

LSTM 모델을 이용한 수소 충전소의 일일 수소 판매량 예측 연구

김민수[†] · 전성탁 · 정태영

한국전력기술 전력기술연구원

A Study on the Daily Demand Forecasting of Hydrogen Charging Station Using LSTM Model

MINSU KIM[†], SUNGTAK JEON, TAEYOUNG JYUNG

Power Technology Research Institute, KEPCO E&C, 269 Hyeoksin-ro, Gimcheon 39660, Korea

[†]Corresponding author :
sriiv@kepc-enc.com

Received 4 September, 2024
Revised 8 October, 2024
Accepted 15 October, 2024

Abstract >> This study analyzes one year of hydrogen sales data from an off-site hydrogen charging station in South Korea to identify factors influencing sales and develop a forecasting model using long short-term memory (LSTM) model. The LSTM model's performance was compared with the existing traditional time series models, seasonal auto-regressive integrated moving average (SARIMA) and SARIMA with exogenous factors (SARIMAX). Key factors such as the day of the week, holidays, and weather significantly impact sales, with comparatively higher sales observed on Fridays, weekends, holidays and clear days. The LSTM model outperformed SARIMA and SARIMAX, effectively capturing nonlinear relationships between sales and factors. In addition, the forecasting model proposed in this study showed improved performance compared to the forecasting model proposed in previous study. Despite some limitations, this research provides valuable insights for improving operational efficiency at hydrogen charging stations.

Key words : Hydrogen charging station(수소 충전소), Hydrogen sales(수소 판매량), Deep learning(딥러닝), Demand forecasting(수요 예측), Time series forecasting(시계열 예측)

1. 서론

현재 온실가스 배출 증가로 인한 급격한 기후 변화가 발생하면서 전 세계적으로 탄소중립과 같은 친환경에너지 사용에 대한 움직임이 가속화되고 있다.

특히 Renewable Energy 100%와 같은 정책적 움직임이 이러한 지구 온난화 문제를 해결하기 위한 대표적인 정책적 사례이다. 이러한 배경에서 수소에너지는 탄소 배출이 거의 없는 청정에너지원으로 주목받고 있으며 산업용 원료뿐만 아니라 발전, 수송, 산업,

건축 분야 등 다양하게 소비될 것으로 기대되고 있다. 특히 수소에너지는 높은 충전 속도와 에너지 저장 밀도 등의 장점으로 인해 수소차, 수소버스와 트럭 등 수송 부문에서 높은 수요가 예상된다¹⁾. 실제로 수소차는 같은 친환경 차량으로 주목받는 전기차와 비교해서도 충전 시간이 짧고 1회 충전 시 주행 거리가 길다. 또한 수소차는 연료 충전 과정이 내연기관 차량과 유사하기 때문에 고객들이 연료 충전 행동 패턴을 크게 변경할 필요가 없다는 장점이 있다^{2,3)}.

이러한 장점에도 불구하고 현재 수소차 시장의 성장이 더딘 주된 이유는 높은 연료 비용과 충전 인프라 부족 때문이며 특히 후자가 수소차 시장 확대에 가장 큰 장애물로 작용하고 있다. 실제로 2020년 말 기준으로 한국은 전 세계에서 가장 많은 대략 1만 대의 수소차를 보유하고 있었지만 수소 충전소는 약 50개에 불과하여 충전소 1기당 수소차 200대를 감당해야 하는 상황이다⁴⁾. 이와 같이 충전소의 개수가 부족할 뿐만 아니라 수소 공급처의 부족과 고장, 지역별 충전 인프라의 불균형으로 인해 줄서기 현상, 1회 충전 시 차량당 충전량 제한 등 여러 문제가 발생하고 있다⁵⁾. 이러한 문제들은 수소차 고객들에게 불편을 초래하고 수소차 보급 확대에도 부정적인 영향을 미치고 있다.

이러한 문제를 해결하기 위하여 본 연구에서는 수소 충전소의 일일 판매량 예측을 제안한다. 일일 판매량 예측을 통해 충전소 운영자는 미래의 수요 변동을 예측하여 효율적인 수소 공급 계획을 수립할 수 있으며 재고 관리를 최적화하여 고객 만족도를 향상시킬 수 있다. 또한 상위 기관이나 수소 공급처에서 담당 충전소의 공급량을 관리할 경우 예측 데이터를 활용해 충전소별로 수소를 적절히 배분하여 충전 대란, 재고 소진으로 인한 운영 중단 등의 문제들을 사전에 방지함으로써 수소차 보급 확대에도 긍정적인 영향을 미칠 수 있을 것이다.

따라서 본 연구에서는 국내의 주거 지역에 위치한 off-site 수소 충전소의 1년간의 수소 판매량 데이터를 분석하여 판매량에 영향을 끼치는 요인을 도출하고 long short-term memory (LSTM) 모델을 활용하여

수소 충전소의 일일 수소 판매량을 예측하였다. 또한 성능 검증을 위해 전통적 시계열 모델인 seasonal autoregressive integrated moving average (SARIMA), SARIMA with exogenous factors (SARIMAX) 모형과 예측 성능을 비교하였다.

2. 선행 연구

현재 수소차 시장은 초기 확장 단계에 있어 수소 판매량 데이터가 충분히 축적되지 않은 상황이다. 또한 보안상의 이유로 데이터에 대한 일반인의 접근이 제한되어 있어 수소 충전소의 일일 판매량 예측과 관련된 선행 연구는 부족한 실정이다. 그럼에도 불구하고 수소 판매량 데이터를 활용하여 예측 모델을 설계한 연구가 국내외에 일부 존재한다(Table 1).

