

MÉTODOS DE OPTIMIZACIÓN DEL STRESS.
COMPARACIONES USANDO DISIMILITUDES TIPO
INTERVALO

WILLIAM CASTILLO* JORGE GONZÁLEZ† OLDEMAR RODRÍGUEZ‡

Recibido (versión revisada): 23 May 2003

Resumen

Usando sobrecalentamiento simulado se construye el algoritmo MDSI-SS para optimizar el Stress definido por Denoeux y Masson [6] para disimilitudes tipo intervalo. En tres juegos de datos se compara el valor del Stress obtenido con MDSI-SS e INTERSCAL, este último método propuesto por Rodríguez et al. [12].

Palabras clave: Stress, Interscal, sobrecalentamiento simulado.

Abstract

By the use of simulated annealing it is defined the algorithm MDSI-SS for the optimization of Stress defined by Donoeux and Masson [6] for interval-type dissimilarities. On three data sets it is compared the value of stress obtained by MDSI-SS and INTERSCAL (the later is the algorithm proposed by Rodríguez et al. [12]).

Keywords: Stress, Interscal, simulated annealing.

Mathematics Subject Classification: 62-07, 91C15.

*CIMPA, Escuela de Matemática, Universidad de Costa Rica. 2060 San José Costa Rica. E-mail: wcastill@cariari.ucr.ac.cr

†CIMPA, Escuela de Matemática, Universidad de Costa Rica. 2060 San José Costa Rica. E-mail: jgonzale@cariari.ucr.ac.cr

‡CIMPA, Escuela de Matemática, Universidad de Costa Rica. 2060 San José Costa Rica. E-mail: orodrigu@una.ac.cr

1 MDS clásico

1.1 MDS de Torgerson y Gower

El Escalamiento Multidimensional, abreviado MDS por sus siglas obtenidas del inglés *MultiDimensional Scaling* fue propuesto por Torgerson [13] y Gower [9]. El problema fue planteado en términos de que dada una matriz de distancias D entre n objetos

$$D = (d_{ij}) \text{ de tamaño } n \times n$$

se trata de encontrar un conjunto $\{\mathbf{x}_1, \dots, \mathbf{x}_n\}$ de puntos del espacio euclídeo \mathbb{R}^p , de modo que las distancias $\|\mathbf{x}_i - \mathbf{x}_j\|$ entre estos puntos reproduzcan lo mejor posible las distancias d_{ij} , entre los correspondientes objetos.

La solución encontrada por Torgerson y Gower consiste en definir a partir de D una matriz de productos internos $B = XX^t$, donde X es la matriz cuyas filas son $\{\mathbf{x}_1, \dots, \mathbf{x}_n\}$, para luego obtener la Descomposición Espectral de esta última matriz. Los valores y vectores propios de B permiten construir la configuración de puntos deseada. Una descripción breve del procedimiento es la siguiente.

- Se define B centrando en filas y columnas la matriz A

$$A = \left(-\frac{1}{2}d_{ij}^2\right), \quad J = I_n - \frac{1}{n}\mathbf{1}\mathbf{1}^t, \quad B = JAJ.$$

- Sean u_1, \dots, u_p los vectores propios ortonormados de B y $\lambda_1, \dots, \lambda_p$ los valores propios correspondientes. Sea además la matriz $U = (u_1 \dots u_p)_{n \times p}$ y Δ_λ la matriz diagonal de los valores propios. Se obtiene la Descomposición Espectral de B

$$B = \sum_{i=1}^p \lambda_i u_i u_i^t = U \Delta_\lambda U^t = QQ^t$$

con $Q = U \Delta_{\sqrt{\lambda}}$ y p es el rango de B .

- La configuración deseada es la definida por las filas de Q .

Una solución aproximada óptima, de rango $r < p$, en el sentido del producto de Hilbert-Schmidt, se obtiene si se eligen solamente los primeros r valores y vectores propios para definir Q .

