Korean Journal of Remote Sensing, Vol.33, No.5-2, 2017, pp.631~640
http://dx.doi.org/10.7780/kjrs.2017.33.5.2.3

ISSN 1225-6161 (Print)
ISSN 2287-9307 (Online)

Article

SSAE 알고리즘을 통한 2003-2016년 남한 전역 쌀 생산량 추정

마종원*, 이경도**, 최기영***, 허준*†

*연세대학교 건설환경공학과, **농촌진흥청 국립농업과학원, ***통계청 농어업통계과

Rice Yield Estimation of South Korea from Year 2003-2016 Using Stacked Sparse AutoEncoder

Jong Won Ma*, Kyungdo Lee**, Ki-Young Choi*** and Joon Heo*†

*School of Civil and Environmental Engineering, Yonsei University

**National Institute of Agricultural Science, RDA

***Fisheries and Agriculture Department, Statistics Korea

Abstract : The estimation of rice yield affects the income of farmers as well as the fields related to agriculture. Moreover, it has an important effect on the government's policy making including the control of supply demand and the price estimation. Thus, it is necessary to build the crop yield estimation model and from the past, many studies utilizing empirical statistical models or artificial neural network algorithms have been conducted through climatic and satellite data. Presently, scientists have achieved successful results with deep learning algorithms in the field of pattern recognition, computer vision, speech recognition, etc. Among deep learning algorithms, the SSAE (Stacked Sparse AutoEncoder) algorithm has been confirmed to be applicable in the field of forecasting through time series data and in this study, SSAE was utilized to estimate the rice yield in South Korea. The climatic and satellite data were used as the input variables and different types of input data were constructed according to the period of rice growth in South Korea. As a result, the combination of the satellite data from May to September and the climatic data using the 16 day average value showed the best performance with showing average annual %RMSE (percent Root Mean Square Error) and region %RMSE of 7.43% and 7.16% that the applicability of the SSAE algorithm could be proved in the field of rice yield estimation.

Key Words : Deep learning, AutoEncoder, Rice yield estimation, Climatic data, Satellite data, Remote sensing

요약 : 쌀 생산량 예측 및 조사는 농가 소득 보전 및 농업 분야 기관에 영향을 주고 수급 조절과 가격 예측 등 정부의 정책 수립과 관련하여 중요한 의미를 갖는다. 이에 따라 작황 추정 모델의 구축이 필요하며 과거

Received August 29, 2017; Revised September 17, 2017; Accepted September 20, 2017; Published online October 16, 2017.

† Corresponding Author: Joon Heo (jheo@yonsei.ac.kr)

This is an Open-Access article distributed under the terms of the Creative Commons Attribution Non-Commercial License (http://creativecommons. org/licenses/by-nc/3.0) which permits unrestricted non-commercial use, distribution, and reproduction in any medium, provided the original work is properly cited.

로부터 기상 자료 및 위성 자료를 통해 경험적 통계 모델 또는 인공신경망 알고리즘을 기반으로 한 연구가 다수 진행되었다. 현재 인공신경망 모델을 기반으로 개발된 딥 러닝 알고리즘이 패턴 인식, 컴퓨터 비전, 음성 인식 등의 분야에서 폭넓게 사용되며 뛰어난 성능을 보이고 있다. 최근 다양한 딥 러닝 알고리즘 중 SSAE 알고리즘이 시계열 자료를 통한 예측 분야에서 적용 가능성이 확인되었으며 본 연구에서는 SSAE를 통해 남한 전역에 대한 쌀 생산량 추정 연구를 진행하였다. 입력 변수로 기상자료와 위성자료를 사용하였으며 남한 벼의 생육 기간을 고려하여 입력 자료를 기간별로 나누고 최적의 입력 자료를 찾고자 하였다. 실험 결과, 5월부터 9월까지의 위성 자료와 16일 평균값을 사용한 기상 자료와의 조합을 사용하였을 경우 평균 연도별 %RMSE, 시군구 %RMSE 각각 7.43%, 7.16%로 가장 좋은 성능을 보였으며 이를 통해 쌀 생산량 추정 분야에 대한 SSAE 알고리즘의 적용 가능성을 확인할 수 있었다.

1. 서 론

쌀 생산량 예측 및 조사는 농가 소득 보전 및 농업 분야 기관에 영향을 주고 수급 조절과 가격 예측 등 정부의 정책 수립과 관련하여 중요한 의미를 갖는다. 이를 위해 국내 통계청에서는 매년 작황조사, 예상 생산량 조사 및 실 수확량 조사를 실시하고 있으며 국내 여러 기관에선 이러한 자료를 활용하여 쌀 생산량 예측을 포함하여 다수의 연구를 진행하고 있다.

