Shape-based visual analysis of solutions for Multiobjective Optimization problems

Christian von Lücken
Facultad Politécnica
Universidad Nacional de Asunción
Email: clucken@pol.una.py

Abrahan D. Fretes Prieto
Facultad Politécnica
Universidad Nacional de Asunción
Email: abrahan.fretes@gmail.com

Arsenio M. Ferreira Vera y Aragón
Facultad Politécnica
Universidad Nacional de Asunción
Email: arse.ferreira@gmail.com

Benjamin Barán
Facultad Politécnica
Universidad Nacional de Asunción
Email: bbaran@pol.una.py

***Resumen*—In general, Multiobjective Optimization Problems (MOPs) with conflicting objectives requires a Decision Maker to select a solution from a set of alternatives in the Pareto Front. A visual approach is a valuable alternative to analyze the several options that may exist. Thus, to support the visual analysis, this work proposes to combine a clustering method based on the shape of the solutions together with the parallel coordinate graph. The shape of a solution is defined as the positions of the ordered values of the normalized objectives that correspond to a solution. The method proposed in this work, allows exploring groups of solutions with similar shapes, to generate visual summaries with a high degree of representativeness, and to calculate statistical data of the groups and to identify the solutions. Also, this paper presents a software tool called Tava. This tool was developed to perform the visual analysis interactively.**

***Keywords*—Multiobjective optimization, visual analysis of solutions**

I. Introducción

Un gran número de problemas presentes en diversas áreas del conocimiento requieren la optimización simultánea de varios objetivos en conflicto [1]. Estos problemas, denominados Problemas de Optimización Multiobjetivos (MOPs), no poseen por lo general una solución única, sino un conjunto de soluciones que representan los distintos compromisos entre los objetivos optimizados. Los conflictos entre objetivos hacen que la mejora de algunos de ellos provoque un empeoramiento en otros. Este conjunto de soluciones en el espacio de búsqueda se conoce como el Conjunto Pareto y su imagen en el espacio objetivo es el Frente Pareto [2]. Los algoritmos evolutivos [3], [4], [5] y otras metaheurísticas basadas en población [6], [7], [8], han sido utilizados efectivamente para encontrar muy buenas aproximacionaes al conjunto de soluciones Pareto óptimas en problemas MOP.

Una vez que un algoritmo encuentra un conjunto de soluciones para un MOP, éstas deben pasar por un proceso de análisis donde serán estudiadas, clasificadas, valoradas, calificadas y finalmente seleccionadas con respecto a una o más características relevantes. La elección de una solución en particular, por lo general, la realiza un *Tomador de Decisiones* quién necesita disponer de herramientas y métodos adecuados para el efecto. Las técnicas de visualización forman parte del conjunto de métodos para el análisis y complementan a las métricas tradicionalmente aplicadas [9].

En los MOP, las técnicas de visualización sirven para ilustrar el Frente Pareto obtenido. La utilización de gráficos resulta de suma importancia puesto que permite descubrir características sobresalientes y peculiares del conjunto visualizado, las cuales podrían ser difícilmente percibidas con otras técnicas de análisis, ofreciendo al mismo tiempo una visión global de las soluciones. Las posibles características reveladas pueden corresponder a relaciones, patrones, agrupaciones y valores atípicos.

En presencia de conjuntos con cantidades grandes de soluciones es una práctica común agrupar los elementos para su estudio. Una división grupal implica doble beneficio para el análisis visual. Por un lado, se asocian soluciones con características similares aislándolas del resto del conjunto, hecho que posibilita una adecuada exploración. Por otro lado, se disminuyen los problemas de desorden, ambigüedad y solapamiento, presentes en algunos gráficos [10]. En un gráfico, los grupos pueden ser resaltados con diferentes colores para una mejor distinción, para de este modo destacar o descartar subconjuntos con propiedades significativas para quien realiza el análisis.

En este trabajo se presenta una estrategia para análisis visual de soluciones para Problemas de Optimización Multiobjetivo basada en la forma de los vectores objetivos. Además, se presenta la herramienta Tava [11], la cual implementa el método de análisis visual presentado. La estrategia de análisis visual se basa en el gráfico de coordenadas paralelas [12] y un agrupador basado en la forma de las soluciones [13]. Mediante el análisis propuesto, se busca explorar grupos de soluciones con formas similares, generar resúmenes visuales con alto grado de representatividad, calcular datos estadísticos de los grupos y resaltar las soluciones que resulten más atractivas para el tomador de decisiones.

El resto del artículo está organizado como sigue. En la sección II se definen algunos conceptos básicos sobre MOP y en la sección III se mencionan algunos trabajos relacionados

978-1-4799-7492-4/15/$31.00 ©2015 IEEE

a la propuesta, luego en la sección IV y V se describen el gráfico de coordenadas paralelas y un método de agrupación de datos, respectivamente. El enfoque de análisis visual propuesto es presentado en las sección VI y la herramienta Tava en la sección VII. Finalmente, en la última sección se presentan las conclusiones y los trabajos futuros.

II. Fundamentos Teóricos

A continuación se presentan algunos conceptos definidos en el campo de la optimización multiobjetivo, tomando como base los trabajos de Coello et al. [14] y Deb et al. [15].

Definición II-.1. ***Problemas de Optimización Mono-objetivo:*** Un Problema de Optimización Monoobjetivo minimiza (o maximiza) una función de la forma $f(\mathbf{x})$ sujeto a $g_i(\mathbf{x}) \leq 0$, $i = \{1, ..., m\}$, y $h_j(\mathbf{x}) = 0$, $j = \{1, ..., p\}$, $\mathbf{x} \in \Omega$. Una solución minimiza (o maximiza) el escalar $f(\mathbf{x}) \in \mathbb{R}$ donde $\mathbf{x}$ denota un vector de variables de decisión n-dimensional $\mathbf{x} = (x_1, ..., x_n) \in \mathbb{R}^n$ de algún universo Ω.

Las restricciones impuestas al problema son representadas por $g_i(\mathbf{x}) \leq 0$ y $h_j(\mathbf{x}) = 0$. Ω contiene todos los valores de $\mathbf{x}$ que satisfacen una evaluación de $f(\mathbf{x})$ y sus restricciones.

