

유비쿼터스 환경에서 자가 치유를 지원하는 하이브리드 예측 모델

유길중, 박정민, 이은석
성균관대학교 컴퓨터공학과
e-mail : {gjyoo, jmpark, eslee}@ece.skku.ac.kr

A Hybrid Prediction Model for Self-Healing in Ubiquitous Environment

Giljong Yoo, Jeongmin Park and Eunseok Lee
Dept. of Computer Engineering, Sungkyunkwan University

요 약

오늘날 분산 컴퓨팅 환경에서 운용되는 시스템이 증가됨에 따라 시스템의 관리작업은 고수준의 자동화를 요구하고 있다. 이에 따라 시스템 관리방식은 전통적인 관리자 중심에서 시스템 스스로가 자신의 문제를 인식하고 상황을 분석하여 해결하는 자율 컴퓨팅 방식으로 변화하고 있으며, 현재 이에 대한 연구가 많은 연구기관에서 다양한 방법으로 이루어 지고 있다. 그러나 이러한 대부분의 연구에서 자율 컴퓨팅의 한 요소인 자가 치유는 문제가 발생한 이후의 치유에 주로 초점이 맞추어져 있다. 이러한 문제를 해결하기 위해서는 시스템 스스로가 동작환경을 인식하고 에러의 발생을 예측하기 위한 예측 모델을 필요로 하게 된다. 따라서 본 논문에서는 자율 컴퓨팅환경에서 자가 치유를 지원하는 4 가지 예측 모델 설계 방법을 제안하며, 본 예측 모델은 ID3 알고리즘, 퍼지추론, 퍼지뉴럴네트워크 그리고 베이저안 네트워크가 각 시스템 상황에 맞추어 적절하게 사용되는 하이브리드 방식이며, 이를 통해 보다 정확하고, 신속한 에러 예측이 가능해진다. 우리는 제안 모델을 평가하기 위해 본 예측 모델을 자가 치유 시스템에 적용하여 기존 연구와 예측 효율을 비교하였으며, 그 결과를 통해 제안 모델의 유효성을 증명하였다.

1. 서론

최근의 분산컴퓨팅환경은 그 운용의 폭이 넓어짐에 따라 시스템의 관리작업에 있어서 고수준(high-level)의 자동화에 대한 요구가 증가하고 있다. 예를 들면, 컴퓨터 이벤트의 실시간 스트림에 기반한 진단과 예측, 경고의 세팅, 그리고 지속적인 모니터링을 수행할 수 있는 것들이 포함된다. 자율 컴퓨팅의 핵심은 차세대 IT 시스템을 위해 최근 제안되고 있는 시스템의 잠재적인 문제를 예측하고 실시간으로 데이터를 분석하여 이에 대처하는 능력이다[1]. 현재 대부분의 자가치유 시스템(Self-Healing system)은 에러가 발생한 후 치유를 수행하는 방식으로, 치유의 효율이 떨어진다. 본 논문에서는 이러한 문제점을 극복할 수 있도록, 시스템이 스스로의 문제를 인식하고 예측하기 위해서 시스템의 상황에 따라서 모델을 선택하여 적용할 수 있

도록 하는 하이브리드 예측 모델을 제안한다. 이를 위해 ID3 알고리즘(ID3 Algorithm), 퍼지로직(Fuzzy Logic), 퍼지뉴럴네트워크(Fuzzy Neural Network) 그리고 베이저안 네트워크(Bayesian Network)를 자가치유시스템에 적용하였다.

