

자가-적응 소프트웨어에서 사전 문제인지를 위한 하이브리드 모델 기반 적응 시점 판단 기법

김혜연·설광수·백두권[†]

A Timing Decision Method based on a Hybrid Model for Problem Recognition in advance in Self-adaptive Software

Hyeyun Kim · Kwangsoo Seol · Doo-Kwon Baik[†]

ABSTRACT

Self-adaptive software is software that adapts by itself to system requirements about the recognized problems without stopping the software cycle. In order to reduce the unnecessary adaptation in the system having the critical points, we propose proactive approach which can predict the future operation after a critical point. In this paper, we predict the future operation after a critical point using a hybrid model to deal with the characteristics of the observed data with the linear and non-linear pattern. The operation of the prediction method is determined on a timing decision indicator based on the prediction accuracy. The two main points of contributions of this paper are to reduce uncertainty about the future operation by predicting the situation after a critical point using hybrid model and to reduce unnecessary adaptation implementation by deciding a timing based on a timing decision indicator.

Keyword : Self-adaptive Software, Proactive Approach, ARIMA, ANN, Hybrid Model, Timing Decision

요약

자가-적응 소프트웨어는 스스로 문제를 인지하여 인지한 문제에 대하여 소프트웨어 사이클이 멈추지 않고 해당 요구사항에 맞게 적응하는 소프트웨어이다. 본 논문에서는 임계점이 존재하는 시스템에서 발생하는 불필요한 적응 수행을 감소시키기 위하여 선행적 방식으로 임계점 이후의 상황을 예측함으로써 문제가 되는 이벤트를 사전에 처리하고자 한다. 실측치는 대부분 선형과 비선형이 모두 나타나기 때문에 하이브리드 모델을 사용하여 임계점 이후를 예측하며, 예측 기법의 사용 여부는 예측의 정확도를 기반으로 하는 적응 시점 판단 지표를 기준으로 한다. 본 논문의 기여점으로는 하이브리드 모델을 MAPE-K에 적용하여 임계점 이후 상황을 예측함으로써 실제 변화에 대한 불확실성을 감소시켰다는 점과 적응 시점 판단 지표를 기반으로 적응 시점을 판단함으로써 불필요한 적응 수행을 줄였다는 데에 있다.

주요어: 자가-적응 소프트웨어, 선행적 방식, ARIMA, 인공신경망, 하이브리드 모델, 적응 시점 판단

이 논문은 2016년도 정부(미래창조과학부)의 재원으로 한국연구재단-차세대정보·컴퓨팅기술개발사업의 지원을 받아 수행된 연구임(No.2012M3C4A7033346).

Received: 2 June 2016, **Revised**: 13 July 2016,
Accepted: 21 July 2016

† Corresponding Author: Doo-Kwon Baik
E-mail: baikdk@korea.ac.kr
Computer Science and Engineering Dept.,
Korea University, Seoul, Korea

1. 서론

스마트홈, 스마트카 등 여러 분야의 사물들이 네트워크로 연결되면서 다양한 정보의 공유가 가능해졌다. 이러한 현상으로 인해 주변 환경이 유동적으로 변화하고 있어 이에 맞는 소프트웨어의 공급이 필요해지고 있다. 하지만 현실적으로 모든 요구사항에 맞는 소프트웨어를 공급하

는 것은 불가능하기 때문에 시시각각 변화하는 상황에 따른 요구사항에 맞추어 스스로 해당 환경에 적응하는 자가-적응 소프트웨어의 도입이 필요하다.

자가-적응 소프트웨어란 실시간 변화에 따른 요구사항에 맞추어 적응을 수행하는 소프트웨어를 의미한다(Salehie and Tahvildari, 2009). 자가-적응 소프트웨어는 실시간으로 데이터를 수집하고 모니터링 하는데, 모니터링 중 이상(Anomaly)이 감지되면 분석하는 단계를 거쳐 적응이 필요한 시점인지 판단하여 적응 수행 여부를 결정하게 된다. 여기에서 이상이란 정의된 개념에 따르지 않는 비정상적인 행동을 의미한다(Chandola et al., 2009). 이때, 적응에 대한 시점을 판단하는 방식에는 2가지가 있는데, 후행적(Reactive) 방식과 선행적(Proactive) 방식이 있다. 후행적 방식은 문제가 되는 이벤트가 발생한 이후에 적응을 수행하는 방식이며 선행적 방식은 문제가 되는 이벤트가 발생하기 전에 해당 이벤트를 처리하는 방식이다. 후행적 방식은 이벤트 발생 후에 처리하기 때문에 해당 문제에 대하여 보다 정확한 적응 전략 선택이 가능하다는 장점이 있지만 불필요한 적응 수행의 가능성이 있어 이에 대한 추가적인 비용이 초래될 수 있다는 한계점을 지닌다. 여기서 불필요한 적응을 수행한다는 의미는 다음과 같다. 임계값이 사전에 정의되어 있는 시스템의 경우, 수집하는 데이터 값이 임계값에 도달하였을 때 적응을 수행하도록 설계되어 있는데, 데이터 값이 임계값에 도달한 후, 바로 정상 범위에 속하게 되는 상황을 의미한다. 따라서 본 논문에서는 이벤트가 발생하기 전 임계값 이후의 상황을 예측하여 불필요한 적응 수행을 감소시킬 수 있는 선행적 방식으로 후행적 방식의 한계점을 다루고자 한다.

또한 본 논문에서는 자가-적응 소프트웨어에서의 한계점 중 하나인 불확실성(Uncertainty)을 다루고자 한다. 이와 관련된 Esfahani and Malek의 연구(2013)에서는 자가-적응 시스템에 대한 불확실성을 모호함이라는 단어로 정의하고 있으며 불확실성에 대하여 다음과 같이 복잡한 시스템을 간소화할 때의 불확실성(Simplifying assumptions), 시스템 간의 느슨한 결합에 대한 불확실성(Model drift), 잡음에 대한 불확실성(Noise), 실제 변화에 대한 불확실성(Parameters in future operation), 사람에 대한 불확실성(Human in the loop), 사용자의 요구사항 표현이나 선호를 끌어내는 데에서 나오는 복잡함에 대한 불확실성(Objectives), 유닛 간 제어에 대한 불확실성(Decentralization), 시스템마다 다른 문맥으로부터 오는 불확실성(Context), 물리적 세계에 대한 불확실성

(Cyber-physical systems)으로 분류하고 있다. 본 논문에서는 앞서 언급된 불확실성 중 실제 변화에 대한 불확실성을 시계열 분석 기반의 예측을 통하여 다루고자 한다. 시계열이란 일정한 간격에 따른 시간의 흐름과 더불어 측정된 관측 값의 계열을 의미하며, 시계열 분석은 과거에 관측된 값들의 패턴에 따라 이후에도 비슷한 양상을 띠 것이라는 전제를 지니고 있다. 실제 관측된 값은 선형과 비선형의 특징을 모두 지니고 있는 경우가 많기 때문에 단일 모델로 예측할 경우, 선형 예측 모델에서는 비선형 예측에 대하여 한계점이 발생하고 비선형 예측 모델에서는 선형 예측에 대하여 한계점이 발생하게 된다. 따라서 본 논문에서는 서로의 한계점을 보완하여 선형과 비선형을 모두 다룰 수 있는 하이브리드 모델을 사용하고자 한다. 사용하는 하이브리드 모델은 선형을 모델링하는 ARIMA(Auto-regressive integrated moving average)와 비선형을 모델링 및 예측하는 인공신경망(ANN; Artificial neural network)의 조합으로 이루어져 있다.

