상관계수 가중법을 이용한 커널회귀 방법

Kernel Regression with Correlation Coefficient Weighted Distance

신호철*, 박문규*, 이재용*, 류석진**
(Moon Ghu Park*, Ho Cheol Shin*, Jae Yong Lee*, and Skin You**)

Abstract - Recently, many on-line approaches to instrument channel surveillance (drift monitoring and fault detection) have been reported worldwide. On-line monitoring (OLM) method evaluates instrument channel performance by assessing its consistency with other plant indications through parametric or non-parametric models. The heart of an OLM system is the model giving an estimate of the true process parameter value against individual measurements. This model gives process parameter estimate calculated as a function of other plant measurements which can be used to identify small sensor drifts that would require the sensor to be manually calibrated or replaced. This paper describes an improvement of auto-associative kernel regression by introducing a correlation coefficient weighting on kernel distances. The prediction performance of the developed method is compared with conventional auto-associative kernel regression.

Key Words : Kernel Regression. Correlation Coefficient, Sensor Drift

1. Introduction

Recently, many on-line approaches to instrument channel surveillance (drift monitoring and fault detection) have been reported worldwide. On-line monitoring (OLM) method evaluates instrument channel performance by assessing its consistency with other plant indications through parametric or non-parametric models [1].

The heart of an OLM system is the model giving an estimate of the true process parameter value against individual measurements. This model gives process parameter estimate calculated as a function of other plant measurements which can be used to identify small sensor drifts that would require the sensor to be manually calibrated or replaced.

This paper describes an improvement of auto-associative kernel regression (AAKR) by introducing a correlation coefficient weighting on kernel distances. The prediction performance of the developed method is compared with conventional auto-associative kernel regression.

2. Methods and Results

2.1 Auto-associative Kernel Regression

Sensor drift monitoring is based on the empirical models developed with historical measurement data to generate reference signals. The reference signal values are compared to the sensor measurements and the differences, called residuals, are monitored to detect sensor degradation. To explain the auto-associative kernel regression, consider the following illustrative description of Hines [2].

The exemplar or memory vectors used to develop the empirical model are stored in a matrix X, where $X_{i,j}$ is the ith observation of the jth variable. For n_m observations of p process variables, this matrix can be written as:

$$\mathbf{X} = \begin{bmatrix} X_{1,1} & X_{1,2} & \cdots & X_{1,p} \\ X_{2,1} & X_{2,2} & \cdots & X_{2,p} \\ \vdots & \vdots & \ddots & \vdots \\ X_{n_{tm},1} & X_{n_{tm},2} & \cdots & X_{n_{tm},p} \end{bmatrix} \quad (1)$$

Using this format, a query vector is represented by a $1\times p$ vector of process variable measurements: x.

$$x = [x_1 \ x_2 \ldots \ x_p] \quad (2)$$

The corrected input is calculated as a weighted average of historical, error-free observations termed memory vectors (X_i). The mathematical framework of this modeling technique is composed of three basic steps. First, the distance between a query vector and each of the memory vectors is computed. There are several distance functions that may be used, but the most commonly used function is the Euclidean distance, whose equation for the

* 전력연구원 원자력발전연구소
** USERS

ith memory vector is as follows:

$$d_i(X_i, x) = \sqrt{\sum_j^p (X_{i,j} - x_j)^2} \quad (3)$$

For a single query vector, this calculation is repeated for each of the nm memory vectors, resulting in an $n_m x1$ matrix of distances: d. Next, these distances are transformed to similarity measures used to determine weights by evaluating the Gaussian kernel, expressed by:

$$w = K_\sigma(d) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp(-d^2 / \sigma^2) \quad (4)$$

where σ is the kernel bandwidth, w are the weights for the n_m memory vectors.

Finally, these weights are combined with the memory vectors to make predictions according to:

$$\hat{x} = \sum_{i=1}^{n_m} (w_i X_i) / \sum_{i=1}^{n_m} w_i \quad (5)$$

The kernel bandwidth σ^2 should be optimized for the trades-off of accuracy and generality of the signal.

