Development of a Big Data Framework for Operation Support of Nuclear Power Plants

Moon-Ghu Park [a*], Seong-Ki Kang [b], Jangbom Chai [c], Ji-Hun Kim [d]
[a]*Nuclear Engr. Sejong Univ., 209 Neungdong-ro, Seoul, Korea 05006*
[b]*Monitoring & Diagnosis, 1556 Duckyoungdaero, Suwon, Korea 16693*
[c]*Mechanical Engr. Ajou Univ.,206 Worldcup-ro, Suwon, Korea 16499*
[d]*UpHeare.ai Korea, 34 Myungdong 2gil, Seoul, Korea 04535*
[*]*Corresponding author: mgpark@sejong.ac.kr*

1. Introduction

This paper proposes a comprehensive big data framework for diagnosis and operation support of nuclear power plants. The framework is composed of selected Hadoop big data echo-system and has three application modules of a data-driven model for plant dynamics, a prognostic model by using a neural network/pattern matching and a dynamic time warping based pattern search algorithm. This paper presents the structure of the big data analytics system and application areas emphasizing the operational support for nuclear power plants.

2. Methods and Results

Big data analytics is becoming an inevitable tool in the foreseeable future. It should be an attractive idea to mining the historical data in the scrambled storage and convert them to knowledge. [1,2]

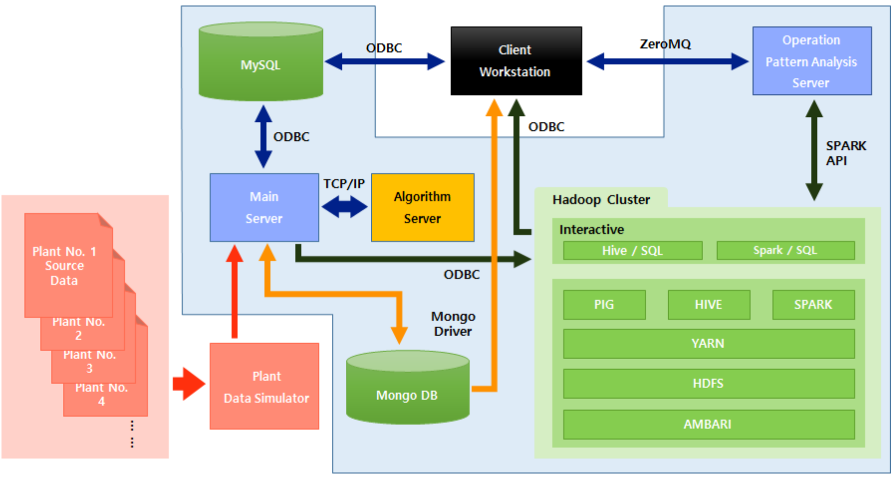

Figure 1. The architecture of BIDAN

However, we need to find a viable way to read and treat big data in a reachable time frame.

In general, the big data platforms are deployed in large scale distributed memory system with many machines using parallel logic software such as Hadoop based MapReduce and their program paradigm. The developed framework of this study is composed of selected Hadoop echo-system which includes data collection (Flume, Sqoop, FluntD), data storage (HDFS, Cassandra, Hbase), data management and job scheduling/monitoring (MapReduce, YARN, Storm, Spark), data analytics (in-house engines and RHive) and data visualization (in-house). Figure 1 shows the architecture of the framework call BIDAN (BIg Data Analytics platform for Nuclear power plants).

In the following subsections, we introduce the application modules based on the BIDAN platform.

2.1 Data-driven Model

The prediction model for system diagnosis is based on the kernel regression by introducing the invariant analysis. The invariant means a mathematical relationship of the input-output model of $\hat{\mathbf{y}} = \mathbf{f}(\mathbf{x})$ where the function is a model constructed by using the measurement data over a certain period of the system operation under an intact condition.[3] The model produces reference signals in order to identify the 'gap' that deviates from those of normal operations. The invariant based methods focusing on the big data platform are also published [1,2]. Figure 2 shows the concept to monitor the gap between invariant and current system status.

The measurement data are extracted from the Generic Pressurized Water Reactor simulator designed similar to APR1400.

2.2 Pattern Classification

Pattern classification method with neural net and feature extraction. (Fig.3) A combination of the conventional multi-layered neural network and pattern matching algorithm is developed to classify failure patterns of components. Training data are produced using simulated scenario. Typical features are extracted from failure data to simplify neural net structure and to preserve robust performance.

