

온실 내 환경데이터 분석을 통한 파프리카 온실의 식별

김나은¹ · 이경근² · 이덕현¹ · 문병은³ · 박재성³ · 김현태^{4*}

¹경상대학교 대학원 바이오시스템공학과 대학원생, ²경상남도농업기술원 농촌지도사,
³경상대학교 스마트팜연구소 학술연구교수, ⁴경상대학교 생물산업기계공학과(스마트팜연구소) 교수

Identification of Sweet Pepper Greenhouse by Analysis of Environmental Data in Greenhouse

Na-eun Kim¹, Kyoung-geun Lee², Deog-hyun Lee¹, Byeong-eun Moon³,
Jae-sung Park³, and Hyeon-tae Kim^{4*}

¹Graduate Student, Department of Bio-Systems Engineering, Graduate School of Gyeongsang National University, Jinju 52828, Korea

²Instructor, Gyeongsangnam-do Agricultural Research & Extension Services, Jinju 52733, Korea

³Research Professor, Institute of Smart Farm, Gyeongsang National University, Jinju 52828, Korea

⁴Professor, Department of Bio-Industrial Machinery Engineering, Gyeongsang National University(Institute of Smart Farm), Jinju 52828, Korea

Abstract. In this study, analysis was performed to identify three greenhouses located in the same area using principal component analysis (PCA) and linear discrimination analysis (LDA). The environmental data in the greenhouse were from 3 farms in the same area, and the values collected at 1 hour intervals for a total of 4 weeks from April 1 to April 28 were used. Before analyzing the data, it was pre-processed to normalize the data, and the analysis was performed by dividing it into 80% of the training data and 20% of the test data. As a result of PCA and LDA analysis, it was found that PCA classification accuracy was 57.51% and LDA classification was 67.06%, indicating that it can be classified by greenhouse. Based on the farmhouse data classified in advance, the data of the new environment can be classified into specific groups to determine the tendency of the data. Such data is judged to be a way to increase the utilization of data by facilitating identification.

Additional key words : big data, classification, LDA, smart farm, PCA

서 론

우리나라의 농업은 다른 농업 선진국들에 비해 경쟁력이 많이 갖춰지지 않은 실정이다. 또한, 국내 농업은 빠른 고령화와 함께 농촌 노동인구도 줄어들고 있으며, 농가 인구는 2010년 306만명(6.2%)에서 2019년 224만명(4.3%)으로 눈에 띄게 줄고 있다(KOSIS, 2019). 이에 대한 방안으로는 최근 지능화된 시스템을 농업에 접목한 융복합 기술인 스마트팜(smart farm)이 각광받고 있다. 스마트팜은 농장에 정보통신기술(ICT)을 접목하여 사물인터넷(IoT), 빅데이터(big data), 인공지능(AI) 등의 기술을 활용할 수 있고, 농산물의 생산·가공 단계에서 최적의 방법으로 제어할 수 있는 지능형 농장이다. 농업경쟁력 확보를 위한 수단으로 스마트팜의 도입은 생산효

율을 높이는 방안으로 매우 중요하게 부각 되고 있으며(Hwang과 Park, 2020), 현재 우리나라는 스마트팜의 빠른 보급과 확산 정책을 실행하여 스마트팜을 농업현장에서 유용하게 사용하기 위하여 노력하고 있다(Lee 등, 2018). 스마트팜은 농장 내부의 실시간 상태에 따라 해당 작물의 생장에 요구되는 빛(태양광, LED 등), 온도, 수분, CO₂ 농도, 배양액 등과 같은 환경 요소의 자동 제어 및 모니터링이 가능하다는 장점이 있다. 온실은 내부의 환경 제어가 중요한 만큼 온실환경과 수확량과의 관계를 분석한 연구(Joung 등, 2018), 온실 환경요인을 이용한 증산량 추정 연구(Nam 등, 2017), 온실용 온습도 예측시스템 연구(Jeong 등, 2020) 등 온실 내 환경과 관련된 다양한 연구가 진행되고 있다. 이와 같은 연구의 결과들을 바탕으로 스마트폰을 활용하여 실제 농장 내부에서만 아니라, 다른 곳에서라도 농장에 관련된 환경 및 필요한 정보들을 확인하거나, 자동이나 원격 제어를 수행할 수 있는 스마트팜 플랫폼의 설계 및 시스템 구현을 위해 노력 중이다(Kim 등,

*Corresponding author: bioani@gnu.ac.kr

Received September 16, 2020; Revised December 23, 2020;

Accepted December 24, 2020

2020). 한국형 스마트팜 모델을 위해 개발되고 있는 다양한 요소 기술들이 있는데, 그중 온실에서 실시간으로 생성되는 환경 및 구동기 데이터를 클라우드 서버에 수집하고 분석하는 기술은 현재의 온실 운영 상태와 과거의 기록도 확인할 수 있다(Jang 등, 2016).

