Chemometrics and Intelligent Laboratory Systems, 7 (1989) 95–100
Elsevier Science Publishers B.V., Amsterdam — Printed in The Netherlands

Aspects of the Analysis of Three-Way Data

LARS STÅHLE

Department of Pharmacology, Karolinska Institute, Box 60 400, S-10401 Stockholm (Sweden)

(Received 6 October 1988; accepted 15 April 1989)

ABSTRACT

Ståhle, L., 1989. Aspects of the analysis of three-way data. *Chemometrics and Intelligent Laboratory Systems*, 7: 95–100.

An algorithm is developed for linear three-way decomposition (LTD) of three-way tables for the case of one dependent block of variables and one independent, predictor block. The algorithm is a generalization of the two-block partial least squares (PLS) algorithm. Using simulation and real data LTD is compared with multi-way PLS with and without decomposition of the weight matrix. LTD and PLS with a one-dimensional decomposition of the weight matrix gives essentially the same results. By means of the cross-validation criterion it is shown that, in practice, optimal prediction may be obtained with a model that is neither completely trilinear nor bilinear. The multi-way PLS method with decomposition of the weight matrix to give the minimum cross-validation prediction error is preferred.

INTRODUCTION

The increasing complexity of chemical data has recently stimulated the development of data-analytic methods that can utilize more information than conventional univariate and multivariate statistical methods [1,2]. A general method for the analysis of multi-way data by bilinear decomposition has been proposed [1]. There is some debate concerning the applicability of various approaches to analysis of e.g. 3-way data [1,3,4]. However, this discussion has been limited to decomposition of one-block data. To the best of our knowledge there is no method available for trilinear 3-way decomposition of 2-block data, i.e. data with one block of predictor variables and one block of predicted responses (Fig. 1).

In order to make possible comparisons between different approaches to the analysis of 3-way tables, an algorithm is developed for relating two blocks of variables where each block may be organized in up to three ways. The algorithm is a generalization of the 2-block partial least squares

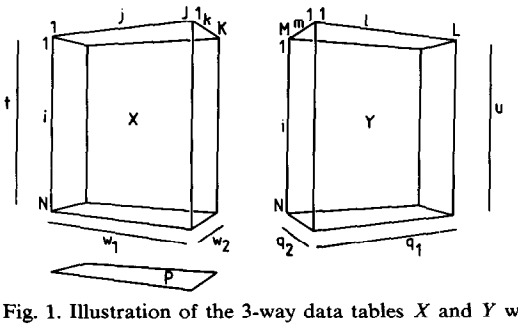

Fig. 1. Illustration of the 3-way data tables X and Y with the decomposition vectors for rows (t and u), columns (w_1 and q_1) and sheet (w_2 and q_2). In addition there is a matrix **P** used to model X and calculate residuals. The indices used in LTD are indicated.

0169-7439/89/$03.50

(PLS) algorithm [5,6] and it will be referred to as the linear three-way decomposition (LTD) algorithm. Numerical examples from the author's pharmacological research as well as simulation data from the literature [1] will be used as an illustration of LTD and for comparison with the multi-way decomposition of Wold et al. [1].

METHODS

Notational preliminaries

The data are subdivided into two blocks (Fig. 1), a dependent block (Y) and an independent block (X). Each block is organized as a 3-way table where the rows are objects. The columns and the sheet are variables classified in two different ways (e.g. columns, or the first way, as wavelengths and sheet; or the second way, as retention times in a diode array chromatogram). The goal of the methods is to find a bilinear (PLS) or trilinear (LTD) decomposition of X that can be used to predict Y. The prediction of Y is carried via latent variables of X and Y. In LTD X and Y are decomposed respectively into vectors corresponding to rows ($\boldsymbol{t}$ and $\boldsymbol{u}$: boldface italic lower case letters indicate vectors), vectors corresponding to columns ($\boldsymbol{w}_1$ and $\boldsymbol{q}_1$, where the index 1 denotes column, or first way, in the 3-way table) and vectors corresponding to sheet ($\boldsymbol{w}_2$ and $\boldsymbol{q}_2$, where 2 denotes sheet, or the second way, in the 3-way table). In PLS there is in the first phase no distinction between columns and sheet and the vectors $\boldsymbol{w}$ and $\boldsymbol{q}$ are used for the unfolded X and Y blocks (Fig. 2). The number of rows is N with index i, the number of columns and sheet in the X block is J (index j) and K (index k) and the number of columns and sheet in Y are L (index l) and M (index m) respectively. The number of dimensions is A (index a).