Definición II-.2. ***Problema de Optimización Multiobjetivo:*** Un MOP es definido minimizando (o maximizando) una función vectorial de la forma $F(\mathbf{x}) = (f_1(\mathbf{x}), ..., f_k(\mathbf{x}))$ sujeto a $g_i(\mathbf{x}) \leq 0$, $i = \{1, ..., m\}$, y $h_j(\mathbf{x}) = 0$, $j = \{1, ..., p\}$ $\mathbf{x} \in \Omega$. La solución de un MOP minimiza (o maximiza) las componentes de un vector $F(\mathbf{x})$ donde $\mathbf{x}$ denota un vector de variables de decisión n-dimensional $\mathbf{x} = (x_1, ..., x_n)$ de algún universo Ω, donde $g_i(\mathbf{x}) \leq 0$ y $h_j(\mathbf{x}) = 0$ representan restricciones impuestas al problema. Ω contiene todos los valores de $\mathbf{x}$ que satisfacen una evaluación de $F(\mathbf{x})$ en Ω.

Definición II-.3. ***Dominancia de Pareto:*** Dados dos vectores $\mathbf{u} = (u_1, ..., u_k)$ y $\mathbf{v} = (v_1, ..., v_k)$ se dice que $\mathbf{u}$ domina a $\mathbf{v}$ (denotado por $\mathbf{u} \prec \mathbf{v}$) si y sólo si $\mathbf{u}$ es mejor o igual que $\mathbf{v}$ para cada uno de sus componentes y estrictamente mejor para al menos uno de ellos, i.e., para un problema de minimización $\forall i \in \{1, ..., k\}$, $u_i \leq v_i \wedge \exists i \in \{1, ..., k\} : u_i < v_i$.

Definición II-.4. ***Optimalidad de Pareto:*** Se dice que una solución $\mathbf{x} \in \Omega$ es Óptimo de Pareto con respecto a Ω si y sólo si no existe $\mathbf{x}' \in \Omega$ para la cual $\mathbf{v} = F(\mathbf{x}') = (f_1(\mathbf{x}'), ..., f_k(\mathbf{x}'))$ domina a $\mathbf{u} = F(\mathbf{x}) = (f_1(\mathbf{x}), ..., f_k(\mathbf{x}))$. Se entiende que el significado de la expresión *Óptimo de Pareto* se refiere al espacio de la variable de decisión, a menos que se especifique lo contrario.

Definición II-.5. ***Conjunto Pareto:*** Dado un MOP, $F(\mathbf{x})$, el Conjunto Pareto, PS, se define como:

$$PS := \{\mathbf{x} \in \Omega \mid \neg \exists \mathbf{x}' \in \Omega \; F(\mathbf{x}') \prec F(\mathbf{x})\}.$$

El Conjunto Pareto (PS) está formado por todas las soluciones en el espacio de búsqueda cuyos vectores asociados del espacio objetivo no son dominados. Las soluciones óptimas de Pareto se clasifican basándose en sus valores funcionales evaluados.

Definición II-.6. ***Frente Pareto:*** Dado un MOP, $F(\mathbf{x})$, y el Conjunto Óptimo de Pareto, (PS), el **Frente Pareto** (PF) se define como:

$$PF := \{\mathbf{u} = F(\mathbf{x}) \mid \mathbf{x} \in PS\}.$$

El Frente Pareto (PF) corresponde al espacio objetivo de todas las soluciones no dominadas. Concretamente, los vectores resultantes de las evaluaciones de los vectores no dominados que componen el Conjunto Pareto forman el Frente Pareto.

III. Trabajos Relacionados

Una visualización adecuada de un conjunto de soluciones puede ser útil para múltiples fines: en el proceso de diseño de un problema dado sería beneficioso identificar objetivos redundantes [16], obtener valores numéricos para restricciones de variables o comparar las soluciones generadas en diferentes iteraciones, por la misma u otras estrategias de solución. Así también, un *Tomador de Decisiones* puede emplear la visualización para explorar el conjunto de soluciones de manera a encontrar grupos o soluciones con ciertas características y seleccionar algunas que el *Tomador de Decisiones* considere el mejor compromiso entre los objetivos.

En la búsqueda constante por alcanzar los beneficios que aportan las técnicas de visualización en el análisis multiobjetivo, se han propuesto varios enfoques basados en agrupación de datos y gráficos multidimensionales [17], [18], [19], [20]. Cada propuesta se centra en algún aspecto particular del conjunto visualizado, con el objetivo de capturar nuevas informaciones potencialmente útiles a la hora de tomar decisiones.

Fua et al. [18], proponen un enfoque que emplea un algoritmo de agrupamiento jerárquico junto con una variación del gráfico de coordenadas paralelas. Cada grupo generado se visualiza como una banda desvanecida desde un centro completamente opaco hasta un borde transparente lo que permite a los usuarios navegar por la estructura jerárquica resultante para localizar un nivel particular de detalle. De esta manera, se genera una vista con multiresolución de los datos que, con una adecuada interacción, puede ayudar al descubrimiento sistemático de tendencias de datos y patrones ocultos.

Novotny [19], utiliza una versión modificada del método de agrupamiento K-means y representa cada grupo como un área poligonal, en la cual emplea diferentes niveles de densidad y textura para distinguir cada grupo. Los polígonos se ordenan por tamaño antes de su visualización para optimizar la visibilidad de los grupos más densos. Novotny emplea el concepto de visualización abstracta cuyo propósito es reducir el volumen de información pero preservando siempre su significado, lo cual es muy útil cuando se visualiza un gran volumen de datos.

Artero et al. [20], aborda el problema de solapamiento de datos con una estrategia que calcula la frecuencia y densidad de la información de un conjunto de datos, y utiliza esa información en un gráfico de coordenadas paralelas para filtrar la información que se va a presentar al usuario, reduciendo así el desorden visual y permitiendo al analista observar patrones en los datos.

Johansson et al. [17], tratan el problema de visualizar un gran número de elementos de datos en un gráfico

de coordenadas paralelas. Resuelven este problema con la construcción de grupos y emplean texturas de alta precisión las cuales se componen de texturas de colores que revelan la estructura de un grupo. También emplean funciones de transferencia las cuales actúan sobre las texturas de alta precisión manipulando los valores de intensidad, esto con el fin de resaltar diferentes aspectos de las características de los grupos. Por último, se introduce la técnica "Feature animation" la cual sirve como una forma intuitiva de visualizar las propiedades estadísticas de los grupos.

IV. Gráfico de Coordenadas Paralelas

Los gráficos multidimensionales permiten representar elementos de datos con múltiples variables. Estos elementos forman un conjunto de observaciones $X = \{x_i\}$, donde cada elemento x_i consiste en un vector con m variables, x_i = (x_{i_1}, ..., x_{i_m}) [21].