Kwon 등⁶⁾은 수소 판매량 데이터를 통해 고객의 소비 패턴을 파악하여 연간 수요를 예측하는 방정식을 도출하여 충전소의 증설을 제안하였다. Kurtz 등⁷⁾은 수소 충전소와 주유소의 판매 데이터를 기반으로 확률 모델과 포아송 과정(poisson process)을 사용하여 수소 충전소의 시간별 수요를 예측하는 방법을 제안하였다. 이 연구에서는 수소차의 충전 빈도와 충전량의 분포를 확률적으로 모델링하여 미래의 수요를 예측하고 마코프 연쇄(Markov chain)를 활용하여 시간별 충전소 상태를 예측함으로써 충전소 운영의 효율성을 극대화 할 수 있는 방법을 제시하였다. Sun 등⁸⁾은 수전해 기반의 on-site 수소 충전소의 최적 구성 및 운영 계획을 위해 계시별 요금제를 활용하여 시간별 수소 생산을 최적화하고 설비 투자 및 운영 비용을 최소화하는 모델을 제안하였다. 이 연구는 전기차의 주행 패턴을 수소차에 적용하여 개인 승용차와 택시의 주행 시간과 거리가 정규 분포와 로그 정규 분포를 따른다는 가정을 바탕으로 충전 확률과 충전량을 예측하고 이를 통해 시간별 수소 수요를 추정하였다. Hwang 등⁹⁾은 수소 충전소의 효율적인 수급 관리를 위해 머신러닝과 딥러닝 기반의 일일 판매량 예측 모델을 제안하였다. 이 연구에서는 강원도 지역의 3개의 수소 충전소 판매 데이터와 기상 데

Table 1. Summary of related works on hydrogen charging station and gas station's sales forecasting

Study	Forecasting for	Method	Features
Kurtz et al ⁷⁾ . (2020)	Hydrogen sales	Poisson process	-
Sun et al ⁸⁾ . (2022)	Hydrogen sales	Normal distribution	-
Hwang et al ⁹⁾ . (2023)	Hydrogen sales	Linear regression, hist gradient boosting, deep neural network	Weather
Pan et al ¹⁰⁾ . (2022)	Gas sales	BiLSTM, genetic algorithm	Weather, date type, holiday, temperature, sales price
Sun et al ¹¹⁾ . (2018)	Gas sales	K-means clustering, decision tree	Weather, date type, holiday, sales price, sales promotion
Henzel et al ¹²⁾ . (2021)	Gas sales	XGBoost, KNN, ARIMA	Date type, holiday

이터를 사용하였고 선형 회귀, 그래디언트 부스팅, 딥러닝 등 다양한 예측 모델의 성능을 비교하였다.

고객 입장에서 수소차는 내연기관 차량과 유사한 방식으로 연료를 충전할 수 있다는 점^{2,3)}을 고려하여 휘발유 주유소의 일일 판매량 예측 관련 선행 연구들도 추가로 조사하였다.

Pan 등¹⁰⁾은 휘발유 주유소의 효율적인 수급 관리를 위해 딥러닝 기반의 일일 판매량 예측 모델을 제안하였다. 이 연구에서는 판매 데이터를 분석하여 요일, 기온, 날씨, 공휴일, 판매 단가가 판매량에 미치는 영향을 확인하였으며 특히 내일의 날씨와 공휴일 여부가 오늘의 판매량에 영향을 미친다는 점을 활용하여 역방향 학습이 가능한 BiLSTM 모델을 사용하였다. 또한 예측 모델의 정확도를 높이기 위하여 유전 알고리즘(genetic algorithm)을 결합하여 하이퍼 파라미터를 최적화하였다. Sun 등¹¹⁾은 미래의 일별 주유소 판매량 예측을 위해 K-means 클러스터링을 사용하여 데이터를 유사성을 가지는 군집으로 분류한 후 decision tree를 활용하는 예측 모델을 제안하였다. 이 과정에서 공휴일 여부, 날씨, 판매 가격, 홍보 활동 여부 등 수요에 영향을 미치는 주요 요인을 기준으로 데이터를 분류하여 예측의 정확성을 높였다. Henzel 등¹²⁾은 시계열 클러스터링 기법이 주유소 연료 판매 예측 결과에 미치는 영향을 분석한 연구를 수행하였다. 이 연구에서는 폴란드에 위치한 25개 주유소에서 수집된 50개의 시계열 데이터를 사용하였으며 특히 XGBoost 알고리즘을 활용하여 여러 주

유소 데이터를 통합한 예측 모델이 단일 시계열 모델보다 뛰어난 성능을 보였음을 입증하였다. 이 연구에서는 공휴일, 판매 가격 등 여러 요인을 반영하여 예측 모델의 정확성을 더욱 강화하였다.

현재까지 조사한 선행 연구에 따르면 주유소 판매량 예측 관련 선행 연구들은 시계열 모형, 머신러닝과 딥러닝 기반의 예측 모델을 사용하였다. 또한 판매량과 여러 요인 간의 상관관계를 파악하여 예측 모델 설계에 사용하였다. 반면에 수소 충전소 판매량 예측 관련 선행 연구들은 Hwang 등⁹⁾의 연구를 제외하고는 통계적 모델링에 의한 추정 방식을 사용하고 있었으며 주유소 판매량 예측 관련 선행 연구들에서 고려한 요일, 평일과 주말 여부, 공휴일 여부, 날씨 등 다양한 요인들이 판매량에 영향을 끼치지만 이에 대한 고려가 이루어지지 않았다. 특히 Philipsen 등¹³⁾이 실시한 전기차와 내연기관 차량의 운전자들의 연료 충전에 대한 설문조사에 따르면 운전자들은 충전 빈도와 충전량을 연료 잔량, 재정 상태, 개인의 충전 습관 등 개인이 처한 다양한 상황에 의거하여 판단한다. 이는 수소차 운전자들에 대해서도 유사하며 이는 충전이 무작위성을 가지고 있어 케이스별 통계적 모델링으로는 예측이 어렵다는 것을 의미한다. 실제로 Kim 등¹⁴⁾과 Kang 등¹⁵⁾이 실시한 수소 충전소 판매량 데이터 분석에 따르면 시간대, 요일, 날씨와 같이 다양한 요인에 의해 영향을 받는다고 한다. 이에 따라 수소 충전소의 수요 예측의 경우에도 주유소 수요 예측과 같은 복잡한 시퀀스 패턴을 포착할 수

있는 방법론 접근이 필요하다. Hwang 등⁹⁾이 머신러닝과 딥러닝 기반의 예측 모델을 제안하였으나 판매량에 영향을 많이 미치는 요일, 공휴일 여부를 고려하지 않았다.

이러한 배경에서 본 연구에서는 운전자의 관점에서 내연기관 차량과 수소차의 충전 과정이 유사함을 고려하여 휘발유 주유소의 판매량 예측 연구의 방법론을 수소 충전소 판매량 예측에 적용하였다. 또한 본 연구는 실제 수소 충전소 판매량을 분석하여 영향을 끼치는 요인을 추가로 고려하여 딥러닝 기반 시계열 예측 모델을 설계함으로써 개선된 예측 성능을 보였다는 점에서 기여가 있다.

