ESCALAMIENTO MULTIDIMENSIONAL: CONCEPTOS Y ENFOQUES

G. Linares, Departamento de Matemática Aplicada, Facultad de Matemática y Computación, Universidad de La Habana

RESUMEN
El nombre de "escalamiento multidimensional" se asocia a un conjunto de técnicas que persiguen como objetivo la representación de datos a través de una configuración de puntos cuando se conoce una determinada información sobre similaridades entre objetos. La historia de las técnicas de escalamiento multidimensional comienzan con el trabajo de Torgeson en 1952, quien introdujo el término y esbozó las primeras ideas. A partir de ese momento muchos investigadores han esbozado otras ideas con un único propósito: obtener una representación euclídea que respete lo más posible las relaciones de proximidad entre los objetos. En el trabajo se exponen los principales enfoques que han surgido alrededor de este tema y se precisan sus aspectos conceptuales. Una aplicación a un problema real del campo psicopedagógico muestra alguna de las posibilidades del escalamiento multidimensional.

Palabras clave: índice de esfuerzo, matriz de similitud, computar métodos intensivos

ABSTRACT
"Multidimensional Scaling" is the name of a set of procedures with the objective of data representation in a graphical plot when some information about similarities between objects is known. The history of these procedures began with Torgeson´s work in 1952 who gave the preliminary ideas. Since then many investigators have developed other works with one only aim: to obtain a euclidean representation most similar to the starting proximities between objects. In this paper we show the main purposes and some essential concepts about this thematic.

Key words: Stress, similarity matrix, computer intensive methods

MSC: 62H99

1. INTRODUCCION

Con el nombre de *escalamiento multidimensional* se conoce un conjunto de técnicas que persiguen como objetivo la representación de datos a través de la construcción de una configuración de puntos cuando se conoce una determinada información sobre proximidades entre objetos.

Aunque en la primera mitad de este siglo surgieron las primeras ideas, la historia de las técnicas de escalamiento multidimensional comienzan con un trabajo de Torgerson en 1952, quien introdujo el término y esbozó las primeras ideas. En 1962, Shepard hizo una formulación bastante precisa del escalamiento multidimensional cuando demostró, empíricamente, que si se conocía una ordenación de las distancias entre puntos, podría encontrarse una configuración de puntos en un espacio euclidiano de baja dimensión cuyas interdistancias euclidianas reproducían prácticamente la ordenación original. Las ideas de Shepard fueron refinadas por Kruskal en los años 60 y desarrolladas por otros autores como Guttman y Lingoes,. Estas técnicas abordan el problema de construir distancias métricas transformando adecuadamente las disimilaridades y se conocen con el nombre de escalamiento multidimensional *no métrico*. En 1966, Gower propuso el método de Análisis de Coordenadas Principales, que puede considerarse un método *métrico* de escalamiento multidimensional, y que evita resolver los procesos iterativos de las técnicas no métricas. Durante las décadas de los 80 y de los 90 numerosos investigadores continuaron la búsqueda de algoritmos que lograran alcanzar una configuración final de puntos cuyas distancias fueran lo "más cercanas posibles" a las disimilitudes apreciadas.

El propósito de este trabajo es introducirnos en el tratamiento de las técnicas de Escalamiento Multidimensional y el campo de investigación actual de las mismas y para ello debemos precisar que existen dos requisitos esenciales para desarrollar un análisis de escalas multidimensionales. Estos requisitos son:

a) Partir de un conjunto de números, llamados *proximidades o similaridades*, que expresan todas o la mayoría de las combinaciones de pares de similaridades dentro de un grupo de objetos, y,

b) Contar con un *algoritmo* implementado computacionalmente para llevar a cabo el análisis.

En lo que sigue desarrollaremos algunos detalles de ambos requisitos, pero antes, en la Sección 2, resumiremos los conceptos y las ideas básicas de estas técnicas. En la Sección 3 expondremos algunos aspectos metodológicos para obtener una matriz de similaridades y brindaremos la clasificación de técnicas propuesta por Gower y Digby (1981). En la Sección 4 se define matemáticamente el problema de Escalamiento Multidimensional y se plantea el problema general de optimización que es objeto de análisis en estas técnicas, exponiéndose también la clasificación de algoritmos numéricos en el Escalamiento Multidimensional debida a Trosset (1993) y los nuevos enfoques de trabajo que aplican métodos de Optimización Estocástica. Un ejemplo clásico de la literatura, el de *percepción de las diferencias entre naciones*, es usado para ilustrar algunos aspectos, sin ninguna pretensión de comparación entre técnicas dado que los resultados mostrados son obtenidos partiendo de elementos diversos y no comparables.

2. ESCALAMIENTO MULTIDIMENSIONAL: CONCEPTOS BASICOS

En las técnicas de escalamiento multidimensional, el primer requisito trata con los conceptos de *objeto-estímulo* y de *similaridad–disimilaridad–distancia* y el segundo requisito se refiere al *procedimiento* para alcanzar una configuración de puntos que refleje las similaridades observadas o percibidas.

Los términos *objeto* y *estímulo* se usan de manera intercambiable. Realmente un *objeto* es simplemente una cosa, un individuo o un evento, mientras que *estímulo* se refiere al objeto percibido, o sea, a la percepción del objeto.

