

데이터 퓨전 전처리와 인공지능망을 활용한 스마트 양봉

권기현¹ · 조창현¹ · 이형봉^{2*}¹강원대학교 전자정보통신공학부²강릉원주대학교 컴퓨터공학과

Smart Beehive using Data Fused Preprocessing and Artificial Neural networks

Ki-Hyeon Kwon¹ · Changhyun Cho¹ · Hyung-Bong Lee^{2*}¹Department of Electronics, Information & Communication Engineering, Kangwon National University, Samcheok 25913, Korea²Department of Computer Science & Engineering, Gangneung-Wonju National University, Wonju 25457, Korea

[요 약]

데이터 퓨전은 다양한 센서 데이터를 조합하여 활용하는 방법으로, 개별 센서가 취득한 데이터보다 정확하고 유용한 자료를 얻을 수 있다. LSTM 네트워크는 RNN(recurrent neural network)의 개선된 모델로, RNN의 장기 의존성 문제를 개선한 형태이다. 이를 통해 시계열 데이터를 분류 또는 데이터 사이 장기간의 의존도를 학습할 수 있다. 본 논문에서는 IoT 양봉 장치를 활용하여 양봉 환경 모니터링 보조 및 양의 분봉 발생을 예측한다. 무게·온도·습도·사운드 센서로 데이터를 취득하고, 온도와 사운드 센서 데이터를 칼만 필터 기반 알고리즘과 게인 퓨전 기반 알고리즘을 통해 퓨전하여 LSTM 네트워크로 학습시켜 분봉 발생 시기 예측이 가능함과 게인 퓨전 기반 알고리즘이 우수한 성능을 보임을 확인한다.

[Abstract]

Data Fusion is a method that combines and utilizes various sensor data to obtain more accurate and useful data than the data acquired by individual sensors. The LSTM network is an improved model of the recurrent neural network (RNN), which improves the long-term dependency problem of RNNs. This allows you to classify time series data or to learn long-term dependencies between time series data. In this paper, the IoT beekeeping device is used to support beekeeping environment monitoring and predict the time of the swarm. Data can be acquired by weight, temperature, humidity, and sound sensors, and the temperature and sound sensor data can be fused through Kalman filter-based algorithm and gain fusion-based algorithm to be learned by LSTM network to predict the time of the swarm. And the gain fusion based algorithm shows more excellent performance than Kalman filter-based algorithm.

색인어 : 양봉 관리, 칼만 필터, 게인 퓨전, 장단기 메모리 네트워크**Key word** : Honey Bee Monitoring, Kalman Filter, Gain Fusion, LSTM<http://dx.doi.org/10.9728/dcs.2019.20.12.2321>

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.

Received 03 October 2019; Revised 13 November 2019

Accepted 05 December 2019

***Corresponding Author; Hyung-Bong Lee**

Tel: +82-33-760-8668

E-mail: hblee@gwnu.ac.kr

1. 서론

4차 산업혁명 기술이 여러 분야에 접목되고 있다. 이를 농업에 적용시켜 편의성과 생산성을 높이기 위해 스마트팜(Smart Farm)이 도입되었다. 사물 인터넷(IoT)과 농업 데이터 마이닝을 활용한 자동제어체계를 도입한 스마트팜은 사람의 행위 없이 최적의 환경에서 높은 생산성 향상을 보여주고 있다 [1].

양봉(beekeeping)은 벌을 기르는 분야로, 밀원의 감소 및 생육환경의 변화 그리고 양봉에 영향을 주는 매개변수(온도, 습도, 소음, 분봉 등)가 다양하여 양봉상자의 관리에 어려움을 겪고 있으며, 사물 인터넷(IoT) 기술이 접목 되는 경우 양봉 관리의 어려움을 해소할 수 있는 스마트팜(Smart Farm)의 좋은 사례가 될 수 있다.

