

스마트팜 에너지소비량 예측 모델 개발에 관한 연구

임청룡, 박영준, 이철성
한국농어촌공사 농어촌연구원
e-mail: chrylim@ekr.or.kr

A Study on the Development of Smart Farm Energy Consumption Prediction Model

Cheong-Ryong Lim, Young-Jun Park, Chul-Sung Lee
Rural Research Institute, Korea Rural Community Corporation

요약

본 논문에서는 이 연구에서는 스마트팜 경영비용 중 많은 부분을 차지하는 에너지소비량에 대한 예측모델을 구축하였으며, 분석결과를 다음과 같이 요약할 수 있다. 첫째, 예측력에 우수한 모델 구축을 위해 전통적인 시계열분석모형인 VAR 모델과, 딥러닝모델인 RNN, LSTM, GRU 모델을 추정하였다. 둘째, 4가지 모델 예측력에 대한 비교결과 예측력이 GRU>LSTM>RNN>VAR 순으로 나타났다.

1. 서론

농산물시장 개방에 따라 식량자급률은 지속적으로 하락하고 있으며, 생산비 증가로 인한 농업소득의 정체와 농가인구의 지속적인 감소와 고령 인구비율의 증가 및 경지면적의 감소 등 여러 요인들이 농업경쟁력 약화를 유발하고 있다.

'03년 대비 '20년 농업소득은 12% 증가한 반면 농업경영비는 86%나 증가하였다. 동기간 농업소득은 매년평균 1.4% 증가한 반면 농업경영비는 연평균 3.8% 증가하여 농업소득의 증가율이 크게 감소하였다. 향후 농업경영비의 지속적인 증가에 따라 농업소득은 지속적으로 감소할 것으로 여겨진다.

지난 15년간 일반 소비자 물가상승율은 평균 2.1% 씩 증가하였으나 농업소득 증가율이 물가상승율에 비해 매년 평균 1% 이상 낮게 나타나 농업생산력 저하 및 신규 농업인력유입을 가로막는 원인이 되고 있다

농촌인구 또한 연간 2.5% 이상씩 지속적으로 감소하고 있으며, 특히 20대~50대 가용노동인구가 매년 크게 감소하고 있음. 반면 60대 이상 고령농의 비율은 지속적으로 증가하고 있으며, '19년에는 20대 이상 농민의 60%를 넘어섰다.

2. 스마트팜 확산

2.1 스마트팜 개요

농촌 위기극복 방안으로 4차 산업혁명 핵심기술인 스마트팜이 빠르게 확산되고 있다. 생산성이 높은 스마트팜의 확산을 통해 농촌의 경쟁력 향상 및 지속가능한 발전에 기여하고, 청년유입 및 농업경쟁력 확보를 위해 스마트팜혁신밸리 조성 과 같은 다양한 스마트팜 도입정책을 추진하고 있다. 따라서 스마트팜 시장규모 연평균 10% 성장하고 있으며 관련 정부 지원 예산도 36% 증가하였다.

스마트팜 재배작물은 주로 시설채소이며 재배면적은 46,089ha이다. 이 중 무가온 온실이 66%, 가온온실이 34%로 조사되었다. 가온온실의 사용연료는 대부분 유류로 나타났으며, 등유 및 경유를 가장 많이 소비하고 있다. 난방비는 전체 경영비의 30%~40%(최고 60%)로 비중이 매우 높다.

에너지 단위비용은 국제 유가의 영향을 받아 농가 경영비 불확실성을 키우며, 에너지 비용의 저감은 농가소득 개선 및 경영안정화에 기여할 수 있으므로 에너지 저감이 필요하다.

2.2 스마트팜 작동방식

스마트팜은 복합환경제어 시스템을 기반으로 온도나 습도, 영양공급등 작물의 생장관리 조건을 정밀·자동 제어하여 최적화할 수 있는 최첨단 온실이다.