La solución de Torgerson y Gower fue posteriormente generalizada sustituyendo la matriz de distancias D por una matriz de disimilitudes $\Delta = (\delta_{ij})$.

En el caso muy frecuente, de que la distancia o la disimilitud no sean euclídeas la matriz B no es semidefinida positiva, hay valores propios negativos y por lo tanto la solución de Torgerson y Gower no es aplicable o bien no es óptima. Una alternativa es minimizar una función denominada Stress.

1.2 MDS minimizando la función de Stress

Dada una matriz de disimilitudes $\Delta = (\delta_{ij})$, el problema se plantea en términos de encontrar una configuración $\{\mathbf{x}_1, \dots, \mathbf{x}_n\}$ de puntos de \mathbb{R}^p que definimos como las filas de una matriz $X = (x_{ij})_{n \times p}$ que minimize la función de Stress:

$$\sigma(X) = \sum_{i < j} (\delta_{ij} - d_{ij}(X))^2$$

donde $d_{ij}(X)$ es la matriz de distancia entre las filas de X .

Diferentes métodos se han utilizado para obtener el mínimo de esta función, que se fundamentan en técnicas tales como: Descenso del gradiente, Smacof, sobrecalentamiento simulado, etc. Ver por ejemplo [3], [5] y [10].

2 MDS con matriz de disimilitud tipo intervalo

En la actualidad es frecuente encontrarse con grandes bases de datos que es necesario reducirlas previamente para poder utilizar alguna herramienta de análisis. Una manera de enfrentar esta dificultad ha sido propuesta por Diday [7] la cual consiste utilizar el Análisis de Datos Simbólicos. Las bases de datos son reducidas en su tamaño al redefinirlas sobre una unidad más compleja que él llamó Objeto Simbólico. A partir de los trabajos de Diday se han redefinido algunas de las técnicas usuales de análisis de datos para ser aplicadas a datos simbólicos, tal es el caso del Escalamiento Multidimensional (MDS) para el que se han propuesto dos procedimientos que corresponden a los dos planteados en las secciones 1.1 y 1.2, pero bajo el supuesto que las entradas de la matriz de disimilitud son intervalos que expresan la variación de la disimilitud entre objetos simbólicos.

2.1 Los datos

En el resto de este artículo supondremos que los datos están dados en la forma de intervalos de variación de disimilitudes entre objetos simbólicos, organizados en una matriz simétrica $\Delta = (\delta_{ij})_{n \times n}$ donde $\delta_{ij} = [\underline{\delta}_{ij}, \bar{\delta}_{ij}]$ representa el intervalo de variación de la disimilitud entre dos objetos simbólicos R_i y R_j .

Datos de esta naturaleza se generan por ejemplo, en experimentos donde los objetos se presentan por pares (R_i, R_j) , a un grupo de jueces quienes, en una escala determinada, emiten una calificación de la “disimilitud” entre dichos objetos. Esta calificación varía de un juez a otro dando lugar a un rango de variación del tipo $\delta_{ij} = [\underline{\delta}_{ij}, \bar{\delta}_{ij}]$.

2.2 Objetivo del MDS simbólico. Caso intervalo

Dada una matriz de disimilitud $\Delta = (\delta_{ij})$ cuyas entradas son intervalos que expresan el rango de variación de la disimilitud entre n objetos simbólicos. Sea $X = ([\underline{x}_{ij}, \bar{x}_{ij}])_{n \times p} = (R_{ij})_{n \times p}$ una matriz cuyas n filas corresponden a la representación de cada uno de estos objetos mediante un hipercubo, cuyos intervalos expresan el rango de variación de las r variables que caracterizan los objetos.

