

기계학습기반 양돈생산성 예측방안

Production Performance Prediction of Pig Farming using Machine Learning

Woongsup Lee¹ · Kil-Young Sung² · Tae-Won Ban^{1*} · Young Hwa Ham³¹Associate Professor, Department of Information and Communication Engineering, Institute of Marine Industry, Gyeongsang National University, 53064, Korea²Professor, Department of Information and Communication Engineering, Gyeongsang National University, 53064, Korea³CEO, Agrirobotech Co., Ltd., 58, Sina-ro, Bubal-eup, Icheon-si, Gyeonggi-do, 17332, Korea

ABSTRACT

Smart pig farm which is based on IoT has been widely adopted by many pig farmers. In order to achieve optimal control of smart pig farm, the relation between environmental conditions and performance metric should be characterized. In this study, the relation between multiple environmental conditions including temperature, humidity and various performance metrics, which are daily gain, feed intake, and MSY, is analyzed based on data obtained from 55 real pig farm. Especially, based on preprocessing of data, various regression based machine learning algorithms are considered. Through performance evaluation, we show that the performance can be predicted with high precision, which can improve the efficiency of management.

Keywords : Machine learning, pig farm, big data, performance metric, smart farm

I. 서 론

최근 국내 양돈업계에서 사물인터넷에 기반을 둔 스마트 돈사의 활용이 점차 늘어나고 있다. 스마트 돈사의 효율적 관리를 위해서는 돈사내의 다양한 환경정보 및

각 개체의 생체정보와 양돈 생산성 지표와의 관계를 파악하는 것이 필요하다.

환경 정보를 기반으로 양돈 생산성을 예측하는 방법으로 돼지의 성장을 수학적인 모델로 근사하는 방안이 있다 [1]. 그 외에도 Net Energy 모델에서는 사료섭취량과 활동량 및 외부 기온을 토대로 돼지의 성장을 예측하였다 [2]. 하지만 이러한 방식에서는 복잡한 생체 물질 대사를 모사해야 하는 문제가 있다.

스마트 돈사에서는 여러 센서를 통해서 환경 및 생체 데이터를 수집할 수 있고 이를 기반으로 양돈관련 빅데이터 구축이 가능하다 [3, 4]. 이렇게 구축된 빅데이터에 기계학습방안을 적용하여 생산성 지표와 환경요소간의 관계를 구축할 수 있다. [4]에서는 실험돈사에서 수집된 데이터에 대표적인 기계학습방안중 하나인 의사결정트리 기법을 적용하여 돈방의 환경요소와 돼지의 성장률간의 관계를 분석하였다. 또한 [5]의 연구에서는 양돈관리 프로그램인 Pigplan을 통해 수집된 데이터를 기반으로 모든의 생산성을 예측하였다. 더불어 실제 돈사에서 측정된 데이터에 기계학습 방안을 적용하여 생산성과 밀접한 연관이 있는 급수량을 예측하는 방안이 [6]에서 개발되었다. 마지막으로 [7]에서는 딥러닝을 이용하여 축사에서 정상 및 질병개체를 분류하는 방안을 제시하였다.

기계학습을 적용하여 양돈 생산성지표를 예측하는 방안들이 많이 시도되었으나 대부분의 연구들에서 돼지의 평균 일별 성장률(일당증체), 모든의 생산성 및 질병발생여부 등의 지표들만 예측되었고, 실제 양돈농가에서 중요하게 고려하는 생산성 지표인 사료요구율 및 기간육성율과 같은 다른 지표들은 고려하지 않았다. 또한 많은 연구들이 실험돈사에서 적은 수의 개체로부터 수집된 인위적 데이터를 기반으로 진행되었으므로 실제 돈사환경과는 맞지 않을 수 있다.

본 연구에서는 55개의 양돈농가에서 수집된 530만개의 돈사 데이터를 기반으로 양돈농가 생산성 지표를 회귀기반 기계학습을 이용하여 예측하는 방안을 개발하

Received 1 November 2019, Revised 7 November 2019, Accepted 2 December 2019

* Corresponding Author Tae-Won Ban(E-mail:twban35@gnu.ac.kr, Tel:+82-55-772-9177)

Associate Professor, Department of Information and Communication Engineering, Institute of Marine Industry, Gyeongsang National University, 53064, Korea

Open Access <http://doi.org/10.6109/jkiice.2020.24.1.130>

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

였다. 또한 성능분석을 통해 높은 정확도로 생산성 지표를 예측할 수 있음을 보였고, 실제 수집된 데이터를 사용하였으므로 실제 축사환경에 더 잘 부합한다. 본 예측방안을 통해 농가의 생산성을 극대화하도록 돈사환경을 조절할 수 있다.

