Transactions of the Korean Nuclear Society Spring Meeting
Jeju, Korea, May 17-18, 2012

Pattern Matching Study Using Principal Component Analysis for the Monitoring of Power Plant Transients

Sangjun Park [a], Jinsoo Shin [a], Jinkyun Park [b], Gyunyoung Heo [a*]
[a] *Kyung Hee University, Yongin-si, Gyeonggi-do, 446-701, Korea*
[b] *Korea Atomic Energy Research Institute, 1045 Daedeokdaero, Yuseong-gu, Daejeon, Korea*
[*] *Corresponding author: gheo@khu.ac.kr*

1. Introduction

The purpose of this research project is to evaluate the effect of human errors during operation and maintenance, particularly related with secondary systems. In previous study, we acquired a database from an OPR1000 simulator, and signal preprocessing was carried out using Principal Component Analysis (PCA). Later on, by generating virtual data, pattern matching was performed using a cosine distance measure [1]. This research suggests an improved version of the algorithm than the previous one. In this study, we performed (1) analysis of trend of transient database to observe how to change the trend of transient in same scenario, and (2) we studied that matching method at same scenario when occurred with an initial condition which was not conventionally acquired. In this study, we will briefly explain a conventional study in the body and supplemented algorithm.

2. Methods and Results

2.1 Review of previous study

As Figure 1 shows a whole framework of previous study, the purpose of the transient urgency monitoring system is to indicate remaining time to a turbine trip, and the possibility of the turbine trip is evaluated by comparing a set of signals in a certain time window and the transient pattern database. The transient database representing the latent abnormal scenarios which occur due to a malfunction in the secondary system is composed by acquiring signals from the simulator of the Nuclear Power Plants (NPPs). The acquired data go through several steps. One of the most important preprocessing steps is a PCA for the dimension compression to reduce the computing time. Consequently, several hundred variables are reduced in size without losing much to the information therein. Operating data in a time window is acquired through a real-time database or DCS. The dimension of the operating data is suppressed by PCA as well. The pattern matching algorithm searches for the most similar pattern among the stored patterns in the transient database for the operating data. One important factor in this process is to utilize the reasonable similarity measurements. To reduce the computing time further, we performed these databases' comparison at 10 seconds interval. In this study, the feasibility of various similarity measures for the pattern matching was verified and a cosine measure was selected finally [2].

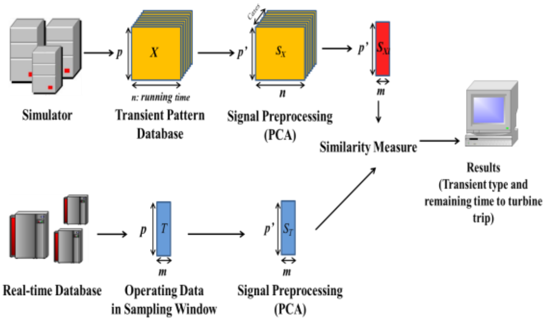

Fig. 1. Signal processing framework

Also, 10 largest principal components are trimmed in the database by using a weighting factor. When a score matrix has the eigenvalues λ_i , the weighting factor for j[th] principal component is as follow [1, 3],

$$Q_j = \frac{\lambda_j}{\sum_{i=1}^{p} \lambda_i} \tag{1}$$

Where Q_j is the weighting factor, p is the total number of principal components. For comparing an operating data with the transient pattern database, similarity measures were necessary. In the study, Cosine distance was tested as the best similarity measure. [2] The cosine measure formula is

$$d(z',t') = \sum_{j=1}^{p'} \left[Q_j \times \cos^{-1} \frac{z_j' \cdot t_j'}{\|z_j'\| \cdot \|t_j'\|} \right] \tag{2}$$

Where is Q_j the weighting factor, z' is the trimmed score matrix of transient database, t' is a score matrix of operating data.

2.2 Score matrix examination

To testify the validity of operating database, transient database is applied by using several noises. Figure 2 show data applied with various noises.

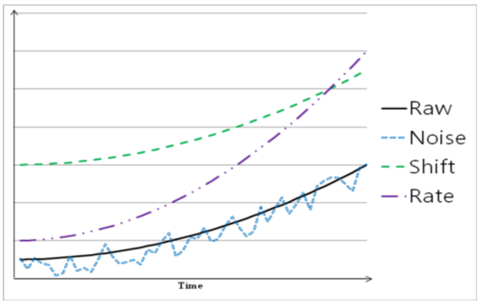

Fig. 2. Pattern examples for validation

The noises we considered are (1) a uniformly distribution noise additionally added from 1% to 10% at raw data value for general noises, (2) a parallel shift additionally added from 1% to 10% at raw data value for different initial conditions, and (3) a rate additionally multiplied from 1% to 10% at raw data value to observe increase or decrease of different initial condition. A uniformly distributed noise has the characteristic of random error and the shift is to reflect a case of adding and reducing the entire value, and rate is to consider a uniformly changing a case since, the size of transient data from NPPs is difference.

