

Multidimensional Scaling (MDS)

Álex Murillo*, Javier Trejos, Eduardo Piza, Mario Villalobos y
Alejandra Jimenez**

CIMPA - Universidad de Costa Rica

*Sede del Atlántico

**Instituto Tecnológico de Costa Rica

26 de abril de 2010

1 Introducción

2 Datos en MDS

- Disimilaridades cuantitativas
- Disimilaridades cualitativas
- Clasificación general de datos analizados en MDS

3 Modelos en MDS

- El método CLÁSICO
- El método ALSCAL
- El método PROXSCAL

4 Bibliografía

Introducción

Uno de los problemas más interesantes en muchas disciplinas se plantea cuando se necesita medir y entender las relaciones entre objetos siendo desconocidas las dimensiones subyacentes de los mismos y especialmente en aquellas situaciones en las que la información disponible se refiere exclusivamente a la semejanza o diferencia entre los objetos que son motivo de estudio.

Introducción

El análisis multidimensional de estructuras mediante proximidades o *Multidimensional Scaling* (MDS) puede definirse como un conjunto de técnicas para la representación de coeficientes de proximidad entre objetos mediante distancias en un espacio de dimensión mínima y el posterior análisis de la configuración obtenida.

Ejemplo

- Suponga, que se dispone de una matriz cuadrada de tamaño $n \times n$, formada por las distancias entre cada pareja de ciudades y se pretende construir el mapa partiendo de dicha matriz.
- Este problema, en dimensión arbitraria, tiene solución exacta en base a los trabajos de Schoenberg (1935) y Young y Houselholder (1938), que se recogen en el siguiente resultado y cuya demostración puede verse en el trabajo de Mardia, Kent, y Bibby (1980).

Teorema

Theorem

Sea $D_{n \times n}$ matriz de distancias entre n puntos en un espacio de dimensión ν y sea $B = HAH$, siendo $H_{n \times n}$ dada por $H = I - n^{-1}\mathbf{1}\mathbf{1}^t$ y $A_{n \times n}$ la matriz cuyos elementos vienen dados a través de $a_{rs} = -\frac{1}{2}d_{rs}^2$. Entonces, D es una matriz de distancias euclídeas si y solo si B es semidefinida positiva.

Además se tiene:

1. Si D es la matriz de distancias euclídeas para $Z_{n \times \nu} = (z_1, \dots, z_n)^t$, entonces $B = (HZ)(HZ)^t$, donde $B \succeq 0$. B será la matriz centrada de productos escalares de Z .

Teorema (continuación)

2. Inversamente, si B es semidefinida positiva de rango ν , entonces puede construirse una configuración asociada a B de la siguiente forma: Sean $\lambda_1 > \dots > \lambda_\nu$ valores propios positivos de B correspondientes a los vectores propios $X_{n \times \nu} = (x_1, \dots, x_\nu)$, normalizados según la condición

$$\forall i = 1, \dots, \nu, \quad x_i' x_i = \lambda_i.$$

Los puntos $x_r = (x_{r1}, \dots, x_{r\nu})^t \in \mathbb{R}^\nu$, donde x_r es la r -ésima fila de la matriz X , tienen matriz de distancias D . Además, esa configuración está centrada en $\bar{x} = 0$ y B es la matriz de productos escalares de esa configuración.

Problema que resuelve MDS

- Existe solución única para distancias euclídeas en un espacio de dimensión $\nu = \text{rang}(B) \leq (n - 1)$, salvo isometrías.
- El problema que resuelve MDS surge cuando se pretende una representación en un espacio de dimensión menor que ν o la información de la que se dispone entre cada par de elementos a representar no viene dada en términos de distancia euclídea sino en términos de pseudo-distancia o en general de *proximidad* respecto a algún criterio mediante coeficientes de *disimilaridad* o *similaridad*.

Ejemplo

En el ejemplo del mapa que se planteo, dicha situación podría corresponder al caso en que se desea reconstruir el mapa (dimensión dos) utilizando las distancias medidas por carretera o coeficientes de proximidad entre las ciudades con base en algún criterio de tipo económico, social, entre otros.

Datos en MDS

- La terminología empleada en MDS ha sido elaborada fundamentalmente en el ámbito de las **ciencias de la conducta** y frecuentemente puede no resultar usual en estadística.
- El término *proximidad* indica el concepto de cercanía en espacio, tiempo o cualquier otro contexto. Desde un punto de vista matemático, ese término hace referencia al concepto de disimilaridad o similaridad entre dos elementos.

Datos en MDS

Sea Ω un conjunto finito o infinito de elementos sobre los que se desea definir una proximidad.

