Applying Ideas from Homogeneity Analysis to Visualize Similarity Data

Nees Jan van Eck   Ludo Waltman

Econometric Institute, Erasmus School of Economics
Erasmus University Rotterdam
P.O. Box 1738, 3000 DR Rotterdam, The Netherlands
Email: {nvaneck,lwaltman}@few.eur.nl

Correspondence Analysis and Related Methods (CARME 2007)

Rotterdam, The Netherlands
June 25–27, 2007

ERASMUS UNIVERSITEIT ROTTERDAM

Outline of the presentation

Homogeneity analysis

VOS

Empirical comparison between VOS, MDS, and DAM

Application to a large data set

Conclusions

Outline of the presentation

Homogeneity analysis

VOS

Empirical comparison between VOS, MDS, and DAM

Application to a large data set

Conclusions

Homogeneity analysis

Homogeneity analysis

- A method for visualizing the relations between categories and objects

Homogeneity analysis

Homogeneity analysis

- A method for visualizing the relations between categories and objects

Objective

- To provide a low-dimensional visualization in which categories and objects are located in such a way that each category is the centroid of all objects that score on the category

Constrained optimization problem

Objective function

$$\sigma(\mathbf{X}, \mathbf{Y}; \mathbf{G}) = \sum_{i<j} g_{ij} \|\mathbf{x}_i - \mathbf{y}_j\|^2$$

where $\mathbf{x}_i$ and $\mathbf{y}_j$ denote, respectively, the location of object i and category j and $g_{ij} \in \{0, 1\}$ denotes whether or not object i scores on category j

Constrained optimization problem

Objective function

$$\sigma(\mathbf{X}, \mathbf{Y}; \mathbf{G}) = \sum_{i<j} g_{ij} \|\mathbf{x}_i - \mathbf{y}_j\|^2$$

where $\mathbf{x}_i$ and $\mathbf{y}_j$ denote, respectively, the location of object i and category j and $g_{ij} \in \{0, 1\}$ denotes whether or not object i scores on category j

Constraints

$$\mathbf{X}'\mathbf{X} = n\mathbf{I}$$
$$\mathbf{1}'\mathbf{X} = 0$$

Outline of the presentation

ERASMUS UNIVERSITEIT ROTTERDAM

VOS

- An abbreviation for *visualization of similarities*
- A new method for visualizing similarities between objects, based on ideas similar to homogeneity analysis

VOS

VOS

- An abbreviation for *visualization of similarities*
- A new method for visualizing similarities between objects, based on ideas similar to homogeneity analysis

Objective

- To provide a low-dimensional visualization in which objects are located in such a way that the distance between any pair of objects reflects their similarity as accurately as possible

Mathematical notation

Input

- n: number of objects
- m: number of dimensions of the solution
- $\mathbf{S} = (s_{ij})$: $n \times n$ similarity matrix (measured on a ratio scale)

Mathematical notation

Input

- n: number of objects
- m: number of dimensions of the solution
- $\mathbf{S} = (s_{ij})$: $n \times n$ similarity matrix (measured on a ratio scale)

Output

- $\mathbf{X}$: $n \times m$ coordinate matrix
  $\mathbf{x}_i = (x_{i1}, \ldots, x_{im})$ contains the coordinates of object i

Constrained optimization problem

Objective function

$$E(\mathbf{X};\mathbf{S}) = \sum_{i<j} s_{ij} \|\mathbf{x}_i - \mathbf{x}_j\|^2$$

Constrained optimization problem

Objective function

$$E(\mathbf{X}; \mathbf{S}) = \sum_{i<j} s_{ij} \|\mathbf{x}_i - \mathbf{x}_j\|^2$$

Constraints

- Homogeneity analysis:

$$\mathbf{X}'\mathbf{X} = n\mathbf{I}$$
$$\mathbf{1}'\mathbf{X} = 0$$

ERASMUS UNIVERSITEIT ROTTERDAM

Constrained optimization problem

Objective function

$$E(\mathbf{X};\mathbf{S}) = \sum_{i<j} s_{ij}\|\mathbf{x}_i - \mathbf{x}_j\|^2$$