El problema se plantea en términos de encontrar X de modo que las distancias euclídeas máximas y mínimas $[\underline{d}_{ij}, \overline{d}_{ij}]$ entre los hipercubos R_i, R_j (filas de X) reproduzcan lo mejor posible las correspondientes $\delta_{ij} = [\underline{\delta}_{ij}, \overline{\delta}_{ij}]$. Las distancias euclídeas $\overline{d}_{ij}, \underline{d}_{ij}$ se definen como

$$\overline{d}_{ij} = \max \{d(x, y) \mid x \in R_i, y \in R_j\} \text{ y } \underline{d}_{ij} = \min \{d(x, y) \mid x \in R_i, y \in R_j\}.$$

Denoeux y Masson probaron en [6] que estas distancias se expresan así:

$$\overline{d}_{ij} = \frac{1}{2} \sqrt{\sum_{k=1}^p [u_{ik} + u_{jk} + |v_{ik} - v_{jk}|]^2} \quad (1)$$

$$\underline{d}_{ij} = \frac{1}{4} \sqrt{\sum_{k=1}^p [u_{ik} + u_{jk} - |v_{ik} - v_{jk}| - |u_{ik} + u_{jk} - |v_{ik} - v_{jk}||]^2} \quad (2)$$

donde $u_{ik} = \overline{x}_{ik} - \underline{x}_{ik}$ y $v_{ik} = \overline{x}_{ik} + \underline{x}_{ik}$.

Dos soluciones se han propuesto al problema anterior, que se corresponden a los casos clásicos descritos anteriormente, adaptados a disimilitudes tipo intervalos. A continuación damos una breve descripción de estos métodos.

2.3 Método INTERSCAL

Rodríguez, O.; Diday, E. y Winsberg, S. [12] propusieron un método que llamaron INTERSCAL (INTERval SCALing) el cual es una generalización del método clásico de Torgerson y Gower. Ellos transforman la matriz de disimilitud tipo intervalo de $n \times n$ en una matriz de disimilitud clásica $\tilde{\Delta}$ de tamaño $2n \times 2n$ definida como $\tilde{\Delta}$.

$$\tilde{\Delta} = \begin{bmatrix} 0 & \overline{\delta}_{11} & \frac{\delta_{12}}{2} & \frac{\overline{\delta}_{12} + \delta_{12}}{2} & \frac{\delta_{13}}{2} & \frac{\overline{\delta}_{13} + \delta_{13}}{2} & \cdots & \frac{\delta_{1n}}{2} & \frac{\overline{\delta}_{1n} + \delta_{1n}}{2} \\ \overline{\delta}_{11} & 0 & \frac{\overline{\delta}_{12} + \delta_{12}}{2} & \overline{\delta}_{12} & \frac{\overline{\delta}_{13} + \delta_{13}}{2} & \overline{\delta}_{13} & \cdots & \frac{\overline{\delta}_{1n} + \delta_{1n}}{2} & \overline{\delta}_{1n} \\ \frac{\delta_{21}}{2} & \frac{\overline{\delta}_{21} + \delta_{21}}{2} & 0 & \overline{\delta}_{22} & \frac{\delta_{23}}{2} & \frac{\overline{\delta}_{23} + \delta_{23}}{2} & \cdots & \frac{\delta_{2n}}{2} & \frac{\overline{\delta}_{2n} + \delta_{2n}}{2} \\ \frac{\overline{\delta}_{21} + \delta_{21}}{2} & \overline{\delta}_{21} & \overline{\delta}_{22} & 0 & \frac{\overline{\delta}_{23} + \delta_{23}}{2} & \overline{\delta}_{23} & \cdots & \frac{\overline{\delta}_{2n} + \delta_{2n}}{2} & \overline{\delta}_{2n} \\ \frac{\delta_{31}}{2} & \frac{\overline{\delta}_{31} + \delta_{31}}{2} & \frac{\delta_{32}}{2} & \frac{\overline{\delta}_{32} + \delta_{32}}{2} & 0 & \overline{\delta}_{33} & \cdots & \frac{\delta_{3n}}{2} & \frac{\overline{\delta}_{3n} + \delta_{3n}}{2} \\ \frac{\overline{\delta}_{31} + \delta_{31}}{2} & \overline{\delta}_{31} & \frac{\overline{\delta}_{32} + \delta_{32}}{2} & \overline{\delta}_{32} & \overline{\delta}_{33} & 0 & \cdots & \frac{\overline{\delta}_{3n} + \delta_{3n}}{2} & \overline{\delta}_{3n} \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ \frac{\delta_{n1}}{2} & \frac{\overline{\delta}_{n1} + \delta_{n1}}{2} & \frac{\delta_{n2}}{2} & \frac{\overline{\delta}_{n2} + \delta_{n2}}{2} & \frac{\delta_{n3}}{2} & \frac{\overline{\delta}_{n3} + \delta_{n3}}{2} & \cdots & 0 & \overline{\delta}_{nn} \\ \frac{\overline{\delta}_{n1} + \delta_{n1}}{2} & \overline{\delta}_{n1} & \frac{\overline{\delta}_{n2} + \delta_{n2}}{2} & \overline{\delta}_{n2} & \frac{\overline{\delta}_{n3} + \delta_{n3}}{2} & \overline{\delta}_{n3} & \cdots & \overline{\delta}_{nn} & 0 \end{bmatrix}. \quad (3)$$