양봉에 사물 인터넷(IoT)을 도입할 경우 양봉상자 내부의 모니터링을 위한 센서의 선택과 상이한 여러 센서 데이터들의 해석, 합성 및 주성분 선택을 통해 분봉 시점을 예측하고 분봉 예측의 정확도를 높이는 것이 과제가 된다.

양봉에 스마트팜 기술을 적용한 기존 연구를 살펴보면, 양봉상자 내부 온도 변화에 관한 연구와 양봉상자 내부 사운드 변화에 관한 연구로 나눌 수 있다.

온도 변화에 관한 연구로, 양봉상자 내부 온도 값을 측정하여 일정 구간에서 온도 변화 발생 시 분봉이 발생함을 보여주었다[2]. 또한, 상자 내부가 고온일 경우 꿀벌의 스트레스 상황을 모니터링 하고 분봉이 발생하는 시점을 탐지할 수 있다고 하였다[3]. 분봉은 상자 내부 온도 변화가 33 °C에서 35 °C로 변할 때 시작되어 실제 분봉이 일어나는 시기는 32 °C로 떨어질 때이다. 상자 내부 온도와 사운드 및 습도 사이 관련성이 크지 않지만 사운드를 보조적으로 활용하여 판단이 필요하다고 하였다[4].

사운드 변화에 관한 연구로, 상자 내부 사운드 패턴 분석을 통해 내부 벌의 비정상적 움직임 발생 탐지가 가능함을 보였다[5]. 그리고 상자 내부 사운드 변화 값을 주성분분석(PCA)을 활용하여 분봉 신호 식별 및 이를 통해 분봉 발생 이전에 예측이 가능하다는 것을 보였다[6].

그리고 양봉 사물인터넷에 관한 연구로 다양한 센서들(무게, 습도, 온도, 사운드)을 분석하고 활용하여 양봉 상태를 낮은 전송 에너지로 관리할 수 있음을 보였다[7].

위의 연구들을 종합하면 양봉상자 온도 변화가 중요하고, 사운드 변화도 중요한 데이터로 사용이 가능하며, 무게와 습도는 모니터링용 보조 데이터로 사용할 수 있음을 알 수 있다.

데이터 합성에 대한 연구로 선형 모델인 경우 개인 퓨전(Gain fusion) 알고리즘, 칼만 필터(Kalman filter) 알고리즘을 통해 퓨전이 가능하고 비선형 모델인 경우 EKF(extended KF)나 DFKF(derivative-free KF)를 통해 퓨전이 가능하다[8].

양봉 데이터와 유사한 시계열 데이터를 분류 또는 데이터 사이 장기간의 의존도를 학습할 수 있는 인공 신경망에 대한 연구로 RNN(Recurrent Neural Network)이 사용되고 있으며, 장기

의존성 문제를 해결하고, 일반화를 잘 할 수 있는 알고리즘인 LSTM(Long Short-Term Memory)이 적용 가능하다[9, 10].

본 논문에서는 양봉 생태환경 모니터링 및 분봉 발생 탐지할 수 있도록 다양한 센서(온도, 습도, 사운드, 무게)를 탑재한 IoT 양봉상자를 구성, 일정 기간 동안 데이터를 취득하였다. 다양한 데이터 분석 및 학습을 원활하게하기 위해 온도 및 사운드 값을 GF(Gain Fusion)를 사용한 데이터 퓨전(data fusion)을 통해 데이터 전처리를 하고 LSTM 학습을 수행하여 에러율을 구한다. 비교를 위해 KF(Kalman Filter) 데이터 퓨전 및 LSTM 학습과 차이가 있는지를 제시한다.

2장에서는 개인 퓨전 기반 알고리즘과 LSTM 네트워크를 정리하고, 3장에서는 구현한 IoT 양봉 장치에 대한 설명 및 데이터 취득 과정에 대하여 설명한다. 4장에서는 실험결과를 통한 결론을 설명한다.