작황 추정 분야에서 진행된 선행 연구를 살펴보면 대다수의 연구에서 입력 변수로 기상 자료 및 위성 자료 등 원격탐사를 통해 취득된 자료를 사용하였다. 위성 자료의 경우 식생의 분광특성을 통해 식생관련 정보를 얻을 수 있으며, 비용과 사용성 측면에서 편익을 제공하기 때문에 작황 추정에 널리 사용되어 왔다(Nuarsa *et al.*, 2011; Fang *et al.*, 2011). Alvarez (2009)는 월별 강수량 및 월별 평균 온도 자료를 통해 아르헨티나의 팜파스(Pampas) 지역 내 밀 생산량 추정을 하였고, 추정값과 생산량 간 결정계수(coefficient of determination)로 0.761을 나타냈다. 그리고 Prasad *et al.* (2006)에선 온도, 강수량 자료와 NDVI(Normalized Difference Vegetation Index) 및 토양 습윤 정보를 입력 자료로 사용하여 미국 아이오와 주를 대상으로 곡물 생산량 추정 연구를 진행하였으며 옥수수와 콩의 경우 각각 결정계수로 0.78과 0.86의 결과를 보였다. Jaikla *et al.* (2008)은 최고 온도(Maximum temperature; Max), 최저 온도(Minimum temperature; Min), 일사량(Solar radiation; Sol) 및 강수량(Precipitation; Pre)을 통해 토양 내 질소, 줄기 무게(stem weight) 및 낟알 무게(grain weight)를 추정하는 연구를 진행하였다. 국내의 경우, Hong *et al.* (2012)은 MODIS(MODerate Resolution Imaging Spectroradiometer) NDVI와 두 가지 종류의 일조시간 및 일사량 자료를 변수로 사용하여 국내 쌀 생산량을 추정하였으며 통계청과의 비교 결과 1.1 kg/10a 및 14.1 kg/10a 차이를 나타내었다. Na *et al.* (2013)에서도 국내 벼 수량 추정 연구를 진행하였으며 해당 연구에서는 일별 일사량 자료와 MODIS 표면 반사율(surface reflectance), 지표면 온도(land surface temperature) 및 FPAR(Fraction of the absorbed Photosynthetically Active Radiation) 자료를 사용하여 NPP(Net Primary Productivity)를 계산하고 이와 벼 수량과의 상관성 분석을 통해 2012년 벼 수량을 대상으로 RMSE(Root Mean Square Error) 9.46 kg/10a를 보였다.

작황 추정 모델의 종류를 살펴보면 과거로부터 경험적 통계 모형을 주로 사용하였다(Hong *et al.*, 2012). 하지만, 재배 지역, 관개 시설, 날씨를 포함하여 곡물 생산량에 영향을 미치는 요인들이 다양하게 존재하는 가운데 이러한 요인들과 생산량과의 비선형(nonlinear) 관계를 설명하는데 한계가 있으며 이를 해결하기 위해 새로운 알고리즘의 필요성이 대두되었다. 이로 인해 입력 자료와 출력 자료 간 두 개 이상의 은닉 층(hidden layer)을 통해 비선형 문제를 해결할 수 있는 인공신경망(artificial neural network) 모델이 적용되었다(Irmak *et al.*, 2006; Uno *et al.*, 2005). Li *et al.* (2007)과 Ji *et al.* (2007)에서는 각각 위성자료와 기상자료를 입력 자료를 통해 인공신경망과 다중 회귀 분석(multiple regression analysis) 모델을 구축하여 비교 및 평가하였으며 두 연구에서 모두 인공신경망이 더 우수한 성능을 보임을 확인하였다.

현재 인공신경망 모델을 기반으로 개발된 딥 러닝

(deep learning) 알고리즘이 뛰어난 성능을 보이며 패턴 인식(pattern recognition), 컴퓨터 비전(computer vision) 등의 다양한 분야에서 널리 사용하고 있다. 최근 딥 러닝 알고리즘 중 SSAE(Stacked Sparse AutoEncoder) 알고리즘이 시계열 자료를 통한 예측 분야에서 성공적인 결과를 보이고 있는데 Khodayar and Teshnehlab (2015)에서는 풍속 예측(wind speed prediction) 분야에 처음으로 SAE(Stacked AutoEncoder)을 적용하였고 인공신경망과의 성능 비교를 통해 SAE가 인공신경망보다 더 좋은 성능을 보임과 동시에 일반화 능력이 뛰어나다는 결과를 보였다. 또한 기후 예측(weather forecasting), 교통량 흐름 예측(traffic flow prediction) 및 에너지 부하(energy load) 분야에서도 SAE 알고리즘의 적용 가능성을 확인할 수 있다(Busseti *et al.*, 2012; Liu *et al.*, 2014; Lv *et al.*, 2015).

본 연구에서는 딥 러닝 알고리즘 중 SSAE 알고리즘을 적용하여 작황 추정 모델을 구축하고 2003~2016년도에 해당하는 기상 자료 및 위성 자료를 입력 자료로 사용하여 남한 쌀 생산량 추정 연구를 통해 작황 추정 분야에서 SSAE 알고리즘의 적용 가능성을 평가 및 확인하고자 한다.