Siguiendo un procedimiento análogo al de Torgerson definen a partir de $\tilde{\Delta}$ una matriz de productos internos B de tamaño $2n \times 2n$. Sea $\{u_1, \dots, u_p\} \subseteq \mathbb{R}^{2n}$ un conjunto de vectores propios ortonormales de B asociados a los p valores propios $\lambda_1 \geq \lambda_2 \geq \dots \lambda_p > 0$. Utilizan la matriz $U \Delta_{\sqrt{\lambda}} = (\sqrt{\lambda_1} u_1, \dots, \sqrt{\lambda_p} u_p)$ de tamaño $2n \times p$, para definir una matriz de hipercubos X de tamaño $n \times p$. Los extremos de los r intervalos que forman el hipercubo

i -ésimo se obtienen tomando los correspondientes máximos y mínimos entre las filas $2i - 1$ y $2i$ de $U\Delta_{\sqrt{\lambda}}$. Es decir $\underline{x}_{ij} = \min\{\tilde{x}_{2i-1,j}, \tilde{x}_{2i,j}\}$ y $\overline{x}_{ij} = \max\{\tilde{x}_{2i-1,j}, \tilde{x}_{2i,j}\}$ donde $\tilde{x}_{2i,j}$ es la entrada $(2i, j)$ de la matriz $Q = U\Delta_{\sqrt{\lambda}}$. Las filas de Q así definidas conforman la configuración buscada.

2.4 Método minimizando el Stress

Una generalización del MDS para el caso de datos tipo intervalos fue dada por Denoeux y Masson ([6]). Para ello redefinieron la función de Stress como

$$\sigma(X) = \sum_{i < j} [(\overline{\delta}_{ij} - \overline{d}_{ij})^2 + (\underline{\delta}_{ij} - \underline{d}_{ij})^2]$$

El problema es planteado en términos de obtener una representación de tipo intervalo $X = ([\underline{x}_{ij}, \overline{x}_{ij}])_{n \times p}$ para los objetos simbólicos, de manera que se minimice el Stress. Denoeux y Masson utilizaron el método del descenso del gradiente para minimizar el Stress.

3 Aplicación del sobrecalentamiento simulado en la optimización del Stress

La optimización del Stress utilizando el descenso del gradiente no garantiza la obtención de un mínimo global, por otra parte el método INTERSCAL no se aplica directamente sobre la matriz de disimilitud tipo intervalo, sino sobre otra matriz que no es tipo intervalo y que además puede no ser semidefinida positiva, afectando la optimalidad de la configuración encontrada por el método de Torgerson.

