

관광업종에 따른 인공지능 시계열 모델 분석 연구*

김선미⁰¹ 공희산² 신성국³ 김동희² 김광수²

¹성균관대학교 인공지능융합학과 / ²성균관대학교 소프트웨어학과 / ³성균관대학교 인공지능학과
seonmikim@g.skku.edu, khtks11@g.skku.edu, davidshyn@skku.edu, ym.dhkim@skku.edu, kim.kwangsu@skku.edu

A Study on the Analysis of Artificial Intelligence Time Series Model by Tourism Industry

Seonmi Kim⁰¹ Heesan Kong² Sung Kuk Shyn³ Donghee Kim² Kwangsu Kim²

¹Department of Applied Artificial Intelligence, Sungkyunkwan University

²Department of Computer Science and Engineering, Sungkyunkwan University

³Department of Artificial Intelligence, Sungkyunkwan University

요약

시계열 예측은 금융, 의료, 제조, 관광 등 다양한 분야에서 활용 가능하다. 코로나19 확산 및 장기화에 따라 미래 성장 동력이라고 평가되던 관광산업이 외생변수에 취약한 것이 밝혀지면서, 데이터에 기반한 예측 및 선제 대응의 중요성이 부각되고 있다. 국내 대형 카드사의 관광업 결제 데이터를 분석해본 결과, 코로나19 이전과 이후 다른 양상을 나타냈으며, 관광업이라는 카테고리가 같다 하더라도 업종별로 다른 소비패턴을 보이고 있다. 본 연구에서는 코로나19로 인해 소비패턴이 다르게 나타난 골프장과 항공운송업에 대해 통계방식의 모델과 인공지능 방식의 모델을 적용하여 소비패턴을 예측하고, 예측한 정확도를 바탕으로 데이터 추이에 따른 모델별 특성을 비교하여 실무에 적용할 수 있을지 시사점을 제공한다.

1. 서론

코로나19 확산 및 장기화에 따라 미래 성장동력이라고 생각했던 관광산업이 외생변수에 얼마나 취약한지를 경험하게 되면서, 데이터에 기반한 예측 및 선제 대응의 중요성이 부각되고 있다. 관광 수입 규모는 그림 1과 같이 2019년 207.5억 달러로 규모가 계속해서 커지고 있었지만, 2020년 발생한 코로나19로 인해 줄어든 관광 수입 규모는 2021년 기준, 2010년 수준으로 회귀해 있는 상태이다[1].

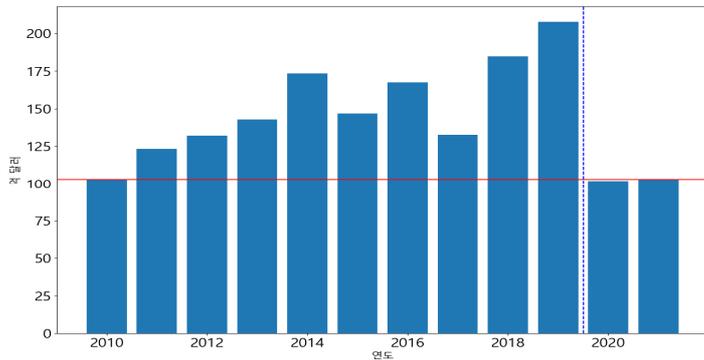


그림 1 2010년-2021년 관광 수입 규모

코로나19로 인해 예상하지 못했던 관광업계의 타격을 높은 정확도의 수요예측을 통한 사업전략 수립으로 코로나19 이전의 관광 수입 규모를 회복하고, 관광객 집계방

식으로 예측하던 관광수요를 실제 관광업종 결제 데이터에 기반하여 예측함으로써 업종별 마케팅 전략 수립에 기여하고자 하였다. 특히 관광업종 중, 코로나19 상황이었던 2020년 결제 규모가 2019년 대비 약 16% 증가한 골프장과 약 70% 감소한 항공운송업 결제금액 데이터에 적용해 보았다.

