

# 온실형 스마트팜에서 환경 데이터를 분석을 활용한 습도 센서 예측 모델

안형태, 안수용, 김재영, 이철원  
한국전자통신연구원

[anten@etri.re.kr](mailto:anten@etri.re.kr), [syong.an@etri.re.kr](mailto:syong.an@etri.re.kr), [jaeyoung@etri.re.kr](mailto:jaeyoung@etri.re.kr), [e10000won@etri.re.kr](mailto:e10000won@etri.re.kr)

## Humidity Sensor Prediction Model using Environmental Data Analysis in Greenhouse Smartfarm

Hyeongtae Ahn, Su-Yong An, Jae-Young Kim, Cheol Won Lee  
Electronics and Telecommunications Research Institute(ETRI)

### 요 약

스마트팜은 정보통신기술을 바탕으로 시공간의 제약 없이 영농 의사결정을 지원하는 첨단화된 농장이다. 온실형 스마트팜은 온실 내·외부의 다양한 센서를 활용하여 온실 환경을 측정하고, 이를 바탕으로 필요한 온실제어를 실행시킨다. 온실 내부습도는 온실제어를 결정하는 중요한 지표 중 하나이며, 온도, 광도, 외부기상 등 온실 환경과 환기창, 커튼, 유동팬, 보일러 등 구동기의 작동에 많은 영향을 받는다. 본 논문은 특정 온실 내부습도 센서의 측정값을 나머지 센서들의 측정값과 구동기들의 작동 상태 정보로 예측할 수 있는지 확인한다. 다중선형회귀분석(Multiple Linear Regression Analysis)을 활용하여, 해당 내부습도 센서의 측정값을 종속변수로 설정하고, 나머지 센서들의 측정값과 구동기들의 작동 상태 정보를 설명변수로 설정하여 온실에서 수집된 환경 데이터를 학습한다. 다중선형회귀분석으로 추정된 내부습도 센서 예측 모델은 내부습도와 나머지 환경 요소 및 구동기 작동 간의 복합적인 관계해석에 활용되며, 디지털 트윈의 가상화된 스마트팜에서 내부습도의 중요 모델로 사용될 것으로 기대한다.

### I. 서 론

스마트팜은 기존의 정보통신기술과 4 차 산업혁명의 인공지능, 사물인터넷, 빅데이터, 로봇 등 첨단기술이 융합된 농장이다. 특히, 온실형 스마트팜은 시공간의 제약 없이 온실 환경을 실시간으로 모니터링하면서 적절한 작물의 성장 환경을 자동으로 유지·관리할 수 있다. 따라서 스마트팜은 농업 인구의 감소 및 고령화, 작물 생산성 증대, 병충해 예방, 기후변화 대응 등 농업에 관련된 각종 현안의 해결책으로 주목받으며, 미래 농업의 핵심 흐름이 되었다. 현재 농림축산식품부는 온실형 스마트팜의 보급을 확대하고 있으며, 2018 년 4,900ha 에서 2022 년까지 7,000ha 보급을 목표로 한다.[1]

디지털 트윈(Digital Twin)은 실존하는 대상을 가상의 세계에서 디지털 쌍둥이로 구현하여 시뮬레이션 할 수 있는 장점이 있다.[2] 가상세계에서 온실형 스마트팜을 구현하면, 외부 환경의 변화와 사용자의 제어 선택에 따라서 온실 환경 및 작물 생장의 변화를 시뮬레이션으로 예측할 수 있다. 또한, 인공지능을 바탕으로 데이터를 학습하면서 최적화된 온실 환경 제어 및 성장 모델을 도출이 가능하다. 하지만 실제 온실형 스마트팜은 수십 개의 센서와 구동기가 존재하여, 복합적인 제어 로직에 의해서 작동한다. 따라서 가상세계에서 온실형 스마트팜의 구현하기 위해서는, 실제 온실 환경을 유사하게 측정되고 작동하는 센서와 구동기의 모델들이 밀반입되어야 한다.

