

온라인 리테일러에서 수요 패턴 예측을 위한 컨볼루션 LSTM 적용 방안

정재원⁺, 임성택, 이승민, 공형준
한양대학교 산업공학과

Application of Convolutional LSTM for Predicting Demand Patterns in Online Retailers

J. Jeong⁺, S. Lim, S. Lee, H. Kong
Department of Industrial Engineering, Hanyang University

Demand pattern forecasting would be surely one of significant modules for successful and profitable on-line retail business. Our work studies the mechanism for demand pattern forecasting at the on-line retailers in a context of e-commerce business. This paper deals with demand pattern forecasting on a basis of time series prediction. We review papers about demand forecasting in a lot of application areas and suggest a new approach to time series prediction using Convolutional LSTM by generating the images for corresponding demand patterns. And it is observed that our proposed approach could elicit more accurate results than the current existing techniques such as seasonal ARIMA and Holt-Winters Technique. Numerical studies using the data for an e-commerce company have been done in order to validate the performance of the suggested demand pattern forecasting mechanism using the Convolutional LSTM. Finally, we provide the meaningful measures ensuring the usefulness and applicability of the proposed demand pattern forecasting mechanism.

Keywords: Demand pattern forecasting, On-line retailing, Convolutional LSTM, Customer orders, Order management

1. 서론

1.1 연구배경 및 목적

온라인 리테일링 운영에 있어서 고객의 수요패턴 예측은 해당 기업의 비즈니스 프로세스 전반에 지대한 영향을 미치는 중요한 요소이다. 즉, 고객의 수요패턴 예측과정에서 발생하는 예측오차는 해당 기업의 매출에 직/간접적으로 막대한 영향을 미치기 때문이다. 또한, 고객수요 패턴에 정확한 예측은 입고, 적재, 오더 피킹, 출고의 기능을 주로 하는 재고관리 시스템의 최적화를 위해 기본적으로 이루어져야 하는 작업이다(Song and Cho, 2015).

본 연구에서는 온라인 리테일 산업의 재고 관리 및 물류 관리 시스템을 위한 수요 예측에 대해 다룬다. 최근 몇 년간 꾸준한 성장세를 보이던 온라인 리테일 산업은 코로나 사태 이후, 더욱 급격하게 성장하게

되었다. 기존에 주로 오프라인으로 판매되던 많은 품목들이 언택트의 필요성에 따라 상당 부분 온라인 판매 품목으로 유입되었고, 그로 인해 온라인 리테일러는 품목의 다양성과 물량증가에 따라 재고 관리에 어려움을 겪게 되었다. 또한, 온라인 리테일 사업의 성장에 따른 물량증가는 물류시스템의 운영과정에서 물류센터에서의 용량 부족 현상으로 나타나고 있다. 예시로, 대한민국 온라인 리테일 사업의 선두주자라고 할 수 있는 쿠팡은 자사 뉴스룸을 통하여 최근 충청북도, 광주광역시, 경상북도에 대규모 첨단물류센터를 설립하기 위해 부지를 매입하였음을 밝혔고, 다른 온라인 리테일사에서도 비슷한 움직임을 보여주고 있다(news.coupang.com, 2020). 실질적인 운영능력의 확보를 위해서는 사전적으로 고객의 수요패턴에 대한 좀 더 정확한 예측기능이 전제가 되어야 하고, 이러한 예측정확도의 개선은 물류시스템의 물리적인 운영용량이 아니라 실질적인 운영역량으로 연계될 수 있을 것이다. 즉, 고객수요의 사전 예측력 제고는 물류시스템의 물리적인 확장에 앞서서 우선적으로 검토되어야 하는 중요한 과제이다. 최근에는 전통적 수요예측 기법인 Holt-윈터스 기법, 계절성 ARIMA(Seasonal ARIMA) 등보다는 인공지능 기반의 수요예측 기법을 많이 사용하는 추세이다. 인공지능 기반 수요예측 기법은 기존 수요예측 기법에 비해 상대적으로 높은 정확도를 보이며 예측실행시간도 덜 소요된다. 또한 독립변수와 종속변수간의 관계를 나타내는 모델을 설정하지 않고 알고리즘을 구축할 수 있다는 장점이 있다(Jeong and Lim, 2019). 이와 같은 장점들이 있기 때문에 수 년 전부터 구글, 패션브랜드, 카드사 등 전 세계 많은 업체에서 더욱 정확한 예측 결과를 위해 LSTM, GRU 등의 인공지능, 인공지능망 기반의 수요예측기법을 이용해오고 있다(Nenni et al., 2013).

본 연구에서는 수리적인 모델에 기반한 수요패턴 예측기법들이 갖는 한계점을 극복할 수 있는 대안으로 딥러닝(Deep Learning) 기반의 예측방법론을 제안하고자 한다. 따라서, 본 연구에서는 온라인 리테일러의 과거 고객주문 데이터를 패턴화하기 위하여 판매품목별 시간구간별 수요량을 이미지 형식으로 변환하고, 이러한 이미지를 이용하여 딥러닝 시계열 예측 모형을 적용하고자 한다. 연구에서 사용한 고객주문 데이터는 UCI의 Online Retail 데이터이며, 영국의 온라인 리테일 회사의 2010년 12월부터 2011년 12월까지의 54만여건의 데이터이다(<https://archive.ics.uci.edu/ml/datasets/online+retail>). 이미지화된 데이터에는 상용화된 많은 영상, 이미지 분석·예측 알고리즘을 그대로 적용할 수 있다는 장점이 있다. 우리는 공간정보를 활용할 수 있는 시계열 예측 모형인 컨볼루션 LSTM에 주문리스트 정보를 이미지 데이터로 변환하여 입력하였을 때, 데이터 안에서 픽셀 간의 관계를 고려하는 이미지 학습의 효과가 수요 예측에도 유의미하게 작용할 것이라는 가정 하에 연구를 수행하였다.

본 연구에서는 전통적 예측 기법과 기존의 인공지능 기반의 수요예측 기법에 대한 논문들을 리뷰하고, 이미지화된 데이터에 컨볼루션 LSTM을 적용하여 수요(판매량)를 예측하는 새로운 방법을 제안하고자 한다. 또한, 온라인 리테일러 회사인 A 회사의 12개월간의 주문데이터에 대해 일별 수요 패턴 예측과 시간대별 수요 패턴 예측을 우리가 제안하는 새로운 방식으로 실험한 결과와 기존의 대표적 수요예측기법인 Holt-윈터스 기법, SARIMA, LSTM, GRU 모형으로 실험한 결과의 RMSE, MAE 등을 비교하여 우리가 제안하는 방법의 예측력 향상에 대해 검증하고자 하였다.

1.2 이론적 배경

1) Holt-윈터스 모형

Holt-윈터스(Holt-Winters) 모형은 Winters가 제안한 방법으로 기법의 간명성과 단순성에도 불구하고 예측능력 만큼은 ARIMA와 같은 모형들에 결코 뒤지지 않는다고 평가받고 있다(Kim, 2009). 또한, 계절성을 추정할 수 없는 일반적인 지수평활 기법과 달리 예측식과 평활식을 이용하여 계절성을 직접 고려하여 시계열의 미래를 예측할 수 있다. 데이터가 가지고 있는 특성에 따라 활용할 수 있는 Holt-윈터스 기법은 승법(multiplicative) 방법과 가법(additive) 방법이 있다(Jeong, 2004). Holt-윈터스 기법에 사용되는 평활식은 평균수준(level), 추세(trend), 계절성(seasonality) 총 3가지 요인들에 대하여 각각 하나씩 총 3개의 평활식으로 이루어져 있다. 아래에서는 Holt-윈터스 기법에 일반적으로 사용되는 용어들을 정의하고 Holt-윈터스 기법의 종류를 설명한다.

Y_t : t 시점에서 관측된 값
 L_t : t 시점에서의 시계열 평균수준
 b_t = 시점 t 에서 시계열 추세성분
 S_t = 시점 t 에서 시계열 계절성분
 $F_{t+m} + m$ = 시점 t 에서 예측한 시점 t + m 의 예측값
 s = 계절성분의 길이
 α, β, γ =평활모수

홀트-윈터스 기법은 시계열 데이터가 지닌 계절성의 특성에 따라서 승법적 모형, 가법적 모형으로 나타낸다. 시계열 데이터 전반에 나타나는 계절성(seasonality) 요인이 큰 변동없이 일정하게 나타나는 경우에는 가법적 모형을, 시계열의 수준에 따라서 계절성 요인이 비례하는 경우 승법적 기법을 사용한다.

