Studying triadic distance models under a likelihood function

Mark de Rooij

Leiden University
Department of Psychology

Summary: Triadic distance models are relatively new. Their merits and demerits are fairly unknown. In the present paper we will study triadic distance models and bring the understanding of those models to a next level. Therefore, the models are studied under a Multinomial sampling scheme and a detailed investigation of the likelihood function results in relationships with multiple correspondence analysis and three-way quasi-symmetry models.

1. Introduction

The analysis of three-way tables has received an enormous amount of attention in the last few decades. In the case of distance models, the attention was initially focused on three-way two-mode data, but recently the attention shifted towards three-way one-mode and three-way three-mode data. The latter data types require a different modeling strategy: we are not after a graphical representation of one mode that is afterwards transformed for each specific instance of the second mode, but instead we are looking for a graphical representation of the entire three-way table in a single Euclidean space. The entries of the three-way table can in some way be viewed as (dis)similarities that relate three categories of (different) variables. Triadic distance models try to represent the three categories as points in a Euclidean space, such that a measure of distance for these three points as closely as possible approximates the dissimilarities in the three-way table. A popular class of triadic distance models is formed by the L_p-transform, where a triadic distance is defined on dyadic distances:

$$d_{ijk} = \left[d_{ij}^p + d_{jk}^p + d_{ik}^p \right]^{1/p}. \quad (1)$$

Here d_{ijk} is the triadic distance between the three points i, j, and k; d_{ij} are the usual (dyadic) distances between two points i and j. The precise definition of the dyadic distances is left for the next section.

In the present paper we will study triadic distances, defined by the L_2-transformation. Therefore, we apply them to contingency tables assuming a Multinomial sampling scheme, and study the likelihood function in more detail. This procedure establishes relationships of triadic distance models with Multiple Correspondence Analysis and Quasi-Symmetry models for three-way tables.

2. Triadic distance models

The study of triadic distance models was initiated by Hayashi (1972). A number of papers have been written since: Cox, Cox, and Branco (1991), Pan and Harris (1991), Joly and Le Calvé (1995), Daws (1996), Heiser and Bennani (1997), De Rooij and Heiser (2000), De Rooij (2001, submitted). Two papers (Joly and Le Calvé, 1995 and Heiser and Bennani, 1997) propose an axiomatic framework for the study of triadic distance models. All these papers assumed a three-way square, i.e., a $K \times K \times K$, proximity matrix.

We will focus on the Generalized Euclidean Model, where $p = 2$ in the L_p-transform. The triadic distance in this case is defined as the square root of the sum of squared dyadic Euclidean distances. The squared triadic distance is given by

$$d_{ijk}^2(\mathbf{X}) = d_{ij}^2(\mathbf{X}) + d_{jk}^2(\mathbf{X}) + d_{ik}^2(\mathbf{X}), \tag{2}$$

where $d_{ij}(\mathbf{X})$ is the usual Euclidean distance between points i and j, i.e.,

$$d_{ij}(\mathbf{X}) = \left[\sum_m (x_{im} - x_{jm})^2\right]^{1/2}. \tag{3}$$

Joly and Le Calvé (1995) show this triadic distance is equal to the square root of the inertia, the sum of squared distances of each of the three points towards their center of gravity. In a one dimensional representation the triadic distance is then equal to the standard deviation of the three points. In a multidimensional representation we can consider the triadic distance as a natural generalization of the standard deviation, and so the squared triadic distance as a natural generalization of the variance of the three points.

For a further development later, it is good to note here that Heiser and Bennani (1997) showed this model (2) can be rewritten as

$$\begin{aligned} d_{ijk}^2(\mathbf{X}) &= tr\mathbf{X}^T\mathbf{A}_{ij}\mathbf{X} + tr\mathbf{X}^T\mathbf{A}_{jk}\mathbf{X} + tr\mathbf{X}^T\mathbf{A}_{ik}\mathbf{X} \\ &= tr\mathbf{X}^T\mathbf{A}_{ijk}\mathbf{X}, \end{aligned} \tag{4}$$

where tr denotes *trace*, the sum of the diagonal elements of a matrix, $\mathbf{A}_{ij} = (\mathbf{e}_i - \mathbf{e}_j)(\mathbf{e}_i - \mathbf{e}_j)^T$, and $\mathbf{e}_i$ is the i-th column of a identity matrix of order K.

