

국내 외래객 출입국 데이터를 활용한 관광객 일별 수요 예측 인공지능 모델 연구

김동건^{*1} · 김동희² · 장승우³ · 신성국⁴ · 김광수¹

¹성균관대학교 소프트웨어학과 · ²Hippo T&C · ³성균관대학교 전자전기공학과 ·

⁴성균관대학교 인공지능학과

A Study on Artificial Intelligence Model for Forecasting Daily Demand of Tourists Using Domestic Foreign Visitors Immigration Data

Dong-Keon Kim^{1,*} · Donghee Kim² · Seungwoo Jang³ · Sung Kuk Shyn⁴ · Kwangsu Kim¹

¹SungKyunKwan University College of Computing ·

²Hippo T&C · ³SungKyunKwan University Department of Electrical and Computer Engineering ·

⁴SungKyunKwan University Department of Artificial Intelligence

E-mail : kdk1996@skku.edu / ym.dhkim@skku.edu / se93wo@gmail.com / davidshyn@gmail.com /

kim.kwangsu@skku.edu

요 약

외래 관광객 수요를 분석하고 예측하는 것은 관광 정책을 수립하고 기획하는데 지대한 영향을 미치기 때문에 관광 산업 분야에서 매우 중요하다. 외래 관광객 데이터는 여러 외적 요인들에 의해 영향을 받기 때문에, 시간에 따른 미세한 변화가 많다는 특징을 갖는다. 따라서, 최근에는 관광객 입국자 수요를 예측하기 위해 경제 변수 등 여러 외적 요인들도 함께 반영하여 예측 모델을 설계하는 연구를 진행하고 있다. 그러나 기존의 시계열 예측에 주로 사용되는 회귀분석 모델과 순환신경망 모델은 여러 변수들을 반영하는 시계열 예측에 있어 좋은 성능을 보이지 못했다. 따라서 우리는 합성곱 신경망을 활용하여 이러한 한계점들을 보완한 외래 관광객 수요 예측 모델을 소개한다. 본 논문에서는 한국관광공사에서 제공한 과거 10개년 외래 관광객 데이터와 추가적으로 수집한 여러 외적 요인들을 입력 변수로 반영하는 1차원 합성곱 신경망을 설계하여 외래 관광객 수요를 예측하는 모델을 제시한다.

ABSTRACT

Analyzing and predicting foreign tourists' demand is a crucial research topic in the tourism industry because it profoundly influences establishing and planning tourism policies. Since foreign tourist data is influenced by various external factors, it has a characteristic that there are many subtle changes over time. Therefore, in recent years, research is being conducted to design a prediction model by reflecting various external factors such as economic variables to predict the demand for tourists inbound. However, the regression analysis model and the recurrent neural network model, mainly used for time series prediction, did not show good performance in time series prediction reflecting various variables. Therefore, we design a foreign tourist demand prediction model that complements these limitations using a convolutional neural network. In this paper, we propose a model that predicts foreign tourists' demand by designing a one-dimensional convolutional neural network that reflects foreign tourist data for the past ten years provided by the Korea Tourism Organization and additionally collected external factors as input variables.

키워드

Leverage Inbound tourist data, Convolutional neural network, Time Series Forecasting, Deep Learning

I. 서 론

한국의 높은 경제 수준과 더불어 K-POP과 같은 한류 콘텐츠의 영향을 받아 한국을 관광 목적으로 방문하는 외국인 입국자 수가 매년 증가하고 있고, 이로 인해 해를 거듭할수록 관광 산업의 중요성은 더욱 두드러지고 있다. 그러므로 외국인 관광객 수요를 분석하고 예측하는 것은 관광 정책을 수립하고 기획하는 데 큰 영향을 미치기 때문에 관광 산업 분야에서 매우 중요한 연구 주제다.

한국에 오는 외국인 입국자 수는 한류 유행에 따른 한국에 관한 관심도 증가, 2020년 코로나-19와 같은 국제적 전염병 추세, 사드(THAAD) 배치에 따른 2017년 한·중 외교 갈등과 같은 정치적 상황 등과 같은 불확실한 외적 요인들의 영향을 직·간접적으로 받는다. 그래서 외국인 입국자 데이터는 추세가 다소 불규칙하고, 시간에 따른 미세한 변화가 많다는 특징을 가지고 있다.

