

Escalamiento Multidimensional empleando Metaheurísticas

Multidimensional Scaling by using Metaheuristics

María Esther Reyes Calzado^{1*}

Resumen Dentro del Análisis Multivariado, el Escalamiento Multidimensional se utiliza para obtener una representación en un espacio de dimensión reducida de ciertos individuos u objetos cuyas similitudes o disimilitudes fueron obtenidas previamente. En este trabajo tomando una matriz de distancias, resultado del MDS-Clásico, que bajo cierto criterio de error (Strain) es óptima, se considera otro criterio de error (STRESS) a minimizar, empleando MDS No Métrico con Metaheurísticas. Se implementan algoritmos en lenguaje MATLAB y los resultados obtenidos son comparados con los de métodos conocidos.

Abstract In Multivariate Analysis, Multidimensional Scaling (MDS) is used to obtain a representation in a lower dimensional space for certain individuals or objects which similarities or dissimilarities were previously collected. In this work a minimization of the error criterion Stress employing Non Metric MDS with Metaheuristics is applied after using the Strain minimization with Classic MDS as an initial distance matrix. Algorithms are implemented in MATLAB language and its results are compared with the corresponding in the MDS known methods.

Palabras Clave

Escalamiento Multidimensional — Metaheurísticas — Proximidades — Stress

¹Departamento de Matemática, Universidad de la Habana, Cuba, m.reyes@matcom.uh.cu

*Autor para Correspondencia

Introducción

Dentro de la estadística descriptiva, el análisis exploratorio de datos multivariados, agrupa los métodos que pretenden representar datos agrupados en una matriz de entrada, que tiene asociada cierta matriz de proximidades. La esencia de estos métodos consiste en hacer un conjunto de transformaciones a la matriz de datos inicial que culminan con la aplicación del teorema de descomposición singular de Eckart & Young (1936).

El Escalamiento Multidimensional (MDS), se encarga de representar ciertos objetos o individuos de los que sólo se conocen similitudes o disimilitudes debidas a su comportamiento respecto a variables cualitativas o mixtas que les han sido medidas previamente, encontrando en un espacio de dimensión reducida, puntos cuyas distancias euclídeas se aproximen lo mejor posible a las relaciones de proximidades originales.

La implementación de una herramienta computacional propuesta en la literatura para la solución de este problema resulta una labor difícil debido a que la información suministrada no es suficiente y muchas veces la descripción del algoritmo no es muy clara de modo que permita entender la técnica para su aplicación posterior.

En este trabajo, se elaboró un paquete de programas en lenguaje MATLAB de algoritmos que combinan el Escalamiento Clásico (cuyo criterio de error a minimizar es el *Strain*) y

el No métrico con Metaheurísticas (minimizando la función *Stress*) con el objetivo de obtener resultados aceptables que superen deficiencias numéricas de los métodos más conocidos.

Se compararon las estrategias programadas a partir de sus resultados en ejemplos de la literatura y reales. En todos los casos se logró minimizar el resultado del MDS Clásico y se comprobó que se mantenían las relaciones de semejanzas originales.

1. Escalamiento Multidimensional

Sean $\Omega = \{\omega_1, \omega_2, \dots, \omega_n\}$ un conjunto de n elementos diferentes, y $\delta_{ij} = \delta(i, j) = \delta(j, i) \geq \delta(i, i) = 0$ una distancia o disimilaridad entre los elementos i, j de Ω . Consideremos la matriz de distancias:

$$\Delta = \begin{pmatrix} \delta_{11} & \delta_{12} & \cdots & \delta_{1n} \\ \delta_{21} & \delta_{22} & \cdots & \delta_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ \delta_{n1} & \delta_{n2} & \cdots & \delta_{nn} \end{pmatrix}$$

$$\delta_{ij} = \delta_{ji} = \delta(i, j) \geq \delta_{ii} = 0$$

El objetivo del **Escalamiento Multidimensional (Multidimensional Scaling, MDS)** es encontrar una matriz X de la forma