Las *medidas de semejanza*, como una aplicación de valores numéricos que permiten expresar numéricamente el vínculo existente entre estímulos, son aquí de importancia capital. Los conceptos de *similaridad, disimilaridad y distancia*, como medidas de semejanza, poseen propiedades específicas que deben tenerse en cuenta. (Linares (1990)).

El *procedimiento*, en términos muy generales, sigue algunas ideas básicas en la mayoría de las técnicas. El punto de partida es una matriz de disimilaridades entre n objetos, con el elemento δ_{ij} en la fila i y en la columna j, que representa la disimilaridad del objeto i al objeto j. También se fija el número de dimensiones, p, para hacer el gráfico de los objetos en una solución particular. Generalmente el camino que se sigue es:

1) Arreglar los n objetos en una configuración inicial en p dimensiones, esto es, suponer para cada objeto las coordenadas $(x_1, x_2, ..., x_p)$ en el espacio de p dimensiones.

2) Calcular las distancias euclidianas entre los objetos de esa configuración, esto es, calcular las d_{ij}, que son las distancias entre el objeto i y el objeto j.

3) Hacer una regresión de d_{ij} sobre δ_{ij} . Esta regresión puede ser lineal, polinomial o monótona. Por ejemplo, si se considera lineal se tiene el modelo

$$d_{ij} = a + b\delta_{ij} + \varepsilon$$

y utilizando el método de los mínimos cuadrados se obtienen estimaciones de los coeficientes a y b, y de ahí puede obtenerse lo que genéricamente se conoce como una “disparidad”

$$\hat{d}_{ij} = \hat{a} + \hat{b}\delta_{ij}$$

Si se supone una regresión monótona, no se ajusta una relación exacta entre d_{ij} y δ_{ij} , sino se supone simplemente que si δ_{ij} crece, entonces d_{ij} crece o se mantiene constante.

4) A través de algún estadístico conveniente se mide la bondad de ajuste entre las distancias de la configuración y las disparidades. Existen diferentes definiciones de este estadístico, pero la mayoría surge de la definición del llamado *índice de esfuerzo* (en inglés: STRESS).
Los criterios más utilizados son los dos siguientes:

$$\text{STRESS1} = \sqrt{\frac{\sum\sum\left(d_{ij} - \hat{d}_{ij}\right)^2}{\sum\sum d_{ij}^2}}$$

y

$$\text{SSTRESS1} = \sqrt{\frac{\sum\sum\left(d_{ij}^2 - \hat{d}_{ij}^2\right)^2}{\sum\sum d_{ij}^4}}$$

Todas las sumatorias sobre i y j van de 1 a p y las disparidades dependen del tipo de regresión utilizado en el tercer paso del procedimiento.

El STRESS1 es la fórmula introducida por Kruskal quien ofreció la siguiente guía para su interpretación:

TAMAÑO DEL STRESS1	INTERPRETACION
0.2	Pobre
0.1	Regular
0.05	Bueno
0.025	Excelente
0.00	Perfecto

5) Las coordenadas (x_1, x_2, ..., x_t) de cada objeto se cambian ligeramente de tal manera que la medida de ajuste se reduzca.

Los pasos del 2 al 5 se repiten hasta que al parecer la medida de ajuste entre las disparidades y las distancias de configuración no puedan seguir reduciéndose. El resultado final del análisis es entonces las coordenadas de los n objetos en las p dimensiones. Estas coordenadas pueden usarse para elaborar un gráfico que muestre cómo están relacionados los objetos. Lo ideal sería encontrar una buena solución en menos de tres dimensiones, pero esto no es siempre posible.

3. LA MATRIZ DE SIMILARIDADES

La entrada básica de un análisis de escalamiento multidimensional son los valores de similaridad o disimilaridad entre todos, o casi todos, los pares de *n* objetos. Estos datos se nombran genéricamente *similaridades* o *proximidades.* Existen diversas maneras de generar estos valores, aunque las dos maneras típicas son (1) preguntar a los sujetos acerca de la similaridad entre todos los pares de estímulos y/o (2) pedir a los sujetos que clasifiquen los estímulos sobre la base de descriptores tales como adjetivos. De la primera manera se obtienen las llamadas *similaridades directas,* mientras que de la segunda manera se tienen las *similaridades derivadas.*

Similaridades directas. Vimos que el término *similaridad directa* se refiere al caso cuando a los sujetos se les presentan pares de estímulos y se les pide que emitan un juicio de su similaridad.

Los juicios de similaridad se pueden obtener de maneras muy diferentes. Veamos algunos de los métodos de registrar los juicios.

- Hacer una marca sobre una recta.
- Estimación de la magnitud directa.
- Colocar o clasificar.
- Ordenar parejas.
- Ternas.
- Determinar el orden de los rangos.

Nótese que la dificultad de la recolección de datos de similaridad está determinada por el número de estímulos considerados. Si el número de estímulos es grande, el número de comparaciones es grande.

Aunque para evitar tener que recoger un número demasiado grande de juicios se pudiera limitar el número de estímulos, es deseable incluir tantos estímulos como prácticamente sea posible. El uso de un número muy pequeño de estímulos hace que las soluciones en pocas dimensiones sean inestables. Algunos autores recomiendan tener de 9 a 12 estímulos para soluciones bidimensionales y de 13 a 18 estímulos para soluciones tridimensionales.

