

수정된 ARIMA 기반 원격교육 학습자 전처리 연구

민연아^o, 백영태^{*}

^o한양사이버대학교 응용소프트웨어공학과,

^{*}김포대학교 멀티미디어과

e-mail: yah0612@hcu.ac.kr^o, hanna@kimpo.ac.kr^{*}

Modified ARIMA-based Distance Learning Learner Preprocessing Study

Min Youn A^o, Baek YeongTae^{*}

^oDept. of Applied SW Engineering, Hanyang Cyber University,

^{*}Dept. of Multimedia, Kimpo University

● 요약 ●

본 논문 원격교육환경에서 학습자가 남긴 개별 데이터에 대한 장기적 관리 및 효율적 학습자 관리를 위한 데이터 전처리 방법으로 전통적인 ARIMA를 수정하여 연구하였다. ARIMA는 과거시점 데이터에 대한 회귀식과 변화율을 현 시점 데이터에 반영하는 방식이며 본 연구에서는 ARIMA 처리과정에서 딥러닝 알고리즘인 RNN의 변형방법인 LSTM을 적용하여 부분 데이터셋의 전처리과정에 대한 정확성과 재현율을 높이기도 하였다. 본 연구의 결과 전통적인 ARIMA 적용시와 대비하여 7~9%의 성능향상을 확인하였다.

키워드: ARIMA, 원격교육, RNN, LSTM

I. Introduction

원격교육이란 교육자와 학습자가 비대면 환경에서 교육을 진행하는 것을 말하며 최근 다양한 환경적 요인에 의해 비대면을 통한 원격교육의 수요가 증가하고 있다[1]. 2022년 SW정책연구원의 ‘2021년 이력산업 실태조사’에 의하면 국내 원격교육의 수요는 2020년 대비 10%증가한 5조원의 수요시장 규모 추이를 형성하고 있으며 이 중 개인 주도의 학습이 대부분을 차지한다[2]. 본 논문에서는 원격교육을 통해 수집된 학습자 개인의 장기간 학습 이력 및 데이터에 대하여 효율적인 학습자 관리를 위한 방법으로 수정된 ARIMA분석을 통한 데이터 전처리를 연구하였다.

영향을 주도록 설계된 모형이다. AR의 식에서 $\theta(\theta)$ 는 자기상관계수이며 $\epsilon(\epsilon)$ 를 white noise로 하고 Time lag ≥ 1 의 조건에 따라 AR의 식을 $\theta * X(t-1) + \epsilon(t)$ 와 같이 나타낼 수 있다[4]. MA는 과거의 시계열적 데이터를 사용하여 그 안에서 데이터의 패턴을 찾아내므로 MA모형은 차수(p)에 따라 사용하는 데이터의 범위가 달라진다. 차수를 확장하여 n번째 시점까지 누적하여 적용하려면 차수를 n으로 설정하면 된다. β 는 이동평균계수, $\epsilon(\epsilon)$ 은 t시점의 오차항이다. $b(\beta)$ 는 이동평균계수이며 $\epsilon(\epsilon)$ 을 t시점의 오차항으로 설정할 때 MA의 식은 $\epsilon(t) - b_1 * \epsilon(t-1)$ 과 같이 나타낼 수 있다[3,4].

II. Preliminaries

1. Related works

1.1 ARIMA 분석

ARIMA는 Autoregressive Integrated와 Moving Average의 통합형태로 과거시점의 데이터에 대한 회귀식과 변화율을 현 시점의 데이터에 반영하는 방법이다[3]. ARIMA의 AR은 자기회귀(Autoregression)를 뜻하며 이전 관측한 값의 오차항이 이후 관측값에