3. 방법론

본 연구는 일일 수소 판매량을 예측하기 위하여 데이터 전처리 및 탐색적 데이터 분석에 따른 feature selection과 feature transformation에 해당되는 embedding layer 및 one-hot encoding을 사용한 범주형 변수의 벡터화, min-max scaling을 이용한 정규화, LSTM 모델을 사용한 최종 예측 그리고 검증 및 분석으로 이루어진다(Fig. 1). 본 절에서는 본 연구에서 사용된 주요 모델들의 설명과 선택 이유에 대해 기술한다.

3.1 Embedding layer

예측 모델에서 feature로 사용되는 요일, 공휴일 여부, 날씨는 범주형 변수로 그 크기와는 관련성이 없는 변수들이다. 이러한 범주형 변수에 대한 처리로 일반적으로 사용하는 방법은 one-hot encoding으로 각 범주에 대해 하나의 이진변수를 생성하고 해당 범주에 속하면 1, 속하지 않으면 0을 부여하는 방법이다. 그러나 범주의 수만큼 차원이 확장되기 때문에 범주가 많아지면 데이터의 차원이 급격하게 증가하여 모델의 복잡성이 증가하는 단점이 있다. 또한 one-hot encoding으로 생성된 데이터는 대부분이 0으로 구성된 희소 행렬로 메모리의 공간을 낭비하는 결과를 초래한다. 마지막으로 범주형 변수 간의 유사

성을 표현할 수 없다. 예를 들어 토요일과 일요일은 주말의 특성으로 수소 판매량이 비슷한 수준일 것이나 그 유사성을 표현할 수 없다.

Embedding layer는 딥러닝 신경망을 이용하여 텍스트의 단어나 토큰을 고차원 벡터 공간에 매핑하는 word to vector (Word2Vec)에 사용되며 매핑된 벡터는 단어 간의 의미론적 유사성을 보존하도록 설계하는 자연어 처리 방식 중 하나이다. 즉 의미론적으로 비슷한 단어들은 공간상에서 서로 가깝게 표현되도록 한다. 이러한 embedding layer는 Word2Vec에 국한되지 않고 categories to vector (Cat2Vec)로도 사용되어 범주형 변수 간 유사도를 계산할 수 있는 연속적인 공간으로 매핑함으로써 범주형 변수 간 유사성을 쉽게 파악할 수 있고 더 작은 차원의 벡터로 표현할 수 있기 때문에 메모리 사용량을 효과적으로 줄일 수 있다는 장점이 있다.

3.2 LSTM

시계열 예측을 위한 딥러닝 모델에는 대표적으로 recurrent neural network (RNN), LSTM, gated recurrent unit (GRU)이 있다. RNN은 입력 데이터의 순차적인 특성을 학습할 수 있는 모델이지만 출력층에서 멀어질수록 그래디언트가 매우 작아지는 그래디언트 소실 현상이 발생하여 시계열 데이터의 장기 의존성을 학습하기엔 어려움이 있다. 이 한계를 극복하기 위해 LSTM이 개발되었으며 이 모델은 RNN의 구조를 개선하여 장기 의존성을 학습할 수 있다.

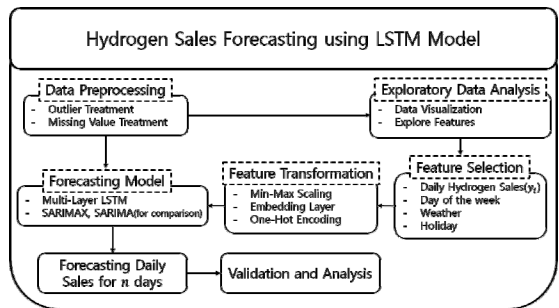


Fig. 1. Flow chart of the proposed forecasting model

GRU는 LSTM의 단순화된 구조로 계산량과 메모리 사용량을 줄여 학습 속도가 빠르다는 장점이 있으나 장기적인 패턴 학습에는 성능이 저하될 수 있다. 본 연구에서는 수소 판매량 데이터에서 확인되는 장기적인 패턴을 보다 정확하게 학습하기 위하여 LSTM 모델을 사용하였다.

LSTM 모델은 이전 정보를 오랫동안 기억할 수 있는 메모리 셀을 가지고 있으며 이를 통해 긴 시퀀스 데이터를 처리할 수 있다는 장점이 있다. 이러한 LSTM의 메모리 셀은 새로운 정보가 셀 상태에 얼마나 반영될지를 결정하는 입력 게이트, 이전 셀 상태의 정보를 얼마나 잊을지를 결정하는 망각 게이트, 셀 상태에서부터 어떤 정보를 출력할지를 결정하는 출력 게이트로 구성된다(Fig. 2)¹⁰⁾.

LSTM은 hidden layer의 수에 따라 single-layer

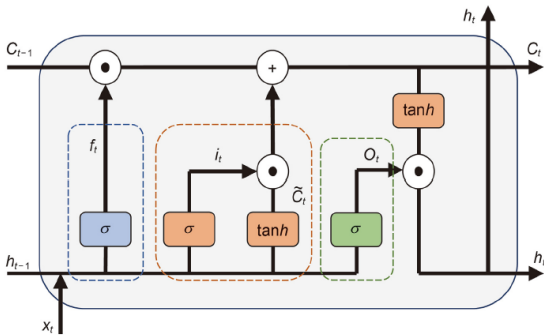


Fig. 2. Structure of LSTM cell¹⁰⁾

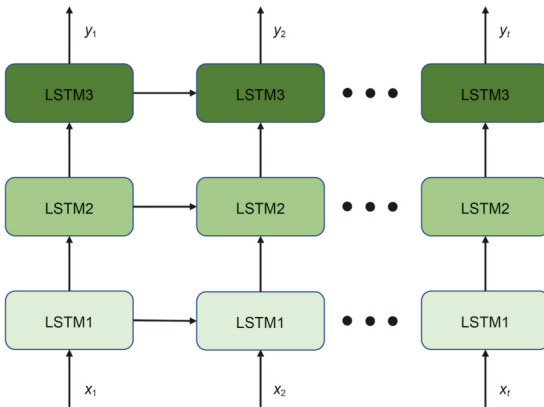


Fig. 3. Structure of multi-layer LSTM model¹⁰⁾

LSTM과 multi-layer LSTM으로 구분된다. Multi-layer LSTM은 여러 개의 LSTM layer를 층층이 쌓아 올린 모델 구조를 의미하며 각 LSTM layer의 출력이 다음 layer의 입력으로 전달되는 방식으로 구성된다. 이를 통해 layer의 개수가 증가할수록 더 복잡한 패턴을 학습할 수 있는 능력이 향상된다. 그러나 layer가 깊어질수록 학습에 소요되는 시간과 계산 비용이 증가하는 단점도 있다. 이러한 multi-layer LSTM의 구조를 Fig. 3에 나타내었다.