온실의 주요 환경 요소인 온도, 습도, CO<sub>2</sub>, 광도는 센서들에 의해서 주기적으로 측정되며, 서로에게 밀접한 영향을 미친다. 또한, 외부기상의 변화나 각종 구동기의 작동에도 많은 영향을 받는다. 내부습도를 예로 들면, 내부

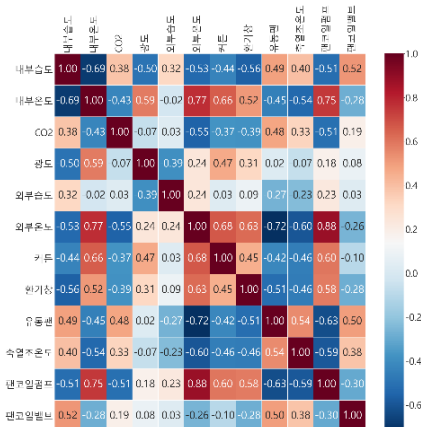


그림 1. 내부습도 센서와 유의미한 피어슨 상관계수를 가지는 센서와 구동기 간의 상관계수를 히트맵(Heat Map)으로 나타냈다. 특히 내부온도, 광도, 외부 온도, 환기창, 팬 코일 구동기는 내부습도와 강한 상관관계를 가진다.

온도가 증가하면 내부습도는 감소하고, 환기창이 개방되면 내부 습기가 외부로 유출되며, 비가 내리면 내부습도는 상승한다. 그림 1 은 온실형 스마트팜에서 수집한 센서들의 측정값과 구동기들의 작동 상태 정보 등 환경 데이터에서 내부습도 센서와 유의미한 피어슨 상관계수(Pearson Correlation Coefficient)를 가지는 센서와 구동기를 보여준다. 다양한 센서의 측정값과 및 구동기의 작동 상태 정보가 내부습도 센서에 상관관계를 있다는 것을 알 수 있다. 따라서 이러한 내부습도 센서의 측정값과 나머지 환경 데이터와 관련성 및 영향력을 파악할 수 있다면, 특정 내부습도 센서의 측정값을 알지 못하더라도, 나머지 센서들의 측정값과 구동기들의 작동 상태 정보를

기반으로 해당 내부습도 센서의 측정값을 예측할 수 있을 것이다.

다중선형회귀분석(Multiple Linear Regression Analysis)은 둘 이상의 설명변수로 종속변수에 대한 선형적 상관관계를 설명 및 예측하는 통계적 기법이다.[3]  $n$  개의 설명변수( $X_1, X_2, \dots, X_n$ )로 종속변수  $Y$ 를 예측하기 위한 다중회귀모형의 수식은 다음과 같다.

$$E[Y] = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \dots + \beta_n X_n \quad (1)$$

여기서  $\beta_0$ 는 절편(Intercept)이며,  $\beta_i$ 는 설명변수  $X_i$ 의 회귀계수로 최소제곱법(Ordinary Least Squares)을 활용해서 추정할 수 있다.

본 논문은 다중선형회귀분석을 활용하여, 특정 내부습도 센서의 측정값을 종속변수로 설정하고, 나머지 센서들의 측정값과 구동기들의 작동 상태 정보를 설명변수로 설정하여 온실형 스마트팜에서 수집된 환경 데이터를 학습한다. 그리고 추정된 다중회귀모형의 예측값과 실제 해당 센서의 측정값을 비교함으로써 예측 정확도를 평가한다. 추정된 특정 습도 센서의 예측 모델은 가상화된 온실형 스마트팜에서 내부습도의 중요 모델로 활용될 것 기대한다.

## II. 본론

본 연구에서 하나의 온실형 스마트팜에서 2019 년 12 월부터 2020 년 11 월까지 1년간 분 단위로 환경 데이터를 수집하였다. 데이터 전처리 과정에서 수집된 환경 데이터를 0.0 에서 1.0 사이의 값으로 정규화를 하였고, 이상치 및 결측치는 제외했다. 해당 온실형 스마트팜은 60 개 이상의 센서와 구동기가 활용되고 있으며, 그중에서 ‘내부습도(1 번)’ 센서의 측정값을 예측하고자 하는 종속변수로 설정하였다. 나머지 센서와 구동기에서 다중공선성(Multicollinearity)을 고려하고, 단계적 변수 선택법(Stepwise Selection Method) 활용하여 최종적으로 37 개의 센서와 구동기의 환경 데이터를 설명변수로 선택되었다.

전처리된 환경 데이터에서 75%는 학습용(Training) 데이터 세트로, 25%는 검증(Validation) 데이터 세트로 임의로 분배하여 다중회귀분석으로 학습하였다. 추정된 다중회귀모형의 조정된 결정계수(Adjust  $R^2$ )는 0.918 로 높은 모형에 대한 설명력을 달성하였다. 표 1 은 추정된 ‘내부습도(1 번)’ 다중회귀모형의 설명변수 목록 및 각 설명변수의 회귀계수를 나타낸다. 모든 회귀계수의 유의확률(P-value)은 0.1% 미만으로 추정된 회귀계수는 통계적으로 유의미함을 확인했다. 또한, 설명변수 간의 다중공선성을 나타내는 주요 지표인 분산팽창요인(Variance Inflation Factor)에서도 모든 회귀계수는 10 미만임을 확인했다.

