Performance Measurement using Hybrid Prediction Model in Ubiquitous Computing*

Giljong Yoo, Jeongmin Park, and Eunseok Lee
School of Information and Communication Engineering, Sungkyunkwan University, Suwon 400-746, South Korea
{gjyoo, jmpark, eslee}@ece.skku.ac.kr

Abstract

Ubiquitous computing devices recently are increasing requirements of high-level performance management automation, and therefore a system management is changing from a conventional central administration to autonomic computing. Many research centers are conducting various studies on self-healing method. However, most existing research focuses on healing after a system error has already occurred. In order to solve this problem, a prediction model is required to recognize operating environments and predict error occurrence. In this paper, we present how to predict the performance of system using hybrid prediction model. This hybrid prediction models adopts a selective healing model according to system context, for self-diagnosis and prediction of errors when using the four algorithms. In this paper, we evaluate the prediction time of the hybrid prediction model prototype and the performance of the target system's workload. In addition, the prediction is compared with existing research and the effectiveness is demonstrated by experiment.

Keywords: Self-Healing, Ubiquitous Computing, Prediction Model, ID3, Fuzzy, FNN (Fuzzy Neural Network), Bayesian NET

1. Introduction

As time goes by, computer systems need to satisfy the requirements of the Autonomic Computing. If the system can configure and reconfigure itself by knowing the operating environments, it protects and heals itself from various failures or malfunctions. In order to know the environments and detect failure, an autonomic system needs the capability of acquiring the information through self-monitoring. Distributed computing systems are continuously increasing in complexity, and system management tasks require significantly higher levels of automation [1]. Examples include diagnosis and prediction approaches, based on real-time streams of computer events, setting alarms, system faults, and performing continuous monitoring. The core of autonomic computing, a recently proposed initiative towards next-generation IT-systems capable of 'self-healing', is the ability to analyze data and to predict potential problems in real-time. Currently, most self-healing systems perform healing after error occurrence. Therefore, healing efficiency is not particularly good. We proposed a hybrid prediction model to solve this problem. This model adopts a selective model, depending on the system situation, for self-diagnosing and prediction of problems. Each algorithm is an ID3, Fuzzy, FNN and Bayesian Network. In this paper it can be seen

* This work was supported in parts by ITRC IITA-2007-(C1090-0701-0046, Basic Research Program of the Korea Science &Engineering Foundation, and the Post-BK21 Project. Corresponding author: Eunseok Lee

how this theme is being unveiled, and characteristics of each algorithm are detailed through a comparison of four prediction model and a demonstration by experiment.

The remainder of this study is organized as follows: Section 2 presents related work. Section 3 presents algorithms that used to predict a system status. Section 4 describes the proposed model. Section 5 describes implementation and evaluation. Section 6 presents the conclusion.

2. Related works

Self-healing is the one of autonomic computing technologies (Self-Optimization, Self-Healing, Self-Configuration, Self-Protection). It monitors the target system, analyzes and diagnoses the monitored data, then solves the problem by itself. In recent years, numerous studies have attempted to find and explore self-healing systems, such as IBM [2], and Sun [3]. IBM and Sun's self-healing system has the following features.

Firstly, the Adaptive Service Framework (ASF) [4] proposed by IBM and CISCO is applied in the form of self-adaptive behaviors. However, the problems in these existing systems can be summarized as follows [5]:

- Inefficient system manager and recovery after system error occurrence
- Disk, CPU and memory usage drastically increases in the process of conversion, due to the complex calculations involved.

Secondly, The Sun produced the Prediction Self-Healing system(Solaris 10). Sun has developed a new architecture for building and deploying systems and services capable of Predictive Self-Healing [3]. However, in recent years, there have been no studies of inference functions resulting in self-healing. In addition, research on self-healing prediction models is still in its infancy, as the brevity of the bibliography attests. Also, this is predicted through limited elements, and healing in the same manner as IBM (vender-dependant). Therefore, we study the prediction model to solve this problem.

