

기계학습을 이용한 돈사 급수량 예측방안 개발

이웅섭¹ · 류종열¹ · 반태원¹ · 김성환^{1*} · 최희철²

Prediction of Water Usage in Pig Farm based on Machine Learning

Woongsup Lee¹ · Jongyeol Ryu¹ · Tae-Won Ban¹ · Seong Hwan Kim^{1*} · Heechul Choi²

^{1*}Department of Information and Communication Engineering, Gyeongsang National University, Tongyeong 53064, Korea

²Livestock Environment Division, National Institute of Animal Science, Korea

요 약

최근 사물 인터넷 센서가 설치된 스마트 돈사의 보급을 통해 돈사 관련 빅데이터 축적이 가능해졌고, 다양한 기계 학습방안들이 수집된 데이터에 적용되어 축산농가의 생산성을 향상시키고 있다. 본 연구에서는 다양한 기계학습 방안을 이용하여 돈사관리에서 가장 중요한 요소 중 하나인 급수량을 예측하였다. 구체적으로 실제 돈사에서 수집된 데이터에 회귀 방안인 선형회귀, 회귀트리 및 아다부스트 회귀 방안과 분류 방안인 로지스틱 분류, 결정트리 및 서포트 벡터 머신 (SVM) 분류방안을 적용하여 돈사의 온도와 습도를 기반으로 급수량을 예측하였다. 성능 분석을 통해서 제안한 방안이 높은 정확도로 급수량을 예측하는 것을 확인할 수 있었다. 제안한 방안은 돈사의 급수시설 이상을 조기에 파악하는데 활용되어 가축을 폐사를 막고 돈사 생산성을 높이는데 활용될 수 있다.

ABSTRACT

Recently, accumulation of data on pig farm is enabled through the wide spread of smart pig farm equipped with Internet-of-Things based sensors, and various machine learning algorithms are applied on the data in order to improve the productivity of pig farm. Herein, multiple machine learning schemes are used to predict the water usage in pig farm which is known to be one of the most important element in pig farm management. Especially, regression algorithms, which are linear regression, regression tree and AdaBoost regression, and classification algorithms which are logistic classification, decision tree and support vector machine, are applied to derive a prediction scheme which forecast the water usage based on the temperature and humidity of pig farm. Through performance evaluation, we find that the water usage can be predicted with high accuracy. The proposed scheme can be used to detect the malfunction of water system which prevents the death of pigs and reduces the loss of pig farm.

키워드 : 기계학습, 급수량, 돈사, 예측

Key word : Machine learning, Water usage, Pig farm, Prediction

Received 04 July 2017, Revised 17 July 2017, Accepted 25 July 2017

* Corresponding Author Seong Hwan Kim (E-mail: seonghwan.kim@gnu.ac.kr, Tel:+82-55-772-9176)

Department of Information and Communication Engineering, Gyeongsang National University, Tongyeong 53064, Korea

Open Access <https://doi.org/10.6109/jkiice.2017.21.8.1560>

print ISSN: 2234-4772 online ISSN: 2288-4165

© 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.
Copyright © The Korea Institute of Information and Communication Engineering.

I. 서 론

최근 농가의 생산성을 높이기 위해 다양한 사물 인터넷 센서들이 설치된 스마트 돈사가 확대 보급되고 있다. 특히 사물인터넷 센서들의 가격이 점차 낮아지면서 스마트 돈사의 보급이 점차 빨라지고 있는 추세이다. 스마트 돈사는 설치되어 있는 다양한 센서들을 이용하여 온도, 습도, 풍속, 이산화탄소, 암모니아 등 다양한 환경요소들을 측정할 수 있을 뿐만 아니라 환경요소를 실시간으로 모니터링 및 수집 할 수 있다. 이를 통해서 돈사환경 및 생체관련 빅데이터구축이 가능해졌고, 구축된 빅데이터에 다양한 통계분석 및 기계학습 분석기법들을 적용하여 농가의 생산성을 높이는 연구가 활발하게 진행되고 있다[1-6].