Similaridades derivadas. Este término se origina del hecho que los datos de similaridades se construyen o derivan de los rangos que los sujetos dan a cada estímulo según un conjunto de descriptores verbales. Hay que señalar que los descriptores verbales son altamente subjetivos y también, a menudo, conceptualmente incompletos, puesto que es bastante improbable que todas las dimensiones relevantes contenidas en las diferencias entre los estímulos puedan lograrse usando adjetivos que las describan.

Típicamente, se le pide a cada sujeto que indique la magnitud en que cada adjetivo describe el estímulo que está siendo evaluado, asignando, digamos un número entre 1 (describe el estímulo muy bien) y 100 (no describe en nada el estímulo). Por ejemplo, se pudieran usar las siguientes frases adjetivas para evaluar marcas de café: sabor fuerte, para personas sociables, imprescindible después de comer, bajo en cafeína, un lindo envase, etc.

El uso de datos de adjetivos con rangos supone que el conjunto de frases adjetivas seleccionadas brinda conceptualmente una lista completa de descriptores verbales que dan razón de las principales causas de las diferencias entre estímulos. Una práctica recomendada es hacer entrevistas, previas al experimento, a grupos de personas típicas para identificar las dimensiones importantes de la comparación.

Una variante del método anterior, conocida con el nombre de bipolar, es pedirle al sujeto que le ponga un valor, generalmente en una escala de -10 a 10, a cada estímulo con respecto a un número de atributos. Esto resulta en un arreglo de tres entradas de los datos: estímulo, atributo y sujeto y tiene la propiedad que los atributos pueden incluirse en el gráfico de los estímulos.

Existen muchas otras técnicas, que tratan de ordenar los estímulos de una u otra manera. Como puede apreciarse el componente metodológico de la investigación va a determinar ciertas características de la matriz de similaridades o de disimilaridades inicial.

3.1. Clasificación de técnicas de escalamiento multidimensional según las características de la matriz de similaridades. (Gower & Digby, 1981)

La matriz de similaridades puede consistir de medidas de similaridad δ_{ij} entre los objetos i y j que son iguales a δ_{ji} (matriz simétrica), pero también puede ser que δ_{ij} no es igual a δ_{ji}. (matriz asimétrica).

Las principales técnicas para matrices simétricas son:

- Escalamiento métrico. Entre estos métodos puede citarse el análisis de coordenadas principales que usa los valores δ_{ij} directamente. El proceso computacional de las coordenadas principales es similar al del análisis de componentes principales, técnica exploratoria de datos que parte de una matriz de datos observados. Un caso especial del análisis de coordenadas principales ocurre cuando los datos son distancias euclidianas obtenidas de la matriz de datos originales Y de n filas y p columnas; en este caso, existe una dualidad entre un análisis de coordenadas principales sobre una matriz de dispersión y un análisis de componentes principales sobre una matriz de distancias.
- "Unfolding". Cuando la matriz de similaridades tiene valores perdidos, el escalamiento métrico es impracticable, pero un caso especial es cuando de la matriz triangular inferior, por ejemplo, se tiene una tabla de dos entradas de dimensión r x s y con sólo estos datos pueden encontrarse las coordenadas de los r + s puntos.
- Escalamiento no métrico. Aunque en esencia estos métodos son similares a su contrapartida métrica en que encuentran localizaciones para los n objetos que mantienen aproximadamente las asociaciones entre los objetos, esto se hace utilizando una escala ordinal de tal manera que el criterio de bondad de ajuste depende sólo del orden de los valores δ_{ij} y no de sus valores absolutos.

Para las matrices asimétricas existen dos clases de métodos, a saber,

- *Métodos de primera clase.* Se modela la tabla completa.

• *Métodos de segunda clase.* Se modela la tabla en dos partes, primero la componente simétrica y
después la asimétrica.

Los trabajos de Harshman **et al** (1982), Kiers **et al** (1990) y Kiers & Takane (1994) tratan el problema de la representación gráfica de las relaciones asimétricas de los datos y explican los algoritmos DEDICOM, GIPSCAL y GIPSCAL Generalizado.

Otros métodos surgen cuando se tiene más de una matriz de similaridades. Claro que pudieran ser analizadas por los métodos mencionados y producir gráficos asociados, pero surge el problema de cómo combinar todos estos gráficos con alguna *configuración promedio* que indique cómo los gráficos individuales difieren del promedio. Estos métodos se conocen como métodos gráficos para tablas de tres entradas y entre ellos cabe mencionar los siguientes:

Escalamiento individual (en inglés: INDSCAL).	Opera sobre m matrices de distancias d_{ijk} (distancia entre los objetos i y j para configuración k) que han sido derivadas de m matrices δ.
Análisis de Procrustes Generalizado	Opera sobre m matrices, cada una de tamaño n x p y da las coordenadas de los n puntos en p dimensiones, las cuales han sido derivadas de las matrices de distancias. Veamos a continuación un ejemplo que ilustra algunas de estas técnicas.