3. Prediction Algorithms

Using prediction models on self-healing system result is described by the ID3, Fuzzy, FNN and Bayesian Network. The ID3 is a learning algorithm that learns from provided data. The Fuzzy logic is the theory of dividing ambiguous situations into different approximate values, when there is no explicit limit of different situations. The FNN is a combination of learning capability in neural networks and fuzzy logic. The FNN can automatically recognize fuzzy rules, and tune the membership functions. Finally, the Bayesian Network is the theory of reasoning reliability under uncertainty conditions. It is represented by a directed acyclic graph (DAG), with conditional independence of bayes theorem. Each node of this graph presents reliability. Thus, the four algorithms clarify each of the features and advantages in Table 1. Each of the Prediction models can be chosen according to situation and related work analysis. The subsequent section presents detailed descriptions that can flexibly predict a hybrid inference model.

Table. 1. Comparison of Prediction Algorithms for self healing

Features	ID3	Fuzzy	FNN	Bayesian Net.
Characteristic	Create Decision Tree with historic data of healing and run reasoning from input data.	Reason level of resource availability of system, and make decision from result.	Supports automatic tuning of Membership Functions with characteristic of fuzzy logic.	Reliability reasoning of each situation (Emergency, Error, Warn and Normal) of system state present based on conditional probability using casual inference and diagnostic inference.
Advantage	Performs correct reasoning in short time	Obtain reasoning result that reflects human's knowledge of system (utilization) situation through several resources types.	Complements the Fuzzy and has a Learning ability	Supports the Backtracking ability
Disadvantage	It is only possible to reason in the Decision Tree. Backtracking is not supported.	Rule generation is inefficiency	Until output is an expected value, it can consume considerable learning time. Implementation of algorithm is difficult.	When composition of network is incorrect, correction is difficult.
Performance (Data:500)	About 67ms	About 1000 ms	Difficulty depends on complexity of data required.	About 60~1100 ms

4. Proposed Approach

Our proposed system can select a proper algorithm, depending on system situation and features described in Table 1. Therefore, it is important to select algorithms and methods that can be applied in the healing system. This healing system is added in the server environment, and controls the utilization of server resources. The following four algorithms are used to prevent occurrence of error and for performance of the system to be kept constant.

In this paper, four algorithms are applied to the proposed architecture. That is, ID3, Fuzzy Logic, FNN and Bayesian Network.

4.1 A proposed Architecture for self-healing system

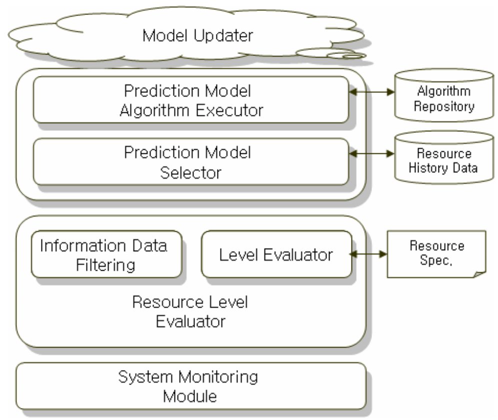

Fig. 1. Architecture for Hybrid Prediction Model

We propose architecture described in Fig. 1, for efficient use of the four inference model algorithms in the self healing system. The architecture consists of five modules and contains a partial function in the Monitoring Agent [5].

The System Monitoring Module monitors a target system. In these models, data collected in the System Monitoring Module is filtered and passed to the Resource Level Evaluator. The Resource Level Evaluator filters an information data and decides on the level of resources in the target system using the Level Evaluator. For example, the resource level's decision is manipulated by the Resource Specification. The Prediction Model Selector chooses the prediction model that can be used most suitably in the present state. For example, it is a Quantity of early studying data or degree of the critical state of an actual system. In this module, it is implemented by applying the TMR method. The Prediction Model Algorithm Executor performs a prediction model decided by the Hybrid Prediction Model Algorithm Repository.

4.2 Applying TMR method

Originally, the Triple Modular Redundancy (TMR) [6] method was used for increasing fault tolerance in safety critical software systems. TMR is built from three equal components (all three components should produce the same output for the same input) and an output selector. TMR configuration tolerates failure of a component. If the one algorithm has different output, it is recognized as a fault. And the others are recognized as a correct value of inference. An initial prediction uses the ID3, Fuzzy and Bayesian Network, and performs the three algorithms at the same time. Exactly the TMR method is not used because each algorithm is an immature model. Over time, the selective hybrid prediction model improves. Then, a latter prediction uses all algorithms with FNN and performs a performance prediction selectively. The Fuzzy neural network is always used in the background because the Fuzzy Neural Network has self-learning properties.