Constraints

- Homogeneity analysis:

$$\mathbf{X}'\mathbf{X} = n\mathbf{I}$$
$$\mathbf{1}'\mathbf{X} = 0$$

- VOS:

$$\sum_{i<j} \|\mathbf{x}_i - \mathbf{x}_j\| = n(n-1)$$

Motivation for the objective function

- Ideal coordinates of object i

$$c_i(\mathbf{X}, \mathbf{S}) = \frac{\sum_{j \neq i} s_{ij} \mathbf{x}_j}{\sum_{j \neq i} s_{ij}}$$

ERASMUS UNIVERSITEIT ROTTERDAM

Motivation for the objective function

- Ideal coordinates of object i

$$c_i(\mathbf{X}, \mathbf{S}) = \frac{\sum_{j \neq i} s_{ij} \mathbf{x}_j}{\sum_{j \neq i} s_{ij}}$$

- Objective function for object i when the coordinates of all other objects are fixed

$$E_i(\mathbf{x}_i; \mathbf{X}, \mathbf{S}) = \sum_{j \neq i} s_{ij} \|\mathbf{x}_i - \mathbf{x}_j\|^2$$

ERASMUS UNIVERSITEIT ROTTERDAM

Motivation for the objective function

- Ideal coordinates of object i

$$c_i(\mathbf{X}, \mathbf{S}) = \frac{\sum_{j \neq i} s_{ij} \mathbf{x}_j}{\sum_{j \neq i} s_{ij}}$$

- Objective function for object i when the coordinates of all other objects are fixed

$$E_i(\mathbf{x}_i; \mathbf{X}, \mathbf{S}) = \sum_{j \neq i} s_{ij} \|\mathbf{x}_i - \mathbf{x}_j\|^2$$

- Minimization of this function results in

$$\mathbf{x}_i = c_i(\mathbf{X}, \mathbf{S})$$

Motivation for the objective function

- Ideal coordinates of object i

$$c_i(\mathbf{X}, \mathbf{S}) = \frac{\sum_{j \neq i} s_{ij} \mathbf{x}_j}{\sum_{j \neq i} s_{ij}}$$

- Objective function for object i when the coordinates of all other objects are fixed

$$E_i(\mathbf{x}_i; \mathbf{X}, \mathbf{S}) = \sum_{j \neq i} s_{ij} \|\mathbf{x}_i - \mathbf{x}_j\|^2$$

- Minimization of this function results in

$$\mathbf{x}_i = c_i(\mathbf{X}, \mathbf{S})$$

- VOS seems to have the tendency to locate objects close to their ideal coordinates

Relationship between VOS and MDS

Objective function of weighted ratio MDS

$$\sigma(\mathbf{X}; \mathbf{D}, \mathbf{W}) = \sum_{i<j} w_{ij}(d_{ij} - \|\mathbf{x}_i - \mathbf{x}_j\|)^2$$

where d_{ij} denotes the dissimilarity between objects i and j and w_{ij} denotes the weight of objects i and j

Relationship between VOS and MDS

Objective function of weighted ratio MDS

$$\sigma(\mathbf{X}; \mathbf{D}, \mathbf{W}) = \sum_{i<j} w_{ij}(d_{ij} - \|\mathbf{x}_i - \mathbf{x}_j\|)^2$$

where d_{ij} denotes the dissimilarity between objects i and j and w_{ij} denotes the weight of objects i and j

Theorem

Let $s_{ij} > 0$ for all i and j $(i \neq j)$.