4. 모델 적용

본 연구의 데이터 분석 및 예측은 Python 3.9.16 (Python Software Foundation, Wilmington, DE, USA) 버전 환경에서 수행되었다. 데이터 시각화를 위하여 matplotlib과 seaborn 라이브러리가 활용되었으며 예측 모델 구축에는 딥러닝 라이브러리인 Keras를 사용하여 LSTM 모델을 구현하였다. 또한 statsmodels 라이브러리를 사용하여 SARIMA와 SARIMAX 모델을 구현하였다.

4.1 데이터 소개 및 특성 분석

본 연구에서 사용한 데이터는 국내의 주거 지역에 위치한 off-site 수소 충전소의 2021년 4월부터 2022년 4월까지 수소 충전소의 SCADA 시스템에서 time-log 형식으로 수집된 약 1년간의 판매 데이터이며 판매된 수소 및 이용 현황을 Table 2에 나타내었다. 수소 충전소에는 일 평균 19.2대의 수소차가 방문하여 일 평균 100.9 kg의 수소가 판매되었다. 일 최대 방문 대수는 35대, 일 최대 수소 판매량은 167 kg이며 수소차 1대당 최대 판매량은 28 kg으로 나타났다. 수집된 기간 동안 총 6,869대의 수소차가 방문하였으며 총 판매량은 36,138 kg에 달하였다.

해당 수소 충전소는 데이터가 수집된 기간 동안 운영 시간이 요일과 상관없이 9-20시이고 연중무휴로 운영되었으며 충전 가능한 차종은 승용차와 버스이고 충전기는 총 3대이다. 본 연구에서는 데이터가

수집된 수소 충전소의 전체 일일 판매량을 예측하기 위하여 충전기와 차종에 구분 없이 일일 기준으로 집계하였으며 운영 시간 외에 기록된 판매량은 이상치로 간주하여 삭제하였다. 또한 확인된 7일의 결측치에 대해서는 다수의 선행 연구에서 언급한 요일의 특성을 유지하기 위하여 해당 요일의 전주와 차주의 평균으로 처리하였다.

2절의 다수의 선행 연구들에서 소개된 바와 같이 주유소의 휘발유 판매량뿐만 아니라 수소 충전소의 판매량 또한 다양한 요인에 의해 변동될 수 있다. 선행 연구들이 소개한 요인을 토대로 본 연구에서 대상으로 하는 수소 충전소의 데이터를 분석하였다.

4.1.1 요일

수소 충전소의 요일에 따른 판매량을 Fig. 4에 나타내었으며 월요일은 85.7 kg, 화요일은 93.7 kg, 수요일은 92.7 kg, 목요일은 95 kg, 금요일은 104.7 kg, 토요일은 117.8 kg, 일요일은 116.5 kg으로 요일에 따라 판매량에 변동이 있음을 확인할 수 있다. 이는 금요일과 주말은 운전자의 이동거리 및 충전할 시간적 여유가 많아 평일 대비 판매량이 많은 것으로 판단된다^{14,15}.

4.1.2 공휴일

수소 충전소의 공휴일에 따른 판매량을 Fig. 5에 나타내었으며 공휴일 여부에 따라 판매량에 변동이 있음을 확인할 수 있다. 공휴일이 아닌 경우에는 100.3 kg,

공휴일의 경우에는 110.2 kg으로 공휴일에 판매량이 대략 10 kg 높은 것을 확인할 수 있다. 이는 금요일과 주말과 마찬가지로 운전자의 이동거리 및 충전할 시간적 여유가 많기 때문인 것으로 판단된다.

4.1.3 날씨

날씨에 따른 판매량의 변동을 확인하기 위하여 기상자료 개방 포털¹⁷)에서 수소 충전소가 위치한 지역과 데이터가 기록된 기간에 대하여 강수량, 풍속, 적설량의 시간별 데이터를 취득하였다. 풍속의 결측치에 대해서는 두 점의 값이 주어졌을 때 그 사이의 값을 직선거리에 따라 선형적으로 계산하는 선형 보간법(linear interpolation)으로 처리하였고 강수량과 적설량의 결측치에 대해서는 0으로 처리하였다.

일별 판매량에 대한 분석이기 때문에 시간별 날씨

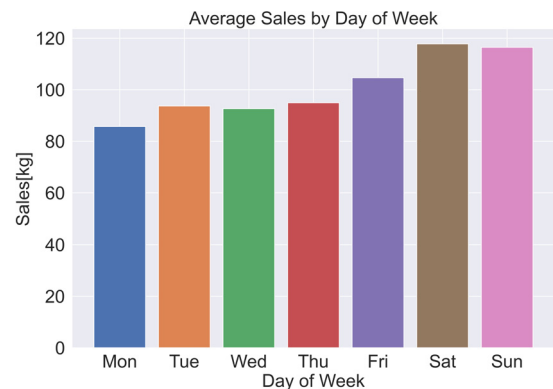


Fig. 4. Average sales by day of week

Table 2. Sales value of hydrogen charging station

Item	Value
Average number of visiting vehicles per day	19.2 vehicles/day
Average hydrogen sales per day	100.9 kg/day
Maximum number of visiting vehicles per day	35 vehicle/day
Maximum hydrogen sales per day	167 kg/day
Maximum sales of hydrogen per vehicle	28 kg/vehicle
Total number of visiting vehicle	6,869 vehicles
Accumulate sum of hydrogen sales	36,138 kg