3.2. EJEMPLO: Percepción de las diferencias entre naciones

Un ejemplo clásico de la literatura es este de la percepción que tienen 18 estudiantes acerca de las diferencias entre 12 naciones. Los datos originales aparecieron en Wish (1970), Wish (1971) y Wish, Deutsch and Biener (1972) y es citado en Kruskall (1976). En la mayoría de los textos que tratan el tema de Escalamiento Multidimensional como Dillon y Goldstein (1984) y Cuadras(1990) lo toman como ejemplo. También es ejemplo de prueba de paquetes estadísticos como el STATISTICA(1998). El planteamiento del mismo es el siguiente: “En un estudio piloto sobre la percepción sobre diferentes naciones llevado a cabo a principio del año 70, cada uno de los 18 estudiantes que participaron en el estudio atribuyó una nota entre 1 (para los países muy diferentes) y 7 (para los muy similares) a cada uno de los 66 pares formados en el conjunto de las 12 naciones consideradas”.

Así, pudo obtenerse una tabla de datos que muestra la media de los rangos de similaridad entre 12 países de acuerdo al criterio de los 18 estudiantes y de la que se obtuvo la siguiente tabla después de hacer las normalizaciones requeridas.

Nótese que en esta manera de proceder, se ignoran las posibles diferencias de percepción entre los sujetos. Puede observarse que los valores más bajos en la tabla indican que, por ejemplo, China y Brasil y USA y el Congo son percibidos con las mayores diferencias, mientras que Yugoslavia y Rusia y USA y Japón se perciben como los más similares ya que tienen los valores más altos.

Utilizando el módulo “Multidimensional Scaling” del paquete STATISTICA (1998) se llevó a cabo el análisis de los datos. Se tomó como configuración inicial la obtenida por el algoritmo de Guttman-Lingoes y a continuación usando un algoritmo iterativo con el método de descenso máximo (en inglés: *steepest desdent*) se alcanzó una configuración final cuyo gráfico 1 se muestra debajo. El STRESS1 obtenido fue de 0.189 que según la guía ofrecida anteriormente puede calificarse de “regular”.

Recordemos que la orientación real de los ejes, en el escalamiento multidimensional, es arbitraria. Así, si se desea, pudiera rotarse la configuración para alcanzar una solución más fácil de interpretar. El gráfico muestra la oposición que existe entre países desarrollados y subdesarrollados y países occidentales y comunistas. Estos datos se recogieron en los años 70, reflejándose el “status socio político” de las naciones estudiadas en esa época. Podemos nombrar estas dimensiones como “Desarrollo Económico” para la 1 y “Alineamiento Político” para la 2.

Tabla 1. Tabla de similaridades de la percepción sobre 12 naciones

	BRA	CON	CUB	EGY	FRA	IND	ISR	JAP	CHI	RUS	USA	YUG
BRA	7.0	4.8	5.3	3.4	4.7	4.5	3.8	3.5	2.4	3.1	5.4	3.2
CON	4.83	7.0	4.6	5.0	4.0	4.8	3.3	3.4	4.0	3.4	2.4	3.5
CUB	5.20	4.56	7.0	5.2	4.1	4.0	3.6	2.9	5.5	5.4	3.2	5.1
EGY	3.44	5.00	5.17	7.0	4.8	5.8	4.7	3.8	4.4	4.4	3.3	4.3
FRA	4.72	4.00	4.11	4.78	7.0	3.4	4.0	4.2	3.7	5.1	5.9	4.7
IND	4.50	4.83	4.00	5.83	3.44	7.0	4.1	4.5	4.1	4.5	4.3	4.0
ISR	3.83	3.33	3.61	4.67	4.00	4.11	7.0	4.8	3.0	4.2	5.9	4.4
JAP	3.50	3.39	2.94	3.83	4.22	4.50	4.83	7.0	4.2	4.6	6.1	4.3
CHI	2.39	4.00	5.50	4.39	3.67	4.11	3.00	4.17	7.0	5.7	2.6	5.1
RUS	3.06	3.39	5.44	4.39	5.06	4.50	4.17	4.61	5.72	7.0	5.0	6.7
USA	5.39	2.39	3.17	3.33	5.94	4.28	5.94	6.06	2.56	5.00	7.0	3.6
YUG	3.17	3.50	5.11	4.28	4.72	4.00	4.44	4.28	5.06	6.67	3.56	7.0

Nota: La matriz triangular inferior se da con dos decimales y la superior con sólo uno)