본 논문에서는 통신망 및 전산망, 단파 통신 등에 영향을 미치는 태양 활동으로 인한 흑점 발생에 따라 영향을 받는 전리층의 foF2 데이터를 사용하여 예측 정확도에 대한 성능과 적응 시점 판단 성능에 대한 실험을 하였으며, 이를 통해 적응 시점을 판단하여 불필요한 적응 수행을 감소시킬 뿐만 아니라 하이브리드 모델을 자가-적응에 적용하여 예측의 정확도를 향상시킴으로써 실제에 대한 불확실성을 감소시켰다는 데에 기여가 있다.

논문의 구성은 다음과 같다. 2장에서는 후행적 및 선행적 방식, 단일 모델, 선형 및 비선형 모델을 조합한 하이브리드 모델에 관련된 연구를 기술하였으며 3장에서는 자가-적응 소프트웨어, ARIMA, 인공신경망, 하이브리드 모델에 대한 연구 배경을 기술하였다. 그리고 4장에서는 제안하는 하이브리드 모델과 예측 기법의 사용 여부를 판단하는 지표인 적응 시점 판단 지표에 대하여 서술하였으며, 5장에서는 4장에서 제안한 기법에 대하여 실험 및 평가한 내용을 기술하였다. 마지막으로, 6장에서는 본 연구의 결론과 향후 연구로 구성되어 있다.

2. 관련 연구

본 장은 이상 현상 등 문제가 되는 이벤트가 발생한 후에 처리하는 후행적 방식과 사전에 처리하는 선행적 방식에 대한 연구, 한 가지의 모델을 사용하는 단일 모델로 예측한 연구 및 선형 모델과 비선형 모델을 조합한 형태인 하이브리드 모델로 예측한 연구에 대하여 서술한다.

2.1 후행적 및 선행적 방식

문제가 발생한 이후에 적응을 수행하는 후행적 방식에 대한 연구(Katseros et al., 2012)에서는 클라우드 환경에서 이질성과 상호운용성을 다루기 위하여 다층 구조를 지니는 적응형 모니터링 메커니즘을 제안한다. 이 메커니즘은 모니터링한 데이터의 양이 항상 충분하지는 않기 때문에 IaaS(Infrastructure as a Service)에서 리소스를 모니터링한 데이터와 SaaS(Software as a Service)에서 어플리케이션을 모니터링한 데이터를 PaaS(Platform as a Service)에서 통합하는 형태를 지니고 있다. 모니터링한 데이터의 양이 충분하지 않을 경우, 정상적인 모니터링이 어려워져 문제를 인식하는 데 있어 문제가 발생할 수 있게 된다. 또한, 후행적 방식으로 적응을 수행하게 될 경우, 불필요한 적응 수행이 발생할 가능성이 있다. 따라서 본 연구에서는 데이터의 부족, 부재 등을 다루기 위하여 시계열 분석과 학습을 이용하며, 불필요한 적응 수행을 감소시키기 위하여 하이브리드 모델을 사용하여 문제가 되는 상황인 임계점 이후를 예측함으로써 이를 다루고자 한다.

데이터를 모니터링하고 이상을 감지하여 사전에 문제를 처리하는 선행적 방식을 사용하기 위해서는 지속적인 모니터링이 필요하며, 이상 현상을 인지하여 그에 맞게 처리하는 것이 중요하다. Ehlers와 Hasselbring(2011)의 연구에서는 퍼포먼스 엔지니어가 모니터링 목표에 맞는 규칙을 구체화하여 Kieker 모니터링 프레임워크를 기반으로 처리하는 컴포넌트 기반 소프트웨어 시스템을 위한 규칙 기반 퍼포먼스 모니터링을 제안한다. 여기에서의 자가-적응 모니터링은 비즈니스에 악영향을 끼치는 잠재적인 실패나 이상 현상을 감소시키는 데에 목적을 두고 있으며 정의해 둔 규칙에 따라 실시간으로 적응을 수행하게 된다. 하지만 예측하는 데에 있어서 정의된 규칙만을 사용할 경우, 불확실성이 존재하는 실제 변화에 대하여 유동적으로 다룰 수 없기 때문에 한계점이 발생하게 된다. 따라서 본 연구에서는 실제 변화에 대한 불확실성을 다루기 위하여 시계열 분석과 학습을 사용한다.

2.2 단일 모델

Ehlers 등(2011)의 연구에서는 자가-적응 모니터링으로 소프트웨어 시스템에서 퍼포먼스 이상을 로컬리제이션하기 위한 접근법을 제안하며 모니터링의 범위를 실시간 자동으로 변화시키면서 잠재적으로 지연 시간을 유발시키는 요소인 응답 시간을 단일 지수 평활법(SES; Single exponential smoothing), 이중 지수 평활법(DES;

Double exponential smoothing), ARIMA를 사용하여 단일 모델로 예측한다. 단일 모델로 예측하는 또 다른 Anaya 등(2014)의 연구에서는 센서 네트워크에서 환경 요소들을 고려하여 N 시간 이후에 발생 가능한 산물의 잠재적인 값을 예측하는 예측 컴포넌트와 추론 컴포넌트를 제안하였다. 이를 위해 사용된 모델은 다층 퍼셉트론(MLP; Multi-layer perceptron), 랜덤 포레스트(RF; Random forest), 확률신경회로망(Probabilistic neural network), 퍼지 규칙 모델(Fuzzy rule model), 서포트 벡터 머신(SVM; Support vector machine), 로지스틱 회귀(Logistic regression), 나이브 베이즈(Naive bayes), Functional trees이며 그중 가장 성능이 높은 다층 퍼셉트론을 사용하였다. Ehlers 등(2011)의 연구와 Anaya 등(2014)의 연구에서는 각각 단일 모델로 예측하였는데, 실제로 관측되는 실측치에서는 선형 패턴과 비선형 패턴이 혼합되어 있는 형태이기 때문에 선형 모델에서는 비선형 패턴을, 비선형 모델에서는 선형 패턴을 잘 다룰 수 없어 한계점이 발생하게 된다. 따라서 본 논문에서는 실측치에서 나타나는 이러한 패턴을 다루기 위하여 단일 모델이 아닌 하이브리드 모델을 이용하여 예측한다.

