JDCS
디지털콘텐츠학회논문지
Journal of Digital Contents Society
Vol. 20, No. 1, pp. 1-6, Jan. 2019

데이터 퓨전 및 LSTM를 사용한 양봉 분봉 예측

권 기 현[1] · 김 진 수[1] · 이 형 봉[2*]
[1]강원대학교 전자정보통신공학부
[2]강릉원주대학교 컴퓨터공학과

Forecast of Bee Swarming using Data Fusion and LSTM

Ki-Hyeon Kwon[1] · Jin-Su Kim[1] · Hyung-Bong Lee[2*]

[1]Department of Electronics, Information & Communication Engineering, Kangwon National University, Samcheok 25913, Korea

[2]Department of Computer Science & Engineering, Gangneung-Wonju National University, Wonju 25457, Korea

[요 약]

데이터 퓨전은 유용한 정보를 창출할 목적으로 여러 센서로부터 취득한 자료를 조합하는 것으로, 취득한 자료를 최소화하는 절차를 거쳐 신뢰성이나 판별력을 개선해 나가는 처리 방법이다. 그리고 LSTM 네트워크는 RNN(recurrent neural network)의 한 형태로서 타임 시리즈 데이터에 대해서 분류 하거나 긴 시간 동안의 의존도를 학습할 수 있다. 본 논문에서는 IoT 양봉 장치를 제작하여, 양봉의 생육 환경 모니터링을 지원하고 양봉의 분봉 발생을 탐지한다. 온습도, 사운드, 무게 센서를 탑재하여 데이터를 취득하고, 취득한 온도와 사운드 센서 데이터를 퓨전하여 데이터 분석에 유용한 데이터를 생성하고 판별력을 검사한다. LSTM 네트웍을 퓨전된 데이터를 사용하여 학습시켜, 분봉 발생 시기 예측이 가능함을 보인다.

[Abstract]

Data fusion(DF) is the process of combining or preprocessing data originated from different sensors or sources to produce a more specific, comprehensive, and unified dataset about event of interest that has been observed. And, LSTM(long short-term memory) network is a type of recurrent neural network(RNN) that can learn classification and long-term dependencies between time steps of sequence data. In this paper, we implement an IoT beekeeping device that attached in the beehive box, monitor the bee living conditions and identify the time of the swarm. We gather the dataset from the IoT beekeeping device equipped with the temperature/humidity, sound, weight sensors and fuse temperature with sound data to produce a comprehensive, and unified dataset. We assess the usability of forecasting the bee swarming by an LSTM network model learned with the fused data.

색인어 : 양봉 관리, 사물인터넷, 스마트팜, 데이터 퓨전, 장단기 메모리 네트워크

Key word : Honey Bee Monitoring, Internet of Things (IoT), Honey Bee Swarming, Data Fusion, LSTM

http://dx.doi.org/10.9728/dcs.2019.20.1.1

This is an Open Access article distributed under the terms of the Creative Commons Attribution Non-CommercialLicense(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 December 2018; **Revised** 13 December 2018
Accepted 20 January 2019

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

Tel: +82-33-760-8668
E-mail: hblee@gwnu.ac.kr

Ⅰ. 서 론

4차 산업혁명의 주요 기술인 사물 인터넷(IoT), 인공지능(AI)을 농업에 접목시켜 전통적인 농업 방식에서 탈피하고 농업의 생산성 및 효율성을 높이려는 움직임이 활발하게 이루어지고 있다. 그 대표사례가 농업기술에 4차 산업혁명 기술을 접목한 스마트팜(Smart Farm) 이다[1]. 농업기술 중에서 양봉(beekeeping)은 꽃가루 매개자로서 식량 공급체계의 중요한 역할을 담당하고 있으면서도, 온습도 관리 및 식량 공급 그리고 분봉 발생 탐지 등 그 관리의 어려움으로 인해 스마트팜 기술 적용이 쉽지 않은 분야이다.