El gráfico de coordenadas Paralelas permite representar conjuntos de datos n-dimensionales en dos dimensiones [12]. Para cada una de las m dimensiones se emplea un eje vertical. Estos m ejes verticales se ubican de manera equidistante y de forma paralelas mientras que los elementos del conjunto de datos son representados por líneas que atraviesan los m-ejes.

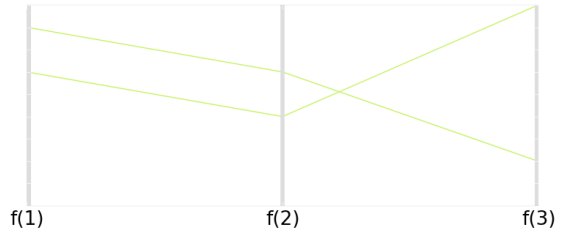

Figura 1: Coordenadas Paralelas: elementos de datos con valores objetivos armónicos y conflictivos para 3 objetivos

El Frente Pareto obtenido para los MOP corresponden a conjunto de datos m-dimensionales. Así, para representar las funciones objetivos en coordenadas paralelas, los objetivos son dispuestos en el eje de ordenadas y se trazan líneas conectando los vértices en los ejes. La relación entre objetivos queda marcada por la línea que une dos ejes adyacentes. Las líneas cruzadas indican la presencia de conflictos y la ausencia de éstas indica la armonía entre objetivos. En la Figura 1, se visualizan dos elementos cuyos objetivos $f(1)$ y $f(2)$ están en armonía, en contraste a los objetivos $f(2)$ y $f(3)$ que son conflictivos.

La fortaleza del gráfico de coordenadas paralelas reside en su simplicidad y naturaleza multidimensional, sin embargo, posee algunas limitaciones. Una de ellas es que sólo puede representar las relaciones entre objetivos contiguos y no más de dos. Cuando algunos elementos tienen valores iguales o cercanos en determinados objetivos aparecen los problemas de ambigüedad y solapamiento, hecho que disminuye la claridad del gráfico e impide una identificación correcta de los mismos, como se ilustra en la Figura 2(a). El resaltado en colores y la agrupación de datos son los principales mecanismos para disminuir los efectos, de ambigüedad y solapamiento, como se muestra en la Figura 2(b).

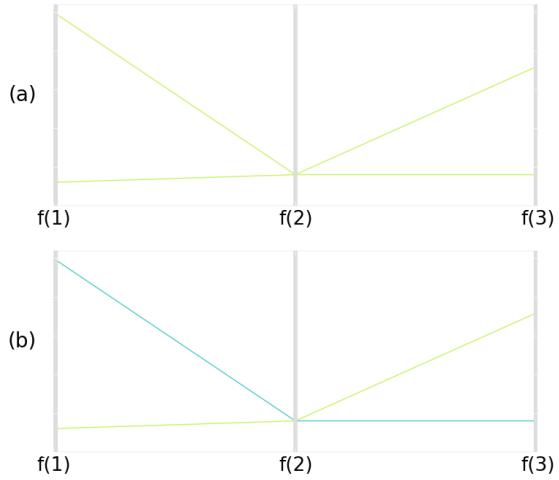

Figura 2: Coordenadas Paralelas: (a) ilustración ambigua de elementos; (b) ilustración clara o desambigua de elementos

En este trabajo se emplea el gráfico de coordenadas paralelas para representar visualmente las alternativas de solución encontradas para MOP y graficar resúmenes del conjunto solución calculados a partir de soluciones que adoptan formas similares.

V. Método de Agrupación de Datos

Los métodos de agrupación de datos permiten clasificar los elementos de un conjunto bajo algún criterio dado para formar grupos o subconjuntos de datos relacionados. Los criterios comúnmente utilizados son la distancia o similitud. La distancia entre elementos se define en términos de una determinada función de distancia y la similitud es generalmente evaluada mediante una medida de correlación. Por lo general, los métodos de agrupamiento son utilizados en presencia de conjuntos de grandes cantidades de datos ya que los divide en fracciones más pequeñas y manejables.

Las soluciones en el espacio objetivo para un MOP son en general vectores m-dimensionales, y cada uno de estos vectores tienen la propiedad llamada ***forma*** [13], definida como sigue:

Definición V-.1. Forma de un vector en $\mathbb{R}^m$: dado un vector $y \in \mathbb{R}^m$, una permutación $\pi(y) = \pi_1, ..., \pi_m$, $\pi_i \in 1, ..., m$, es la forma de y si:

$$y_{\pi_i} \leq y_{\pi_j}, \forall i < j$$

La ***similitud de forma*** de los vectores objetivos se puede utilizar como criterio de agrupación, para generar grupos de elementos integrados por soluciones de un MOP que poseen formas similares. Supongamos que se tiene el siguiente conjunto solución Y con seis elementos, cada uno, con tres objetivos:

- $y_1 = [0.2, 0.5, 0.4]$
- $y_2 = [0.3, 0.1, 0.6]$
- $y_3 = [0.3, 0.2, 0.5]$
- $y_4 = [0.4, 0.2, 0.1]$

- $y_5 = [0.6, 0.5, 0.2]$
- $y_6 = [0.5, 0.7, 0.4]$

La forma de cada vector es como sigue:

- $\tau(y_1) = [0, 2, 1]$
- $\tau(y_2) = [1, 0, 2]$
- $\tau(y_3) = [1, 0, 2]$
- $\tau(y_4) = [2, 1, 0]$
- $\tau(y_5) = [2, 1, 0]$
- $\tau(y_6) = [2, 0, 1]$

La cantidad posible de formas diferentes en este conjunto es igual a seis (3!). Utilizando la ***igualdad de forma*** como medida de similitud, el conjunto queda agrupado como se muestra en la Figura 3. En este gráfico, cada forma existente en el conjunto Y queda representada por un grupo.

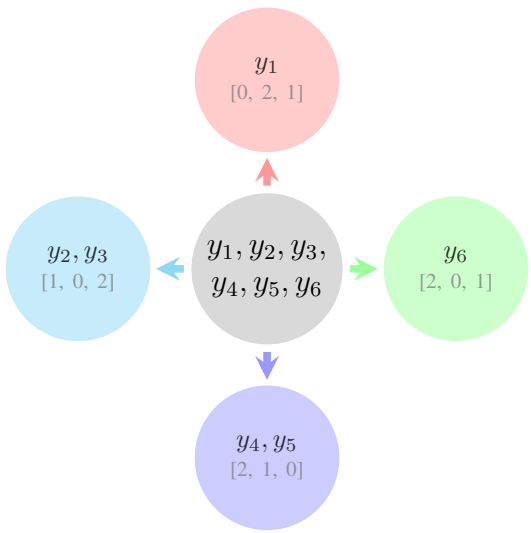

Figura 3: Agrupaciones generadas empleando como criterio de agrupación la ***igualdad de forma*** de los vectores objetivos.