II. 데이터 수집 방안

본 연구에서는 애그리로보텍의 돈사관리 프로그램(LEMS)을 사용하는 55개의 양돈농가에서 2017년 10월부터 2018년 9월까지 수집된 530만개의 데이터를 사용하였다.

우선 돈사의 환경데이터로서 돈사의 온도(Ambient temperature) 및 습도(Humidity)를 측정하여 사용하였다. 그 외에 온도와 습도를 기반으로 실효온도(Effective temperature)를 계산하여 분석에 사용하였다. 또한 돼지의 현재 체중에 따라서 결정되는 최적의 돈사온도(Comfort temperature)를 계산하여 데이터로 활용하였다. 다음으로 일령(Initial age), 전입평체(Initial body weight), 전입두수(Number of units), 사육밀도(Stocking density)와 같은 생체특성을 데이터로 사용하였다.

본 연구에서는 실제 양돈농가 및 관리업체와의 컨설팅을 통해 양돈에서 제일 중요한 3가지 생산성 지표를 고려하였다. 우선 돼지의 일별 체중변화(일당증체, Average daily gain, ADG)를 고려하였다. 본 지표는 양돈농가에서 제일 중요하게 고려하는 지표로서 일당증체가 낮을 경우 돼지를 기르는데 더 오랜 기간이 걸리므로 양돈 농가의 생산성이 저하될 수 있다. 또한 1마리의 개체가 매일 소비하는 사료의 양인 사료요구율(Feed intake, FI)을 고려하였다. 사료요구율이 높을 경우 농가의 농장유지비가 증가하므로 생산성에 악영향을 준다. 그 외에도 입식 개체들 중 실제 도축된 개체의 비율을 나타내는 기간육성율(Marketed pigs per sow per year, MSY)을 양돈 생산성 지표로 고려하였다.

본 연구에서 고려한 데이터의 통계적인 특성은 표1에 나타내었다. 표1의 데이터에서 확인할 수 있듯이 최적 돈사온도에 비해서 돈사의 온도가 대체적으로 높게 유지되는 것을 확인할 수 있는데 이는 본 연구에서 사용한 무창돈사의 특성 때문이다. 또한 일당증체의 평균값이 0.74kg으로 양돈에서 일반적으로 고려하는 일당증체

값인 0.8kg에 비해서 다소 작을 것을 확인할 수 있다.

Table. 1 Attribute of data for pig farming.

Explanation	Average	Standard deviation	Minimum	Maximum
Ambient temperature (°C)	28.11	3.23	0.00	37.00
Humidity (%)	75.04	9.15	0.00	99.90
Effective temperature (°C)	23.40	4.94	-3.28	34.07
Comfort temperature (°C)	17.11	1.36	16.00	21.70
Initial age (days)	70.98	8.31	62	93
Initial body weight (kg)	29.58	5.19	24.00	46.00
Number of units (units)	492	360	200	1844
Stocking density (units/m ²)	1.03	0.20	0.53	1.80
ADG (kg)	0.74	0.11	0.46	1.00
Feed intake (kg)	3.05	0.3	2.54	3.88
MSY (%)	92.45	7.20	69.00	99.00

III. 데이터 전처리와 사용된 기계학습 방안

본 연구에서 사용한 온습도 데이터의 경우 1분에 1번씩 측정이 되므로 1개의 생산성 지표 데이터에 대해서 매우 많은 양의 특성(feature)이 존재하게 되고 기계학습 적용 시 과적합(overfitting)이 발생할 수 있다. 이를 해결하기 위해서 일정 시간간격으로 온습도 데이터를 샘플링하고 해당영역에서의 통계적인 특성, 즉 평균, 표준편차, 최소, 최댓값만을 사용하였다. 본 연구에서는 10분(10min), 2시간(2H), 1일(1D) 단위의 시간간격을 고려하였고 각 시간간격에서 10개, 120개, 720개의 온습도 데이터가 샘플링 되어서 사용된다. 이러한 데이터 전처리과정을 통해서 데이터의 양을 효과적으로 줄일 수 있고 과적합의 문제를 해결할 수 있다. 또한 분산팽창계수(Variance Inflation Factor, VIF)를 고려하여 다중공선성 진단을 통해 사용하는 특성을 선택하였다. 구체적으로 돈사의 온도 값에 대한 VIF값이 100이상으로 크게 계산되어 실제 기계학습방안에서는 돈사 온도데이터를 사용하지 않았다.