For the Generalized Euclidean Model we give an example of a graphical representation in Figure 1. In this figure we see 5 points a, b, c, d and e. The triadic distance is defined as the square root of the sum of squared dyadic distances. Comparing some distances we find that $d_{abe} < d_{ade}$ since $d_{ab} < d_{ad}$ and the other dyadic distances are equal; $d_{abd} < d_{acd}$, since $\sqrt{(2^2+2^2+4^2)} < \sqrt{(3^2+1^2+4^2)}$.

De Rooij and Heiser (2000) extend the work of Heiser and Bennani to the unfolding situation, and discuss restrictions on triadic unfolding models to visualize trends in longitudinal studies. For the triadic unfolding model any $I \times J \times K$-matrix with proximity data can be used. The restricted unfolding

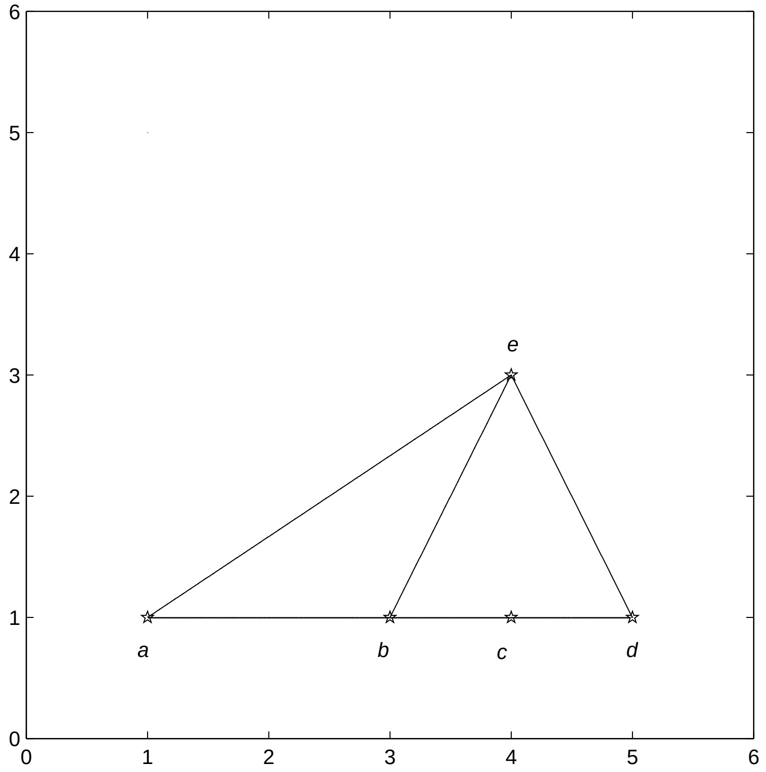

Figure 1: An configuration with triadic distances.

models, proposed by De Rooij and Heiser, can only be applied to $K \times K \times K$-tables. In the present paper we will study the (restricted) triadic unfolding models in more detail.

In the triadic unfolding model we estimate a coordinate matrix for each way. For three-way square tables the triadic unfolding model allows for three-way asymmetry. Still the triadic distance is defined as the square root of the sum of dyadic distances, but the dyadic distances are defined by

$$d_{ij}(\mathbf{X}; \mathbf{Y}) = \left[\sum_m (x_{im} - y_{jm})^2 \right]^{1/2}. \tag{5}$$

The squared triadic unfolding distance is then defined as

$$d^2_{ijk}(\mathbf{X}; \mathbf{Y}; \mathbf{Z}) = d^2_{ij}(\mathbf{X}; \mathbf{Y}) + d^2_{jk}(\mathbf{Y}; \mathbf{Z}) + d^2_{ik}(\mathbf{X}; \mathbf{Z}). \tag{6}$$

In the triadic unfolding model the subscript i is attached to the first way, and thus to the coordinates x; j is attached to the second way, and thus to the coordinates y; k is attached to the coordinates for the third way, z. The triadic unfolding model has the same interpretation as shown in Figure 1, but in the triadic unfolding model distances between points representing categories of one way are not related to observations.