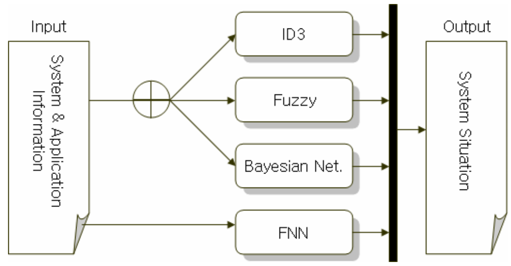

Fig. 2. Hybrid Prediction Model using the Triple Modular Redundancy

In our system, when four prediction algorithms cannot select a suitable algorithm in the specified situation, the Monitoring Agent predicts the system situation by comparison among each algorithm using TMR of Fig. 2. Four algorithms are flexibly changed and replaced according to the system situation.

4.3 Hybrid Prediction algorithm for Self-healing

4.3.1 ID3 algorithm

The ID3 is a learning algorithm that learns from provided data. The resulting tree is used to classify future system situation. A decision tree has several attributes and belongs to a class (e.g. yes or no). The leaf nodes of the decision tree contain the class name whereas a non-leaf node is a decision node. The decision node is an attribute test with each branch (to another decision tree) being a possible value of the attribute. ID3 uses information gain to help decide which attribute goes into a decision node. The advantage of learning a decision tree is that a program, rather than knowledge engineer, elicits knowledge from an expert.

The ID3's input data is fundamentally provided through a system log (e.g., the 'error_log' of the Apache server). When a log occurs, the ID3 contains the filtered historic data of the system situation (emergency > error > warning > normal), including system resource status. A given data can also be represented as a set of "if-then" rules, through learning. If new data is input, the system situation is predicted using a decision tree. This can cause objective and correct conclusions, because of the characteristics of the ID3 algorithm, such as relatively fast calculation compared to other algorithms introduced in this paper. This algorithm can obtain the fastest and most effective result using the four algorithms when the early historic data is rich.

The ID3 algorithm is used in the following situation.

- In the early stages, in the case that there is much data to make the Decision tree
- In the case that the system requires quick prediction, such as a bank's web server

4.3.2 Fuzzy Inference

It can be a dangerous method that applies fuzzy logic in an actual system. But, ambiguousness can be useful because the exact boundary of error cannot be defined in the real world. The characteristic of fuzzy inference is the following.

- Linguistic information + numeric information
- Easy to handle nonlinearity
- Human-readable knowledge

In the proposed system, the level of the Membership function is divided into five levels. MH(Max High), H(High), N(Normal), L(Low) and ML(Max Low), created by Formula (1). The duration degree of Formula (1) is the keeping time's degree of resources by [0.0, 1.0]. For example, '1.0' contentiously means a fixed resource.

$$Level = \frac{\text{ResourceUtilization}}{100} \times \text{Duration degree} \quad (1)$$

The developer or user must directly input the degree of Membership Function for generation of the rules as Fig. 3.

R1: If *CPU* is *MH* and *RAM* is *N* and *Bandwidth* is *N* then Situation is Warn

R2: If *CPU* is *H* and *RAM* is *L* and *Bandwidth* is *N* then Situation is Normal

R125: …

Fig. 3. Method to describe the fuzzy rule

4.3.3 Fuzzy Neural Network

Fuzzy neural network can support autonomic tuning of Membership Functions with characteristics of fuzzy logic. Figure 4 describes the general Fuzzy neural network used in our system. This figure specifies an example when the measured parameter is CPU, RAM, and Bandwidth. This algorithm decides a state of each resource, and then each state compares the matching degree between 'Rule1' and 'Rule125'. The resulted values are transferred to each of neurons.

The fuzzy neural network is used in the following situation.

•In the case that a developer cannot intervene in a system

•In the case that the number of rules is not sufficient to understand the status of the system

It is an adaptable FNN where the membership functions of the fuzzy predicates as well as the fuzzy rules inserted before training adaptation may adapt and change according to the training data. But, in the training of neural networks there is no guarantee that all of the domain specific information is being used and, furthermore, inspecting for specific features is highly nontrivial. Therefore, we also use the Bayesian Network. Both algorithms have many opposed characteristics.

