

ANÁLISIS DE PROXIMIDADES MÉTRICO USANDO BÚSQUEDA TABÚ

MARIO VILLALOBOS* — JAVIER TREJOS**

Recibido: 4 Abril 2000

Resumen

Se aplica la técnica de búsqueda tabú (BT) en análisis de proximidades, obteniéndose buenos resultados, comparables a los obtenidos con sobrecalentamiento simulado. Un estado en BT es una configuración de n puntos en un espacio p dimensional, y un vecino se define por la traslación de longitud h de una o más de las coordenadas de un punto.

Palabras clave: escalamiento multidimensional métrico, análisis de proximidades, búsqueda tabú, optimización combinatoria, discretización.

Abstract

We apply tabu search (TS) in metric multidimensional scaling, obtaining good results comparable to those obtained with simulated annealing. A state in TS is a configuration of n points in a p dimensional space, and a neighbour is defined by the translation of length h of one or more coordinates of a point.

Keywords: metric multidimensional scaling, tabu search, combinatorial optimization, discretization.

Mathematics Subject Classification: 62-07, 90C27, 92G15

1. Introducción

En el análisis de proximidades métrico (conocido como *Multidimensional Scaling* en inglés y abreviado MDS) se tiene un conjunto Ω de objetos o individuos y una tabla

*CIMPA–PIMAD, Escuela de Matemática, Universidad de Costa Rica, 2060 San José, Costa Rica; Tel. +(506) 207 5574; Fax: +(506) 207 4397; E-Mail: mvillalo@cariari.ucr.ac.cr

**Misma dirección; E-mail: jtrejos@cariari.ucr.ac.cr

$(\delta_{ij})_{n \times n}$ de las disimilitudes entre estos individuos, y se busca un conjunto de n puntos en \mathbb{R}^p , $X = (\vec{x}_1, \vec{x}_2, \dots, \vec{x}_n)$, de manera que las distancias euclídeas, $(d_{ij}(X))_{n \times n}$ entre estos puntos sean lo más “cercanas” posibles de la tabla de disimilitudes dada $(\delta_{ij})_{n \times n}$.

Para medir qué tan cercanos están estas dos matrices se utiliza la siguiente función, llamada Stress normalizado:

$$\sigma^2(X) = \frac{\sum_{i=1}^n \sum_{j=1}^n w_{ij} [\delta_{ij} - d_{ij}(X)]^2}{\sum_{i=1}^n \sum_{j=1}^n w_{ij} \delta_{ij}^2},$$

donde w_{ij} son los pesos asociados con las disimilitudes, que se utilizan para dar alguna información extra, como la ausencia de las disimilitudes ($w_{ij} = 0$ si no se tiene δ_{ij}).

2. Introducción

Torgerson en 1958 resolvió este problema para el caso en que las disimilitudes son distancias euclídeas, mediante diagonalización. En este caso p es el número de valores propios no nulos de la matriz que se diagonaliza, ver Bourouche (1997) o Cox & Cox (1995). En Kruskal (1964a) y Kruskal (1964b) se da un algoritmo y un programa que da una solución para el caso general, definiendo una función a optimizar llamada Stress. El método de Kruskal esá basado en un descenso del gradiente y obtiene una solución subóptima.

Debido a que no se conoce una solución exacta al problema planteado, diversos autores han aplicado diferentes técnicas de optimización para tratar de obtener el óptimo global.

Algunos ejemplos de estas técnicas son Mayorización Iterativa, por de Leeuw (1988), en la que se busca una función que mayorice la función a optimizar, en el sentido de que sea mayor en general pero que sea igual a la función que se está optimizando en el punto de la iteración, y que sea más fácil de encontrar el óptimo para ella, este último óptimo será el siguiente valor de la iteración. El método de *Tunneling* usado por Groenen (1993) en el que se pretende salir de un óptimo local encontrando otro punto con el mismo valor de la función que el óptimo local y luego seguir descendiendo, para ello utiliza una función llamada la función de Tunneling.

El problema de estas técnicas de optimización es que no se puede garantizar la convergencia al óptimo global.

Con respecto a técnicas de optimización estocásticas, Trejos & Villalobos (1998), Trejos & Villalobos (1999), Villalobos (1998), Trejos et al. (2000) aplicaron Sobrecalentamiento Simulado al problema obteniendo buenos resultados. En Groenen et al. (2000) se aplica el sobrecalentamiento simulado, algoritmos genéticos, mayorización iterativa y alisamiento a un problema de gran escala concerniente a la localización de emisores de señales para teléfonos celulares, obtiéndose muy buenos resultados para el algoritmo genético. También han aplicado algoritmos genéticos Mathar (1995) y N’Gouenet (1995).