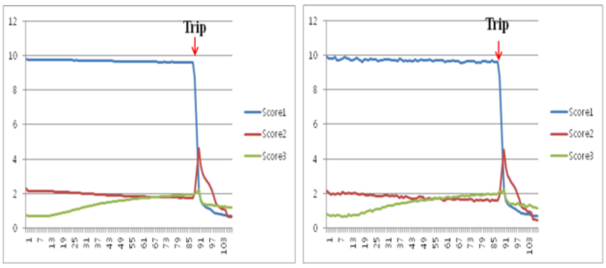

Fig. 3. A graph showing a trend applied noise at 1% and 10%

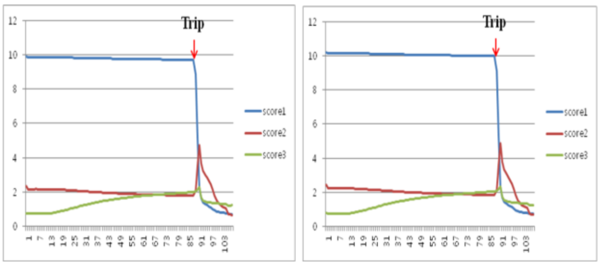

Fig. 4. A graph showing a trend applied shift at 1% and 5%

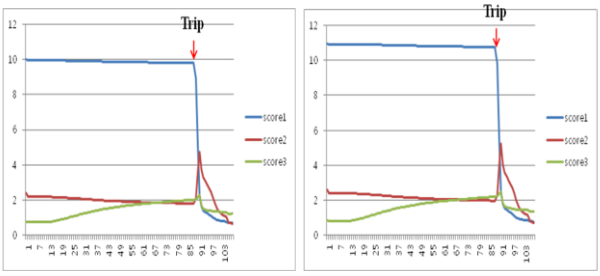

Fig. 5. A graph showing a trend applied rate at 2% and 12%

Figure 3, 4, and 5 is showing the graghs applied to various noises. A y-axis is number of score, and a x-axis stands for time. This situation reflects a SG1 pipe rupture. The result of analysis is that uniformly distribution noise and shift show the similar trend, but rate is creates a difference in the score values for the same scenario when applied to various noise. We have to use the other pattern matching methods due to the rate situation, therefore, when dealing with the states of a other initial conditions that conventionally not acquired.

2.4 Pattern fusion technique

In the study, we have to consider another situation that does not belong to the conventionally acquired transient databases. This study suggests a pattern fusion technique. If transient occurs with a different initial condition, it is matched by using the average between the value of initial condition of conventional transient database and transient database having other initial condition since all of transient database have a similar trend. For example, if occur a 3% initial condition related with a scenario when we assumed that this scenario has 1% and 5% initial condition, a pattern matching suggested in the study will choose a closest value by using the average of between other initial condition not belong to the conventional acquired initial condition and a closest value of conventionally acquired value. Therefore, if accident in same scenario occurred at a NPP, a pattern matching will recognize a closest similar pattern database in a scenario on the basis of initial conditions. Its validation is going on at present.

3. Conclusions

This paper was focused on analyzing physical behavior of NPPs under transient states through the pattern matching. To improve the validity, we tested the technique for analyzing a transient database having a different severity by using the average value of two closer initial conditions for the same scenario. And in order to check the trends for the same scenarios, we analyzed the trends of transient database for same scenario by applying signal preprocessing and various noises to the rawdata.

REFERENCES

[1] C.K Park, S.G Yu, J.K Hong, N.K Jung, K.D Kim, J.S Kim, S.G Lee, G.S Shin, S.B Choi the G.H Chang, "The development of pre-service inspection guidelines for safety of NPPs secondary system", Korea Institute of Nuclear Safety, 589, 2008

[2] S.J Park, J.S Shin, J.K Park, H.J Kim, S.Y Park, and G.Y Heo, "Development of Transient Urgency Monitoring System for Power Plants", International Symposium on Future I&C for nuclear Power Plants Cognitive System Engineering in Process Control International Symposium on Symbiotic Nuclear Power Systems ICI 2011, pp. 1175, Aug. 2011

[3] H. Abid, L.J. Willams, "Principal Component Analysis," Wiley Interdisciplinary Review; Computational Statistics, volume 2, Issue 4, pp. 433-459, July/August 2010