본 논문에서는 이러한 불규칙하고 미세한 변화가 많은 외국인 입국자의 향후 추이를 예측할 수 있는 1차원 합성곱 신경망(1D-CNN) 기반의 외국인 입국자 예측 모델을 제안한다. 설계한 1D-CNN 모델은 과거 외국인 입국자 추세 외에도 입국자 수 변화에 영향을 주는 외적 요인들의 패턴을 합성곱 신경망을 통해 특성을 추출하고, 이를 바탕으로 정확도 높은 외국인 입국자 수를 예측한다.

II. 활용 데이터

2.1 외국인 입국자 데이터

한국에 입국하는 외국인 입국자 데이터는 한국관광공사에서 제공한 2010년 1월 1일부터 2020년 9월 30일까지의 21개 국가의 방한 외래객 데이터를 활용한다.

2.2 외적 요인 데이터

정확도 높은 외국인 입국자 예측을 위해 과거 외국인 입국자 통계뿐만 아니라 한국 관광에 영향을 주는 외적 요인도 입력 변수로 활용한다. 사용한 외적 요인 변수 요약은 표 1에 제시되어 있다.

정치 요인 한국의 외교·정치적 상황을 대표하는 외적 요인 변수로 대(對)중 정치 이슈인 한한령을 선정한다. 한한령을 반영하는 변수는 더미 변수(Dummy Variable)로 설정하고, 한한령이 유효한 기간인 2017년 3월 1일부터 2020년 9월 30일까지는 1, 그 외의 기간에는 0으로 변수값을 설정한다.

계절 요인 외국인 입국자 수 추이는 1년 주기로 계절성(Seasonality)이 뚜렷하기 때문에 계절에 대한 변수도 입력 변수로 반영한다. 계절 변수는 [봄, 여름, 가을, 겨울]의 총 4차원의 벡터를 가지는 더미 변수로 설정하였으며, 각 계절 기간에는 1, 그

외의 기간에는 0으로 변수값을 설정한다.

질병 요인 세계적 전염병 상황을 대표하는 외적 요인 변수로 메르스, 코로나-19 발생 기간에 따른 더미 변수 및 Google, Apple에서 제공하는 코로나-19 유행 기간 이동성 지수를 선정한다.

한류 요인 한국 관광 및 한류를 대표하는 외적 요인 변수로 Google Trend를 선정하고, 키워드는 관광과 연관이 있는 “Seoul Hotel”, “Korea Tour”, “Incheon Airport”, “Myeongdong”을 선정한다.

요인	변수명	유형	비고
정치	한한령	더미변수	
계절	계절	더미변수	사계절
질병	질병	더미변수	코로나, 메르스
	이동성 지수	수치형	
한류	구글 트렌드	수치형	

표 1. 입력 변수로 활용한 외적 요인 변수 요약

III. 예측 모델

3.1 문제 정의

본 논문에서는 과거 30일 동안의 정치, 계절, 질병, 한류 총 4개의 외적 요인들의 데이터와 입국자 데이터를 입력 변수로 가지고, 현재 시점에서 향후 30일 동안의 외국인 입국자 예측 데이터를 출력으로 하는 예측 프레임워크를 구성한다. 예측 모델에 입력되는 변수와 결과 출력에 대한 도식화된 그림은 그림 1에 제시되어 있다.

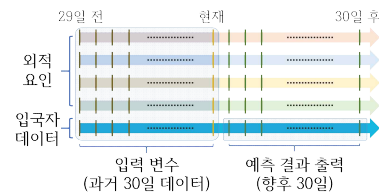


그림 1. 예측 모델 입력 변수 및 출력 결과 요약

3.2 1D-CNN 예측 모델

본 논문에서는 외국인 입국자 예측 모델로 일차원 합성곱 신경망(1D-CNN) 모델을 설계한다. 일차원 합성곱 신경망은 필터를 통해 시계열 데이터의 특징한 패턴이나 주기, 계절성을 두드러지게 할 수 있다는 장점이 있다[1]. 제안한 예측 모델은 본 논문에서 사용하는 주 데이터인 외국인 입국자 수 데이터와 같이 다양한 패턴을 가지고 있는 데이터를 처리하기에 적합하다. 모델의 전체적인 구조는 그림 3에 제시되어 있다.

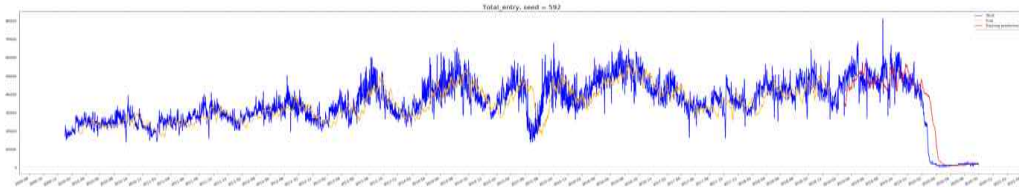


그림 2. 제시한 예측 모델 출력 결과 그래프

데이터에 대한 예측값은 주황색, 테스트 데이터에 대한 예측값은 빨간색으로 표시되어 있다.