II. 데이터 퓨전과 LSTM

2-1 데이터 퓨전(data fusion)

데이터 퓨전은 다양한 센서 데이터를 조합하여 활용하는 방법으로, 개별 센서가 취득한 데이터보다 정확하고 유용한 자료를 얻을 수 있다. [11].

양봉상자의 여러 센서 중에서 온도 센서와 사운드 센서를 활용하면 분봉 탐지가 가능함을 확인했기 때문에 이 두 개의 센서 데이터를 퓨전하여, 분봉 발생 탐지 가능성과 신뢰성을 향상시키고자 한다.

예측을 위한 데이터 퓨전에서 칼만 필터(Kalman Filter, KF) 기반 알고리즘과 개인 퓨전(Gain Fusion, GF) 기반 알고리즘이 있다. 데이터 손실이 없는 상태에서는 KF 기반 알고리즘이 GF 기반 알고리즘보다 근소하게 우위를 가지고 있고, 데이터 손실이 있는 상태에서는 GF 기반 알고리즘이 KF 기반 알고리즘보다 유의미하게 우위를 가짐을 보여주었다[11].

개인 퓨전은 시간 흐름에 따른 전역 예측치를 다음과 같이 설정한다(식 1, 2). x 는 상태 추정치이고 P 는 공분산이다. F 는 상태전이행렬, G 는 공정노이즈 이득행렬, Q 는 공정노이즈의 분산이다(공정노이즈의 평균은 0이다).

$$\tilde{x}_f(k+1) = F\hat{x}_f(k) \tag{1}$$

$$\tilde{P}_f(k+1) = F\hat{P}_f(k)F^T + GQG^T \tag{2}$$

지역 필터는 다음과 같이 리셋 한다(식 3,4).

$$\tilde{x}_m(k+1) = \tilde{x}_f(k+1) \tag{3}$$

$$\tilde{P}_m(k+1) = \tilde{P}_f(k+1) \tag{4}$$

지역 이득(K_m)과 지역상태의 측정값(\hat{x}_m) 업데이트는 다음 식으로 구해진다(식 5, 6). 센서별로 다른 노이즈 레벨을 가

지기 때문에 센서별 측정 노이즈 레벨 상수를 γ_m 으로 설정하고, $\sum_1^m \frac{1}{\gamma^m} = 1$ 을 만족한다. H 는 관찰행렬, R 은 측정노이즈 분산행렬, k_m 은 센서의 손실을 반영한 칼만이득(Kalman Gain), z 는 측정값에 공정노이즈와 측정노이즈를 더한 값이다.

$$K_m = (1/\gamma_m)\tilde{P}_f(k+1)H^T \left[H\tilde{P}_f(k+1)H^T + (1/\gamma_m)R_m \right]^{-1} \quad (5)$$

$$\hat{x}_m(k+1) = \tilde{x}_f(k+1) + k_m [z_m(k+1) - H\tilde{x}_f(k+1)] \quad (6)$$

m 개의 지역 예측치의 전역 퓨전은 다음과 같다(식 7, 8).

$$\hat{x}_f(k+1) = \sum^m \hat{x}_m(k+1) - (m-1)\tilde{x}_f(k+1) \quad (7)$$

$$\hat{P}_f(k+1) = \left[I - \sum^m K_m H \right] \tilde{P}_f(k+1) \left[I - \sum^m K_m H \right]^T + \sum^m K_m R_m (K_m)^T \quad (8)$$

2-2 LSTM(long short-term memory)

온도와 사운드 센서 데이터를 퓨전하고, 이를 LSTM 학습을 통해 분봉 발생 예측이 가능함을 확인하고자 한다.