Multi-way PLS method

The multi-way PLS method is described in detail elsewhere [1]. A short account is given here assuming that the reader is familiar with the PLS-method which has been described by others in extenso [5–10].

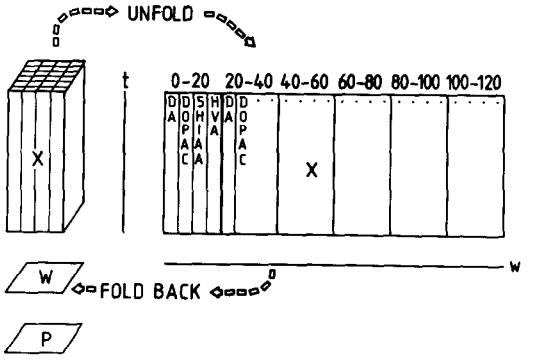

Fig. 2. Unfolding a 3-way table into a 2-way table to prepare for the PLS analysis. The unfolded weights may be folded back to form a 2-way matrix that can be decomposed by singular-value decomposition.

The PLS-algorithm proceeds as follows:

A. Unfold the 3-way tables X ($N \times J \times K$) and Y ($N \times L \times M$) into the usual two-way matrices $\mathbf{X}$ ($N \times JK$) and $\mathbf{Y}$ ($N \times LM$). Center and scale as required.

B. Make an initial guess of $\boldsymbol{u}$, e.g. as the first column of $\mathbf{Y}$.

C. Proceed with the usual PLS algorithm
   1. $\boldsymbol{w}' = \mathbf{X}'\boldsymbol{u}/\boldsymbol{u}'\boldsymbol{u}$
   2. Optionally $\boldsymbol{w}$ may be folded back to form a matrix $\boldsymbol{W}$ ($J \times K$) which is subjected to principal component decomposition by e.g. the NIPALS method [11] into a smaller number of components. The decomposition of $\mathbf{W}$ thus obtained is then used as a model of $\mathbf{W}$. Finally $\mathbf{W}$ is unfolded to $\boldsymbol{w}$.
   3. $\| \boldsymbol{w} \| = 1$
   4. $\boldsymbol{t} = \mathbf{X}\boldsymbol{w}'/\boldsymbol{w}\boldsymbol{w}'$
   5. $\boldsymbol{q}' = \mathbf{Y}'\boldsymbol{t}/\boldsymbol{t}'\boldsymbol{t}$
   6. Optionally $\boldsymbol{q}$ may be folded into a matrix $\mathbf{Q}$ and then decomposed in the same way as $\mathbf{W}$. In case $\mathbf{Q}$ and $\mathbf{W}$ have sheet or column directions in common it is possible to take $\mathbf{W}$ as the predictor block and $\mathbf{Q}$ as the predicted block and run the usual PLS algorithm with the common direction as the latent variables $\boldsymbol{t}_w$ and $\boldsymbol{u}_q$ (Wold, personal communication). In this case step 2 is not run.

7. $\|\boldsymbol{q}\| = 1$
8. $\boldsymbol{u} = \mathbf{Y}\boldsymbol{q}'/\boldsymbol{q}\boldsymbol{q}'$
9. repeat from 1, until convergence as tested on $\boldsymbol{u}$.
10. $b = \boldsymbol{u}'\boldsymbol{t}/\boldsymbol{t}'\boldsymbol{t}$ where b is the regression coefficient between $\boldsymbol{u}$ and $\boldsymbol{t}$.