Una ventaja de estas últimas técnicas es que se puede garantizar la convergencia asintótica al óptimo global. En Trejos et al. (1998) se hace uso de estas técnicas de optimización estocásticas, en particionamiento, con muy buenos resultados. En Trejos & Castillo se usa el sobrecalentamiento simulado en el problema de particionamiento bimodal, con mejores resultados que para otras técnicas clásica. Esto nos ha dado la idea de usar BT en MDS y este es el objeto del presente trabajo.

3. Búsqueda tabú

La BT (Glover (1989) y Glover (1990)) pretende obtener el óptimo global generando una sucesión de estados pasando de un estado a otro estado “vecino”, manteniendo una lista tabú de los últimos lugares o movimientos que se han realizado para tratar de evitar que se vuelva a pasar por los mismos lugares varias veces. Esta lista tiene una longitud finita, que es un parámetro del problema. Este principio se puede cambiar adicionando el criterio de aspiración que establece que si un movimiento no permitido hace que se mejore el valor actual del criterio, el movimiento se aceptará.

Para la aplicación de BT es necesario definir para cada estado el conjunto de estados vecinos a los cuales se puede trasladar en una iteración.

La lista tabú puede contener los estados recién visitados o preferentemente una codificación de los movimientos que permitieron pasar de un estado a otro.

Es importante también definir la manera de actualizar la lista es decir cuándo un elemento entra o sale de ella.

En vista de que se deben evaluar todos los vecinos de un estado, y éstos pueden ser muchos, es conveniente tener una forma rápida de evaluar la variación de la función a optimizar. En casos de dimensiones muy grandes. algunos autores sugieren utilizar una muestra de los vecinos.

4. Implementación

Como se mencionó, lo primero que se debe definir es un estado. Como se busca una configuración de n puntos en \mathbb{R}^p que representen los objetos, en MDS definimos un estado como:

$$I = [\vec{x}_1^I, \vec{x}_2^I, \dots, \vec{x}_n^I],$$

en donde cada $\vec{x}_i^I \in (h\mathbb{Z})^p$, con $h\mathbb{Z} = \{x \in \mathbb{R}/x = hr, r \in \mathbb{Z}\}$, $h \in \mathbb{R}$ fijo, o sea se ha discretizado \mathbb{R}^p por medio de un mallado de ancho h .

Los estados vecinos de un estado $I = [\vec{x}_1^I, \vec{x}_2^I, \dots, \vec{x}_n^I]$ son los que tienen los mismos vectores que I salvo uno de ellos que es una variación de una de las coordenadas del punto que está en I , esto es: J es un vecino de I si existen $l \in \{1, \dots, n\}$ y $\vec{\alpha} \in \{-1, 0, 1\}^p$, $\|\vec{\alpha}\| = 1$ (sólo una de las coordenadas de $\vec{\alpha}$ es 1 ó -1) tal que:

$$\vec{x}_i^J = \begin{cases} \vec{x}_i^I & \text{si } i \neq l, i \in \{1, \dots, n\}, \\ \vec{x}_i^I + h\vec{\alpha} & \text{si } i = l. \end{cases}$$

Nótese que el número de vecinos de un estado dado es $2np$.

La función a optimizar será el Stress normalizado que se denota por:

$$\sigma^2(I) = \frac{\sum_{i=1}^n \sum_{j=1}^n w_{ij} [\delta_{ij} - d_{ij}(I)]^2}{\sum_{i=1}^n \sum_{j=1}^n w_{ij} \delta_{ij}^2},$$

donde I es un estado y $d_{ij}(I) = d(\vec{x}_i^I, \vec{x}_j^I) = \|\vec{x}_i^I - \vec{x}_j^I\|$.