스마트팜은 복합환경제어시스템이 실시간으로 데이터를 입력 받아 설정온도를 맞추기 위해 내부 장치를 제어하게 된다. 입력받는 데이터는 외부환경 및 내부환경 데이터이며, 이러한 데이터들은 다양한 센서를 통해 측정되어 복합환경제어 시스템으로 전송된다. 복합환경제어시스템 사전에 설정된 온도 및 습도를 유지시키기 위해 다양한 장치를 제어하게 된다.

3. 에너지소비량 예측방법 및 분석자료

3.1 예측방법

시계열 데이터를 활용한 예측에 있어서 전통적인 통계분석 방법으로는 단일변수 분석모형인 ARIMA(Auto-regressive Integrated Moving Average) 모형이나, 다변량 시계열 분석 모형인 벡터자기회귀분석(Vector Auto Regression, VAR) 모형이 많이 사용되고 있다.

VAR 모델은 예측할 변수의 과거 값뿐만 아니라 예측할 변수와 의존성이 있는 변수 들까지 고려하여 선형 함수로 나타내는 확률적 과정이라고 할 수 있다. 모델분석을 위해 정상성 여부와 최적시차결정 및 잔차항에 대한 자기상관성 확인 등이 필요하다. 독립변수가 종속변수에 미치는 영향을 확인할 수 있는 장점이 있지만, 딥러닝 모델에 비해 정확도 낮으며 단기간 예측에 비교적 적합한 특징을 가지고 있다.

최근 들어 인공지능의 빠른 발전과 컴퓨터성능의 향상으로 딥러닝 기법인 순환신경망(Recurrent Neural Network, RNN), 장단기메모리(Long Short-Term Memory, LSTM), 게이트순환유닛(Gated Recurrent Unit, GRU) 등 모델들이 활발하게 활용되고 있다.

RNN(Recurrent Neural Network)은 입력과 출력을 시퀀스 단위로 처리하는 시퀀스(Sequence) 모델이다. 인공 신경망의 한 종류로, 유닛간의 연결이 순환적 구조를 갖는 특징을 갖고 있다. RNN은 가장 기본적인 인공 신경망 시퀀스 모델이다[1].

LSTM(long short-term memory, 장단기 메모리)는 기울기 소실 문제를 해결하기 위해 고안된 딥 러닝 시스템이다. LSTM은 망각 게이트(forget gate)라 부르는 게이트를 추가적으로 가진다.LSTM은 은닉층의 메모리 셀에 입력게이트, 망각게이트, 출력게이트를 추가하여 불필요한 기억을 지우고, 기억해야 할 것들을 결정한다. LSTM은 은닉 상태(hidden state)를 계산하는 방식이 전통적인 RNN보다 조금 더 복잡해

졌으며, 셀 상태(cell state)라는 값이 추가되었다[2].

GRU는 LSTM에서는 출력, 입력, 삭제 게이트라는 3개의 게이트가 존재. 반면, GRU에서는 업데이트 게이트와 리셋 게이트 두 가지 게이트만 존재한다. GRU는 LSTM보다 학습 속도가 빠르다고 알려져 있으며, 여러 평가에서 GRU는 LSTM과 비슷한 성능을 보인다[3].

3.2 분석자료 및 변수선택

스마트팜 에너지소비량 예측을 위해 2021.12.01.~2022.02.28.사이 분단위 측정데이터를 활용하였다. 수집자료는 외부환경변수들인 일사량(W/m²), 외기온(°C), 외기습도(%), 풍속(m/s), 풍향(°)이며 내부환경변수들인 온도(°C), 습도(%), 수평 차광스크린(%), 수평 차열스크린(%), 수평 에너지스크린(%), 수직 에너지스크린(%), 배기팬(on/off), 천창개폐(%), 유동팬(on/off), 팬코일(on/off)이다. 설정온도에는 환기온도(°C), 난방온도(°C) 값을 수집하였으며, 에너지소비량에 대한 측정변수로 온수 열공급량(kWh)을 수집하였다.

분단위 데이터 과도한 변동성 줄이고, 예측목적에 부합하는 결과 도출을 위해 1시간 단위로 데이터 처리 이후 상관성 확인을 통해 분석변수를 선택하였다.