2. 연구 대상 지역 및 자료 수집

1) 연구 대상 지역

남한 지역은 동아시아에 북위 33~38°, 동경 126~131°에 위치하고 있으며 약 100,210 km^{2}의 면적을 차지한다. 행정적으로는 서울특별시와 6개의 광역시, 8개의 도, 1개의 특별시, 1개의 특별자치시 및 1개의 특별자치도로 구성되며, 본 연구에서는 Fig. 1과 같이 총 162개의 시군구에 대해 쌀 생산량 추정 연구를 진행하였다. 남한은 사계절이 뚜렷하며 1981년~2010년 평년값 기준 연평균기온은 10~15°C를 나타내고 가장 더운 8월엔 23~26°, 가장 추운 1월엔 -6~-3°C를 보인다(KMA, 2017). 여름엔 주로 장마 기간을 포함하며 연 강수량의 50~60%가 여름에 나타난다.

2) 기상 자료 및 위성 자료

본 연구에서 사용된 일별 최고, 최저온도 및 평균 온도(Average temperature; Avg), 일사량 및 강수량 자료('03~'16)는 기상청에서 제공하는 89개 지점의 관측소 및 599개 지점의 방재기상관측장비에서 취득되었으며, Cressman 기법을 사용하여 3km의 격자간격을 가지도록 객관 분석이 수행되었다(Ahn *et al.*, 2010). 위성자료는 MODIS에서 제공하는 자료 중 Terra와 Aqua 위성으로부터 제공되는 MOD13 및 MYD13 NDVI 및 EVI(Enhanced Vegetation Index)와 MOD15 및 MYD15의 LAI(Leaf Area Index) 및 FPAR(Fraction of absorbed Photosynthetically Active Radiation) 자료를 사용하였다. 또한, 해당 자료를 본 연구에서 사용하기 위해 WGS-84 Sinusodial 좌표계를 UTM Zone 52N 좌표계로 변환하고 보정계수를 적용하였으며 이에 대한 자세한 설명은 Hong *et al.* (2012)와 Na *et al.* (2013)에 나와 있다. NDVI와 EVI의 시간 해상도 및 공간 해상도는 각각 16일, 250 m이며, LAI와 FPAR의 경우엔 8일, 1 km의 해상도를 가진다.

3) 쌀 생산량 통계 자료 및 논 마스크 영상

2003년부터 2016년까지의 쌀 생산량 통계 자료는 통계청 웹사이트에서 수집가능하며(KOSIS, 2017), 이는 남한 시군별 논벼의 10a 당 생산량 정보를 포함한다. 이후, 2003년부터 2016년까지의 쌀 생산량 통계값을 모두 포함하고 있는 시군구만을 추출하였으며 최종적으로 본 연구에서는 152개의 시군구에 대해 쌀 생산량을 추정하고자 한다. 또한, 벼논 지역을 대상으로 기상 자료 및 위성자료 값을 추출하기 위해 논 마스크 영상을 사용하였으며(Fig. 1) 이를 제작하기 위해 2011년 상반기

Fig. 1. The 162 SiGunGu and rice fields (green dot) in South Korea.

기준의 지적도를 사용하였다(Hong *et al.*, 2012; Na *et al.*, 2013). 이는 벡터 자료이며 벼논 지역만을 추출하여 30 m 격자로 논 마스크를 제작하였고, 남한 전역에 대해 총 11,499,215개의 화소 수를 가진다.

3. 쌀 생산량 추정 모델 입력 자료 전처리

남한 쌀 생산량 추정을 위해 수집된 자료를 가공하였으며 전처리 과정은 Fig. 2와 같다. 먼저 처리 과정을 보다 효율적으로 진행하기 위해 기존의 벼논 마스크 영상을 보유하고 있는 자료 중 가장 높은 해상도를 갖고 있는 NDVI 및 EVI의 250m로 재추출(Resample) 하였다. 이후 래스터(raster) 형태의 자료를 벡터 형태로 변환하였으며, 쌀 생산량 정보를 포함하고 있는 남한 시군구 쉐이프(Shape) 자료와의 공간 결합(spatial join)을 통해 각 벼논 지역에 시군구 번호를 부여하였다. 이를 통해 남한 전역에 대해 총 263,538개의 쌀 생산 표본점 자료가 구축되었으며 이는 시군구 번호와 경위도 좌표를 포함한다. 이후, 아핀변환(affine transformation)을 활용하여 각 표본점에 대한 기상 자료 및 위성 자료 값을 추출하였다.

남한에서 주로 재배되는 쌀의 생태형은 중만생종이며(Kim *et al.*, 2013; Yun, 2003), 이앙 기간은 약 4월 20일부터 6월 20일까지 진행되고 수확 시기는 약 8월 25일부터 9월 25일까지를 나타낸다. 하지만 2003년부터 2016년까지 매년 동일한 이식 및 수확 시기를 결정할 수 없으며 쌀 생산량 추정 모델에 가장 적합한 입력 자료를 선정하기 위해 기상 자료 및 위성 자료를 대상으로 다양한 기간을 고려하였다. 또한, 선행 연구를 살펴보면 기상 자료를 대상으로 일별 자료를 사용하기 보단 7일, 8일, 10일 또는 15일 등의 평균값을 곡물 생산량 추정에 사용한 사례들을 확인할 수 있었는데(Kaul *et al.*, 2005; Na *et al.*, 2013; Ji *et al.*, 2007; USDA, 2017) 본 연구에서는 위성 자료의 NDVI, EVI 및 LAI, FPAR의 시간 해상도와 동일한, 기상 자료의 8일 및 16일 평균값을 추출하였으며 앞서 언급한 특정 기간에 해당하는 자료의 총 갯수에 대해 8일 또는 16일 마다의 평균값을 구하였다. 이 과정에선 쌀 생산 표본점 자료의 시군구 번호에 맞춰 평균값을 계산하였으며 앞의 모든 과정이 진행된 이후엔 쌀 생산량 추정 모델의 입력 변수로 적용하기 위해 값들을 0과 1 사이의 값을 가지도록 하는 정규화(normalization) 과정을 진행하였다. 최종적으로 본 연구에서 남한 쌀 생산량 추정 모델을 위해 제작한 입력 자료는 Table 1과 같다.