우선 기계학습을 이용하여 축사의 환경요소와 가축의 생산성간의 관계를 분석하여 축사의 생산성을 높이는 연구가 진행되었다. [1]에서는 스마트 돈사에서 수집된 축사의 환경데이터와 돼지의 체중데이터를 기반으로 기계학습방안중 하나인 의사결정트리기법을 적용하여 돈사의 환경데이터(온도, 습도, 풍속)와 돼지의 일당 증체간의 연관성 모델을 도출하였다. 또한 국내의 대표적인 양돈관리 프로그램인 Pigplan을 통해 수집된 데이터를 기반으로 다양한 기계학습 방안을 이용하여 모돈의 생산성을 예측하는 연구가 [2]에서 진행되었다. 더불어 젖소의 생체데이터와 농장의 환경데이터에 다양한 기계학습기법(결정트리, 랜덤 포레스트, 나이브 베이즈)을 적용한 연구가 미국 위스콘신 대학 연구팀에 의해 이루어졌다[3].

축사에서 환경요소와 생산성간의 관계분석 이외에도 질병개체를 조기에 파악하여 축산농가와 사회에 큰 피해를 미치는 전염성 질병을 조기에 파악하는 연구도 진행되고 있다. [4]에서는 스마트축사에서 수집되는 다양한 종류의 가축 생체데이터를 활용하여 딥러닝기반 정상개체 및 질병개체 분류 방안을 제안하였다. 또한 돼지 개체의 움직임을 실시간으로 모니터링하여 각 객체의 질병여부를 파악하는 연구가 [5]에서 진행되었고, [6]에서는 닭의 움직임 데이터를 수집하고 베이지안 및 회귀 방안을 통해서 문제가 발생한 개체를 파악하는 연구가 진행되었다.

현재 가축 및 양돈분야에서의 기계학습 관련 연구는 주로 환경 데이터를 이용하여 각 개체의 성장을 예측하

거나 질병에 걸린 개체를 파악하는 것에 집중하고 있다. 하지만 실제 양돈 및 축산 농가에서는 각 개체의 성장뿐만 아니라 돈사의 급수량을 예측하는 것도 중요하다. 구체적으로 돈사에서 사용되는 물사용량이 크기 때문에 돈사의 급수량은 축산농가의 소득과 직접적인 연관이 있고, 음수량은 개체의 성장 및 질병과 밀접한 관계가 있다[7].

특히 실제 양돈농가에서는 여름철 돈사의 급수시설이 고장이 나서 돼지들이 집단폐사하는 경우가 자주 발생하고 급수시설의 누수로 인해 큰 경제적 손실을 보는 경우도 발생한다. 따라서 돈사의 급수량을 예측하는 방안의 개발이 필요하다. 하지만 현재 돈사의 급수와 관련된 연구는 주로 폐수처리 시설 쪽에 집중이 되어있고[8] 환경데이터를 기반으로 음수량을 예측하는 방안은 현재 연구되지 않았다. 따라서 본 연구에서는 실제 양돈농가에서 수집된 데이터에 다양한 기계학습 방안들을 적용하여 돈사 급수량을 예측하는 방안을 제안한다.

본 논문의 2장은 연구에서 사용한 데이터 및 분석방안에 대해서 설명하고 3장에서는 기계학습 기반 음수량 예측방안에 대해서 기술하고 제안 방안의 성능을 알아본다. 마지막으로 4장에서는 논문을 마무리 한다.

II. 사용된 데이터 및 분석방안

본 연구에서는 B농장의 육성비육돈사에서 2016년 2월 3일부터 2016년 8월 1일까지의 수집된 환경 데이터를 사용하였다. 음수량 데이터는 10분 간격으로 측정되었고 각 날짜의 최종 음수량만을 사용하였다(Water). 또한 해당 날짜에 수집된 온도와 습도 데이터도 사용하였다. 각 날짜에서 5분 간격으로 수집된 온도와 습도 데이터 중 각 날짜의 최대/최소/평균 데이터만을 사용하였다 (Temp_avg, Temp_max, Temp_min, Hum_avg, Hum_max, Hum_min). 또한 본 연구에서는 회귀방식 (Regression) 기반의 급수량 예측뿐만 아니라 분류방식 (Classification) 기반의 급수량 예측도 진행하였다. 분류방식 기계학습방안에서는 분류기준을 3500L로 설정하여 급수량이 3500L이하일 경우에는 0, 초과일 경우에는 1로 두었다 (Water_bin). 본 연구에서 활용한 데이터의 통계적 특성은 표1에 나타내었다.