이 연구에서는 서로 높은 상관관계가 보이는 외부온도(Otemp), 외부광량(Sun), 내부온도(Itemp), 열공급량(Energy) 변수들을 이용하여 시간대별 추이 확인하면 다음과 같이 나타 낼 수 있다.



[그림 3] 에너지예측모델 활용될 변수 시간대별 추이

4. 분석결과 및 예측력비교

4.1 분석결과

4.1.1 VAR모델

분석에 사용된 변수들의 정상성 관련하여 ADF 단위근 검

정을 하였으며, 검정결과 단위근이 존재하지 않았으며, 최적 시차 선정을 위해 1-9차 까지 모형분석 이후 BIC값이 최소인 시차 4를 선택하였다.

VAR(4)모형 추정결과는 다음 표와 같이 나타나며, 잔차항에 대한 자가상관성 확인을 위한 DW통계량이 2.0으로 나타나 잔차항의 자기상관성이 존재하는 것을 알 수 있었다.

[표 2] VAR 모델 추정결과

| Variables | coefficient | std. error | t-stat | prob |
|-----------|-------------|------------|--------|-------|
| const | 19.363 | 3.292 | 5.882 | 0 |
| L1.Otemp | -0.427 | 0.265 | -1.612 | 0.107 |
| L1.Sun | -0.031 | 0.004 | -7.656 | 0 |
| L1.Itemp | -1.394 | 0.245 | -5.682 | 0 |
| L1.Energy | 0.645 | 0.023 | 28.115 | 0 |
| L2.Otemp | 0.330 | 0.480 | 0.688 | 0.492 |
| L2.Sun | 0.041 | 0.007 | 6.08 | 0 |
| L2.Itemp | 1.090 | 0.368 | 2.964 | 0.003 |
| L2.Energy | -0.042 | 0.027 | -1.537 | 0.124 |
| L3.Otemp | 0.122 | 0.480 | 0.253 | 0.8 |
| L3.Sun | 0.003 | 0.007 | 0.37 | 0.711 |
| L3.Itemp | -0.568 | 0.355 | -1.599 | 0.11 |
| L3.Energy | 0.073 | 0.027 | 2.662 | 0.008 |
| L4.Otemp | 0.039 | 0.260 | 0.149 | 0.882 |
| L4.Sun | -0.012 | 0.005 | -2.597 | 0.009 |
| L4.Itemp | 0.209 | 0.215 | 0.971 | 0.332 |
| L4.Energy | 0.016 | 0.023 | 0.686 | 0.493 |

D.W. : 2.0
 ADF test statistic Otemp: -4.633 p-value: 0.000
 ADF test statistic Sun: -5.436, p-value: 0.000
 ADF test statistic Itemp: -5.452, p-value: 0.000
 ADF test statistic Energy: -6.953, p-value: 0.000

4.1.2 RNN모델

RNN모델 추정을 위해 과거 50개의 유닛을 활용하여 예측을 하며, 은닉층은 1개로 하였으며, 에폭(Epoch) 횟수는 50회, Batch size는 72개로 설정하여 총 2,801개의 파라미터를 추정하였다.

[표 1] RNN 모델 적합

| Layer (type) | Output Shape | Params |
|------------------------|--------------|--------|
| simple_rnn (SimpleRNN) | (None, 50) | 2,750 |
| dense (Dense) | (None, 1) | 51 |

Total params: 2,801

에폭 회수 증가에 따라 MSE가 지속적으로 감소하지만 40회 이후에는 감소폭이 크지 않으므로 에폭수가 적절한 것으로 나타났다.