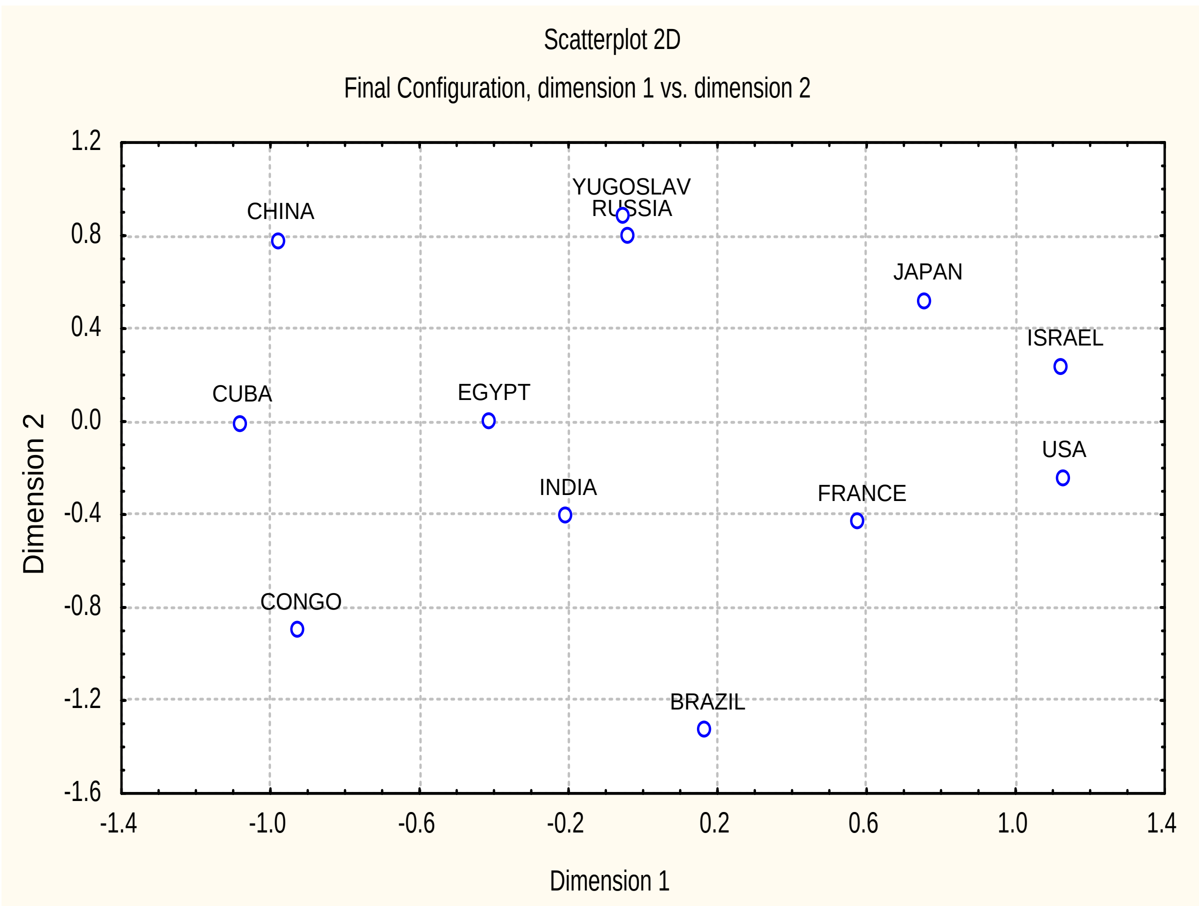

Gráfico 1. Configuración final de la percepción de 12 naciones.

4. LOS ALGORITMOS

Vamos a definir primero el problema de Escalamiento Multidimensional en una forma matemática para ello decimos que:

"Existe un conjunto de n objetos o estímulos que denotaremos por Ω y se conoce una matriz de disimilaridades $\Delta = (\delta_{ij})$ que expresa las disimilaridades entre cada pareja de estos objetos. Fijando una dimensión p, generalmente pequeña (p = 1, 2 ó 3) se desea obtener una matriz de configuración de puntos $X_{pxn,}$ donde el vector columna $x_i \in R^p$. La distancia euclidiana p dimensional d_{ij} es la distancia entre los puntos x_i y x_j, y $D_{nxn} = (d_{ij}) = D(X)$ es la matriz de distancias euclidianas interpuntos. El conjunto de todas las matrices D_{nxn} se denota por $D_n(p)$. Limitándonos al caso de matrices de disimilaridad de dos entradas. El problema consiste en construir una matriz de configuración de puntos X_{pxn} para la cual las distancias interpuntos D(X) se aproximen lo más posible a las disimilaridades Δ".

Obsérvese que en el problema anterior se plantea la solución de problemas de optimización de la forma

$$\text{minimizar } \rho\,(D, \Delta)$$

$$\text{sujeto a } D \in D_n(p),$$

donde ρ es alguna medida de discrepancia entre las matrices de distancia y las matrices de disimilaridades. Existe en la literatura una gran variedad de estas medidas pero las más populares son los criterios STRESS y SSTRESS que mencionamos antes.

4.1. Taxonomía según el problema de optimización (Trosset, 1993)

Trosset (1993) propone una taxonomía para los problemas de Escalamiento Multidimensional según los problemas de optimización involucrados y plantea que los algoritmos numéricos de estas técnicas pueden verse como algoritmos para problemas de optimización numérica con ciertas características específicas. Así, hace una distinción entre disimilaridades que representan datos observados y disimilaridades que sirven como variables libres en problemas de optimización.

En el primer caso incluye el primer enfoque riguroso de Escalamiento Multidimensional propuesto por Torgerson en 1952 y llamado por Gower en 1966 *análisis de coordenadas principales* y corrientemente conocido por escalamiento *clásico.* En este caso se obtiene un mínimo global al problema de optimización planteado. En contraste al escalamiento clásico, las técnicas que requieren la minimización de STRESS o SSTRESS requieren de algoritmos iterativos para la optimización numérica. Numerosos algoritmos han sido propuestos para minimizar los criterios citados anteriormente, pero casi siempre quedan atrapados en óptimo locales.