2.3 하이브리드 모델

하이브리드 모델은 단일 모델을 사용하여 예측할 때 발생하는 문제를 여러 모델을 조합함으로써 서로의 한계점을 보완해주는 장점이 있다. 이 절에서는 여러 하이브리드 모델 중 선형 패턴은 ARIMA로 모델링하고 비선형의 패턴은 인공신경망으로 모델링 및 예측하는 연구에 대하여 서술한다. Katris와 Daskalaki(2014)의 연구에서는 인터넷 트래픽의 LRD(Long range dependence)와 SRD(Short range dependence)를 예측한다. 이 연구에서는 예측하는 과정에서 발생하는 모델 선택(Model selection) 문제를 White neural network test의 결과 값을 기반으로 선형 패턴과 비선형 패턴을 분류한 뒤, 그에 맞는 모델을 사용함으로써 다루고 있다. Herbst 등(2014)의 연구에서는 의사결정 트리와 피드백 사이클을 기반으로 주어진 문맥에 맞는 예측 방법을 선택하여 스스로 적응하는 접근법을 제안한다. Katris와 Daskalaki(2014)의 연구와 Herbst 등(2014)의 연구에서는 하이브리드 모델을 사용하여 예측함으로써 단일 모델에서 발생하는 문제를 다루고 있지만 Katris와 Daskalaki(2014)의 연구에서는 White neural network test의 결과에 따라 선형 패턴은 선형 모델로 비선형 패턴은 비선형 모델로 다루고 있어 선형과 비선형 패턴을 모두 지니는 실측치에서 한계

가 발생하게 되며, Herbst 등(2014)의 연구에서는 선형 모델로만 이루어진 하이브리드 모델로 예측하기 때문에 비선형 패턴을 다루는 데에 있어서 한계점을 지니게 된다. 따라서 본 논문에서는 선형과 비선형 패턴을 모두 다루기 위하여 먼저 ARIMA로 모델링한 후, ARIMA를 거쳐 나온 선형 및 비선형의 값들과 지수 평활화에 대한 값을 인공지능경망으로 다루는 하이브리드 모델을 사용한다.

3. 연구 배경

3.1 자가-적응 소프트웨어

자가-적응 소프트웨어는 스스로 문제를 인지하며 소프트웨어 사이클이 멈추지 않고, 인지한 문제에 대하여 해당 요구사항에 맞게 적응을 수행하는 소프트웨어를 의미한다. 자가-적응 소프트웨어는 IBM에서 제안한 MAPE-K 루프(Jeffrey et al., 2003)를 기반으로 하는데 이에 대한 그림은 Fig. 1과 같다.

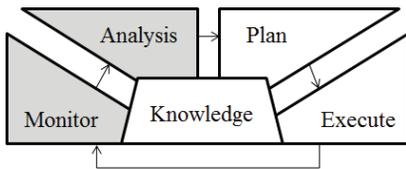


Fig. 1. MAPE-K loop

MAPE-K 루프는 수집하는 데이터를 모니터링하는 모니터 단계(M; Monitor), 이상 현상을 분석하여 문제로 인지하는 분석 단계(A; Analysis), 해당 문제에 대한 적응 전략을 수립하는 계획 단계(P; Plan), 계획 단계에서 세운 적응 전략에 따라 적응을 수행하는 수행 단계(E; Execute)로 구성된다. 그중 본 논문에서는 모니터 단계와 분석 단계를 중점적으로 다루고 있으며, 적응 시점을 판단하여 불필요한 적응 수행을 줄이기 위하여 다음과 같은 프로세스를 따른다. 우선, 데이터를 수집하여 모니터링을 하고, 모니터링 하는 값이 미리 정의해 놓은 임계점 아래에 위치한 예비점에 도달하였을 때, 임계점 이후 상황을 예측하여 적응 시점을 판단하여 적응 수행이 필요한 시점이라 판단되면 문제로 인지하게 된다.

3.2 시계열 예측 모델

시계열 분석은 시간의 흐름에 따라 변화하는 패턴에 따라 이후에도 비슷한 양상을 띠게 될 것이라는 전제를 두고 있으며 과거에서 일정한 시간의 간격으로 수집한

기록 데이터를 기반으로 미래의 값을 예측하는 데에 목적을 두고 있다.

3.2.1 ARIMA

ARIMA(p, d, q) 모델은 자기 회귀(AR; Auto-regressive)와 이동 평균(MA; Moving average)이 통합된(Integrated) 모형이며 p는 자기 회귀에 대한 차수, d는 차분한 횟수, q는 이동 평균에 대한 차수를 의미한다. 자기 회귀는 현재 값은 과거 p만큼의 영향을 받아 나타난다고 가정하며 이동 평균은 주기 q에 따라 이동하는 평균을 의미한다. ARIMA는 AR과 IMA의 수식이 합쳐진 형태로 다음과 같이 나타난다.

$$z_t = \alpha_1 Z_{t-1} + \alpha_2 Z_{t-2} + \dots + \alpha_p Z_{t-p} + \varepsilon_t - \beta_1 \varepsilon_{t-1} - \beta_2 \varepsilon_{t-2} - \dots - \beta_q \varepsilon_{t-q} \quad (1)$$

ARIMA인 z_t 를 나타내는 위의 식 (1)에서 Z는 차분을 통하여 계절성을 제거한 정상 시계열을 의미하며 α, β 는 상수를 의미하고 ε 는 오차 항을 의미한다.

시계열 분석의 대표적인 Box-Jenkins 모형(Box et al., 2015)은 모형 식별(Identification), 파라미터 추정(Parameter estimation), 진단(Diagnostic checking)의 단계로 이루어져 있으며 ARIMA는 이 모형을 따른다. 모형 식별 단계는 주어진 모형이 어느 모형에 적합하지 판단하는 단계이며, 파라미터 추정 단계는 추정된 모형으로부터의 예측치와 잔차가 적합도가 판단하는 단계이다. 여기서 잔차는 실측치와 예측치 간의 차이를 의미한다. 마지막으로 진단 단계는 최종적으로 추정된 모형이 맞는지 진단하는 단계를 의미한다.

3.2.2 인공지능경망

인공지능경망은 다층 구조로 이루어진 퍼셉트론(Perceptron)으로 구성되어 있으며 신경망(Neural network) 또는 다층 퍼셉트론이라고도 불린다. 인공지능경망의 구성은 입력층, 은닉층, 출력층으로 이루어져 있으며 식 (2)와 같이 나타낼 수 있다.

$$y_t = w_0 + \sum_{j=1}^Q w_{0,j} g(w_{0,j}) + \sum_{i=1}^P w_{i,j} y_{t-i} + e_t \quad (2)$$

y_t 는 출력층을 의미하며 y_{t-i} 는 입력층, w 는 가중치, g 는 활성화 함수, Q 는 은닉 노드의 수, P 는 입력 노드의 수를 의미한다. 인공지능경망은 입력 값과 연결의 강도를 의미하는 가중치에 대한 가중 함수를 계산한 후에, 활성화 함수를 거쳐 출력 값이 결정된다. 활성화 함수의 종류에는 출력 값이 0 또는 1로 이루어진 계단 함수(Step function), 출력 값이 -1 또는 1로 이루어진 사인 함수(Sign function),

가중 합을 그대로 출력 값으로 가지는 항등 함수(Identity function), 출력 값이 [0,1]인 시그모이드 함수(Sigmoid function), 출력 값이 [-1,1]인 하이퍼 탄젠트 함수(Hyper tangent function)가 있다. 이 중 시그모이드 함수와 하이퍼 탄젠트 함수는 미분이 가능하다는 장점을 지니고 있다. 본 연구에서는 활성화 함수로 은닉층에서는 시그모이드 함수를 사용하고, 출력층에서는 항등 함수를 사용한다.