[그림 5] 에폭 수 증가에 따른 RNN모델 개선추이

4.1.3 LSTM모델

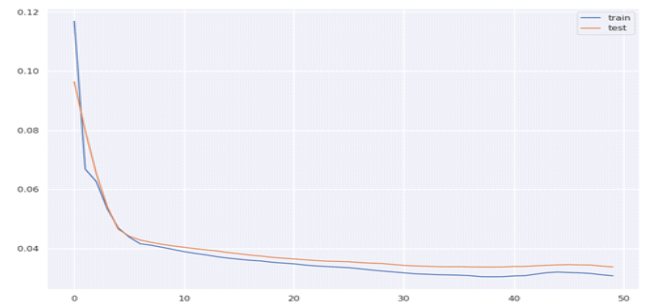
LSTM모델 추정을 위해 과거 50개의 유닛을 활용하여 예측을 하며, 은닉층은 1개로 하였으며, 에폭(Epoch) 횟수는 50회, Batch size는 72개로 설정하여 총 11,000개의 파라미터를 추정하였다.

[표 2] LSTM 모델 적합

| Layer (type) | Output Shape | Params |
|---------------|--------------|--------|
| lstm (LSTM) | (None, 50) | 11,000 |
| dense (Dense) | (None, 1) | 51 |

Total params: 11,051

에폭 회수 증가에 따라 MSE가 지속적으로 감소하지만 40회 이후에는 감소폭이 크지 않으므로 에폭수가 적절한 것으로 나타났다.



[그림 6] 에폭 수 증가에 따른 LSTM모델 개선추이

4.1.4 GRU모델

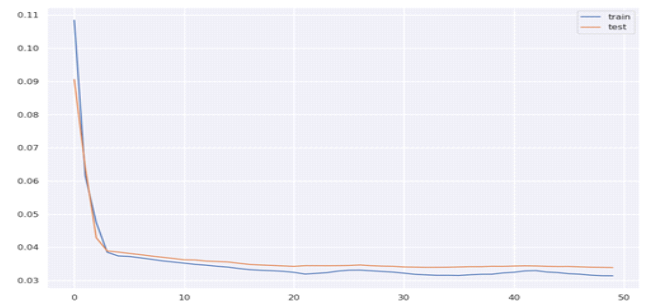
GRU모델 추정을 위해 과거 50개의 유닛을 활용하여 예측을 하며, 은닉층은 1개로 하였으며, 에폭(Epoch) 횟수는 50회, Batch size는 72개로 설정하여 총 8,250개의 파라미터를 추정하였다.

[표 1] GRU 모델 적합

| Layer (type) | Output Shape | Params |
|---------------|--------------|--------|
| gru (GRU) | (None, 50) | 8,250 |
| dense (Dense) | (None, 1) | 51 |

Total params: 8,301

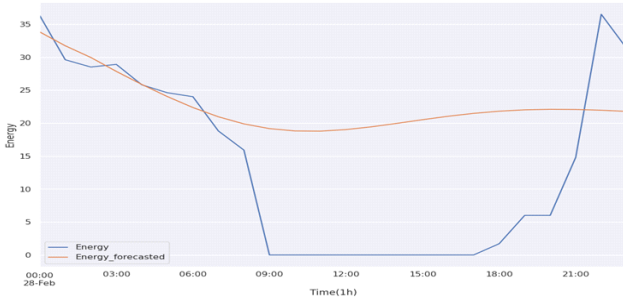
에폭 회수 증가에 따라 MSE가 지속적으로 감소하지만 40회 이후에는 감소폭이 크지 않으므로 에폭수가 적절한 것으로 나타났다.



[그림 7] 에폭 수 증가에 따른 GRU모델 개선추이

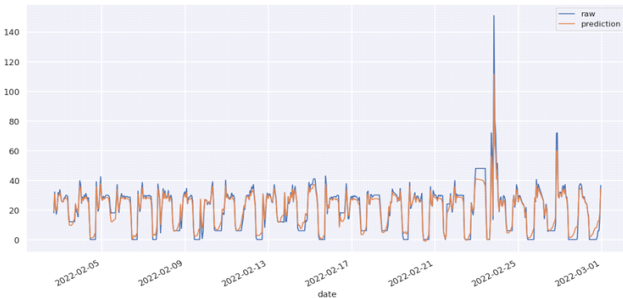
4.2 예측모델 간 예측력비교

VAR모델을 이용하여 미래 24시간을 예측할 경우 MSE가 201.94, RMSE가 14.2107로 나타나 단기예측은 비교적 우수하지만 전반적인 예측력은 좋지 않은 것으로 나타났다.