본 연구에서는 스마트 돈사에서 수집된 데이터를 이용하여 각 생산성 지표를 예측하는 방안을 기계학습을 기반으로 개발한다. 특히 실제 생산성 지표의 값을 예측하는 방안을 고려하였고 이를 위해서 회귀방식 기계학습 방안을 이용하였다. 본 연구에서 고려한 회귀방안에서는 선형회귀방식(Linear regression), 회귀트리방식(Regression Tree) 및 에이다부스트방식(AdaBoost regression)을 고려하였다. 구체적으로 선형회귀방식에서는 선형방정식을 이용하여 각 생산성을 예측하게 하였다. 회귀트리방식에서는 결정트리구조를 이용하여 값을 예측하였고, 트리의 최대 깊이는 6으로 설정하였다. 마지막으로 에이다부스트방식은 예측이 잘못된 값들을 적응적으로 고려하는 앙상블 알고리즘으로 본 연구에서는 base classifier로 깊이가 4인 회귀트리를 사용하였다. 성능 분석에서 평균제곱근 오차 (Mean Squared Error)를 성능지표로 사용하였다.

IV. 결과분석

본 장에서는 개발된 양돈 생산성 예측방안의 성능을 분석하였다. 성능 분석에서는 10-겹 교차검정 (10-fold validation)을 통해서 학습데이터를 변화시켜가면서 성능을 분석하였다. 표 2-4에는 각 생산성 지표 예측방안의 정확도를 나타내었다. 본 결과에서는 평균을 이용하여 값을 예측하는 ZeroR 예측방안을 이용하여 기준 성능을 계산하였다. 성능분석 결과에서 확인할 수 있듯이 에이다부스트방식이 낮은 오차를 보이고 회귀트리방식이 높은 오차를 보이는 것을 확인할 수 있었다.

Table. 2 MSE for average daily gain prediction.

Algorithm	10 min	2H	1D
Linear regression	0.0127	0.0082	0.0123
Regression tree	0.0139	0.0137	0.0135
Adaboost regression	0.0082	0.0081	0.0078
Baseline (ZeroR)	0.0121	0.0127	0.0124

Table. 3 MSE for feed intake prediction.

Algorithm	10 min	2H	1D
Linear regression	0.1057	0.1055	0.1063
Regression tree	0.1921	0.1945	0.1941
Adaboost regression	0.1002	0.0928	0.1076
Baseline (ZeroR)	0.0919	0.0911	0.0906

Table. 4 MSE for MSY prediction.

Algorithm	10 min	2H	1D
Linear regression	58.6	51.2	50.3
Regression tree	73.1	64.4	71.3
Adaboost regression	39.9	36.4	37.4
Baseline (ZeroR)	53.9	53.6	54.7

구체적으로 일당증체는 1일 단위 샘플링의 에이다부스트 방식에서, 사료요구율과 기간육성율은 2시간 단위 샘플링의 에이다부스트 방식에서 제일 낮은 오차를 보이는 것을 확인할 수 있었다. 이를 통해서 데이터의 전처리도 예측의 정확도에 큰 영향을 주는 것을 확인할 수 있었다. 더불어 사료요구율은 제안 방안의 성능이 기준 결과 비해서 낮은 것을 확인할 수 있고 본 지표는 제안 방안으로 예측이 어려움을 확인할 수 있다.

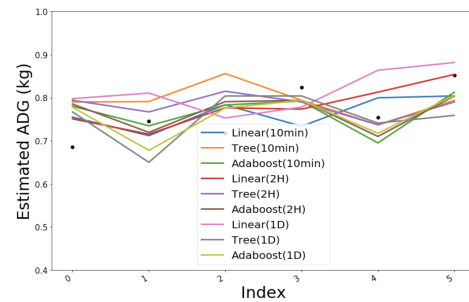


Fig. 1 Performance comparison ADG prediction.