이처럼 농업의 미래는 다양한 첨단기술의 융합과 복합이 절대적으로 필요한 부분이며(Cho와Kang, 2019), 이를 활용하기 위해서는 다양한 형태의 데이터 수집과 가공이 중요한 부분일 것으로 예상된다. 그래서 현재 농촌진흥청에서는 각 도 농업기술원과 함께 2017년부터 시설원에 스마트팜 온실에서 표준화된 데이터를 수집 중이다(Park 등, 2019). 그 외 다른 사례들로는 온실 내 환경의 특성을 파악하여 작물의 생육을 예측하거나(Joung 등, 2018), 실내 병해충관리 방법(Yang 등, 2018) 등과 같은 온실 내부의 데이터 수집에 관한 다양한 연구들이 진행 중이다. 그리고 센서의 발전으로 다양한 형태의 데이터가 대량으로 수집되고 있으므로 데이터 분석기술의 개발과 발전을 통하여 수집된 데이터 활용을 질적으로 높일 수 있다(Choi, 2019). 이처럼 방대한 규모의 정보와 생성 주기가 짧은 데이터의 가치는 빅데이터라는 이름으로 이미 높아져 있는 실정이고(Kim 등, 2018), 데이터들을 목적에 맞게 분석하여 가공된 데이터가 또 다른 종류의 데이터로 생성된다. 이처럼 빅데이터는 활용 분야가 매우 다양하지만, 생성된 데이터라도 데이터 공유에 어려운 점이 있고 용도에 맞는 데이터의 분류 자체가 어려운 경우가 많이 발생하기 때문에 새로운 데이터가 실시간으로 쌓이는 현재도 빅데이터 산업 발전은 느리게 진행되어가는 상황이다(Kim, 2015). 이처럼 농업 분야의 데이터셋은 활용도가 떨어지고, 데이터의 효율적인 분류 방법에 대한 연구도 미흡한 실정이다.

본 연구에서의 분석을 통해 데이터를 각 그룹별로 분류를 수행하고, 분류된 데이터는 새로운 데이터가 추가되어도 어느 집단으로 식별될 것인지의 예측을 용이하게 해줄 방법 제시와 평가를 하고자 한다. 따라서 본 연구의 목적은 파프리카 온실의 환경데이터를 주성분 분석과 판별분석을 통해 분류하고 식별하는 방법을 도출함으로써 방대한 데이터를 가공하고 데이터의 활용도를 높이는 데 있다.

재료 및 방법

1. 실험 재료

1.1 실험 장소

본 연구의 실험 장소는 경남 함안군 범수면 지역의 파프리카 재배 온실 3곳이다. 온실 내부 환경 데이터는 설치된 파프리카 재배용 복합환경 제어시스템을 통해 수집하였다. 이들 온실

의 상호간 직선거리는 2.0km 이내에 위치하여 외부 기상환경 편차가 적으며, 온실의 형태, 재배작목 및 품종이 동일하다. 재배면적은 약 4,950~5,280m²이며, 각 온실 면적은 같다. 온실 설치 방향은 두 온실(A, B)이 남-북 방향으로, 나머지 한 온실(C)은 동-서 방향으로 설치되어 있다. 대상 온실은 농촌진흥청 원예특작시설 내재해형 규격의 자동화비닐하우스(10-자동화-1형)이며, 동고가 7.4m, 측고가 5.4m로 파프리카와 토마토 등의 다단 유인재배에 용이한 형태의 온실이다.

1.2 대상 온실 특징

온실 내에 재배되는 작물은 파프리카이며, 3곳 모두 동일한 품종인 나가노이다. 조사대상 온실의 조사 시점은 전체 재배 기간의 2/3 정도가 경과 된 시기이며, 정식 후 220±10일이다. 그리고 조사 시점까지의 생육상황이 균일하다고 가정하였고, 온실 내 식물체 생육 상황도 40±5마디 정도에서 개화, 결과, 수확이 진행되는 생육상황이었다. 파프리카의 재배 간격은 약 1.6m, 정식간격은 약 20.0cm이며, 작물 재배에 사용되는 용수는 지하수로 조사대상 3개 온실 모두 동일하였다.

온실의 환경측정은 대상 온실 3곳 모두 Fig. 1 같이 국내 복합환경제어기(Shinhan A-TEC, GK-3000, Korea)와 Fig. 2



Fig. 1. Composite environment controller.



Fig. 2. Instrumentation sensor.

와 같은 계측 센서를 활용하여 환경데이터를 수집하였다(Lee, 2020).

1.3 환경데이터 전처리

수집된 모든 데이터는 온실 내의 데이터 로거에 기록된 값이다. Fig. 3처럼 수집 항목은 온도(°C), 상대습도(%), CO₂ 농도(ppm)이며, 4월 1일부터 4월 28일까지 총 4주간의 매일 1시간마다 측정된 값을 수집하여 사용하였다. 사용된 데이터 중 이상치는 Fig. 3(c)의 마지막 부분과 같이 분석을 수행하기 전에 제외하였다.