Definition

Dados dos puntos $o_i, o_j \in \Omega$ y sea $\delta : \Omega \times \Omega \longrightarrow \mathbb{R}$, se dirá que δ es una *disimilaridad* si verifica:

- 1 $\forall i, j, \delta_{ij} = \delta_{ji},$
- 2 $\forall i, j, \delta_{ii} \leq \delta_{ij},$
- 3 $\forall i, \delta_{ii} = \delta_0.$

- La primera condición podría eliminarse.
- Suelen establecerse que $\delta_0 = 0$.

Disimilaridades cuantitativas

Existen diferentes medidas para el cálculo de disimilaridades entre un par de variables u objetos.

Si se considera una matriz de datos $(x_{ij})_{n \times \nu}$, obtenida de n objetos sobre ν variables, y se denota por x_i una fila, algunos de los coeficientes de disimilaridad cuantitativos más utilizados son:

- **Distancia euclídea** ponderada, que viene determinada por:

$$d_{M,ij} = \left((x_i - x_j)^t M (x_i - x_j) \right)^{1/2}, \quad (1)$$

donde la matriz M es simétrica y semidefinida positiva.

Disimilaridades cuantitativas

- **Distancia euclídea clásica**, cuando $M = I$ en (1):

$$d_{ij} = \left[\sum_{k=1}^{\nu} (x_{ik} - x_{jk})^2 \right]^{1/2}. \quad (2)$$

Apropiada si las variables tienen escalas equivalentes, varianzas iguales y son independientes dos a dos.

- **Distancia de Mahalanobis**, usada análisis de datos:

$$d_{V,ij} = \left((x_i - x_j)^t V^{-1} (x_i - x_j) \right)^{1/2},$$

donde V es la matriz de varianzas y covarianzas.

Disimilaridades cuantitativas

- **Distancia de Minkowski**, depende del parámetro $p \geq 1$:

$$d_{p,ij} = \left[\sum_{k=1}^{\nu} |x_{ik} - x_{jk}|^p \right]^{1/p}. \quad (3)$$

Para $p = 2$ es la distancia euclídea clásica dada en (2).

- **Distancia city-block**, con $p = 1$ en Minkowski (3):

$$d_{1,ij} = \sum_{k=1}^{\nu} |x_{ik} - x_{jk}|. \quad (4)$$

Disimilaridades cuantitativas

- **Distancia de Chebychev**, $p \rightarrow \infty$ en Minkowski (3):

$$d_{\max,ij} = \max_{k=1,\dots,\nu} |x_{ik} - y_{jk}|.$$

- **Distancia chi-cuadrado** (χ^2), en tablas de contingencia y frecuencia, en el análisis factorial de correspondencias:

$$d_{\chi^2,ij} = \left[\sum_{k=1}^{\nu} \frac{1}{x_{\cdot k}} \left(\frac{x_{ik}}{x_{i\cdot}} - \frac{x_{jk}}{x_{j\cdot}} \right)^2 \right]^{1/2},$$

$$\text{donde } x_{\cdot k} = \sum_{i=1}^n x_{ik}, \quad x_{i\cdot} = \sum_{k=1}^{\nu} x_{ik}.$$

Disimilaridades cuantitativas

- Separación angular:

$$d_{a,ij} = 1 - \frac{\sum_{k=1}^{\nu} x_{ik}x_{jk}}{\left[\sum_{k=1}^{\nu} x_{ik}^2 \sum_{k=1}^{\nu} x_{jk}^2 \right]^{1/2}}.$$

Disimilaridades cualitativas

- Los coeficientes cualitativos más utilizados están relacionados con datos de tipo psicológico, procedentes del campo de aplicación en el que surgió MDS.
- Los más importantes es la obtención de *juicios directos de disimilaridad* entre objetos, primero se definen los individuos y los estímulos que serán tratados. Luego se genera una muestra aleatoria o estratificada de los estímulos, eliminando aquellos que no sean familiares a los individuos encuestados.

Disimilaridades cualitativas

- El procedimiento de *clasificación categórica* más utilizado para obtener juicios directos de disimilaridad consiste en presentarle al sujeto un par de estímulos y se le pide que lo clasifique en una escala como:

Muy similar, , Muy disimilar.

Clasificación general de datos analizados en MDS

Se clasifican en relación a la escala de medida de la variable.

- 1 **Escala nominal:** son categóricos y solamente se distinguirá entre clases diferentes, por ejemplo: color de ojos, color de pelo.
- 2 **Escala ordinal:** pueden ser ordenados pero no son datos cuantitativos, por ejemplo: la leche de la botella tres es de mejor calidad que la de la botella cuatro.
- 3 **Escala de tipo intervalo:** son cuantitativos con diferencias significativas entre valores, no existe un cero significativo, por ejemplo: la temperatura en grados.
- 4 **Escala de razón:** son similares a los de tipo intervalo salvo que posee un cero, por ejemplo: pesos, alturas.