2. 관련 연구

2.1 기존의 자가치유시스템

자가 치유는 시스템 스스로가 자신의 정보를 분석 및 진단하고, 그에 대응 할 수 있는 시스템으로써, 자율 컴퓨팅의 기술 중 하나이다[2]. 현재 자가치유시스템은 IBM[3], Sun[4]과 같은 기업에 의해 논문들이 발표되고, 활발한 연구가 진행되고 있다. 하지만 IBM의 경우 자사제품에서 발생하는 로그를 기반으로 습득된 데이터를 분석하여 이에 대처하는 방식으로써, 사후

<표 1> 예측 능력을 가지는 자가치유시스템을 위한 4 가지 알고리즘에 대한 비교

	ID3 Algorithm	Fuzzy Inference	Fuzzy Neural Network	Bayesian Network
특징	치유에 대한 히스토리 정보 (History Information)를 바탕으로 Decision Tree를 생성하여 입력된 데이터에 대한 추론 수행	시스템의 리소스(CPU, RAM, Bandwidth) 사용률에 대한 수준(level)을 추론하여 그 결과로부터 시스템의 상황을 판단	퍼지의 특징을 바탕으로 Membership Function에 대한 자동 튜닝을 지원하여 학습적인 효과를 볼 수 있음	인과추론과 진단추론을 사용하여 조건부확률을 기반으로 현재 시스템 상태의 각 상황(Emergency, Error, Warn, Normal)에 대한 신뢰도를 추론
장점	단시간에 정확한 추론을 수행할 수 있음	여러 리소스 타입에 대한 애매 모호한(utilization의 정도) 시스템의 상황에 대해서 인간의 지식을 반영해 원하는 추론 결과를 얻을 수 있음	- Fuzzy Inference의 단점 보완 - 학습 가능	- 부분적인 Backtracking 기능을 지원 - 즉, 인과추론 이외에 진단 추론을 이용하여 결과로부터 원인을 추론하는 것이 가능함
단점	- Decision Tree로 정해져 있는 범위 안에서만 추론 가능 - Backtracking이 지원되지 않음	- 개발자의 개입이 많음 (Membership Function 입력) - 규칙 생성의 비효율성	- 원하는 출력 값이 나오기까지 학습의 시간이 오래 걸릴 수 있음 - 구현의 어려움	네트워크의 구성이 잘못되었을 때, 수정이 어려움
수행 속도	67 Milliseconds	1000 Milliseconds	원하는 데이터의 복잡성에 따라 다름	67 ~ 1000 Milliseconds

처리 중심으로 인한 뒤늦은 복구로 시스템 관리의 비효율성이라는 단점을 갖고 있다[5]. 이는 긴박한 상황에 대해서 즉각적인 행동이 부족하기 때문에 많은 치유 시간을 소비하게 된다. Sun의 자가치유시스템은 Solaris 10의 내부에 사전에 예측 가능한 시스템을 연구, 개발하고 있다. 하지만 IBM과 마찬가지로 Sun에서 개발된 시스템을 대상으로 자가치유를 하기 때문에 한정된 요소에 의해 예측되며, 그를 위한 특정 예측 모델만을 사용하는 것이 특징이다. 이를 해결하기 위해 우리의 제안 시스템에서는 여러 시스템에서 범용성을 갖는 예측 모델을 적용한 자가치유시스템을 설계, 구현하였다.

2.2 자가치유시스템을 위한 예측 알고리즘

자가치유시스템에서 이용될 수 있는 예측 알고리즘은 ID3 알고리즘, 퍼지로지, 퍼지뉴럴네트워크, 베이지안네트워크가 있다. ID3 알고리즘은 주어진 데이터를 사용하여 결정 트리(Decision Tree)를 학습하는 알고리즘이다. 퍼지로지는 개념(concept)이 적용되거나 적용되지 않는 상황 사이에 분명한 경계가 존재하지 않을 때, 애매 모호한 상황을 여러 근사값으로 구분 지어 놓는 논리이다. 퍼지뉴럴네트워크는 뉴럴네트워크가 가지는 학습 능력과 퍼지를 결합한 것이다. 이를 통해 일반적인 퍼지추론의 단점을 보완할 수 있다. 퍼지뉴럴네트워크는 퍼지 규칙을 자동적으로 식별할 수 있으며, 역전파 알고리즘(Backpropagation algorithm)을 통해 연결 가중치를 수정함으로써 소속함수를 튜닝할 수 있다[6]. 마지막으로 베이지안 네트워크는 불확실한 조건하에 부분적인 신뢰도를 추론하기 위한 이론이다. 베이지안 정리의 조건부 독립을 확률이 부여된 방향성 비순환 그래프(DAG:Directed Acyclic Graph)로 표시되며, 인과네트워크(Casual Network), 믿음 네트워크(Belief Network)라고도 한다. 이 그래프의 각 노드는 신뢰도를 나타내고, 이는 확률론의 규칙에 따라서 조합되고 조종된다.

