ESANN'2002 proceedings - European Symposium on Artificial Neural Networks
Bruges (Belgium), 24-26 April 2002, d-side publi., ISBN 2-930307-02-1, pp. 507-512

Non-linear Canonical Correlation Analysis Using a RBF Network

Sukhbinder Kumar, Elaine B Martin and Julian Morris

Center for Process Analytics and Control Technology
University of Newcastle, Newcastle upon Tyne,
NE1 7RU, England

Abstract: A non-linear version of the multivariate statistical technique of canonical correlation analysis (CCA) is proposed through the integration of a radial basis function (RBF) network. The advantage of the RBF network is that the solution of linear CCA can be used to train the network and hence the training effort is minimal. Also the canonical variables can be extracted simultaneously. It is shown that the proposed technique can be used to extract non-linear structures inherent within a data set.

1. Introduction

Over the past decade, a number of techniques have been proposed for the extraction of non-linear features inherent within process data including the multivariate statistical technique of principal component analysis [1-4]. More recently a non-linear variant of Canonical Correlation Analysis (CCA) has been proposed [5] through the integration of a Multi-Layer Perceptron (MLP) network. A drawback of this approach is that the optimisation problem is non-linear and thus suffers from the potential problem of becoming trapped within a local minimum. Hsieh [5] addressed this issue by training an ensemble of neural networks. Although not a serious limitation of the methodology, it does require major training effort. The other limitation is that when using a MLP network, the canonical variables cannot be extracted simultaneously. This has two repercussions. First the number of MLP networks to be trained (hence the training effort) increases with the number of canonical variables and secondly since the MLP networks are trained on the residuals, the extraction of subsequent canonical variables becomes difficult because of the reduction in signal to noise ratio. In this paper an alternative method of implementing non-linear CCA using a Radial Basis Function (RBF) network is proposed.

2. Linear Canonical Correlation Analysis

Canonical Correlation Analysis (CCA) is a multivariate statistical technique that identifies a linear relationship between two sets of variables $\mathbf{x} \in \mathbf{R}^{\mathbf{m}}$ and $\mathbf{y} \in \mathbf{R}^{\mathbf{n}}$. Linear CCA seeks to find vectors $\mathbf{a} \in \mathbf{R}^{\mathbf{m}}$ and $\mathbf{b} \in \mathbf{R}^{\mathbf{n}}$ such that the linear combinations:

$$u_1 = \mathbf{a}^{\mathbf{T}}\mathbf{x} \text{ and } v_1 = \mathbf{b}^{\mathbf{T}}\mathbf{y} \tag{1}$$

ESANN'2002 proceedings - European Symposium on Artificial Neural Networks
Bruges (Belgium), 24-26 April 2002, d-side publi., ISBN 2-930307-02-1, pp. 507-512

have maximum correlation. The vectors $\mathbf{a}$ and $\mathbf{b}$ are the canonical correlation vectors and u_1 and v_1 are the canonical variables. The above problem is solved as follows. Let $\Sigma_{\mathbf{xx}}$ and $\Sigma_{\mathbf{yy}}$ be the covariance matrices of $\mathbf{x}$ and $\mathbf{y}$ respectively and $\Sigma_{\mathbf{xy}}$ be the cross covariance matrix between $\mathbf{x}$ and $\mathbf{y}$. Let matrix , $\mathbf{K}$, be defined as:

$$\mathbf{K} = \Sigma_{\mathbf{xx}}^{-1/2}\ \Sigma_{\mathbf{xy}}\ \Sigma_{\mathbf{yy}}^{-1/2} \tag{2}$$

If k is the rank of matrix $\mathbf{K}$, then by singular value decomposition, $\mathbf{K}$ can be decomposed as:

$$\mathbf{K} = (\boldsymbol{\alpha}_1, \boldsymbol{\alpha}_2 \ldots, \boldsymbol{\alpha}_k)\, \mathbf{D}\, (\boldsymbol{\beta}_1, \boldsymbol{\beta}_2, \ldots, \boldsymbol{\beta}_k)^T \tag{3}$$

where α_i and β_i are the eigenvectors of the matrices $\mathbf{KK^T}$ and $\mathbf{K^TK}$ respectively and $\mathbf{D}$ is a block diagonal matrix comprising the square root of the k non-zero eigenvalues. Letting:

$$\mathbf{a}_i = \Sigma_{\mathbf{xx}}^{-1/2}\, \alpha_i \quad \text{and} \quad \mathbf{b}_i = \Sigma_{\mathbf{yy}}^{-1/2}\, \beta_i \ \text{ for } i = 1,\ 2. \ldots k \tag{4}$$

then $\mathbf{a}_i$ and $\mathbf{b}_i$ are the (k) canonical vectors.