We can rewrite the triadic unfolding model also in matrix terms. Therefore, first define the matrix $\mathbf{S}$ with $\mathbf{X}$, $\mathbf{Y}$, and $\mathbf{Z}$ concatenated vertically, that is, $\mathbf{S}$ is defined as

$$\mathbf{S} = \begin{pmatrix} \mathbf{X} \\ \mathbf{Y} \\ \mathbf{Z} \end{pmatrix}. \tag{7}$$

Equation 6 can then be rewritten as

$$\begin{aligned} d^2_{ijk}(\mathbf{X}; \mathbf{Y}; \mathbf{Z}) &= tr\mathbf{S}^T\mathbf{A}_{ij}\mathbf{S} + tr\mathbf{S}^T\mathbf{A}_{jk}\mathbf{S} + tr\mathbf{S}^T\mathbf{A}_{ik}\mathbf{S} \\ &= tr\mathbf{S}^T\mathbf{A}_{ijk}\mathbf{S}, \end{aligned} \tag{8}$$

where in this case $\mathbf{A}_{ij} = (\mathbf{e}_i - \mathbf{e}_{I+j})(\mathbf{e}_i - \mathbf{e}_{I+j})^T$; $\mathbf{A}_{ik} = (\mathbf{e}_i - \mathbf{e}_{I+J+k})(\mathbf{e}_i - \mathbf{e}_{I+J+k})^T$; $\mathbf{A}_{jk} = (\mathbf{e}_{I+j} - \mathbf{e}_{I+J+k})(\mathbf{e}_{I+j} - \mathbf{e}_{I+J+k})^T$ and $\mathbf{e}_i$ is the i-th column of a identity matrix of order $I + J + K$.

The generalized slide vector models proposed by De Rooij and Heiser are restricted triadic unfolding models. These restricted models can only be applied to three-way tables where each way refers to the same variable. So the three-way matrix needs to be square. The imposed restrictions are $z_{jm} = y_{jm} - v_m$, and $y_{jm} = x_{jm} - u_m$, for the slide-2 model. Consequently, $z_{jm} = x_{jm} - u_m - v_m$. The square slide-2 distance model is then defined by

$$d^2_{ijk}(\mathbf{X}; \mathbf{u}; \mathbf{v}) = d^2_{ij}(\mathbf{X}; \mathbf{u}) + d^2_{jk}(\mathbf{X}; \mathbf{v}) + d^2_{ik}(\mathbf{X}; \mathbf{u}; \mathbf{v}), \tag{9}$$

where $d_{ij}^2(\mathbf{X};\mathbf{u}) = \sum_m (x_{im} - x_{jm} + u_m)^2$, the slide vector model as defined by Zielman and Heiser (1993). A further restriction is $v_m = u_m$, for the slide-1 model. We obtain the triadic distance model by imposing the restriction $v_m = u_m = 0$.

In matrix terms the constraints can be written as $\mathbf{S} = \mathbf{ER}$, where $\mathbf{E}$ is a known design matrix, and $\mathbf{R}$ is a matrix with the coordinates $\mathbf{X}$ and possibly the slide vectors $\mathbf{u}$ and $\mathbf{v}$ concatenated vertically. Define the three arrays $\mathbf{E}$ for the slide-2 model, the slide-1 model and the triadic distance model, respectively, with identity matrices of order K ($\mathbf{I}$), and $K \times 1$ vectors with ones ($\mathbf{1}$) and zeros ($\mathbf{0}$):

$$\mathbf{E}_2 = \begin{pmatrix} \mathbf{I} & \mathbf{1} & \mathbf{0} \\ \mathbf{I} & \mathbf{0} & \mathbf{0} \\ \mathbf{I} & \mathbf{0} & -\mathbf{1} \end{pmatrix}, \quad \mathbf{E}_1 = \begin{pmatrix} \mathbf{I} & \mathbf{1} \\ \mathbf{I} & \mathbf{0} \\ \mathbf{I} & -\mathbf{1} \end{pmatrix}, \quad \mathbf{E}_0 = \begin{pmatrix} \mathbf{I} \\ \mathbf{I} \\ \mathbf{I} \end{pmatrix}. \tag{10}$$

We will make use of these design matrices in the next section to further analyze the generalized slide vector models.