3.2.3 하이브리드 모델

하이브리드 모델은 단일 모델의 결함을 극복하려는 방법 중 하나로서 상호작용할 수 있는 모델을 조합한 모델을 말한다. 가령 선형과 비선형의 특성이 모두 나타나는 시계열이 있을 때, 또는 순수한 선형이나 순수한 비선형의 패턴이 나타나게 될 때, 이를 단일 모델로 예측할 경우에 선형 모델로 예측하면 비선형 패턴에 대하여 잘 다룰 수 없게 되며, 비선형 모델로 예측하면 선형 패턴에 대하여 잘 다룰 수 없게 되기 때문에 예측에 대한 정확도가 떨어지게 된다. 이러한 이유로 선형 모델과 비선형 모델을 조합하여 서로의 단점을 보완해 주는 형태의 하이브리드 모델 연구가 등장하게 되었다. Pai와 Lin(2005)의 연구에서는 ARIMA와 SVM(Support vector machine)을

조합한 하이브리드 모델을 사용하여 주식을 예측하였고 Khashei와 Bijari(2011), Khashei와 Bijari(2010), Zhang(2003)의 연구에서는 ARIMA와 인공지능망을 조합한 하이브리드 모델을 사용하여 Wolf의 흑점, 잡힌 스피라소니의 수, 파운드와 달러의 환율을 예측하였다. 여기에서 ARIMA는 선형에 대하여 모델링하며 이때, ARIMA를 거친 후에는 선형 형태의 값만 얻어지는 것이 아니라 비선형 형태의 정상적 시계열과 잔차 또한 얻어지게 되는데, 이 값들은 비선형을 다루어 예측하는 SVM 또는 인공지능망을 거치게 된다. 본 논문에서는 선형 예측 모델로 보편적으로 이용되는 ARIMA와 비선형 및 예측에 대하여 수학적으로 좋은 성능을 가진다는 사실이 입증된 인공지능망을 조합한 하이브리드 모델을 사용한다(Zhang et al., 1998).

4. 제안 기법

자가-적응 소프트웨어에서 후행적 방식은 수집하는 데이터를 실시간으로 모니터링하여 미리 정의해 둔 임계점에 도달하게 되면 적응을 수행하게 된다. 하지만 임계점을 초과한 이후 데이터 값이 위험 범위에 속하게 되어 적

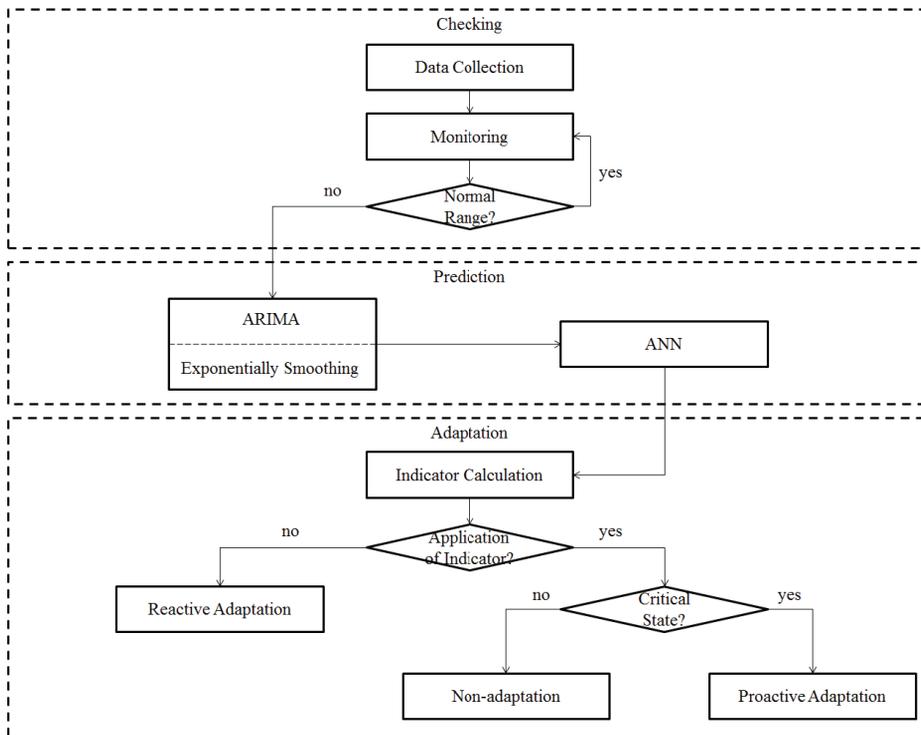


Fig. 2. Overview of timing decision process

응 수행이 필요한 경우 외에 임계점을 초과하였지만, 그 후 바로 정상 범위에 속하게 되는 경우는 불필요한 적응을 수행하는 경우라 할 수 있다. 이러한 경우, 불필요한 적응 수행으로 인하여 추가적인 비용이 발생하게 된다. 또한, 때에 따라 불필요한 적응에 대한 또 다른 적응 수행을 초래할 수 있어 임계점 이후의 상황을 예측하여 적응이 필요한 시점인지 판단하는 절차가 필요하다. 따라서 본 논문에서는 문제가 되는 이벤트가 발생하기 전에 처리하는 방식인 선형적 방식을 통해 임계점 이후의 상황을 예측함으로써 불필요한 적응 수행을 줄이고자 한다.

Fig. 2는 본 제안 기법의 전체적인 프로세스를 나타낸다. 우선, 실시간으로 데이터를 수집하며 모니터링 한다. 이 과정에서 수집하는 데이터가 정상범위를 벗어날 경우 임계점 이후의 상황을 예측하는 예측 단계로 넘어가게 되며, 여기서 정상범위를 판단하는 기준은 임계점 아래에 미리 정의해 둔 예비점에 도달하는지에 대한 여부이다. 예비점에 도달하였을 경우, 예측 단계에서 임계점 이후의 상황을 예측하게 되는데 선형 모델인 ARIMA, 지수 평활화와 비선형 모델인 인공신경망을 이용한 하이브리드 모델을 통해 예측하는 과정을 거치게 된다. 본 논문의 제안 기법으로 예측한 값은 예측의 정확도를 기반으로 하여 기법 적용 여부를 판단하는 지표의 결과 값에 따라 기법의 적용 여부를 결정하게 된다. 기법을 적용하지 않을 경우에는 기존의 후행적 방식에 따라 적응을 수행하게 되고, 기법을 적용하는 경우에는 예측한 결과에 따라 적응 수행 여부를 결정하게 된다. Fig. 2에서 Checking은 MAPE-K 루프 중 모니터 단계에 해당하며, Adaptation 은 분석 단계에 해당한다.