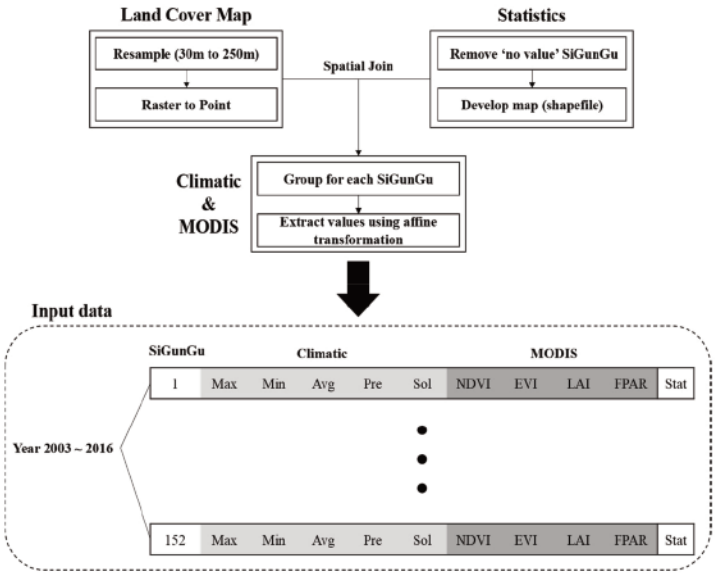

Fig. 2. Flowchart for data preprocessing.

Table 1. The number of scenes of input data per year for rice yield estimation model

Period	Temporal resolution	Product									Total
		Climatic					MODIS				
		Max	Min	Avg	Pre	Sol	NDVI	EVI	LAI	FPAR	
April ~ September	8 day	22	22	22	22	22	11	11	22	22	176
	16 day	11	11	11	11	11	11	11	22	22	121
April ~ August	8 day	18	18	18	18	18	9	9	18	18	144
	16 day	9	9	9	9	9	9	9	18	18	99
May ~ September	8 day	18	18	18	18	18	9	9	18	18	144
	16 day	9	9	9	9	9	9	9	18	18	99
May ~ August	8 day	14	14	14	14	14	7	7	14	14	112
	16 day	7	7	7	7	7	7	7	14	14	77
June ~ September	8 day	14	14	14	14	14	7	7	14	14	112
	16 day	7	7	7	7	7	7	7	14	14	77
June ~ August	8 day	10	10	10	10	10	5	5	10	10	80
	16 day	5	5	5	5	5	5	5	10	10	55

4. 쌀 생산량 추정 모델 및 정확도 평가 방법

1) Stacked Sparse AutoEncoder (SSAE)

남한 전역 쌀 생산 추정 모델을 위해 SAE를 쌓아 올려 구축한 SSAE 알고리즘을 적용하였다. 본 연구에서 사용된 모델은 Fig. 3과 같으며 이는 먼저 주어진 입력 자료를 대상으로 무감독 학습으로 AE(AutoEncoder) 모델을 훈련시킨 뒤 동일한 과정을 한 번 더 진행한 이후 마지막엔 선형 회귀층(linear regression layer)을 통해 각 표본점에 대한 쌀 생산량을 추정한다. 전형적인 AE는 은닉층을 통해 입력 자료와 동일한 값을 표현하도록 학습한다. 이는 항등함수와 동일한 역할을 하는 것으로 간주될 수 있지만, 은닉 뉴런(hidden neuron)의 개수를 제한하거나 다양한 모델 내 하이퍼 파라미터(hyperparameter)를 조정하는 등의 제약 조건을 지정함으로써 입력 자료에 대해 흥미로운 구조를 발견할 수 있다(Ranzato *et al.*, 2007). 이와 비슷하게 은닉 뉴런의 수가 입력 자료의 크기보다 큰 경우에도 은닉 뉴런에 희박성(sparsity) 제한을 두는 방법으로 또한 자료의 특징을 발견할 수 있으며 이 경우의 모델을 SAE로 표현한다. SAE에서 사용되는 비용 함수는 다음과 같이 계산한다(Ng *et al.*, 2011).

$$J_{sparse}(\mathrm{W,b}) = \mathrm{J(W,b)} + \beta\sum_{j=1}^{S_2} KL(\rho\|(\hat{\rho}_j) \qquad (1)$$

$$\mathrm{J(W,b)} = \left[\frac{1}{m}\sum_{i=1}^{m}\left(\frac{1}{2}\|h_{W,b}(x^{(i)}) - y^{(i)}\|^2\right)\right] + \frac{\lambda}{2}\sum_{l=1}^{n_l-1}\sum_{i=1}^{s_l}\sum_{j=1}^{s_l+1}(W_{ji}^{(l)})^2$$