Relationship between VOS and MDS

Objective function of weighted ratio MDS

$$\sigma(\mathbf{X}; \mathbf{D}, \mathbf{W}) = \sum_{i<j} w_{ij}(d_{ij} - \|\mathbf{x}_i - \mathbf{x}_j\|)^2$$

where d_{ij} denotes the dissimilarity between objects i and j and w_{ij} denotes the weight of objects i and j

Theorem

Let $s_{ij} > 0$ for all i and j ($i \neq j$).
Let similarities be transformed into dissimilarities using $d_{ij} = s_{ij}^{-1}$ ($i \neq j$).

Relationship between VOS and MDS

Objective function of weighted ratio MDS

$$\sigma(\mathbf{X};\mathbf{D},\mathbf{W}) = \sum_{i<j} w_{ij}(d_{ij} - \|\mathbf{x}_i - \mathbf{x}_j\|)^2$$

where d_{ij} denotes the dissimilarity between objects i and j and w_{ij} denotes the weight of objects i and j

Theorem

Let $s_{ij} > 0$ for all i and j $(i \neq j)$.

Let similarities be transformed into dissimilarities using $d_{ij} = s_{ij}^{-1}$ $(i \neq j)$.

Let the weight of objects i and j equal $w_{ij} = s_{ij}$.

Relationship between VOS and MDS

Objective function of weighted ratio MDS

$$\sigma(\mathbf{X}; \mathbf{D}, \mathbf{W}) = \sum_{i<j} w_{ij}(d_{ij} - \|\mathbf{x}_i - \mathbf{x}_j\|)^2$$

where d_{ij} denotes the dissimilarity between objects i and j and w_{ij} denotes the weight of objects i and j

Theorem

Let $s_{ij} > 0$ for all i and j ($i \neq j$).

Let similarities be transformed into dissimilarities using $d_{ij} = s_{ij}^{-1}$ ($i \neq j$).

Let the weight of objects i and j equal $w_{ij} = s_{ij}$.

VOS and weighted ratio MDS are then equivalent in the sense that solutions from these methods differ only by a multiplicative constant.

ERASMUS UNIVERSITEIT ROTTERDAM

Outline of the presentation

ERASMUS UNIVERSITEIT ROTTERDAM

Empirical comparison between VOS, MDS, and DAM (1)

Real-world bibliometric data sets

- Cocitation data of 376 journals in the field of economics and management (obtained from CWTS, Leiden University)
- Co-occurrence data on 332 concepts from the field of computational intelligence (obtained from Elsevier Scopus)
- In each data set around 75% of the similarities equal zero

Empirical comparison between VOS, MDS, and DAM (1)

Real-world bibliometric data sets

- Cocitation data of 376 journals in the field of economics and management (obtained from CWTS, Leiden University)
- Co-occurrence data on 332 concepts from the field of computational intelligence (obtained from Elsevier Scopus)
- In each data set around 75% of the similarities equal zero

Methods

- VOS
- Unweighted ordinal MDS (SPSS PROXSCAL)
- Distance association model (DAM; De Rooij and Heijser, 2005)

Empirical comparison between VOS, MDS, and DAM (2)

VOS/MDS

- Co-occurrences/cocitations normalized using

$$s_{ij} = \frac{c_{ij}}{\left(\sum_k c_{kj}\right)\left(\sum_k c_{ik}\right)}$$

where c_{ij} denotes the co-occurrence/cocitation of objects i and j ($c_{ii} = 0$)

ERASMUS UNIVERSITEIT ROTTERDAM

Empirical comparison between VOS, MDS, and DAM (2)

VOS/MDS

- Co-occurrences/cocitations normalized using

$$s_{ij} = \frac{c_{ij}}{\left(\sum_k c_{kj}\right)\left(\sum_k c_{ik}\right)}$$

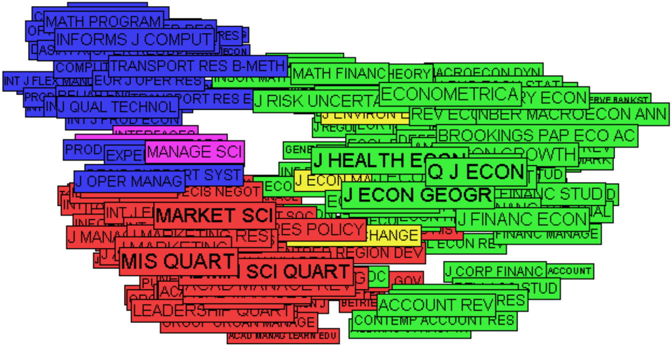
where c_{ij} denotes the co-occurrence/cocitation of objects i and j ($c_{ii} = 0$)

