

ARIMA 모델을 이용한 데이터 흐름 예측 기법

김동현⁰, 김민우*, 이병준*, 김정태*, 윤희용**

⁰성균관대학교 정보통신대학 전자전기컴퓨터공학과

**성균관대학교 소프트웨어대학 소프트웨어학과

e-mail: {kdh7263O, kimmw95*, byungjun*}@skku.edu, kyungtaekim76@gmail.com*,
youn7147@skku.edu**

Data Flow Prediction Scheme using ARIMA Model

Dong-Hyun Kim⁰, Min-Woo Kim*, Byung-Jun Lee*, Kyung-Tae Kim*, Hee-Yong Youn**

⁰Dept. of Electrical and Computer Engineering, Sungkyunkwan University

**Dept. of Software, Sungkyunkwan University

● 요약 ●

기존 데이터의 패턴 예측에는 통계를 기반으로 한 수학적 모델이 주로 사용되었으나 새로운 데이터에 대한 피드백이 부족하기 때문에 장기간의 데이터 예측에 한계가 있다. 또한 데이터의 특성이 다양하고 복잡한 경우에는 수학적 모델의 결합 및 계산과정이 어려워진다. 따라서 본 논문에서는 데이터의 학습 및 예측에 기존 정적 모델이 아닌 기계학습 중 시계열 데이터 분석 (Time Series Analysis) 을 기반으로 연구를 진행하였다. 기계학습은 복잡한 특성을 가진 데이터를 학습하여 미래의 데이터 값을 예측하거나 분류하는데 있어서 정확도 및 처리시간 측면에서의 성능을 향상시킬 수 있다.

키워드: 데이터 패턴(data pattern), 예측(prediction), 시계열(time series)

I. Introduction

데이터의 패턴 예측에 있어서 기존에는 통계에 의한 수학적 모델이 주로 사용되었으나, 정확도 측면에서의 한계가 있고 새로운 데이터에 대한 피드백이 부족하기 때문에 장기간의 데이터를 예측 하는데 한계가 있다. 또한 데이터의 특성이 다양하고 복잡한 경우에는 수학적 모델의 결합이 매우 복잡하며, 비선형의 데이터 패턴 예측에 한계가 있다 [1]. 기계학습은 복잡한 특성을 가진 데이터를 학습하여 미래의 값을 예측하거나 분류하는데 있어 성능이 우수하다. 최근 이러한 기계학습의 특성을 이용하여 데이터의 분석 및 예측에 사용되고 있다.

어 분석 및 예측하는 것이다. 시계열 분석에 사용되는 모델은 Autoregressive(AR) 모델, Moving Average(MA)모델, Autoregressive Moving Average(ARMA) 모델, Autoregressive Integrated Moving Average(ARIMA) 모델이 있다.

1.1.1 Autoregressive (AR)

AR 모델은 데이터에 대해 특정시간 전후의 데이터를 비교하여 상관관계가 있을 때 예측이 가능하다. 동일한 Series의 과거 데이터와 현재 Series의 데이터를 선형회귀로 표현한다. AR(P) 모델은 다음과 같이 정의된다.

$$Y_t = c + \sum_{i=1}^p \phi_i X_{t-i} + \epsilon_t$$

위 수식에서 ϕ_1, \dots, ϕ_p 는 매개변수, c는 상수, ϵ_t 는 백색 잡음이다.

1.1.2 Moving Average (MA)

MA 모델은 시간의 경과에 따라 Variable의 평균값이 지속적인

II. Preliminaries

1. Related works

1.1 Time series

시계열(Time series)은 일정 간격으로 수집된 데이터를 의미하며, 이러한 데이터의 분석에 사용되는 방법을 시계열 분석(Time series analysis) 이라고 한다. 시계열 분석은 시계열에서 일정 규칙을 찾아내

증가 또는 감소하는 경향이 발생하는 원리를 기본으로 한다. 과거의 예측 오류를 변수로 사용하며, 이동평균을 시계열로 표현한 것을 의미한다.

$$Y_t = \mu + \epsilon_t + \sum_{i=1}^q \theta_i \epsilon_{t-i}$$

위 수식에서 $\theta_1, \dots, \theta_q$ 는 매개변수, ϵ_t 는 백색잡음이며, μ 는 Y_t 의 기댓값이다.