Table. 1 Attribute of environmental data in the pig-farm.

Mark	Explanation	Average	Standard deviation	Unit
Temp_avg	Average Temperature	28.35	1.89	°C
Temp_max	Maximum Temperature	29.85	2.26	°C
Temp_min	Minimum Temperature	26.98	1.82	°C
Hum_avg	Average Humidity	58.55	10.54	%
Hum_max	Maximum Humidity	66.89	10.40	%
Hum_min	Minimum Humidity	48.28	12.55	%
Water	Water	4155	704.64	Liter
Water_bin	0: Water < 3500 1: Otherwise	0.78	0.41	Binary

표 1의 환경 데이터에서 확인할 수 있듯이 온도는 평균 28도를 유지하는 것을 볼 수 있고 표준편차가 작으므로 돈사 내 온도변화가 크지 않는 것을 확인할 수 있다. 습도는 평균 60% 정도를 유지하고 온도에 비해서 변화량이 큰 것을 확인할 수 있다.

또한 돈사의 급수량은 평균 4155 리터임을 확인할 수 있고 돈사 물사용량이 큰 것을 확인할 수 있고, 돈사의 급수량이 농가지출의 큰 부분을 차지하고 있음을 유추할 수 있다.

다음으로 온도/습도/급수량의 산점도(Scatter plot)를 그림 1에 표시하였다. 산점도에서는 평균온도/평균습도만을 사용하였고 주대각선 (diagonal)에는 누적분포함수를 표시하였다. 산점도에서 확인할 수 있듯이 평균 온도 및 습도와 돈사급수량이 양의 상관관계(positive correlation)를 가지고 있는 것을 확인할 수 있고 상관도가 큰 것을 확인할 수 있다. 또한 습도와 온도도 양의 상관관계를 가지는 것을 확인할 수 있다.

본 연구에서는 회귀방식(Regression)과 분류방식(Classification)의 기계학습 방안을 이용하여 분석을 진행하였다. 회귀방식 알고리즘에서는 돈사급수량을 정확하게 예측하는 것을 목표로 하였고 분류방식 알고리즘에서는 이진데이터인 Water_bin을 정확하게 예측하는 것을 목표로 하였다.

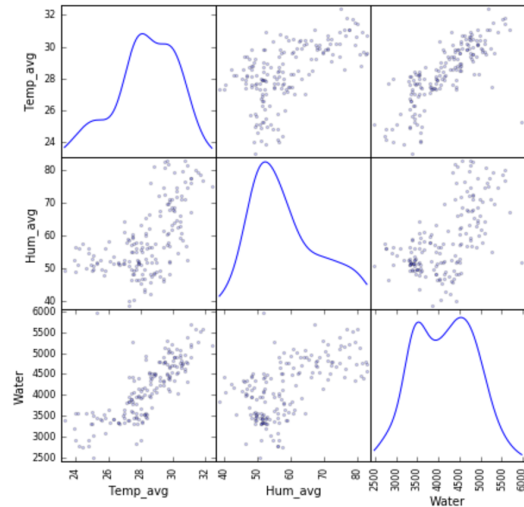


Fig. 1 Scatter plot of Average temperature, average humidity and water usage.

