Available online at www.sciencedirect.com
ScienceDirect
Advanced Engineering Informatics 21 (2007) 257–267
ADVANCED ENGINEERING INFORMATICS
www.elsevier.com/locate/aei

Inference of power plant quake-proof information based on interactive data mining approach

Yufei Shu *

Center for Promotion of Computational Science and Engineering, Japan Atomic Energy Agency, 6-9-3 Higashiueno, Taitou-ku, Tokyo 110-0015, Japan

Received 1 November 2006; accepted 22 December 2006

Abstract

This paper presents a nonlinear structural health inference technique, based on an interactive data mining approach. A mining control agent emulating cognitive process of human analysts is developed and integrated in the data mining loop, analyzing and verifying the output of the data miner and controlling the data mining process to improve the interaction between human users and computer system. Additionally, an artificial neural network method, which is adopted as a core component of the proposed interactive data mining method, is evolved by adding a novelty detecting and retraining function for handling complicated nuclear power plant quake-proof data. Based on proposed approach, an information inference system has been developed. To demonstrate how the proposed technique can be used as a powerful tool for inferring of structural health status in unclear power plant, quake-proof testing data have been applied.

Keywords: Interactive data mining; Cognitive model; Information inference; Structural health assessment

1. Introduction

In Japan, there are 53 Nuclear Power Plant (NPP) units in operation, with designed around 40-years lifespan. Among these NPP units, seven units have been in operation for more than 30 years, and 20 units have been in operation for more than 20 years. Some of these power plants are entering the final phases of their designed life and are becoming structurally deficient, as professor Fujino claimed that the engineers are in the "maintenance age" [1]. To maintain a higher level of reliability and safety of the NPP, it is essential to understand the true structure condition and rate of degradation of each significant part or subsystem of the NPP infrastructure. The fields of structural health assessment have had explosive growth in the past two or three decades. However, despite exponential technological progress, the use of these technologies to deal with nuclear power plant is still practically limited. For example, signal acquisition and process techniques provide real time assessment, but for the large scale infrastructure, i.e. NPP, we only can afford to deploy limited sensors. Also model-based finite element method (FEM) [2] is widely employed to analyze structural behavior. For the objective of identifying if damage is present, focusing on individual parts in the NPP, a model-based FEM analysis can work very well. However, if an infrastructure with thousands of parts, like a real case in NPP, the answer is hard to determine. For example, within large-scale three-dimensional virtual earthquake test-bed (3D-VET), which is constructing by Japan Atomic Energy Agency (JAEA) on a grid computational environment to numerically simulate the behavior of the entire NPP, under static load excitation, the computation time took 7633,232 s by using 64PE CPUs, and generated massive data set of 3.5 GB [3]. In addition, there are still some unsolved issues for model-based numerical simulation, for instance, models are limited by the capability of their mathematical description. Most of physical models are simplified or just stripped out such factors [4]. To increase the accuracy of FEM, the

* Present address: RISTEX, Japan Science and Technology Agency, 1-1-2 Otemachi, Chiyoda-ku, Tokyo 100-0004, Japan. Tel.: +81 3 3210 1265; fax: +81 3 3210 1301.
E-mail address: shuyufei@ristex.jst.go.jp

1474-0346/$ - see front matter © 2007 Elsevier Ltd. All rights reserved.
doi:10.1016/j.aei.2006.12.005

large mesh is used. However, the larger the mesh is, the longer of computing time and the larger of data size would be. At another hand, it is important to note that there are huge amount of plant quake-proof data already available in the nuclear industry. For many years, data have been accumulated which were generated by model-base analysis, and collected from experiment or sensor data. Clearly, the main issue facing for a structural health assessment is not the lack of data, but rather how to process, and analyze these collected massive amount of raw data. Therefore, the challenging tasks of NPP infrastructure health assessment are: (1) to develop model-free statistical data analysis method to represent and predict behavior of plant infrastructure, and furthermore, (2) to employ such model-free data analysis method to screen real plant structure condition and integrate it into model-based analysis platform to provide reliable information regarding the integrity of the plant structure, rapidly and accurately.

Among various data analyzing and interpreting approaches, the data mining technique is considered as an advanced means [5]. The core components of data mining technology have been under development for decades and various models have been developed, for example classical statistics regression analysis, decision tree [6], rough sets [7], and the machine learning based method (e.g. generic programming, SVM, artificial neural networks, etc.). These technologies have already been proven successfully in a variety of application domains, such as stock market assessment and prediction [8], medical diagnosis and protein structure discovery [9,10], aircraft component failure detection [5], analysis of insurance packages [11], and structural health monitoring [12,13]. However, the application of the data mining technique to handle massive nuclear power plant quake-proof information is a new attempt. Compared to other application domains, the nuclear industry has more complex databases with more indigestible physical implication related to plant parameters. In addition, for power plant infrastructure health assessment, it is impossible to either acquire ground truth data for all possible damage scenarios that are likely to occur and also to acquire in advance for some critical situations. These characteristics of plant quake-proof data require that human experts collaborate more closely to the computer based data mining system to guide data analysis process and to evaluate hypotheses.

To meet the requirements of using data mining technology to interpret nuclear power plant quake-proof information, the first objective of this research program is to establish an interactive data mining approach. To improve the interaction between users and computers, the data mining control agent based on human cognitive model will be integrated into an automated data mining loop. The second objective is to develop an intelligent quake-proof information inference system based on the proposed interactive data mining method. This inter-disciplinary computer science and structural engineering information inference system is expected to synergize the resolution of the basic technical challenge and speed up the discovery of knowledge related to infrastructure deterioration in nuclear power plant. It aims to contribute to the vitality of the NPP, as rehabilitation, renewal, replacement, and maintenance of infrastructure are required to mitigate the effects of man-made events (i.e. human operating errors, terrorism), or natural disasters (i.e. earthquake, tsunami).

Following, in Section 2, the framework of proposed interactive data mining method is explained. In Section 3, based on this method, information inference system is developed. To validate the proposed method, two kinds of typical structural NPP deterioration data are applied to the developed information inference system in Section 4. Finally, the conclusion is given in Section 5.

2. Conceptual framework of interactive data mining method

Various machine learning techniques for instance, genetic programs (GPs), neural networks (NN), and case-based reasoning (CBR) have been employed to handle wide spectrum of data mining tasks. Among these techniques, neural networks (NN) offer the advantage of dealing with unseen but similar data. Neural networks have powerful capability as a universal approximation of nonlinear mapping function. It has been successfully applied to different fields of science and engineering, for example, detecting the defect of a chemical processing plant [14], technical analyzing the stock market [8], predicting typhoon damage to electric power system [15]. There has also been increasing interest in using it as a statistical pattern classifier to identify and predict behavior of structural systems [16–18]. In this study, the neural networks approach is adopted as a core component of the interactive data mining method to represent and predict structure damage typically encountered in the NPP.

2.1. *Adverseness of neural network in interpreting plant quake-proof data*

However, neural network approach has an inherent disadvantage because of its black-box nature. For example, information or knowledge embedded in trained neural networks is hard to be verified or interpreted by human beings. Lack of facility for explanation becomes the major barrier of using neural network to interpreting NPP quake-proof data, which is desirable to enable the extraction of rules from trained neural network for analysts to gain a greater understanding of the problem at hand.

Furthermore, after training, the neural network structure is fixed. The networks treat new type of data as “unknown”. The point of controversy among researchers is that in the realistic NPP, it is impossible to get ground true data for all possible damage cases or certain critical situations. Therefore, when the new situations occur, the trained NN cannot recognize those data and classify them. It becomes another disadvantage of using neural networks