D. The loadings of $\mathbf{X}$ are calculated as $\boldsymbol{p}' = \mathbf{X}'\boldsymbol{t}/\boldsymbol{t}'\boldsymbol{t}$

E. Appropriate scaling of $\boldsymbol{p}$ (unit length), $\boldsymbol{t}$ (multiply by $\|\boldsymbol{p}\|$) and $\boldsymbol{w}$ (divide by $\|\boldsymbol{p}\|$) may be performed and the residuals are calculated as $\mathbf{X} = \mathbf{X} - \boldsymbol{t}\boldsymbol{p}$ and $\mathbf{Y} = \mathbf{Y} - b\boldsymbol{t}\boldsymbol{q}$ before the next component is extracted. In case step 2 is performed the score and loading vectors of $\mathbf{W}$ may, for consistency, be scaled by division by $\sqrt{\|\boldsymbol{p}\|}$ when $\boldsymbol{p}$ is scaled to unit length.

Linear 3-way decomposition (LTD)

This method is defined as an algorithm and it is a generalization of the usual two-block PLS algorithm [5,6]. Below, the steps in the algorithm are given, with motivations at each step where necessary.

1. Make an initial guess for $\boldsymbol{u}$, e.g. as the first column in the first sheet of $\mathbf{Y}$. In addition, make an initial guess for $\boldsymbol{w}_1$, $\boldsymbol{w}_2$, $\boldsymbol{q}_1$ and $\boldsymbol{q}_2$, e.g. $1/\sqrt{J}$ for $\boldsymbol{w}_1$ and correspondingly for the others.

Comment: starting guesses of the $\boldsymbol{w}$ and $\boldsymbol{q}$ vectors are necessary in LTD since, in steps (2), (3), (5) and (6) below, a value from the previous iteration is necessary.

2. $w_{1aj} = \sum_{i=1}^{N}\sum_{k=1}^{K} x_{ijk}u_{ia}w_{2ak}$; give $\boldsymbol{w}_1$ length 1.0

3. $w_{2ak} = \sum_{i=1}^{N}\sum_{j=1}^{J} x_{ijk}u_{ia}w_{1aj}$; give $\boldsymbol{w}_2$ length 1.0

Comment: on the right-hand side of (2) and (3) the values of the elements in $\boldsymbol{w}_1$ and $\boldsymbol{w}_2$ from the previous iteration are used.

4. $t_{ia} = \sum_{j=1}^{J}\sum_{k=1}^{K} x_{ijk}w_{1aj}w_{2ak}$

5. $q_{1al} = \sum_{i=1}^{N}\sum_{m=1}^{M} y_{ilm}t_{ia}q_{2am}$; give $\boldsymbol{q}_1$ length 1.0

6. $q_{2am} = \sum_{i=1}^{N}\sum_{l=1}^{L} y_{ilm}t_{ia}q_{1al}$; give $\boldsymbol{q}_2$ length 1.0

Comment: in steps (5) and (6) $\boldsymbol{q}_1$ and $\boldsymbol{q}_2$ are taken from the previous iteration as with $\boldsymbol{w}_1$ and $\boldsymbol{w}_2$ above.

7. $u_{ia} = \sum_{l=1}^{L}\sum_{m=1}^{M} y_{ilm}q_{1al}q_{2am}$

8. Check for convergence, if two successive iterations give approximately the same values for all vectors the procedure has converged. In case of convergence go to step (9), otherwise go to step (2).

Comment: it is not sufficient to check convergence at $\boldsymbol{u}$ alone since e.g. $\boldsymbol{w}_1$ and $\boldsymbol{w}_2$ can be modified in spite of the fact that $\boldsymbol{u}$ has converged. This is of practical importance when Y has only one column and one sheet.

9. $b = \sum_{i=1}^{N} t_{ia}u_{ia} \Big/ \sum_{i=1}^{N} t_{ia}^2$

Comment: b is the inner relation in the same way as in PLS i.e. the regression coefficient between $\boldsymbol{t}$ and $\boldsymbol{u}$.