회귀방식 알고리즘으로는 선형회귀 (Linear regression), 회귀트리 (Regression Tree) 및 아다부스트 (AdaBoost regression) 회귀방안을 고려하였다. 선형회귀 방식은 환경데이터의 선형방정식으로 예측 식을 도출하는 방식이다. 선형회귀 방식에서는 \hat{Y} 가 예측 값이고 X_i 가 각 환경요소일 때 계수 W_i 와 바이어스 B 를 구하고 아래의 식과 같이 값을 예측한다. 아래의 수식에서 N 은 전체 feature의 수를 나타낸다.

$$\hat{Y} = \sum_{i=1}^N W_i X_i + B \tag{1}$$

회귀트리는 트리구조를 회귀알고리즘에 사용한 방식이다. 마지막으로 아다부스트는 기계학습메타알고리즘으로 잘못 예측된 샘플을 적응적으로 수정하여 더 좋은 성능을 보이는 앙상블 알고리즘이다. 본 연구에서 고려한 아다부스트 방안에서는 base classifier로 선형회귀트리를 사용하였다.

회귀방안에서는 성능분석 지표로 평균제곱근 오차 (MSE) 및 R2-score를 사용하였다. 실제급수량이 Y 이고 예측된 급수량이 \hat{Y} 일 때 평균 제곱근 오차는 다음과 같이 계산할 수 있다. 아래의 식에서 m 은 테스트 데이터의 샘플수를 나타낸다. 예측방안의 정확도가 높을수록 평균 제곱근 오차는 0에 가까워진다.

$$\frac{1}{m} \sum_{j=1}^m |Y - \hat{Y}|^2 \quad (2)$$

또한 R2-score는 다음과 같이 계산할 수 있다. 예측방안의 정확도가 높을수록 R2-score는 1에 근접하게 된다.

$$1 - \frac{\sum_{j=1}^m |Y - \hat{Y}|^2}{\sum_{j=1}^m |Y - E[Y]|^2} \quad (3)$$

분류방식 알고리즘으로는 로지스틱 분류, 결정트리 및 서포트 벡터 머신 (SVM) 방안을 사용하였다. 로지스틱 분류방식은 아래와 같은 로지스틱 함수를 이용하여 0/1의 바이너리 결과를 예측한다.

$$\frac{1}{1 + e^{-\left(\sum_i w_i X_i + B\right)}} \quad (4)$$

결정트리는 트리알고리즘을 분류에 사용한 방식이고 서포트 벡터 머신 방안에서는 선형커널을 사용하여 분류를 하도록 학습시켰다. 분류 방안에서는 성능분석 지표로 정밀도 (precision)와 재현율 (recall) 및 이 두 값을 평균한 F1-score를 사용하였다.

III. 결과분석

본 장에서는 기계학습 방안을 이용하여 개발된 돈사 급수량 예측방안의 성능을 분석한다. 전체 데이터 중 90%는 학습에 사용하고 나머지 10%의 데이터를 이용하여 성능을 분석하였고, 정확도 분석에는 10-Fold 교차검증을 사용하였다. 결과분석에서는 총 173개의 인스턴스를 사용하였다. 돈사급수량을 예측한 기존의 연구가 없으므로 본 연구에서 제안한 방안들의 성능만 분석하였다.

우선 회귀방식의 성능분석을 진행하였다. 선형회귀방식의 경우 급수량이 $-45 \cdot \text{Temp_avg} + 165 \cdot \text{Temp_max} + 75 \cdot \text{Temp_min} - 48 \cdot \text{Hum_avg} + 51 \cdot \text{Hum_max} + 13 \cdot \text{Hum_min} - 2731$ 으로 예측이 되었다. 선형 회귀식

을 통해 돈사급수량이 일최대온도에 가장 큰 영향을 받는 것을 확인할 수 있다. 회귀트리방안의 경우 아래의 그림 2와 같은 트리구조를 구할 수 있었다. 그림 2의 트리구조에서 일최대온도가 급수량에 가장 큰 영향을 주는 것을 확인할 수 있었다. 아다부스트 회귀방식의 경우 시각화가 어려우므로 성능만을 나타내었다.

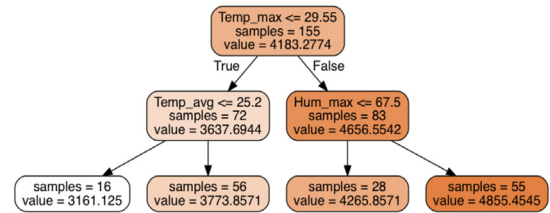


Fig. 2 Regression tree for water usage prediction in pig farm.