2.2 AAKR with Correlation Coefficient Weighting

σ^2 is the bandwidth of the kernel which controls how wide the influencing measurements are spread around a query point. Bandwidth can also control the smoothness or roughness of a density estimate. Increasing the kernel width σ^2 means further away points get an opportunity to influence the query point. [3] In this paper, an improved performance of the AAKR method with correlation coefficient weighting is demonstrated in view of auto-sensitivity and accuracy.

Let's recall the normalized correlation coefficient vector assessing the linear dependence between random variables as :

$$p_j = \begin{bmatrix} \frac{\sigma_{j1}}{\sigma_j \sigma_1} & \frac{\sigma_{j2}}{\sigma_j \sigma_2} & \cdots & \frac{\sigma_{jj}}{\sigma_j \sigma_j} \end{bmatrix} / \left(\sum_{i=1}^{j} \frac{\sigma_{ji}}{\sigma_j \sigma_i} / j \right) . \quad (6)$$

where j is the index of the number of redundant sensors.

The correlation coefficient assesses the linear dependence between two random variables. It is equal to the covariance divided by the largest possible covariance and has a range $-1<p_{xy}<1$. A negative correlation coefficient simply means the relationship is inverse, or as one goes up, the other tends to go down.

The correlation coefficient weighting on distance metric is performed as follows :

$$d_i^p(X_i, x) = \sqrt{\sum_j^p (X_{i,j} - x_j)^2 \times |p_j|} \quad . \quad (7)$$

This improvement makes the variables with close linear relationship have a long range of memory vectors.

2.3 Performance Comparison

2.3.1 Accuracy

The accuracy metric is simply defined as the mean squared error (MSE) between the model's predictions and the target values. It is important to note that this metric compares the un-faulted, or error corrected, predictions with the target, or error free, data. The equation for a single variable is simply [2]:

$$A = \frac{1}{N} \sum_i^N (\hat{x}_i - x_i)^2 \quad (8)$$

where N is the number of test observations, $\hat{x}_i$ is the model prediction of the ith test observation, x_i is the ith observation of the test data. Figure 1 shows the improved accuracy of the correlation coefficient weighted AAKR over the conventional AAKR.

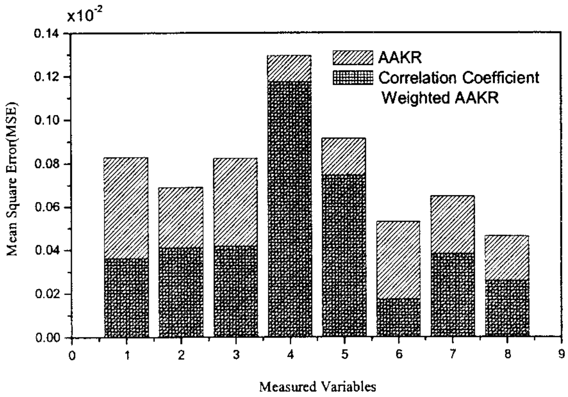

Figure 1. Comparison of accuracy of the correlation coefficient weighted AAKR with conventional AAKR.

2.3.2 Auto-Sensitivity

The auto sensitivity is a measure of an empirical model's ability to make correct sensor predictions when the respective sensor value is incorrect due to some sort of fault. Therefore, this metric involves the following values: the un-faulted prediction $\hat{x}_i$, the drifted prediction $\hat{x}^{drift}$, the un-faulted input variable x, the drifted input x^{drift}, and the index of the artificially drifted variable k. Using these definitions, the auto sensitivity for sensor k is given by:

$$S_{Ai} = \frac{1}{N} \sum_{k=1}^{N} |\hat{x}_{ki}^{drift} - \hat{x}_{ki}| / |x_{ki}^{drift} - x_{ki}| . \quad (9)$$

An auto sensitivity value of 0 is desirable and means the model is impervious to the input fault. The auto sensitivity metric is of great importance to OLM. If a model's auto sensitivity is 1, then the model's prediction follows the fault, resulting in a residual of zero, and the fault cannot be detected. If a model's auto sensitivity value is non-zero, its prediction will underestimate the size of the sensor fault and the OLM system drift limits may need to be adjusted to reflect this fact. Figure 2 shows the improved performance of the correlation coefficient weighted AAKR over the conventional AAKR.