$$X = \begin{pmatrix} x_{11} & x_{12} & \cdots & x_{1p} \\ x_{21} & x_{22} & \cdots & x_{2p} \\ \vdots & \vdots & \ddots & \vdots \\ x_{n1} & x_{n2} & \cdots & x_{np} \end{pmatrix}$$

tal que $\delta_{ij}^2 = \sum_{\alpha=1}^p (x_{i\alpha} - x_{j\alpha})^2 = (x_i - x_j)'(x_i - x_j)$, donde $x_1, x_2, \dots, x_n \in \mathbb{R}^p$ serán las coordenadas de los puntos que representan a los elementos de Ω . Si esto se cumple, decimos que Δ es una matriz de distancias euclídeas. [5]

Los modelos de MDS pueden clasificarse fundamentalmente en Métricos y No Métricos. En el Escalamiento Métrico se utiliza una función paramétrica, y es recomendable cuando la matriz de disimilaridades es bien condicionada. El Escalamiento No Métrico trabaja con una función monótona e intenta aproximar la mejor representación en el espacio euclídeo. [1] [4]

El caso más simple de MDS es el Escalamiento Clásico, también llamado Análisis de Coordenadas Principales (Análisis de Componentes Principales aplicado a una matriz de distancias), donde se optimiza por vía algebraica una función de pérdida, conocida como *Strain*. [8] [4]

En el Escalamiento No Métrico, la representatividad de una solución del MDS en p dimensiones se halla generalmente mediante la función de pérdida *Stress*. Los algoritmos Iterativos de MDS encuentran una configuración de MDS con óptimo *Stress*, para lo cual, re-escalan los datos siguiendo ciertas restricciones. [1] [2]

1.1 Escalamiento Clásico

Sea Δ , una matriz de distancias, $A = -\frac{1}{2}\Delta^{(2)}$ y $B = HAH$ donde $\Delta^{(2)}$ es la matriz que contiene en la posición i, j el cuadrado del elemento correspondiente a la matriz Δ ; $H = I - n^{-1}11'$ y 1 es el vector columna conformado por n unos. Se cumple el siguiente teorema:

Teorema 1: La matriz de distancias Δ es euclídea si y solo si B es semidefinida positiva.

Sea ahora $B = P\Lambda P'$ la descomposición espectral de B , donde P es una matriz de $n \times p$ vectores propios ortonormales de B correspondientes a los valores propios ordenados de la matriz Λ : $\lambda_1 \geq \dots \geq \lambda_p \geq \lambda_{p+1} = 0$.

Obsérvese que $B1 = 0$, y por tanto $\lambda_{p+1} = 0$ es valor propio de B correspondiente al vector propio 1 . También, se tiene que si $X = P\Lambda^{\frac{1}{2}}$ cumple que $B = XX'$.

Luego, las coordenadas del elemento i de Ω serán

$$x_i' = (x_{i1}, \dots, x_{ip})$$

donde x_i es la fila i -ésima de X . Éstas reciben el nombre de coordenadas principales y cumplen que

$$\delta_{ij}^2 = \sum_{\alpha=1}^p (x_{i\alpha} - x_{j\alpha})^2 = (x_i - x_j)'(x_i - x_j)$$

La función de pérdida más conocida para el MDS Clásico se llama *Strain* y viene dada por la expresión:

$$Strain = \|B - B(k)\| = tr(P_{(n-p)}\Lambda_{(n-p)}^2 P_{(n-p)}') = \sum_{i=1}^n \lambda_i^2 - \sum_{i=1}^p \lambda_i^2$$

En algunas aplicaciones, especialmente en Biología y Psicología, en lugar de una matriz de distancias Δ , se dispone de una matriz con coeficientes que miden el grado de similitud entre cada par de individuos. Para aplicar las técnicas de MDS es necesario transformar las similitudes en distancias, que pueden ser euclídeas o no. (Ver [8])

1.2 Escalamiento No Métrico

Supongamos que la matriz de distancias Δ es no euclídea. Entonces la matriz B (por el Teorema 1) tiene valores propios negativos: $\lambda_1 \geq \dots \geq \lambda_p \geq 0 > \lambda_{p+1} \geq \dots \geq \lambda_n$.