선형회귀 방식의 성능을 표2에 나타냈다. 표에서 확인할 수 있듯이 아다부스트 회귀방식이 가장 좋은 성능지표를 보이는 것을 확인할 수 있고 모든 방안에서 높은 정확도로 물사용량을 예측하는 것을 확인할 수 있다.

Table. 2 Accuracy of regression algorithms for water usage prediction of pig farm.

Algorithm	MSE	R2-score
Linear regression	388.21	0.29
Regression tree	470.56	0.42
Adaboost regression	367.84	0.49

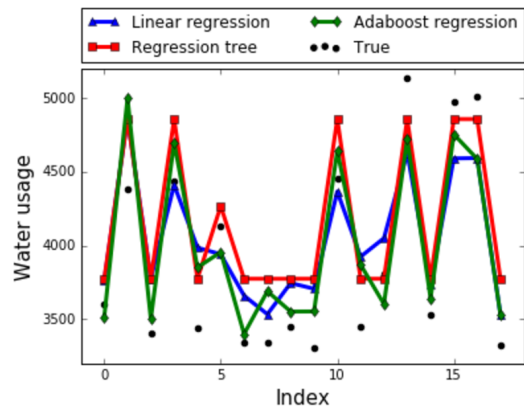


Fig. 3 Performance comparison of regression algorithms for water usage prediction in pig farm.

개발된 예측방안의 성능을 좀 더 자세히 살펴보기 위해 실제 돈사급수량과 예측된 급수량을 비교하였다. 성능비교는 그림 3에 표시하였다. 그림 3에서 가로축은 테스트에 사용된 인스턴스들의 인덱스를 의미한다. 성능분석에서 확인할 수 있듯이 본 연구에서 사용한 3가지 방식 모두 실제 돈사급수량을 잘 예측하는 것을 확인할 수 있다.

다음으로 분류방안의 성능분석을 진행하였다. 로지스틱 분류방안에서 평균온도, 최대온도, 최소온도, 평균습도, 최대습도, 최소습도에 대응하는 계수는 각각 0.38, 0.66, -0.25, -0.31, 0.2, 0.11 로 계산되었고 이를 통해서 돈사급수량에 가장 큰 영향을 주는 환경요소는 돈사최대온도임을 확인할 수 있다. 결정트리방안의 경우 아래의 그림과 같은 트리구조를 구할 수 있었다. 그림 4의 트리구조에서 확인할 수 있듯이 일최대온도가 가장 큰 영향을 주는 것을 확인할 수 있었다. 서포트 벡터 머신 분류방식의 경우 시각화가 어려우므로 성능만을 나타내었다.

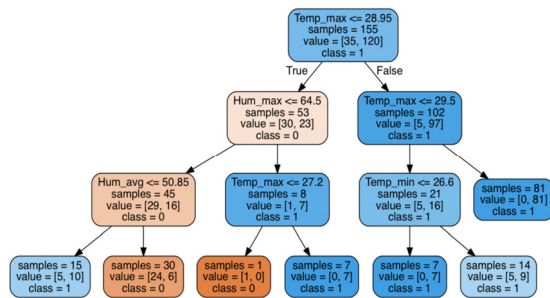


Fig. 4 Decision tree for water usage prediction in pig farm.

분류 방식들의 성능을 표 3에 나타내었다. 표에서 확인할 수 있듯이 로지스틱 분류방식이 가장 좋은 성능을 보이는 것을 확인할 수 있다.

Table. 3 Accuracy of classification algorithms for water usage prediction of pig farm.

Algorithm	Precision	Recall	F1-score
Logistic classification	0.85	0.94	0.88
Decision tree	0.53	0.7	0.60
Linear SVM	0.85	0.87	0.85

특히 88% 이상의 높은 정확도로 돈사급수량을 예측하는 것을 확인할 수 있고 이를 통해 제안방안이 실제 양돈농가에서 효율적으로 활용될 수 있을 것으로 예측할 수 있다.