3. Non Linear CCA using a RBF Network

Non-linear Canonical Correlation Analysis (CCA) is similar to linear CCA except that the linear transformation applied to the variables, $\mathbf{x}$ and $\mathbf{y}$, is replaced by a non-linear transformation. In this paper a RBF network replaces the linear transformation. Non-linear CCA is performed in two stages. First the variables, $\mathbf{x}$ and $\mathbf{y}$, are projected from the higher dimensional space down onto a lower dimensional space and then the latent variables are transformed back to the original variables. The second step is termed self-consistency [6].

3.1 Mapping from Original Data Space to the Canonical Variables

The mapping of $\mathbf{x}$ and $\mathbf{y}$ to canonical variables u_1 and v_1 is from $\mathbf{R^m} \to \mathbf{R}^k$ and $\mathbf{R^n} \to \mathbf{R}^k$ respectively. For simplicity, k, the number of canonical variables, is taken to be unity. The situation where k is greater than unity is a straightforward extension of the described methodology. Given the centers $\mathbf{c}_{xi}$, $\mathbf{c}_{yi}$ and the widths $\sigma_{\mathbf{x}i}$, $\sigma_{\mathbf{y}i}$ of the radial basis functions for the two mappings, the canonical variables u_1 and v_1 can be defined as:

ESANN'2002 proceedings - European Symposium on Artificial Neural Networks
Bruges (Belgium), 24-26 April 2002, d-side publi., ISBN 2-930307-02-1, pp. 507-512

$$u_1 = \sum_{i=1}^{p} w_{xi} f_i(\mathbf{x}) = \mathbf{w_x}^T \mathbf{f} \quad \text{and} \quad v_1 = \sum_{j=1}^{q} w_{yi} g_j(\mathbf{y}) = \mathbf{w_y}^T \mathbf{g} \tag{5}$$

where $\mathbf{f} = [f_1, f_2 \, . \, . \, . \, f_p]^T$ and $\mathbf{g} = [g_1, g_2 \, . \, . \, . \, g_q]^T$ are RBF vectors and $\mathbf{w_x} = [w_{x1}, w_{x2}, \, . \, . \, . \, w_{xp}]^T$ and $\mathbf{w_y} = [w_{y1}, w_{y2}, \, . \, . \, . w_{yq}]^T$ are the weight vectors for the mapping from $\mathbf{x}$ to u_1 and $\mathbf{y}$ to v_1 respectively. Non-linear CCA then reduces to that of finding the weight vectors $\mathbf{w_x}$ and $\mathbf{w_y}$ such that there is maximum correlation between u_1 and v_1. This problem is similar to the linear case except that the vectors $\mathbf{x}$ and $\mathbf{y}$ are replaced by $\mathbf{f}$ and $\mathbf{g}$ respectively. If $\mathbf{A_{xx}}$ and $\mathbf{A_{yy}}$ denote the covariance matrices of the radial basis functions $\mathbf{f}$ and $\mathbf{g}$ respectively and $\mathbf{A_{xy}}$ is the cross covariance matrix, then similar to equations (2) and (3):

$$\mathbf{M} = \mathbf{A_{xx}}^{-1/2} \mathbf{A_{xy}} A_{\mathbf{yy}}^{-1/2} \tag{6}$$

can be decomposed using singular value decomposition:

$$\mathbf{M} = [\mathbf{p_1}, \mathbf{p_2} \, . \, . \, . \, \mathbf{p}_k] \Lambda_k [\mathbf{q}_1, \mathbf{q_2}, \, . \, . \, . \, \mathbf{q}_k]^T \tag{7}$$

The weight vectors $\mathbf{w_x}$ and $\mathbf{w_y}$ are calculated as follows:

$$\mathbf{w_x} = \mathbf{A_{xx}}^{-1/2} \mathbf{p}_1 \tag{8}$$

$$\mathbf{w_y} = \mathbf{A_{yy}}^{-1/2} \mathbf{q}_1 \tag{9}$$

In the case where more than one canonical variable is required, the weight vectors for the network for transforming the variables $\mathbf{x}$ and $\mathbf{y}$ into successive canonical variables can be obtained using vectors $\mathbf{p}_i$ and $\mathbf{q}_i$ for $i = 2,3 \, . \, . \, k$ in equations (8) and (9) respectively. Thus the canonical variables can be obtained simultaneously without solving any non-linear optimization problem. The basis function for the mapping from $\mathbf{x}$ to u_1 is chosen such that $\mathbf{y}$ is predicted from $\mathbf{x}$, that is:

$$\mathbf{y} = \sum_{i=1}^{p} \gamma_i f_i(\mathbf{x}) + \varepsilon_y$$

where ε_y is the prediction error and γ_i is a coefficient . Similarly the basis functions for the mapping from $\mathbf{y}$ to v_1 are selected so that $\mathbf{x}$ is predicted from $\mathbf{y}$ with minimum error.