① 홀트-윈터스 승법적 모형 (Holt-Winters Multiplicative Model)

홀트-윈터스 승법적(multiplicative) 모형에서의 식은 식 (1.1)~(1.4)과 같다.

$$L_t = \alpha \frac{Y_t}{S_{t-s}} + (1 - \alpha)(L_{t-1} + b_{t-1}), \quad (1.1)$$

$$b_t = \beta(L_t - L_{t-1}) + (1 - \beta)b_{t-1}, \quad (1.2)$$

$$S_t = \gamma \frac{Y_t}{L_t} + (1 - \gamma)S_{t-s}, \quad (1.3)$$

$$F_{t+m} = (L_t - b_t m) S_{t-s+m}. \quad (1.4)$$

홀트-윈터스의 승법적 모형은 로그변환을 통해 가법모형으로 변형시킬 수 있다.

② 홀트-윈터스 가법적 모형 (Holt-Winters Additive Model)

홀트-윈터스 가법적(additive) 모형에서의 식은 식 (2.1)~(2.4)과 같다.

$$L_t = \alpha(Y_t - S_{t-s}) + (1 - \alpha)(L_{t-1} + b_{t-1}), \quad (2.1)$$

$$b_t = \beta(L_t - L_{t-1}) + (1 - \beta)b_{t-1}, \quad (2.2)$$

$$S_t = \gamma(Y_t - L_t) + (1 - \gamma)S_{t-s} \quad (2.3)$$

$$F_{t+m} = L_t + b_t m + S_{t-s+m} \quad (2.4)$$

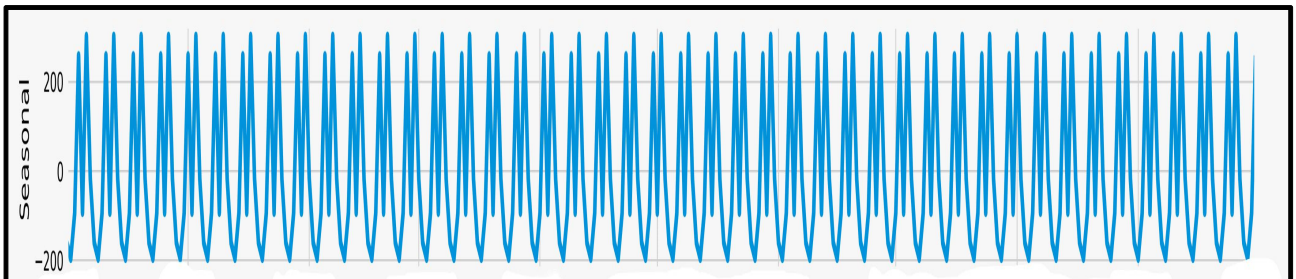


Figure 1 : Seasonality of daily demand of top 77 products

본 연구에서 분석대상이 되는 수요데이터가 계절성이 시계열의 수준에 따라서 비례하는 경우를 가정하여 적용하는 승법적 홀트-윈터스 기법이 더 적합하다고 판단하여 승법적 방법으로 비교군을 생성한다.

2) SARIMA (Seasonal Autogressive Integrated Moving Average)

대표적인 확률적 시계열모형에는 AR, MA, 그리고 ARMA 모형이 있다. 이러한 모형은 모두 정상적 시계열 자료, 즉 시점에 관계없이 평균과 분산이 일정한 시계열에 사용한다. 따라서, 비정상 시계열 자료에 대해서는 시계열을 차분하여 정상화하는 ARIMA 모형 등을 이용해야한다. 비정상 시계열 Y_t 를 d 번 차분한 시계열 y_t 를 ARMA(p,q) 모형으로 나타낼 수 있으며, 차분 전 시계열 자료 Y_t 를 적분된 자기회귀 이동평균(Autogressive Integrated Moving Average), 즉 ARIMA(p,d,q) 모형이라고 한다. 일반적으로 ARIMA(p,d,q) 모형은 식(3)와 같이 나타낼 수 있다(Lee and Kwon, 2004).

$$\phi_p(L)(1-L)^d y_t = \theta_q(L)u_t \quad (3)$$

, where

- p = 자기회귀(AR) 항의 차수
- q = 이동평균(MA)의 차수
- d = 정상 시계열이 될 때까지 차분한 횟수
- L = 시차변수(lag operator)
- y = 수요변수
- u = 평균이 0, 분산이 σ^2 인 교란변수

일반적으로 본 연구의 실험에서 사용된 데이터와 같이 주기적인 특성을 가지는 시계열 자료에 대해서는 계절성을 고려할 수 있는 Seasonal ARIMA (이하 SARIMA) 모형을 사용하는 것이 적합하다. SARIMA(p, d, q)(P, D, Q)s 모형은 식(4)와 같다.

$$\phi_p(L)\phi_p(L)(1-L)^d(1-L^s)^D y_t = \theta_q(L)\theta_q(L)u_t \quad (4)$$

, where

- s = 분기 수
- D = 계절 단위근(Seasonal Unit Root)의 차수
- P = 계절 자기회귀(Seasonal Autogressive, SAR)항의 차수
- Q = 계절 이동평균(Seasonal Moving Average)항의 차수
- $\theta_q(L)$ = 계절 이동평균항(SMA term)
- $\theta_q(L)$ = 계절 자기회귀항(SAR term)

기존 연구에서 SARIMA 모형은 주로 주기적 특성이나 계절성이 뚜렷한 전력, 교통이나 관광 사업에서의 수요 예측 및 분석에 사용된다. Chang and Liao (2010)의 연구에서는 대만에서 홍콩, 일본, 미국으로 가는 여행객 수의 월별 데이터를 SARIMA를 적용하여 예측하였다. Kim et al.(2014)은 2002년부터 2010년 까지 주간 최대수요 시계열 데이터에 SARIMA 모형을 적용하여 예측결과를 제시하였다. Lee and Kwon (2011)에서는 장래 교통수요 예측에 있어, 국내 교통실적 자료들의 짧은 축적기한으로 인한 데이터 부족을 해결하기 위해 월별 자료를 활용하고, 이에 따라 나타나는 계절성을 고려할 수 있는 SARIMA 모형을 적용하였다. 그러나, ARIMA 또는 SARIMA 모형과 같은 시계열 예측 모델은 기존 데이터의 변동 형태를 파악하여 일정한 패턴이 있는 데이터 값을 예측하기 때문에 데이터가 복잡하거나 비선형성을 가질 때에는 높은 정확도를 기대할 수 없다는 한계점이 있다. 따라서, 복잡하거나 비선형적인 데이터에는 인공지능 기법을 이용한 예측모형을 활용하는 것이 효과적이다 (Lee and Kim, 2020).

3) LSTM (Long Short-Term Memory)

LSTM 모형은 Hochreiter and Schmidhuber(1997)에 의해 제안된 모형으로 기존의 RNN(Recurrent Neural Network) 모형에서 장기 의존성을 해결하기 위해 만들어진 RNN의 변형 모델이다. RNN의 은닉층에 cell state 구조를 추가한 것으로, 셀 안의 forget gate와 input gate로 입력 정보에 대해 보존 여부를 업데이트하여 학습한다. 이로 인해 RNN에서 사용하는 정보와 사용 지점의 거리가 길 때 학습능력이 저하되는 단점을 극복할 수 있는 대안으로 사용될 수 있다. 일반적인 LSTM 모형의 구조는 다음의 Figure 2 와 같다.

$$f_t = \sigma(W_f \times [h_{t-1}, x_t] + b_f) \quad (5.1)$$

$$i_t = \sigma(W_i \times [h_{t-1}, x_t] + b_i) \quad (5.2)$$

$$\tilde{C}_t = \tanh(W_C \cdot [h_{t-1}, x_t] + b_C) \quad (5.3)$$

$$C_t = f_t \times C_{t-1} + i_t \times \tilde{C}_t \quad (5.4)$$

$$\sigma_t = \sigma(W_o \cdot [h_{t-1}, x_t] + b_o) \quad (5.5)$$

$$h_t = \sigma_t \times \tanh(C_t) \quad (5.6)$$

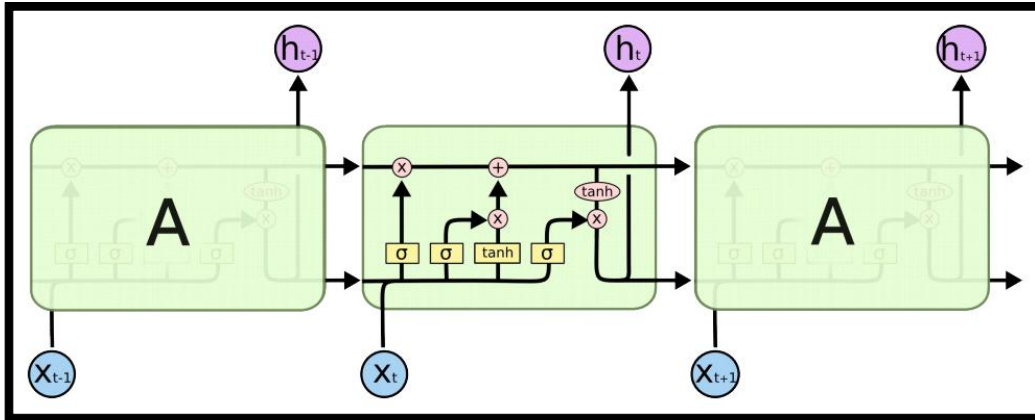


Figure 2. LSTM framework (Olah, 2015)