4.1 예측 기법

$$y_t = L_t + N_t \quad (3)$$

하이브리드 모델은 선형 컴포넌트와 비선형 컴포넌트 로 이루어져 있는데, 이 절에서는 제안하는 바를

Zhang(2003)의 표현식을 기반으로 하여 나타내고자 한다. 여기서 L_t 는 선형 컴포넌트, N_t 는 비선형 컴포넌트를 의미한다.

$$\hat{L}_t = \sum_{i=1}^p \phi_i Z_{t-i} - \sum_{j=1}^q \theta_j \varepsilon_{t-j} - \varepsilon_t \quad (4)$$

$$N_{t1} = f_1(z_{t-1}, z_{t-2}, \dots, z_{t-n}) + \varepsilon_t \quad (5)$$

$$N_{t2} = f_2(e_{t-1}, e_{t-2}, \dots, e_{t-n}) + \varepsilon_t \quad (6)$$

위 식에서 z_t 는 정상적 시계열, ε_t 는 잔차, ε_t 는 백색잡음을 의미하며, f_t 는 비선형 함수를 의미한다. 우선, 식 (4)에서 나타난 바와 같이 ARIMA로 선형 컴포넌트를 모델링 한다. 그 다음, ARIMA는 선형에 대한 모델이지만 ARIMA를 거쳐 나온 값에는 비선형 형태의 값인 정상적 시계열과 잔차를 포함하고 있어 이를 식 (5)와 식 (6)과 같이 인공신경망으로 모델링하는 과정이 필요하다.

$$E_{t+1} = E_t + \alpha(A_t - E_t) \quad (7)$$

여기서 E_t 는 선형 컴포넌트를 의미하며 지수 평활화에 대한 값을 의미한다. 또한 α 는 상수, A_t 는 실측치를 나타낸다. 이 값은 식 (8)에서 나타난 것과 같이 ARIMA로 예측한 값, ARIMA를 통해 얻은 정상적 시계열, 잔차와 더불어 인공신경망의 입력 값이 된다.

$$y_t = f(N_1 + \hat{L} + N_2 + E_t) \quad (8)$$

앞서 언급한 내용을 다시 정리해보자면, 본 예측 기법은 2단계로 이루어져 있다. 첫 번째 단계는 선형 부분에 대해서 ARIMA와 지수 평활화를 통하여 모델링한 값을 구한다. 두 번째 단계는 첫 번째 단계에서 모델링한 두 값과 ARIMA를 통하여 얻은 잔차와 정상적 시계열을 인공신경망의 입력 값으로 설정하여 출력 값을 구한다. 이 과정은 Fig. 3과 같이 나타낼 수 있으며 w 는 연결 가중치를 의미한다. 본 기법에서는 활성화 함수를 은닉층에서는 시그모이드 함수, 출력층에서는 항등 함수로 사용한다. 여기에서의 출력 값은 임계점 이후의 상황에 대하여 예측한 값을 의미한다.

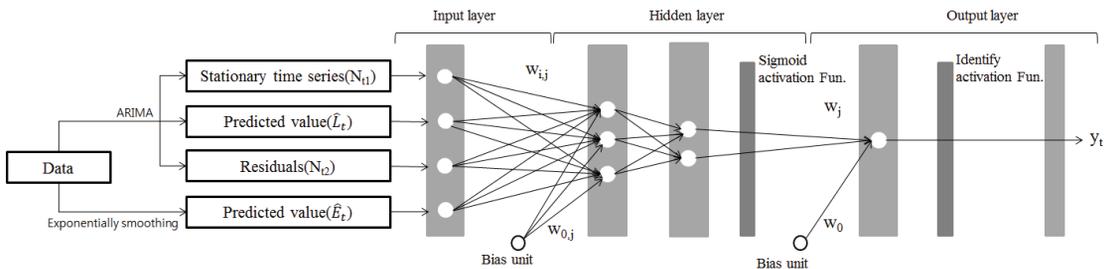


Fig. 3. Structure of the proposed prediction method

4.2 적응 시점 판단 지표

본 논문에서는 임계점이 존재하는 시스템에서의 문제점을 다루고 있으며, 이 시스템에서의 문제는 수집하는 데이터가 임계점에 도달하는 상황이다. 따라서 임계점에 도달할 때 적응을 수행하게 되는데 이러한 경우, 불필요한 적응 수행이 발생할 수 있으므로 임계점 아래에 예비점을 두어 예비점에 도달할 때 임계점 이후의 상황을 예측하여 적응 시점을 판단하고자 한다. 여기에서 예비점은 단순히 임계점 아래의 값과 더불어 임계점에 도달하기 전에 발생하는 전조 등을 포함한다.

예측에 있어서의 정확도는 항상 같은 값을 가지지 않아 어느 구간에서는 높은 예측 정확도를 보이며, 다른 구간에서는 낮은 예측 정확도를 보인다. 따라서 본 논문에서는 적응 시점을 판단하는 기법의 사용 여부에 대한 지표를 제안하여 이 지표를 기반으로 높은 정확도를 나타낼 때는 기법을 적용하고, 그렇지 않을 경우에는 기법을 적용하지 않고 기존의 방식인 후행적 방식, 즉, 임계점에 도달했을 때 적응을 수행하는 방식을 취하고자 한다. 적응 시점 판단 지표는 식 (9)와 같다.

$$\begin{aligned}
 INDICATOR = & \alpha \left\{ \frac{\sum(F_{t_2} - Z_{t_2})}{\sum F_{t_2} - Z_{t_2}} - \frac{\sum(F_{t_3} - Z_{t_3})}{\sum F_{t_3} - Z_{t_3}} \right\} \\
 & + \beta \left\{ \frac{\sum(F_{t_1} - Z_{t_1})}{\sum F_{t_1} - Z_{t_1}} - \frac{\sum(F_{t_2} - Z_{t_2})}{\sum F_{t_2} - Z_{t_2}} \right\} \\
 & + \gamma \left\{ \frac{\sum(F_t - Z_t)}{\sum F_t - Z_t} - \frac{\sum(F_{t_1} - Z_{t_1})}{\sum F_{t_1} - Z_{t_1}} \right\} \quad (9)
 \end{aligned}$$

여기에서 F_t 는 실측치, Z_t 는 예측치를 나타내며, 이 지표는 예측의 정확도를 판단하는 지표인 추적지표(Tracking signal)를 응용하여 구간마다 다른 가중치를 부여한 식이다. 예측하는 구간은 학습하고자 하는 기간에 부합하는 지점에서 시작하여 임계점 이후까지의 구간이며, 적응 시점 판단 지표를 기반으로 기법 적용 여부를 판단하는 구간은 예측을 시작한 지점에서부터 임계점 사이의 구간이다. 이 구간에 있어서 예측 정확도가 높은 추세로 나타난다면 이후에도 예측 성능이 높게 나타나고, 낮은 추세로 나타난다면 이후에도 예측 성능이 낮게 나타난다는 전제와 예측하려는 시점과 시간상으로 가까울수록 예측하려는 해당 시점에 영향을 많이 끼친다는 전제를 기반으로 하고 있다. 따라서 가중치에 대한 상수 α , β , γ 는 $0 < \alpha < \beta < \gamma$, $\alpha + \beta + \gamma = 1$ 과 같은 관계로 나타낼 수 있다. 예비점인 t_3 은 예측하려는 시점과 가장 먼 지점을 의미하며 임계점인 t_1 은 예측하려는 지점과 가장 가까운 지점을 의미한다. 이 지표의 값을 통해 예측의 정확도의

추세를 파악하여 제안 기법의 적용 여부를 판단하게 된다. 이 지표에 의하여 제안 기법을 적용하지 않을 시, 후행적 방식으로 적응을 수행하게 된다.