시계열 예측에는 전통적인 시계열 알고리즘도 많이 활용되지만, 컴퓨팅 리소스의 성능이 좋아지고, 활용할 수 있는 데이터가 증가함에 따라 딥러닝을 이용한 시계열 예측 알고리즘에 대해 많은 연구가 진행되고 있다[2]. 본 연구에서는 시계열 데이터 예측을 위하여 다양한 모델을 활용하고 비교하면서 그 결과를 적용할 수 있을지 확인하고자 한다.

2. 관련 연구

시계열 예측은 중요한 연구분야 중 하나로 ARIMA(Autoregressive Integrated Moving Average)[3], SVM(Support Vector Machine)[4], Neural Network 등 다양한 방법들이 있다.

이 중 AR, ARIMA 등 전통적 시계열 분석 방법들은 데이터를 시간의 흐름과 관계없이 평균과 분산이 일정한 정상성을 전제로 한다. 하지만 현실에서 생산되는 데이터는 보통 정상성을 가지고 있지 않고, 시계열 특징을 충분히 추출하기 어렵기 때문에 예측 정확도가 제한되며, 특히 갑작스러운 변동이 있는 데이터 세트에서는 더욱 그러한 경향을 보인다[5]. 그래서 데이터에 차분 또는 로그 변환의 전처리를 통해 정상성을 갖도록 규칙성을 부여한 후 분석하게 된다.

이와는 다르게 SVM, Neural Network 등 인공지능 기

* 이 논문은 2022년도 정부(과학기술정보통신부)의 재원으로 정보통신기획평가원의 지원을 받아 수행된 연구임 (No. 2020-0-00973, VR·AR 콘텐츠 비가시 영역 영상 복원 기술 개발)

반의 시계열 예측 방식은 규칙성을 부여하지 않은 원 데이터로부터 학습하는 과정을 통해 자동으로 데이터의 특징을 추출하고 개선해 나간다. 특히 최근에는 여러 가지 기법들을 통해 길이가 긴 데이터에서도 특징을 잘 추출하는 방식으로 발전하고 있다. 이렇듯 데이터의 특징을 반영하는 방법론을 사용함으로써 심층 신경망은 복잡한 데이터 표현을 학습할 수 있어 수동 엔지니어링 및 모델 설계의 필요성을 완화할 수 있다[6].

3. 연구결과

본 논문에서 활용한 데이터는 2021년 사업보고서[7] 기준 국내 점유율 1위 신한카드와 국내 점유율 8위 비씨카드로, 2018년부터 2022년 7월까지 데이터를 정규화하여 활용하였으며, 업종은 해당 카드사의 분류기준을 참고하였다. 또한, 데이터의 특성에 따른 결과를 비교해보고자, 관광업종 중에서도 코로나19 시기에 일정하게 증가했던 골프장 결제 데이터와 같은 시기에 가장 타격을 많이 받은 항공운송 결제 데이터를 비교함으로써 데이터의 추세 변화에 따른 모델들의 결과 차이도 확인할 수 있었다. 활용한 모델로는 전통적인 시계열 모델인 ARIMA와 인공지능 기반의 Prophet[8], 그리고 최근 다양한 분야에서 좋은 성능을 보여주고 있는 Transformer[9] 모델을 활용하였다. 성능에 대한 평가는 실제값과 예측값과의 차이를 제공하여 평균을 구하는 MSE(Mean Squared Error)를 사용하여 확인하였다.

3.1. ARIMA 예측 결과

전통적인 시계열 예측 알고리즘으로 활용한 ARIMA 모델은 먼저 데이터 분석을 통해 데이터가 정상성을 갖고 있는지를 확인하고 규칙성을 부여한 후, 모델에 입력할 필요가 있다. 따라서 두 업종의 결제 데이터를 추세, 계절, 순환, 불규칙 요인으로 분해하여 분석해 본 결과, 평균과 분산이 일정하지 않았기 때문에 로그와 차분을 적용해 정상성을 부여하였다. 예측된 값을 MSE로 평가한 결과, 안정적인 주기를 가지는 골프장 데이터는 성능이 0.003, 변동이 심했던 항공운송업은 0.019의 오차로 데이터 추이에도 좋은 성능을 보여주었는데, 이는 데이터에 규칙성을 부여하고 예측했기 때문이라고 해석된다. 결과를 나타낸 그림 2, 3에서 파란색 선은 실제 데이터, 주황색 선은 모델이 예측한 값을 나타낸다.