El aumento del número de objetivos implica un crecimiento factorial en la cantidad de formas, por ejemplo, para 3 objetivos hay 6 formas diferentes, pero para 6 objetivos hay 720 formas posibles. Sin embargo, en un conjunto de soluciones, se encuentra como máximo un número de formas igual al tamaño del conjunto y varias soluciones podrían corresponder a una misma forma.

Cuando se desea generar una cantidad predefinida de grupos, se debe utilizar una métrica basada en la similitud de forma como se muestra en [13], donde se emplea la distancia de Sperman-rho para agrupar los elementos con formas similares. Esta distancia se define como sigue:

Definición V-.2. Distancia de Spearman-rho: Dadas dos permutaciones τ y τ', e interpretando π_i como la posición del elemento i en π, la distancia de Spearman rho se define como:

$$\rho(\pi, \pi') = (\sum_{i=i}^{m} |\pi_i - \pi'_i|^2)^{(1/2)}$$

El Algoritmo 1 presenta el procedimiento definido en [13] para el agrupamiento de funciones objetivo en τ grupos considerando sus formas. Primeramente, en la línea 1 se inicializan una cantidad de τ grupos. En la línea 2, cada elemento x de la población P es normalizado con la función $\hat{\mathbf{F}}(\mathbf{x})$ y se calcula la forma ($\pi_{\mathbf{x}}$) del valor normalizado $\pi(\hat{\mathbf{F}}(\mathbf{x}))$. Seguidamente, en la línea 3 se seleccionan aleatoriamente τ diferentes formas del conjunto de formas de valores objetivos normalizados en un conjunto S. Si eventualmente hay menos que τ diferentes formas, el conjunto S se completa con valores repetidos. Un índice Id correspondiente a cada grupo es asignado a cada forma en S. A continuación, por cada elemento x perteneciente a P, en la línea 5 se va poblando un conjunto S'_x con los índices de las formas en S que tienen la mínima distancia de Spearman rho respecto a la forma π_x. El conjunto S'_x es utilizado para seleccionar el índice Id del grupo en el cual el elemento x será incluido. Si la cardinalidad del conjunto S'_x es igual a uno, entonces en la línea 7 se asigna al índice Id el único elemento del conjunto S'_x; de otra manera, en la línea 9 se selecciona al azar uno de los índices del conjunto S'_x. Por último, en la línea 11 el elemento x es asignado al grupo G_{Id}. El procedimiento finaliza cuando todos los elementos de P son asignados a uno de los τ grupos.

Algoritmo 1 Algoritmo Shape utilizado para agrupar soluciones en τ grupos considerando sus formas

1: Inicializar $G_1, G_2, ..., G_\tau$ grupos. $G_{Id} = \emptyset$ para $Id \in [1, ..., \tau]$.
2: Para cada $\mathbf{x} \in P$ obtenemos su valor objetivo normalizado $\hat{\mathbf{F}}(\mathbf{x})$ y la forma $\pi_{\mathbf{x}} = \pi(\hat{\mathbf{F}}(\mathbf{x}))$.
3: Seleccionar aleatoriamente τ diferentes formas $S = sh_1, sh_2, ..., sh_\tau$ de las formas de $\mathbf{x} \in P$.
4: **for each** $\mathbf{x} \in P$ **do**
5: $\quad S' = \{Id | sh_{Id} \in min_{sh \in Sh}\{\rho(\pi_{\mathbf{x}}, sh)\}\}$
6: $\quad$ **if** $|S'| == 1$ **then**
7: $\quad\quad Id$ es el elemento de S'
8: $\quad$ **else**
9: $\quad\quad$ Seleccionar Id al azar de S'
10: $\quad$ **end if**
11: $\quad G_{Id} = G_{Id} \cup \mathbf{x}$
12: **end for**

A continuación, se muestra tres grupos ($\tau = 3$) generados con el algoritmo definido en [13] a partir del conjunto solución de ejemplo. Las agrupaciones quedan integradas como se muestra en la Figura 4. En este caso, no todas las formas quedan representadas por los grupos, ya que existen cuatro formas que se combinan en tres grupos, esta combinación se realiza utilizando la distancia de Spearman-rho para asociar elementos con formas similares.

La agrupación por ***igualdad de forma*** divide el conjunto solución en una cantidad de grupos igual al número de diferentes formas adoptadas por los elementos. En este trabajo, se aprovecha esta característica de las soluciones con el fin de ofrecer una exploración del conjunto y apoyar el análisis de soluciones para MOP.

VI. Enfoque de Análisis Visual Propuesto

En esta sección se plantea un procedimiento de análisis visual de soluciones encontradas para un MOP, basado en una exploración por formas de los elementos del conjunto solución, cuya finalidad es mostrar características relevantes e identificar soluciones.

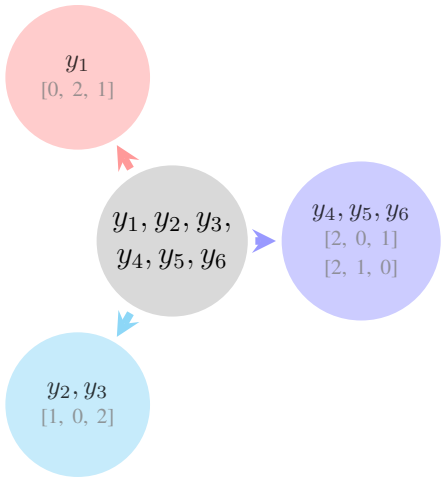

Figura 4: Tres grupos generados a partir de cuatro formas distintas utilizando ***similitud de forma***

El proceso de análisis del enfoque se realiza de la siguiente manera. Inicialmente, se escoge un conjunto de soluciones, del cual, se generan $\tau grupos$ utilizando **formas**, luego a cada grupo obtenido se asigna un color, se calculan datos estadísticos y se genera un resumen. Por último, los grupos son trazados total o parcialmente en un gráfico de coordenadas paralelas mediante una herramienta llamada TAVA, creada especialmente para el efecto [11].