5. 실험 및 평가

이 장에서는 제안 기법의 예측 정확도에 대한 성능과 적응 시점 판단에 대한 성능에 대하여 평가한다. 예측 정확도에 대한 실험에서는 단일 모델인 ARIMA와 인공신경망, 하이브리드 모델인 Khashei와 Bijari(2011)의 기법과 비교하며, 적응 시점 판단에 대한 실험에서는 이를 사용했을 시 발생하는 불필요한 적응 수행 감소율에 대하여 실험하였다. 사용한 프로그램은 IBM에서 개발한 통계 프로그램인 SPSS[17]이며, 데이터는 국립전파연구원 우주전파센터[18]에서 제공하는 2014년에서 2015년 6월까지의 기간에 해당하는 선형 패턴과 비선형 패턴을 모두 지니는 전리층에서의 foF2를 사용하였다. foF2는 전리층의 F2층에 대한 상태 변화를 나타내는 지수인데, 태양흑점의 수와 양성적 비례관계를 갖는다. 태양흑점의 수는 태양의 활동이 활발할수록 증가하게 되는데, 이로 인하여 통신망 또는 전력망 등에 손상이 발생할 수 있으며, 군 및 항공기 통신 등에 사용되는 단파(High Frequency) 통신이 중지되는 경우가 발생하게 된다(박준규와 김민규, 2012). 따라서 발생할 문제의 사전 대응을 위하여 문제 상황을 예측하여 적응 시점을 판단하는 것이 필요하다(김민수 등, 2003).

5.1 예측 정확도에 대한 성능

2015년 1월 1일부터 6월30일까지의 데이터로 2층의 은닉층으로 구성된 인공신경망으로 은닉층에서는 시그모이드 함수, 출력층에서는 항등 함수를 활성화 함수로 사용하여 예측 정확도에 대한 실험을 하였으며, 실측치와 예측치를 비교하여 일부를 표현한 그래프는 Fig. 4에서 Fig. 7과 같다. 이는 각각 ARIMA로 예측한 값, 인공신경망으로 예측한 값, Khashei와 Bijari(2011)의 기법으로 예측한 값, 제안한 기법으로 예측한 값과 foF2의 값을 비교한 그래프이다. 이것을 수치로 나타낸 표는 Table. 1과 같으며 사용한 지표는 정밀도를 표현하는 데에 적합한 평균 제곱근 편차(RMSE; Root mean square deviation), 잔차의 실제치에 대한 비율을 절대값한 값들의 평균을 나타내는 평균 절대 백분율 오차(MAPE; Mean absolute percentage error), 잔차의 절대값을 평균을 의미하는 평균 절대 예측 오차(MAE; Mean absolute error)이다.

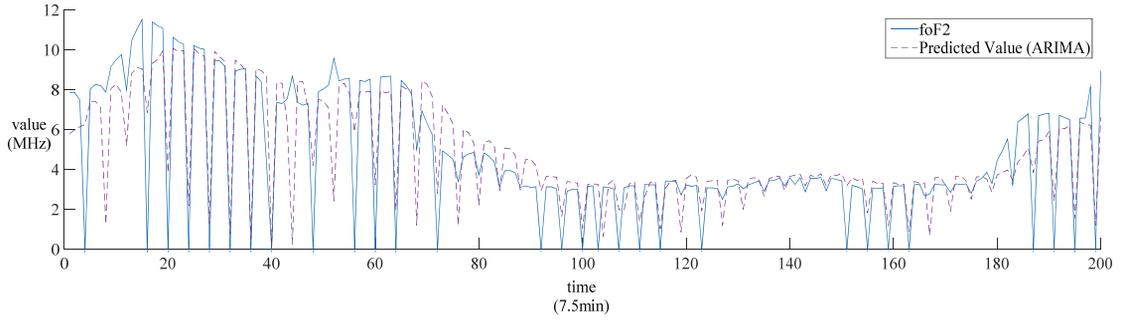


Fig. 4. Comparison between the predicted value using ARIMA and the foF2

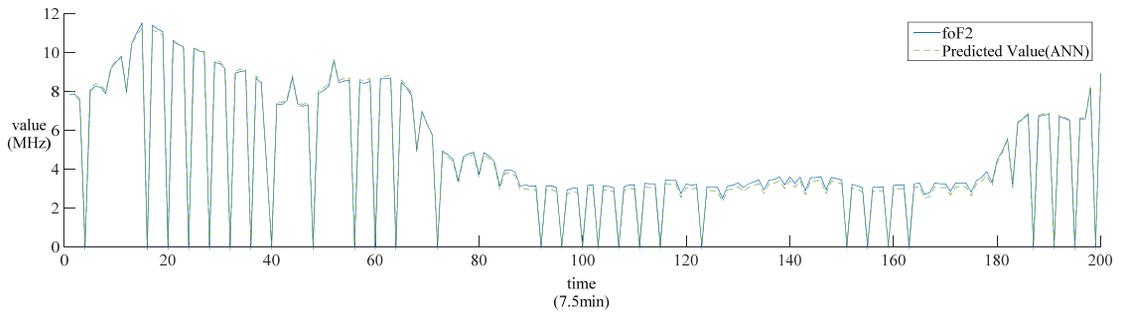


Fig. 5. Comparison between the predicted value using ANN and the foF2

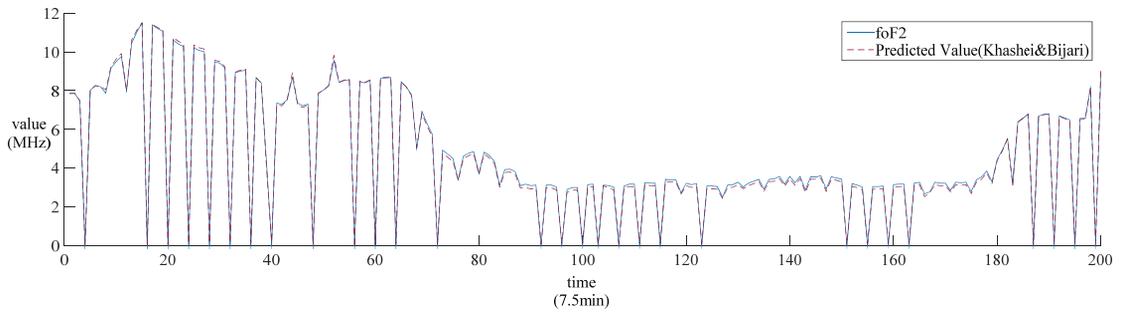


Fig. 6. Comparison between the predicted value using Khashei and Bijari(2011) and the foF2

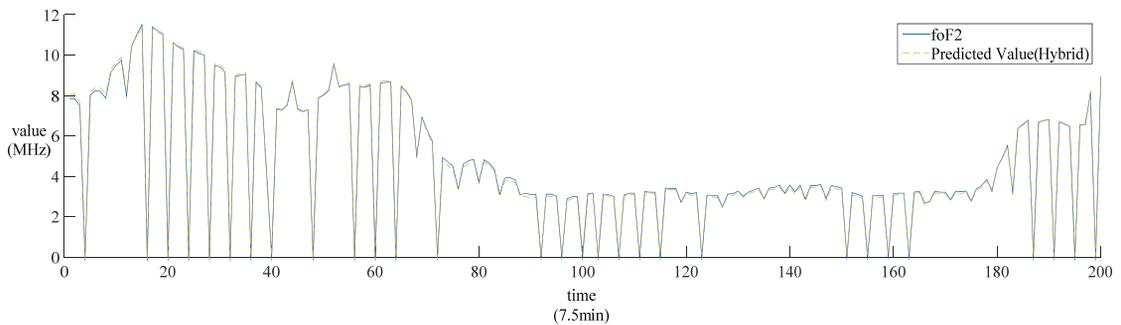


Fig. 7. Comparison between the predicted value using proposed method and the foF2