1.1.3 Autoregressive Moving Average (ARMA)

AR 모델과 MA 모델을 결합한 것으로 광범위 한 시계열 데이터를 표현 할 수 있다. ARMA 모델은 현재의 관측점을 과거 ρ 까지의 자기회귀를 하고 난 후의 잔차에 대한 상관관계를 이용하여 백색잡음을 만든다. 시간 t 에서의 관측 값 Y_t 가 과거의 관측 값 $Y_{t-1}, Y_{t-2}, \dots, Y_{t-p}$ 와 과거 오차 $e_{t-1}, e_{t-2}, \dots, e_{t-q}$ 로 표현 될 때 ARMA(p,q)를 따른다.

$$Y_t = c + \epsilon_t + \sum_{i=1}^p \phi_i Y_{t-i} + \sum_{i=1}^q \theta_i \epsilon_{t-i}$$

1.1.4 Autoregressive Integrated Moving Average

ARMA 모델은 과거의 데이터만을 사용하는 반면, ARIMA 모델은 상관관계(Correlation)를 가지는 ARMA 모델에 데이터가 가지고 있는 추세(Momentum)까지 반영한 모델이다. 데이터의 추세관계란 두 변수 a, b 간의 추세관계가 0보다 클 때 시간 $t+1$ 에서의 a 값이 t 에서의 a 값에 비해 증가하면 b 값도 증가하며, 이와 반대로 $t+1$ 의 a 값이 이전 t 에서의 a 값보다 증가하면 b 값은 감소한다.

$$Y_t = a_1 Y_{t-1} + \epsilon_t$$

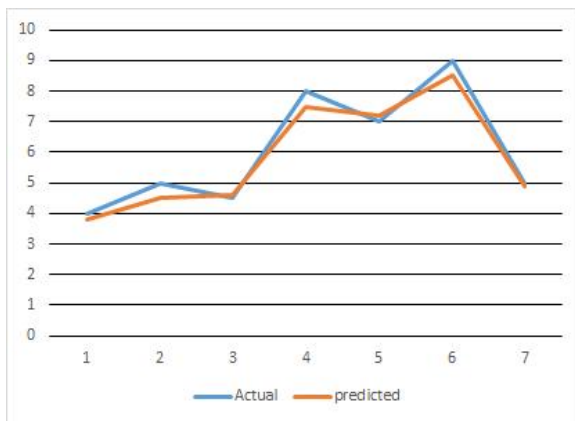


Fig. 1. The example of the ARIMA model

위 그림 1은 ARIMA 모델의 예제이다.

IV. Conclusions

본 연구에서는 기계학습을 이용한 데이터의 분석 및 패턴 예측기법에 대해 연구하였다. 기존 통계학을 기반으로 한 수학적 모델링 기법과 비교하였을 때 분석 및 예측 과정에서 정확도 및 속도 측면에서의 성능이 우수하다.

ACKNOWLEDGMENT

본 연구는 과학기술정보통신부 및 정보통신기술진흥센터의 정보통신-방송연구 개발 사업(No. 2016-0-00133, 초연결 IoT 노드의 군집 지능화를 통한 Edge Computing 핵심 기술 연구), SW중심대학지원사업(2015-0-00914), 한국연구재단 기초연구사업(No.2016R1A6A3A11931385, 실시간 공공안전 서비스를 위한 소프트웨어 정의 무선 센서 네트워크 핵심기술 연구, 2017R1A2B2009095, 실시간 스트림 데이터 처리 및 Multi-connectivity를 지원하는 SDN 기반 WSN 핵심 기술 연구), 삼성전자, BK21PLUS 사업의 일환으로 수행되었음.

REFERENCES

- [1] H. NIE et al. "Hybrid of ARIMA and SVMs for Short-Term Load Forecasting," International Conference on Future Energy, Environment, and Materials, pp.1455-1460, 2012