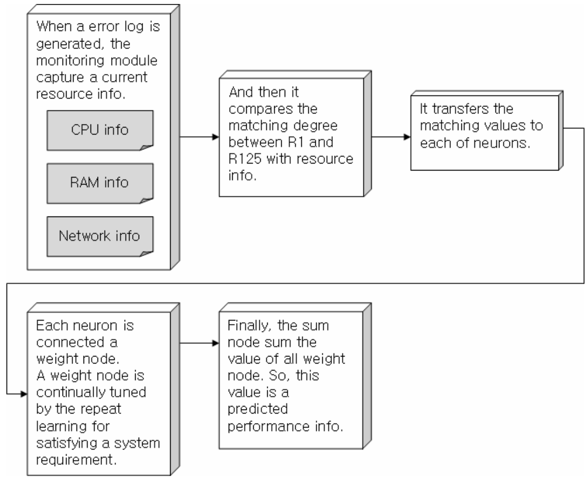

Fig. 4. Performance prediction flow using FNN

4.3.4 Bayesian Network

Bayesian networks are based on probability theory. A primer on Bayesian Networks is found in Ref. [7, 8].

This paper uses the casual inference and the diagnostic inference technique, using various inference methods. The casual inference reasons a result variable from a given cause and the diagnostic inference reasons a cause from a given result. This characteristic of the diagnostic inference is a good point. The following is a result that applies to the Bayesian Network for self-healing in this system.

"What is the probability that a system experiences an emergency when CPU is MH(Max High)?"

It is assumed that CPU and RAM each have the following reliability in the above problem. Formula (2) is used to calculate the given example conditions.

$$P(CPU\_MH) = 0.9$$
$$P(RAM\_H) = 0.6$$

CPU_MH	RAM_H	Emergency
T	T	0.95
T	F	0.8
F	T	0.4
F	F	0.1

$$
\begin{aligned}
&p(Emergency = true \mid CPU\_MH = true) \\
&= P(Emergency = true, RAM\_H = true \mid CPU\_MH = true) \\
&+ P(Emergency = true, RAM\_H = false \mid CPU\_MH = true) \\
&P(Emergency = true \mid CPU\_MH = true) \\
&= P(Emergency = true \mid RAM\_H = true, CPU\_MH = true) \\
&P(RAM\_H = true \mid CPU\_MH = true) \\
&+ P(Emergency = true \mid RAM\_H = false, CPU\_MH = true) \\
&P(RAM\_H = false \mid CPU\_MH = true) \\
&= (0.95 \cdot 0.6) + (0.8 \cdot 0.4) = 0.89
\end{aligned} \tag{2}
$$

Four models are presented in this paper. Each model can be used in a hybrid manner, depending on the inference function and characteristics of the system.

5. Implementation and Evaluation

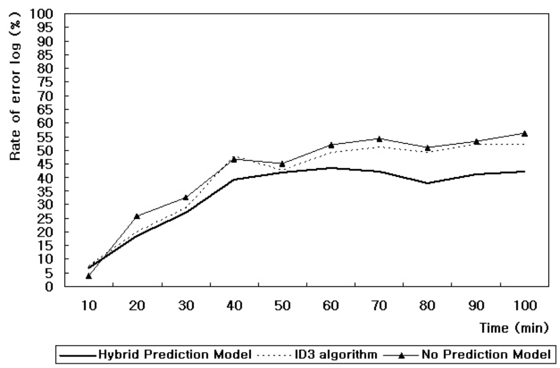

Fig. 5. Error rate changed by the time flow

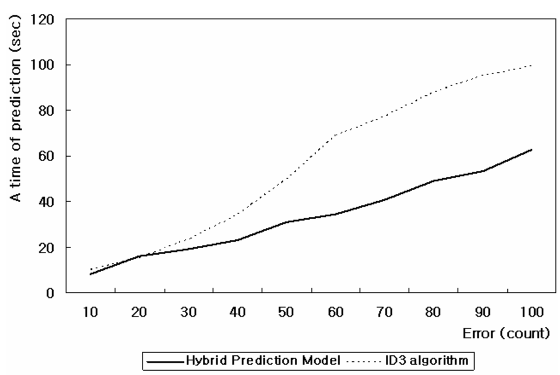

Fig. 6. A processing time to predict the system situation by the number of error

In order to evaluate hybrid inference model's performance, firstly, a prototype of a self-healing system is implemented, secondly, this prototype is applied to a self-healing system, and thirdly, it is compared with the ID3 algorithm. We assume that a traditional algorithm is the ID3 algorithm with weak prediction ability. Finally, we conduct experiments of following two cases.