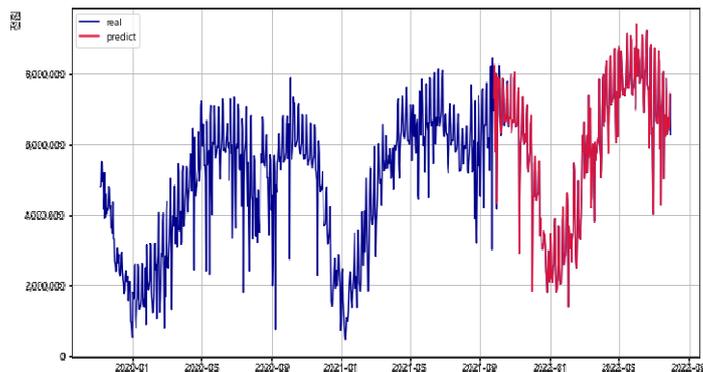


그림 2 ARIMA로 예측한 골프장 결제금액

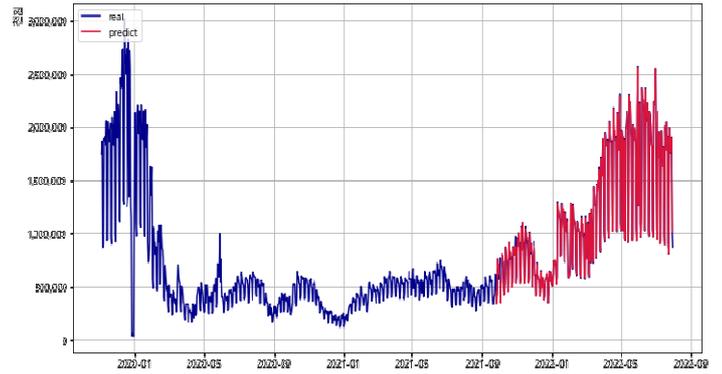


그림 3 ARIMA로 예측한 항공운송 결제금액

3.2. Prophet 예측 결과

Prophet은 시계열을 추세, 주기, 휴일 요소로 분해하고 각 요소를 비선형관계로 추론하는 일반화가법모형의 아이디어를 사용하여, 직관적인 파라미터로 모델 수정이 쉽고 계절성 데이터에 강한 장점을 가지고 있다. 또한, 다른 모델과 다르게 입력으로 요구되는 값이 일자와 일자에 해당하는 값 2개뿐이라는 점이 특징인데, 다루기 쉬우면서도 높은 정확도의 Prophet 특징을 이용해 따로 피쳐를 수정하지 않고도 그림 4, 5와 같은 결과를 확인할 수 있었다. 파란색 선은 실제 데이터, 주황색 선은 모델이 예측한 값을 나타내며, MSE 평가 결과, 골프장 데이터는 0.007, 항공운송업은 0.016의 오차를 나타냈다. 이를 통해 Prophet 모델은 데이터가 주기성을 가지고 있을 경우, 더 높은 예측 정확도를 보이는 것을 알 수 있다.