식 (5.1) ~ (5.6)에서 W_f 는 forget gate의 가중치 행렬을 의미하고, $[h_{t-1}, x_t]$ 는 t-1에서 output이 t에서의 input 으로 반영되는 것을 의미한다. f_t (forget gate vector)에서는 정보를 제거하는 가중치를 결정하고, i_t (input gate vector)는 새로운 정보를 획득하는 가중치를 결정한다. o_t (output gate vector) 는 output의 후보를 결정한다. x_t 는 입력 벡터, h_t 는 출력 벡터, c_t 는 셀 상태 벡터, W, U, b 는 각각 파라미터 행렬 및 벡터이다(Weng et al., 2019). LSTM 모형은 외부 변수 통합과 자동 특징 추출로 인한 용이성으로 전기사용량, 풍력발전량, 주가, 주택가격 예측 뿐 만 아니라 시퀀스 예측 작업인 음성 인식, 음악 작곡 등 다양한 분야에서 이용된다(Wu et al., 2018). 그러나, 이 모형은 시공간적 특성을 고려하지 못해 이미지나 영상 데이터를 사용할 때는 flatten된 데이터를 사용해야한다는 한계점을 가진다(Ai et al., 2018).

4) GRU (Gated Recurrent Unit)

GRU는 LSTM과 같이 게이트의 개념을 이용한 순환 신경망의 일종으로 LSTM의 구조를 단순화하여 연산속도가 LSTM보다 빠르다(Lee et al., 2019). GRU는 LSTM의 망각게이트(forget gate vector, f_t)와 입력게이트(input gate vector, i_t)를 갱신게이트(z_t)로 통합하고, 셀 상태(c_t)와 은닉 상태(h_t)를 h_t 하나로 통합하

였다 (Kim, 2017). 이처럼 LSTM보다 단순한 구조로 가중치 수가 작으므로 학습이 더 빠르지만, LSTM과 거의 동일한 성능을 보인다(Chaudhuri et al., 2016).

1.3 분석 방법론

1) 컨볼류션 LSTM (Convolutional Long Short-Term Memory)

컨볼류션 LSTM (이하 ConvLSTM) 은 FC-LSTM(Fully-Connected Long Short-Term Memory) 을 확장한 모델로 컨볼류션 레이어를 LSTM에 추가한 기법이다(Shi et al., 2015). 기본적으로 FC-LSTM은 셀의 input, output, state가 모두 1차원 벡터이고, FC-LSTM의 기본 구조는 아래의 식 (6.1)~(6.5)과 같다.

$$i_t = \sigma(W_{\xi}X_t + H_{(t-1)} + W_{ci} \circ C_{(t-1)} + b_i) \quad (6.1)$$

$$f_t = \sigma(W_{xf}X_t + W_{hf}H_{t-1} + W_{cf} \circ C_{t-1} + b_f) \quad (6.2)$$

$$C_t = f_t \circ C_{t-1} + i_t \circ \tanh(W_{xc}X_t + W_{hc}H_{t-1} + b_c) \quad (6.3)$$

$$o_t = \sigma(W_{xo}X_t + W_{ho}H_{t-1} + W_{co} \circ C_t + b_o) \quad (6.4)$$

$$H_t = o_t \circ \tanh(C_t) \quad (6.5)$$

이와는 달리 ConvLSTM은 셀의 input, output, state가 3차원 텐서를 가지고 있다. ConvLSTM은 “Input to State”와 “State to State”에 컨볼류션 연산을 사용하는 특징을 가진다. 컨볼류션 연산을 사용함으로써 각 셀에 존재하는 웨이트 수가 FC-LSTM보다 확연히 적어지는 효과를 기대할 수 있다. 다음의 식(7.1)~(7.5)에서의 * 은 컨볼류션 연산을 의미한다. 이러한 특성을 제외한 이외의 수식은 FC-LSTM과 동일하다.

$$i_t = \sigma(W_{\xi}^*X_t + W_{hi}^*H_{t-1} + W_{ci} \circ C_{t-1} + b_i) \quad (7.1)$$

$$f_t = \sigma(W_{xf}^*X_t + W_{hf}^*H_{t-1} + W_{cf} \circ C_{t-1} + b_f) \quad (7.2)$$

$$C_t = f_t \circ C_{t-1} + i_t \circ \tanh(W_{xc}^*X_t + W_{hc}^*H_{t-1} + b_c) \quad (7.3)$$

$$o_t = \sigma(W_{xo}^*X_t + W_{ho}^*H_{t-1} + W_{co} \circ C_t + b_o) \quad (7.4)$$

$$H_t = o_t \circ \tanh(C_t) \quad (7.5)$$

ConvLSTM 구조는 Figure 3과 같으며 LSTM 셀에서 공간적인 의미와 시간적인 의미를 동시에 반영할 수 있다(Shi et al., 2015).

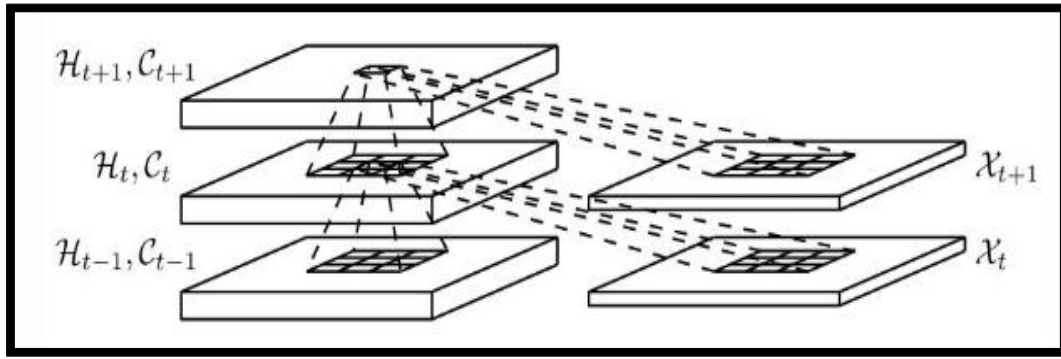


Figure 3. Convolutional LSTM framework (Shi *et al.*, 2015)

2. 본론

2.1 연구 방안

1) 실험 환경

본 연구에서는 진행된 수치실험은 고객 수요패턴 예측을 위해 적용한 컨볼류션 LSTM의 성능비교를 위하여 동일 데이터를 이용하여 딥러닝 모델(GRU, LSTM)과 시계열 예측 모델(SARIMA, Holt-윈터스 기법)로 구분되었다. 구체적인 실험환경 설정사항으로 딥러닝 모델들의 경우는 Google-Colab의 TPU 런타임에서 Python 3.6.9의 Tensorflow 2.3.0, Keras2.4.3 등의 기계학습 라이브러리를 사용해서 진행되었다. 또한, SARIMA 와 Holt-윈터스 기법의 실험환경은 Python 3.6.9의 Pmdarima, Statsmodel, Pandas 라이브러리와 마이크로소프트 Excel 등을 사용하였다. 수치실험에 사용된 하드웨어 사양은 Intel® Core™ i5-7500, DDR4 RAM 16.0GB 이다.