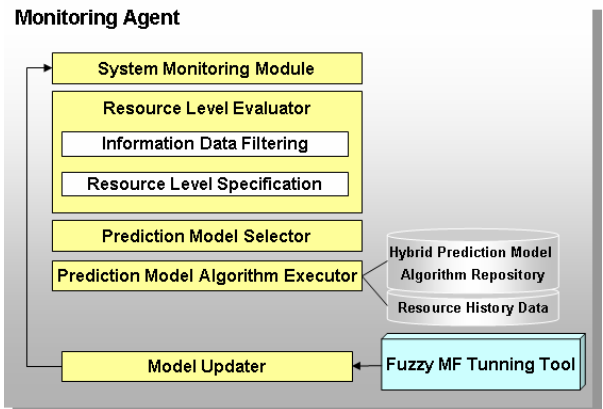
위 4 가지의 알고리즘은 각각의 특징과 장점이 명확하며, <표 1>과 같이 구분이 될 수 있다.

우리는 이와 같은 관련 연구 분석을 바탕으로, 다양한 환경에서 적용 가능하도록 예측모델들을 상황

에 맞게 선택할 수 있으며, 이를 이용하여 효과적으로 시스템을 치유 및 관리 할 수 있는 자가 치유 시스템을 설계, 구현 하였다. 다음 장에서는 이에 대한 구체적인 언급을 할 것이다.

3. 제안 모델

시스템에 발생될 에러를 미연에 찾아내어 시스템의 성능을 항상 일정하게 유지시켜주기 위해 다음의 4 가지 알고리즘을 사용하게 된다. 본 제안 시스템은 <표 1>과 같이 기술된 특성에 따라 주어진 시스템의 상황에 맞추어 적절한 알고리즘을 선택할 수 있으며, 그러한 알고리즘을 선택하여 자가 치유 시스템에서 적용될 수 있는 방법을 기술하였다. 본 논문에서는 자가 치유 시스템에서 예측 모델의 4 가지 알고리즘의 효율적인 사용을 위하여 (그림 1)과 같은 아키텍처를 제안한다. 이 아키텍처는 우리의 자가치유시스템의 아키텍처 중 Monitoring Agent의 내부 기능 중 일부로 들어가게 된다[5].



(그림 1) 모니터링 에이전트 내부의 하이브리드 예측 모델을 지원하는 아키텍처

위 아키텍처는 System monitoring Module, Resource Level Evaluator, Prediction Model Selector, Prediction Model Algorithm Executor, Model Updater로 총 5 개의 모듈로 구성된다. 이 중, System Monitoring Module에서 수집되는 데이터는 필터링되고 Resource Level Evaluator에게 전달된다. Prediction Model Selector는

현재의 상태에서 가장 최적으로 사용될 수 있는 예측 모델을 선택하게 된다. 리소스 수준에 대한 정보 이외에 시스템 전반적인 정보(i.e. 초기 학습데이터의 양, 실제 시스템의 위급한 상태에 대한 정도, etc.)를 이용하게 되며, 제안 시스템에서는 퍼지뉴럴네트워크의 경우는 시스템에서 어느 모델이 선택되더라도 항상 시스템의 상태를 고려하도록 설계되었다. 이는 뉴럴네트워크의 특징인 자가학습기능 때문이다.