Table. 1. Accuracy of the forecasting performance

	RMSE	MAPE	MAE
ARIMA	1.902	12.350	1.179
ANN	0.133	1.678	0.113
Khashei & Bijari(2011)	0.109	1.229	0.088
Proposed method	0.108	1.087	0.073

세 가지 지표 모두 0에 가까울수록 예측 성능이 높음을 나타낸다. ARIMA와 인공신경망은 단일 모델이다. ARIMA의 경우에는 선형 예측 모델이기 때문에 비선형 패턴에 대하여 잘 다룰 수 없고, 인공신경망의 경우에는 비선형 예측 모델이기 때문에 선형 패턴에 대하여 잘 다룰 수 없다. 따라서 하이브리드 모델인 Khashei와 Bijari(2001)의 기법과 본 제안 기법보다 예측의 정확도가 낮은 결과를 보였다. 그리고 Khashei와 Bijari(2001)의 기법과 달리, 본 제안 기법에서는 시간상으로 최근과 가까운 데이터에 가중치를 더 주는 지수평활화한 값을 인공신경망의 출력층에 추가하였다. 따라서 Khashei와 Bijari(2011)의 기법에서는 1.229%가 예측치에서 벗어나는 결과를 보였고, 제안하는 기법은 1.087% 벗어나는 결과를 보였다.

5.2 적응 시점 판단에 대한 성능

이 절에서는 전리층을 사용하여 통신하는 단파통신에서 태양의 활동으로 인하여 전리층이 불안정해짐으로써 발생하는 통신 중지가 나타나기 전에 나타나는 전조를 예비점, foF2 값에 심한 교란이 오는 경우를 임계점으로 하여 적응 시점 판단에 대한 성능을 보이고자 한다. 따라서 주파수를 변경하는 시스템에서 후행적 방식과 제안하는 하이브리드 모델을 적용한 예측치를 적응 시점 판단 기법으로 적용 여부를 판단하여 적응을 수행했을 때의 적응 수행 횟수를 비교한다. 적응 시점 판단 지표를 적용한 구간은 예비점에서 임계점까지이며 이 지표를 적용한 값에 따라 적응 수행 여부를 판단한다. 사용한 데이터는 2014년 1월 1일부터 2015년 2월 28일까지이며, 2015년 2월을 중심으로 실험하였다. 총 데이터의 수는 5,347개이며 적응 수행 횟수에 대한 비교표는 Table. 2와 같고 적응 시점 판단 지표에 따라 적응을 수행한 결과에 대한 그래프는 Fig. 8과 같다.

우선, 후행적 방식의 경우에는 152회의 적응 수행이 나타났으며, 제안 기법을 적용하였을 경우에는 99회의 적응 수행이 나타나 제안 기법을 적용하였을 경우에 후행

Table. 2. Comparison between reactive and proposed method about the number of adaptation implementation

	Reactive method	Proposed method
# of adaptation	152	99
# of unnecessary adaptation	111	58

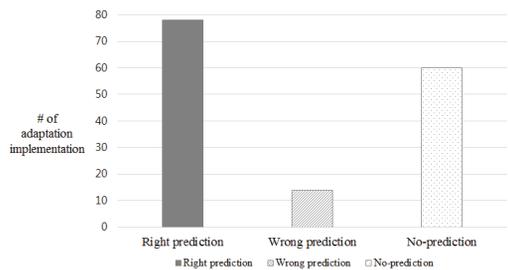


Fig. 8. Timing decision result

적 방식과 비교하여 34%의 적응 수행 횟수의 감소율을 보였다. 제안 기법의 경우 지표에 의해 예측하지 않는 경우를 제외하고, 적응 수행이 필요할 때 적응을 수행해야 하는 시점이라고 판단한 경우가 78%, 적응 수행이 불필요할 때, 적응 수행이 불필요한 시점이라고 판단한 경우가 88%로 총 올바르게 예측한 경우가 84%로 나타났다. 또한, 적응 수행이 필요할 때 불필요한 적응 시점으로 판단하는 경우가 21%, 적응 수행이 불필요할 때 적응 수행이 필요하다고 판단하는 경우가 11%로 총 잘못된 예측을 한 경우가 15%로 나타났다. 결과적으로 적응 시점 판단 지표를 적용함으로써 47%의 불필요한 적응 수행을 감소시켰다.

6. 결론

본 논문에서는 임계점이 존재하는 시스템에서 발생하는 불필요한 적응 수행을 줄이고자 임계점 이후의 상황을 예측하여 적응 시점을 판단하는 기법을 제안하고자 한다. 단순히 임계점 아래의 지점 또는 임계점 전에 발생하는 전조 등을 포함하는 예비점에 도달하였을 때, 임계점 이후를 예측하게 된다. 실측치에서는 선형이나 비선형 중 하나의 패턴을 지니기보다 대부분 둘의 양상을 같이 띠기 때문에 단일 모델을 사용했을 때 한계점이 나타나게 된다. 따라서 이러한 한계점을 다루기 위하여 선형 패

턴과 비선형 패턴을 모두 다룰 수 있는 ARIMA와 인공 신경망이 조합된 하이브리드 모델을 예측 기법으로 사용한다. 이 모델에서 ARIMA는 선형 모델링을 하고 인공신경망은 비선형 모델링 및 예측을 하는데, 이에 대한 예측치는 적용 시점 판단 지표에 따라 제안 기법의 사용 여부를 결정하게 된다. 만약 제안 기법을 사용하지 않을 경우, 후행적 방식에 따라 임계점에 도달하였을 때 적응을 수행하게 된다. 그리고 본 연구에서는 통신망 및 전산망, 단파 통신의 두절 등을 일으킬 수 있는 태양 활동과 관련된 있는 전리층의 foF2를 예측하는 실험을 하였다. 이 실험에서는 단일 모델인 ARIMA와 인공신경망, 하이브리드 모델인 Khashei와 Bijari(2011)가 제안한 기법과의 비교를 통하여 이 중 가장 높은 예측 정확도를 보였으며, 후행적 방식과 적응 시점 판단 지표를 적용했을 때를 비교하여 이 지표를 사용했을 때의 불필요한 적응 수행의 감소시켜 제안 기법의 우수성을 보였다. 본 논문의 기여점으로는 단일 모델에서 발생하는 한계점을 다룰 수 있는 하이브리드 모델을 자가-적응 소프트웨어의 MAPE-K에 적용하여 예측함으로써 실제 변화에 대한 불확실성을 감소시켰다는 점과 임계점 이후의 상황을 예측하여 적응 시점 판단 지표를 통하여 적응 시점을 판단함으로써 불필요한 적응 수행을 감소시켰다는 데에 있다.

향후 연구로는 시계열 분석에서 간과되거나 추정이 불가능한 요소를 보정해주는 판단보정기법(JA; Judgmental adjustment)을 적용하여 예측 정확도를 향상시키는 연구가 있으며, 대부분 실측치에서는 선형 패턴과 비선형 패턴이 함께 존재하지만 때에 따라 순수한 선형 또는 순수한 비선형 패턴이 존재할 수 있어 하이브리드 모델보다 단일 모델로 예측할 때의 성능이 더 좋을 수 있기 때문에 실측치에서 나타나는 패턴을 기반으로 적합한 모델을 선택하는 문제에 대한 연구가 필요하다.