2.3 Dynamic Time Warping

The most desirable feature of the big data platform is to extract the overall data resources very fast not a snapshot based treatment. Operators normally want to consult the best practice previously implemented under a similar operating environment with current operational maneuvering or maintenance works. Otherwise, they can find past mis-operation and unpredicted plant behavior recorded could happen with their operation process. This is not that simple task for huge data base and semantic structure of the inter-related variables. The key factor is to locate the very similar operation record in a fast way. Dynamic time warping (DTW) technique is applied to find an optimal alignment between two given (time-dependent) sequences under certain restrictions. [5]

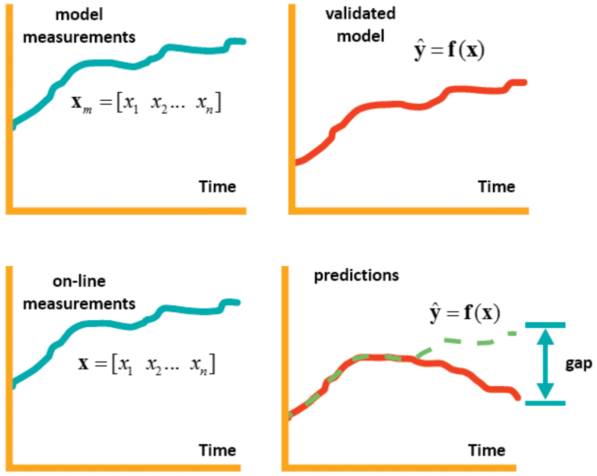

Figure 2. Prediction model for diagnosis

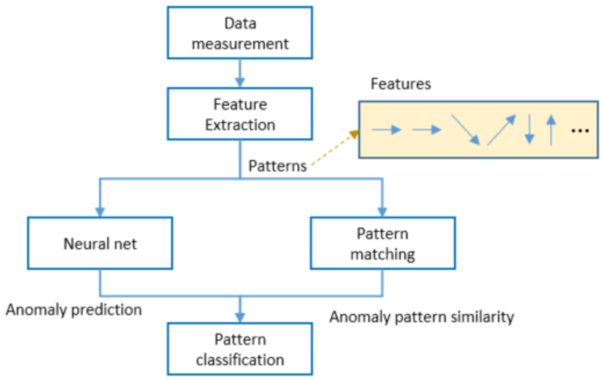

Figure 3. Pattern classification with neural net and feature extraction

Euclidean distances are used for matching criterion between the *i*-th point on one time series with the *i*-th point on the other.(Figure 4). To find the best alignment between *C* and *Q* we needs to find the path through the grid which minimizes the total distance between them. DTW algorithm is implemented with distributed Python code on the Spark-Hadoop platform. Figure 5 shows the power maneuvering plan query result from past operational record search by DTW giving other operational variables at that time.

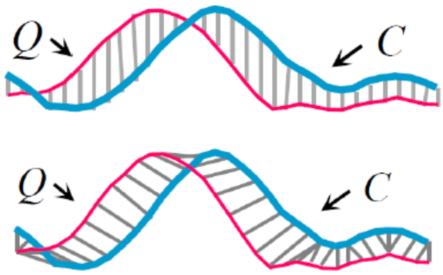

Figure 4. Nonlinear alignment of data pattern for DTW

Figure 5. Query screen of operational record search by DTW

3. Conclusions

A general framework for big data management system is developed for operation support of nuclear power plants. We investigated the Hadoop echo-system and constructed an optimized platform for plant data treatment. The application sub-modules are data-driven prediction model for plant dynamics, diagnosis and very fast operation data search algorithm to integrate the overall operation data resource of multiple units of power plants. The platform can easily become complete operation support system by just replacing the big data source constructed from plant simulator by real plant measurements.

REFERENCES

[1] [1] Fukushima K. *et al.*, Failure Sign Monitoring System for Large-scale Plants Applying System Invariant Analysis Technology (SIAT). NEC Technical Journal 9(1), 115-118, 2015.

[2] Chen P. *et al.*, InvarNet-X. A Comprehensive Invariant based Approach for Performance Diagnosis in Big Data Platform. Big Data Benchmarks, Performance Optimization, and Emerging Hardware Springer, 124-139, 2014.

[3] Park M. G. "Data Quality Enhanced Prediction Model for Massive Plant Data", Transactions of the Korean Nuclear Society Autumn Meeting, 2016.

[4] Hines J., Seibert R., Technical Review of On-Line Monitoring Techniques for Performance Assessment NUREG/CR-6895 Vols. 1, 2 and 3 US NRC, 2006.

[5] Müller M., Information Retrieval for Music and Motion, Springer, 2007.