A continuación se detalla el procedimiento propuesto, empleando como ejemplo, un conjunto compuesto por 188 soluciones y considerando 3 objetivos de soluciones no comparables entre si para el problema de optimización KP (Knapsack Problem - Problema de la mochila) [22].

Antes de empezar el proceso de análisis se ilustra el conjunto completo de datos en la Figura 5. En el gráfico se puede visualizar una gran diversidad de soluciones con valores objetivos muy variados. Al mismo tiempo, se observa la presencia de conflictos entre objetivos representados por las líneas cruzadas entre columnas. Además, existen líneas solapadas que agregan ambigüedad a la visualización. En esta figura es muy difícil de distinguir las formas adoptadas por las soluciones, por lo que, no se puede identificar los grupos con formas similares existentes en el conjunto.

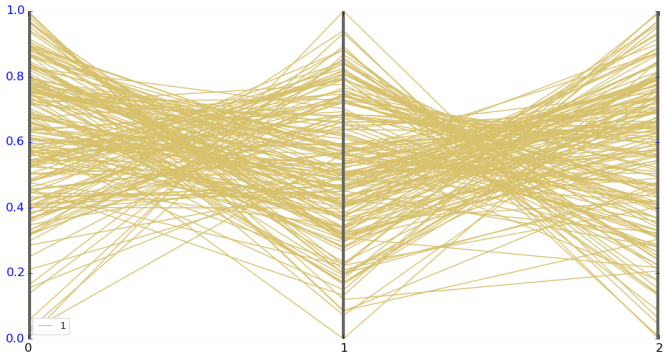

Figura 5: Población con 188 individuos generados para el problema de optimización KP.

Debido a la ambiguedad ilustrada en la Figura 5, este trabajo propone inicialmente generar grupos del conjunto utilizando agrupación por ***forma***, luego, se asocia un color a cada grupo obtenido y el resultado se visualiza en la Figura 6. Esta figura divide la población inicial en colores de líneas diferentes, en total se observan seis colores donde cada color corresponde a un grupo asociado a una forma. El hecho de disponer las soluciones de esta manera, ilustra la existencia de un grupo de formas al que le corresponde un número mayor de soluciones y que los integrantes de un grupo no necesariamente forman aglomeraciones acentuadas. Igualmente, se mejora la visualización de las soluciones y se permite notar con mayor claridad la relación de conflicto entre los objetivos. Si bien, se propicia una mejor identificación de las soluciones y resalta la forma que adoptan, aún resulta difícil realizar un análisis visual adecuado.

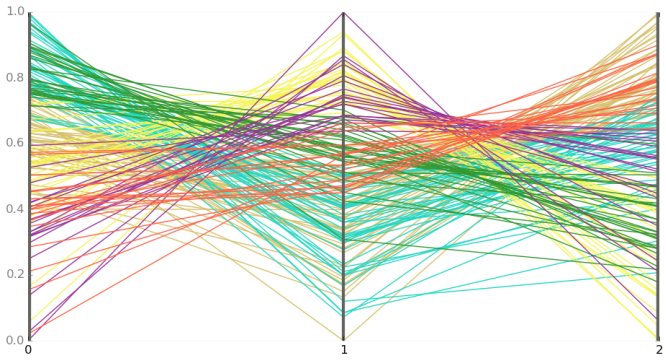

Figura 6: Resaltado en colores de las 6 formas encontradas en la Población.

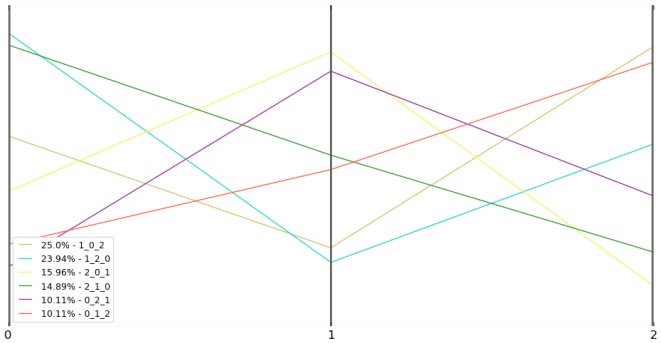

Figura 7: Resumen visual y datos estadísticos de la población, cada resumen representa a un grupo en particular

Al terminar de generarse las agrupaciones, además de asignar color a cada grupo, se crea un resumen y se realiza el cálculo de datos estadísticos para cada uno. El resumen está constituido por un vector que se consigue al promediar las observaciones incluidas en cada grupo, un objetivo a la vez. Para cada grupo se determina la cantidad de integrantes del grupo y el porcentaje de elementos en cada grupo respecto al número total de soluciones. La Figura 7 muestra los resúmenes de los grupos del conjunto representados con seis líneas, cada una, pintada con el mismo color que el grupo al

cual representa, ilustrado antes en la Figura 6. También, se visualizan unas etiquetas que exponen las formas asociadas a los grupos y el porcentaje de soluciones respecto al total. Un análisis realizado a partir de este gráfico, permite notar con facilidad los grupos con mayor cantidad de soluciones que adoptan una misma forma y se consigue visualizar los resúmenes, altamente representativos, de los grupos generados.

Posteriormente y como parte del proceso de análisis visual, se aplica un filtro al conjunto de datos seleccionando alguno de los grupos asociados a formas. La Figura 8 muestra solamente el grupo con mayor número de integrantes. Cuando se filtran las soluciones se obtiene un gráfico más claro, se reduce significativamente las ambigüedades y solapamientos, lo cual, mejora en gran medida la identificación de cada solución. Además, la exploración realizada, a través del filtro de grupo, permite aislar las soluciones pertenecientes a los grupos escogidos y enfocar el análisis en ellas.

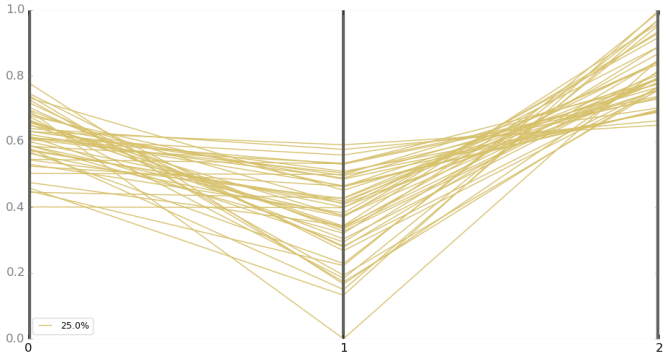

Figura 8: Filtrado del grupo que representa a la forma más adoptada por los elementos del conjunto.