2) 실험 설계

① 데이터 전처리

본 연구에서는 영국의 온라인 리테일 회사의 2010년 12월부터 2011년 12월까지의 1년 동안의 고객주문 데이터를 이용했으며, 확보된 고객주문건수는 모두 54만 여 건이다. 개략적으로는 공휴일과 주말을 제외하고 실제적으로는 총 305일에 해당하는 영업일 기준의 데이터로 구성되어있다. 데이터 항목체계는 '주문번호', '제품식별코드', '제품설명', '주문수량', '주문시간', '단위가격', '고객ID', 그리고 '주문국가'로 구성되어 있다. 단, 고객 주문시각은 초 단위로 설정되어 있다. 본 연구의 목적에 따라서, 품목별 시간대별 고객 주문 수량을 예측하기 위해 '제품식별코드', '주문시각', '주문수량'만을 선택적으로 발췌하여 실험을 진행하였으며, 데이터 전처리는 1) 이상치 및 결측치 제거, 2) 일자별 구간화, 3) 주문 시간대 별 구간화, 그리고 4) 주문데이터의 이미지화의 순서로 진행되었다. 이상치는 논리적 이상치와 통계적 이상치로 나누어서 처리하였다. 먼저, 논리적 이상치에 해당하는 '해외배송비용', '반품비용', 그리고 '반품처리'를 모두 제거하였다. 또한, 주문 데이터 중 주문 수량에서 상위 0.03%에 해당하는 1,000 이상의 주문수량을 가진 경우를 통계적 이상치로 판단하여 논리적 이상치와 함께 제거하였다. 이러한 품목별 시간구간별 수요정보를 본 연구에서는 각 행(row)은 시간, 열(column)은 품목을 의미하는 2차원 공간상의 이미지 체계를 구성하였다. 이러한 구조에 기반하여 2차원 공간상에 나타나는 각 픽셀의 색상(Color)은 주문수량을 나타낼 수 있도록 이미지를 구성함으로써 시간적 요소와 공간적 요소를 모두 반영할 수 있는 이미지 데이터로 구성하였다. 이를 위한 고객주문 데이터의 이미지화 과정에 적용한 절차는 다음과 같다. 아래 과정을 도식화 한 것은 Appendix a에 첨부하였다.

[단계-1] 고객주문 데이터를 시간대 별로 구간화 한다. 이상치와 결측치를 제거한 총 53만 여 건의 데이터에서 가장 주문이 일찍 들어온 시각(오전 7시 40분)과 가장 늦게 들어온 시각(오후 8시 20분) 사이의 데이터를 15분 간격으로 나누어 총 77개 구간으로 구간화한다.

[단계-2] 수요패턴을 이미지로 변환할 품목을 선정한다. 이미지의 row와 column 수를 일치시키기 위하여 총 4082개의 품목 중 주문 수량 상위 77개 품목의 주문 데이터만 활용하였다.

[단계-3] 15분 간격으로 설정된 각 시간구간별 품목들의 주문 수량 데이터를 RGB값으로 변환한다. 주문 수량을 16진법으로 변환시킨 후 1,2번째 자리의 수는 Green값으로, 3,4번째 자리의 수는 Red값으로 설정해주었다.

이와 같은 고객주문 데이터의 이미지 변환 과정을 통해서 시간구간별 고객주문 수량을 RGB값으로 변환함으로써 이미지(77x77x3)로 만들 수 있었다. 고객주문 정보와 이를 이미지 형태로 변환한 예시는 Appendix에 제시하였다. 최종 데이터는 주문일자, 시간대, 품목, 주문수량 순으로 (305x77x77x3) 구조의 4차원 텐서 형태로 구성하였다. 또한, 비교군의 모형과 ConvLSTM 모형에 대해 공통적으로 시계열 데이터의 마지막 28일 중 14일을 Validation 데이터로, 나머지 14일을 Test 데이터로 사용하였으며, 이를 제외한 부분을 Training 데이터로 사용하였다.

② 컨볼루션 LSTM 모형 수립

본 연구에서는 Tensorflow 라이브러리의 ConvLSTM2D 함수를 사용해서 ConvLSTM을 구현하였다. ConvLSTM2D의 필터 개수와 커널 크기, 레이어의 깊이를 변화시키면서 모델의 성능을 확인하고, 이를 바탕으로 모델을 수립하였다. 필터 개수는 10, 20, 40, 60, 80개 순으로 증가시켰으며, 커널 크기는 1x1, 2x2, 3x3 순으로 증가시켰고, 레이어의 깊이는 1, 2, 3단으로 증가시켰다. 최종 결과 값을 이미지의 형태로 만들기 위해서 모델의 마지막 레이어는 Conv3D 레이어를 사용하였으며, 필터는 RGB값을 맞춰주기 위해서 3개를 사용하였고, 커널의 크기는 3x3으로 고정하였다.

모델 학습에서는 두 이미지의 차이를 나타내는 지표로 알려진 Binary cross entropy를 Loss로 사용하였으며, 최적화 함수는 adadelat를 사용하였다(Creswel et al., 2017). 검증 데이터에서의 Loss값이 30회 동안 1퍼센트 이상 줄어들지 않으면 훈련을 종료하고 이미지 형태로 나온 최종 결과값을 위에서 설명한 주문 데이터에서 이미지를 만드는 순서의 역순으로 변환해주고, 이를 실제 값과 비교하면서 최적 파라미터를 찾았다. 이를 아래와 같이 도식화할 수 있다.

커널의 크기가 1x1, 2x2인 경우 필터의 개수와 레이어의 깊이와 상관없이 정상적으로 학습이 이루어지지 않았다. 따라서 커널의 크기가 3x3인 경우만 고려하면서 최적 파라미터를 찾았다.

위 실험을 통해서 레이어의 깊이가 깊어질수록 동일한 파라미터의 경우 좋은 성능을 내는 것을 알 수 있었으며, 커널의 크기가 일정 크기 이상인 경우에만 정상적인 학습이 이루어 지는 것을 알 수 있었다. 본 연구의 최종 컨볼루션 LSTM 모델은 레이어를 3단으로 쌓았으며, 각각의 레이어는 필터 개수 80개, 커널 크기 3x3의 파라미터를 가지고, 마지막 레이어는 Conv3D를 사용한다.

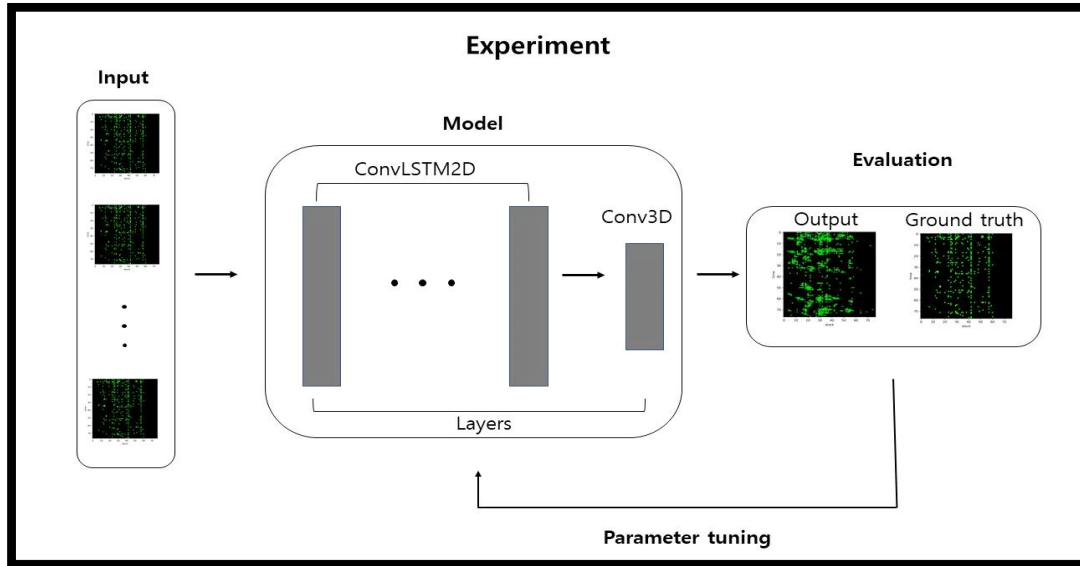


Figure 4. Framework of ConvLSTM for demand prediction

③ 비교군(Holt-Winters, SARIMA, LSTM, GRU) 모형 추정 및 수립

비교군 중 SARIMA의 실험은 다음과 같은 과정으로 진행되었다. 일별 예측에서는 Pmdarima의 auto_arima를 이용하였고, 시간대별 예측은 SARIMAX 라이브러리를 이용하였다. auto_arima를 이용한 실험에서는 먼저 시계열 데이터를 statsmodel의 seasonal_decompose 함수로 추세와 계절성 정도를 확인한 후, 파라미터 d와 D의 값은 1로 고정한 뒤, p와 q에 대해서는 0~10의 범위 내에서, P와 Q는 0~3의 범위 내에서 최적의 값을 탐색하여 AIC 값을 최소화하는 최적의 파라미터들을 추정하였다. 이러한 과정을 거쳐 일별 주문량 예측에는 ARIMA(0, 1, 1)(2, 1, 1)₁₁ 모형을 적용하였다. 시간대별 예측 실험에서는 데이터 특성상 auto_arima를 이용하여 좋은 결과를 얻기 어려웠기 때문에 일별 예측에서 사용한 파라미터들을 기준으로 변경해가며 실험하였고, 그 중 가장 낮은 RMSE 값을 보이는 값은 일별 예측에서와 동일한 모형인 ARIMA(0, 1, 1)(2, 1, 1)₁₁ 모형임을 확인할 수 있었다.