La solución que se propone en este trabajo consiste en aplicar la técnica de Optimización Estocástica conocida como Sobrecalentamiento Simulado (Simulated Annealing) que abreviamos como *SS*, introducida por Kirkpatrick, Gelatt y Vecchi 1982-1983) [1] la cual utiliza un algoritmo conocido como la regla de Metropolis que garantiza una convergencia asintótica a un óptimo global, con probabilidad 1.

Aplicamos *SS* para minimizar la función de Stress $\sigma(X)$ con respecto a X directamente sobre la matriz de disimilitud tipo intervalo.

3.1 Descripción del algoritmo

Para poder utilizar el método de Sobrecalentamiento Simulado, es necesario discretizar el espacio R^p , para ello se define un mallado de paso h .

El objetivo es minimizar la función de Stress $\sigma(X)$ entre todas las configuraciones X de tamaño $n \times p$ cuyas filas son puntos del mallado. El procedimiento comienza eligiendo una configuración inicial $X_0 = ([\underline{x}_{ij}, \overline{x}_{ij}])_{n \times p}$ sujeta a las restricciones $\underline{x}_{ij} < \overline{x}_{ij} \quad \forall i, j$, de tal manera que para h, p y X_0 dados queda determinado un conjunto de estados posibles $H(X_0)$, por medio del sistema de generación de estados que se describe en 3.1.1.

El procedimiento para elegir de $H(X_0)$ la configuración óptima no solo requiere definir un sistema de generación de estados, si no además determinar la temperatura inicial c_0 y el largo l_c de la cadena de Markov.

3.1.1 Sistema de generación de estados

Sea X una configuración inicial de tamaño $n \times p$, $D = \{\pm e_i | i = 1, 2\}$ el conjunto de direcciones canónicas de R^2 y $h > 0$ el tamaño del paso.

Un estado vecino de X es la matriz que se obtiene de X , perturbando una fila cualquiera de ella de acuerdo a lo siguiente:

1. Se selecciona al azar $(i, j) \in \{1, \dots, n\} \times \{1, \dots, p\}$ con probabilidad $\frac{1}{np}$ y $d \in D$ con probabilidad $\frac{1}{4}$ perturbar la fila x_i sustituyendo el j -ésimo intervalo por $(\underline{x}_{ij}, \overline{x}_{ij}) + hd$
2. Si \tilde{x}_i denota la fila perturbada, el nuevo estado vecino de X es la matriz \tilde{X} obtenida al sustituir la fila x_i de X por \tilde{x}_i . Denotamos como $\mathcal{V}(X)$ al conjunto de todos los vecinos de X . La probabilidad de generar un vecino de X es $\frac{1}{4np}$.

3.2 El algoritmo MDSI-SS

La implementación computacional del método se basa en el siguiente esquema algorítmico que denominamos MDSI-SS, donde ϵ es un umbral elegido por el usuario, γ es y $nmax$ es el número máximo de iteraciones.

1. Se escoge al azar una configuración X tipo intervalo. Elegir un valor de h (por ejemplo 0.1) y calcular la temperatura inicial c_0 y definir $t = 0$.
2. Repetir l_c veces 2.1 y 2.2.
 - (a) Generar un vecino \tilde{X} de X de acuerdo con el sistema de generación definido anteriormente.
 - (b) Calcular la diferencia del Stress $\Delta\sigma = \sigma(X) - \sigma(\tilde{X})$ y aceptar (o no aceptar) \tilde{X} de acuerdo con la regla de Metropolis, es decir: si $\Delta\sigma > 0$ entonces se define $X := \tilde{X}$. Si $\Delta\sigma \leq 0$ entonces se escoge $\alpha \in [0, 1]$ al azar con probabilidad uniforme y si $e^{-\frac{\Delta\sigma}{c_t}} > \alpha$ entonces se define $X := \tilde{X}$.
3. Si $c_t < \epsilon$ o $t = nmax$, terminar. Si no definir $t := t + 1$, $c_t = \gamma c_{t-1}$ y regresar al paso 2.

