

선형 판별 분석을 이용한 비파괴적 새싹삼 생장 분석 방법

박근호*, 최강인*, 권기현**, 백승훈**, 정성환*

*한국전자기술연구원 IT응용연구센터

**한국식품연구원

khpark0224@keti.re.kr, cki921@keti.re.kr, kkh@kfri.re.kr, bsh917@naver.com, shjeong@keti.re.kr

Non-destructive Sprout Ginseng Growth Analysis Method Using Linear Discriminant Analysis

Keunho Park*, Kangin Choi*, Ki-Hyun Kwon**, Seunghoon Baek**, Sunghwan Jeong*

*IT Application Research Center, Korea Electronics Technology Institute

**Korea Food Research Institute

요약

최근 비파괴분석 방법으로 작물의 생장 및 병충해를 분석하는 초분광 영상을 이용한 분석방법에 관한 연구가 이루어지고 있다. 하지만 원격방식인 비파괴분석 방법은 파괴분석 방법에 비해 정확성과 신뢰성이 떨어지는 단점을 극복해야 한다는 한계를 지니고 있다. 본 논문은 초분광 영상을 이용하여 수경재배 새싹삼의 생장을 분석한다. ‘수소 수’, ‘펩타이드 0.1ppm’, ‘펩타이드 1ppm’, 그리고 ‘펩타이드 5ppm’ 네 가지 재배조건이 다른 실험군을 분류 비교하였다. 초분광 영상 데이터를 선형 판별 분석을 이용하여 차원 축소하여 생장 환경별로 다른 분류를 보이는 것을 확인하였고 값을 정량화하여 실제 파괴분석 방법으로 형질 분석 및 성분분석 결과와 비교하여 비파괴분석 방법의 유효성을 입증하였다.

1. 서론

작물 생장 및 병충해를 분석하여 출하 시기를 결정하거나 병충해를 예방하는 일은 농가의 수입을 증가시키고 유통을 안정화할 수 있어 농업에 미치는 영향이 매우 크다. 작물의 형질 및 성분을 분석하는 방법은 파괴분석 방법, 비파괴분석 방법 두 가지로 나뉘는데 화학적, 물리적 검사로 인하여 작물에 변형을 가하는 방법인 파괴분석 방법보다 원격방식으로 작물을 분석하는 비파괴분석 방법이 선호된다[1, 2]. 하지만 비파괴분석 방법은 파괴분석 방법에 비해 정확성과 신뢰성이 떨어지는 단점이 있다. 최근 비파괴분석 방법의 일종인 초분광 영상을 이용하여 작물의 생육상태를 분석하는 연구가 이루어지고 있다[3, 4]. 본 논문에서는 거리 계량 학습을 기반으로 새싹삼의 생장을 분석하는 연구를 수행하였다. 선형 판별 분석을 이용하여 새싹삼을 실험군별로 분석하고 새싹삼의 형질 분석과 성분분석 결과와 비교하여 실험의 유효성을 입증하였다.

2. 비파괴적 새싹삼 생장 분석 방법

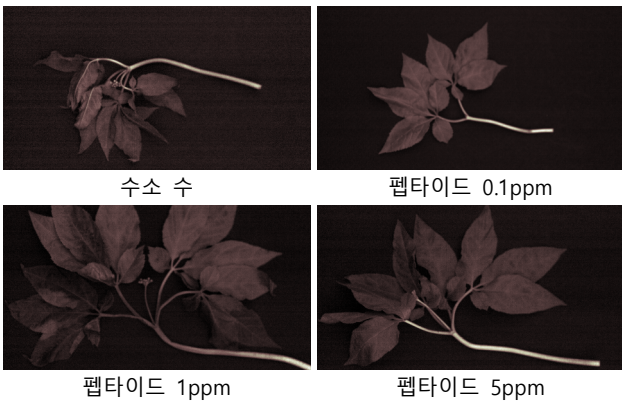
본 논문에서는 거리 계량 학습(distance metric learning)을 기반으로 새싹삼의 생장을 비파괴적으로 분석하는 연구를 수행하였다. 거리 계량 학습이란 데이터 간의 유사도를 학습하는 방법을 의미하며 분류, 회귀, 탐지 등 다양한 머신러닝 분