마지막으로 개발된 예측방안의 성능을 좀 더 자세히 살펴보기 위해 실제 돈사급수량(이진데이터)과 예측된 이진 데이터값을 비교하였다. 성능비교는 그림 5에 나타났다. 그림 5에서 가로축은 테스트에 사용된 인스턴스들의 인덱스를 의미한다. 성능분석에서 확인할 수 있듯이 로지스틱 분류방식과 결정트리 방식은 실제 돈사급수량을 잘 예측하는 것을 확인할 수 있다. 다만 서포트 벡터 머신기반의 분류방안은 모든 경우 1로 예측하는 것을 확인할 수 있다. 이를 통해서 선형 서포트 벡터 머신보다는 로지스틱 분류나 결정트리 방식이 좀 더 좋은 성능을 보이는 것을 확인할 수 있다.

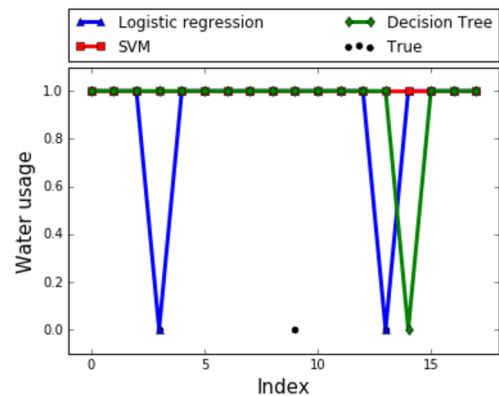


Fig. 5 Performance comparison of classification algorithms for water usage prediction in pig farm.

IV. 결 론

본 연구에서는 스마트돈사에서 수집된 온도 및 습도 데이터를 기반으로 다양한 기계학습 방안을 이용하여 돈사의 급수량을 예측하는 방안을 제안하였다. 선형 회귀, 회귀 트리 및 아다부스트 회귀방식을 이용하여 급수량을 예측하는 방안과 로지스틱 분류, 결정트리 및 서포트 벡터 머신을 이용하여 이진 급수량 데이터를 예측하는 방식을 고려하였다. 성능분석을 통해 제안방안이 높은 정확도로 급수량을 예측하는 것을 확인하였다.

제안 방안의 사용을 통해서 양돈농가에 큰 영향을 미치는 급수시설의 문제를 조기에 파악가능하고 이를 통해 양돈의 생산성을 향상시킬 수 있다.

ACKNOWLEDGMENTS

This work was carried out with the support of “Cooperative Research Program for Agriculture Science & Technology Development (Project title: The Study of ICT Standard Setup and Big-data Utilization in Livestock Field, and Realtime Cow Health Monitoring Module, Project No. PJ012299)” Rural Development Administration, Republic of Korea.

REFERENCES

- [1] K. Han, W. Lee, and K. Sung, “Development of a model to analyze the relationship between smart pig-farm environmental data and daily weight increase based on decision tree,” *Journal of Korea Institute of information and communication engineering*, vol. 20, no. 12, pp.2348-2354, Dec. 2016.
- [2] M. S. Lee and Y.C. Choe, “Forecasting Sow’s Productivity using the Machine Learning Models,” *Journal of Agricultural Extension & Community Development*, vol. 16, no. 4, pp. 939-965, Dec. 2009.
- [3] S. Shahinfar, D. Page, J. Guenther, V. Cabrera, P. Fricke and K. Weigel, “Prediction of insemination outcomes in Holstein dairy cattle using alternative machine learning algorithms,” *Journal of dairy science*, vol. 97, no. 2, pp.731-742, Feb. 2014.
- [4] W. Lee, S. Kim, J. Ryu, and T. Ban, “Fast Detection of Disease in Livestock based on Deep Learning,” *Journal of Korea Institute of information and communication engineering*, vol. 21, no. 5, pp.1009-1015, May 2017.
- [5] M. A. Kashiha, C. Bahr, S. Ott, C. P. Moons, T. A. Niewold, F. Tuytens and D. Berckmans, “Automatic monitoring of pig locomotion using image analysis,” *Livestock Science*, vol. 159, no. 1, pp.141-148, Jan. 2014.
- [6] S. J. Roberts, R. Cain and M. S. Dawkins, “Prediction of welfare outcomes for broiler chickens using Bayesian regression on continuous optical flow data,” *Journal of the Royal Society interface*, vol. 9, no. 77, pp.3436-3443, Sep. 2012.
- [7] K. Kim, K. Kim, J. Kim, K. Seol, J. Hong, Y. Jung, J. Park, and Y. Kim, “Changes of serum electrolytes and hematological profiles in Yorkshire at a high ambient temperature” *Journal of Agriculture and Life Science*, vol. 49, no. 1, pp.103-113, Nov. 2014.
- [8] B. Fridrich, D. Krčmar, B. Dalmacija, J. Molnar, V. Pešić, M. Kragulj, and N. Varga, “Impact of wastewater from pig farm lagoons on the quality of local groundwater,” *Agricultural Water Management*, vol. 135, no. 1, pp. 40-53, Mar. 2014.