Observación Para la rapidez del algoritmo es importante encontrar una expresión sencilla de la diferencia del Stress, en nuestro caso, asumiendo que se sustituye la k -ésima fila x_k por \tilde{x}_k se tiene que

$$\begin{aligned} \Delta\sigma_k = & \sum_j [(\bar{\delta}_{kj} - \bar{d}(x_k, x_j))^2 + (\underline{\delta}_{kj} - \underline{d}(x_k, x_j))^2 \\ & - (\bar{\delta}_{kj} - \bar{d}(\tilde{x}_k, x_j))^2 - (\underline{\delta}_{kj} - \underline{d}(\tilde{x}_k, x_j))^2] \\ & - (\bar{\delta}_{kk} - \bar{d}(\tilde{x}_k, x_k))^2 - (\underline{\delta}_{kk} - \underline{d}(\tilde{x}_k, x_k))^2 \\ & + (\bar{\delta}_{kk} - \bar{d}(\tilde{x}_k, \tilde{x}_k))^2 + (\underline{\delta}_{kj} - \underline{d}(\tilde{x}_k, \tilde{x}_k))^2 \end{aligned}$$

3.3 Acerca de la convergencia

Para garantizar la convergencia asintótica con probabilidad uno al óptimo global del Stress, es necesario verificar ciertas propiedades de generación de nuevos estados (ver [1]).

Simetría: Sea G_{ij} la probabilidad de generar el estado j a partir del estado i . Entonces $G_{ij} = G_{ji}$. En efecto, todas las rutas o cadenas en $H(X_0)$ para ir de i a j son sucesiones de estados vecinos de probabilidad de generación constante. Así la probabilidad de generar una cadena de i a j es igual a la probabilidad de generar la cadena inversa de j a i . Como G_{ij} es la suma de las probabilidades de las distintas cadenas de i a j , es claro que $G_{ij} = G_{ji}$.

Conexidad: para todo par de estados i, j de $H(X_0)$, existe $\{l_0, l_1, \dots, l_p\} \subset H(X_0)$, tal que $i = l_0$, $j = l_m$ y $G_{l_s l_r} > 0$ para todo $l_s, l_r \in \{l_0, l_1, \dots, l_m\}$. Esta propiedad es evidente puesto que cualquier cadena del estado i al estado j sirve como $\{l_0, l_1, \dots, l_m\}$.

4 Comparación entre MDSI-SS e INTERSCAL

Se utilizaron las matrices de disimilitudes correspondientes a tres juegos de datos:

1. Matriz de disimilitudes δ entre ocho tipos de aceite. La matriz δ es construida a partir de una matriz de tamaño 8×4 usando las distancias definidas en (1), (2) ([4]).
2. Matriz de disimilitudes δ entre 20 rectángulos. La matriz δ es construida directamente de un experimento: Esto es sin tener previamente una matriz de individuos por variables ([14]).
3. Matriz de disimilitudes δ entre 10 pacientes. Construida a partir de la matriz de tamaño 10×3 usando las distancias (1), (2) ([11]).

Se calcularon las configuraciones por el método INTERSCAL usando el paquete *PI-MAD SIMBOLICO* y con la técnica de Sobrecalentamiento Simulado (MDSI-SS) programado con *Mathematica 4.0*, tomando $p = 2$ y $D = \{(1, 0), (-1, 0), (0, 1), (0, -1)\}$. A continuación se reportan las configuraciones en \mathbb{R}^2 y el Stress correspondiente en ambos métodos y además el porcentaje en que MDSI-SS disminuyó el stress respecto de INTERSCAL. Las configuraciones obtenidas se presentan en las figuras 1, 2 y 3.