야의 알고리즘은 어떤 거리 계량(distance metric)을 선택하는가에 따라 성능의 영향을 크게 받는다. 본 논문의 실험에 사용된 초분광 영상은 초분광 카메라 Snapscan VNIR(IMEC Inc.), 할로겐 램프(FOMEX HI000, Halogen 650W, 2Set)가 설치된 실내에서 획득하였다. 초분광 영상의 전체 볼륨 크기는 1280(Width)×800(Height)×150(Band)×32(Bit)이다. 실험에 사용된 데이터는 ‘수소 수(hydrogen water)’, ‘펩타이드(peptide) 0.1ppm’, ‘펩타이드 1ppm’, 그리고 ‘펩타이드 5ppm’ 네 가지 조건으로 분류하였다. 표 1은 각 실험군을 수경 재배할 때 사용한 양액의 배합조건을 나타낸다. ‘수소 수’ 샘플의 경우 물과 수소 수를 1:1의 비율로 혼합하여 양액을 제조하였고, ‘펩타이드 0.1ppm’, ‘펩타이드 1ppm’, ‘펩타이드 5ppm’의 경우 20L의 물에 펩타이드를 각각 2mg, 20mg, 100mg 혼합하여 양액을 제조하였다. 표 2는 13일간 수경재배 환경에서 생장한 새싹삼의 초분광 영상을 나타낸다.

[표 1] 새싹삼 실험군별 양액 배합조건

	물	수소 수	펩타이드
수소 수	10L	10L	-
펩타이드 0.1ppm	20L	-	2mg
펩타이드 1ppm	20L	-	20mg
펩타이드 5ppm	20L	-	100mg

[표 2] 새싹삼의 초분광 영상



2.1 초분광 영상

초분광 데이터는 RGB 데이터에 비해 많고(many), 연속적이고(continuous), 파장폭이 좁은(narrow) 정보를 내포한다. 따라서 초분광 데이터는 다양한 방법으로 분석이 가능하며 현재 초분광 데이터를 분석하는 다양한 기술들이 존재한다. 초분광 데이터는 400~2,500nm의 육안으로 인식이 가능한 가시광선부터 인식이 불가능한 적외선 영역의 주파수 대역을 포함하고 있다. 이러한 특성으로 육안으로 분별 불가능한 정보도 분석 가능하여 항공기에 탑재하여 광물자원의 분포를 파악하거나 암석 종류 구분 등 지질분야에 주로 활용되었다. 식생 분야에서는 식물의 생화학적 구성 인자들의 추정에 많은 연구가 수행되고 있으며 더 나아가 식물에 나타나는 각종 스트레스 탐지에 활용된다. 또한 수질 모니터링이나, 도시지역에서 각종 인공물의 분류 등 그 활용 분야가 확대되고 있다.

2.2 선형 판별 분석

선형 판별 분석(linear discriminant analysis, LDA)[5]은 주 성분 분석법과 더불어 대표적인 특징벡터 차원축소기법 중 하나이다. 선형 판별 분석은 클래스 간 분산을 최대화하고 클래스 내 분산을 최소화하는 방식으로, 특징 벡터의 차원을 축소하는 방법이라고 할 수 있다. 클래스 C_1 에 속하는 n_1 개와 클래스 C_2 에 속하는 n_2 개의 d 차원 표본 집합 $x = \{x_1, x_2, \dots, x_n\}$ 이 있다고 가정했을 때, 표본 집합의 1차원 선형 투영은 수식 과 같다.

$$y = W^T x \tag{1}$$

클래스간 분산을 최대화하기 위해서는 투영 벡터들의 평균 벡터를 계산하여 클래스의 평균을 크게 유지하여야 하며, 클래스 내 분산을 최소화하기 위해서는 클래스 내 분산이라는 척도로 평균간의 차이를 정규화하여야 한다. 이를 위해 원 데이터의 표본 x 의 평균 벡터 μ_i 와 투영된 표본 y 의 평균 벡터 $\tilde{\mu}_i$ 는 수식 (2)와 같이 정의할 수 있고, 각 클래스에 대한 분산은 수식 (3)과 같이 정의할 수 있다.