양봉 관리에 스마트팜 기술을 적용한 연구는 양봉 상자내의 온도 변화에 관한 연구와 사운드 변화에 관한 연구로 구분된다.

온도 변화에 관한 연구로서, 분봉 발생 탐지를 위해서 양봉 상자의 온도 값을 측정하여, 일정 구간의 온도 변화가 있는 경우 분봉 발생 을 탐지할 수 있다고 하였다[2]. 또한, 양봉 상자의 내부 온도가 고온일 때 꿀벌의 스트레스 상황을 모니터링 하고 분봉이 발생하는 시점을 탐지할 수 있다고 하였다[3]. 그리고 양봉에서 분봉은 양봉 상자의 내부 온도가 33 ℃에서 35 ℃일 때 개시되며 실제 분봉이 일어날 때는 온도가 32 ℃로 떨어질 때이다. 양봉 상자의 내부 온도와 사운드 및 습도와는 상관관계는 높지 않으나 양봉 상자의 사운드를 보조적으로 결합하여 판단하는 것이 필요함을 제시하였다[4].

사운드 변화에 관한 연구로서, 양봉 상자의 벌의 움직임을 관찰하기 위해서 사운드 패턴을 분석하여 비정상적인 변화가 발생하는 경우를 탐지하는 실험을 진행하였다[5]. 또한, 양봉 상자 내 사운드 변화를 취득하고 주성분분석(PCA)을 통해 분봉 신호를 식별하고 이 신호를 통해 분봉 수일 이전에 분봉 예측이 가능함을 보였다[6].

그리고 양봉 사물인터넷에 관한 연구로서, 양봉 상자의 무게, 온도, 습도, 사운드 등의 센서 값을 취득하여 여러 관점에 맞게 분석하고 의사 결정 알고리즘을 사용하여 양봉의 상태를 파악하며 최소의 전송 에너지로 관리 가능함을 보였다[7].

관련 연구를 통해 양봉 상자에서 온도의 변화가 주된 분석 대상이고 사운드 변화가 보조적으로 활용될 수 있으며, 습도 및 무게는 양봉의 생태를 모니터링 하는 보조적인 데이터로 활용됨을 알 수 있다.

본 논문에서는 양봉의 분봉 발생 탐지 및 양봉의 생태를 모니터링 할 수 있도록 온습도, 사운드, 무게 센서를 탑재한 IoT 양봉 상자를 구성하고 일정 기간 동안 데이터를 취득하였다. 여러 종류의 데이터를 분석하고 학습하는 어려움을 극복하기 위해서 온도와 사운드 데이터에 대해서 데이터 퓨전(data fusion)을 통해 데이터를 합성하고 LSTM(long short-term memory) 네트워크로 학습 시켜 어느 정도 데이터 퓨전의 활용도가 있고 LSTM 학습 효과가 있는지를 제시한다.

2장에서는 데이터 퓨전과 LSTM에 대해 정리하고, 3장에서는 IoT 양봉 장치 구현 및 데이터 취득에 대해 설명한다. 4장에서는 실험결과를 제시하고 결론을 맺는다.

Ⅱ. 데이터 퓨전과 LSTM

2-1 데이터 퓨전(data fusion)

데이터 퓨전은 유용한 정보를 창출할 목적으로 여러 센서로부터 취득한 자료를 조합하는 것으로, 취득한 자료를 최소화하는 절차를 거쳐 신뢰성이나 판별력을 개선해 나가는 처리 방법이다[8].

관련 연구[4, 5]로부터, 양봉 상자의 여러 센서 값 중에서 온도와 사운드는 분봉 발생 탐지를 위해서 상호 보완적이므로, 온도와 사운드 센서 값을 퓨전(fusion)한 데이터를 사용하여, 분봉 발생 탐지의 신뢰성이나 판별력을 개선하고자 한다.