LSTM 네트워크는 RNN(recurrent neural network)의 개선된 모델로, RNN의 장기 의존성 문제를 개선한 형태이다. 시계열 데이터를 분류 또는 데이터 사이 장기간의 의존도를 학습할 수 있다. LSTM 네트워크는 입력 층과 LSTM층으로 구분할 수 있는데 입력 층은 시계열 데이터를 입력받고, LSTM 레이어는 시계열 데이터 사이 장기간 의존도를 학습한다[11-13].

다음은 시간 t 에서 LSTM 데이터 흐름을 나타낸 것이다. x_t 는 관측 값, c_t 는 셀 상태, h_t 는 히든 상태를 의미한다. 입력 게이트 i_t , 리셋 게이트 f_t , 후보 게이트 g_t , 출력 게이트 o_t 에 의해 c_t 와 h_t 가 결정된다(그림 1).

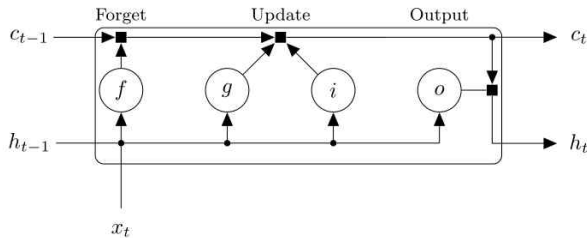


그림 1. LSTM 데이터 흐름도
Fig. 1. The Data Flow of LSTM

LSTM의 구성요소를 공식화하면 다음과 같다(W : Weight,

R : Recurrent Weight, b : Bias).

$$i_t = \sigma_g(W_i x_t + R_i h_{t-1} + b_i) \quad (8)$$

$$f_t = \sigma_g(W_f x_t + R_f h_{t-1} + b_f) \quad (9)$$

$$g_t = \sigma_c(W_g x_t + R_g h_{t-1} + b_g) \quad (10)$$

$$o_t = \sigma_g(W_o x_t + R_o h_{t-1} + b_o) \quad (11)$$

셀 상태와 히든 상태는 다음과 같다.

$$c_t = f_t \odot c_{t-1} + i_t \odot g_t \quad (12)$$

$$h_t = \sigma_t \odot c_t \quad (13)$$

여기서 \odot 는 Element-wise 곱을 의미한다.

III. IoT 양봉 장치 구현 및 데이터 취득

3-1 IoT 양봉 장치

양봉 관리 및 분봉 발생 탐지 데이터 취득을 위해 컨트롤러로 라즈베리 파이를 사용하였고 무게, 온도, 습도, 사운드 센서를 장착하고 취득된 데이터를 스마트폰으로 전송하는 IoT 양봉 장치를 구현하였다(그림 2). 사용한 센서 및 하드웨어 스펙을 표 1에 나타내었다.

표 1. 하드웨어 스펙

Table. 1. Hardware Specification

Item	Value
Controller	• Raspberry PI 3B+
Sensors	• Load Cell: capacity: 40kg, HX711 Module • Temperature/Humidity: analog DHT11 • Sound: sampling frequencies of 44.1 kHz
Touch Screen	• 7" Touch Screen, 800 × 480 pixels

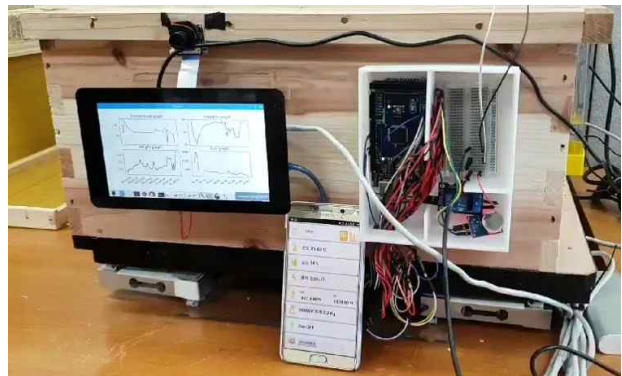


그림 2. IoT 양봉 장치
Fig. 2. IoT Beehive Monitoring Device

3-2 초기 데이터 취득 및 정규화

구현된 IoT 양봉 장치를 양봉상자에 설치하여 5월 한 달 동안 60초 간격으로 데이터를 취득하였으며 분봉의 변화가 감지된 구간의 1200개의 데이터를 취득하고 서로 다른 물리량의 범위를 동일하게 맞추기 위해 리스케일링 전처리 후 실험에 사용한다.