Para calcular el valor de $\sigma^2(J)$ en los vecinos de un estado se debería hacer una gran cantidad de cálculos, del orden de n^2 , pero como los vecinos de I se obtienen a partir de éste modificando sólo un punto, las distancias entre los demás puntos de I se mantiene igual, por lo que en la función de Stress sólo cambian los valores que contengan los términos $d_{il}(I) = d_{li}(I)$, donde l es el índice del punto que varía en el estado I , para obtener el vecino. Entonces para evaluar $\sigma^2(J)$, con J un vecino de I , basta con restarle a $\sigma^2(I)$, los valores en los que aparezcan los términos $d_{il}(I)$ ó $d_{li}(I)$ y sumarle los valores correspondientes del estado J . Así se reduce al orden n el cálculo de $\Delta_{IJ}(\sigma^2) = \sigma^2(J) - \sigma^2(I)$:

$$\Delta_{IJ}(\sigma^2) = \sum_{i=1}^n \left\{ w_{il} ([\delta_{il} - d_{il}(J)]^2 - [\delta_{il} - d_{il}(I)]^2) + w_{li} ([\delta_{li} - d_{li}(J)]^2 - [\delta_{li} - d_{li}(I)]^2) \right\}.$$

Los elementos de la lista tabú serán los movimientos inversos de los movimientos que se han realizado para llegar al estado actual, es decir que si para pasar del estado I al estado J se movió el punto con índice l , en la dirección $\vec{\alpha}$, entonces en la lista tabú se agregará el par $(l, -\vec{\alpha})$.

Al inicio de las iteraciones, se toma una configuración al azar y la lista tabú es vacía. Cada iteración, consiste en evaluar el cambio en el Stress para cada estado consistente en perturbar una coordenada de algún punto de la configuración actual. Se escoge la nueva configuración con el menor stress en el vecendario si ella es la mejor encontrada hasta ahora, sino, se escoge la mejor configuración tal que el movimiento que la genera no está en la lista tabú. Una vez realizado el movimiento se actualiza la lista tabú.

Inicialmente se utilizó una lista tabú de longitud pequeña, pero no dió buenos resultados (ver 5.1) debido a que se necesitan muchos movimientos para salirse de un óptimo local, por lo que la longitud de la lista tabú debe ser suficientemente larga para poder permitir que los puntos tengan suficiente movilidad para poder salirse de los óptimos locales, puede tomarse como el 80% del total del número de vecinos.

5. Resultados obtenidos

Con lo anterior se implentó un algoritmo y programa, TABU_SCAL, con el que se han corrido los ejemplos que se presentan a continuación.

Los nueve puntos

En este ejemplo las disimilitudes que se tienen son las distancias entre nueve puntos localizados en las intersecciones de una malla de 3×3 . Al utilizar TABU_SCAL con una lista tabú pequeña de un total de 100 corridas del programa se localizaron varios óptimos

locales y en el 74 % se obtuvo el óptimo global, que es similar al 72 % que reporta Groenen (1993) al utilizar Multistart, por lo que se decidió aumentar la longitud de la lista tabú al doble y en este caso se obtuvo que en un 86 % de los casos se llegó al óptimo global y el 14 % restante convergió al segundo mejor valor reportado. Por lo anterior se recomienda utilizar una lista tabú de longitud suficientemente grande. En Villalobos (1998), utilizando sobrecalentamiento simulado (SS) se reporta que de un total de 273 ejecuciones se obtuvo el óptimo global en el 95.47 % de los casos.

Las Distancias 0–1

Este es un ejemplo en el que las disimilitudes corresponde a la distancia 0,1; en la cual la distancia entre dos individuos es 1 si son diferentes y 0 si son iguales. Para el caso en que se tienen 4 individuos ($n = 4$) se ejecutó 100 veces y de estas en 98 de ellas se llegó al óptimo global, 0.028595. Cuando se tienen 6 individuos de las 100 veces en 87 % se obtuvo el óptimo global, 0.071457, semejante al valor reportado por Villalobos (1998) 88.8 % al utilizar sobrecalentamiento simulado.

Los Países del Mundo

En este caso se tienen las similitudes entre 12 países del mundo, de acuerdo con el criterio de 18 estudiantes, fue reportado por Kruskal (1976). De esta tabla de similitudes se obtiene una de disimilitudes que se puede encontrar en Villalobos (1998) (p. 63). A esta tabla se le aplicó TABU_SCAL obteniéndose que el mejor valor fue 0.0474600 que es similar al reportado por Villalobos (1998) 0.047419 al utilizar SS. Este valor se obtuvo en el 50 % de los casos que se ejecutó el programa, mientras que Villalobos (1998) reporta un 46.2 %.