3. Application to contingency tables

De Rooij and Heiser (2000) apply their triadic distance model to a three-way contingency table, assuming that the frequencies are a measure of similarity. Here we will also apply the models to contingency tables, but in the present paper we will assume a Multinomial sampling scheme for the observed frequencies. In general, the kernel of the log-likelihood function for a model under Multinomial sampling can be written

$$\mathcal{L} = \sum_{ijk} f_{ijk} \log(\pi_{ijk}), \tag{11}$$

where π_{ijk} are the expected probabilities under a specified model. The f_{ijk} are the observed frequencies. The simple model we will study is that the expected probabilities are given by

$$\pi_{ijk} = \exp(-d_{ijk}^2), \tag{12}$$

that is the expected probabilities are related to distances in Euclidean space by the Gaussian transform. The larger the distance the smaller expected probability; the smaller the distance the larger the expected probability. If a combination of categories often occur the categories are close in Euclidean space, which is in line with our assumption that frequencies are a measure of similarity.

If we insert our model into the likelihood function we obtain

$$\mathcal{L} = -\sum_{ijk} f_{ijk} d_{ijk}^2. \tag{13}$$

We will use different distance models and develop the likelihood function. This will give a more detailed view on triadic distance models as developed by now.

3.1 Triadic unfolding models

For unfolding models we saw in the previous section that the distance can be written as

$$d^2_{ijk}(\mathbf{X};\mathbf{Y};\mathbf{Z}) \quad = \quad tr\mathbf{S}^T\mathbf{A}_{ijk}\mathbf{S}. \qquad (14)$$

Inserting this definition in the log-likelihood function (13), and developing we obtain

$$\begin{aligned} \mathcal{L}_u &= -\sum_{ijk} f_{ijk} \times tr\mathbf{S}^T\mathbf{A}_{ijk}\mathbf{S} \\ &= tr\mathbf{S}^T\mathbf{C}_u\mathbf{S}, \end{aligned} \qquad (15)$$

where

$$\mathbf{C}_u \quad = \quad \begin{pmatrix} -2\mathbf{F}_i & \mathbf{F}_{ij} & \mathbf{F}_{ik} \\ \mathbf{F}_{ij}^T & -2\mathbf{F}_j & \mathbf{F}_{jk} \\ \mathbf{F}_{ik}^T & \mathbf{F}_{jk}^T & -2\mathbf{F}_k \end{pmatrix}. \qquad (16)$$

Here $\mathbf{F}_{ij}$ is a two-way matrix obtained by summing the three-way table over the third way k, and similarly for the other marginal arrays, $\mathbf{F}_i$ denotes the diagonal matrix with elements f_{i++}.

The matrix $\mathbf{C}_u$ has the same form as the Burt matrix that is decomposed in Multiple Correspondence Analysis (MCA) (Greenacre, 1984; Gifi, 1990). For the triadic unfolding model the diagonal blocks have matrices defined by minus two times the univariate margin, where in MCA the margin itself is used. However, there has been a discussion lately about these diagonal blocks (Greenacre, 1988; Boik, 1996; Tateneni and Brown, 2000). A new method called Joint Correspondence Analysis (JCA) is devised to reduce the influence of the diagonal blocks on the result of MCA. Further research could be done in the field of correspondence analysis whether the replacement of the diagonal blocks as proposed here would add to the understanding of the method. At least then the interpretation of MCA can be done in terms of triadic Euclidean distances.

3.2 Triadic slide vector models

The triadic unfolding model could be written as a trace function. The slide vector models and the symmetric model are restricted unfolding models, where the restriction has the form $\mathbf{S} = \mathbf{ER}$. We will use these restrictions in our development of the likelihood function.