여기서 J(W,b)의 첫 번째 항은 오차의 평균 제곱합을 의미하며 m은 입력 자료의 크기, W는 가중치(weight), b는 편차(bias), x와 y는 입력 자료를 나타내며 $h_{W,b}$는 가중치와 편차를 변수로 갖는 활성 함수를 나타낸다. 본 연구에서는 활성 함수로 시그모이드(sigmoid) 함수를 사용하였다. J(W,b)의 두 번째 항목은 가중치 조절(weight decay) 항을 의미하며 n_l은 출력층, s_l은 l번째 층의 은닉 뉴런 수를 나타낸다. 여기서 λ는 가중치 조절 파라미터(weight decay parameter)로 가중치의 크기를 감소시킴으로써 두 항목간의 상대적 중요도를 결정하고 과적합(overfitting) 현상을 방지하는데 도움을 준다. J_{sparse} (W,b)의 $KL(\rho\|(\hat{\rho}_j)$항은 각각 ρ와 $\hat{\rho}_j$를 평균으로 갖는 두 베르누이 확률 변수 분포의 차이를 나타낸다. 여기서 ρ는 희박성 매개변수(sparsity parameter)로 모델의 sparsity 정도를 결정하며 이는 특정 뉴런에서의 활성 정도에 제한을 두어 은닉 뉴런의 수가 입력 자료의 크기보다 클 경우에도 입력 자료의 특징을 표현하는 데 도움을 준다. β는 희박성 제한 항(sparsity penalty term)의 가중치를 의미한다. 본 연구에서는 남한 쌀 생산량 추정을 위한 SSAE 모델의 최적 파라미터 조합을 찾기 위해 Table 2과 같이 파라미터 값들을 실험해 보았으며, 실험 결과 희박성, 가중치 조절 및 희박성 제한 매개변수 값으로 0.6, 0.1, 10을 사용하기로 결정하였다. 은닉 뉴런 수의 경우 Table 1에 명시한 12개의 입력 자료를 대상으로 각 은닉층에 대해 100, 150, 200개를 지정하였으며 총 9개의

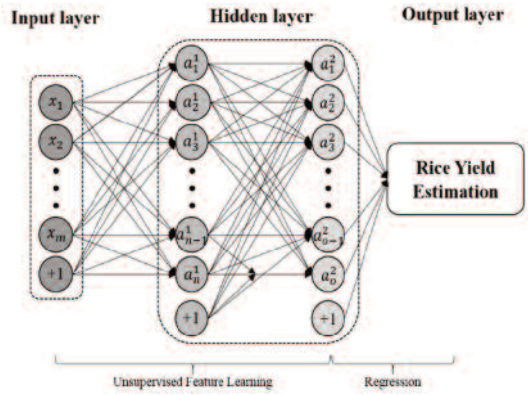

Fig. 3. SSAE model for the rice yield estimation in South Korea.

Table 2. The testing parameter set for SSAE model

Parameters	Values
Sparsity parameter (ρ)	0.1, 0.2, 0.3, 0.4, 0.5, 0.6, 0.7, 0.8, 0.9
Weight-decay parameter (λ)	0.001, 0.01, 0.1, 1, 10, 100
Sparsity penalty parameter (β)	1, 10, 20, 30, 40, 50, 60, 70, 80, 90, 100

모델을 구축하여 제일 높은 성능을 보이는 입력 자료 조합을 찾아보았다.

2) 정확도 평가

남한 쌀 생산량 추정 모델의 정확도를 평가하기 위해 RMSE(Root Mean Square Error)와 %RMSE를 사용하였으며 시공간적 분석을 위해 연도별 정확도와 시군구별 정확도를 계산하였고 이는 다음과 같다.

$$\mathrm{RMSE} = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(\hat{r}_i - r_i)^2}$$
$$\mathrm{RMSE\%} = \frac{RMSE}{\bar{r}} \times 100\% \qquad (2)$$

여기서 먼저 연도별 정확도의 경우 $\hat{r}_i$은 i번째 시군구의 쌀 생산 추정값, r_i은 i번째 시군구의 쌀 생산 통계값. $\bar{r}$은 전체 152개 시군구의 쌀 평균 통계값을 나타내고 n은 전체 시군구 수(152개)를 의미한다. 시군구별 정확도의 경우 $\hat{r}_i$, r_i는 위와 동일하며, $\bar{r}$는 i번째 시군구의 2003~2016년도 쌀 평균 통계값을 나타내고 n은 전체 연도 수(14개년)를 의미한다.

본 연구에서는 교차 검증(cross validation)을 실시하였으며 보유하고 있는 2003년부터 2016년까지 총 14개 년도에 대해 13개 년도를 훈련 자료로, 나머지 한 개 년도를 검증 자료로 지정하여 모델을 총 14번 구동하여 정확도 평가를 실시하였다.