Ejemplo de catadores de vinos

El conjunto de objetos podría ser diez botellas de vino tinto correspondientes a distintas bodegas. La disimilaridad δ_{ij} entre cada par de botellas (i, j) debería ser una puntuación entera entre cero y diez significando la comparación entre la i -ésima y la j -ésima botella por un experto catador de vino. El juicio podría venir dado comparando una copa del vino i con una del vino j , clasificando sus diferencias en una escala en la que el 0 representaría que los vinos son idénticos y el 10 que son completamente diferentes. Si en lugar de uno se supone que hay varios catadores, entonces los datos serán δ_{ijr} donde i, j se refieren a los vinos y r al r -ésimo catador.

Clasificación del MDS

Según los dos conjuntos mencionados se clasifica en:

- 1 **Número de modas.** Se denomina moda a cada conjunto de objetos subyacente en los datos para el análisis con MDS. Así, las disimilaridades δ_{ijr} del ejemplo de los vinos son datos bimodales, una moda corresponde a los vinos y la otra a los catadores.
- 2 **Número de vías.** Se denomina vía a cada índice en la medida entre objetos. Así, en el ejemplo anterior los datos son a tres vías.

Si solo se considera un catador, los datos serían unimodales y a dos vías.

Modelos en MDS

- Para describir los diferentes modelos de MDS se hace referencia al tipo de datos más usual que son: disimilaridades, unimodal y a dos vías.
- Sean n objetos y una matriz de disimilaridades δ_{ij} . El objetivo del MDS será encontrar una configuración de puntos en un espacio de dimensión ν de forma que las distancias d_{ij} , no necesariamente euclídeas, entre cada par de puntos (i, j) representen tan bien como sea posible las disimilaridades δ_{ij} . Las diferentes formas de *representación* establecen dos grandes grupos: métrico y no métrico.

MDS métrico

Toman en cuenta, además del orden, el valor intrínseco de las disimilaridades . Suele estar asociado a datos de tipo intervalo y de razón y trata de encontrar un conjunto de puntos en un espacio de forma que cada punto represente uno de los objetos y las distancias entre los puntos sean aproximadamente iguales cuantitativamente a las correspondientes disimilaridades transformadas mediante la relación

$$d_{ij} \approx f(\delta_{ij}),$$

donde f es una función continua, paramétrica y monótona. Dentro de estos modelos pueden destacarse principalmente los siguientes:

MDS métrico

- *MDS clásico*: trata las disimilaridades directamente como distancias euclídeas mediante la relación $d_{ij} = \delta_{ij}$, usando la descomposición espectral de una matriz doblemente centrada para obtener la matriz de configuración. Algunos de los más importantes son el de Torgerson (1958) e INDSCAL de Carroll y Chang (1970).
- *MDS mínimo cuadrático*: obtienen la configuración ajustando mediante mínimos cuadrados distancias $\{d_{ij}\}$ a disimilaridades transformadas $\{f(\delta_{ij})\}$. Entre los más importantes están ALSCAL y PROXSCAL (Young, Takane y Lewyckyj (1978) y De Leeuw y Heiser (1980)).

MDS métrico

- *MDS desviación absoluta*: obtienen la configuración ajustando mediante desviaciones absolutas distancias $\{d_{ij}\}$ a disimilaridades transformadas $\{f(\delta_{ij})\}$, con f una función del tipo descrito anteriormente (Brusco (2001)).
- *MDS máximo verosímil*: bajo la hipótesis de una distribución estadística de los datos, permite ajustar las disimilaridades a las distancias mediante una función de verosimilitud, como el modelo de Ramsay (1982), Vera y González (1996) y Orellana, Vera, Gonzalez, Pascual y Rufián (2004).

MDS no métrico

Si se abandona la naturaleza métrica de la transformación de las disimilaridades se obtienen modelos de MDS no métricos. La transformación f en este caso puede ser arbitraria y solo obedece a la restricción monótona

$$\delta_{ij} < \delta_{rs} \quad \Rightarrow \quad f(\delta_{ij}) \leq f(\delta_{rs}).$$

Por tanto, solo las ordenaciones de las disimilaridades deben preservarse por la transformación y de ahí el término no métrico. Entre los modelos más importantes de este tipo están: mínimo cuadrático y máximo verosímil.

MDS no métrico

- 1 *MDS mínimo cuadrático*. Obtienen la configuración ajustando mediante mínimos cuadrados distancias a disimilaridades transformadas. Entre los más utilizados se tienen PROXSCAL y ALSCAL (Kruskal (1964b)).
- 2 *MDS máximo verosímil*. El modelo de Takane (1981) bajo hipótesis de normalidad en los errores permite la estimación por máxima verosimilitud de los parámetros. Una extensión puede verse en García, Vera y González (1995).