DAM

- One-mode DAM with symmetric associations is used
- Row and column margins are assumed equal
- Distances are transformed into associations using the Gaussian function
- Likelihood under independent Poisson sampling is maximized

ERASMUS UNIVERSITEIT ROTTERDAM

Journal cocitation map constructed using VOS



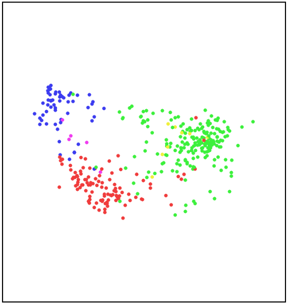
ERASMUS UNIVERSITEIT ROTTERDAM

Journal cocitation map constructed using MDS

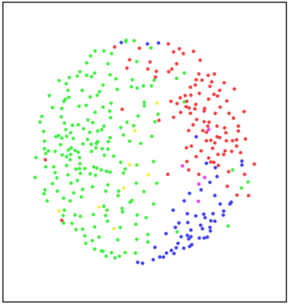
Journal cocitation map constructed using DAM

Comparison of journal cocitation maps

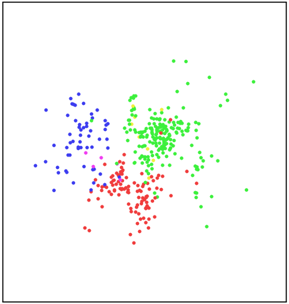
VOS



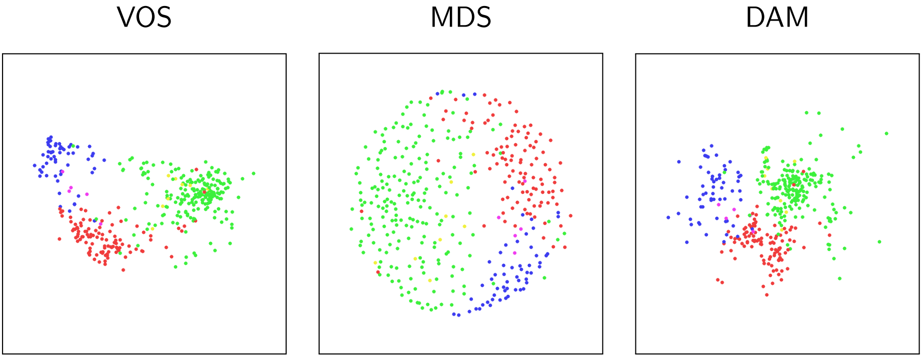
MDS



DAM



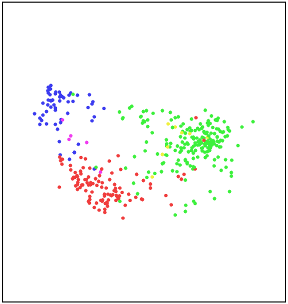
Comparison of journal cocitation maps



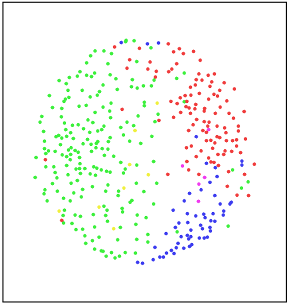
- MDS locates all objects roughly uniformly distributed within a circle