데이터 퓨전에 있어 선형 모델인 경우 칼만 필터(KF: Kalman Filter)를 통해 퓨전이 가능하고 비선형 모델인 경우 EKF(extended KF)나 DFKF(derivative-free KF)를 통해 퓨전이 가능하다.

온도와 사운드 센서 값은 선형 모델로 볼 수 있으므로 칼만 필터를 통해 퓨전한다. 칼만 필터는 각 센서 데이터에 독립적으로 적용되어 필터링된 결과, 즉 시간 흐름에 따른 상태 추정치(state estimate) X와 공분산 P을 구한다(식 1, 2)[8].

$$\widetilde{X}_{(k+1)} = \Phi \hat{X}_{(k)} \tag{1}$$

$$\widetilde{P}_{(k+1)} = \Phi \widehat{P_{(k)}} \Phi^T + GQG^T \tag{2}$$

다음으로 칼만게인 K를 계산하고 측정값 Z로 상태 추정치와 공분산을 보정하며(식 3, 4, 5), 추정과 보정의 과정을 반복한다.

$$K_{(k+1)} = \widetilde{P}_{(k+1)} H^T \left[H\widetilde{P}_{(k+1)} H^T + R \right]^{-1} \tag{3}$$

$$\hat{X}_{(k+1)} = \widetilde{X}_{(k+1)} + K_{(k+1)} \left[Z_{(k+1)} - H\widetilde{X}_{(k+1)} \right] \tag{4}$$

$$\hat{P}_{(k+1)} = \left[I - K_{(k+1)} H \right] \widetilde{P}_{(k+1)} \tag{5}$$

퓨전은 두 독립적인 상태 추정치의 가중 합을 사용하여 상태 추정치들을 결합하여 구하며, 사용되는 가중 인자는 공분산 매트릭스를 사용한다.

퓨전된 상태와 공분산 매트릭스는 다음 식으로 구해진다(식 6, 7).

$$\widehat{X^f} = \widehat{X^1} + \widehat{P^1}(\widehat{P^1} + \widehat{P^2})(\widehat{X^2} - \widehat{X^1}) \tag{6}$$

$$\widehat{P^f} = \widehat{P^1} - \widehat{P^1}(\widehat{P^1} + \widehat{P^2})^{-1} \widehat{P^1}^T \tag{7}$$

- $\widehat{X^1}$과 $\widehat{X^2}$는 각각 센서 1, 2의 측정치를 필터 벡터 1, 2로 필터링하여 추정한 상태 벡터들이다.

- $\widehat{P^1}$과 $\widehat{P^2}$는 필터 1과 2로부터 추정된 상태 에러 공분산(estimated state error covariance)이다.

2-2 LSTM(long short-term memory)

온도와 사운드 센서 값을 퓨전(fusion)한 데이터에 대하여, LSTM 네트워크를 사용하여 학습하고 분봉 발생을 예측하고자 한다.

LSTM 네트워크는 RNN(recurrent neural network)의 한 형태로서 타임 시리즈 데이터에 대해서 분류 하거나 긴 시간 동안의 의존도를 학습할 수 있다. LSTM 네트워크는 입력 층과 LSTM 층으로 구성된다. 입력 층은 타임 시리즈 데이터를 입력하는 층이고 LSTM층은 타임 시리즈 데이터간의 긴 시간 동안 의존도를 학습하는 층이다[9-11].

다음 그림은 시간 t에서 LSTM 데이터 흐름을 나타낸 것이다. x_t는 관측값, c_t는 셀 상태, h_t는 히든 상태를 의미한다. 입력게이트 i_t, 리셋 게이트 f_t, 후보 게이트 g_t, 출력 게이트 o_t에 의해 c_t와 h_t가 결정된다(그림 1).