그림 3을 통해 사운드와 온도의 경우 출력결과가 선형으로 유사하여 칼만 필터(KF: Kalman Filter) 기반 알고리즘과 개인 퓨전(Gain fusion) 기반 알고리즘으로 퓨전이 가능함을 알 수 있다.

연구 [4-6]을 토대로, 온도 데이터에 사운드 데이터를 칼만 필터(KF: Kalman Filter) 기반 알고리즘과 개인 퓨전(Gain fusion) 기반 알고리즘으로 퓨전하여 분봉 예측 데이터로 사용하면 어떤 알고리즘이 개선된 결과를 도출할 수 있을지를 확인하기 위해, 계열 데이터에 특성이 있는 LSTM 네트워크를 통해 분봉 발생 예측을 학습하여 칼만 필터(KF: Kalman Filter) 기반 알고리즘과 개인 퓨전(Gain fusion) 기반 알고리즘 중에서 어떤 알고리즘이 분봉 예측 데이터로서 개선된 결과를 보이는지 실험을 통해 확인한다.

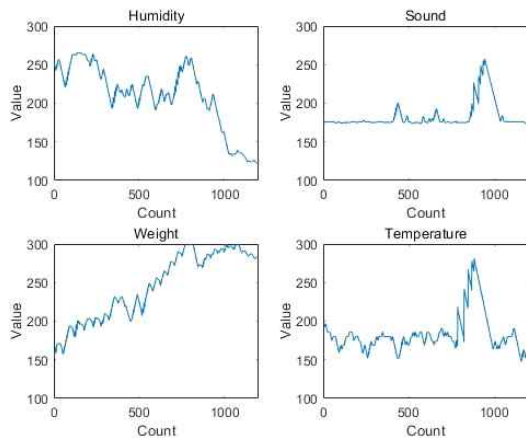


그림 3. IoT 양봉장치 센서 데이터 취득
Fig. 3. IoT Beehive Sensor Data Logging

IV. 실험 결과

4-1 온도, 사운드 데이터 퓨전

온도와 사운드 데이터를 칼만 필터 기반 알고리즘과 개인 퓨전 기반 알고리즘으로 데이터 퓨전을 하였다. 칼만 필터 기반 알고리즘을 이용한 퓨전(그림 4) 보다 개인 퓨전 기반 알고리즘을 이용한 퓨전(그림 5)에서 온도 및 사운드 데이터 값과 퓨전된 데이터와의 오차가 최소화되고 정합 정도가 우수하여 퓨전이 잘 되고 있는 것으로 판단된다. 칼만 필터 기반 알고리즘과 개인 퓨전 기반 알고리즘 두 경우 모두 퓨전 과정에서 초기 온도와 사운드 데이터의 값도 변경 되어 그림 4, 그림 5와 같이 온도, 사운드와 퓨전된 데이터가 표현 된다.

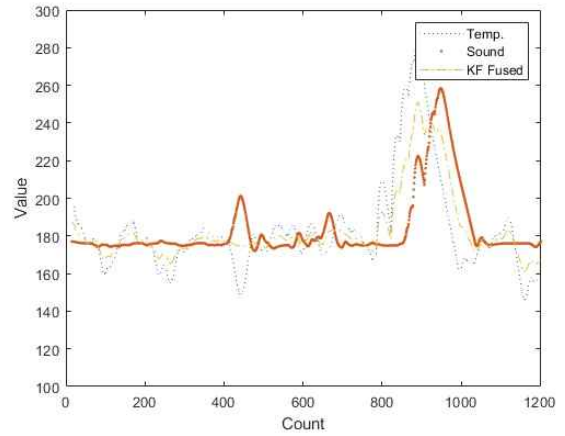


그림 4. 칼만 필터 기반 알고리즘을 이용한 퓨전된 데이터
Fig. 4. Fused Data using Kalman Filter-based Algorithms