홀트-윈터스 기법을 적용하기 위해 계절성의 주기인 s (=계절성분의 길이)를 6일, 7일, 12일, 14일, 30일, 42일로 바꾸어가며 최적의 s 값을 탐색하였다. 또한 그 중 가장 작은 MSE, MAE, MAPE, RMSE 값을 보이는 12일을 계절 성분 주기 s 로 선정하여 모형을 수립하였다. 그리고 홀트-윈터스 기법을 사용하기 위해서는 데이터가 시작되는 초기 구간의 계절성, 추세, 평균수준을 정해주어야 한다. 이때, 첫 번째 주기의 S_t (계절성)은 식 (8.1)과 같다.

$$S_t = \frac{Y_t}{\text{구간 내 } Y_t \text{의 평균}} \quad (8.1)$$

그리고 초기 L_t (평균수준)과 b_t (추세)는 두 번째 주기의 첫 기간에 설정하게 되는데, L_1 (두 번째 구간의 첫 번째 평균수준)은 식 [7.2]과 같이,

$$L_1 = \frac{Y_{1+s}(\text{두 번째 구간의 첫 번째 수요})}{S_1} \quad (8.2)$$

정해주고, b_1 은 다음 식 [7.3]과 같이 정한다.

$$b_1 = L_1 - \frac{Y_s(\text{첫 번째 구간의 마지막 수요})}{S_s(\text{첫 번째 구간의 마지막 계절성})} \quad (8.3)$$

그 이후 예측을 위해 1.2의 1)에 서술한 Holt-윈터스 승법모형을 이용하여 예측연산을 전개하였다. 그리고 Holt-윈터스 기법에서 사용되는 평활모수 α, β, γ 는 RMSE를 최소화하는 값으로 최적화하여 실험을 진행하였다.

2.2 실험 결과 및 분석

본 실험에서는 적용대상이 되는 상위 77개 품목의 수요합계를 일별 수요패턴 예측과 시간구간별 수요패턴 예측으로 구분하여 진행하였다. 즉, 일별 수요패턴은 77개 품목에 대해서 동일 일자에 발생한 수요량의 합계치를 산정하고, 해당하는 일별 수요량을 예측하는 실험으로 진행하였다. 본 실험에서 적용하는 딥러닝 기반의 예측방식과 시계열 기반의 예측방식에 대한 상대적인 성능 비교를 위하여 수요정보의 집계수준 (Aggregation level)에 따른 적용방법 등의 상대적인 수요예측 성능을 분석하고자 한다. 기존 예측기법과의 비교를 위해 77개 품목 각각에 대해 설정 구간별로 수요량을 예측하고, 이를 모든 품목에 대해 합산하여 구간별 총 예측 수요량으로 예측 정확도를 비교하였다. 또한, 본 연구에서 제안하는 컨볼루션 LSTM기반의 수요예측 방법의 장/단점을 분석하기 위하여 수치실험 조건을 차별화 하였다.

1) 일별 수요패턴 예측 (Daily demand pattern prediction)

컨볼루션 LSTM이 일자별 예측에서 MAE기준으로 가장 좋은 성능을 보여주었으며 MAE를 제외한 지표들에서는 SARIMA가 가장 좋은 성능을 보여주었다. Table 1에서 컨볼루션 LSTM과 SARIMA는 다른 방법론들과 비교했을 때, MAPE 지표에서 특히 큰 차이를 보인다. GRU와 LSTM은 지표에서 전반적으로 비슷한 값을 보였으며, Holt-윈터스 기법의 경우 다른 방법론들과 비교하였을 때 전체적인 지표에서 모두 낮은 성능을 보여주고 있다.

일자별 예측의 경우 큰 오차를 보이고 있는 Holt-Winters 방법을 제외한 4가지 방법론 모두 준수한 오차를 보였으며, Figure 5와 figure 6a ~ 6d 에서 나타나는 것처럼 수요의 패턴을 어느 정도 예측하는 것을 확인할 수 있다. 하지만 ConvLSTM 의 경우 급격한 변화에 민감하게 반응하지만, GRU, LSTM, SARIMA의 경우 급격한 변화를 잘 예측하지 못하고 예측치가 일정한 수준에서 크게 변하지 않는 것을 볼 수 있다. 이는 Table 2 에서도 나타나는데 실제 값과 ConvLSTM은 유사한 값을 가지나, GRU, LSTM, SARIMA 등은 예측의 평균치에서 크게 달라지지 않는다.

Table 1. Performance comparison of daily prediction results

Model	Accuracy	MSE	RMSE	MAE	MAPE
ConvLSTM		3.228570e+06	1796.822	<u>1460.428</u>	37.128
GRU		3.530255e+06	1878.897	1585.552	44.720
LSTM		3.432641e+06	1852.739	1573.307	44.975
SARIMA		<u>3.084248e+06</u>	<u>1756.203</u>	1594.036	<u>35.612</u>
Holt-Winters		5.034231e+06	2243.710	1894.353	53.646

Table 2. Mean, Max, Min, and STD values of daily prediction results

Model	Statistics	Mean	Max	Min	STD
Actual(True Data)		5488.286	8041.000	1948.000	1881.077
ConvLSTM		5388.749	8192.064	1347.813	1802.190
GRU		4764.260	5674.180	3104.324	704.366
LSTM		4832.886	5896.279	3441.449	762.360
SARIMA		4826.330	5788.540	3635.976	662.229
Holt-Winters		5170.456	8685.643	2815.270	1774.032