Comparison of journal cocitation maps

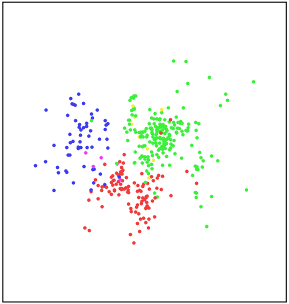
VOS



MDS



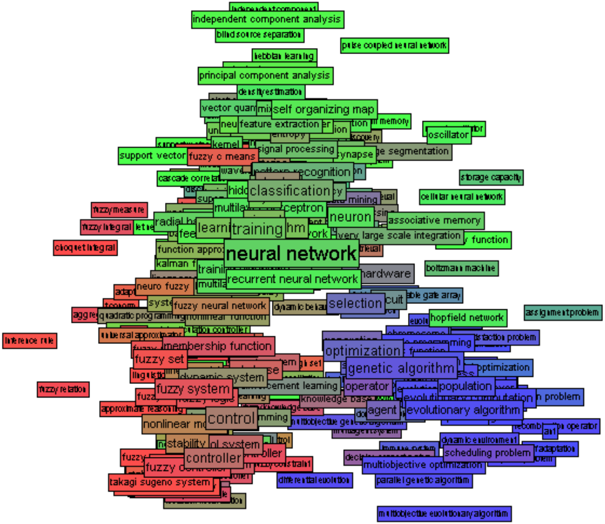
DAM



- MDS locates all objects roughly uniformly distributed within a circle
- VOS better separates the fields of economics, management, and operations research from each other than MDS and DAM

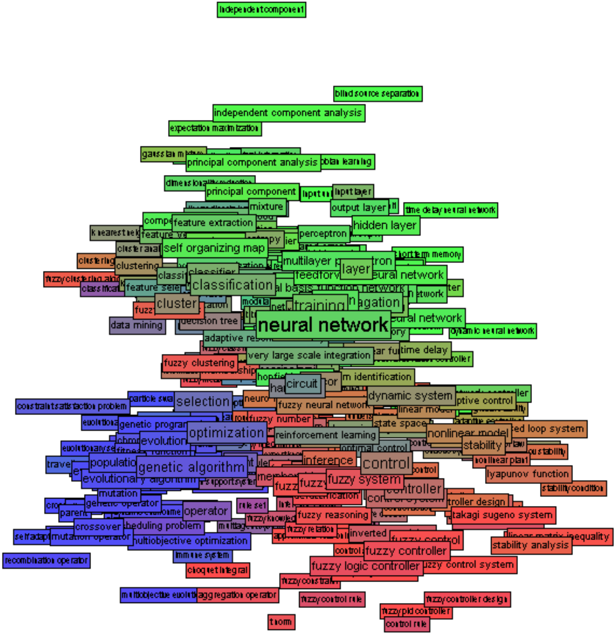
ERASMUS UNIVERSITEIT ROTTERDAM

Concept co-occurrence map constructed using VOS



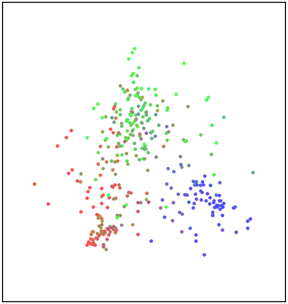
Concept co-occurrence map constructed using MDS