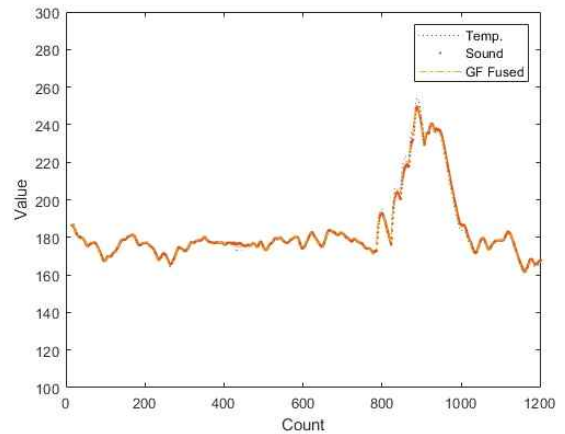


그림 5. 개인 퓨전 기반 알고리즘을 이용한 퓨전된 데이터
Fig. 5. Fused Data using Gain Fusion-based Algorithms

4-2 LSTM 학습 및 예측

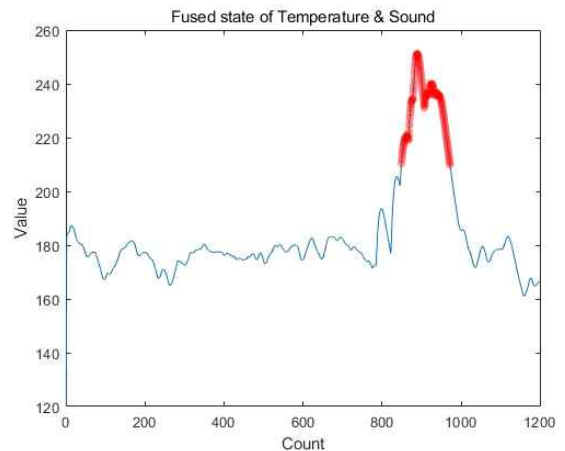


그림 6. 퓨전 데이터의 분봉 예측 영역
Fig. 6. Forecast Area of Swarming for Fused Data

그림 4와 5는 온도와 사운드를 퓨전한 데이터로, LSTM 네트워크 학습 데이터로 사용하여 분봉을 예측한다. 그림 6의 적색으로 표시한 부분은 퓨전 데이터에서 상자 내부 온도가 33 °C에서 35 °C로 증가하다가 32 °C로 떨어질 때를 의미하며, 분봉 발생이 예측되는 부분이다.