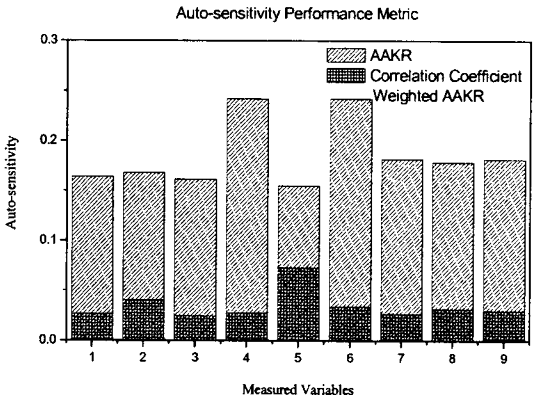

Figure 2. Comparison of auto-sensitivity of the correlation coefficient weighted AAKR with conventional AAKR.

The plots presented in Fig. 3 and 4. Fig. 2 shows the trend of measured plant variables. There is a drifting signal on the top of Fig. 3 since other variables represent the plant is being operated in a steady state. Figure 4 shows the capability of providing the reference signal predicted by the correlation coefficient weighted AAKR to identify the sensor drift.

3. Conclusion

This paper has presented an improvement in sensor drift monitoring method using the correlation coefficient weighted AAKR. The performance is demonstrated in view of its accuracy and sensitivity to identify the sensor drift. The further work would be assurance of the theoretical foundations.

ACKNOWLEDGEMENTS

We would like to express our gratitude to Dr. J. Wesley Hines at The University of Tennessee, Knoxville for his guidance and providing useful information on OLM.

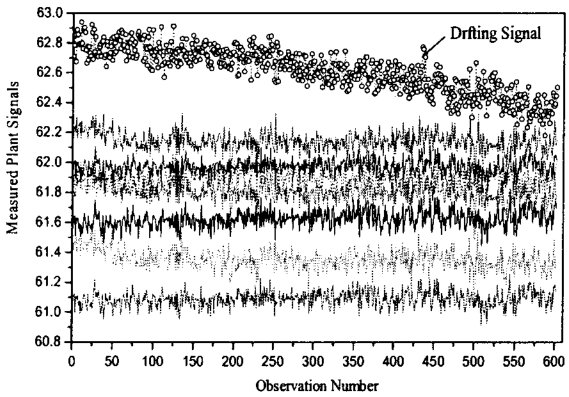

Figure 3. Trend of measured plant variables including a drifting signal

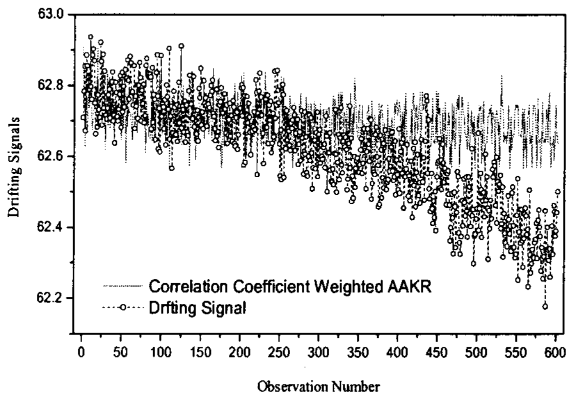

Figure 4. Signal predicted by the correlation coefficient weighted AAKR identifying the sensor drift

REFERENCES

[1] "On-line monitoring of instrument channel performance,"EPRI, Palo Alto, CA, Tech. Rep. TR-104965, Sep. 2000.

[2] J. W. Hines and D. R. Garvey, "Development and Application of Fault Detectability Performance Metrics for Instrument Calibration Verification and Anomaly Detection," Journal of Pattern Recognition Research 1, 2006.

[3] M. G. Park, H. C. Shin, Y. K. Lee, and S. You, "Filtering Performance Comparison of Kernel and Wavelet Filters for Reactivity Signal Noise," KNS Spring Meeting, May 2006.