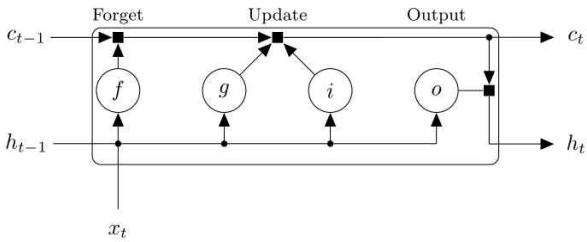

그림 1. LSTM 데이터 흐름도
Fig. 1. The Data Flow of LSTM

LSTM의 구성요소를 공식화하면 다음과 같다(W: Weight, R: Recurrent Weight, b: Bias).

$$i_t = \sigma_g(W_i x_t + R_i h_{t-1} + b_i) \tag{8}$$
$$f_t = \sigma_g(W_f x_t + R_f h_{t-1} + b_f) \tag{9}$$
$$g_t = \sigma_c(W_g x_t + R_g h_{t-1} + b_g) \tag{10}$$
$$o_t = \sigma_g(W_o x_t + R_o h_{t-1} + b_o) \tag{11}$$

셀 상태와 히든 상태는 다음과 같다.

$$c_t = f_t \odot c_{t-1} + i_t \odot g_t \tag{12}$$
$$h_t = \sigma_t \odot \sigma_c(c_t) \tag{13}$$

Ⅲ. IoT 양봉 장치 구현 및 데이터 취득

3-1 IoT 양봉 장치

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

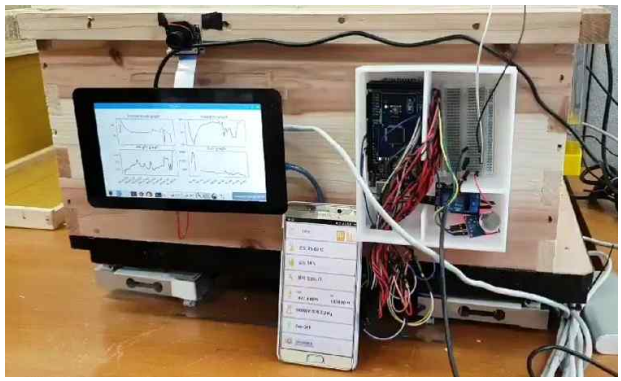

그림 2. IoT 양봉 장치
Fig. 2. IoT Beehive Monitoring Device

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.

3-2 초기 데이터 취득 및 정규화

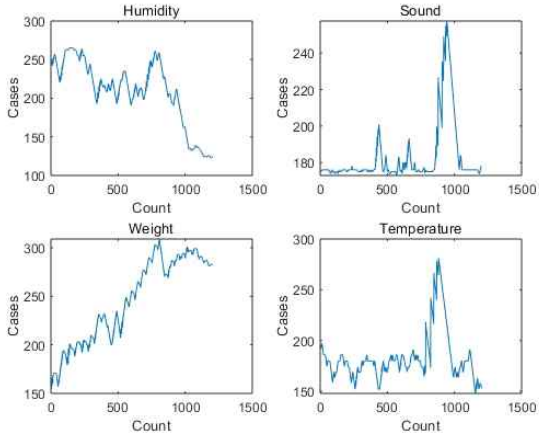

그림 3. IoT 양봉장치 센서 데이터 취득
Fig. 3. IoT Beehive Sensor Data Logging

IoT 양봉 장치를 양봉 상자에 부착하여 분봉이 가장 많이 일

어나는 5월 한 달간 60초 간격으로 데이터를 취득하고 정규화를 통해 전처리하였다. 그림 3을 통해 온도와 사운드는, 상호 보완적이고 선형 모델로 볼 수 있으므로, 칼만 필터를 통해 퓨전(fusion)하고, 퓨전된 데이터를 LSTM 네트워크를 사용하여 학습하여 분봉 발생 예측에 사용한다.

Ⅳ. 실험 결과

4-1 온도, 사운드 데이터 퓨전

온도와 사운드 데이터 각각에 대해서 칼만 필터를 적용하여 추정 값과 측정값을 표시하였고 잔차(residual)를 나타내었다(그림 4).