먼저, 다중 정규성 검정을 통해 정규성을 확인한 후 Box-cox

변환을 이용한 데이터 전처리를 수행하였다. Box-cox 변환은 정규분포가 아닌 데이터들을 분산 안정화 및 정규화를 위해 모수(parameter)인 λ 값을 식(1)과 같이 사용하여 정규분포에 가장 가까운 형태를 취하는 값을 찾아서 사용한다. 이 변환은 멱변환(power transformation)으로도 알려져 있다(Na, 2017).

$$y^{(\lambda)} = \begin{cases} \frac{y^\lambda - 1}{\lambda} & \text{if } \lambda \neq 0, \\ \ln y & \text{if } \lambda = 0 \end{cases} \quad (1)$$

전처리를 거친 데이터의 약 80%는 훈련자료(training data)로, 나머지 약 20%는 테스트자료(test data)로 만들어 분석을 진행하였다.

2. 통계적 방법

본 논문에서는 환경데이터 3가지를 사용하여 주성분 분석과 판별 분석을 통해 환경 데이터의 식별을 위해 분류를 수행하였다. 앞의 2가지 방법 모두 차원 축소 개념으로 시각화하여 정보를 보여주는 것에 매우 유용하다. 하지만 주성분 분석은 군집의 정보를 사용하지 않는 비지도 학습이고, 판별 분석은 군집의 정보를 사용하는 지도 학습으로 더 나은 자료의 구분을 제공한다는 것을 기대(Na, 2017)하고, 두 가지 방법을 사용하여 분석을 진행하였다.

2.1 주성분 분석

주성분 분석(PCA, principal component analysis)은 변수 간의 상관관계가 있는 고차원의 데이터를 정보 손실을 최소화시켜 효율적으로 저차원의 데이터로 변환하는 방법 중 하나이다. 하나의 변수를 알면 그 변수와 상관관계를 가지는 다른 구성요소들의 값을 추정 가능하기 때문에, 작은 분산 값을 갖는 주성분을 제거하게 되면 원래 데이터와의 상관관계도 크게 잃지 않으면서 원형변수들의 개수보다 적은 수의 특성 변수들을 이용해 데이터를 얻을 수 있게 된다(Lee 등, 2014). 이는 수집된 샘플 3차원 데이터를 2개의 주성분만을 사용하여 표현할 수 있다는 의미이다(Wise 등, 1996). 이처럼 적은 개수의 변수로 다차원의 데이터를 분석 가능하다는 것이 주성분 분석에서 차원 축소(dimension reduction)의 개념이며(Yu 등, 2018), 고차원의 데이터에서 저차원의 특성화된 값으로 데이터의 차원을 줄이면서 고차원 문제를 해결하는 것이 PCA의 원리이다.

2.2 판별 분석

판별 분석의 종류는 여러 가지 방법이 있다. 하지만 본 연구

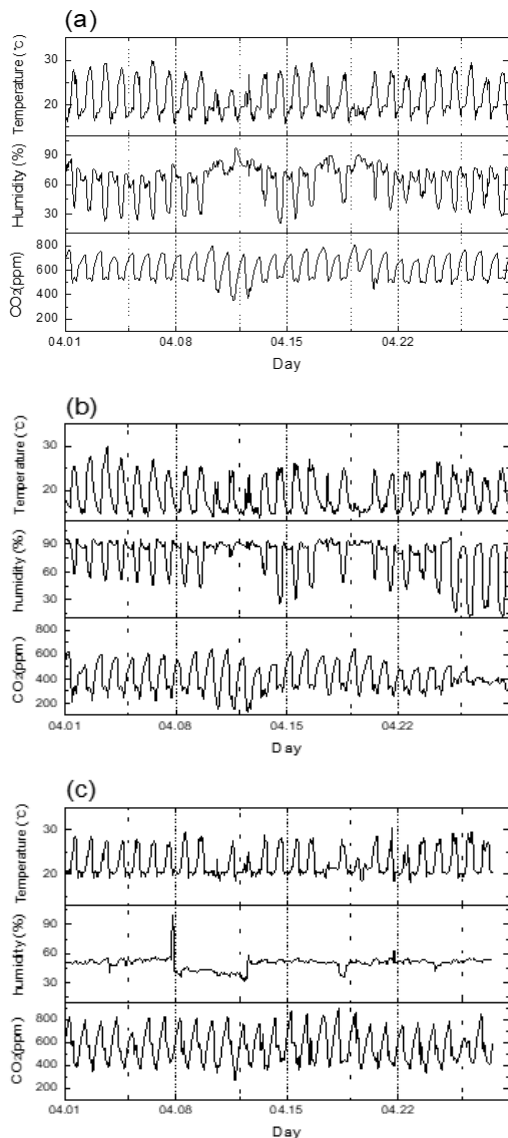


Fig. 3. Environment data; (a) A Greenhouse; (b) B Greenhouse; (c) C Greenhouse.