Inserting $\mathbf{S} = \mathbf{E}_2\mathbf{R}$ in the likelihood function for the triadic unfolding model (15) we obtain the likelihood function for the slide-2 model, that is

$$\begin{aligned} \mathcal{L}_{sv2} &= tr\mathbf{R}^T\mathbf{E}_2^T\mathbf{C}_u\mathbf{E}_2\mathbf{R} \\ &= tr\mathbf{R}^T\mathbf{C}_{sv2}\mathbf{R}, \end{aligned} \tag{17}$$

where $\mathbf{C}_{sv2}$ has the form

$$\mathbf{C}_{sv2} = \begin{pmatrix} \mathbf{C}_s & \mathbf{n}_1 & \mathbf{n}_2 \\ \mathbf{n}_1^T & -2f_{+++} & -f_{+++} \\ \mathbf{n}_2^T & -f_{+++} & -2f_{+++} \end{pmatrix}, \tag{18}$$

in which the vector $\mathbf{n}_1$ has elements $\{n_i^1\}$ defined by $n_i^1 = f_{++i}+f_{+i+}-2f_{i++}$, and the vector $\mathbf{n}_2$ has elements $\{n_i^2\}$ defined by $n_i^2 = 2f_{++i} - f_{+i+} - f_{i++}$. The matrix $\mathbf{C}_s$ has elements $\{c_{ij}^s\}$ defined by $c_{ij}^s = \sum_k g_{ijk}$ if $i \neq j$, else $c_{ij}^s = \sum_j c_{ij}^s$, and the g_{ijk} are given by $g_{ijk} = \frac{1}{6}(f_{ijk}+f_{ikj}+f_{jik}+f_{jki}+f_{kij}+f_{kji})$.

Using the same steps as above, the likelihood function for the slide-1 model is given by

$$\begin{aligned} \mathcal{L}_{sv1} &= tr\mathbf{R}^T\mathbf{E}_1^T\mathbf{C}_u\mathbf{E}_1\mathbf{R} \\ &= tr\mathbf{R}^T\mathbf{C}_{sv1}\mathbf{R}, \end{aligned} \tag{19}$$

where $\mathbf{C}_{sv1}$ has the form

$$\mathbf{C}_{sv1} = \begin{pmatrix} \mathbf{C}_s & \mathbf{n}_3 \\ \mathbf{n}_3^T & -6f_{+++} \end{pmatrix}, \tag{20}$$

in which $\mathbf{n}_3$ has elements $\{n_i^3\}$ defined by $n_i^3 = 3(f_{++i} - f_{i++})$.

For the symmetric generalized Euclidean model the design matrix $\mathbf{E}_0$ can be used and the likelihood function obtains the following form

$$\begin{aligned} \mathcal{L}_{s} &= tr\mathbf{R}^T\mathbf{E}_0^T\mathbf{C}_u\mathbf{E}_0\mathbf{R} \\ &= tr\mathbf{R}^T\mathbf{C}_s\mathbf{R}. \end{aligned} \tag{21}$$

The matrix $\mathbf{C}_s$ is the same as above in the likelihood functions for the triadic slide vector models. The matrix $\mathbf{R}$ in this case is equal to the coordinate matrix $\mathbf{X}$.

Both the slide-1 and the slide-2 model will fit the same coordinate matrix as the symmetric triadic distance model, corresponding to the matrix $\mathbf{C}_s$. The slide-1 model represents in addition the difference between the third margin and the first margin. The slide-2 model represents two differences: (1) The difference of the first margin compared with the second and the third margin; (2) The difference of the third margin compared with the first and the second margin.

We can compare these models to quasi-symmetry models for three-way tables. The quasi-symmetry model for three-way tables is given by

$$\log(\pi_{ijk}) = \lambda + \lambda_i^R + \lambda_j^C + \lambda_k^P + \lambda_{ijk}, \tag{22}$$

where $\lambda_{ijk} = \lambda_{ikj} = \lambda_{jik} = \lambda_{jki} = \lambda_{kij} = \lambda_{kji}$, i.e., the interaction term is three-way symmetric. In the triadic slide vector models the symmetric part is modeled by the distances between the points in Euclidean space given in the coordinate matrix $\mathbf{X}$. The main effect terms λ_i^R, λ_j^C, and λ_k^P represent the occurences of each of the categories. In the generalized slide vector models the differences of these main effect parameters are represented by vectors, which are attached to the dimensions of the Euclidean space. This does give us a nice representation of, for example, the trends in longitudinal research. For the symmetric generalized Euclidean model we find no marginal differences, i.e., the model is equal to the model of three-way symmetry with metric constraints.