Comment: in order to get orthogonal $\boldsymbol{t}$-vectors in successive dimensions it is necessary to use a vector (or a matrix) with $J \times K$ elements to calculate the residuals in exactly the same way as in the multi-way PLS algorithm. Denoting the matrix $\mathbf{P}$ we write

10. $p_{ajk} = \sum_{i=1}^{N} x_{ijk}t_{ia} \Big/ \sum_{i=1}^{N} t_{ia}^2$

11. $\Theta = \|\mathbf{P}\|$: give $\mathbf{P}$ length 1.0

Comment: It is necessary to divide by $\boldsymbol{t}'\boldsymbol{t}$ to preserve the correct scale when $\mathbf{P}$ is used to model X. $\|\mathbf{P}\|$ is the Frobenius norm of a matrix which is the usual vector norm when $\mathbf{P}$ is unfolded.

12. $t_{ia} = t_{ia}\Theta$

13. $w_{1aj} = w_{1aj}/\sqrt{\Theta}$

14. $w_{2ak} = w_{2ak}/\sqrt{\Theta}$

Comment: The steps (14)–(16) are necessary to give correct prediction of Y, but only if step (11) is carried out. It is possible to have w_1 and w_2 with unit length instead.

15. $x_{ijk} = x_{ijk} - t_{ia} p_{ajk}$

Comment: Because of the isomorphism between matrices and vectors with the same dimensionality this way of calculating residuals will make the successive t-vectors orthogonal regardless of the choice of weights as pointed out by Höskuldsson [10].

16. $y_{ilm} = y_{ilm} - bt_{ia} q_{1al} q_{2am}$

17. If the last component was insignificant the procedure may be stopped. Otherwise the next component can be calculated with the residuals calculated in (15) and (16) and incrementing the dimension index a by 1.

This completes the basic algorithm. Additional facilities such as cross-validation for testing the number of components and missing value handling can be made in the same way as in the usual PLS algorithm [5,6].

RESULTS

The algorithms were tested on real as well as simulated data. The experimental data are taken from research in our laboratory dealing with the effect of drugs on the composition of the extracellular fluid in the rat brain. Four neurochemicals, dopamine (DA), 3,4-dihydroxyphenylacetic acid (DOPAC), homovanillic acid (HVA) and 5-hydroxyindoleacetic acid (5HIAA) were measured in samples of the extracellular fluid. The sampling was accomplished by microdialysis, as has been described elsewhere [12]. Samples were collected in 20-min fractions and a number of samples (6 and 14 in the present examples) were analyzed following administration of drugs. The drugs given were amphetamine and α-methyl-p-tyrosine.

Simulation data

The simulation data are taken from Wold et al. [1] and are given in Table 1 together with the results from LTD and PLS. PLS was applied with and without decomposition of the weight matrix **W**. With decomposition of **W**, PLS extracts the same vectors as LTD in the first dimension while the complete bilinear decomposition by PLS gives a slightly different result. Observe that the data are analysed without centering and scaling.

TABLE 1

Simulation data from Wold et al. [1]

The data set consists of three rows, two columns and two sheets in X and two columns in Y. The solutions are calculated for unscaled and uncentered data. The results from the PLS are taken from Wold et al. [1] but the scaling of w_1 and w_2 is adjusted to unit length for each. The 1d indicates results obtained from PLS with one-dimensional decomposition of **W**.

Raw data

x_{i11}	x_{i12}	x_{i21}	x_{i22}	y_{i11}	y_{i12}
0.424264	0.565685	0.565685	0.424264	1.0	1.0
0.565685	0.424264	0.424264	0.565685	2.0	1.5
0.707101	0.707101	0.707101	0.707101	3.0	2.0

Analysis by LTD

w_{111}	w_{212}	w_{111}	w_{212}	q_{111}	q_{112}
0.707	0.707	0.707	0.707	0.806	0.592

t_{i1}
0.990
0.990
1.414

Analysis by PLS

w_{111}	w_{212}	w_{111}	w_{212}	q_{111}	q_{112}
0.720	0.694	0.694	0.720	0.806	0.592
1d 0.707	0.707	0.707	0.707	0.806	0.592

t_{i1}	1d
0.987	0.990
0.992	0.990
1.414	1.414

Effect of amphetamine

In this experiment there was one group of control animals receiving saline and one group receiving amphetamine. The method for analyzing such two-group data with PLS described previously [13] was applied, i.e. the measured variables were collected in X and a one-dimensional Y was constructed with $y_i = 0$ for controls and $y_i = 1$ for