El fundamento del MDS no métrico es transformar las distancias δ_{ij} para convertirlas en euclídeas, pero conservando las relaciones de proximidad entre los elementos del conjunto Ω .

Sea $X_{(n \times p)}$ una configuración en \mathbb{R}^p , la expresión de X como solución del MDS en p dimensiones se halla generalmente mediante la función de pérdida *Stress* [1] [2] dada por:

$$Stress = \sqrt{\frac{\sum_{i < j} (d_{ij} - \hat{d}_{ij})^2}{\sum_{i < j} d_{ij}^2}}$$

donde los d_{ij} son las nuevas distancias euclídeas, obtenidas de las nuevas coordenadas para los puntos de Ω .

Esta función toma valores en el intervalo $[0, 1]$ y expresa si la representación obtenida refleja correctamente las relaciones de distancia originales. Una solución perfecta del MDS tiene *Stress* = 0. Se considera que una solución aproximada es suficientemente buena si el *Stress* es menor que 0.2. [8]

Otras variantes de la función de *Stress* son: el *Stress-normado* y el *S-Stress*. [1] [2]

Los algoritmos iterativos de MDS encuentran una configuración con óptimo *Stress*, para lo cual, re-escalan los datos siguiendo ciertas restricciones. Estos algoritmos presentan dos fases. En cada fase se fija un conjunto de parámetros y se modifica otro conjunto de argumentos de manera que se reduzca el *Stress*.

Si luego de t iteraciones, este proceso no reduce el valor de *Stress* hasta cierta cantidad prefijada, se detiene el algoritmo de búsqueda y X_t se toma como la solución óptima. Debe tenerse en cuenta que estos algoritmos iterativos no siempre garantizan que se encuentre una solución óptima global, debido a que los cambios en cada paso son pequeños y puede detenerse en un mínimo local de la función *Stress*.

Los métodos más conocidos son de tipo descenso de gradiente, pero muchas veces conducen a óptimos locales del *Stress*. Otros métodos, basados en una función de mayorización (método SMACOF) o el método de Tunelado (Tunneling)

tampoco garantizan que se llegue a un óptimo global; y otras implementaciones han demostrado ser un poco lentas. [1] [2]

2. Estrategias y procedimientos empleados

A partir de la configuración inicial dada por la solución del MDS Clásico, el proceso iterativo de MDS no métrico sigue las dos fases mencionadas anteriormente, en tanto no se cumpla alguno de ciertos criterios de parada previamente prefijados.

Denotaremos $X_{(0)}$ a la solución del MDS Clásico (solución inicial del algoritmo iterativo), $X_{(k)}$ la solución de la iteración k y $s_{(k)}$ el valor del stress para la solución $X_{(k)}$.

En la primera fase se generan posibles soluciones siguiendo una estrategia de vecino más cercano [7] en la que se relaciona cada elemento particular i con aquellos cuya distancia es mínima. En la segunda fase se calculan las nuevas distancias y procedemos a escoger la nueva solución a partir de la estrategia determinada por la metaheurística que se decida aplicar.

2.1 Recocido Simulado en MDS No Métrico

El algoritmo de Recocido Simulado se basa en el proceso de recocido del acero y cerámicas, una técnica que consiste en calentar y luego enfriar lentamente el material para variar sus propiedades físicas. El calor causa que los átomos aumenten su energía y que puedan así desplazarse de sus posiciones iniciales (un mínimo local de energía); el enfriamiento lento les da mayores probabilidades de recrystalizar en configuraciones con menor energía que la inicial (mínimo global).