En el segundo caso incluye aquellos problemas donde existe un número muy grande de objetos y es necesario imponer restricciones a la matriz de disimilaridades Δ. También en este caso ha surgido gran variedad de técnicas entre las que puede citarse el popular algoritmo ALSCAL que es el procedimiento utilizado en el paquete estadístico SPSS. (Statistical Package for the Social Sciencies)

4.2. Enfoques actuales

Como hemos podido apreciar anteriormente, los algoritmos existentes para resolver el problema de Escalamiento Multidimensional no garantizan la obtención de mínimos globales, por lo que es necesario aplicar otros métodos que sí lo garanticen. En los últimos años se vienen aplicando métodos de Optimización Estocástica ya que en la mayoría de estos métodos se puede demostrar la convergencia asintótica al óptimo global. Entre los más conocidos se encuentran el de Sobrecalentamiento Simulado (en inglés: *Simulated Annealing)* y el de Algoritmos Genéticos.

El método de Sobrecalentamiento Simulado se basa en el conocido algoritmo de Metropolis que utiliza técnicas de Monte Carlo para generar cadenas de Markov. El interés de este método es encontrar el estado más probable de una cadena de Markov. Así si este estado es k, entonces la probabilidad π_k de este estado es mayor que la de cualquier otro estado i para todo $i \neq k$. Para acentuar el peso dado al estado k, la distribución de equilibrio π puede reemplazarse por la distribución de probabilidad

$$\pi_i^{(\tau)} = \frac{\pi_i^{1/\tau}}{\sum_j \pi_j^{1/\tau}}$$

sobre el estado i. Aquí τ es un parámetro positivo pequeño llamado tradicionalmente temperatura. Con una cierta densidad simétrica, la distribución $\pi_i^{(\tau)}$ puede obtenerse al correr una cadena con la probabilidad de aceptación de Metropolis

$$a_{ij} = \min\left\{ \left(\frac{\pi_j}{\pi_i}\right)^{1/\tau}, 1 \right\}.$$

Lo que de hecho hace el método de Sobrecalentamiento Simulado es correr la cadena haciendo decrecer gradualmente τ hacia 0. (Lange, K., 1999)

Villalobos (1998) elaboró un procedimiento que aplica este método a la solución del problema de Escalamiento Multidimensional. En el Algoritmo SCAL-SS (SCALing con Sobrecalentamineto Simulado) la función a optimizar, Stress, es la siguiente:

$$\sigma^2(I) = \sum\sum w_{ij} (\delta_{ij} - d_{ij}(I))^2$$

donde I es un estado que representa una determinada configuración, $d_{ij}(I) = d(x_i, x_j) = || x_i - x_j ||$ y las sumatorias son para i, j =1, ..,n.

Los Algoritmos Genéticos se utilizan para obtener el máximo de una función positiva pero pueden adaptarse para obtener mínimos. Gouenet(1995) propone una solución genética al problema del Escalamiento Multidimensional.

4.2.1. *Otras soluciones al ejemplo de la percepción de las naciones*

Al aplicar el Algoritmo Genético a la matriz de proximidades del ejemplo de la diferencia de percepción sobre 12 naciones, Gouenet (1995) reporta un STRESS = 0.0013, obtenido con 500 iteraciones a partir de una configuración aleatoria. La configuración final, también en dos dimensiones, es similar a la obtenida anteriormente. La interpretación dada por el autor a esta configuración es:

"Existe un posicionamiento muy interesante de los países sobre el plano:

- USA, Japón, Francia e Israel son países capitalistas desarrollados.
- Congo, Egipto e India son países subdesarrollados.
- Cuba, China, Rusia y Yugoslavia son países comunistas".

Obsérvese, que el valor del STRESS puede calificarse de "excelente", aunque la configuración final obtenida permite la misma interpretación que anteriormente.

Por otra parte, al aplicar el algoritmo SCAL-SS al ejemplo de las naciones el mejor valor de STRESS obtenido fue 0.0474 y el segundo mejor valor fue 0.0477. Este algoritmo garantiza la localización de los óptimos globales o valores muy cercanos en la mayoría de las ejecuciones.

A pesar de que no pueden hacerse comparaciones entre los algoritmos ya que se utilizan diferentes criterios es evidente el mejoramiento en las soluciones que utilizan los métodos de Optimización Estocástica.

5. UNA APLICACION EN PSICOPEDAGOGIA

En el análisis exploratorio de datos las técnicas de Escalamiento Multidimensional tienen gran aplicación. Múltiples son los campos donde estas técnicas contribuyen al esclarecimiento del fenómeno bajo estudio. En particular, en las investigaciones psicopedagógicas pueden ser de gran utilidad para los investigadores ya que a través de ellas se descubren relaciones complejas difíciles de observar directamente. En una investigación llevada a cabo en el Centro de Perfeccionamiento para la Educación Superior de la Universidad de La Habana (CEPES-UH) sobre la formación psicopedagógica del profesorado universitario, es de interés establecer un diagnóstico sobre los aspectos de reflexión y criticidad sobre la labor que desempeña el profesorado. Para ello se seleccionaron aleatoriamente 12 profesores que representaban los tres factores siguientes:

Especialidad con tres niveles: 1-Química, 2-Lengua Inglesa y 3- Idioma (servicio)

Categoría Docente con dos niveles: 1-Principal, 2- No Principal

Experiencia Docente con dos niveles: 1- Muy experimentado, 2- Menos experimentado.

