

LSTM을 이용한 도서관 이용자 예측

김 경 섭

(동국대학교 컴퓨터공학과 연구원, 010-5451-9846, kyskks@doguk.edu)

김 택 수

(제주대학교 경제학과, 010-8638-1097, mario109745@gmail.com)

김 정 옥

(대덕대학교 컴퓨터공학과, 010-2715-0235, qpitem@gmail.com)

채 성 민

(한밭대학교 컴퓨터공학과, 010-4645-5450, fire5450@naver.com)

요 약(영문 또는 국문)

사회에는 경제활동, 물리적 현상, 경영활동, 인구, 생산관리 그리고 사회생활과 관련된 사회 전반에 시계열 자료들이 존재한다. 본 연구에서는 도서관 이용자 혼잡도 해결을 위해 경제학에서도 매우 많이 쓰이는 방법론인 ARIMA(Autoregressive Integrated Moving Average Model) 모델과, 최근 다양한 시계열연구에서 사용되어진 딥 러닝 모델 LSTM과 비교해 성능을 테스트 한다.

Key Words : prediction, deep learning, machine learning, LSTM, ARIMA

I. 서 론

전국 도서관 연간 평균이용자 수는 2억 8천만 명 규모이다. 반면 전국 1관 당 평균 좌석 수는 해마다 감소하고 있다. 좌석 수 감소에 따라 도서관은 이용 혼잡도가 올라갈 것으로 예상된다. 미래예측을 하는 것은 도서관을 찾는 사람들에게 도움이 될 수 있다.

이로 인해 도서관 이용자 수 예측은 이용자에게 도서관 혼잡도를 제시함으로써 도서관 이용의 편의를 제공 할 수 있다. 도서관 좌석 수와 이용자를 공급과 수요로 생각하여 수요 예측 알고리즘을 사용한다. 기존 수요 예측에 자주 사용되는 ARIMA를 사용한다. ARIMA는 대표적인 시계열 분석 방법론으로 자기회귀 모형과 이동평균 모형을 혼합한 것에 차분 절차를 포함하여 비정상 시계열 자료를 정상화시켜 분석하는 모형이다.[1] 그리고 수요 예측 알고리즘인 LSTM(Long Short-Term Memory)은 딥러닝 RNN 알고리즘의 데이터 양이 많을 때 오랜 과거의 시계열 데이터가 반영이 안 되는 문제점이 생기는데 이를 해결하기 위해 개발된 알고리즘이다.[2]

II. 관련 연구

ARIMA모델은 시계열 분석 기법의 변동형태를 파악하고 이를 통해 예측이 가능하다는 장점으로 증권시장 등 경제 분야와 수산자원 관리에 많이 응용 되고 있다. 단일변량 ARIMA 분석 기법의 특징은 어떤 시계열에도 적용이 가능하며 시간의 흐름에 따라 자료의 변동이 빠를 때 민감하게 반영할 수 있다.[3] 시계열 분석 기법의 한 종류로, 과거의 데이터들을 사용하는 것

을 넘어서서 과거의 데이터가 지니고 있는 추세(Momentum)까지 반영 한다.

본 논문에서는 사용 한 데이터는 2017.01부터 2019.07 까지 국립세종도서관 일별 이용자수 데이터를 대상으로 했으며, 이용자수와 관련된 인자들을 사용해 분석하였다. 2017년 1월부터 2019년 6월까지를 기계학습을 위해 사용하였고, 나머지 1개월을 예측 결과의 정확성을 확인하기 위해 사용하였다. 그리고 LSTM 모델을 사용하여 도서관이용자의 수를 예측하는 것을 제안 한다.

III. LSTM 모델 적용

LSTM은 Hochreiter와 schmidhuber(1997)에 의해 제안된 방법으로 RNN의 기울기 손실문제를 해결할 수 있다. 셀 스테이트(cell state)를 통해 과거 학습결과를 전달하는 구조로 장기 의존성 문제를 해결하는 부분은 LSTM 기법의 핵심이라 할 수 있다. 셀 스테이트는 3개 요소인 망각게이트(forget gate), 입력 게이트(input gate), 출력게이트(output gate)로 구성된다. 이 요소들을 활용하여 정보를 더하거나 제거하는 기능을 수행하고, 선택적으로 정보를 반영할 수 있다.

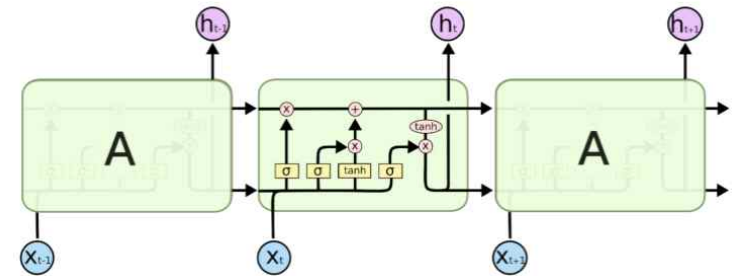


그림 1 structure of LSTM(chritopher, 2015)

LSTM은 순환 신경망을 기반으로 하는 모델로써 여러 개의 게이트(gate)가 붙어있는 셀(cell)로 이루어져 있으며 이 셀들이 하는 역할은 정보를 버리고, 저장하고, 업데이트하고, 내보내는 기능들로 이루어져 있다. 각 셀은 셀에 연결된 게이트의 값을 보고 무엇을 저장할지, 언제 정보를 내보낼지, 언제 쓰고 언제 지울지를 결정한다. LSTM은 이러한 셀 구조를 가짐으로써 순환 신경망의 대표적인 문제인 기울기 소실 및 발산 문제와 장기 의존성 문제를 해결할 수 있다. [4]

데이터는 국립세종도서관에서 이용자수 데이터를 2017년 1월부터 2019년 7월 데이터를 제공 받았고, 이용자수와 관련된 요일, 휴일, 휴관일, 날씨(습도, 온도, 강수량, 풍속)를 사용하여 분석하였다.