References

- [1] Salehie. M., Tahvildari. T., “*Self-adaptive software: Landscape and research challenges*”, ACM Transactions on Autonomous and Adaptive Systems (TAAS), Vol.4, No.2, 2009, 14.
- [2] Chandola, V., Banerjee, A., & Kumar, V.. “*Anomaly Detection: A Survey*, ACM Computing Surveys (CSUR)”, Vol.41, No.3, 2009, 15.
- [3] Esfahani, N., & Malek, S., “*Uncertainty in Self-Adaptive Software Systems*”, Software Engineering for Self-Adaptive Systems II, Vol.7475, 2013, 214-238.
- [4] Katsaros, G., Kousiouris, G., Gogouvitis, S. V., Kyriazis, D., Menychtas, A., Varvarigou, T., “*A Self-adaptive hierarchical monitoring mechanism for Cloud*, Journal of Systems and Software”, Vol.85, No.5, 2012, 1029-1041.
- [5] Ehlers, J., Hasselbring, W., “*Self-Adaptive Software Performance Monitoring*”, In Software Engineering 2011, Lecture Notes in Informatics (LNI), 2011, 51-62.
- [6] Ehlers, J., van Hoorn, A., Waller, J., & Hasselbring, W., “*Self-adaptive software monitoring for performance anomaly localization*”, Proceedings of the 8th ACM International Conference on Autonomic Computing (ICAC), 2011, 197-200.
- [7] Anaya, I. D. P., Simko, V., Bourcier, J., Plouzeau, N., Jézéquel, J. M., “*A Prediction-Driven Adaptation Approach for Self-Adaptive Sensor Networks*”, In Proceedings of the 9th International Symposium on Software Engineering for Adaptive and Self-Managing Systems (SEAMS), 2014, 145-154.
- [8] C. Katris and S. Daskalaki. “*Combining Time Series Forecasting Methods for Internet Traffic*”, Stochastic Models, Statistic and Their Applications, Springer International Publishing, Vol.122, 2015, 309-317.
- [9] Herbst, N. R., Huber, N., Kounev, S., Amrehn, E., “*Self-adaptive workload classification and forecasting for proactive resource provisioning*”, Concurrency and computation: practice and experience, Vol.26, No.12, 2014, 2053-2078.
- [10] Jeffrey O. Kephart, David M. Chess. “*The Vision of Autonomic Computing*”, IEEE Computer Society, Vol.36, No.1, 2003, 41-50.
- [11] Box, George EP, G. M., Reinsel, G.C., Ljung, G.M., *Time series analysis: forecasting and control*, John Wiley & Sons, 2015.
- [12] Pai, P. F., Lin, C. S., “*A hybrid ARIMA and support vector machines model in stock price forecasting*”. Omega, Vol.33, No.6, 2005, 497-505.
- [13] Khashei, M., & Bijari, M., “*A novel hybridization*

- of artificial neural networks and ARIMA models for time series forecasting*”, Applied Soft Computing, Vol.11, No.2, 2011, 2664-2675.
- [14] Khashei, M., Bijari, M., “*An artificial neural network (p,d,q) model for timeseries forecasting*”. Expert Systems with Applications, Vol.37, No.1, 2010, 479-489.
- [15] Zhang, G. P., “*Time series forecasting using a hybrid ARIMA and neural network model*”, Neurocomputing, Vol.50, 2003, 159-175.
- [16] Zhang, Guoqiang, B. Eddy Patuwo, Michael Y. Hu., “*Forecasting with artificial neural network: The state of the art*”, International Journal of Forecasting, Vol.14, No.1, 1998, 35-62.
- [17] <http://www.ibm.com/analytics/us/en/technology/spss/>
- [18] <http://spaceweather.rra.go.kr/?lang=ko>
- [19] 박준규, 김민규, “태양흑점 활동에 따른 GPS 위치해석 특성 분석”, 한국지적정보학회지, 제 14권, 2호, 2012, 91-98.
- [20] 김민수, 유치형, 이해수, 정찬수, “개선된 PSFS를 이용한 태양흑점 시계열 데이터의 예측”, 대한전기학회 하계학술대회 논문집 D, 2003, 2750-2752.
- [21] 장석철, 석상문, 이주상, 이상욱, 안병하, “신경망 시계열 분석 및 판단보정 기법을 이용한 교통량 예측”, 한국경영과학회/대한산업공학회 춘계공동학술대회, 2005(단일호), 773-777.
- [22] Taskaya-Temizel, T., Casey, M. C., “*A comparative study of autoregressive neural network hybrids*”, Neural Networks, Vol.18, No.5, 2005, 781-789.
- [23] 설광수, 백두권, “자가적응형 소프트웨어의 문제인지를 위한 학습 기반 적응 시점 판단 기법”, Proceedings of the 17th Korea Conference on Software Engineering(KCSE), 제 17권, 1호, 2015.
- [24] 김혜연, 설광수, 백두권, “자가 적응 소프트웨어에서 사전 문제인지를 위한 적응 시점 판단 기법”, 한국정보과학회 2015년 동계학술발표회 논문집, 2015, 801-803.”



김혜연 (hyeyun0323@korea.ac.kr)

2015 백석대학교 정보보호학 학사
2015~ 현재 고려대학교 컴퓨터학과 석사과정

관심분야 : 자가-적응 소프트웨어, 소프트웨어 공학, 시뮬레이션



설광수 (seolks@korea.ac.kr)

2012 고려대학교 컴퓨터정보학과 학사
2012~ 현재 고려대학교 컴퓨터학과 석·박통합과정

관심분야 : 소셜 네트워크 서비스, 자가-적응 소프트웨어

백두권 (baikdk@korea.ac.kr)

1974년 고려대학교 수학과(학사)
1977년 고려대학교 산업공학과(석사)
1983년 Wayne State Univ. 전산학과(석사)
1985년 Wayne State Univ. 전산학과(박사)
현 재 고려대학교 정보대학 컴퓨터학과교수



1989년~1991년 고려대학교 전산학과과 학과장
1990년~1991년 미국 Arizona대학교 객원 교수
1991년~2013년 ISO/IEC JTC1/SC32 전문위원회 위원장
1993년~1999년 한국과학기술원 객원책임연구원
1993년~1999년 한국DB진흥센터 표준연구위원
1996년~1997년 고려대학교 컴퓨터과학기술연구소(초대소장)
1997년~1998년 고려대학교 정보전산원 원장
1998년~1999년 한국정보과학회 전산교육연구회 운영위원장
1999년~2001년 정보통신진흥협회 데이터기술위원회 의장
2002년~2004년 고려대학교 정보통신대학(초대학장)
2002년~2003년 한국시뮬레이션학회 회장
2003년~현재 정보통신부 컴퓨터프로그램보호위원회 위원
2004년~2005년 한국정보처리학회 부회장
2005년~2008년 한국소프트웨어진흥원 이사
2009년~2010년 고려대학교 정보통신대학 학장

관심분야 : 메타데이터, 소프트웨어공학, 데이터공학, 컴포넌트기반 시스템, 메타데이터 레지스트리, 프로젝트 매니지먼트