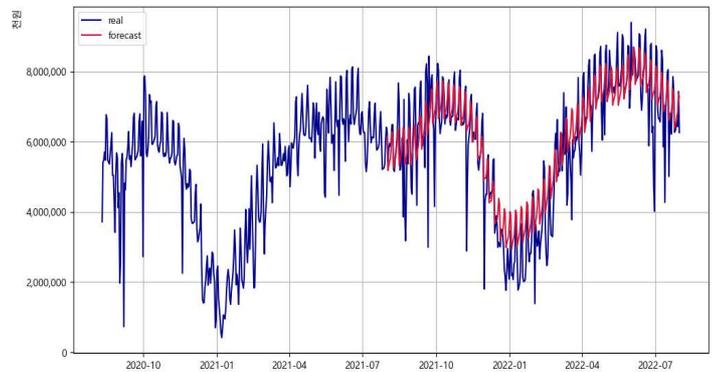


그림 4 Prophet으로 예측한 골프장 결제금액

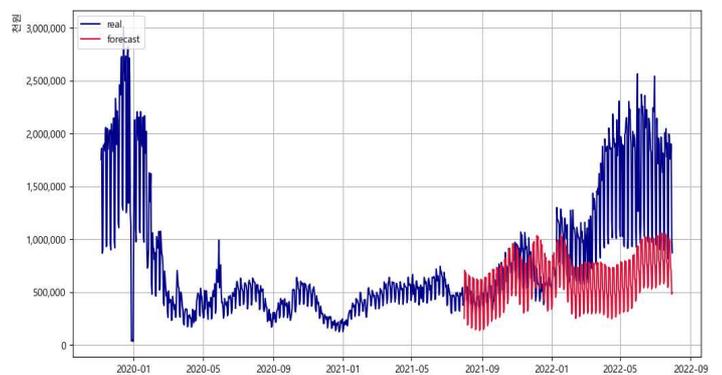


그림 5 Prophet으로 예측한 항공운송 결제금액

3.3. Transformer 예측 결과

Transformer 모델은 데이터의 길이가 길어져도 중요한 부분의 데이터에 더 많은 가중치를 부여하는 방식을 사용함으로써 시계열에서도 좋은 성능을 보여주고 있다. 본 실험에서는 다양한 크기의 Input Window를 Sliding 시키면서 실험하였는데, 각각 10, 2로 설정하였을 때가 31, 16 등 크게 설정했을 때보다 좋은 성능을 보여주었다. MSE 평가 결과, 골프장 데이터는 0.016, 항공운송업은 0.013으로 가장 좋은 결과를 볼 수 있었는데, 이는 데이터를 치밀하게 학습할 수 있었기 때문이라고 해석된다. 그림 6, 7의 파란색 선은 실제 데이터, 주황색 선은 모델이 예측한 값을 나타내며, Transformer의 경우 데이터 추이와 관계없이 비슷한 예측 정확도를 확인할 수 있었다.