‘내부습도(1 번)’ 센서의 추정된 다중회귀모형의 예측 정확도를 평가하기 위해서, 2021 년 3 월의 환경 데이터를 추가로 수집한 시험(Test) 데이터 세트를 사용해서 예측 정확도를 평가하였다. 그림 2 는 2021 년 3 월에 수집된 ‘내부습도(1 번)’ 센서의 실제 측정값과 추정된 다중회귀모형의 예측값을 보여준다. 실제 측정값과 매우 비슷하게 예측했음을 확인할 수 있다. 조정된 결정계수는 1.0 에 근접한 0.9371 로 추정된 다중회귀모형의 높은 설명력을 나타내며, 평균 절대 오차(Mean Absolute Error)는 1.78%, 평균 절대 퍼센트 오차(Mean Absolute Percentage Error)는 3.55%, 평균 제곱근 오차(Root Mean Square Error)는 1.67 의 정확도를 달성했다.

표 1. ‘내부습도(1 번)’의 다중회귀모형에서 선택된 설명변수 및 설명변수별 추정된 회귀계수.

설명변수( $X_i$ )	회귀계수( $\beta_i$ )	설명변수( $X_i$ )	회귀계수( $\beta_i$ )
CO2(2 번)	0.0122	상커튼	0.0027
내부습도(2 번)	0.8215	중커튼	-0.0114
내부온도(1 번)	0.0510	하커튼	0.0067
국부난방온도	-0.0504	천창(동)	-0.0206
주난방펌프온도	0.0335	천창(서)	-0.0257
보조난방펌프온도	-0.0805	주난방펌프	0.0219
축열조상온도	0.0864	보조난방펌프	-0.0095
축열조하온도	0.0466	생장점순환펌프	-0.0281
팬코일펌프온도	-0.0874	레이난방펌프	-0.0090
광도	-0.0386	팬코일펌프	0.0802
감우	-0.0146	주난방밸브	0.0073
강우	0.3357	보조난방밸브	0.0384
대기압	-0.0444	생장점난방밸브	-0.0063
외부습도	0.0853	레이난방밸브	0.0180
풍속	-0.0365	팬코일난방밸브	0.0383
풍향	0.0054	유동팬	-0.0065
동측면커튼	0.0057	탄산	0.0129
남측면커튼	-0.0030	팬코일	0.0376
북측면커튼	0.0236	Intercept( $\beta_0$ )	0.0493



그림 2. 2021 년 3 월에 수집된 ‘내부습도(1 번)’ 센서의 측정값과 예측값 비교. 파랑 선은 해당 내부습도 센서의 실제 측정값을 나타내며, 주황 선은 예측값을 나타낸다.

## III. 결론

본 연구에서 실제 온실형 스마트팜의 환경 데이터를 수집하여 다중선형회귀분석으로 ‘내부습도(1 번)’ 센서의 다중회귀모형을 추정했다. 실제 내부습도 센서의 측정값과 예측 모델의 예측값 간의 평균 절대 오차는 약 1.8 % 미만으로, 높은 예측 정확도를 달성했다. 본 연구를 발전시켜 온실형 스마트팜의 각종 센서와 구동기에 대한 예측 모델을 만든다면, 디지털 트윈의 가상화된 스마트팜에서 온실 환경을 시뮬레이션 하는데 많은 이바지를 할 것이다.

## ACKNOWLEDGMENT

본 연구 논문은 한국전자통신연구원 연구운영지원사업의 일환으로 수행되었음. [21ZD1150, 인공지능기반 스마트팜 통합 솔루션 기술 개발]

## 참 고 문 헌

- [1] 장영주, 김태우, "스마트팜 확산·보급 사업 현황과 과제: 농업분야 ICT 융복합사업을 중심으로," *국회입법조사처 NARS 현안분석*, 제 95 호, pp. 1-24, Dec. 2019.
- [2] Quan, Y. and 박상찬, "Review on the Application of Industry 4.0 Digital Twin Technology to the Quality Management," *Journal of the Korean Society for Quality Management*, 45(4), pp. 601-610, Dec. 2017.
- [3] Keith, T.Z., *Multiple regression and beyond*, 2nd Ed., Routledge, New York, New York, 2015.