4. Conclusions

We studied triadic distance models under a Multinomial sampling scheme. For the triadic unfolding model we found an interesting relationship with MCA. In MCA the distances between points are defined via the underlying subject points. In triadic unfolding models we have a direct distance definition between the three points. In both models no three-way information is represented (see De Rooij, 2001, submitted), which is clear from the Burt matrix and the matrix obtained in (16). This might be a disadvantage of both procedures in case three way relations are of specific interest. However, many multivariate data anayses techniques only consider bivariate relationships and these bivariate relationship are often more interesting compared to the trivariate or higher order relationships. An advantage of looking only at bivariate relationships is that large tables can be handled without many complications. In log-linear analysis, for example, one should always be careful analyzing large sparse tables.

Distance models for three-way two-mode data, as the well known INDSCAL model (Carroll and Chang, 1970) do represent three-way relationships (see De Rooij, 2001). In the INDSCAL model, however, only one of the three bivariate relationships has a distance representation. The other two bivariate relationships are hard to grasp from the results of the model. Moreover, the INDSCAL model has a very specific interpretation, which is only nice when there is a specific interest in differences between individuals or groups of individuals. Many three-way contingency tables do not have that nature, and in that case our triadic distance models have a more natural interpretation. It would be interesting to combine the two approaches, that is a topic of current research.

References:

Boik, R. J. (1996), "An efficient algorithm for joint correspondence analysis," *Psychometrika,* **61**, 255-269.

Carroll, J. D., and Chang, J. J. (1970), “Analysis of individual differences in multidimensional scaling via an N-way generalization of ‘Eckart-Young’ decomposition,” *Psychometrika,* **35**, 283-319.

Cox, T. F., Cox, M. A., and Branco, J. A. (1991), “Multidimensional scaling for n-tuples,” *British Journal of Mathematical and Statistical Psychology ,* **44**, 195-206.

Daws, J. T., (1996), “The analysis of free-sorting data: Beyond pairwise cooccurences,” *Journal of Classification,* **13**, 57-80.

De Rooij, M. (2001, submitted), “Distance models for three-way tables and three-way information: a theoretical note,”

De Rooij, M., and Heiser, W. J. (2000), “Triadic distance models for the analysis of asymmetric three-way proximity data,” *British Journal of Mathematical and Statistical Psychology ,* **53**, 99-119.

Gifi, A. (1990), *Nonlinear multivariate analysis.* Chichester, England: Wiley.

Greenacre, M. J. (1984), *Theory and applications of correspondence analysis.* New York: Academic Press.

Greenacre, M. J. (1988), “Correspondence analysis of multivariate categorical data by weighted least squares,” *Biometrika,* **75**, 457-467.

Hayashi, C. (1972), “Two dimensional quantifications based on a measure of dissimilarity among three elements,” *Annals of the Institute of Statistical Mathematics,* **25**, 251-257.

Heiser, W. J., and Bennani, M. (1997), “Triadic distance models: Axiomatization and least squares representation,” *Journal of Mathematical Psychology ,* **41**, 189-206.

Joly, S., and Le Calvé, G. (1995), “Three-way distances,” *Journal of Classification,* **12**, 191-205.

Pan, G., and Harris, D. P. (1991), “A new multidimensional scaling technique based upon association of triple objects pijk and its application to the analysis of geochemical data,” *Mathematical Geology,* **23**, 861-886.

Tateneni, K., and Browne, M. W. (2000), “A noniterative method of joint correspondence analysis,” *Psychometrika,* **65**, 157-165.

Zielman, B., and Heiser, W. J. (1993), “The analysis of asymmetry by a slide-vector,” *Psychometrika,* **58**, 101-114.