There exist many techniques to adjust the centers and widths σ_{xi} and σ_{yi} of the radial basis functions [7-8]. Here the centres are determined by fitting a Gaussian

ESANN'2002 proceedings - European Symposium on Artificial Neural Networks
Bruges (Belgium), 24-26 April 2002, d-side publi., ISBN 2-930307-02-1, pp. 507-512

mixture model with circular covariances using the EM algorithm with the widths set equal to the maximum inner centre distance.

3.2 Mapping from Canonical Variables to Original Data Space

The scores u_i and v_i calculated in the first stage of the algorithm should be a good approximation of the original vectors, $\mathbf{x}$ and $\mathbf{y}$. The next stage is to apply an inverse transformation, again using a RBF network. The parameters of the mapping from the scores u_i to $\mathbf{x}$ and v_i to $\mathbf{y}$ are adjusted such that the sum of squared prediction errors are minimised:

$$E_{\mathbf{x}} = \sum_{i=1}^{N} \left\| \mathbf{x}_i - \hat{\mathbf{x}}_i \right\|^2 \quad \text{and} \quad E_{\mathbf{y}} = \sum_{i=1}^{N} \left\| \mathbf{y}_i - \hat{\mathbf{y}}_i \right\|^2 \tag{10}$$

The centres and the widths are calculated as described in section 3.1 and the weights are determined by least squares. To avoid overfitting, a regularization term is added to the sum of squares of the errors while finding the parameters of the network.

4. Test Example

The proposed approach to non-linear CCA is applied to the test problem given in [5]. The variables $\mathbf{x}$ and $\mathbf{y}$ are three dimensional:

$$\mathbf{x} = \left[x_1 + x_1', x_2 + x_2', x_3 + x_3'\right] \text{ and } \mathbf{y} = \left[y_1 + y_1', y_2 + y_2', y_3 + y_3'\right],$$

$$x_1 = t - 0.3t^2, x_2 = t + 0.3t^2, \; x_3 = t^2; \tag{11}$$

$$y_1 = t^3, y_2 = -t + 0.3t^3, \; y_3 = t + 0.3t^2; \tag{12}$$

$$x_1' = -s - 0.3s^2, \; x_2' = s - 0.3s^3; \; x_3' = -s^4; \tag{13}$$

$$y_1' = \operatorname{sech}(4s), y_2' = s + 0.3s^2, \; y_3' = s - 0.3s^2; \tag{14}$$

where t and s are independent and uniformly distributed over $[-1, 1]$. The plots of mode 1 and 2 in the $\mathbf{x}$ and $\mathbf{y}$ - space is shown in Fig. 1.

The data set was generated by adjusting the variance of the canonical variable as one third of the first canonical variable. Gaussian noise with standard deviation equal to 10% of the signal standard deviation was added. The variables were then auto-scaled and non-linear CCA was applied. The number of neurons in the projection stage were optimised through cross-validation to reproduce the vectors $\mathbf{x}$ and $\mathbf{y}$. For the test problem, the number of neurons was fifteen. Two canonical variables explained approximately 95% of the variance in $\mathbf{X}$ and $\mathbf{Y}$. In the inverse mapping, from

ESANN'2002 proceedings - European Symposium on Artificial Neural Networks
Bruges (Belgium), 24-26 April 2002, d-side publi., ISBN 2-930307-02-1, pp. 507-512

canonical variables to original variables, the number of neurons was twelve. The plots of mode 1and mode 2 in **x** and **y** space extracted from the data are shown in Figs. 2 and 3 respectively. Comparing Fig. 1 with Figs. 2 and 3, the proposed technique is able to extract Mode 1 and Mode 2 reasonably well from the data.

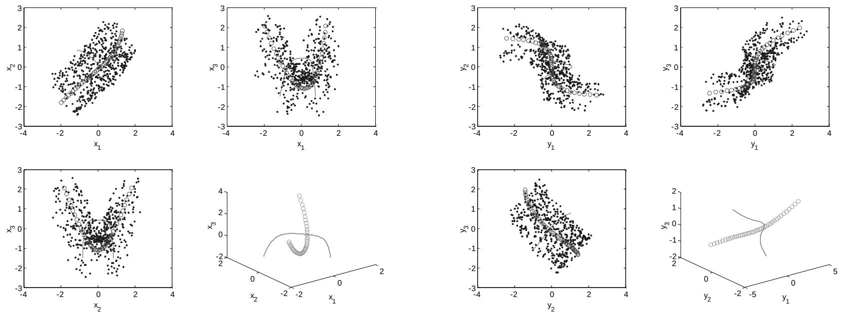

Fig. 1. Modes 1 and 2 in **x** -space (LHS) and **y** -space (RHS). ('……' - data; 'o' – Mode 1; '——' - Mode 2)