El método consiste en, dado x_k , generar un punto y_k tal que, si el valor de la función objetivo disminuye, $x_{k+1} = y_k$. Si no, la aceptación de la solución responderá a cierta ley de probabilidad. Esta ley dependerá de un parámetro de temperatura t que se irá disminuyendo a medida que pasa el tiempo y teniendo en cuenta que $\Delta_k = f(y_k) - f(x_k) > 0$. Para un t fijo, a mayor Δ_k , menor probabilidad de tomar y_k como nuevo punto de iteración. La probabilidad de aceptar una solución que no mejore la actual es proporcional a la temperatura e inversamente proporcional al cambio en la función objetivo. Si consideramos $\{x_k\}$ una sucesión de variables aleatorias, puede demostrarse que el algoritmo converge en probabilidad al óptimo. (Ver [11] [3] [14])

Para nuestra solución al problema del MDS no métrico, en la iteración k , la solución $X_{(k)}$ se aceptó o no de acuerdo a una distribución de probabilidad normal ($e^{s/T}$), donde el parámetro de temperatura inicial $T^{(0)} = 1$, disminuye a medida que pasa el tiempo a razón de 0.95, es decir $T^{(k+1)} = 0.95 * T^{(k)}$.

2.2 Búsqueda tabú en MDS No Métrico

El algoritmo de Búsqueda Tabú tiene asociada una lista cuyos elementos no se pueden considerar (lista tabú). Utiliza un procedimiento de búsqueda local para moverse iterativamente desde una solución x hacia una solución x' en la vecindad de

x que no esté en la lista tabú, hasta satisfacer algún criterio de parada. Este método aumenta el rendimiento de búsqueda local mediante el uso de estructuras de memoria ya que una vez que una solución potencial queda determinada, se marca como "tabú" de modo que el algoritmo no vuelva a visitar esa posible solución. (Ver [11] [13])

Para aplicar este algoritmo al problema del MDS no métrico, se discretizó el espacio de dimensión reducida, dibujando una cuadrícula imaginaria de $h * h$, donde h se selecciona según la cantidad de elementos y la escala de la representación en \mathbb{R}^2 . En cada iteración, si el valor de s disminuye, $X_{(k+1)}$ será tomado como nueva solución inicial. Si no, nos movemos hacia otra cuadrícula siguiendo el siguiente criterio:

- Si en el movimiento anterior el punto se mantuvo en la misma cuadrícula, entonces nos movemos a la cuadrícula más cercana en la misma dirección del movimiento anterior.
- Si en el movimiento anterior el punto cambió de cuadrícula, entonces nos movemos en el sentido opuesto a la dirección del movimiento anterior.

2.3 Algoritmo Genético en MDS No Métrico

Un Algoritmo Genético es un método de búsqueda dirigida basada en probabilidad. Bajo una condición muy débil (que el algoritmo mantenga elitismo, es decir, guarde siempre al mejor elemento de la población sin hacerle ningún cambio), se puede demostrar que el algoritmo converge en probabilidad al óptimo. En otras palabras, al aumentar el número de iteraciones, la probabilidad de tener el óptimo en la población tiende a 1. Estos algoritmos hacen evolucionar una población de individuos sometiéndola a acciones aleatorias semejantes a las que actúan en la evolución biológica (mutaciones y recombinaciones genéticas), así como también a una "selección" de acuerdo con algún criterio, en función del que se decide cuáles son los individuos más adaptados, que sobreviven, y cuáles los menos aptos, que son descartados. (Ver [11] [9] [6] [12])

En la búsqueda de mejores soluciones al problema en cuestión, en cada iteración, a partir de la solución ($X_{(k)}$) se generó una población de manera aleatoria, empleando la distribución uniforme en el intervalo $[0,1]$. Cada individuo se evalúa en la función objetivo para conocer su aptitud. Se seleccionan las dos soluciones con mejor *Stress* y se realiza el cruzamiento para hallar $X_{(k+1)}$ que generará la nueva población.