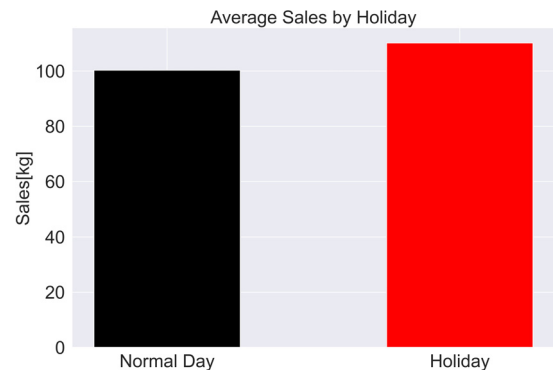


Fig. 5. Average sales by holiday

데이터를 변환하여 해당 일의 날씨를 대표할 수 있도록 하는 변환이 필요하다. 우선 수소 충전소의 운영 시간에서 기록된 시간별 강수량, 풍속, 적설량의 데이터를 일별 평균값으로 설정하였다. Pan 등¹⁰⁾과 Sun 등¹¹⁾에 따르면 주유소 판매량의 경우에는 태풍, 천둥 번개와 같은 극한의 날씨가 판매에 큰 영향을 줄 수 있다고 하나 본 연구에서 대상으로 하는 수소 충전소가 위치한 지역에서는 태풍, 눈보라 등의 극한 날씨가 존재하지 않고 전반적으로 온화한 날씨를 가지므로 Table 3과 같이 날씨를 상대적 기준으로 분류하였다. 강수량과 적설량이 0인 맑은 날씨는 sunny 클래스로, 풍속과 강수량이 상위 50%인 상대적 극한 날씨는 intense 클래스로, 나머지 기타 날씨는 other 클래스로 분류하였다. 단 적설량의 경우에는 기록된 기간이 많지 않은 관계로 기타 날씨에 포함시켰다. Table 3의 날씨 분류 기준에 의한 평균 판매량을 Fig. 6에 나타내었는데 intense 클래스는 92.4 kg, other 클래스는 101.1 kg, sunny 클래스는 101.5 kg으로 intense 클래스의 판매량이 비교적 저조한 것을 확인할 수 있다.

4.2 Feature

4.2.1 Feature selection

수소 충전소 미래의 일일 판매량 예측을 위하여 training set과 test set을 80:20으로 분류하였으며 선정한 feature는 Table 4와 같다. y_i 는 시계열 데이터 그 자체인 시간별 수소 판매량이고 판매량에 영향을 끼치는 요인은 4.1절의 데이터 특성 분석 시 소개한 요일, 공휴일, 날씨를 최종 선정하였다. 이외에 판매

Table 3. Weather classification based on relative criteria

Weather class	Description	Days
Sunny	A day when precipitation and snow cover are 0	273
Intense	A day when wind speed and precipitation are in the top 50%	24
Other	Days not included in other criteria	68

단가를 추가적으로 고려할 수 있으나 데이터가 수집된 기간 동안 변동이 없어 제외하였다.

4.2.2 Feature transformation

Table 4에서 선정한 feature는 모두 범주형 변수이다. 요일은 월요일부터 일요일까지의 7개 범주, 공휴일은 공휴일 여부로 2개의 범주, 날씨는 3개의 범주로 총 12개의 범주형 변수를 가진다. 범주가 오직 2개인 공휴일 여부는 기존처럼 one-hot encoding을 적용하고 요일과 날씨에 대해서는 embedding layer를 사용하여 벡터로 변환한다.

Embedding layer를 사용할 때 출력되는 벡터의 차원을 적절히 설정하는 것이 중요하다. Wagner 등¹⁸⁾에 따르면 벡터의 차원이 너무 높으면 과적합이 발생할 수 있고 반대로 너무 낮으면 범주형 변수의 의미를 포착하지 못할 수 있다. 따라서 Wagner 등¹⁸⁾이 제안한 범주형 변수의 개수인 15-35% 사이에 가장 좋은 성능을 가지는 값을 벡터의 차원으로 결정하였다. 이에 따라 training set의 요일과 날씨 변수를 y_i 와 학습시켜 각각 3차원과 1차원의 벡터가 출력되었으

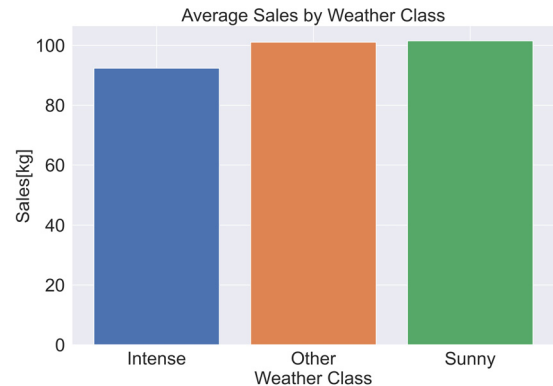


Fig. 6. Average sales by weather class

Table 4. Features for forecasting using LSTM model

Features	Description
y_i	Daily hydrogen sales
Day of the week	Monday to Sunday
Holiday	Holiday, not holiday
Weather	Sunny, intense, other

며 이를 Fig. 7에 산점도로 나타내었다.

각 feature에 대한 벡터는 벡터 간 거리 차이가 적을수록 유사도가 높는데 Fig. 7에서 나타난 유사도는 수소 판매량과 각 변수의 관계를 나타낸 Fig. 4와 Fig. 6과 유사함을 확인할 수 있다. 예를 들어 Fig. 4에서 판매량이 금요일과 주말에 높고 다른 평일은 낮으므로 금요일과 주말, 다른 평일까지 판매량 관점에서 유사도가 높다고 할 수 있는데 Fig. 7(a)에서 금요일과 주말, 다른 평일까지 유사도가 높은 것이 반영되었음이 확인된다. 이는 날씨에서도 동일하게 확인된다.

4.3 LSTM 모델을 이용한 예측

4.3.1 모델 설계

본 연구에서 제안한 LSTM 예측 모델은 3.2절에서 소개한 보다 복잡한 시퀀스 패턴을 학습할 수 있는 multi-layer LSTM 모델로 구성하였으며 설정한 하이퍼 파라미터는 Table 5에 나타내었다. 예측 모델의 feature는 one-hot encoding이 적용된 공휴일, embedding layer를 통해 학습된 3차원의 요일 벡터, 1차원의 날씨 벡터가 사용되었다.