에서는 식별이 쉬운 선형판별분석(LDA, Linear Discriminant Analysis)을 수행할 것이다. 선형판별분석은 2개 이상의 그룹을 구분하는 식별 기법으로서, 그룹 내의 원소들의 분산을 최소화하고 그룹 사이의 분산을 최대화하는 축으로 특징 벡터를 사영(projection)시켜 중첩영역이 최소화되게 하는 새로운 축을 찾는 것이다(Kim 등, 2018). 두 그룹인 A와 B가 있다고 가정한다면, 두 그룹의 원소에 대한 각각의 측정값을 x와 y라고 한다. 새로운 축은 그룹 A와 그룹 B의 두 변수의 집합을 판별 점수로 나타내고 x와 y의 선형결합을 찾음으로써 그 결과를 판별 함수로 사영할 수 있다(G. Baudat과 F. Anouar, 2000). 판별분석법의 최종 결과는 새로운 데이터를 특정 집단의 분류에 최적화된 점인 판별함수를 찾는 것이다. 판별 분석을 통하여 도출하려는 독립변수들의 판별함수는 식(2)와 같이 판별 함수를 만들어 분류하고자 하는 각 대상들의 특징을 대입해서 각 대상들이 속하는 집단을 찾아낸다(Kim 등, 2020).

$$D = d_0 + d_1X_1 + d_2X_2 + \dots + d_nX_n \quad (2)$$

위의 식에서 D=판별점수, d_0 =상수, X=독립변수, d=가중치이다.

그리고 판별분석은 두 가지의 목적을 가진다. 하나는 알려진 모집단으로부터 관측값들의 집합을 구별하는 목적과 구분되지 않은 관측값을 특정 그룹으로 구분 지어 어떤 집단에 분류될 것인가를 결정하는 방법론이다(Kim 등, 2020).

이러한 통계 분석은 오픈소스인 R studio (4.0.2, R-studio, 250 Northern Ave, Boston, MA 02210)의 내장된 패키지

prcomp 함수(Lee, 2018)와 MASS 라이브러리의 lda 함수(Na, 2017)를 활용하여 주성분 분석과 선형 판별 분석을 수행하였다.

결과 및 고찰

1. 환경데이터 특성

온실 내의 환경데이터의 분포 특성을 알아보기 위해 각 항목 별로 히스토그램을 Fig. 4과 같이 그렸다. 데이터의 각 항목의 그래프를 보면 왼쪽부터 A, B, C 온실의 온도, 상대습도, CO₂ 농도 순서이다. 각각의 변수들은 단위도 모두 다르고 같은 변수라도 온실마다 환경 데이터의 분포 특징도 다른 것을 알 수 있다.

Fig. 4의 그림을 보면, 온도의 최빈값은 A 온실이 19~20°C, B 온실 15~16°C, C 온실 20~21°C였고, 상대습도의 최빈값은 A 온실 70~75%, B 온실 80~90%, C 온실 50~55%, CO₂ 농도는 A 온실이 500~550ppm, B 온실 300~350ppm, C 온실 400~450ppm으로 나타났다. 이처럼 Fig. 3과 Fig. 4의 C 온실의 상대습도를 본다면, 다른 온실보다 편차가 매우 적고 다른 패턴의 환경특성이 보여지는데, 이는 C 온실은 다른 온실과 달리 습도 관리하는 온실로써 상대습도의 편차가 적고 특정 온도의 빈도가 높게 나타났다.

2. 주성분 분석결과

먼저 PCA 분석을 수행한 결과를 Table 1에 나타내었다.