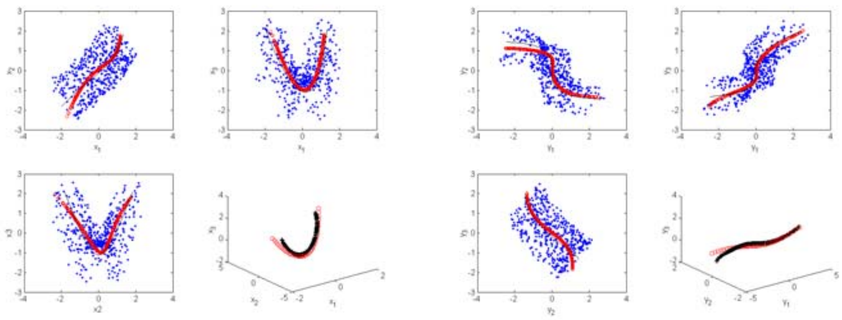

Fig. 2. Extraction of Mode 1 in **x** -space (LHS) and **y** -space (RHS). ('…' - Data; 'o' – Extracted Mode 1; '——' - Actual Mode 1).

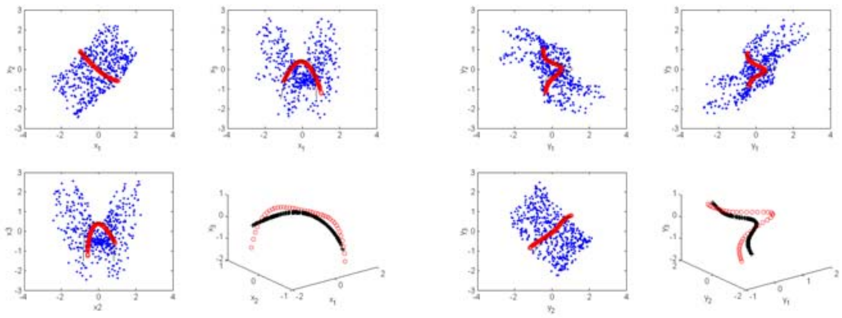

Fig. 3. Extraction of Mode 2 in **x** -space (LHS) and **y** -space (RHS). ('…' - Data; 'o' – Extracted Mode 2; '——' - Actual Mode 2).

The correlation between u_1 and v_1 is 0.9937 and between u_2 and v_2 is 0.9844. The MSE of $\mathbf{x}$ after the extraction of the first canonical variable is 0.8715 and for $\mathbf{y}$ is 0.6930. After extraction of both canonical variables, the MSE in $\mathbf{x}$ and $\mathbf{y}$ are 0.0916 and 0.1509 respectively. After the non-linear CCA model is built, the model is used to predict $\mathbf{y}$ from the given values of $\mathbf{x}$. The average MSE for the prediction of $\mathbf{y}$ for 100 new data sets, given $\mathbf{x}$ is 0.2029. These results are comparable with those reported in [5].

6. Conclusions

In this paper non-linear CCA using a radial basis function network has been proposed. For this method the training effort is less because of the near linear nature of the problem. Also the canonical variables can be extracted simultaneously. However, the issue of how many canonical variables to be retained to build the model remains unresolved. The model has been tested on synthetic data. The aim of the methodology is to use it for fault detection and diagnosis.

7. Acknowledgements

S. Kumar would like to acknowledge the financial support of the Centre for Process Analytics and Control Technology (CPACT) and the EPSRC for financial support.

8. References

1. M. A. Kramer: Nonlinear principal component analysis using autoassociative neural networks. AIChE, 37(2), 233-243 [1991].
2. D. Dong, T. J. McAvoy: Nonlinear principal component analysis-based on principal curves and neural networks. Computers and Chemical Engineering, 20(1), 65-78 [1996].
3. S. Tan, M.L. Mavrovouniotis: Reducing data dimensionality through optimizing neural network inputs. AIChE, 41(6), 1471-1480 [1995].
4. F. Jia, E.B. Martin, A.J. Morris: Non-linear principal components analysis with application to process fault detection. International Journal of System Science, 31(11), 1473 - 1487 [2000].
5. W. W. Hsieh: Nonlinear canonical correlation analysis by neural networks. Neural Networks, 13, 1095-1105 [2000].
6. D. J. H. Wilson, G.W. Irwin, G. Lightbody: RBF principal manifolds for process monitoring. IEEE Transactions on Neural Networks, 10(6), 1424-1434 [1999].
7. C. M. Bishop: Neural networks for pattern recognition. Oxford University Press [1995].
8. T. Kohonen: Self-organization and associative memory. Springer Series in Information Sciences. Springer-Verlag [1984].