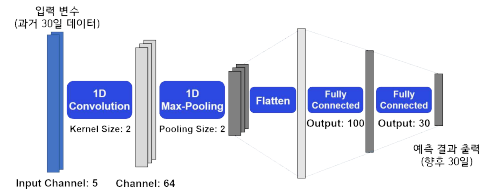


그림 3. 외국인 입국자 수 예측을 위한 1D-CNN 모델

표 2. 예측 모델의 실험 결과 (MAPE)

CNN	LSTM	양방향 LSTM
0.17%	3%	5.1%

그림 2의 테스트 데이터에 대한 예측 결과를 보면 일별 세부적인 패턴을 예측하는데 정확도가 다소 떨어지지만, 개략적인 주기와 추세, 그리고 우측의 코로나-19 발생 기간의 입국자 추이를 정확도 높게 따라가고 있는 것을 확인할 수 있다.

표 2의 결과를 보면, 순환 신경망 계열의 모델보다 합성곱 신경망을 활용하는 예측 모델의 예측 정확도가 더 우수함을 보인다. 그런데 테스트 기간 전체 오차에 대한 평균을 구하는 MAPE 산출식 특성상 그림 2에 보이는 실제 일일 예측 결과보다 더 좋은 결과를 보인다. 이는 제시한 모델이 각 일별 입국자 예측 정확도가 떨어지지만, 장기간의 일별 입국자 예측 성능은 우수하다고 할 수 있다.

IV. 실험

4.1 실험 준비

실험에는 데이터 전체 기간인 2010년 1월 1일부터 2020년 9월 30일까지의 데이터 중 2019년 1월 1일을 기준으로 이전의 데이터는 훈련 데이터, 그 이후의 데이터는 테스트 데이터로 사용한다.

본 논문에서는 제시한 예측 모델 실험 외에도 다른 인공지능 모델로 실험을 진행한다. 활용한 모델은 순차적 입력과 순차적 출력 구조를 가지는 (Seq2seq) 두 순환 신경망 모델, LSTM과 양방향 LSTM이다. LSTM은 길이가 긴 입력 정보들을 해석하기 위한 순환 신경망이고, 양방향 LSTM은 순차적 정보의 전후 관계를 해석하여 결과를 출력하는 LSTM의 한 종류다. 두 모델 모두 순차적 데이터를 처리하는 분야인 자연어 처리와 시계열 예측에 주로 활용된다. 제시한 예측 모델의 성능 평가 및 비교를 위해 제시한 모델과 두 순환 신경망 모델로 예측 실험을 진행한다.

4.2 성능 평가 지표

예측 결과에 대한 평가 지표로는 MAPE(평균절대비오차)를 사용한다. MAPE는 분류 모델의 주로 사용하는 평가 지표인 오차율과 유사하게 0~100% 사이의 값을 가져 직관적으로 이해하기 쉽다는 장점이 있다. 본 논문에서 진행한 실험에서는 테스트 데이터 전체 기간 동안 일별로 예측한 값에 대한 MAPE를 산출한다.

4.3 실험 결과

제시한 예측 모델과 순환 신경망을 활용한 예측 모델의 실험 결과는 표 2에, 제시한 예측 모델의 출력 결과를 도시한 그래프는 그림 2에 제시되어 있다. 그림 2에서 실제 입국자 값은 파란색, 훈련

V. 결 론

본 논문에서는 정치, 질병, 한류 등 다양한 외적 요인들과 일차원 합성곱 신경망 모델을 활용한 외국인 입국자 예측 프레임워크를 제시한다. 본 논문에서 보인 실험 결과들을 통해 외국인 입국자 데이터와 같은 불규칙하고 다양한 패턴을 가진 데이터에는 합성곱 신경망을 활용한 예측 모델의 성능이 우수함을 확인할 수 있다.

Acknowledgement

본 연구는 과학기술정보통신부 및 정보통신기획평가원의 지역지능화혁신인재양성(Grand ICT연구센터) 사업의 연구결과로 수행되었음 (IITP-2021-2015-0-00742)

References

- [1] Yoon Kim, "Convolutional Neural Networks for Sentence Classification" *Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, Doha, Qatar. 1746-1751, Oct. 2014.