$$\mu_i = \frac{1}{n} \sum_{x \in C_i} x, \tilde{\mu}_i = \frac{1}{n} \sum_{x \in C_i} y = \frac{1}{n} \sum_{x \in C_i} W^T x = W^T \mu_i \tag{2}$$

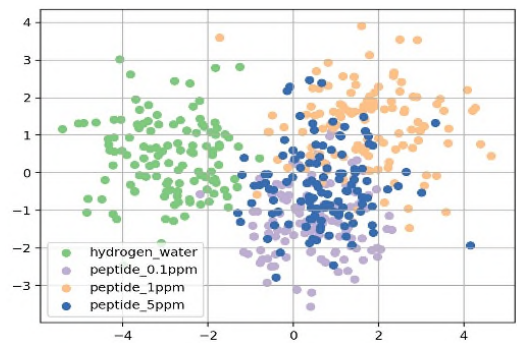
$$\tilde{s}_i^2 = \sum_{x \in C_i} (y - \tilde{\mu}_i)^2 \tag{3}$$

선형 판별 분석은 수식 (4)의 목적함수를 최대화하는, 즉 동일한 클래스의 표본들은 인접하게 투영하고, 동시에 클래스 간의 투영은 중심이 가능한 멀리 떨어지게 하는 선형 변환 행렬 W 를 찾아내는 것이다.

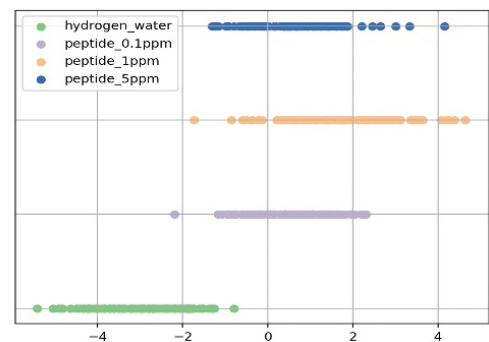
$$J(W) = \frac{|\tilde{\mu}_1 - \tilde{\mu}_2|^2}{(\tilde{s}_1^2 + \tilde{s}_2^2)} \tag{4}$$

2.3 실험 및 결과

본 논문에서는 선형 판별 분석을 이용하여 새싹삼의 생장을 분석하였다. 실험을 위해 표 2와 같이 수소 수, 펩타이드 0.1ppm, 1ppm, 그리고 5ppm의 네 가지 혼합 양액 조건에서 재배한 새싹삼의 잎을 초분광 카메라로 촬영한 영상에서 조건별로 120개의 임의의 지점에서 초분광 카메라의 밴드 수와 일치하는 150 크기의 백터를 추출하였다. 추출한 480개의 샘플들을 선형 판별 분석을 이용하여 2차원으로 축소한 결과는 그림 1과 같다. 그림 1에서 ‘수소 수’ 범주에 속하는 샘플들과 ‘펩타이드 0.1ppm’, ‘펩타이드 1ppm’, ‘펩타이드 5ppm’와 같이 ‘펩타이드’의 범주에 속하는 샘플들은 확연히 다른 분류를 보이는 것을 확인할 수 있다.



[그림 1] 새싹삼 초분광 영상의 2차원 LDA 결과



[그림 2] 새싹삼 초분광 영상의 1차원 LDA 결과

선형 판별 분석을 이용하여 샘플들을 1차원으로 축소한 결과는 그림 2와 같다. 그림 2에서 ‘수소 수’ 범주에 속하는 샘플

들은 가로축을 기준으로 낮은 값을 보이는 것을 볼 수 있고, ‘웹타이드’의 범주에 속하는 샘플들은 가로축을 기준으로 높은 값을 보이는 것을 볼 수 있다.