2.4 Evolución Diferencial en MDS No Métrico

La Evolución Diferencial se caracteriza por el uso de vectores de prueba, los cuales compiten con los individuos de la población actual a fin de sobrevivir. El algoritmo asume que las variables del problema a optimizar están codificadas como un vector de números reales y que el dominio de las variables del problema está restringido por ciertas cotas definidas para cada variable. (Ver [11])

Para aplicar este algoritmo al problema del MDS no métrico, empleando la distribución uniforme en el intervalo $[0,1]$, se generaron aleatoriamente n vectores de Perturbación (Vector de Diferencias) que se recombinan con $X_{(k)}$ para generar una nueva población. La mejor solución (solución con menor *Stress*) se convertirá en la nueva solución inicial($X_{(k+1)}$), y pasará a conformar la nueva población.

3. Resultados

Para comprobar los resultados de los métodos implementados, fueron aplicados a ejemplos de la literatura y problemas reales realizando simulaciones en las que se variaron los parámetros de perturbación de la solución y el número de iteraciones.

El primer ejemplo, extraído de Borg & Groenen (2013), presenta las frecuencias por estado de los diferentes crímenes en Estados Unidos.

Al comparar la solución con *stress* = 0 dada por los autores con la obtenida por los métodos implementados se pudo observar que aunque los valores de *Stress* no son tan pequeños como podría esperarse, la representación bidimensional es suficientemente cercana a la realidad. Se lograron observar dos vecindades principales: crímenes donde se daña a las personas (asesinato, asalto, violación); y crímenes sobre la propiedad (estafa, robo, robo de auto). El atraco se encuentra en medio de estas vecindades, posiblemente porque en caso de ser violentos, no solo dañan las propiedades de las personas sino también sus cuerpos.

Mederos, Linares & Miret (2000) describen el estudio realizado con siete expertos del CENSA-ISCAH, que juzgaron 10 pares de ácaros que pueden formarse teniendo en cuenta diferentes variables como la morfología externa, la morfología interna, tipo de alimentación, etc.

Según los expertos, los ácaros 2, 4 y 5 se alimentan de células epidérmicas de las hojas y otros tejidos verdes de la planta, por lo que tienen comportamientos parecidos y deben estar cerca en la representación final. Sin embargo, el ácaro 1 se alimenta de restos de materias orgánicas y el 3 es depredador de insectos, por tales razones deben estar separados del resto, y no muy separados entre sí.

Al aplicar los métodos implementados a estos datos pudo verse que aunque a partir de las metaheurísticas no se llega a alcanzar un óptimo global, sí se logra establecer entre los elementos la relación expuesta previamente por los expertos. Además, se pudo notar que, aunque la evolución diferencial converge con mayor velocidad, se detiene en un óptimo local diferente del global.

Una de las aplicaciones más conocidas del MDS es la reconstrucción de mapas [10]. Dadas la matriz de distancias por carretera entre las principales ciudades en Cuba e Inglaterra, se aplicaron los algoritmos descritos anteriormente para obtener una representación en el plano euclídeo. Estas son

matrices de disimilitudes no euclídea, ya que las distancias por carretera no son en línea recta, y sus mediciones están sujetas a errores debido al relieve, entre otros factores, sin embargo, como estas disimilitudes se acercan bastante a las distancias euclídeas, el MDS Clásico nos brinda una muy buena solución. Al aplicar posteriormente MDS No Métrico con metaheurísticas no logramos que esta solución mejorase mucho más, aunque puede verse que la Evolución Diferencial es el algoritmo que mejor optimiza el *stress* para estos ejemplos.

Durante la experimentación, en todos los casos se logró reducir la pérdida de información por *Stress* respecto a la del MDS Clásico (por *Strain*) y se comprobó que se mantenían las relaciones de semejanzas originales. Se pudo apreciar que los algoritmos evolutivos convergen al mínimo más rápido que las heurísticas de búsqueda local. En la figura 1 se presenta un resumen de los resultados obtenidos.