Concept co-occurrence map constructed using DAM

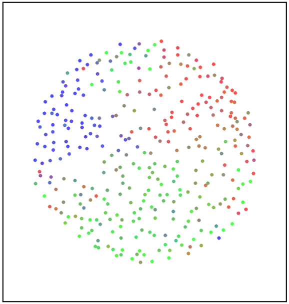


Comparison of concept co-occurrence maps

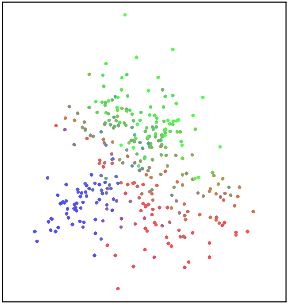
VOS



MDS



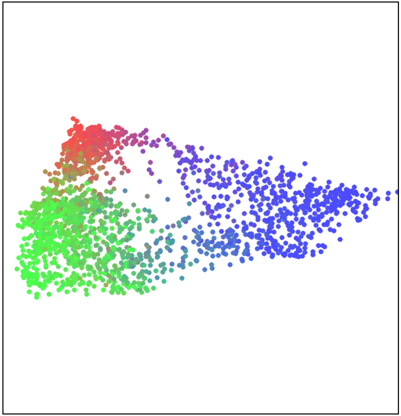
DAM

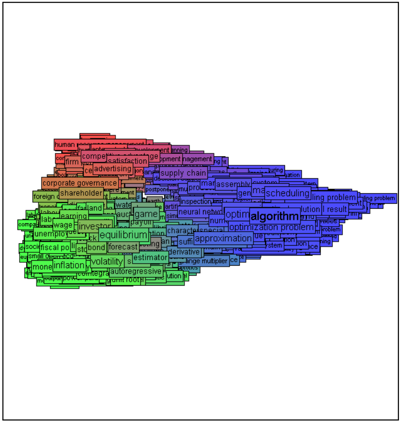


Outline of the presentation

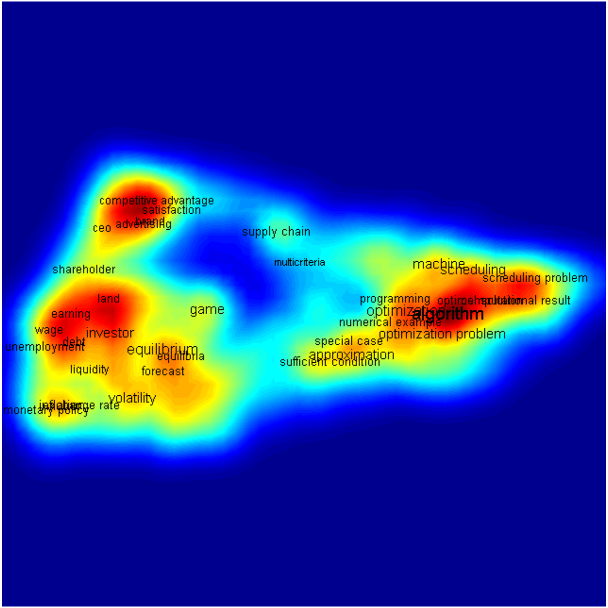
Application to a large data set

- Co-occurrence data on 2000 concepts from the field of economics and management (obtained from CWTS, Leiden University)
- Around 89% of the similarities equal zero





Concept density map



ERASMUS UNIVERSITEIT ROTTERDAM

Outline of the presentation

Homogeneity analysis

VOS

Empirical comparison between VOS, MDS, and DAM

Application to a large data set

Conclusions

ERASMUS UNIVERSITEIT ROTTERDAM

Conclusions

VOS

- A new method for visualizing similarities between objects
- The objective function is based on similar ideas as in homogeneity analysis

Conclusions

VOS

- A new method for visualizing similarities between objects
- The objective function is based on similar ideas as in homogeneity analysis

Empirical comparison between VOS, MDS, and DAM

- Unweighted ordinal MDS locates objects roughly uniformly distributed within a circle
- Solutions from VOS show better separated clusters than solutions from MDS and DAM

Conclusions

VOS

- A new method for visualizing similarities between objects
- The objective function is based on similar ideas as in homogeneity analysis

Empirical comparison between VOS, MDS, and DAM

- Unweighted ordinal MDS locates objects roughly uniformly distributed within a circle
- Solutions from VOS show better separated clusters than solutions from MDS and DAM

Future research

- The effect on homogeneity analysis of using the VOS constraint

ERASMUS UNIVERSITEIT ROTTERDAM

More information on VOS

References to our papers and a computer implementation of VOS are available at:

`http://www.neesjanvaneck.nl/vos/`

ERASMUS UNIVERSITEIT ROTTERDAM