221–236.

Taboada, H. A. y D. W. Coit. 2007. Data Clustering of Solutions for Multiple Objective System Reliability Optimization Problems. *Qual. Technol. Quant. Manag.*, 4(2):191–210.

Veerappa, V. y E. Letier. 2011. Understanding Clusters of Optimal Solutions in Multi-Objective Decision Problems. En *19th Requirements Engineering Conference*, páginas 89–98. IEEE.

Wan, X. 2008. An exploration of document impact on graph-based multi-document summarization. En *Proceedings of the Conference on Empirical Methods in NLP*, páginas 755–762. ACL.

Wu, L., X. Xu, X. Ye, y X. Zhu. 2015. Repeat and near-repeat burglaries and offender involvement in a large Chinese city. *Cartogr. Geogr. Inf. Sci.*, 42(2):178–189.

Zajic, D. M., B. J. Dorr, y J. Lin. 2008. Single-document and multi-document summarization techniques for email threads using sentence compression. *Inf. Process. Manage.*, 44(4):1600–1610.

Zhao, L., Z. Lu, W. Yun, y W. Wang. 2017. Validation metric based on Mahalanobis distance for models with multiple correlated responses. *Reliab. Eng. Syst. Saf.*, 159:80–89.