A los profesores seleccionados se les aplicó instrumentos elaborados al efecto donde se midieron las características "reflexión sobre su labor pedagógica" en tres dimensiones y "criticidad sobre su labor pedagógica" en dos dimensiones. Estas dimensiones fueron medidas en una escala ordinal de 1 a 3 que reflejaban las categorías "alto", "medio" y "bajo". A partir de la tabla de datos originales se obtuvo una matriz de disimilaridades utilizando el coeficiente "porciento de desacuerdo" que posee el sistema STATISTICA en su módulo de análisis de conglomerados jerárquico. La tabla No. 2 muestra esta matriz de disimilaridades a partir de la que se llevó a cabo un análisis de Escalamiento Multidimensional con las características explicadas en el epígrafe 3.2. Obsérvese que las entradas de esta tabla son los 12 profesores que caracterizan los tres factores mencionados anteriormente y que por simplicidad se denotan con el triplo (abc) donde a es el nivel de la Especialidad, b es el nivel de la Categoría Docente y c es la Experiencia Docente.

Tabla 2. Tabla de disimilaridades de un estudio psicopedagógico sobre formación del profesorado universitario. (Datos proporcionados por el CEPES-UH, agosto del 2000).

	(111)	**(112)**	**(121)**	**(122)**	**(211)**	**(212)**	**(221)**	**(222)**	**(311)**	**(312)**	**(321)**	**(322)**
(111)	.00	.71	.85	1.0	.14	.85	.57	1.0	.00	.71	.00	1.0
(112)		.00	.28	1.0	.57	.57	.42	.28.	.71	.42	.71	1.0
(121)			.00	.85	.85	.42	.42	.28	.85	.42	.85	.85
(122)				.00	1.0	.42	1.0	1.0	1.0	.85	1.0	.14
(211)					.00	.85	.42	.85	.14	.57	.14	1.0
(212)						.00	.71	.71	.85	.71	.85	.42
(221)							.00	.42	.57	.28	.57	1.0
(222)								.00	1.0	.42	1.0	.85
(311)									.00	.71	.00	1.0
(312)										.00	.71	.85
(321)											.00	1.0
(322)												.00

Puede observarse que dado que este coeficiente de disimilaridad toma valores de 0 a 1, los valores pequeños muestran individuos mas parecidos, mientras que los valores grandes señalan diferencias. Así por ejemplo, además de los ceros de la diagonal principal, existen ceros que asocian los individuos (311) y (321) con (111) y valores muy pequeños de .14 que relacionan a los individuos (322) y (122) y (321) y (211). La configuración final se muestra en el Gráfico 2 que se muestra a continuación la que se alcanzó con un STRESS1 de 0.060 que podemos calificar de "bueno".

El gráfico muestra en el primer eje oposición entre los profesores de mucha experiencia y aquellos de menos. Nótese la cercanía entre los puntos (322) y (122) a la derecha que representan los profesores de menos experiencia, mientras que a la izquierda los puntos (111), (311), (321) y (211) representan los profesores muy experimentados. Con respecto al segundo eje los puntos más distantes son el (122) y el (322) situados en la parte superior del gráfico y el punto (222) en la parte inferior, que apuntan hacia una diferenciación entre las especialidades de Química y de Idioma (servicios) y la especialidad de Lengua Inglesa. Los resultados del diagnóstico obtenido utilizando una de las técnicas del Escalamiento Multidimensional pueden corroborarse con otras técnicas estadísticas univariadas y multivariadas y reflejan las ideas intuitivas de los investigadores de este problema.