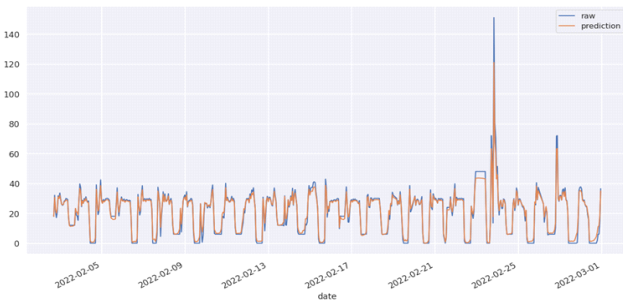
[그림 8] VAR모델을 활용한 에너지 예측결과

학습된 RNN모델을 Test set에 적용결과 MSE가 13.001, RMSE가 3.606로 Test set에 대한 예측력 우수한 것으로 나타났다.



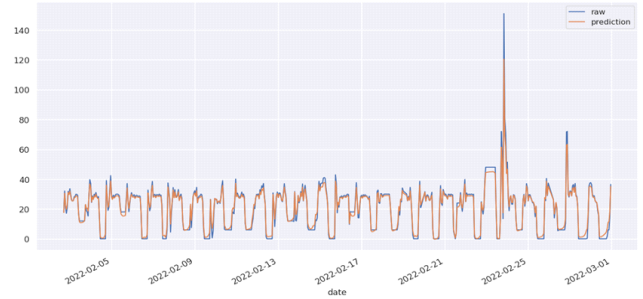
[그림 9] RNN모델을 활용한 에너지 예측결과

학습된 LSTM모델을 Test set에 적용결과 MSE가 5.913, RMSE가 2.432로 Test set에 대한 예측력이 RNN모델보다 우수한 것으로 나타났다.



[그림 10] LSTM모델을 활용한 에너지 예측결과

학습된 GRU모델을 Test set에 적용결과 MSE가 5.659, RMSE가 2.379로 Test set에 대한 예측력 LSTM보다 우수한 것으로 나타났다.



[그림 11] GRU모델을 활용한 에너지 예측결과

5. 결론 및 시사점

이 연구에서는 스마트팜 경영비용 중 많은 부분을 차지하는 에너지소비량에 대한 예측모델을 구축하였으며, 분석결과를 다음과 같이 요약할 수 있다. 첫째, 예측력에 우수한 모델 구축을 위해 전통적인 시계열분석모형인 VAR모델과, 딥러닝모델인 RNN, LSTM, GRU모델을 추정하였다. 둘째, 4가지 모델 예측력에 대한 비교결과 예측력이 GRU>LSTM>RNN>VAR 순으로 나타났다.

이러한 분석결과로부터 다음과 같은 시사점을 도출할 수 있다. 첫째, 에너지소비에 영향을 미치는 요인이 아니 예측 정확도를 목적으로 할 경우 딥러닝 모델 활용이 적합하다. 둘째, 딥러닝 모델별 예측력에도 차이 있으므로, 다양한 모델 적합이후 최적모델 선택하는 것이 효과적임을 알 수 있다. 보다 우수한 딥러닝 모델 구축을 위해서는 적절한 하이퍼파라미터에 대한 설정이 필요하다.

참고문헌

- [1] Dupond, Samuel. "A thorough review on the current advance of neural network structures." 《Annual Reviews in Control》 14: 200 - 230, 2019.
- [2] Gers, Felix A.; Schraudolph, Nicol N.; Schmidhuber, Jürgen. "Learning Precise Timing with LSTM Recurrent Networks" (PDF). 《Journal of Machine Learning Research》 3: 115 - 143, 2002.
- [3] Heck, Joel; Salem, Fathi M. "Simplified Minimal Gated Unit Variations for Recurrent Neural Networks", 2017.