Datos	Stress con INTERSCAL	Stress con MDSI-SS	% Disminución
Aceites	1497	860	43
Lyn	83.7	10.1	88
Rectángulos	361085	94181	74

Tabla 1: Comparación entre los algoritmos MDSI-SS e INTERSCAL

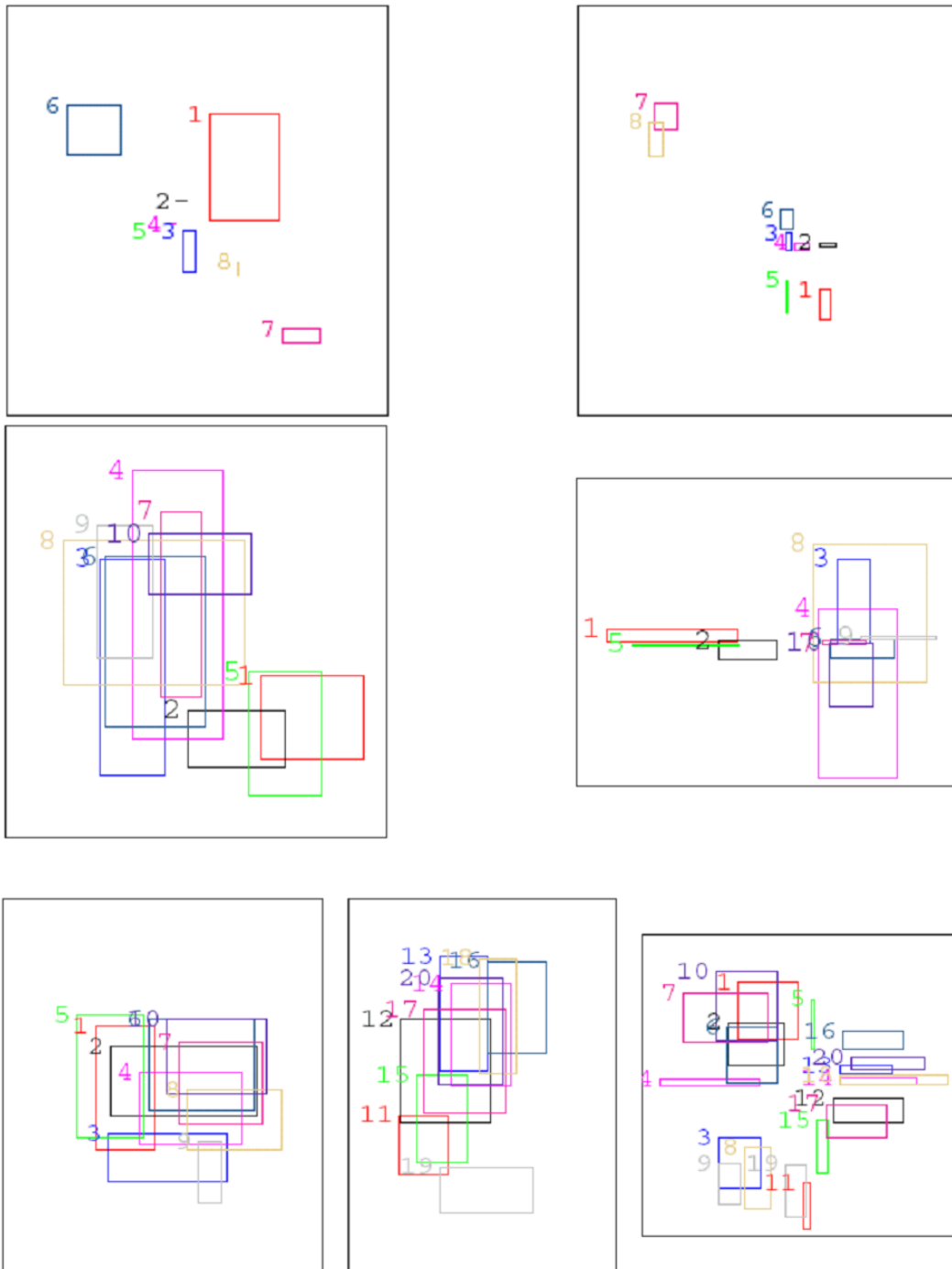




Figura 3: Rectángulos MDSI-SS, INTERSCAL

5 Conclusiones

De la Tabla 1 vemos que el Stress obtenido por MDSI-SS es en todos los casos, significativamente menor que el encontrado por INTERSCAL, sin embargo en la configuración encontrada por este último los rectángulos parecen más separados, lo que permite una mejor lectura de los posibles agrupamientos. Pensamos que eso se debe a que las direcciones calculadas por INTERSCAL son los vectores propios asociados a los primeros valores propios de $\tilde{\Delta}$ y como se sabe, estas direcciones son las de máxima dispersión.

Por otra parte las distancias definidas en (1) y (2) no reflejan adecuadamente las diferencias en ciertas posiciones. Por ejemplo las distancias-intervalo entre los dos pares siguientes de rectángulos  y  son iguales, sin embargo claramente la situación no es la misma.

Creemos que es necesario investigar un poco más acerca de las disimilitudes con valor intervalo en el contexto del escalamiento multidimensional simbólico.

Referencias

- [1] Aarts, E.; Korst, J. (1989) *Simulated Annealing and Boltzmann Machines*. John Wiley & Sons, Chichester.
- [2] Bock, H-H.; Diday, E. (eds.) (2000) *Analysis of Symbolic Data. Exploratory Methods for Extracting Statistical Information from Complex Data*. Springer Verlag, Heidelberg.
- [3] Borg, I.; Groenen, P. (1997) *Modern Multidimensional Scaling*. Springer, New York.