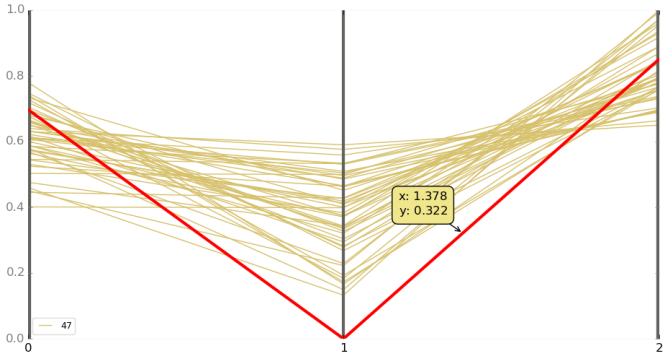

Figura 9: Resaltado de un elemento del conjunto solución.

Por último, se resalta una solución del conjunto como se muestra en la Figura 9, en la cual, se destaca un elemento en particular a partir del conjunto previamente filtrado. Esta solución es la que presenta el valor más bajo en el objetivo 1 y valor medio para arriba en los objetivos 0 y 2 con respecto al grupo representado. Esta información se mantenía oculta bajo la ambigüedad de los gráficos anteriores y es ahora percibida gracias al resaltado. Al disponer una solución de esta manera se elimina toda incertidumbre, se exponen todas sus características y se alcanza la plena identificación.

El procedimiento de análisis visual propuesto permite la exploración del conjunto solución mediante, el agrupamiento por formas, coloreado de elementos y filtrado de grupos. También, provee resúmenes del conjunto y destaca una solución en particular para su plena identificación a través del resaltado, lo que es de suma utilidad en la toma de decisiones, especialmente en problemas con muchos objetivos (many-objective problems)

VII. Herramienta de Visualización Tava

En esta sección se presentan las principales características y funcionalidades proporcionadas por la herramienta **Tava** [11], las cuales, buscan apoyar el análisis visual de soluciones de un MOP. Esta aplicación, se basa en el lenguaje de programación Python en su versión 2.7 [23] y fue desarrollada especialmente para implementar el proceso de análisis visual de soluciones generadas para MOP conforme a lo expuesto en la sección anterior. Además, Tava ofrece algunas utilidades que pueden servir de apoyo al que realiza el análisis en el momento de llevar a cabo el análisis visual.

A continuación se resumen las utilidades que ofrece Tava, para lo cual, se emplea un conjunto solución compuesto por 400 individuos con 10 objetivos, generados por el algoritmo NSGA-II [3] en la iteración 400 para el problema de optimización DTLZ1 [15].

VII-A. Representación de Datos

El método de normalización *Feature Scaling* [24] es utilizado para normalizar los datos. Esta función toma el conjunto solución y transforma los valores iniciales a un rango de valores entre 0 y 1. La Figura 10 ilustra el conjunto de datos normalizado. El método se define como sigue:

Definición VII-A.1. *Feature Scaling*: consiste en reescalar el rango de características para escalar el rango en [0, 1]. La selección del rango objetivo depende de la naturaleza de los datos. La fórmula general es dada como:

$$x' = \frac{x - min(x)}{max(x) - min(x)}$$

donde x es un valor original, x' es el valor normalizado.

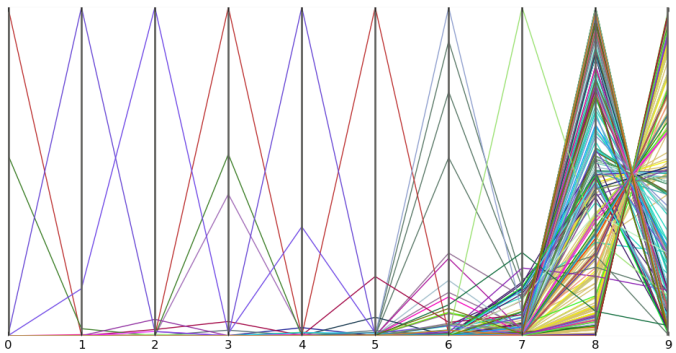

Figura 10: Agrupaciones con datos normalizados.

La normalización hace que los datos se dispersen consiguiendo aprovechar mejor el espacio disponible en el

gráfico, hecho que posibilita una representación más limpia de los valores.

VII-B. Selección o Filtro de Grupos

Tava permite la selección de grupos de manera manual o automática para su visualización. Esta última, resulta de suma utilidad a medida que crece la cantidad de grupos. Con respecto a los filtros automáticos, se manejan dos opciones:

- Representatividad en Cantidad: opción que permite la selección de grupos más relevantes en cuanto al número de integrantes, ya sea, de mayor o menor cantidad.
- Representatividad en Valores Objetivos: opción que permite la selección de grupos cuyos integrantes poseen algún valor extremo respecto a uno o varios de sus objetivos.

Al emplear la selección automática en la Figura 11, se visualizan las dos tendencias dominantes de la población. Análogamente, la Figura 12 ilustra los grupos correspondientes a los valores extremos del objetivo 7.

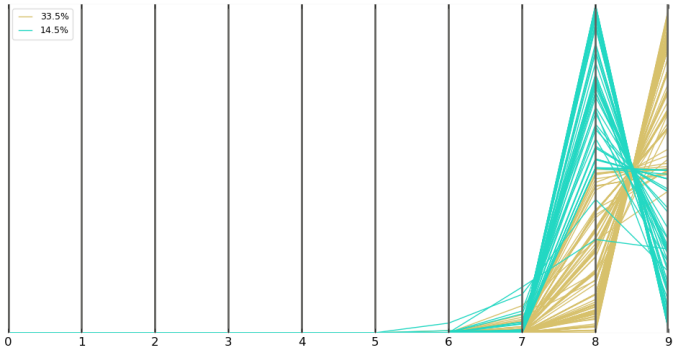

Figura 11: Agrupaciones seleccionadas de manera automática.

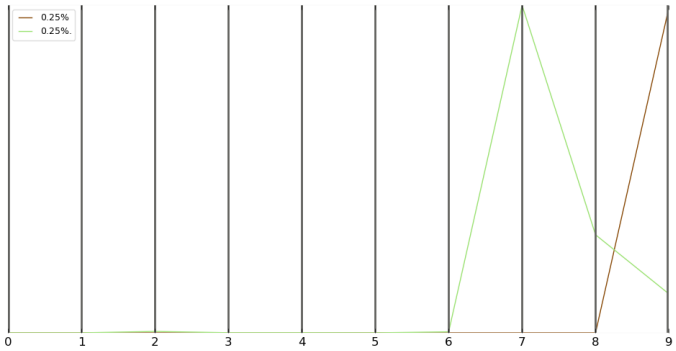

Figura 12: Grupos integrados por individuos con valor más alto y más bajo para el objetivo 7.