Table 1은 주성분 분석의 전체 분석값을 보여주며, 각 주성

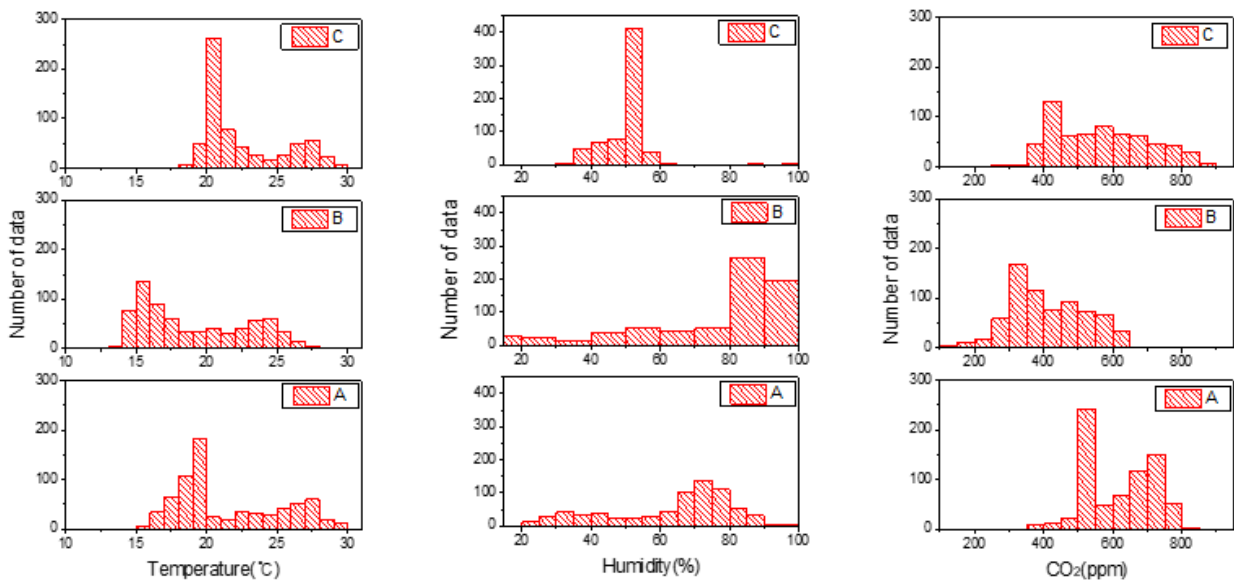
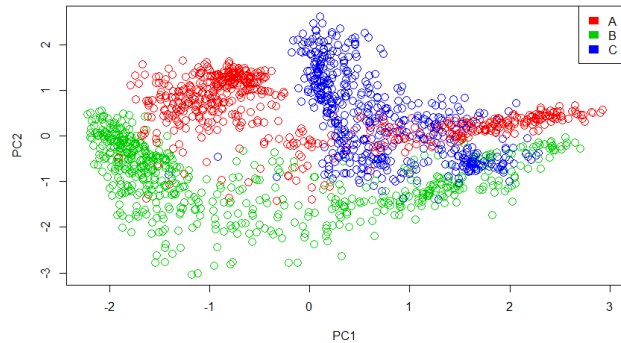
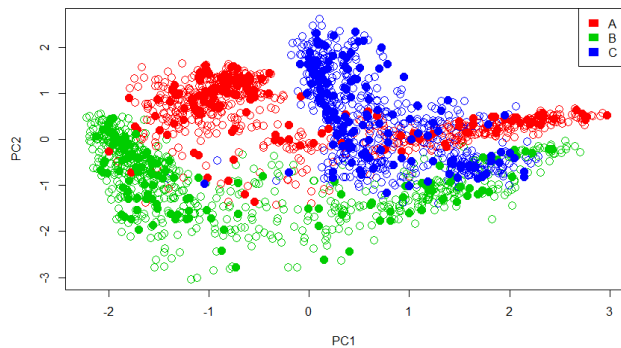


Fig. 4. Environmental data frequency.

Table 1. PCA (principal component analysis) value.

Variable	PC1	PC2	PC3
Temperature	0.7118112	-0.04941046	0.7006307
Humidity	-0.6830065	-0.28131111	0.6740669
CO ₂ concentration	-0.1637892	0.95834371	0.2339881

**Fig. 5.** PCA Result Evaluation (training data).**Fig. 6.** PCA Result Evaluation (training data+test data).

분을 의미하는 PC1, PC2, PC3에 대한 변수들의 참여도를 나타낸 것이다. 예를 들어 PC1, PC2, PC3를 본다면, $PC1 = 0.7118112 \times \text{Tem.} - 0.6830065 \times \text{Humi.} - 0.1637892 \times \text{CO}_2$ 의 식을 가지는 것이고, $PC2 = -0.04941046 \times \text{Tem.} - 0.28131111 \times \text{Humi.} + 0.95834371$ 그리고 $PC3 = 0.7006307 \times \text{Tem.} + 0.16740669 \times \text{Humi.} + 0.2339881 \times \text{CO}_2$ 의 식이 만들어진 것이다. 이렇게 계산된 PC1 값은 Fig. 5의 그래프에 x로 표현되는 점의 좌표가 된다.

Fig. 6는 주성분인 PC1과 PC2를 x축과 y축으로 나타내어 군집이 이루어지는지 분석결과를 평가해 본 그래프이고, A 온실, B 온실, C 온실의 쉬운 구분을 위하여 각각 다른 색으로 표시하였다. 이처럼 주성분 분석결과와 요약 통계량을 나타내는 함수를 사용하여 주성분 PC1, PC2, PC3의 표준편차 (standard deviation)와 각각의 기여율(proportion of variance) 그리고 누적 기여율(cumulative proportion)을 Table 2을 통

Table 2. PCA (principal component analysis) summary statistic.

Classification	PC1	PC2	PC3
Standard deviation	1.3135	1.0117	0.50121
Proportion of variance	0.5751	0.3412	0.08374
Cumulative proportion	0.5751	0.9163	1.00000

Table 3. Discriminant function coefficient.

Classification	LD1	LD2
Temperature	0.8622565	0.7648161
Humidity	-0.1805741	1.4024534
CO ₂ concentration	1.4018140	0.5539150
Constant	0.03040701	-0.01542835

해 확인할 수 있다. 여기서 기여율이란 각 항목이 전체에 대해 어느 정도의 영향력을 가지는지를 의미한다.

Table 2는 각각의 표준편차와 각각의 기여율 그리고 누적 (cumulative)된 기여율을 확인하는 표이다. 이는 PC1 성분 하나로 57.51%를 설명할 수 있고, PC2는 34.12%, PC3는 8.374% 정도를 설명할 수 있다는 의미이다(Lee, 2018).

3. 판별 분석결과

다음은 선형판별분석을 수행한 결과를 Table 3에 나타내었다.