- [4] Cazes, P.; Chouakria, A.; Diday, E.; Schektman, Y. (1997) “Extension de l’analyse en composantes principales à des données de type intervalle”, *Rev. Statistique Appliquée* **45**(3): 5–24.
- [5] Cox, T.; Cox, M. (1994) *Multidimensional Scaling*. Chapman & Hall, London.
- [6] Denoeux, T.; Masson, M. (1999) “Multidimensional scaling of interval-valued dissimilarity data”. Université de Technologie de Compiègne, France.
- [7] Diday, E. (1998) “Symbolic data analysis: a mathematical framework and tool for data mining”, in: A. Rizzi, M. Vichi & H. Bock (Eds.): *Advances in Data Science and Classification*, Springer-Verlag, Heidelberg: 409–416.
- [8] Diday, E. (1999) *An Introduction to Symbolic Data Analysis and its Application to the Sodas Project: Purpose, History and Perspective*. Paris IX–Dauphine University, France.
- [9] Gower, J.C. (1966) “Some distances properties of latent root and vector methods using multivariate analysis”, *Biometrika*, **53**: 325–338.

- [10] Trejos J. et al. (2000) “Application of simulated annealing in some multidimensional scaling problems”, in: Kiers et al. (Eds.): *Data Analysis, Classification and Related Methods*. Springer, Heidelberg: 297–302.
- [11] Raju, S.R.K. (1997) “Symbolic data analysis in cardiology”, in: E. Diday & K.C. Gowda (Eds.): *Symbolic Data Analysis and its Applications*, CEREMADE, Paris: 245–249.
- [12] Rodríguez, O.; Diday, E.; Winsberg, S. (2000) “Multidimensional scaling for interval data: INTERSCAL”, submitted paper.
- [13] Torgenson, W.S. (1958) *Theory and Methods of Scaling*. Wiley, New York.
- [14] Winsberg, S.; De Soete, G. (1997) “Multidimensional scaling with constrained dimensions: CONSCAL”, *British Journal of Mathematical and Statistical Psychology*, **50**: 55–72.