TABLE 2

Result from the analysis of the amphetamine experiment

Treatment	t_{i1}	Variable	w_{11j}	Time	w_{21k}
Analysis by LTD					
ctrl	−3.982	DA	0.465	0– 20	0.413
ctrl	−3.899	DOPAC	−0.625	20– 40	0.375
ctrl	−3.579	5HIAA	−0.012	40– 60	0.411
ctrl	−3.317	HVA	−0.611	60– 80	0.406
ctrl	−3.475			80–100	0.415
amph	2.811			100–120	0.405
amph	3.842				
amph	3.472				
amph	5.354				
amph	2.772				
Analysis by PLS					
ctrl	−3.982	DA	0.465	0– 20	0.413
ctrl	−3.899	DOPAC	−0.625	20– 40	0.375
ctrl	−3.579	5HIAA	−0.012	40– 60	0.411
ctrl	−3.317	HVA	−0.611	60– 80	0.406
ctrl	−3.475			80–100	0.415
amph	2.811			100–120	0.405
amph	3.842				
amph	3.472				
amph	5.354				
amph	2.772				

amphetamine treated rats. In X there were four columns (DA, DOPAC, HVA and 5HIAA) and six sheet (the six fractions consecutive in time) and the rows correspond to the experimental objects (there were 5 rats in each group, hence there were 10 rows in X). All data were normalized to zero mean and unit variance before subjecting them to the PLS and LTD algorithms.

Two significant components were found by both LTD and PLS. The extracted vectors are virtually identical in PLS with a one-dimensional decomposition of **W** and LTD (the first dimensions are given in Table 2). PLS with decomposition of **W**

TABLE 3

Cross-validation/standard deviation ratio for decomposition of **W** into the indicated number of components

Dimension	Amphetamine	α-Methyl-*p*-tyrosine
1	0.2277	0.5560
2	0.2423	0.5256
3	0.2432	0.5280
4	0.2432	0.5276

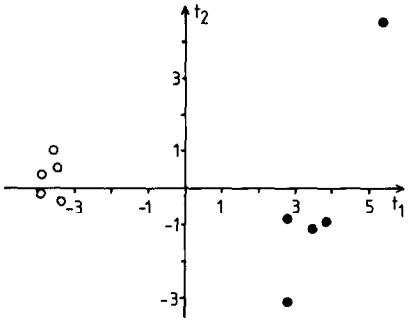

Fig. 3. Result from the analysis of the amphetamine example. Scatterplot of the scores of the two first components extracted by LTD (or PLS with a one-dimensional decomposition of the weight matrix).

into 1, 2 or 3 dimensions or the complete **W** reveals that the cross-validation shows a minimum at the first dimension (see Table 3). Plots of the scores of the first two components extracted by LTD are given in Fig. 3.

The other set of real data is not shown in detail. The most interesting finding from the point of view of the principle of the method was that decomposition of **W** shows a minimum at 2 dimensions as shown in Table 3.

DISCUSSION

The present results suggests that the LTD algorithm presented in this paper has the intended convergence properties. In the limited number of practical cases tested so far convergence was attained in the first 1–3 components. In the amphetamine example the variance in X explained by LTD was similar to that explained by PLS which is expected in view of the fact that decomposition of **W** to one dimension produced the smallest prediction error (Table 3).

In comparison with analysis of one-block 3-way tables, the LTD algorithm corresponds to one of the simplifications of the Tucker 3-mode factor analysis described by Geladi [4], in which the Kronecker product of three vectors, one for each way, is taken as the model of a 3-way table. However, LTD gives orthogonal score vectors due to the bilinear step in the calculation of the loading matrix **P** which was used to find the residuals.

Decomposition of **W** in PLS did, in the two cases tested, reduce the prediction error (Table 3) as measured by the cross-validation/standard deviation ratio [6,14]. In the amphetamine example decomposition of **W** to one dimension gave the smallest prediction error and thus corresponds to a trilinear model. For the same set of data LTD gave virtually the same result as PLS (Table 2). However, the α-methyl-*p*-tyrosine data required two dimensions of **W** to minimize the prediction error (Table 3). This result supports the conclusion of Wold et al. [1] that non-trilinear data may be encountered sufficiently often in practice to avoid the use of strictly trilinear decompositions.