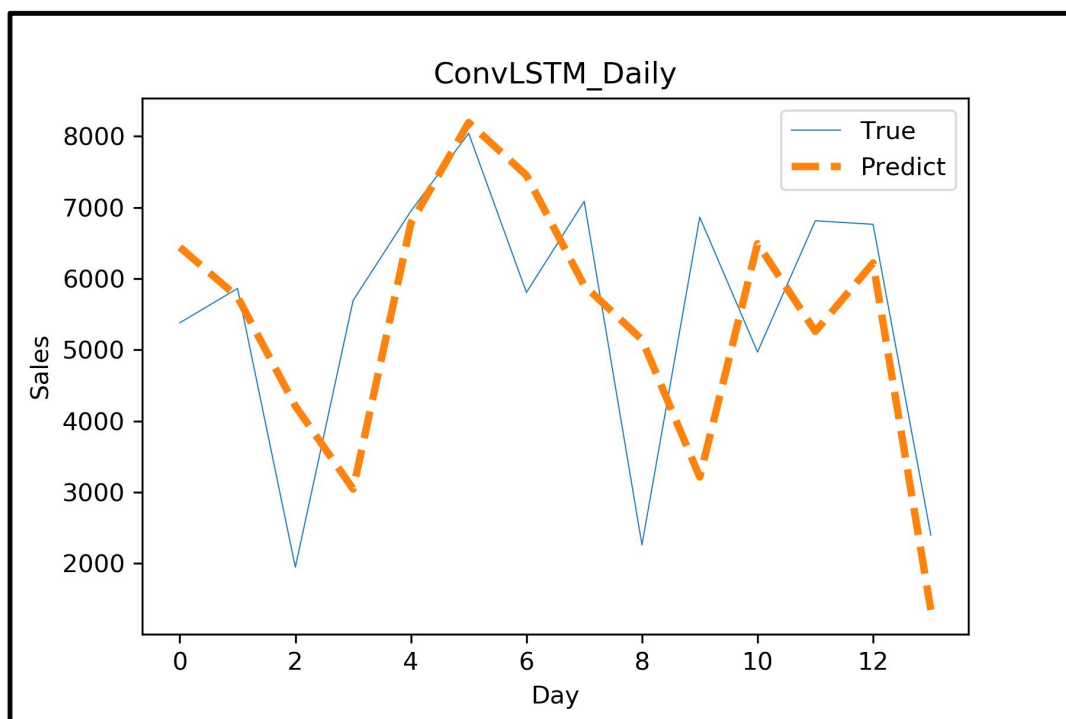


Figure 5. Results of daily prediction by ConvLSTM

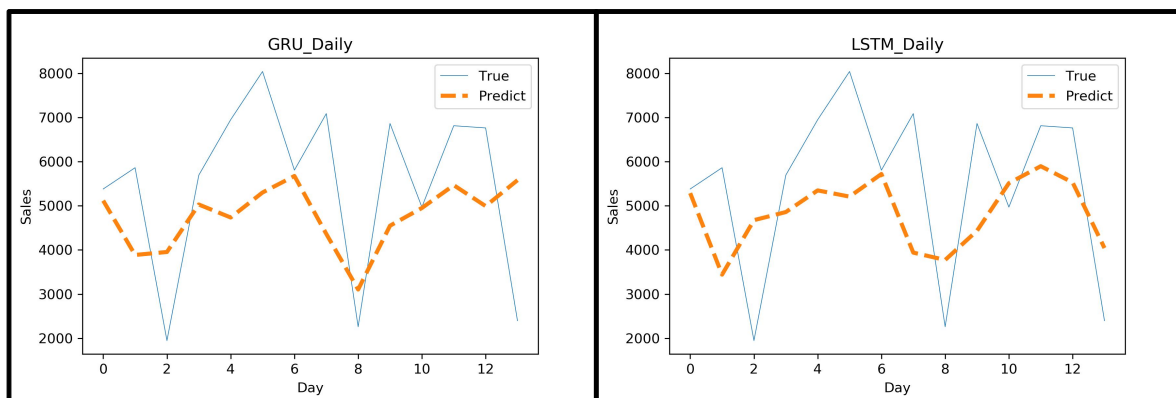


Figure 6a. Results of daily prediction by GRU

Figure 6b. Results of daily prediction by LSTM

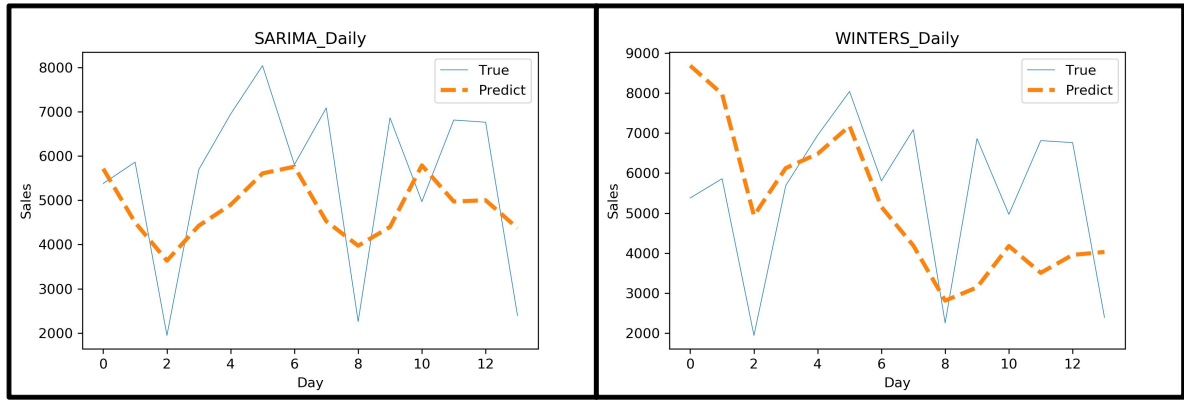


Figure 6c. Results of daily prediction by SARIMA Figure 6d. Results of daily prediction by Holt-Winters

2) 시간구간별 수요패턴 예측 (Hourly demand pattern prediction)

앞서 진행한 일별 수요패턴 예측실험과는 달리 좀 더 세분화된 시간구간 단위별 수요패턴에 대한 예측 실험을 진행하였다. Table 3에서 확인할 수 있듯이 ConvLSTM이 MSE, RMSE, MAE에서 모두 가장 좋은 성능을 보여주었고, 일자별 예측에서 MAE지표를 제외하고 가장 좋은 성능을 보여주었던 SARIMA는 GRU와 LSTM과 비교하였을 때 상대적으로 큰 차이로 낮은 성능을 보여주고 있다. GRU와 LSTM의 성능이 시간대별 수요 예측에서 개선된 것으로 보이며, 이는 시간대별 예측에서는 하루를 13시간으로 보고 1시간 단위로 데이터를 나누었기 때문에 기존의 305일치 데이터에서 3965시간으로 데이터의 양이 많아졌고, 학습 데이터가 많아짐에 따라서 딥러닝 시계열 예측 모형인 GRU와 LSTM의 성능이 SARIMA보다 높게 나온 것으로 추정된다.

Figure 7과 8a~8c 에서 보이는 바와 같이 일자별 수요 데이터 예측과 동일하게 ConvLSTM은 수요의 급격한 변화를 어느 정도 예측하지만, GRU, LSTM, SARIMA에서는 급격한 변화를 예측하지 못하고, 이는 <Table 4>에서도 확인할 수 있다. GRU, SARIMA의 예측 값에는 음수 값이 존재하며, LSTM의 경우 위 두 가지 모델보다는 낮은 오차를 보여주었지만, 실제 데이터에서 주문량이 0인 시점들은 예측하지 못했다. ConvLSTM의 경우 이미지의 RGB값이 0~255 사이의 숫자로 표현되기 때문에 음수 값이 나오지 않았으며, 주문량이 0인 시점을 예측하는 모습을 볼 수 있다.

Table 3. Performance comparison of hourly prediction results

Model	Accuracy	MSE	RMSE	MAE
ConvLSTM		<u>146601.811</u>	<u>382.886</u>	<u>243.581</u>
GRU		159026.510	398.781	270.749
LSTM		150086.307	387.410	271.107
SARIMA		211561.362	459.958	344.083

Table 4. Mean, Max, Min, and STD values of hourly prediction results

model	statistics	Mean	Max	Min	STD
Actual(True Data)		422.176	2236.000	0	457.748
ConvLSTM		414.519	1463.672	0	381.911
GRU		286.941	929.322	-127.195	266.181
LSTM		366.271	800.940	45.019	234.055
SARIMA		342.325	749.677	-387.563	137.723