실험결과를 정량화하기 위해서 1차원 선형 판별 분석의 결과값 중 같은 데이터에 속하는 샘플들의 평균값은 표 3의 마지막 행에 해당한다. 네 가지 재배 조건 중 ‘웹타이드 1ppm’이 1.17으로 가장 높고 ‘수소 수’가 -2.91으로 가장 낮다. 이 값들의 유효성을 평가하기 위해 새싹삼의 형질 측정과 성분분석 결과는 표 3의 두 번째 열부터 여덟 번째 열에 해당한다. 형질 측정은 전체 길이, 뿌리 길이, 줄기 길이, 잎 개수 네 가지를 측정하였고, 성분분석은 진세노사이드 Rb1, 진세노사이드 Rg1, 진세노사이드 Rg3 세 가지를 분석하였다. 분석결과 전체 길이, 줄기 길이, 진세노사이드 Rb1 세 가지 항목에서 ‘웹타이드 1ppm’이 가장 높은값을 나타내었고 다른 항목에서도 상위권에 속하는 값을 가지고 있어 선형 판별 분석의 유효성을 입증하였다.

[표 3] 새싹삼 성장 분석 결과

	수소 수	웹타이드 0.1ppm	웹타이드 1ppm	웹타이드 5ppm
전체 길이 (cm)	24.17	24.04	25.36	24.53
뿌리 길이 (cm)	14.01	14.26	14.10	13.92
줄기 길이 (cm)	10.16	9.78	11.26	10.61
잎 개수(개)	13.05	12.48	12.87	14.52
진세노사이드 Rb1(mg/g)	0.05	0.02	0.07	0.03
진세노사이드 Rg1(mg/g)	0.48	0.66	0.49	0.45
진세노사이드 Rg3(mg/g)	0.02	0.02	0.02	0.02
선형 판별 분석(ID)	-2.91	0.48	1.17	0.94

3. 결론

본 논문에서는 거리 계량 학습을 기반으로 새싹삼의 생장을 비파괴분석 방법으로 분석하는 연구를 수행하였다. ‘수소 수’, ‘웹타이드 0.1ppm’, ‘웹타이드 1ppm’, 그리고 ‘웹타이드 5ppm’ 네 가지 조건으로 수경 재배된 새싹삼을 1280×800×150 볼륨 크기의 초분광 영상으로 촬영하였고 각 실험군 별로 120 개의 임의의 지점에서 초분광 카메라의 밴드 수와 일치하는 150 크기의 벡터를 추출하였다. 추출한 480개의 샘플들을 선형 판별 분석을 이용하여 2차원으로 시각화하여 ‘수소 수’, ‘웹타이드’ 재배 조건 간의 차별성을 보였으며, 선형 판별 분석을 이용하여 1차원으로 정량화하여 형질 측정과 성분분석 결과

값들과 비교하여 선형 판별 분석의 유효성을 입증하였다.

이 논문은 2020년도 정부(산업자원통상부)의 재원으로 한국산업기술진흥원의 지원을 받아 수행된 연구임 (No.P0004795)

참고문헌

- [1] F. Martinelli, R. Scalenghe, S. Davino, S. Panno, G. Scuderi, P. Ruisi, P. Villa, D. Stroppiana, M. Boschetti, and L. R. Goulart, "Advanced methods of plant disease detection. a review", *Agronomy for Sustainable Development*, Vol. 35, No. 1, pp. 1-25, Jan. 2015.
- [2] Y. Fang and R. P. Ramasamy, "Current and prospective methods for plant disease detection", *Biosensors*, Vol. 5, No. 3, pp. 537-561, Aug. 2015.
- [3] K. I. Choi, H. M. Noh, S. H. Jeong, C. J. Yoo, "Classification of Growth Conditions in Crops Using Hyperspectral Images and Deep Neural Network : Case Study of Paprika Leaf", *The Journal of Korean Institute of Information Technology*, Vol. 17, No. 12, pp. 1-12, Dec. 2019.
- [4] K. Park, Y. K. Hong, G. H. Kim, and J. Lee, "Classification of apple leaf conditions in hyper-spectral images for diagnosis of Marssonina blotch using mRMR and deep neural network", *Computers and Electronics in Agriculture*, Vol. 148, pp. 179-187, May 2018.
- [5] A. M. Martinez, A. C. Kak, "PCA versus LDA", *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. 23, No. 2, pp. 228 - 233, 2001.