In the experiment, the operating system used was Windows XP professional(with Apache server), the CPU was an Intel Pentium 3GHz, with 512Mbyte's of RAM, and the network capacity was 100Mbps. The self-healing system was installed on the target system [5], and then we evaluated the efficiency of the hybrid algorithm.

In the first experiment, comparison of the existing system was performed, which only used the ID3 algorithm to predict the proposed system. As a result, it was concluded that the error log rate was reduced through use a suitable inference model, according to system requirements. The above experiment result is presented in Fig. 6.

Figure 6 presents the time taken for inference in the second experiment. First, we assume that learned data is of sufficient quantity. So, the efficiency the model is presented and proven through comparison of the number of errors occurring in the target system. If the number of errors is lower, ID3's result displays similar performance such as the result of applying a hybrid model. However, as the number of errors increases, the shorter the inference time of the hybrid model.

6. Conclusions

Previous monitoring technologies of the self-healing system left unsatisfied the requirement for the ubiquitous computing. Currently, the autonomic computing requires a prediction model for more effective and rapid monitoring. Therefore, this paper presents algorithms for designing a prediction model for self-healing systems, and describes an approach for predicting the possibility of failure using a tool-kit of multiple models, including ID3, Fuzzy, FNN, and Bayesian Network. The main objective is to enable fuzzy logic, neural network, and Bayesian methods to be used in the early phases of the operation of an autonomic distributed system, and then switch over to an efficient decision tree algorithm once a sufficient amount of historical data has been gathered. Four algorithms proposed in this paper are clearly divided, according to characteristics and can be used in hybrid prediction models, depending on the system request. We described the situation where each algorithm can be used. And, we explain how to use each algorithm. We designed architecture to support the hybrid prediction model for the self-healing system. Also, we only use three types to predict a system performance. But we will use various systems, application parameter and OS environment variables to predict a target system performance in the future.

7. References

[1] R.K. Sahoo, A. J. Oliner, I.Rish, M. Gupta, J.E. Moreira, S. Ma, "Critical Event Prediction for Proactive Management in Large-scale Computer Clusters", the ninth ACM SIGKDD international conference on Knowledge discovery and data mining , pp. 426-435, 2003.

[2] B. Topol, D. Ogle, D. Pierson, J. Thoensen, J. Sweitzer, M. Chow, M. A. Hoff-mann, P. Durham, R. Telford, S. Sheth, T. Studwell, "Automating problem determination: A first step toward self-healing computing system", IBM white paper, Oct. 2003.

[3] Sun Microsystems: Predictive Self-Healing in the Solaris 10 Operating System, http://www.sun.com/bigadmin/content/selfheal.

[4] J. Baekelmans, P. Brittenham, T.Deckers, C.DeLaet, E.Merenda, BA. Miller, D.Ogle, B.Rajaraman, K.Sinclair, J. Sweitzer, "Adaptive Services Framework CISCO white paper", Oct.2003.

[5] Jeongmin Park, Giljong Yoo and Eunseok Lee, "Self-Management System based on Self-Healing Mechanism", LNCS4238, pp.369-378, Sep.2006.

[6] J. Von Neumann, 'Probabilistic logics and synthesis of reliable organisms from unreliable components' in Automata Studies, C. E. Shannon and J. McCarthy, Eds. Princeton, NJ: Princeton Univ. Press, pp. 43–98, 1956.

[7] Quinlan,J.R., 'C4.5: Programs for Machine Learning', Morgan Kauffman, 1993.

[8] D.J. Speigelhalter, A.P. Dawid, S.L. Lauritzen, R.G. Cowell, "Bayesian analysis in expert systems", Statistical Science, Vol.8, No.3, pp.219-247, 1993.