3.1 ID3 알고리즘(ID3 Algorithm)

ID3 알고리즘에서 사용되는 데이터는 기본적으로 시스템의 로그를 통해서 얻을 수 있다. 로그가 발생된 시점에 시스템의 리소스 상태와 그 상태에 대한 시스템 상황(응급, 에러, 경고, 정상)을 저장하여 히스토리 데이터로 갖게 된다. 이렇게 주어진 데이터는 ID3 알고리즘을 학습을 통해 “if-then” 규칙으로 나뉜다.

“if-then” 규칙은 결정트리의 코드화와 같은 의미를 갖는다. 이 알고리즘은 주어진 초기 히스토리 데이터가 풍부할 때 4 가지 알고리즘 중에 가장 빠르고 효과적인 결과를 얻을 수 있다.

3.2 퍼지로지(Fuzzy Logic)

실제 시스템에서 퍼지를 적용 하는 것은 위험한 방법일 수 있다. 하지만 에러의 경계를 확실하게 정할 수 없는 실제 세계에서는 이런 애매모호함이 유용하게 사용될 수 있다.

제안 시스템에서 이용하는 퍼지 추론의 소속함수의 정도는 총 5 가지 단계(MH, H, N, L, ML)로 나뉘게 된다. 식(1)의 Duration degree 는 리소스가 그 상태를 유사하게 유지하는 시간을 [0.0, 1.0]으로 나타내는 정도이다.

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

여기서 소속함수의 정도와 그로부터 생겨나는 룰들은 개발자 및 사용자가 직접 입력을 해줘야 한다. 퍼지 추론을 하기 위한 룰은 3 가지의 리소스 타입이 5 개의 레벨로 나뉘기 때문에 총 125 개가 생성된다.

이 룰을 가지고 퍼지 추론을 하기 위한 순서는 조건부 만족도 검사(Fuzzy Matching), 추론(Inference): Mamdani model, 퍼지결론 조합(Combining Fuzzy Conclusion), 비퍼지화(Defuzzification)의 4 단계로 구분된다[6].

3.3 퍼지뉴럴네트워크(Fuzzy Neural Network)

퍼지뉴럴네트워크는 소속함수와 가중치에 대한 자동튜닝을 지원한다. 이 네트워크의 최종 출력 값은 식(2)과 같다.

$$O_k = \frac{\alpha_1 C_1 + \alpha_2 C_2 + \dots + \alpha_{125} C_{125}}{\alpha_1 + \alpha_2 + \dots + \alpha_{125}} = \frac{\alpha_1}{\sum \alpha_i} C_1 + \frac{\alpha_2}{\sum \alpha_i} C_2 + \dots + \frac{\alpha_{125}}{\sum \alpha_i} C_{125} \quad (2)$$

$\frac{\alpha_1}{\sum \alpha_i}$ 는 그림의 β_1 을 말한다. O_k 는 실제 출력 값으로써 기대되는 출력 값과 다를 수가 있다. 이것은 위

에서 기술된 에러 역전파 학습알고리즘으로 해결될 수 있다. 이것은 조금씩 가중치에 변화를 주어 반복적으로 퍼지뉴럴네트워크를 수행함으로써 에러를 줄이게 된다.

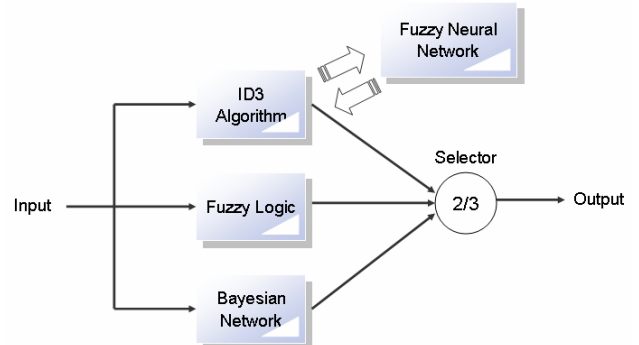
3.4 베이지안네트워크(Bayesian Network)

본 논문은 주어진 신뢰도를 바탕으로 시스템의 상황을 예측하는 방법에 인과추론과 진단추론을 이용하였다. 인과추론(casual inference)은 주어진 원인변수로부터 결과변수를 추론하는 것을 말하며, 진단추론(diagnostic inference)은 주어진 결과로부터 원인을 추론하는 것을 말한다.