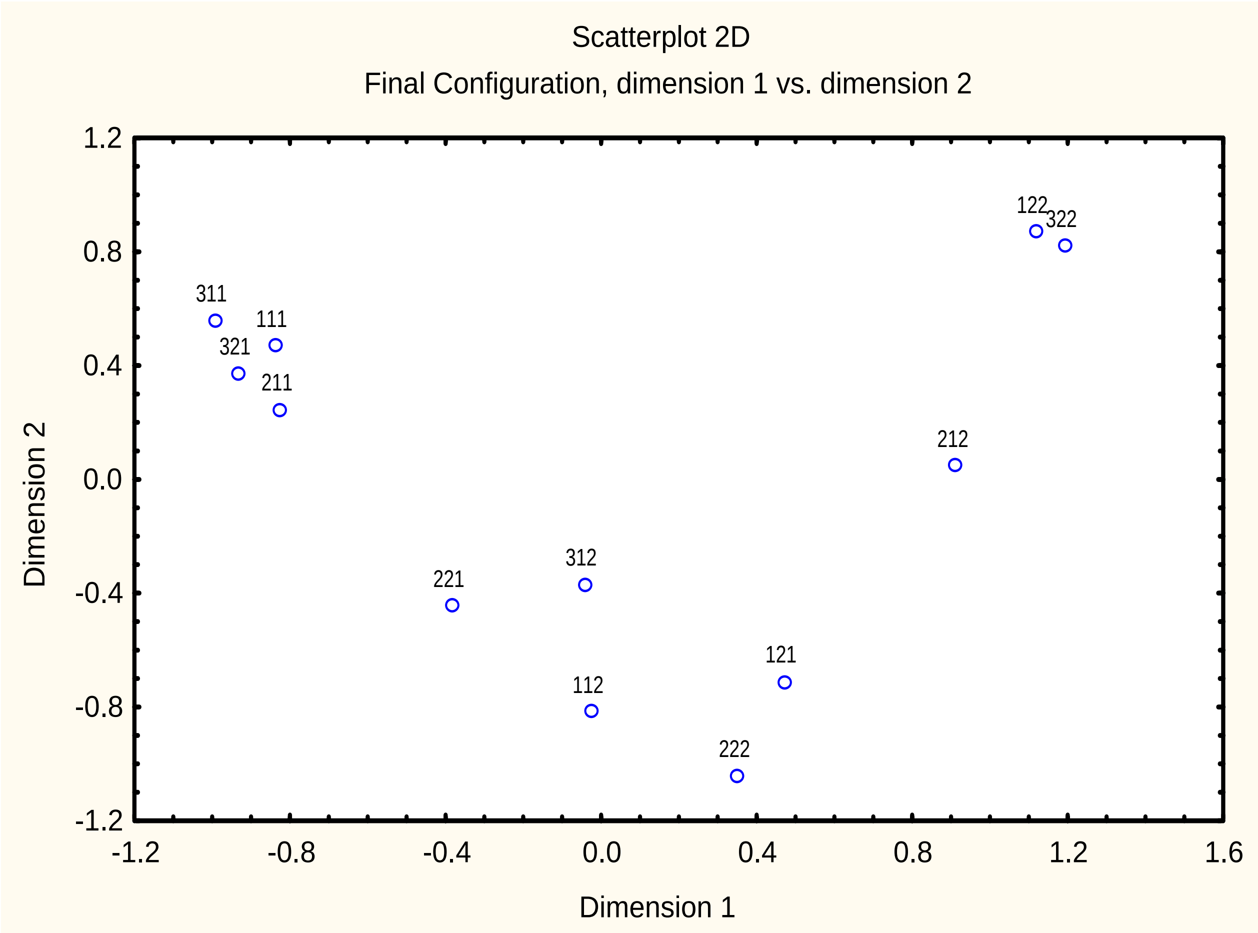

Gráfico 2. Configuración final de un estudio psicopedagógico.

6. CONCLUSIONES

Las técnicas de Escalamiento Multidimensional son una herramienta estadística importante para las investigaciones empíricas ya que permiten la representación gráfica de relaciones complejas. Un vasto campo de investigación está abierto para idear mejores técnicas que permitan conseguir su objetivo esencial: lograr configuraciones de puntos cuyas interdistancias se acerquen lo más posible a las similaridades percibidas. La aplicación desarrollada en una investigación psicopedagógica permite apreciar las posibilidades de estas técnicas.

REFERENCIAS

CUADRAS, C.M. (1991): "Métodos de Análisis Multivariante", PPU. Barcelona. España, 424-427.

DILLON, W.R. and M. GOLDSTEIN (1984): "Multivariate Analysis. Methods and Applications", John Wiley & Sons. New York, 111–115.

GOUENET, R.F. (1995): "Une Solution Genetique de la MDS". En: Analyse geometrique des donnees de disissimilarite par la multidimensional scaling: une approche parallele basee sur les álgorithmes genetiques. Application aux sequences biologiques". These de Doctorat, Universite de Rennes I. Francia, 94-95.

GOWER, J.C. and P.G.N. DIGBY (1981): "Expressing Complex Relationships in Two Dimensions". En: **Interpreting Multivariate Data** Vic Barnett (Editor). John Wiley. Chichester.

HARSHMAN, R.A.; P:E. GREEN; Y. WIND and M.E. LUNDY (1982): "A Model for the Analysis of Asymmetric Data in Marketing Research", **Marketing Science**, 1, 205-242.

KIERS, H.A.L.; J.M.F TEN BERGE, Y. TAKANE and J. de LEEUW (1990): "A Generalization of Takane's Algorithm for DEDICOM", **Psychometrika***,* 55, 151-158.

KIERS, H.A.L., and Y. TAKANE (1994): "A Generalization of GIPSCAL for the Analysis of Nonsymmetric Data", **Journal of Classification***,* 11, 79-99.

KRUSKAL, J. (1976): "More factors than subjects, test and treatment: an indeterminacy theorem for canonical decomposition and individual differences" , **Psychometrika** 41, 281-293.

LANGE, K. (1999): **Numerical Analysis for Statisticians**, Springer-Verlag, New York, Inc, 339.

LINARES, G. (1990): Análisis de Datos. Universidad de La Habana, MES, Cuba, 349-354.

STATISTICA for Windows (1998): StatSoft, Inc. EEUU.

TROSSET, M.W. (1993): "Numerical Algorithms for Multidimensional Scaling" En: **Information and Classification***,* R. Klar & O. Opitz (Eds). Springer, Heidelgerg, 81-92.

VILLALOBOS, M.A. (1998): Optimización Estocástica para el Análisis de Proximidades. Tesis de Magister Scientiae. Universidad de Costa Rica. San José, Costa Rica.

WISH, M. (1970): "Comparisons among multidimensional structures of nations based on differents measures of subjective similarity", **General Systems** 15, 55-65 (L. Von Bertalanffy y A. Rapoport, eds.)

WISH, M. (1971): "Individual differences in perceptions and preferences among nations". En: **Attitude Research in Transition***,* Russell I. Haley (Eds.). Chicago: American Marketing Association.

WISH, M.; M. DEUTSH and L. BIENER, L. (1972): "Differences in perceived similarity of nations". En: **Multidimensional Scaling: Theory and Applications in the Behavorial Sciencies***,* 2. A.K. Romney; R.N. Shepard and S. Nervole (Eds). New York: Seminar Press, 289–313.