이용섭(Woongsup Lee)

2006년 한국과학기술원 전기및전자공학과 학사
 2011년 한국과학기술원 전기및전자공학과 박사
 2012년 2월~2013년 2월 한국과학기술원 박사후 연구원
 2013년 2월~2014년 2월 독일 FAU 박사후 연구원
 2014년 3월~ 현재 경상대학교 정보통신공학과 조교수
 ※관심분야 : 차세대 이동통신 시스템, 인지 라디오 시스템, 스마트 그리드, 빅 데이터 분석, 딥러닝



류종열(Jong Yeol Ryu)

2008년 2월 충남대학교 전기정보통신공학부 학사
 2010년 2월 KAIST 전기및전자공학과 석사
 2014년 2월 KAIST 전기및전자공학과 박사
 2014년 4월~2016년 8월 싱가포르기술디자인대학교 박사후 연구원
 2016년 9월~국립경상대학교 정보통신공학과 조교수
 ※관심분야 : 보안 통신 시스템, 차세대 이동통신 시스템, 사용자 릴레이 통신 등



반태원(Tae-Won Ban)

1998년 2월 경북대학교 전자공학과 학사
2000년 2월 경북대학교 전자공학과 석사
2010년 2월 KAIST 전기및전자공학 박사
2009년 6월~2010년 10월 KT 연구 엔지니어
2011년 2월~2012년 8월 KT Project Manager
2012년 9월~2016년 8월 경상대학교 정보통신공학과 조교수
2016년 9월~현재 경상대학교 정보통신공학과 부교수
※관심분야 : 차세대 이동통신, 무선 자원 관리, Cognitive Radios, Relay Systems, OFDM/MIMO



김성환(Seong Hwan Kim)

2006년 8월 고려대학교 전기전자전파공학부 학사
2008년 8월 KAIST 전자전산학과 석사
2013년 2월 KAIST 전기전자공학과 박사
2013년 8월~2016년 2월 맥길대학교 박사후 연구원
2016년 3월~ 국립경상대학교 정보통신공학과 조교수
※관심분야 : 차세대 이동통신 시스템, 사물인터넷



최희철(Hee-Chul Choi)

1979.3-1986.2월 충남대학교 낙농학과 농학사
1989.3-1991.8월 건국대학교 가축질병학과 수의학석사
2000.3-2003.8월 충남대학교 낙농학과 농학박사
1986.6-1993.7 공주군농촌지도소 농촌지도사
1993.7-2004.9월 축산시험장, 축산기술연구소 축산연구사
2004.10-2012.1월 축산과학원, 축산환경과, 가금과 농업연구관
2012.2월 ~2014.3월 농촌진흥청 국립축산과학원 가금과 과장
2014.4-현재 국립축산과학원 가축시설복지연구, ICT 연구총괄
※관심분야 : 스마트 양돈, 스마트팜, ICT, 축사시설, 축사환경관리, 동물복지