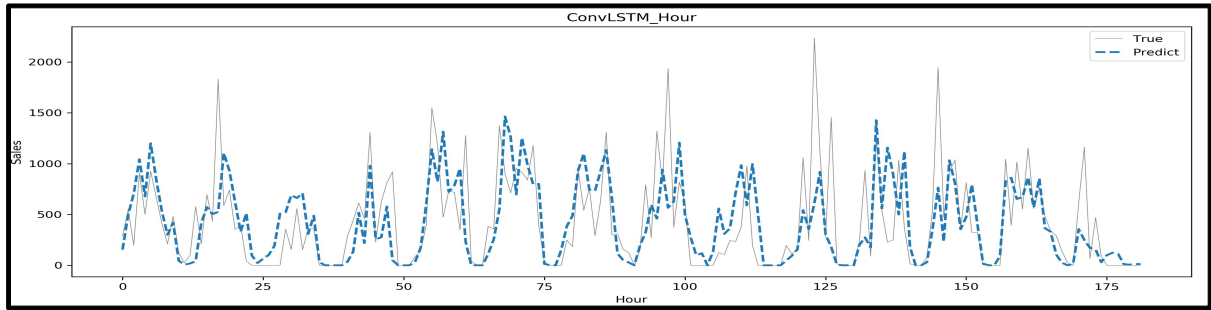


Figure 7. Results of hourly prediction by ConvLSTM

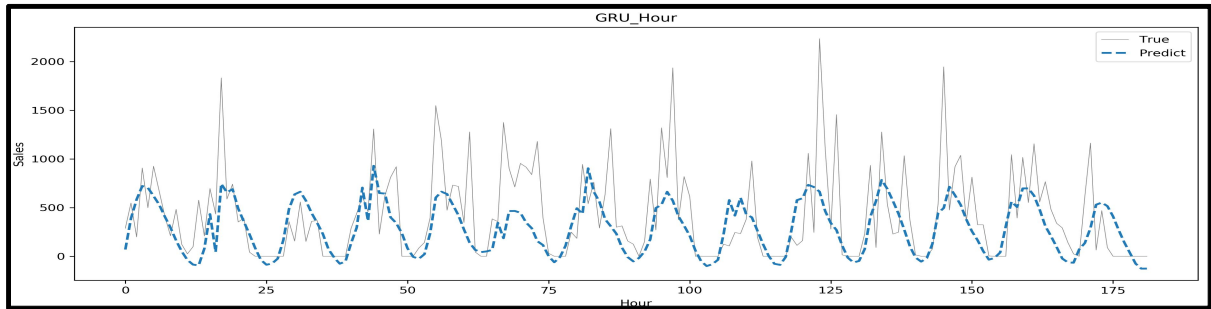


Figure 8a. Results of hourly prediction by GRU

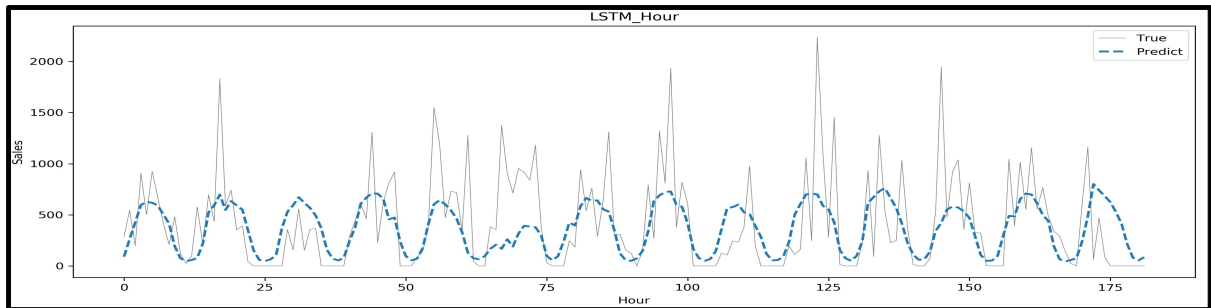


Figure 8b. Results of hourly prediction by LSTM

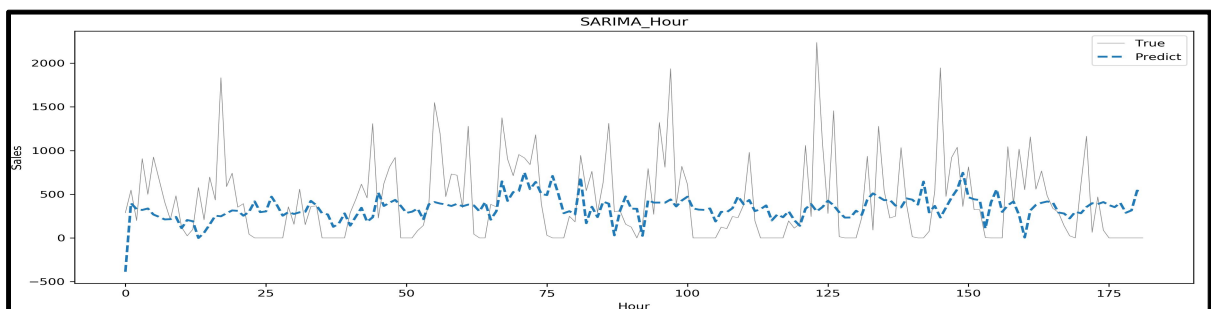


Figure 8c. Results of hourly prediction by SARIMA

3. 결론 및 향후 과제

본 연구에서는 온라인 리테일러에게 접수되는 수요 데이터 중 상위 77개 품목에 대한 시간대별 수요패턴의 예측력을 제고하기 위하여 이미지 시계열 분야에서 활용되고 있는 컨볼루션 LSTM 기법을 활용하였다. 즉, 시간적 특성과 품목별 특성으로 정의된 2차원 공간상에서 수요량을 이미지화하는 과정을 통한 해당 기법의 도입 방안을 제시하였다. 연구에서 진행한 구체적인 적용단계는 1) 고객 주문 데이터를 15분 단위로 품목별 합을 계산하고, 2) 구간별 수요를 RGB값으로 대응 시켜 3) 2차원 이미지 형태로 바꾸어 주었고, 4) 이미지화 된 데이터를 통한 수요예측을 진행하였다. 또한, 제안된 방식의 수요패턴 예측성능을 검증

하고자 Holt-윈터스 기법, SARIMA과 함께 딥러닝 시계열 모델인 GRU, LSTM 과의 예측 정확도를 비교하였다.

고객주문 데이터를 이용하여 진행한 일별 수요예측 및 시간대별 수요예측 등 2가지 수치실험 결과에 따르면, 본 연구에서 제안하는 방식의 ConvLSTM 모형의 일별 수요패턴 예측 성능은 기존의 시계열 분석 기법인 SARIMA와 비슷한 수준이었으며, 다른 딥러닝 모델들과도 비슷하거나 약간의 성능 향상이 있었다. 반면에, 시간대 별 고객수요에 대한 예측성능은 고려한 다른 방법/모델들과 비교하였을 때 모든 지표에서 우수한 성능을 보여주었다. 이는 ConvLSTM이 이미지의 시간적인 특성뿐만 아니라 공간적인 특성도 예측에 반영하기 때문인데, 수요 데이터를 이미지화 시키면 품목별 15분간의 주문 수량이 공간적인 특성으로 바뀌어서 상대적으로 우수한 예측성능을 낼 수 있었던 것으로 추정된다. 따라서, 본 연구에서 제안하는 ConvLSTM 기법을 통해서 단기 구간(예: 2시간 이내의 예측 수요패턴)에서의 우수한 수요예측력을 활용한다면 온라인 리테일 회사에서의 단기 구간 수요에 대한 사전준비를 통해서 물류시스템 실행력 제고가 가능할 것으로 예상된다. 구체적으로는 실시간 수요에 대한 예측력에 기반하여 고객수요 충족을 위한 대응력을 향상하고, 물류시스템 운영자원의 효율적인 운영에도 의미 있는 개선을 도출할 수 있을 것으로 기대된다. 또한 본 연구에서는 비교군과의 비교를 위하여 불가피하게 품목별로 예측한 결과를 전체 품목에 대해 합산하여 예측 정확도를 평가하였지만, 실험 중 확인한 결과 품목별 수요패턴 예측에서도 높은 성능을 보여주었다. 이를 활용하면 품목별 실시간 수요에 대한 예측도 가능할 것으로 예상된다.

본 연구에서는 온라인 리테일 회사의 1년 치 데이터를 사용하였는데, 딥러닝 모델의 특성상 많은 수의 파라미터를 추정하기 때문에 데이터가 적은 경우 파라미터를 적절하게 추정하지 못한다. 이는 일자별 수요예측과 시간대별 수요예측의 성능 비교에서도 확인할 수 있는데, 일자별 수요예측의 경우 305개의 데이터를 사용했고, 시간대별 수요예측은 3965개의 데이터를 사용하였기 때문에 일자별 수요예측에서 SARIMA에 비해서 낮은 성능을 보이던 GRU와 LSTM이 시간대별 수요예측에서 높은 성능을 보여주었다. ConvLSTM의 경우 하루를 하나의 이미지로 만든 데이터를 사용하였고, 이미지 데이터 안에 시간대별 수요가 포함되어 있기 때문에 시간대별 예측에서도 동일한 305개의 데이터만 사용할 수 있었다. 따라서 추후 더 많은 양의 데이터를 사용해서 학습한다면 보다 더 높은 성능을 보일 수 있을 것으로 예상된다. 또한 위의 <데이터 전처리> 부분에서 밝힌 바와 같이, 본 연구에서 사용된 원본 데이터에는 총 4082개의 품목이 있었지만 15분 단위로 품목별 주문 수량을 합해주면 77개의 구간이 나오기 때문에 행의 개수와 열의 개수가 일치하도록 하기 위하여 상위 77개의 품목만을 예측에 사용하였다. 구간을 더 작은 단위의 시간으로 나누고, 더 많은 품목을 사용하여 예측을 실시한다면 더 좋은 성능을 낼 수 있을 것으로 기대한다. 또한 본 연구에서는 실험 환경의 메모리 제약으로 커널의 크기를 3×3 보다 크게 하거나 레이어의 깊이를 4단 이상으로 할 경우 실험을 진행할 수 없었다. 따라서 파라미터를 다르게 설정함으로써 모델이 향상된 성능을 낼 가능성이 있음을 기대한다.