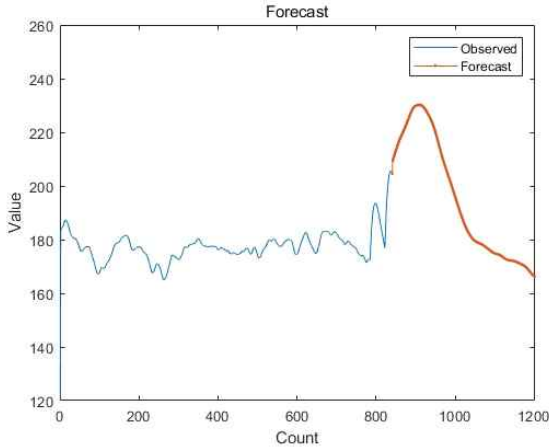


그림 7. 퓨전 데이터의 분봉 예측
Fig. 7. Forecast of Swarming for Fused Data

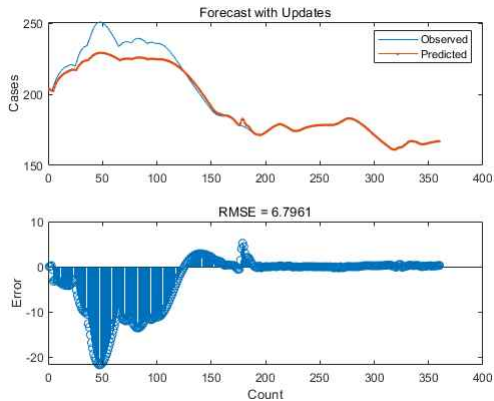


그림 8. 칼만 필터기반 알고리즘 예측 에러(RMSE)
Fig. 8. RMSE of Kalman Filter-based Algorithms

LSTM 학습 환경은, 입력 데이터의 90% 데이터는 학습용 데이터로, 10% 데이터를 검증용 데이터로 사용하였고, LSTM 히든 레이어는 200개, 학습 횟수는 250회, 초기 학습율은 0.005, 그리고 매 125회 학습시마다 0.2씩 학습율을 낮추도록 구성하였다.

그림 7을 통해 데이터 퓨전을 통한 분봉 시기 예측이 가능함을 확인할 수 있고, 그림 8, 9를 통해 칼만 필터 기반 알고리즘을 통한 예측 값의 RMSE(Root Mean Square Error)는 6.7961, 게인 퓨전 기반 알고리즘을 통한 예측 값의 RMSE는 2.5189가 나오는 것으로 보아 게인 퓨전 기반 알고리즘이 더 정확하게 예측할 수 있음을 보여준다.

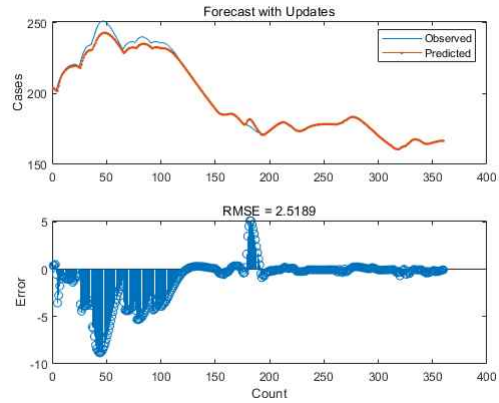


그림 9. 게인 퓨전 기반 알고리즘 예측 에러(RMSE)
Fig. 9. RMSE of Gain Fusion-based Algorithms

V. 결론 및 평가

IoT 양봉상자를 양봉 관리 및 분봉 발생 탐지에 활용 가능성이 있음을 보였다. 특히 분봉과 관련한 센서 데이터 중에서 사운드와 온도 데이터의 출력결과가 선형으로 유사하여 퓨전이 가능하고, 게인 퓨전(Gain fusion) 기반 알고리즘으로 퓨전하여, LSTM 네트워크 학습의 분봉 예측 데이터로 사용한 경우, 칼만 필터를 사용한 경우에 비해 RMSE가 낮아 정확도가 높아지는 개선된 결과를 보였다.

LSTM 네트워크 학습 전 데이터 퓨전을 통해 분봉을 정확하게 예측할 수 있었던 것을 개선하여, 이번 연구를 통해 LSTM 네트워크 학습 전 데이터에 알맞은 데이터 퓨전 알고리즘 선택에 따라 데이터 및 학습결과와 신뢰성 향상이 가능함을 제시하였고, 이를 활용하여 분봉 예측 정확성 개선 및 학습 데이터에 맞는 알맞은 데이터 퓨전 알고리즘 선택 필요성을 제시하였다.

감사의 글

이 논문은 2017년도 정부(과학기술정보통신부)의 재원으로 한국연구재단 -현장맞춤형 이공계 인재양성 지원사업의 지원을 받아 수행된 연구임(No. 2017H1D8A1028271).