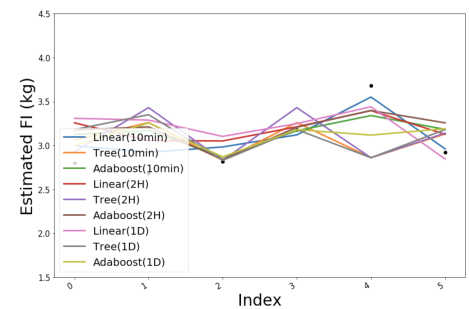


Fig. 2 Performance comparison for FI prediction.

다음으로 그림 1-3에서는 기계학습으로 예측한 생산성 지표값(실선)과 실제 측정된 생산성 지표값(점)을 함께 표시하였다. 결과에서 확인할 수 있듯이 3가지 회귀방식 알고리즘 모두 생산성을 비교적 정확하게 예측하는 것을 확인할 수 있었고 특히 에이다부스트 방식이 높은 정확도를 보이는 것을 확인할 수 있었다.

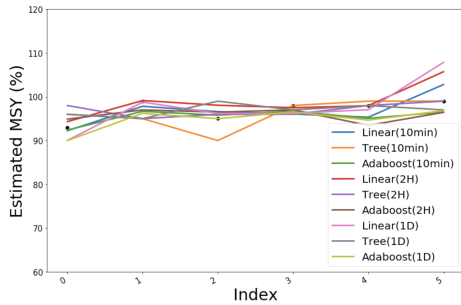


Fig. 3 Performance comparison for MSY prediction.

V. 결론

본 연구에서는 돈사에서 수집된 환경 및 개체의 생체 특성을 기반으로 회귀기반 기계학습방안을 사용하여 양돈 생산성지표를 예측하는 방안을 제시하였다. 성능 분석을 통해서 높은 정확도로 양돈의 생산성을 예측할 수 있음을 확인하였다. 본 예측 방안은 스마트 돈사에서 양돈의 생산성을 최대화 하도록 돈사 환경요소를 조절 하는데 사용할 수 있다. 현재 더 많은 데이터를 수집하여 정확도를 증가시키고 회귀방식 이외에 분류 기반 생산성지표 예측방안도 개발 중에 있다.

ACKNOWLEDGEMENT

This work was carried out with the support of "Cooperative Research Program for Agriculture Science & Technology Development (Project title: Development of swine management model with animal-metric for livestock welfare, Project No. PJ010541022016)" Rural Development Administration, Republic of Korea and also supported by Basic Science Research Program through the National Research Foundation of Korea(NRF) funded by the Ministry of Education(NRF-2018R1D1A1B07040796).

REFERENCES

- [1] W. Lee, K. -H. Han, H. -T. Kim, H. Choi, Y. Ham, and T. -W. Ban, "Prediction of Average Daily Gain of Swine based on Machine Learning," *Journal of Intelligent Fuzzy Systems*, vol. 36, no. 2, pp. 923-933, Mar. 2019.
- [2] S. Nitikanchana, S. Dritz, M. Tokach, J. DeRouchey, R. Goodband, and B. White, "Regression Analysis to Predict Growth Performance from Dietary Net Energy in Growing-finishing Pigs," *Journal of Animal Science*, vol. 93, no. 6, pp. 2826-2839, Jun. 2015.
- [3] M. Ryu, J. Yun, T. Miao, I. Ahn, S. Choi, and J. Kim, "Design and Implementation of A Connected Farm for Smart Farming System," in *Proc. of IEEE SENSORS*, Busan, South Korea, Nov. 2015.
- [4] 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.
- [5] 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.
- [6] W. Lee, J. Ryu, T. Ban, S. Kim, and H. Choi, "Prediction of Water Usage in Pig Farm based on Machine Learning," *Journal of Korea Institute of Information and Communication Engineering*, vol. 21, no. 8, pp.1560-1566, Aug. 2017.
- [7] W. Lee, S. Kim, J. Ryu, and T. Ban, "Fast Detection of Disease in Livestock based on Deep Learning," *Journal of the Korea Institute of Information and Communication Engineering*, vol. 21, no. 5, pp. 1009-1015, May. 2017.