El método CLÁSICO

Sea Φ una aplicación arbitraria,

$$\Phi : \Omega \mapsto \mathbb{R}^\nu$$

de forma que $\Phi(o_i) = x_i$, es un vector de dimensión ν que contiene las coordenadas del objeto o_i en el espacio de representación que se supone de dimensión ν . Sea d_{ij} la distancia entre dos puntos $o_i, o_j \in \Omega$ de \mathbb{R}^ν .

El objetivo es encontrar una aplicación Φ para la que $d_{ij} \approx \delta_{ij}$, $\forall i, j$. El método clásico minimiza la dispersión mediante la transformación de Young y Houselholder, Teorema 1, para productos escalares. Es un modelo anidado según Román, González y Vera (1999).

ALSCAL (Alternating Least Squared sCALing)

ALSCAL es un modelo que permite analizar datos mediante un modelo euclídeo simple o ponderado que pueden ser:

- 1 nominales, ordinales de intervalo o razón;
- 2 poseer observaciones faltantes;
- 3 simétricos o asimétricos;
- 4 condicionales o incondicionales;
- 5 con o sin réplicas y
- 6 discretos o continuos.

ALSCAL (Alternating Least Squared sCALing)

ALSCAL es el más completo ya que es el único que permite realizar análisis métricos y no métricos, siendo sus resultados métricos similares a los de INDSCAL, mientras que en el caso de ajustar modelos euclídeos simples sus resultados son similares a los del modelo de Kruskal.

ALSCAL (Alternating Least Squared sCALing)

Es un método iterativo. Cada iteración consiste en dos fases principales,

- transformar los datos, *fase de representación óptima* y
- estimar las coordenadas de los estímulos y los pesos de los sujetos (cuando sea el caso), *fase de estimación del modelo*.

Ambas fases van alternándose hasta que se obtiene la convergencia global del proceso. Puesto que cada fase es un procedimiento mínimo cuadrático, es por lo que recibe ese nombre.

ALSCAL (Alternating Least Squared sCALing)

El criterio de optimización es denominado SSTRESS, que es un criterio mínimo cuadrático dado por:

$$SS(X, W, \hat{D}^2) = \sum_{r=1}^R \sum_{i=1}^n \sum_{j=1}^{i-1} (d_{ijr}^{\hat{2}} - d_{ijr}^2)^2, \quad (5)$$

salvo un factor de normalización. Este criterio es semejante en su forma al criterio STRESS de Kruskal, salvo que utiliza las disparidades y las distancias al cuadrado. De ahí su nombre *Squared STRESS*.

PROXSCAL (PROXimity SCALing)

Es desarrollado por Commandeur y Heiser (1993) siendo una integración y extensión de los trabajos pioneros de Kruskal, Guttman y Carroll, tratando el problema de MDS para datos de proximidades encontrando la representación de mínimos cuadrados de los objetos en un espacio de dimensión baja, que mediante un algoritmo de mayorización garantiza convergencia monótona para una transformación óptima, con datos métricos y no métricos sobre una variedad de modelos y restricciones.

PROXSCAL (PROXimity SCALing)

El criterio de optimización es denominado STRESS, que es un criterio mínimo cuadrático dado principalmente en función de las distancias de la forma,

$$S(X, W, \hat{D}) = \sum_{r=1}^R \sum_{i=1}^n \sum_{j=1}^{i-1} (\hat{d}_{ijr} - d_{ijr})^2, \quad (6)$$

salvo un factor de normalización.

ALSCAL y PROXSCAL

ALSCAL y PROXSCAL han sido implementados en SPSS, de ahí que su utilización se haya difundido pasando a ser los modelos más empleados. Sin embargo, ambos son propensos a óptimos locales según anota De Leeuw (2000), pues claramente al ser (5) y (6) polinomios multivariantes de cuarto y segundo grado, respectivamente, plantean un gran número de óptimos locales; por lo que es necesario aplicar una técnica novedosa para la búsqueda de óptimos globales de las funciones $SStress$ y $Stress$, como lo estudió Brusco (2001) y Murillo, Vera y Heiser (2005 y 2007).

Bibliografía

-  Murillo, A. (2004). *Representación de asociaciones mediante MDS y su tratamiento computacional*. Tesis doctoral, Universidad de Granada, España.
-  Murillo, A., Vera, J.F. y Heiser, W. (2005). A Permutation-Translation Simulated Annealing Algorithm for L1 and L2 Unidimensional Scaling. *Journal of Classification*, 22: 119-138.
-  Vera, J.F., Heiser, W. y Murillo, A. (2007). Global Optimization in Any Minkowski Metric: A Permutation-Translation Simulated Annealing Algorithm for Multidimensional Scaling. *Journal of Classification*, 24: 277-301.