참고문헌

- [1] J. Muangprathub, "IoT and agriculture data analysis for smart farm, *Computers and Electronics in Agriculture*, Vol.156 pp.467-474, January, 2019
- [2] A. Zacepins, "Remote detection of the swarming of honey bee colonies by single-point temperature monitoring," *Biosystems Engineering*, Vol.148 pp. 76-80, 2016.
- [3] D. S. Kridi, "Application of wireless sensor networks for beehive monitoring and in-hive thermal patterns detection,"

- Computers and Electronics in Agriculture*, Vol. 127 pp. 221-235, 2016.
- [4] S. Ferrari, "Monitoring of swarming sounds in bee hives for early detection of the swarming period," *Computers and Electronics in Agriculture*, Vol. 64, Issue 1, pp. 72-77, 2008.
- [5] M. Bencsik, "Identification of the honey bee swarming process by analysing the time course of hive vibrations," *Computers and Electronics in Agriculture*, Vol. 76 pp. 44-50, 2011.
- [6] N. Pérez, "Continuous monitoring of beehives' sound for environmental pollution control," *Ecological Engineering*, Vol. 90 pp. 326-330, 2016.
- [7] F. Edwards-Murphy, "b+WSN: Smart beehive with preliminary decision tree analysis for agriculture and honey bee health monitoring," *Computers and Electronics in Agriculture*, Vol. 124, pp. 211-219, 2016.
- [8] B. Paik, "Gain fusion algorithm for decentralised parallel Kalman filters", *IEE Proc.-Control Theory Appl.*, Vol. 147, pp. 97-103, 2000
- [9] H. Sepp, "Long short-term memory", *Neural Computation*, Vol. 9, pp. 1735-1780, 1997
- [10] K. Kwon, "Forecast of Bee Swarming using Data Fusion and LSTM", *Journal of Digital Contents Society*, Vol. 20, pp. 1-6, 2019
- [11] J. R. Raol, *Multi-Sensor Data Fusion with MATLAB*, Baton Rouge: CRC Press, 2010.
- [12] The MathWorks, Inc. Long Short-Term Memory Networks - LSTM Layer Architecture. Available: <https://www.mathworks.com/help/deeplearning/ug/long-short-term-memory-networks.html>.
- [13] L. Donghyun, "LSTM RNN-based Korean Speech Recognition on System Using CTC," *Journal of Digital Contents Society*, Vol. 18, No. 1, pp. 93-99, 2017.



권기현(Ki-Hyeon Kwon)

1993년 : 강원대학교 컴퓨터학과(학사)
1995년 : 강원대학교 대학원 컴퓨터학과(석사)
2000년 : 강원대학교 대학원 컴퓨터학과(박사)

1998년 ~ 2002년: 동원대학 인터넷정보과 교수
2002년~ 현재: 강원대학교 정보통신공학과 교수
※관심분야 : 패턴 인식 (Pattern Recognition), 사물 인터넷 (IoT), 기계학습



조창현(Chang-hyun Cho)

2013년 : 건국대학교 전기공학과(학사)
2019년 ~ 현재: 강원대학교 전자정보통신공학과(석사과정)

※관심분야 : 강화학습(Reinforcement Learning), GAN(Generative Adversarial Network), OT보안(Operational Technology Security)



이형봉(Hyung-Bong Lee)

1984년 : 서울대학교 계산통계학과(학사)
1986년 : 서울대학교 대학원 계산통계학과(석사)
2000년 : 강원대학교 대학원 컴퓨터학과(박사)

1986년 ~ 1994년: LG전자 컴퓨터연구소
1994년 ~ 1999년: 한국디지털(주)
2004년 ~ 현재: 강릉원주대학교 컴퓨터공학과 교수
※관심분야 : 무선 통신 (Wireless Networks), 센서 네트워크 (Sensor Networks), 임베디드 시스템 (Embedded Systems), 사물 인터넷 (IoT)