Base de datos	MDS Clásico	Recocido Simulado	Búsqueda Tabú	Algoritmo Genético	Evolución Diferencial
Crimen USA [2]	0.2767	0.21	0.20	0.16	0.15
Ácaros [3]	0.2142	0.20	0.20	0.19	0.18
Reconstrucción del mapa de Cuba [3]	0.0152	0.0151	0.0151	0.0147	0.0116
Ciudades de Inglaterra [4]	0.0493	0.0491	0.0491	0.0465	0.0402

Figura 1. Resultados de la experimentación.

Conclusiones

Se elaboró un paquete de programas en lenguaje MATLAB que reúne algoritmos que combinan MDS No Métrico con Metaheurísticas. Se compararon las estrategias programadas a partir de sus resultados en ejemplos de la literatura y reales obteniendo resultados prometedores, particularmente, cuando se disponga de bases de datos para las cuales el MDS Clásico no sea eficiente por los problemas de mal condicionamiento de las matrices de distancias que procesa y la configuración inicial no sea lo suficientemente confiable de los datos que representa, en cuyo caso las metaheurísticas se encargarán de ofrecer al usuario la mejor configuración.

Agradecimientos

A mi tutora Dra. Elina Miret Barroso por el tiempo y esfuerzo dedicado para que este trabajo se llevara a cabo, muchas gracias por su apoyo.

Referencias

- [1] Ingwer Borg and Patrick J. F. Groenen. *Modern Multidimensional Scaling. Theory and Applications*. Springer., 2005.
- [2] Ingwer Borg, Patrick J. F. Groenen, and Patrick Mair. *Applied multidimensional scaling*. Springer., 2013.

- [3] William Castillo, Jorge González, and Oldemar Rodríguez. Métodos de optimización del stress. comparaciones usando disimilitudes tipo intervalo. *Revista de Matemática: Teoría y Aplicaciones. CIMPA.*, 2003.
- [4] Trevor F. Cox and Michael A. A. Cox. *Multidimensional Scaling*. Chapman & Hall., 1994.
- [5] Carles M. Cuadras. *Nuevos Métodos de Análisis Multivariante*. CMC Editions., 2008.
- [6] Stefan Etschberger and Andreas Hilbert. Multidimensional scaling and genetic algorithms : A solution approach to avoid local minima. *Arbeitspapiere zur mathematischen Wirtschaftsforschung.*, No. 181, 2003.
- [7] Teuvo Kohonen. New developments of nonlinear projections for the visualization of structures in nonvectorial data sets. *Aalto University publication series SCIENCE + TECHNOLOGY.*, 2011.
- [8] K. V. Mardia, J. T. Kent, and J. M. Bibby. *Multivariate Analysis.*, volume Tenth printing. 1995. Academic Press Inc., 1979.
- [9] Rudolf Mathar and Antanas Zilinskas. On global optimization in two-dimensional scaling. *Kluwer Academic Publishers*, 1993.
- [10] Elina Miret. Un enfoque unificado para técnicas de representación euclidiana. *Tesis de Doctorado.*, 2005.
- [11] El-Ghazali Talbi. *Metaheuristics From Design to Implementation*. John Wiley & Sons., 2009.
- [12] P. Tecuanhuehue-Vera, Jesús Ariel Carrasco-Ochoa, and José Fco. Martínez-Trinidad. Genetic algorithm for multidimensional scaling over mixed and incomplete data. *National Institute for Astrophysics, Optics and Electronics.*, 2012.
- [13] Mario Villalobos and Javier Trejos. Análisis de proximidades métrico usando búsqueda tabú. *Revista de Matemática: Teoría y Aplicaciones. CIMPA*, 2000.
- [14] Mario Villalobos and Javier Trejos. Application of simulated annealing in metric multidimensional scaling. *Revista Investigación Operacional.*, Vol. 22, No. 3., 2001.