베이지안 네트워크는 다음과 같은 경우에 이용된다.

- 모든 상황(emergency, error, warn, normal)에 대한 신뢰도를 판단하고 싶을 때
- 사후처리를 하기 위해 결과로부터 원인 분석을 하고 싶을 때

위에서 소개된 4 가지 알고리즘은 특정 상황에 맞는 알고리즘을 선택할 수 없는 상태가 되었을 때, (그림 2)와 같이 TMR 방식 [7]을 사용하여 각 알고리즘의 출력 값을 비교하여 예측을 수행하게 된다. 4 가지의 알고리즘이 유동적으로 교체되면서 출력 값을 비교하게 되며, 모두 동일한 입력 값과 동일한 출력 값을 만들어야 됨을 가정한다. 그래서 출력 중 하나가 다른 결과가 나왔을 때, 의도되지 않은 잘못된 결과라고 인식하고 나머지 두 개가 올바른 추론을 했다고 판단하게 된다.



(그림 2) TMR 방식을 적용한 하이브리드 예측 모델

4. 평가

본 논문에서 제안한 하이브리드 예측 모델의 성능을 평가하기 위해, 자가치유시스템의 프로토타입을 구현하여, 본 알고리즘을 적용하였으며, ID3 알고리즘만을 수행하였을 경우와 비교를 수행하였으며, 다음과 같은 항목을 평가하였다.

- 시스템을 운영하면서 시간의 흐름에 따른 에러수의 비교
- 시스템을 운영하면서 발생하는 상황에 대한 예측 시간

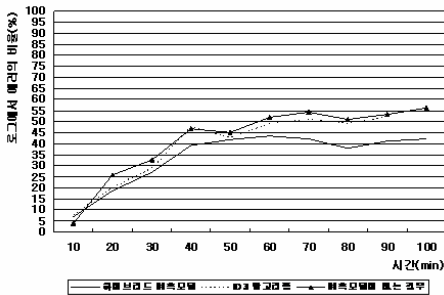
본 실험에는 OS 는 Windows XP professional 이고, CPU 는 Intel Pentium 3GHz, RAM 512Gbyte 의 성능을 가지며, 100.0Mbps 를 지원하는 유선랜으로 연결되어

있는 데스크톱 PC 를 실험환경으로 하였다.

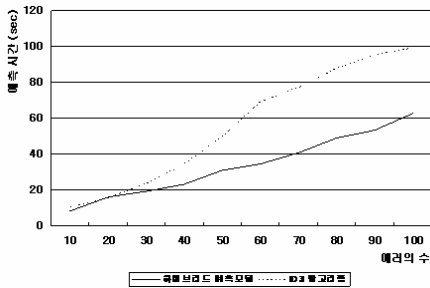
위와 같은 실험환경에서 이전에 제안한 자가치유 시스템[5]을 설치하여 하이브리드한 알고리즘 사용에 대한 효율성을 평가하였다.

본 논문에서는, 교내 웹 서버를 대상으로 한 자가치유시스템에서의 실험결과를 통해, 예측의 정확도, 시간의 양, 시스템 부하에 따라 초기 히스토리 데이터가 100 개 미만의 경우 퍼지로그직을 사용하는 것이 가장 효율적임을 알 수 있었다. 이후 350~500 개가 넘어서는 시점에 ID3 알고리즘을 적용하게 되었을 때 다른 알고리즘에 비해 우수한 예측 성능을 기대할 수 있었다. 물론 퍼지뉴럴네트워크는 히스토리 데이터의 개수에 의존하는 정도가 낮으며, 개발자가 처음에 제공하는 네트워크에 따라 스스로 학습을 하기 때문에, 개발자는 정기적으로 output 의 결과만을 모니터링하여 기대된 예측 결과와 매칭되는지 비교해 주고, 그에 따라 네트워크를 수정할 수 있어야 한다. 따라서 초기에 만들어 지는 네트워크 모델이 아주 중요한 것이 퍼지뉴럴네트워크이다.