위의 표 Table 3에서 각 2개 선형판별식(LD1, LD2)의 선형결합계수가 나온다. 표에서 만들어지는 식은 $LD1 = 0.8622565 \times \text{Tem.} - 0.1805741 \times \text{Humi.} + 1.4018140 \times \text{CO}_2 + 0.03040701$ 와 $LD2 = 0.7648161 \times \text{Tem.} + 1.4024534 \times \text{Humi.} + 0.55391503 \times \text{CO}_2 - 0.01542835$ 이다. 3개의 변수들을 두 식에 대입한 뒤 그 결과 값(판별 점수)이 높은 식을 사용하여 분석을 수행한다. 둘 중 판별 점수가 높은 식은 LD1이며, LD1 식으로 판별된 값을 시각적으로 표현하기 위해 판별 차원상의 분포를 밑의 Fig. 7로 표시하였다.

Fig. 7에서 판별식으로 나온 A 온실의 데이터는 -2.47108~2.419918의 범위를 가지며 중앙값은 1.069775이고, B 온실의 데이터는 -5.03431~0.276817의 범위를 가지며 중앙값은 -1.61118이다. C 온실의 데이터는 -2.71655~3.763126의 범위를 가지며 중앙값은 0.816473이다. 이렇게 판별된 값은 테스트 자료 또는 새로운 데이터가 어느 부분으로 식별될 것 인지의 기준이 된다. LD1 식으로 판별된 값의 예측 정확도를 확인하기 위해 모델 평가를 수행한다. 모델 평가는 모든 온실의 데이터 80%를 훈련자료로 사용하였고, 이렇게 학습된 LDA 모델에 전체 데이터의 20%인 테스트 자료를 넣어 예측

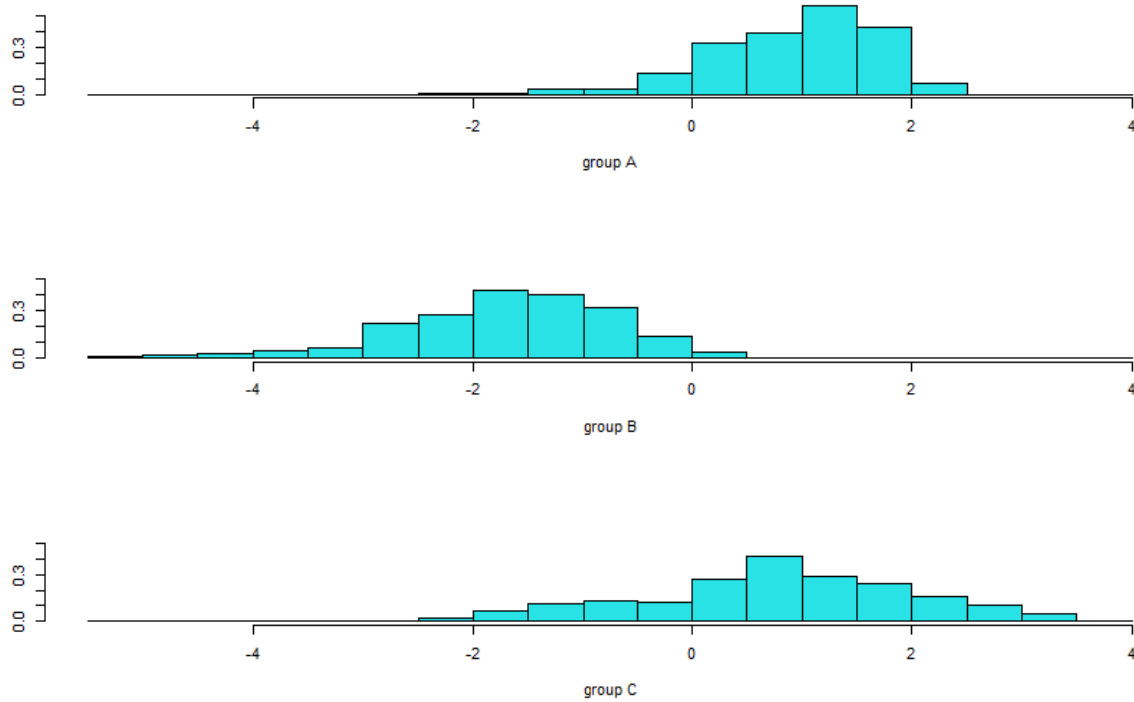


Fig. 7. Discriminant dimensional distribution of LD1 equation.

Table 4. Classification of test results.

Test (input) Training (output)	A Greenhouse data	B Greenhouse data	C Greenhouse data
A Greenhouse	110	2	58
B Greenhouse	11	116	21
C Greenhouse	37	21	55

소속집단의 분류 결과를 Table 4로 나타내었다.

분류 결과를 보면, 훈련자료 바탕으로 만들어진 모델에 A의 테스트 자료 170개 중 110개가 A로 분류되었고 B의 테스트 자료는 148개 중 116개가 B로 분류되었으며, C의 테스트 자료는 101개 중 55개가 C로 분류되었다. 미리 예측된 모델 집단 유형에 속하는 데이터의 예측 정확도는 약 0.6706로 이 자료에서의 제1 선형판별식(LD1)이 자료의 군집 간 분산의 67.06%를 설명한다는 의미이다.