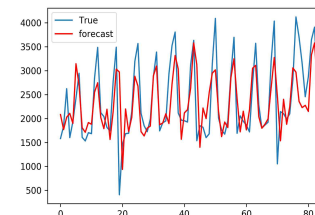


그림 2 Result of ARIMA Model

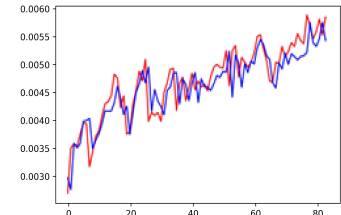


그림 3 Result of LSTM

그림 2와 그림 3은 x축은 시간 y축은 방문자수이다.

회귀문제는 입력 값이 고차원인 특징 영역에서 나타난다. RMSE는(Mean Square Error)MSE의 절대 값인 표준 에러로 성능에 대한 에러율을 확인함으로써 회귀문제를 해결할 수 있다.[5]

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^n (p_i - y_i)^2} \quad (1)$$

pi=yi일 경우 100%의 성능을 갖게 되며 이 때 RMSE는 0이 된다. 즉, RMSE가 작을수록 성능은 향상 된다.

예측 결과와 실제 결과의 차이를 절댓값을 이용해서 구한 후 평균을 구한 값이다. MAPE는 퍼센티지 (percentage)로 표현되어 이해가 쉽다는 장점이 있지만, At와 Ft의 차이가 클 경우 100%를 넘는 오차 율이 나오는 단점이 있다. 또한 At을 포함하는 데이터는 비교 대상에 포함할 수 없다는 단점이 존재한다.[6]

$$MAPE = \frac{100}{N} \sum_{t=1}^n \left| \frac{A_t - F_t}{A_t} \right| \quad (2)$$

각각 ARIMA 모델과 LSTM의 정확도를 확인하기 위해 방정식 1로 표시된 MAPE(평균 절대 백분율 오류)와 방정식 2로 표시된 RMSE(루트 평균 제곱 오류)를 사용했다[6]. 또한 RMSE(root mean squared performance)의 검정까지 확인하여 예측 값의 정확도를 파악하였다.[7]

ARIMA와 LSTM의 모델 예측 정확도 결과는 표 1과 같다.

	MAPE	RMSE
ARIMA	37.72 %	593.058
LSTM	3.69 %	0.011

표 1 도서관 이용자 예측 모델 정확도

ARIMA, LSTM 모델의 MAPE 측정 결과는 37.72%, 3.69%로 측정되었다. RMSE의 결과는 593.058, 0.011로 각각 측정 되었다.

IV. 결 론

본 연구는 도서관의 동시 이용자 포화상태를 해결하기 위해 이용자수 예측모델을 만들었으며, ARIMA 모델과 LSTM 모델의 적용 구현하였다. LSTM 모델의 경우 도서 대출 건수와 날씨를 변수로 설정하였다. 실험 결과는 모델들의 그래프와 함께 제시했다. 예측 결과는 LSTM모델이 MAPE, RMSE 지표에서 향상된 성능으로 측정 되었다. 향 후, 시간대별 데이터를 제공 받았으면 용도 가치가 더 높은 이용자 수 예측을 할 수 있겠다는 생각이 들었다.

참 고 문 헌

- [1] 유정아, 노병준, 장성주, “ARIMA 모형을 이용한 일일 건물 난방 전력 수요 예측”, 2017.
- [2] 정호철, 선영규, 이동구, 김수현, 황유민, 심익삭, 오상근, 송승호, 김진영, ”에너지인터넷에서 1D-CNN 과 양방향 LSTM을 이용한 에너지 수요예측“ 2019.

[3] Box, George EP, et al. Time series analysis: forecasting and control. John Wiley & Sons, 2015.

[4] 김형래, “LSTM 기반의 강남역 시간별 지하철 탑승 인원 예측”, 한국정보과학회 학술발표논문집, 2019.

[5] 김은미, 이배호, “모멘트의 동적 변환에 의한 Kernel Relaxation의 성능과 RMSE.” 멀티미디어학회논문지, 6(5), 788-796. 2003.

[6] 김한호, 탁해성, 조환규. “순환신경망을 이용한 태양광 발전량 예측 모델 설계”, 정보과학회논문지, 46(6), 506-514. 2019.

[7] 임찬영, 김희철. (2019). ARIMA 모형을 활용한 예금은행 주택담보대출 분석 및 예측 연구. 한국정보전자통신기술학회 논문지, 12(3), 265-272.