5. 결과 분석

1) 남한 쌀 생산량 추정을 위한 최적 입력 자료 선정

본 연구에서는 남한 전역을 대상으로 쌀 생산량 추정을 실시하였으며, 추정 모델을 구축하기 위해 SSAE 알고리즘을 적용하였다. 남한 벼의 생육 기간을 고려하여 입력 자료는 14개를 구축하였으며(Table 1), 두 개의 은닉층에 대한 은닉 뉴런 수를 100, 150, 200개씩 지정하여

Table 3. The comparison result of SSAE model for each input data

Period	Temporal resolution	Average Accuracy		Hidden neuron #	
		Year %RMSE	SiGunGu %RMSE	Hidden layer 1	Hidden layer 2
April ~ September	8 day	8.39%	8.18%	100	200
	16 day	7.71%	7.51%	200	100
April ~ August	8 day	8.46%	8.39%	150	150
	16 day	7.76%	7.60%	200	150
May ~ September	8 day	7.74%	7.52%	150	200
	16 day	7.43%	7.16%	150	100
May ~ August	8 day	7.81%	7.57%	100	200
	16 day	7.46%	7.22%	200	100
June ~ September	8 day	7.71%	7.44%	200	100
	16 day	7.64%	7.31%	200	200
June ~ August	8 day	8.39%	8.18%	100	200
	16 day	7.71%	7.40%	150	100

총 9개의 SSAE 모델을 생산하였고 정확도 평가 결과 가장 좋은 성능을 보이는 경우를 Table 3에 정리하였다.

실험 결과, 5월부터 9월까지의 위성 자료와 해당 기간 동안의 16일 평균값을 사용한 기상 자료와의 조합에 대한 150, 100개의 은닉 뉴런을 사용한 SSAE 모델이 평균 연도별 %RMSE, 시군구 %RMSE 각각 7.43%, 7.16%를 보이며 가장 좋은 성능을 보였다. 생육 기간을 고려하였을 때, 9월까지의 기상 자료 및 위성 자료를 사용하는 것이 8월까지의 자료를 사용하는 것보다 더 좋은 결과를 나타내었으며 모든 기간에 대해 16일 평균값을 사용한 경우가 8일 평균값을 사용하였을 때보다 더 좋은 정확도를 나타내었다.

2) 남한 쌀 생산량 추정 결과에 대한 시공간적 분석

앞서 선정된 입력 자료를 통해 시공간적 분석을 수행하였으며 2003년부터 2016년까지의 연도별 %RMSE는 Fig. 4와 같다.

Fig. 4를 보면 전체 14개년도 중 2003, 2007, 2012 그리고 2016년에 SSAE 모델의 추정 정확도가 타 연도에 비해 비교적 낮게 나옴을 확인할 수 있다. 여러 변수들 중 예외적인(anomal) 날씨, 재해 또는 병충해 등이 당해년도 쌀 생산량에 악영향을 끼치는 가운데 실제로 2003년, 2012년에 태풍이 발생하여 곡물 수확에 심각한 영향을 주었으며, 2007년에는 벼 줄무늬잎마름병(rice stripe virus)으로 인해 전국적으로 큰 피해가 있었음을 확인할 수 있었다(Kim *et al.*, 2008). 2016년의 경우, 2015년과 비교하였을 때 기상 자료와 위성 자료의 값이 상대적으로 낮게 나왔음에도 불구하고 연도별 평균 쌀 생산량은 2016년과 2015년 각각 529kg/10a, 532kg/10a로 서로 근사한 결과를 보여 이는 본 연구에서 사용한 입력 자료 외의 변수가 작용하여 이러한 결과를 나타낸 것으로 보이며 이에 대한 분석이 필요하다.

Fig. 4. The rice yield estimation result from year 2003 to 2016.

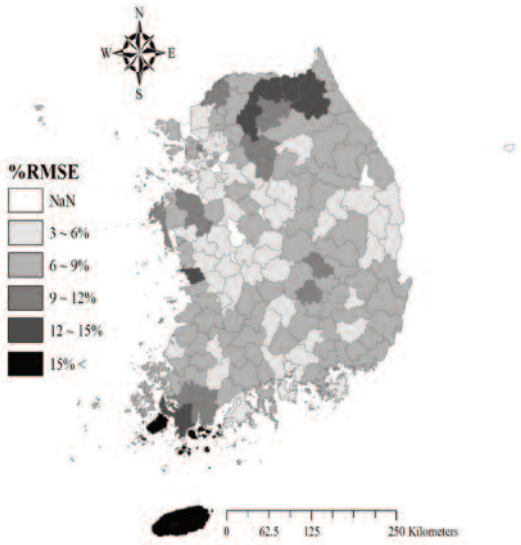

Fig. 5. The average SiGunGu %RMSE from year 2003 to 2016.

SSAE 모델을 통한 쌀 생산량 추정 결과를 공간적으로 분석하기 위해 시군구별 정확도를 지도로 표현하였으며 이는 Fig. 5와 같다. 결과를 보면, 대부분 지역에서의 쌀 생산량 추정 결과가 15%보다 좋게 나온 반면 제주도와 전라도의 남서부 지역인 진도군, 완도군의 경우 각각 16.90%, 20.30%, 16.22%로 상당히 안 좋은 정확도를 보임을 확인할 수 있었다. 그 이유로 제주도, 해남군, 진도군의 경우 해안가에 위치하고 있으며 이에 따라 주요 재배지와는 다른 품종이 재배되거나 이앙시기가 상이하여 이러한 결과를 나타내는 것으로 보인다.