상기 설계한 모델을 토대로 training set과 test set

의 epochs에 따른 손실함수의 결과를 Fig. 8에 나타내었으며 두 데이터 세트에 대하여 손실함수의 값이 작은 값으로 수렴함을 확인하였다.

4.3.2 예측 결과 및 모델 간 성능 비교

제안된 예측 모델의 성능을 검증하기 위해 전통적 시계열 모델로 외부 요인을 고려할 수 없는 선형 모델인 SARIMA, 외부 요인을 고려할 수 있는 선형 모델인 SARIMAX를 비교 모델로 선정하였다. 두 모델 모두 pmdarima 패키지의 auto_arima를 사용하여 최적의 파라미터를 선정하였으며 SARIMAX 모델의 외부 요인은 LSTM에서 사용한 feature를 동일하게 사용하였다. 그리고 본 연구에서 제안한 딥러닝 기반의 비선형 모델인 LSTM 모델과 예측 성능을 각각

Table 5. Hyperparameters of LSTM

Variables	Description
Number of hidden layers	2
Number of cells	50
Optimizer	Adam
Loss function	Mean squared error (MSE)
Epochs	50
Batch size	20

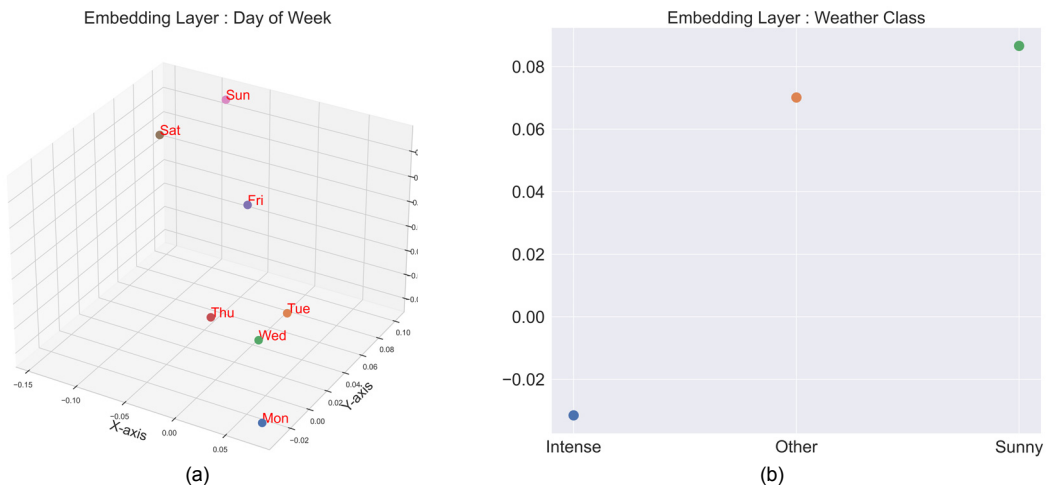


Fig. 7. Scatter plot of embedding layer. (a) Day of week. (b) Weather class

비교하였다.

3가지 알고리즘 모두 training set으로 학습한 모델로 test set에 해당되는 기간에 대해 예측하였으며 실제값과 예측값의 비교 그래프를 Fig. 9에 나타내었다. 각 알고리즘의 예측 성능을 수치적으로 비교하기 위해서 실제값과 예측값을 비교하여 오차율을 구하는 평가 기준인 root mean squared error (RMSE)와 mean absolute percentage error (MAPE)를 각각 계산하였으며 그 결과를 Table 6에 나타내었다.

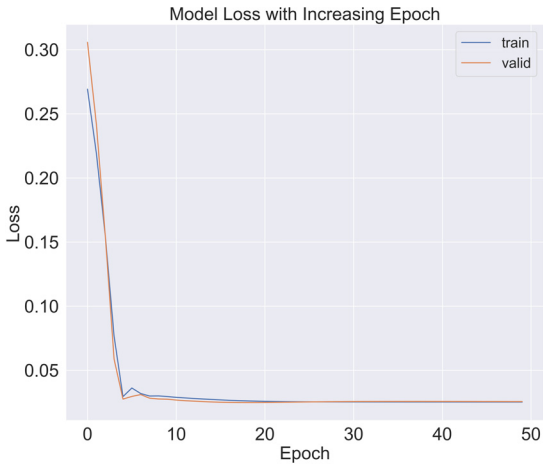


Fig. 8. Model loss with increasing epoch

$$MAPE = \frac{100}{n} \sum_{i=1}^n \left| \frac{y_i^{actual} - y_i^{predicted}}{y_i^{actual}} \right|$$

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

Table 6의 예측 성능 분석 결과에서 LSTM 모델이 SARIMA 및 SARIMAX 모델보다 예측 성능이 우수함이 확인된다. 이는 일반적으로 단기 예측에 해당하는 일일 수소 판매량이 비선형적 특성을 가지며 features와도 비선형적인 관계를 가지기 때문이다. 이러한 특성으로 인하여 비선형 모델인 LSTM이 선형 모델인 SARIMA 및 SARIMAX 모델보다 우수한 예측 성능을 보여준 것으로 판단된다. 또한 SARIMAX 모델이 SARIMA 모델보다 예측 성능이 우수한 것으로 확인되는데 이는 외부 요인을 포함하여 예측하는 것

Table 6. RMSE and MAPE performance of each model

Model	RMSE	MAPE
LSTM	19.83 kg	16.37%
SARIMA (2,0,2)(0,1,7)	21.10 kg	19.03%
SARIMAX (1,0,1)(0,1,2,7)	20.66 kg	17.98%