The fact that both sets of real data showed a minimum prediction error for a **W** decomposed to one or two components also shows that the complete bilinear decomposition obtained with a straightforward unfolding of X and Y and application of PLS may be too little constrained. Obviously, any constraint that can be introduced into a model may improve its prediction properties. Hence, decomposition of **W** has an effect on 3-way table analyses as compared to unfolded PLS, which is similar to the effect of PLS in comparison with multiple regression.

The interpretation of the results from the amphetamine experiment is also made easy by LTD (or PLS). It may readily be seen that amphetamine causes an increase in the levels of the neurotransmitter dopamine and a decrease in the metabolites DOPAC and HVA of DA. This effect can be detected soon after injection of the drug (see Fig. 3 and Table 2).

In conclusion, analysis of 3-way tables is best done with PLS with decomposition of the weight matrix. Cross-validation on the predicted variables (Y) is useful in determining the number of dimensions in the decomposition of the weight matrix. Pure trilinear or bilinear decomposition tends to yield suboptimal results from the prediction point of view.

ACKNOWLEDGEMENTS

The present study was supported by grants from Svenska Läkarsällskapet and the Karolinska Institutet. The generous advice of professors Svante Wold and Urban Ungerstedt is gratefully acknowledged.

REFERENCES

1 S. Wold, P. Geladi, K. Esbensen and J. Öhman, Multi-way principal components- and PLS-analysis, *Journal of Chemometrics,* 1 (1987) 41–56.
2 E. Sanchez, L.S. Ramos and B.R. Kowalski, Generalized rank annihilation method. I. Application to liquid chromatography–diode array ultraviolet detection data, *Journal of Chromatography*, 385 (1987) 151–164.
3 J. Carroll, S. Pruzansky and J. Kruskal, CANDELINC: a general approach to multidimensional analysis of many-way arrays with linear constraints on parameters, *Psychometrika*, 45 (1980) 3–24.
4 P. Geladi, Analysis of multi-way (multi-mode) data, *Chemometrics and Intelligent Laboratory Systems,* 7 (1989) 11–30.
5 S. Wold, H. Martens and H. Wold, The multivariate calibration problem in chemistry solved by the PLS method, in B. Kågström and A. Ruhe (Editors), *Matrix Pencils. Lecture Notes in Mathematics,* Springer, Berlin, 1983, pp. 286–293.
6 S. Wold, A. Ruhe, H. Wold and W.J. Dunn III, The collinearity problem in linear regression, the partial least squares (PLS) approach to generalized inverses, *SIAM Journal of Statistics and Computations,* 5 (1984) 735–743.
7 H. Wold, Soft modeling, the basic design and some extensions, in K.G. Jöreskog and H. Wold (Editors), *Systems under Indirect Observation,* Vol. II, North-Holland, Amsterdam, 1982, pp. 1–54.
8 H.A. Martens, *Multivariate Calibration,* Thesis, Technical University of Norway, Trondheim, 1985.
9 A. Lorber, L.E. Wangen and B.R. Kowalski, A theoretical foundation for the PLS algorithm, *Journal of Chemometrics,* 1 (1987) 19–32.
10 A. Höskuldsson, PLS regression methods, *Journal of Chemometrics,* 2 (1988) 211–220.
11 H. Wold, Nonlinear estimation by iterative least squares procedures, in F.N. David (Editor), *Research Papers in Statistics. Festschrift for J. Neyman,* Wiley, New York, 1966, pp. 411–444.
12 U. Ungerstedt, Measurement of neurotransmitter release by intracranial dialysis, in C.A. Marsden (Editor), *Measurement of Neurotransmitter Release in vivo*, Wiley, London, 1984, pp. 81–105.
13 L. Ståhle and S. Wold, Partial least squares analysis with cross-validation for the two-class problem: a Monte Carlo study, *Journal of Chemometrics,* 1 (1987) 185–196.
14 S. Wold, Estimation of the number of components in principal component and factor analysis by cross-validation, *Technometrics,* 8 (1978) 397–405.