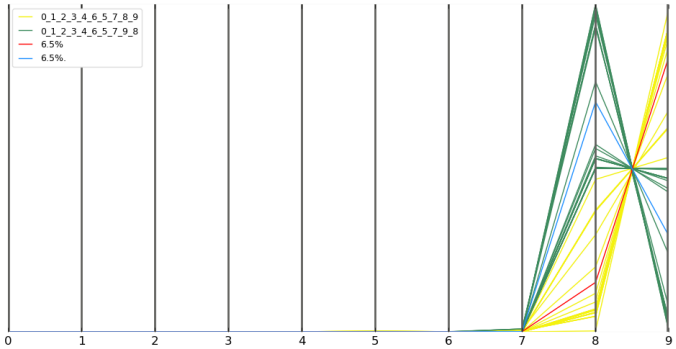

Figura 13: Representación de grupos, resúmenes y leyendas.

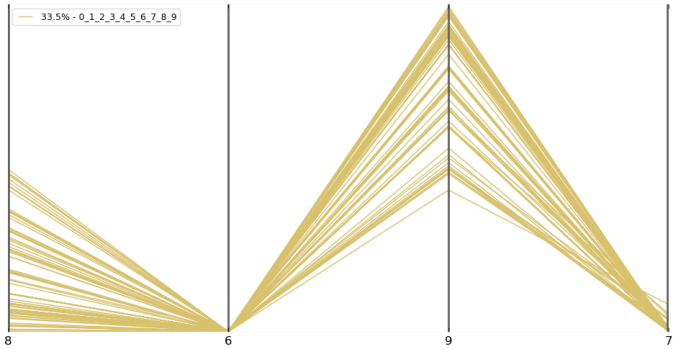

Figura 14: Selección y ordenamiento de objetivos.

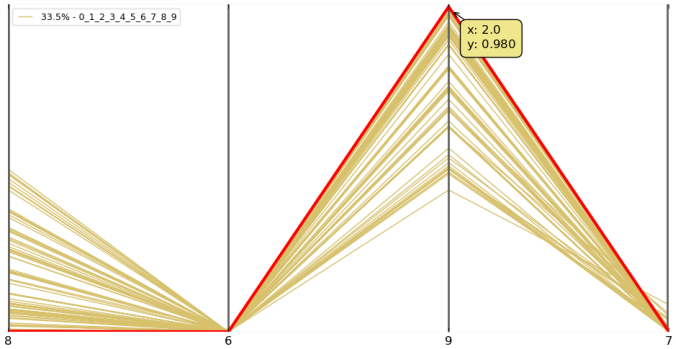

Figura 15: Selección y ordenamiento de objetivos y resaltado de solución.

VII-C. Leyendas y Visualización Combinada de Grupo y Resumen

En los gráficos se pueden agregar leyendas informativas de los grupos visualizados, las alternativas que se manejan son las citadas en la sección VI, así como la forma correspondiente al grupo y un nombre establecido por defecto. Además, una misma figura puede ilustrar grupos y resúmenes de manera combinada. La Figura 13 ilustra dos grupos seleccionados manualmente, pertenecientes a la tercera y cuarta posición en cuanto a representatividad en cantidad, junto con sus resúmenes y leyendas respectivamente.

VII-D. Selección y Reordenamiento de Objetivos

La selección de objetivos es una funcionalidad de la aplicación que permite el filtrado de objetivos a visualizar. Esta característica junto con el reordenamiento agrega flexibilidad a la hora de analizar soluciones con muchos objetivos. Además, en un gráfico como coordenadas paralelas, permite intercambiar de lugar a los objetivos posibilitando la visualización de relación entre objetivos de posiciones inicialmente no contiguos. La Figura 14 muestra los objetivos filtrados en el siguiente orden 8, 6, 9, 7, descartando el resto; además, la figura proporciona las etiquetas de porcentaje y forma inicial adoptadas por el grupo.

VII-E. Resaltado de Solución

Tava permite realizar el resaltado de un elemento en cualquiera de las vistas proporcionada, es decir, un elemento se puede resaltar en el momento de visualización del conjunto total, cuando se tienen filtros de grupos o después de realizar selección y reordenamiento de objetivos, hasta se puede destacar los resúmenes proporcionados, además, permite destacar el valor del punto (x, y) donde se realiza el clic de selección. La Figura 9, muestra un elemento resaltado luego de filtrar grupos del conjunto, en este caso, el punto de selección (x, y) es $(2.0, 0.980)$.

VIII. Conclusión y Trabajos Futuros

El enfoque de análisis visual propuesto plantea una nueva manera de explorar las soluciones que componen el Frente Pareto obtenido para un MOP, basados en agrupación por formas y coloreado de grupos. Esta estrategia, además de distinguir las relaciones estructurales que poseen las soluciones, posibilita generar resúmenes representativos de los grupos permitiendo así informar sobre propiedades o valores estadísticos de los mismos. Igualmente, el enfoque propuesto identifica y destaca una solución particular respecto a las demás por medio del resaltado.

La herramienta Tava implementa el procedimiento de análisis visual. Es un software de código abierto y especializado en el proceso propuesto para visualizar conjuntos de soluciones no-comparables resultantes de la resolución de problemas con múltiples objetivos posiblemente contradictorios. Además, Tava incorpora utilidades que sirven de apoyo al momento de llevar a cabo el análisis y la toma de decisión.

Las informaciones recabadas en la exploración por grupos, obtenida por la técnica aplicada en este trabajo, son distintas a las proveídas por métodos basados en agrupadores que emplean directamente los valores objetivos en el cálculo de similitud. Así, se pone de manifiesto nuevas características y propiedades inmersas en grupos de soluciones encontradas para un MOP, mejorando la percepción visual y apoyando la toma de decisión.

Este trabajo ha desarrollado el método propuesto hasta los niveles presentados en las distintas secciones. El método presentado ofrece desafíos muy interesantes para continuar analizando su aplicación con otros métodos de agrupación y gráficos multidimensionales, para lo cual se propone algunos trabajos a futuro como:

- realizar una comparación de exploración de datos conseguida por el enfoque ante otras estrategias existentes. Por ejemplo, una estrategia que emplea los métodos de agrupación k-means [25] o Spectral Clustering [26];
- emplear gráficos multidimensionales diferentes a coordenadas paralelas para el enfoque en la visualización de los diferentes valores objetivos.