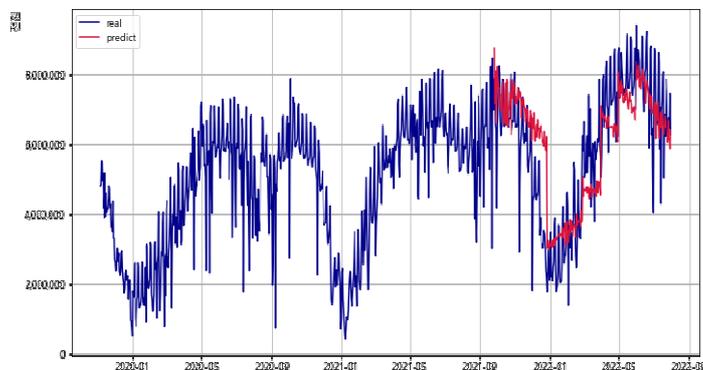


그림 5 Transformer로 예측한 골프장 결제금액

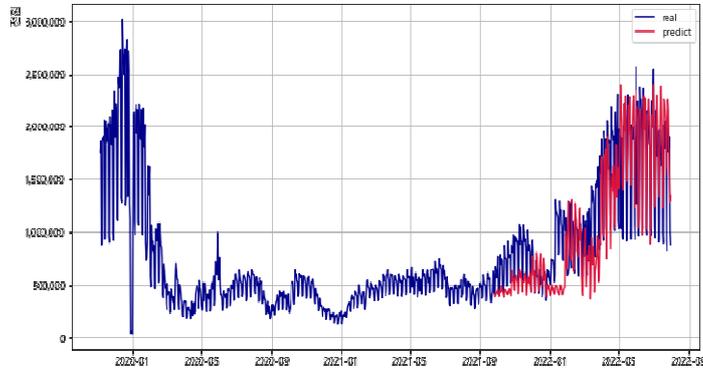


그림 6 Transformer로 예측한 항공운송 결제금액

4. 결론 및 시사점

본 연구에서는 4년간의 실제 관광업종 결제금액을 시계열 데이터로 활용하고 다양한 모델을 통하여 1년간의 업종별 소비를 예측함으로써, 관광객 집계를 통한 관광 수요 예측보다 관광업종별 직관적인 예측을 시도하였다. 연구의 결과는 다음과 같다. 첫째, 전통적인 시계열 예측과 인공지능 모델을 비교했을 때, 데이터에 대한 분석이 선행되고 정상성을 가지도록 규칙성을 부여한다면, 인공지능 모델보다 좋은 성능을 나타냈다. 둘째, Prophet 모델은 계절성 데이터에 강한 점이 특징인 만큼 데이터에 주기가 일정하다면 특별한 Hyperparameter 변경없이도 좋은 성능을 보여주었지만, 데이터의 추이가 갑자기 변하였을 때는 예측 추이가 떨어져 Hyperparameter 변경

등 추가 변수 수정이 필요하다는 것을 알 수 있었다. 셋째, Transformer는 데이터 추이 변화에 큰 영향을 받지 않고 비슷한 정확도를 확인할 수 있었다.

결론적으로 시계열 예측시, 전통적인 방식의 ARIMA 모델은 데이터 분석 및 가공이 필수적이었고, 인공지능 방식의 Prophet과 Transformer 모델은 원 데이터를 가공 없이 사용할 수 있어서 편리했으나, 데이터 특성에 따른 모델별 정확도 차이를 확인할 수 있었다.

해당 연구는 실제 코로나19 시기의 관광업종 결제 데이터를 활용해 시계열 예측함으로써, 데이터 변동 추이에 따른 다양한 모델별 예측 정확도를 비교해보았다는 것에 의의가 있다. 다만, 활용한 데이터는 특정 카드사 데이터를 활용한 것으로, 해당 데이터를 기반으로 모수 추정 하거나 국내 모든 카드사의 합산금액을 활용한 것이 아니기 때문에 특정 카드사의 주 사용자층, 할인 혜택 등으로 인한 소비편향이 있을 수 있다는 점이 한계이다. 나아가 본 연구에서 수행했던 업종별 예측을 지역별, 업종별로 범위를 세분화해서 예측하고, 추가 실험을 통해 일반화할 수 있는 모델을 확인한다면 실무에 더 많은 도움을 줄 수 있을 것으로 기대한다.

참고 문헌

- [1] e-나라지표(<https://www.index.go.kr>, 통계청).
- [2] 최영준, 김대근, 딥러닝 기반 시계열 예측 알고리즘의 모델 크기에 따른 성능 비교 및 분석, 한국통신학회(하), 657-659, 2022.
- [3] George E. P. Box, Gwilym M. Jenkins, Time Series Analysis, Forecasting and Control, Journal of Marketing Research, 1977.
- [4] M.A. Hearst, Support vector machines, IEEE Intelligent Systems, 18-28, 1998.
- [5] Zhipeng Shen, A novel time series forecasting model with deep learning, Elsevier, 2019.
- [6] Bryan Lim, Stefan Zohren, Time-series forecasting with deep learning:a survey, royalsocietypublishing, 2021.
- [7] 전자공시시스템(<https://dart.fss.or.kr>, 금융감독원).
- [8] Sean J. Taylor, Benjamin Letham, Forecasting at Scale, Peerj, 37-45, 2017.
- [9] Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N Gomez, Lukasz Kaiser, and Illia Polosukhin, Attention Is All You Need, In NeurIPS, 2017.