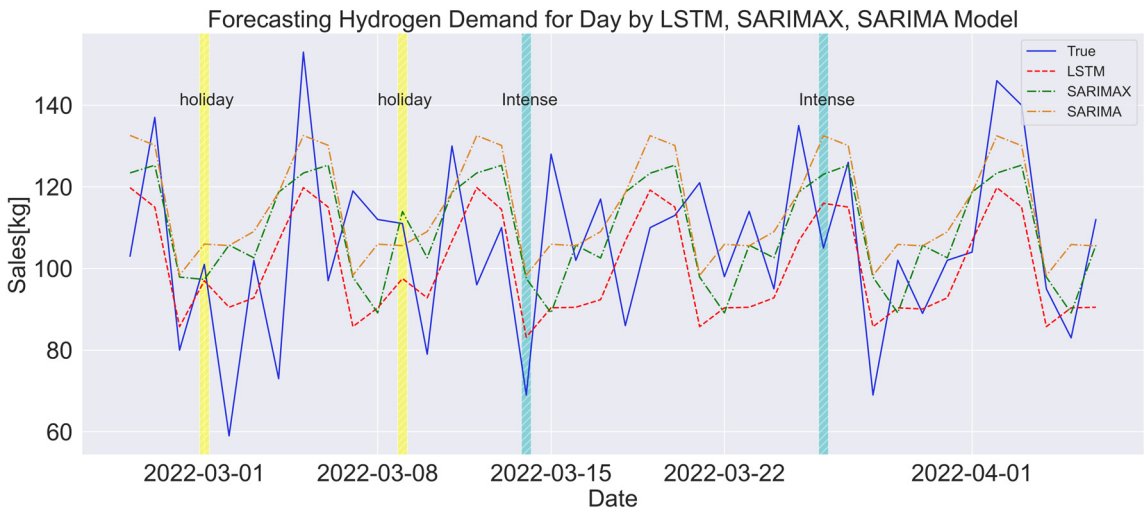


Fig. 9. Daily hydrogen sales prediction by forecasting model, LSTM, SARIMX, and SARIMA

이 포함하지 않는 것보다 판매량에 영향을 끼치는 외부 요인을 추가적으로 고려할 수 있어 관계를 더 잘 학습할 수 있기 때문인 것으로 판단된다.

이러한 모델들의 학습 여부는 Fig. 9에서 확인할 수 있다. 세 가지 모델 모두 학습된 패턴이 7일 주기로 나타난다. SARIMA 모델은 예측 기간 동안 변동 없이 반복적인 패턴을 보였으며 SARIMAX와 LSTM 모델은 선정한 feature에 의해 패턴에 일부 변동이 있음이 확인된다. 평일보다 판매량이 상대적으로 높은 것으로 확인된 공휴일인 3월 1일(삼일절)과 3월 9일(대통령 선거일)에 SARIMAX 모델과 LSTM 모델 모두 기존 패턴 대비 높은 판매량이 나타났다. 그러나 맑은 날씨인 sunny 클래스보다 판매량이 상대적으로 낮은 것으로 확인된 상대적 극한 날씨인 intense 클래스에 해당되는 3월 14일과 3월 26일에 LSTM 모델은 기존 패턴 대비 판매량이 낮아지는 것이 확인되었으나 SARIMAX는 이러한 변동을 반영하지 못하였다. 이는 SARIMAX 모델이 날씨 클래스와 판매량의 관계를 학습하지 못했음을 의미하며 실제로 SARIMAX 모델에서 날씨 클래스 feature에 대한 *p*-value가 0.947로 통계적으로 유의미하지 않다는 결과도 이를 뒷받침한다.

4.3.3 선행 연구와의 예측 성능 비교

본 연구에서 개발한 예측 모델은 Hwang 등⁹⁾이 제안한 날씨를 고려한 머신러닝 및 딥러닝 기반 예측 모델보다 더 높은 예측 정확도를 보였다. 해당 연구에서 MAPE 값이 명확히 제시되지 않았으나 RMSE는 50.91-77.26 kg 범위로 나타났다. 수소 충전소의 일일 판매량이 비슷한 수준임을 고려했을 때 요일의 특성과 공휴일 여부를 추가로 고려한 본 연구의 모델은 RMSE가 19.83 kg으로 선행 연구의 예측 모델보다 예측 성능을 효과적으로 개선한 것을 확인할 수 있다.

5. 결론

본 연구에서는 국내의 주거 지역에 위치한 off-site

수소 충전소의 1년간의 수소 판매량 데이터를 분석하여 판매량에 영향을 끼치는 feature를 선정하고 이를 통해 LSTM 모델을 설계하여 예측하였다. 또한 기존의 전통적 시계열 모델인 SARIMA와 SARIMAX 모델과 예측 성능을 비교하였다.

예측 모델 설계 시 feature 선정을 위하여 수소 충전소의 판매량에 영향을 미치는 다양한 요인을 분석하였다. 요일별로는 금요일과 주말에 판매량이 상대적으로 높고 월요일에 가장 낮은 판매량을 기록하였다. 또한 공휴일에는 비공휴일보다 약 10 kg 높은 판매량을 확인하였으며 날씨와 관련해서는 맑은 날씨에 판매량이 높고 상대적으로 극한 날씨에는 판매량이 저조한 경향을 확인하였다.

예측 결과 LSTM 모델이 선형 모델이면서 외부 요인을 포함한 SARIMAX 모델과 외부 요인을 포함하지 않는 SARIMA 모델보다 예측 성능이 우수함을 확인하였다. LSTM 모델의 우수한 성능은 비선형 모델로서 feature와 판매량 간의 복잡한 비선형 관계를 더 잘 포착할 수 있기 때문으로 분석되었다. 특히 공휴일이나 날씨와 같은 요인을 고려한 예측이 수소 충전소의 일일 판매량을 보다 정확하게 예측하는 데 필수적임을 확인하였다. 이를 통해 수소 충전소의 일일 판매량 예측에는 영향 요인을 고려한 비선형 모델이 적합함을 확인하였다.

또한 본 연구에서 제안한 예측 모델은 선행 연구에서 제안한 예측 모델 대비 개선된 성능을 보였다. 선행 연구에서는 날씨 변수만을 고려한 반면 본 연구에서는 공휴일 여부와 요일의 특성을 feature에 새롭게 추가하여 예측 정확도를 개선하였다.

그러나 본 연구에는 몇 가지 한계점이 존재한다. 예측 모델은 요일, 공휴일, 날씨 등의 요인을 고려하였지만 데이터 수집 기간 동안 판매 단가에 변화가 없는 관계로 판매 단가와 주변 수소 충전소의 존재 유무, 대기 차량, 충전을 하지 못하고 되돌아가는 차량 등의 변수들을 포함하지 못하였다. 이러한 변수들은 수소 충전소의 실제 판매량에 중요한 영향을 미칠 수 있으며 이들을 고려하지 않음으로 인하여 예측 모델의 성능이 제한될 수 있다. 또한 데이터 수집

기간이 1년으로 짧아 예측 모델이 데이터의 패턴을 학습하기에도 충분하지 않았다.