본 연구에서는 온실 3곳의 식별을 위해 통계적인 방법으로 분류를 하고자 주성분 분석과 선형 판별 분석을 수행하였다. 먼저 주성분 분석의 결과 평가를 확인해 보면, 뚜렷한 구분이 된다는 부분도 있지만 겹치는 부분이 많다는 것을 알 수 있다. 그것은 복잡한 환경데이터의 특징도 있고, 분석을 수행한 데이터의 종류와 개수가 다양하지 않았기 때문에 뚜렷한 구분을 나타내기에는 부족했다고 판단된다(Lee 등, 2018). 여기서 수행한 데이터인 온도, 습도, CO₂ 농도 이외에 pH 값이나 조도

나 일사량 등 더 다양한 데이터의 종류가 있다면, PCA의 장점인 데이터 축소를 더욱 용이하게 사용할 수 있었다고 판단된다(Lee 등, 2014). 그리고 데이터의 종류 외에 훈련데이터들의 개수도 많았다면 모집단에 가까운 데이터 값이므로 특징의 보존율이 높아 더욱 뚜렷한 분류 결과를 기대할 수 있다(Yu 등, 2018). 두 번째로 수행한 판별분석의 결과 평가를 확인해 보면, Table 4의 C 온실의 데이터의 예측 정확도가 매우 떨어진다는 것을 확인할 수 있는데 이는 Fig. 7에서의 group C는 범위가 넓고 다른 그룹과 맞물리는 부분이 상당 부분을 차지하기 때문에 예측 정확도가 떨어졌다고 판단된다. 이 경우 비선형 함수를 이용하여 분리가 가능한 일반화된 판별분석 또는 커널판별분석 등 더욱 고도화된 분석 방법을 통해 정확도가 높은 분리가 가능할 것으로 판단된다(Kim 등, 2018).

본 연구를 통해서 온실 내부 환경데이터를 이용한 온실의 분류 기준 설정을 해보았다. 분석결과로는 완벽한 분류가 이루어지지는 않았으나 이 문제는 분석한 데이터의 종류의 다양성

과 더 방대한 데이터의 양으로 고도화된 분석을 진행한다면, 좀 더 정확도가 높은 결과가 도출될 것이다. 향후 온실 내 데이터의 활용도가 높아져 농업의 다양한 분야에 공유가 가능해진다면, 가공된 데이터들은 빅데이터에 기반을 둔 스마트팜 환경에서 첨단기술의 융합과 복합이 이루어지는 데에 활용될 것이다. 이는 수많은 데이터가 쌓이고 있는 요즘 데이터의 분류를 통해 쌓여있는 데이터는 물론이고 새롭게 수집되는 데이터들의 활용도를 높여, 농업 분야의 스마트팜 보급을 통해 농업 경쟁력을 높일 것이다.

적 요

본 연구에서는 같은 지역에 위치한 온실 3곳의 식별을 위해 통계적인 방법으로 분류를 하고자 주성분 분석(PCA)과 선형 판별 분석(LDA)을 수행하였다. 온실 내의 환경데이터는 같은 지역의 온실 3곳을 대상으로 4월 1일부터 4월 28일 총 4주간 1시간 간격으로 수집된 값을 사용하였다. 데이터를 분석하기 전, 데이터 정규화를 시키는 전처리를 거쳤으며, 전체의 80%인 훈련자료(training data)와 20%인 테스트 자료(test data)로 나누어 분석을 수행하였다. 분석을 수행한 결과, PC1은 57.51%의 설명력으로 $PC1 = 0.7118112 \times Tem. - 0.6830065 \times Humi. - 0.1637892 \times CO_2$ 의 식을 가지며, LD1은 67.06%의 설명력으로 $LD1 = 0.8622565 \times Tem. - 0.1805741 \times Humi. + 1.4018140 \times CO_2 + 0.03040701$ 의 식을 가지는 것으로 나타났다. 이렇게 미리 분류시켜 놓은 온실의 데이터를 바탕으로 새로운 환경의 데이터를 입력하였을 때 특정 그룹으로의 분류가 가능함으로써 데이터의 성향을 파악할 수 있다. 이러한 데이터는 식별을 용이하게 함으로써 데이터의 활용도를 높여주는 방법이라고 판단된다.

추가주제어: 분류, 빅데이터, 스마트팜, 주성분 분석, 판별분석

사 사

본 결과물은 농림축산식품부의 재원으로 농림식품기술기획평가원의 농식품기술융합창의인재양성사업의 지원을 받아 연구되었음(717001-7).