1.2 LSTM과 데이터 전처리 방법

LSTM의 처리과정은 간단한 모듈로 처리되는 기본 RNN 처리와는 달리, 4개의 상호작용하는 레이어로 구성된 반복모듈로 처리가 되며 이를 통하여 과거의 데이터 소실에 대한 문제를 감소시킬 수 있다[5, 6]. 데이터 전처리를 위한 전통적인 방법으로는 누락데이터 및 불필요한 데이터를 제거하는 Data Cleaning, 데이터셋에 텍스트가 있는 경우 수치 데이터로 인코딩 하는 방법인 Handling Text and Categorical Attributes, 사이킷런의 다양한 데이터 변환기를 통해

데이터 전처리 하는 Custom Transformers, 수치 데이터로 바꿀 때 제대로 학습할 수 있도록 데이터의 범위를 줄여주는 방법인 Feature Scaling, 데이터 처리 컴포넌트들의 순서를 정의해 놓은 Transformation Pipelines 등을 들 수 있다[6].

III. The Proposed Scheme

전통적인 ARIMA 분석을 통해 현재 시점과 과거시점의 오차관계를 파악할 수 있다. 특히 차수(t)가 n일 경우, t-1 ~ t-n 시점까지의 오차관계 파악이 가능하여 시계열적으로 오차의 폭이 큰 데이터에 대한 이상치 처리가 가능하다.

본논문에서 제안한 수정된 ARIMA 처리과정은 Table 1과 같다.

Table 1. Proposed ARIMA processing process

<ul style="list-style-type: none"> • Collecting subset(individual learner's learning data) • Cell-state processing for each subset (a) • Filtering and processing of missing values in consideration of learning progress rate and existing learning history for prominent outliers for one element – about subset • Manage only with meaningful elements

Table 2는 전통적인 ARIMA(A)과 본 논문에서 제안한 수정된 ARIMA(B)에 대한 평가결과에 대한 정확도(Accuracy), 재현율(Recall), 정밀도(Precision)를 비교한 것이다.

Table 2. Performance evaluation comparison

Divide	Accuracy	Recall	Precision
ARIMA(A)	0.07	0.61	0.05
ARIMA(B)	0.75	0.67	0.53

IV. Conclusions

본 연구에서는 원격교육에서 효율적 학습자 전처리 및 관리를 위하여 ARIMA 분석과정을 수정하여 연구하였다. 본 연구를 통하여 전통적인 ARIMA 대비 8-9% 정도의 성능향상을 확인하였다. 향후 보다 다양한 크기와 형상의 데이터셋을 대상으로 하여 정밀도 부분의 성능향상이 가능하도록 연구할 계획이다.

REFERENCES

[1] Eun-Young Oh, "Mediating Effect of Need of Cognition on Learning Engagement in Distance Education", Journal of Convergence for Information Technology, Vol. 12, No. 5, pp. 51-61, 2022, DOI :<https://doi.org/10.22156/CS4SM B.2022.12.05.051>

[2] 2021 Survey of Korean e-learning industry, <https://spri.kr/>

posts/view/23212?code=research

[3] Mo, Soo-Won, "Forecasts of the 2011-BDI Using the ARIMA-Type Models", Journal of Korea Port Economic Association, Volume. 26, No. 4, pp. 207-218, 2010.

[4] Chu, F.L. et al., "Forecasting Tourism Arrivals: Nonlinear Sine Wave or ARIMA", Journal of Travel Research, Vol. 36, 1998, pp. 79-84, DOI : <https://doi.org/10.1177/004728759803600309>

[5] Noh, Young-Dan, et al., "A Text Content Classification Using LSTM For Objective Category Classification", Journal of the Korea Society of Computer and Information, Vol.26, No.5, pp.39-46, 2021, DOI : <https://doi.org/10.9708/jksci.2021.26.05.039>

[6] Hong, Sungjae et al., "Development of Surface Weather Forecast Model by using LSTM Machine Learning Method", Atmosphere , Vol.31, No.1, pp.73-83, 2021, DOI:<https://doi.org/10.14191/Atmos.2021.31.1.073>