Referencias

[1] Sawaragi, Yoshikazu, Nakayama, Hirotaka, Tanino y Tetsuzo. *Theory of Multiobjective Optimization (Mathematics in Science and Engineering Vol. 176)*. Academic Press, October 1985.

[2] Andrzej Osyczka. *Multicriterion optimization in engineering with FORTRAN programs*. Chichester, West Sussex : E. Horwood ; New York : Halsted, 1984. Translated from the Polish.

[3] Deb, K., Pratap, A., Agarwal, S. y Meyarivan, T. A fast and elitist multiobjective genetic algorithm: NSGA-II. *Trans. Evol. Comp*, 6(2):182–197, April 2002.

[4] E. Zitzler y L. Thiele. Multiobjective Evolutionary Algorithms: A comparative Case Study and the Strength Pareto Approach. *IEEE Transactions on Evolutionary Computation*, 3(4):257–271, Nov 1999.

[5] E. Zitzler, M. Laumanns y L. Thiele. SPEA2: Improving the Strength Pareto Evolutionary Algorithm for multobjective optimization. *In Evolutionary Methods for Design. Optimization and Control*, 3242:95–100, 2002.

[6] Marco Dorigo, Mauro Birattari y Thomas Stützle. Ant colony optimization - Artificial Ants as a Computational Intelligence Technique. *IEEE COMPUT. INTELL. MAG*, 1:28–39, 2006.

[7] J. Kennedy y R. Eberhart. Particle swarm optimization. In *Neural Networks, 1995. Proceedings., IEEE International Conference on*, volume 4, pages 1942–1948 vol.4, Nov 1995.

[8] Gheraibia, Youcef y Moussaoui, Abdelouahab. *Penguins Search Optimization Algorithm (PeSOA)*, pages 222–231. Springer Berlin Heidelberg, Berlin, Heidelberg, 2013.

[9] N. Riquelme, C. Von Lücken y B. Baran. Performance metrics in multi-objective optimization. In *2015 Latin American Computing Conference (CLEI)*, pages 1–11, Oct 2015.

[10] Zhou, H., Yuan, H., Qu, H., Cui, W. y Chen, B.n. Visual Clustering in Parallel Coordinates. *Computer Graphics Forum*, 27:1047–1054, May 2008.

[11] Abrahan Fretes, Arsenio Ferreira y Christian Von Lücken. Tava Software. https://github.com/abrahanfretes/tava.git, 2015. [Análisis de Soluciones para Problemas de Optimización Multi-Objetivo basado en la forma de la Solución].

[12] Inselberg, Alfred y Dimsdale, Bernard. Parallel coordinates: A tool for visualizing multi-dimensional geometry. In *Proceedings of the 1st Conference on Visualization '90*, VIS '90, pages 361–378, Los Alamitos, CA, USA, 1990. IEEE Computer Society Press.

[13] C. Von Lücken, B. Barán y C. Brizuela. Clustering based parallel Many-objective Evolutionary Algorithms using the shape of the objective vectors. *Lecture Notes in Computer Science*, 9019:50–64, March 2015.

[14] Coello, Carlos A. Coello, Lamont, Gary B. y Veldhuizen, David A. Van. *Evolutionary Algorithms for Solving Multi-Objective Problems (Genetic and Evolutionary Computation)*. Springer-Verlag New York, Inc., Secaucus, NJ, USA, 2006.

[15] Deb, Kalyanmoy, Thiele, Lotha, Laumanns, Marco y Zitzler, Eckart. *Scalable Test Problems for Evolutionary Multiobjective Optimization*, pages 105–145. Springer London, London, 2005.

[16] Alan R.R. de Freitas, Peter J. Fleming y Frederico G. Guimarães. Aggregation trees for visualization and dimension reduction in many-objective optimization. *Information Sciences*, 298:288–314, 2015.

[17] J. Johansson, P. Ljung, M. Jern y M. Cooper. Revealing structure within clustered parallel coordinates displays. In *IEEE Symposium on Information Visualization, 2005. INFOVIS 2005.*, pages 125–132, Oct 2005.

[18] Ying-Huey Fua, M. O. Ward y E. A. Rundensteiner. Hierarchical parallel coordinates for exploration of large datasets. In *Visualization '99. Proceedings*, pages 43–508, Oct 1999.

[19] Matej Novotny. Visually effective information visualization of large data. In *In 8th Central European Seminar on Computer Graphics (CESCG 2004)*, pages 41–48. CRC Press, 2004.

[20] Artero, Almir Olivette, de Oliveira, Maria Cristina Ferreira y Levkowitz, Haim. Uncovering clusters in crowded parallel coordinates visualizations. In *Proceedings of the IEEE Symposium on Information Visualization*, INFOVIS '04, pages 81–88, Washington, DC, USA, 2004. IEEE Computer Society.

[21] Wong, Pak Chung y Bergeron, R. Daniel. 30 years of multidimensional multivariate visualization. In *Scientific Visualization, Overviews, Methodologies, and Techniques*, pages 3–33, Washington, DC, USA, 1997. IEEE Computer Society.

[22] Eckart Zitzler y Marco Laumanns. Test problems and test data for multiobjective optimizers [multiobjective 0/1 knapsack problem]. ftp://ftp.tik.ee.ethz.ch/pub/people/zitzler/TESTDATA/ZT1999a/knapsack.100.3.tar.gz. accedido el 11/8/2016.

[23] Guido Rossum. Python reference manual. Technical report, pub-CWI, Amsterdam, The Netherlands, The Netherlands, 1995.

[24] Piotr Juszczak, David M.J. Tax y Robert P. W. Duin. Feature scaling in support vector data description. In E.F. Deprettere, A. Belloum, J.W.J. Heijnsdijk, and F. van der Stappen, editors, *Proc. ASCI 2002, 8th Annual Conf. of the Advanced School for Computing and Imaging*, pages 95–102, 2002.

[25] T. Kanungo, D. M. Mount, N. S. Netanyahu, C. D. Piatko, R. Silverman y A. Y. Wu. An efficient k-means clustering algorithm: analysis and implementation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 24(7):881–892, Jul 2002.

[26] Luxburg y Ulrike. A tutorial on spectral clustering. *Statistics and Computing*, 17(4):395–416, December 2007.