Literature Cited

- Baudat G. and F. Anouar. 2000. Generalized Discriminant Analysis Using a Kernel Approach. *Neural Computation* 12:2385-2404.
- Cho H.J. and C.H. Kang. 2019. A study on the architectural planning of vertical smart farm. *J. Korean Archit. Inst. Korea's Acad. Presentation Conference*. 39:114-117 (in Korean).
- Choi G.H. 2019. A study on Strategies and Relevant Legislation for Responding to Climate Change by Using Core Technologies of the 4th Industrial Revolution – Focusing on Big Data -. *Kangwon Law Rev.* 58:779-815 (in Korean).
- Hwang D.H. and G.J. Park. 2020. An Analysis on the Educational Needs for the Smart Farm: Focusing on SMEs in Jeon-nam Area. *J. Korea Acad. Coop. Soc.* 21:649-655 (in Korean).
- Jang I.H., Y.C., Choe, Y. Kwack, C.H. Chun, and T.W. Kim. 2016. Environmental Management Status of Domestic Field Greenhouses Through Analyzing the Aggregated Environmental and Actuator Data from Smartfarm Greenhouses. *Hortic. Abstracts*. p. 222-223 (in Korean).
- Jeong Y.K., J.G. Lee, E.K. Ahn, J.S. Seo, and Y.C. Yoon. 2020. Performance Improvement of a Temperature and Humidity Measuring System for Strawberry Cultivation Greenhouse. *Protected Hort. Plant Fac.* 29:110-119 (in Korean).
- Joung K.H., H.J. Jin, J.U. An, H.S. Yoon, S.S. Oh, C.S. Lim, Y.C. Um, H.D. Kim, and S.M. Park. 2018. Analysis of Growth Characteristics and Yield of Paprika Cupra and Fiesta for Yield Prediction. *Hortic. Abstracts*. p. 216-217 (in Korean).
- Kim D.W., T.H. Kim, J.W. Seok, and K.S. Bae. 2018. Sonar Target Classification using Generalized Discriminant Analysis. *J. Korea Inst. Information Comm. Eng.* 22:125-130 (in Korean).
- Kim H.J., H.K. Noh, and T.H. Kang. 2020. Discriminant analysis to detect fire blight infection on pear trees using RGB imagery obtained by a rotary wing drone. *Korean J. Agri. Sci.* 47:349-360.
- Kim J.H., E.S. Lee, D.C. Choi, M.S. Kim, S.J. Kim, N.J. Choi, J.H. Choi, and J.D. Lee. 2020. Design and Implementation of Automatic Control Smartfarm Platform using IOT Technology. *Proceedings of the Korean Society Computer Information Conference*. 28:71-72 (in Korean).
- Kim M.H. 2015. Big Data Reality, Big Difference with Expectations. *LGERI Report*. p. 2-15 (in Korean).
- Kim T.S., D.S. Park, and H.J. Kim. 2018. Study of smart farm model using Big Data. *Proceedings of Symposium of the Korean Institute of communications and Information Sciences*. p. 1408-1409 (in Korean).
- Korean Statistical Information Service (KOSIS). 2019. Farm population survey <http://kosis.kr/> (in Korean).
- Lee J.G., Y.K. Jeong, S.W. Yun, M.K. Choi, H.T. Kim, and Y.C. Yoon. 2018. Field Survey on Smart Greenhouse. *Protected Hort. Plant Fac.* 27:166-172 (in Korean).
- Lee K.G. 2020. Growth Interpretation of Paprika for Greenhouse Cultivation by Major Environmental Factors. *PhD Dissertation, Gyeongsang National Univ.* p. 10-24 (in Korean).
- Lee K.J., B.W. Lee, D.H. Choi, T.O. Kim, and D.I. Snin. 2014. A Study on Fault Detection Monitoring and Diagnosis System of CNG Stations based on Principal Component

- Analysis(PCA). J. Korean Inst. Gas, 18:53-59 (in Korean).
- Lee T.G. 2018. Analysis and visualization of basic environmental data using R. Korea. Book Publishing CIR. p. 107-110 (in Korean).
- Na J.H. 2017. R Application Multivariate Analysis. JaYoo-Academy. Korea. p. 129-138 (in Korean).
- Nam D.S., J.W. Lee, T.W. Moon, and J.E. Son. 2017. Transpiration Prediction of Sweet Peppers Hydroponically-grown in Soilless Culture via Artificial Neural Network Using Environmental Factors in Greenhouse. Protected Hort. Plant Fac. 26:411-417 (in Korean).
- Park S.H., H.R. Lee, S.J. Park, and Y.B. Cho. 2019. Comparative Study on Environment and Growth Management of Strawberry Smart Farm by Greenhouse Type. Hortic. Abstracts. p. 186-186 (in Korean).
- Wise B.M. and N.B. Gallagher. 1996. The Process Chemometrics Approach to Process Monitoring and Fault Detection. J. Process Control 6:329-348.
- Yang H.C., J.S. Lee, H.D. Lee, and H.S. Kim. 2018. Automatic Detection of Paprika Diseases/Pests Outbroken during the Hydroponic Cultivation in Greenhouse using Artificial Intelligence. J. Inst. Control. Robot. Syst. 24:1020-1024 (in Korean).
- Yu Y.J., S.H. Moon, and S.H. Park. 2018. Measurement of the Visibility of the Smoke Images using PCA. J. Korea Inst. Information Comm. Eng. 22:1474-1480 (in Korean).