먼저 첫 번째 실험을 위해 우리는 예측을 위해 ID3 알고리즘만을 이용하였던 기존 시스템과의 비교를 수행하였다. 실험 결과, 우리는 하이브리드 예측 모델을 사용함으로써 시스템의 요구에 따라 적합한 알고리즘을 적용함으로써 전체 로그에서 경고이상의 로그 비율(warn, error, emergency)이 감소함을 알 수 있었다. 위 실험의 결과는 (그림 3)과 같다.



(그림 3) 시간의 흐름에 따른 에러 수의 변화



(그림 4) 에러 수 변화에 따른 예측 시간

두 번째 실험을 위해 우리는 예측하는데 걸리는 시간을 측정하였다. 먼저 학습된 데이터가 일정 수준 이상을 갖추고 있다고 가정하고, 측정을 위해 ID3 알고리즘만을 사용한 경우에 에러의 수에 대해 시스템의 상황을 판단하는 정도의 비교를 통해 제안 모델의 효율성을 증명하였다. ID3의 특성에 맞게 주어진 데이터에 대해 하이브리드 모델을 적용한 것과 마찬

가지로 에러의 수가 적을수록 비슷한 성능을 나타내지만 에러의 수가 많아 질수록 리소스의 정보뿐만 아니라 시스템의 특성을 반영하여 예측을 하는 하이브리드 모델의 예측시간이 더 단축되는 것을 알 수 있다. 실험의 결과는 (그림 4)과 같다.

5. 결론

지금까지 우리는 자율 컴퓨팅 환경에서 자가 치유 시스템을 위해 시스템의 상황을 예측할 수 있는 예측 모델을 설계하기 위한 알고리즘을 정리하였다. 본 논문에서 제안된 4 가지의 알고리즘은 서로 가지고 있는 특징이 명확히 구분되며 시스템의 요구에 따라 하이브리드한 예측 모델을 설계하기 위해 사용될 수 있다. 향후 연구에 대한 방향은 다음과 같다.

- Context-aware Prediction Model: 모니터링 기능을 강화하여 리소스와 시스템의 단편적인 요소들뿐만 아니라, 애플리케이션의 성능에 관계된 파라미터, 데이터의 흐름과 같은 것들을 인식하고 이러한 정보를 예측 모델에 적용하기 위한 연구를 진행하고 있다.
- Self-growing Prediction Model: 이 연구의 핵심 알고리즘은 퍼지뉴럴네트워크다. 학습의 기능을 강화하여 Context-aware 의 기능을 결합하여, 스스로 학습하고 자가 성장 할 수 있는 예측 모델을 연구하고 있다.

참고문헌

- [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", ninth ACM SIGKDD international conference on Knowledge discovery and data mining, pp. 426-435, 2003
- [2] A. G. Ganek and T. A. Corbi, "The dawning of the autonomic computing era", IBM Systems Journal, Mar.2003
- [3] B. Topol, D. Ogle, D. Pierson, J. Thoensen, J. Sweitzer, M. Chow, M. A. Hoffmann, P. Durham, R. Telford, S. Sheth, T. Studwell, "Automating problem determination: A first step toward self-healing computing system", IBM white paper, Oct. 2003
- [4] Sun Microsystems: Predictive Self-Healing in the Solaris 10 Operating System, <http://www.sun.com/bigadmin/content/selfheal>
- [5] Jeongmin Park, Giljong Yoo and Eunseok Lee, "Proactive Self-Healing System based on Multi-Agent Technologies", ACIS International Conference on Software Engineering Research, Management & Application(SERA 2005), IEEE, pp.256-263, Aug.2005
- [6] Kwang H.Lee, 'First Course on Fuzzy Theory and Applications', Advances in Soft Computing, Springer, 2005
- [7] 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