6. Perspectivas y conclusiones

Con la utilización de BT en MDS se han obtenido buenos resultados, comparables con los obtenidos por Villalobos (1998) utilizando SS, sin embargo se deben hacer más estudios con respecto a la longitud de la lista tabú y ver la revisión de solo una parte de los vecinos, como recomiendan Murillo, Trejos & Piza (2000). También se debe hacer un análisis sobre la escogencia de una configuración inicial ya sea por medio de un descenso o como recomienda Mathar (1997) haciendo un intercambio de etiquetas de los puntos. Finalmente, se deben hacer comparaciones exhaustivas con SS y otras técnicas de optimización en cuanto al tiempo de ejecución, dificultad de programación y el algoritmo.

Referencias

- Bouroche, J.-M. et al. (1977) *Analyse des Données en Marketing*. Masson, Paris.
- Cox, T.; Cox, M. (1995) *Multidimensional Scaling*. Chapman & Hall, Londres.
- de Leeuw, J. (1988) “Convergence of the majorization method for multidimensional scaling” *Journal of Classification* **5**: 163–180.

- Glover, F. (1989) “Tabu search - Part I”, *ORSA J. Comput.* 1: 190–206.
- Glover, F. (1990) “Tabu search - Part II”, *ORSA J. Comput.* 2: 4–32.
- Groenen, P.J.F. (1993) “A comparison of two methods for global optimization in multidimensional scaling”, in O. Opitz, B. Lausen & R. Klar (Eds.) *Information and Classification*. Springer-Verlag, Berlin: 145–155
- Groenen, P.J.F.; Mathar, R.; Trejos, J. (2000) “Global optimization methods for multidimensional scaling applied to mobile communications”, in W. Gaul, O. Opitz, M. Schader (Eds.) *Data Analysis. Scientific Modeling and Practical Application*. Springer-Verlag, Berlin: 459–469.
- Kruskal, J. (1964a) “Multidimensional scaling by optimizing goodness of fit to a nonmetric hypothesis” *Psychometrika* **29**: 1–27.
- Kruskal, J. (1964b) “Non-metric multidimensional scaling: a numerical method” *Psychometrika* **29**: 115–129.
- Kruskal, J. (1976) “More factors than subjects, test and treatments: an indeterminacy theorem for canonical decomposition and individual differences”, *Psychometrika* **41**: 281–293.
- Mathar, R. (1995) “A genetic algorithm for multidimensional scaling”. Internal Report, RWTH, Aachen.
- Mathar, R. (1997) “A hybrid global optimization algorithm for multidimensional scaling”, Preprint, Institut für Statistik, RWTH Aachen.
- Murillo, A.; Piza, E.; Trejos, J. (2000) “Combinatorial optimization heuristics in partitioning with non euclidean distances”, Preprint, CIMPA, San José.
- Ngouenet, R.F. (1995) *Analyse Géométrique des Données de Dissimilarité par la Multidimensionnal Scaling: une Approche Parallèle Basée sur les Algorithmes Génétiques. Application aux Séquences Biologiques*. Thèse de Doctorat, Université de Rennes I, Rennes.
- Torgerson, W. S. (1958) *Theory and Methods of Scaling*. John Wiley & Sons, New York.
- Trejos, J.; Murillo, A.; Piza, E. (1998) “Global stochastic optimization for partitioning”, in A. Rizzi et al. (Eds.) *Advances in Data Science and Classification*, Springer-Verlag, Berlin: 185–190.
- Trejos, J.; Villalobos, M. (1998) “Análisis de proximidades usando sobrecalentamiento simulado” in W. Castillo & J. Trejos (Eds.) *XI Simposio Internacional de Métodos Matemáticos Aplicados a las Ciencias*, UCR-ITCR, Costa Rica.
- Trejos, J.; Villalobos, M. (1999) “Use of simulated annealing in metric multidimensional scaling”, in: M. Greenacre & J. Blasius (Eds.) *Large Scale Data Analysis Conference*, Cologne, May 1999.
- Trejos, J.; Castillo, W.; González, J.; Villalobos, M. (2000) “Application of simulated annealing in some multidimensional scaling problems”, in: H.A.L. Kiers, J.-P. Rasson, P.J.F. Groenen & M. Schader (Eds.) *Data Analysis, Classification, and Related Methods*, Springer-Verlag, Berlin: 297–302.
- Trejos, J. & Castillo, W. (2000) “Simulated annealing optimization for two-mode partitioning”, in: W. Gaul & R. Decker (Eds.) *Classification and Information Processing at the Turn of the Millennium*, Springer-Verlag, Berlin: 135–142.
- Villalobos, M. (1998) *Optimización Estocástica para el Análisis de Proximidades*. Tesis de Maestría, Universidad de Costa Rica, San José.