5. 결 론

본 연구에서는 쌀 생산량 추정 분야에 대한 딥러닝 알고리즘의 성능을 판단하였으며 이 과정에서 SSAE 알고리즘을 적용하여 남한 전역을 대상으로 쌀 생산량 추정 연구를 진행하였다. 이 연구에서 구축한 쌀 생산량 추정 모델은 SAE 모델을 무감독 학습으로 훈련시킨 뒤 동일한 과정을 한 번 더 진행한 이후 마지막에 선형 회귀층을 통해 각 표본점에 대한 쌀 생산량을 추정한다. 은닉 뉴런의 수가 입력 자료의 크기보다 큰 경우에도 여러 매개변수들 통해 제한을 둘 경우 입력 자료에 대한 흥미로운 구조 및 특징을 발견할 수 있으며 매개변수들

의 조합을 변경하면서 모델의 성능을 평가하였고 남한 쌀 생산량 추정을 위한 SSAE 모델의 최적의 매개변수를 발견하였다. SSAE를 통해 구축된 쌀 생산량 추정 모델의 입력 변수로 최고, 최소 및 평균 온도, 강수량, 일사량의 기상자료와 NDVI, EVI, LAI 및 FPAR의 위성자료를 사용하였다. 이 과정에선 남한 벼의 생육 기간을 고려하여 입력 자료의 기간을 4~9월, 4~8월, 5~9월, 5~8월, 6~9월, 6~8월로 나누었으며 기상 자료의 경우 선행 연구를 통해 8일과 16일 평균값을 추출하여 입력 자료를 구축하였다. 정확도 평가로는 RMSE 및 %RMSE를 통해 연도별 정확도와 시군구별 정확도를 계산하였다. 연구 결과, 5월부터 9월까지의 위성 자료와 해당 기간 동안의 16일 평균값을 사용한 기상 자료와의 조합을 사용하였을 경우 평균 연도별 %RMSE, 시군구 %RMSE 각각 7.43%, 7.16%를 보이며 가장 좋은 성능을 보였으며 SSAE 알고리즘의 적용 가능성을 확인하였다. 이후, 앞서 선택된 입력 자료를 통해 남한 전역에 대한 시공간적 분석을 진행하였다. 연도별 %RMSE의 경우 2003년, 2007년, 2012년 및 2016년도와 시군구별 %RMSE의 경우 제주도, 진도군 및 완도군에서 타 경우보다 상대적으로 낮은 정확도가 나타났고 이는 자연 재해 및 질병과 토양 특성 및 관개 시설과 같은 생육 환경의 차이로 인해 이러한 결과를 나타낸 것으로 확인할 수 있었다.

하지만 공간적 분석 과정에서도 발견되었듯이 본 연구에서 사용된 입력 자료인 기상 자료 및 위성 자료 외 쌀 생육 과정에 영향을 주는 타 변수들을 고려할 필요가 있으며, 공간적 단위를 시군구와 더불어 농업기후지역을 기준으로 세분화하여 정확도를 평가할 필요가 있다(Na *et al.*, 2012; Yoo *et al.*, 2013). 또한, 딥러닝 알고리즘 중 SSAE 알고리즘 이외에 회선신경망(convolutional neural network) 및 순환신경망(recurrent neural network) 등의 알고리즘들의 성능을 비교 및 평가하는 연구가 향후 요구된다.

감사의 글

본 논문은 농촌진흥청 공동연구사업(과제번호: PJ009978)의 지원을 받았으며, 이에 감사드립니다.

References

Alvarez, R., 2009. Predicting average regional yield and production of wheat in the Argentine Pampas by an artificial neural network approach, *European Journal of Agronomy*, 30(2): 70-77.

Ahn, J., J. Hur, and K. Shim, 2010. A Simulation of Agro-Climate Index over the Korean Peninsula Using Dynamical Downscaling with a Numerical Weather Prediction Model, *Korean Journal of Agricultural and Forest Meteorology*, 12(1): 1-10 (in Korean with English abstract).

Busseti, E., I. Osband, and S. Wong, 2012. Deep learning for time series modeling, *Technical report*, Stanford University, Stanford, CA, USA.

Fang, H., S. Liang, and G. Hoogenboom, 2011. Integration of MODIS LAI and vegetation index products with the CSM CERES Maize model for corn yield estimation, *International Journal of Remote Sensing*, 32(4): 1039-1065.

Hong, S. Y., J. Hur, J. B. Ahn, J. M. Lee, B. K. Min, C. K. Lee, … and K. M. Shim, 2012. Estimating rice yield using MODIS NDVI and meteorological data in Korea, *Korean Journal of Remote Sensing*, 28(5): 509-520 (in Korean with English abstract).

Irmak, A., J. Jones, W. Batchelor, S. Irmak, K. Boote, and J. Paz, 2006. Artificial neural network model as a data analysis tool in precision farming, *Transactions of the American Society of Agricultural and Biological Engineeres*, 49(6): 2027-2037.

