The Reduction of the Dimensionality of Redundant Sensor Data Using Principal Component Analysis

Ho Cheol Shin, Moon Gue Park, Eun Gi Lee, Sung Man Bae

Korea Electric Power Research Institute, hcshin@kepri.re.kr

1. Introduction

The safety related sensors of nuclear power plant are redundant. Redundant sensor-systems achieve fault tolerance by duplication of components. It increases the ability of systems to interact with their environment by combining independent sensor readings into logical representations. Sensor integration of highly redundant systems offers these advantages: 1) Multiple inaccurate sensors can cost less than a few accurate sensors; 2) Sensor reliability may increase; *3)* Sensor efficiency and performance can be enhanced; 4) Self-calibration can be attained. But the feasibility of the systems requires attention be paid to both reliability bounds and cost. Several on-line monitoring techniques have been developed that calculate the parameter estimate using only the measurements from a group of redundant instrument channels. These techniques are commonly referred to as redundant sensor calibration monitoring models. In this paper, we reduced the dimensionality of redundant sensor data using principal component analysis.

2. Principal Component Analysis

Principal Component Analysis (PCA) is a method used to reduce the dimensionality of an input space without losing a significant amount of information (variability). The method also makes the transformed vectors orthogonal and uncorrelated. These transformed vectors can be used by regression techniques without having the problems of collinearity. A lower dimensional input space will also usually reduce the time necessary to train a neural network and the reduced noise will improve the mapping. The objective of PCA is to reduce the dimensionality and preserve as much as the relevant information as possible. PCA can also be thought of as a method of preprocessing data to extract uncorrelated features from the data.

The PCA method involves linearly transforming the input space into an orthogonal space that can be chosen to be of lower dimension with minimal loss of information. Suppose we have m samples of n random variables in a matrix x where the n columns are the variables and the m rows are the observations. We want to transform this n-dimensional space into a dimension m where k < m.

Where: n = dimensionality of original space
k = dimensionality of the reduced PC space
m = number of observations in either space

For one observation we can write the equation as follows:

$$\begin{bmatrix} t_{11} & t_{12} & \Lambda & t_{1k} \\ t_{21} & t_{22} & \Lambda & t_{2k} \\ M & M & M & M \\ t_{m1} & t_{m2} & \Lambda & t_{mk} \end{bmatrix} = \begin{bmatrix} x_{11} & x_{12} & \Lambda & x_{1n} \\ x_{21} & x_{22} & \Lambda & x_{2n} \\ M & M & M & M \\ x_{m1} & x_{m2} & \Lambda & x_{mn} \end{bmatrix} * \begin{bmatrix} p_{11} & p_{12} & \Lambda & p_{1k} \\ p_{21} & p_{22} & \Lambda & p_{2k} \\ M & M & M & M \\ p_{n1} & pn_{n2} & \Lambda & p_{nk} \end{bmatrix}$$
$$(mxk) \quad = \quad (mxn) \quad * \quad (nxk)$$

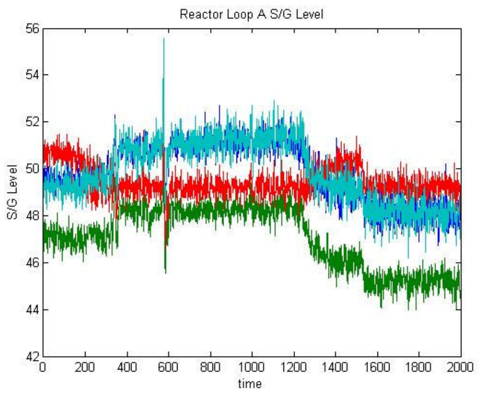
We can do this sequentially by first forming a new variable (t_1) that is a linear combination of the original p variables that has a maximum variance. For the first Principal component, this can be written as:

$$t_1 = xp_1 = x_1 p_{11} + x_2 p_{21} + ... + x_n p_{n1} = \sum_{j=1}^{n} x_j p_{j1} \ .$$

The first observation vector of raw data is labeled x and is of dimension nx1, while the transformed observation is labeled z and is kx1. Next we look for another linear function of the original variables that has maximum variance and is uncorrelated with z_1.

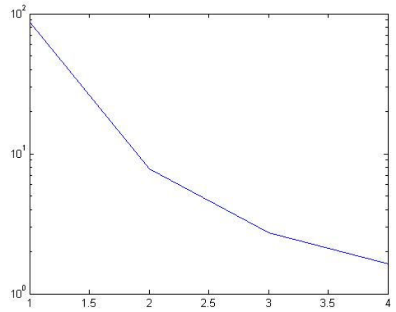
We can continue this until a maximum of n new variables called Principal Components are found. These PCs are uncorrelated and arranged in order of decreasing variance. Hopefully, some k<<n PCs can be found that will contain most of the information of the original data set. We calculate the PCs from the covariance matrix, Σ of the original data matrix x. Recall that the covariance matrix is a nxn matrix where the diagonal elements are the variances of the columns (variables) of x and the off diagonal elements (i,j) are covariances between columns i and j. Since the true covariance matrix is seldom known, we replace it with the sample covariance matrix S.

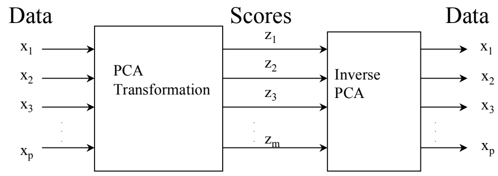
We will now use the above procedures to find the number of principal components of some highly correlated sensor data. The data to be used is from real plant. We will use the steam generator level of the reactor loop A Flow data which consists of four redundant sensors.



Now to find out how many of the principal components (or factors) are important several indicator functions can be used. Each eigenvalue is proportional to the variance in the data that the corresponding eigenvalue accounts for. If a set of eigenvalues span only random noise, the eigenvalues should be small and equal. We can plot the eigenvalues to determine how many of the principal components contain useful information. The values will be arranged in descending order with the magnitude representing the amount of information contained in that principal component (PC).

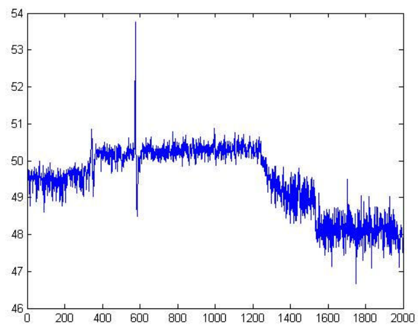
We can transform back to the data space and call this a PC prediction **Xp** using all the PCs. When using all of the PCs, the prediction equals the original data. The figure below shows the PC extraction and transformaiton from the PC scores (**z**) back to the data (**x**). When some PCs are not used to transform back to the original space, the data has effectively been filtered.





PCA Filtering

To get what is probably the best estimate of the correct sensor value, we can use the one important PC and average the results with a median filter.



3. Conclusion

The major advantages of using the PCA over the direct averaging techniques are

1. Its ability to properly model common noise.
2. Its ability to reduce spillover effects
3. Its ability to uncertain in the parameter estimate.

REFERENCE

[1] Jun Ding, A.V. Gribok, J.W. Heines, B. Rasmussen, Redundant Sensor Calibration Monitoring Using ICA and PCA, Real-Time System,27,27-47,2004.