본 연구에서 사용된 이미지 시계열 예측 기법인 ConvLSTM은 Shi et al.(2015)의 연구에서도 지적인 예측 단위가 멀어질수록 이미지에 노이즈가 발생하는 한계점이 있다. 본 연구의 실험에서도 위와 같은 현상을 확인할 수 있었으며, 이는 예측일자가 멀어지면 Red와 Green만으로 표현된 데이터에서 Blue값이 조금씩 나오는 형태로 나타났다. 본 연구에서는 이를 해결하기 위해서 예측일자를 바로 다음 일자로 설정하였으나, 노이즈를 해결하고 예측일자를 늘리기 위해서는 추후 논의가 필요하다.

참고문헌

- Creswel, A. Arulkumaran, K. Bharath, A. A.(2017), *On denoising autoencoders trained to minimise binary cross-entropy*, arXiv
- Olah, C.(2015), *Understanding LSTM Networks*, <http://colah.github.io/posts/2015-08-Understanding-LSTMs/>
- Jung, D(2004), *Prediction of the number of domestic breeding hog Using Winters Model*, Kyung-sung University, Republic of Korea
- Lee, H. Kim, K. Jeong, H. Lee, H. Kim, H. Park, J.(2020), *A Study on Wind Power Forecasting Using LSTM Method*, *The transactions of The Korean Institute of Electrical Engineers*, 69(8), 1157-1164
- Kim, H.(2017), *Prediction of Time Series Data Using LSTM/GRU Recurrent Neural Network*, Korea National Open University, Republic of Korea
- Hochreiter, S. and Schmidhuber, J. (1997). Long short-term memory. *Neural Computation*, 9(8), 1735-1780
- Jeong, H. , Lim, C.(2019), *A review of artificial intelligence based demand forecasting techniques*, *The Korean Journal of Applied Statistics*, 32(6), 795-835
- Song, I., Cho, N.(2015), *Forecasting based Optimal Storage Policy Model of a Warehouse Management System*, *Journal of the Korean Society of Supply Chain Management*, 15(1), 1-7
- Lee, J., Kwon, Y.(2011), *A Study on Dynamic Change of Transportation Demand Using Seasonal ARIMA Model*, *Journal of Korean Society of Transportation*, 29(5), 139-155
- Jong-Tae-Kim(2009), *Forecasting number of student by Holt-Winters additive model*, *Journal of the Korean Data And Information Science Society*, 20(4), 685-694
- Nenni, M. E., Giustiniano, L., Pirolo, L.(2013) *Demand forecasting in the fashion industry: A review*, *International Journal of Engineering Business Management*
- Lee, S., Sun, Y., Lee, J., Lee, D., Cho, E., Park, D., Kim, Y., Sim, I., Kim, J.(2019), *Short-term Power Consumption Forecasting Based on IoT Power Meter with LSTM and GRU Deep Learning*, *The Journal of The Institute of Internet, Broadcasting and Communication*, 19(5), 79-85
- Kim, S., Jung, H., Park, J., Baek, S., Kim, W., Chon, K., Song, K.(2014), *Weekly Maximum Electric Load Forecasting for 104 Weeks by Seasonal ARIMA Model*, *Journal of the Korean Institute of Illuminating and Electrical Installation Engineers*, 28(1), 50-56
- Chaudhuri, T. , Ghosh, I.(2016), *Artificial Neural Network and Time Series Modeling Based Approach to Forecasting the Exchange Rate in a Multivariate Framework*, *Journal of Insurance and Financial Management*, 1(5), 92-123
- Weng, T., Liu, W., Xiao, J.(2019), *Supply chain sales forecasting based on lightGBM and LSTM combination model*, *Industrial Management & Data Systems*, 120(2), 265-279
- Shi, X., Chen, Z., Wang, H., Yeung, D.(2015), *Convolutional LSTM network: A machine learning approach for precipitation nowcasting*, *Advances in neural information processing systems 28(NIPS2015)*, 802-810
- Ai, Y., Li, Z., Gan, M., Zhang, Y., Yu, D, Chen, W., Ju, Y.(2018), *A deep learning approach on short-term spatiotemporal distribution forecasting of dockless bike-sharing system*, *Neural Computing and Applications*, 31, 1665-1677

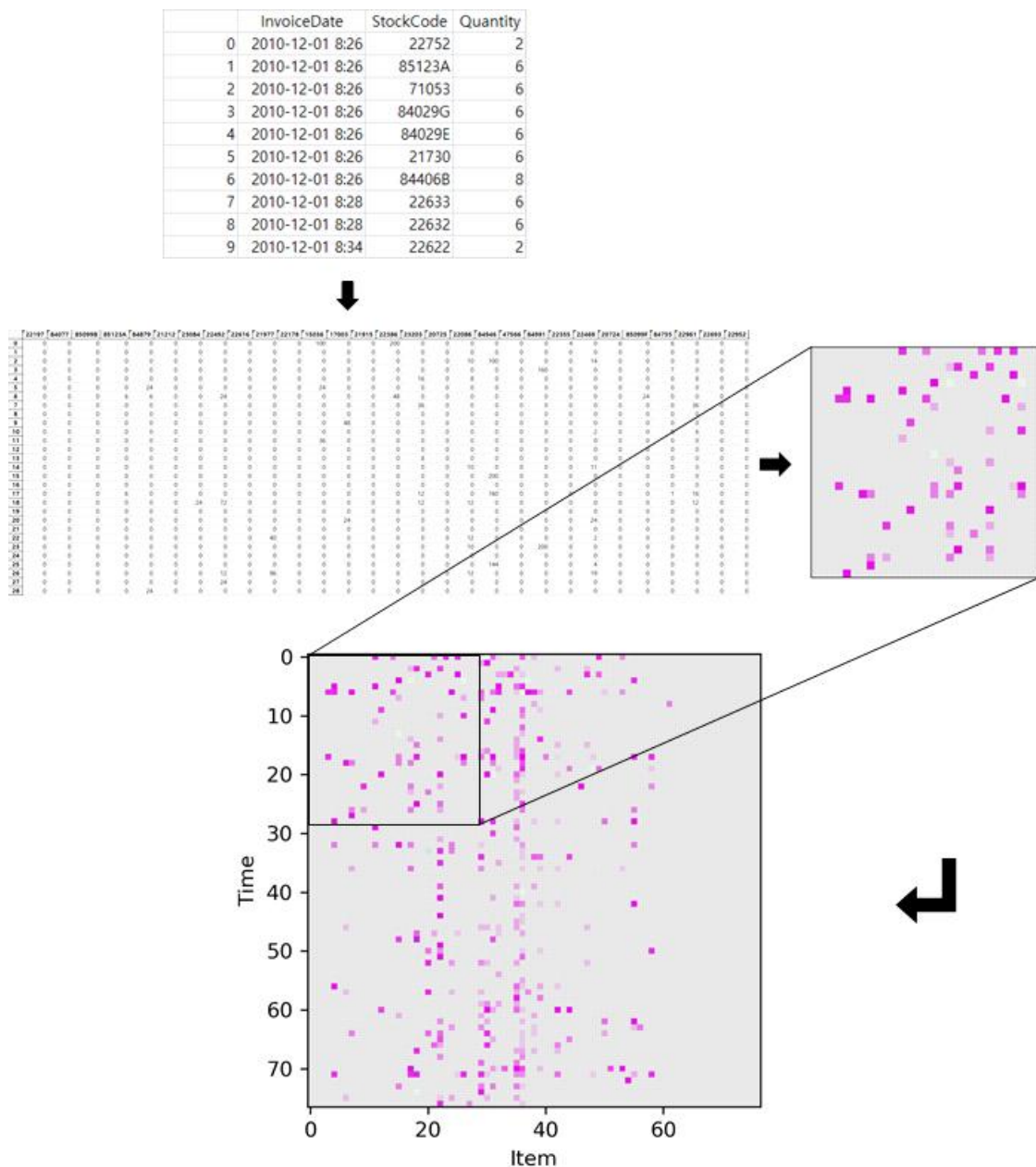
Chang, Y., Liao, M.(2010), *A Seasonal ARIMA Model of Tourism Forecasting: The Case of Taiwan*, *Asia Pacific Journal of Tourism Research*, 15(2), 215-221

Wu, Y. , Wu, Q. , Zhu, J.(2019), *Improved EEMD-based crude oil price forecasting using LSTM networks*, *Physica A: Statistical Mechanics and its Applications*, 516, 114-124

Coupang(2020), *Established a cutting-edge distribution center in Gimcheon-si*, <https://news.coupang.com/archives/5058>

UCI(2015), *Online Retail Data Set*, <https://archive.ics.uci.edu/ml/datasets/online+retail>

Appendix



Appendix a. Process of generating the images for corresponding demand patterns