Jaikla, R., S. Auephanwiriyakul, and A. Jintrawet, 2008. Rice yield prediction using a support vector regression method, Proc., Electrical Engineering/Electronics, Computer, Telecommunications and Information Technology, 5th International Conference on, IEEE, Krabi, Thailand, May 14-17, vol. 1, pp. 29-32.

Ji, B., Y. Sun, S. Yang, and J. Wan, 2007. Artificial neural networks for rice yield prediction in mountainous regions, *The Journal of Agricultural*

Science, 145(3): 249-261.

Kaul, M., R. L. Hill, and C. Walthall, 2005. Artificial neural networks for corn and soybean yield prediction, *Agricultural Systems*, 85(1): 1-18.

Khodayar, M. and M. Teshnehlab, 2015. Robust deep neural network for wind speed prediction, Fuzzy and Intelligent Systems (CFIS), 2015 4th Iranian Joint Congress, Zahedan, Iran, Sep. 9-11, pp. 1-5.

Kim, J. S., S. H. Lee, H. S. Choi, G. S. Choi, J. D. Cho, and B. N. Chung, 2008. Survey of viral diseases occurrence on major crops in 2007, *Research in Plant Disease*, 14(14): 1-9 (in Korean with English abstract).

Kim, Y., H. Lee, and S. Hong, 2013. Continuous monitoring of rice growth with a stable ground-based scatterometer system, *IEEE Geoscience and Remote Sensing Letters*, 10(4): 831-835.

KMA, 2017. Korea Meteorological Administration – Domestic climatic information, http://www.kma.go.kr/weather/climate/average_south.jsp, Accessed Jul. 1, 2017.

KOSIS, 2017. Korean Statistic – Administrative District rice yield production, http://kosis.kr/statHtml/statHtml.do?orgId=101&tblId=DT_1ET0034, Accessed Jul. 1, 2017.

Li, A., Liang, S., Wang, A., and J. Qin, 2007. Estimating crop yield from multi-temporal satellite data using multivariate regression and neural network techniques, *Photogrammetric Engineering & Remote Sensing*, 73(10): 1149-1157.

Liu, J. N., Y. Hu, J. J. You, and P. W. Chan, 2014. Deep neural network based feature representation for weather forecasting, Proceedings on the International Conference on Artificial Intelligence (ICAI), The Steering Committee of The World Congress in Computer Science, Computer Engineering and Applied Computing (WorldComp), Las Vegas, NV, Jul. 21-24, p. 1.

Lv, Y., Y. Duan, W. Kang, Z. Li, and F. Y. Wang, 2015. Traffic flow prediction with big data : a deep learning approach, *IEEE Transactions on Intelligent Transportation Systems*, 16(2): 865-873.

Na, S.-I., J. H. Park, and J. K. Park, 2012. Development of Korean Paddy Rice yield Prediction Model (KRPM) using meteorological element and MODIS NDVI, *Journal of the Korean Society of Agricultural Engineers*, 54(3): 141-148 (in Korean with English abstract).

Na, S. I., S. Y. Hong, Y. H. Kim, K. D. Lee, and S. Y. Jang, 2013. Prediction of rice yield in Korea using paddy rice NPP index-Application of MODIS data and CASA model, *Korean Journal of Remote Sensing*, 29(5): 461-476 (in Korean with English abstract).

Ng, A., 2011. Sparse autoencoder. CS294A Lecture notes, Stanford University, Stanford, CA, USA.

Nuarsa, I. W., F. Nishio, and C. Hongo, 2011. Rice yield estimation using Landsat ETM+ data and field observation, *Journal of Agricultural Science*, 4(3): 45-56.

Prasad, A. K., L. Chai, R. P. Singh, and M. Kafatos, 2006. Crop yield estimation model for Iowa using remote sensing and surface parameters, *International Journal of Applied Earth Observation and Geoinformation*, 8(1): 26-33.

Ranzato, M, Huang, F. J., Boureau, Y. L., and Y. LeCun, 2007. Unsupervised learning of invariant feature hierarchies with applications to object recognition, Computer Vision and Pattern Recognition (CVPR), 2007 IEEE Computer Society, Minnesota, USA, Jun.18-23, pp. 1-8.

Yoo, S. H., J. Y. Choi, S. H. Lee, Y. G. Oh, and D. K. Yun, 2013. Climate change impacts on water storage requirements of an agricultural reservoir considering changes in land use and rice growing season in Korea, *Agricultural Water Management*, 117: 43-54.

Uno, Y., S. O. Prasher, R. Lacroix, P. K. Goel, Y. Karimi, A. Viau, and R. M. Patel, 2005. Artificial neural networks to predict corn yield

from Compact Airborne Spectrographic Imager data, *Computers and Electronics in Agriculture*, 47(2): 149-161.

USDA, 2017. Crop Explorer, http://www.pecad.fas.usda.gov/cropexplorer/, Accessed Jul. 1, 2017.

Yun, J. I., 2003. Predicting regional rice production in South Korea using spatial data and crop-growth modeling, *Agricultural Systems*, 77(1): 23-38.