이러한 한계에도 불구하고 본 연구는 수소 충전소의 일일 판매량 예측을 위해 딥러닝 기법을 적용한 연구로 수소 충전소 운영자, 상위 기관, 수소 공급처에게 중요한 인사이트 제공 및 운영의 효율성 증대에 기여할 수 있는 중요한 기반을 마련하였다는 데 가치가 있다.

후 기

본 연구는 국토교통부 해외 수소기반 대중교통 인프라 기술개발 사업(과제 번호: RS-2021-KA163280)의 연구비 지원에 의해 수행되었습니다.

References

1. Hydrogen Council, "Hydrogen meets digital", Hydrogen Council, 2018. Retrieved from <https://hydrogencouncil.com/wp-content/uploads/2018/10/Hydrogen-Council-Hydrogen-Meets-Digital-2018.pdf>.
2. K. Wipke, S. Sprik, J. Kurtz, T. Ramsden, C. Ainscough, and G. Saur, "National fuel cell electric vehicle learning demonstration final report", National Renewable Energy Laboratory, 2012. Retrieved from https://www.energy.gov/sites/default/files/2014/03/f11/learning_demo_final_report.pdf.
3. M. Robinius, J. Linßena, T. Grubea, M. Reußa, P. Stenzela, K. Syranidis, P. Kuckertza, and D. Stolten, "Comparative analysis of infrastructures: hydrogen fueling and electric charging of vehicles", Jülich, 2018. Retrieved from https://juser.fz-juelich.de/record/842477/files/Energie_Umwelt_408_NEU.pdf.
4. M. Genovese and P. Fragiaco, "Hydrogen refueling station: overview of the technological status and research enhancement", *Journal of Energy Storage*, Vol. 61, 2023, pp. 106758, doi: <https://doi.org/10.1016/j.est.2023.106758>.
5. P. Martin, "Three quarters of hydrogen refuelling stations in South Korea closed amid H₂ supply crash", *Hydrogeninsight*, 2023. Retrieved from <https://www.hydrogeninsight.com/transport/three-quarters-of-hydrogen-refuelling-stations-in-south-korea-closed-amid-h2-supply-crash/2-1-1559610>.
6. O. J. Kwon, H. J. Jo, H. H. Chung, and K. J. Myong, "Analysis and modeling of hydrogen sales at hydrogen filling stations", *Transactions of the Korean Society of Automotive Engineers*, Vol. 27, No. 2, 2019, pp. 93-100, doi: <https://doi.org/10.7467/KSAE.2019.27.2.093>.
7. J. Kurtz, T. Bradley, E. Winkler, and C. Gearhart, "Predicting demand for hydrogen station fueling", *International Journal of Hydrogen Energy*, Vol. 45, No. 56, 2020, pp. 32298-32310, doi: <https://doi.org/10.1016/j.ijhydene.2019.10.014>.
8. J. Sun, Y. Peng, D. Lu, X. Chen, W. Xu, L. Weng, and J. Wu, "Optimized configuration and operating plan for hydrogen refueling station with on-site electrolytic production", *Energies*, Vol. 15, No. 7, 2022, pp. 2348, doi: <https://doi.org/10.3390/en15072348>.
9. M. W. Hwang, Y. Ha, and S. Park, "Machine learning-based hydrogen charging station energy demand prediction model", *Journal of Internet Computing and Services*, Vol. 24, No. 2, 2023, pp. 47-56, doi: <https://doi.org/10.7472/jksii.2023.24.2.47>.
10. S. Y. Pan, Q. Liao, and Y. T. Liang, "Multivariable sales prediction for filling stations via GA improved BiLSTM", *Petroleum Science*, Vol. 19, No. 5, 2022, pp. 2483-2496, doi: <https://doi.org/10.1016/j.petsci.2022.05.005>.
11. L. Sun, X. Xing, Y. Zhou, and X. Hu, "Demand forecasting for petrol products in gas stations using clustering and decision tree", *Journal of Advanced Computational Intelligence and Intelligent Informatics*, Vol. 22, No. 3, 2018, pp. 387-393, doi: <https://doi.org/10.20965/jaciii.2018.p0387>.
12. J. Henzel, J. Bularz, and M. Sikora, "Impact of time series clustering on fuel sales prediction results", *Annals of Computer Science and Information Systems*, Vol. 26, 2021, pp. 13-21, doi: <http://dx.doi.org/10.15439/2021F129>.
13. R. Philipsen, T. Brell, W. Brost, T. Eickels, and M. Ziefle, "Running on empty - users' charging behavior of electric vehicles versus traditional refueling", *Transportation Research Part F: Traffic Psychology and Behaviour*, Vol. 59, Pt. A, 2018, pp. 475-492, doi: <https://doi.org/10.1016/j.trf.2018.09.024>.
14. M. Kim, S. Jeon, and T. Jyung, "Analysis of hydrogen sales data at hydrogen charging stations", *Journal of Hydrogen and New Energy*, Vol. 34, No. 3, 2023, pp. 246-255, doi: <https://doi.org/10.7316/JHNE.2023.34.3.246>.
15. B. M. Kang, Y. T. Kang, S. H. Lee, N. S. Kim, K. E. Yi, M. J. Park, C. H. Jeong, and D. W. Jeong, "Analysis of hydrogen sales volume in Changwon", *Journal of Hydrogen and New Energy*, Vol. 30, No. 4, 2019, pp. 356-361, doi: <https://doi.org/10.7316/KHNES.2019.30.4.356>.
16. B. Kim, S. Jung, M. Kim, J. Kim, H. Lee, and S. Kim, "Solar power generation forecasting based on LSTM considering weather conditions", *Journal of Korean Institute of Intelligent Systems*, Vol. 30, No. 1, 2020, pp. 7-12, doi: <https://doi.org/10.5391/JKIIS.2020.30.1.7>.
17. Korea Meteorological Administration (KMA), "Observation data of automated synoptic observing system", KMA, 2022. Retrieved from <https://data.kma.go.kr/data/grnd/selectAsos>

[RltmList.do?pgmNo=36](#).

18. A. Wagner, E. Ramentol, F. Schirra, and H. Michaeli, “Short- and long-term forecasting of electricity prices using em-

bedding of calendar information in neural networks”, *Journal of Commodity Markets*, Vol. 28, 2022, pp. 100246, doi: <https://doi.org/10.1016/j.jcomm.2022.100246>.