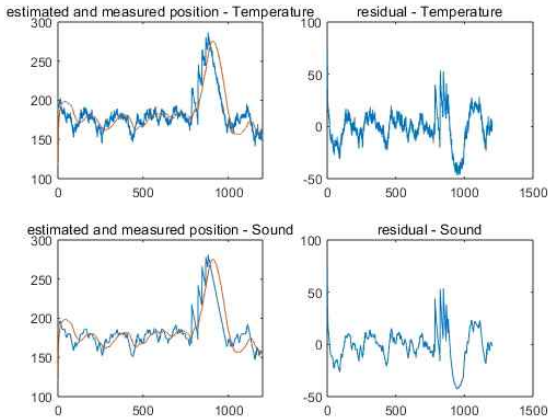

그림 4. 온도, 사운드 데이터의 추정 및 예측 값과 잔차
Fig. 4. Estimated and Measured Position and Residual of Temperature, Sound

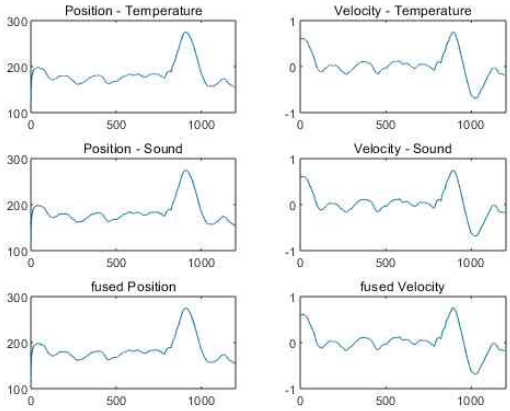

그림 5. 온도, 사운드, 퓨전 데이터의 상태(position)와 속도(velocity)
Fig. 5. Position and Velocity of Temperature, Sound and Fused Data

온도와 사운드 데이터에 대해서 위치(position)와 데이터 이동 속도(velocity)를 구하고, 데이터 퓨전(data fusion)을 통해 온도와 사운드 데이터를 합성하였다(그림 5).

그림 5는 온도, 사운드 그리고 퓨전된 데이터에 대한 공분산을 표시한 것으로, 퓨전된 데이터에 대한 공분산이 온도와 사운드 데이터의 공분산 보다 낮으므로, 데이터 퓨전을 통해 예측정확도가 개선되고 불확실성이 감소되었음을 의미한다(그림 6).

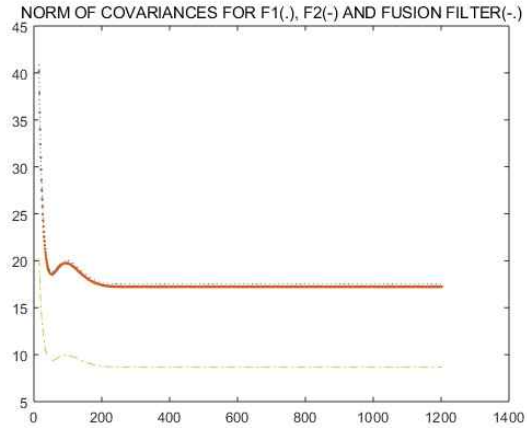

그림 6. 온도, 사운드, 퓨전 데이터의 공분산 값
Fig. 6. Norm of Covariances for Temperature, Sound and Fused Filter

4-2 LSTM 학습 및 예측

그림 7은 그림 5에서 온도, 사운드를 퓨전한(fused Position) 데이터 플롯으로, LSTM의 입력으로 사용되어, 데이터 값이 피크 파형인 경우를 예측하는 데이터로 사용된다. 점선으로 표시한 부분은, 양봉 상자의 내부 온도가 33 °C에서 35 °C로 증가하다가 32 °C로 떨어질 때로 분봉 발생으로 예측되는 영역이다.

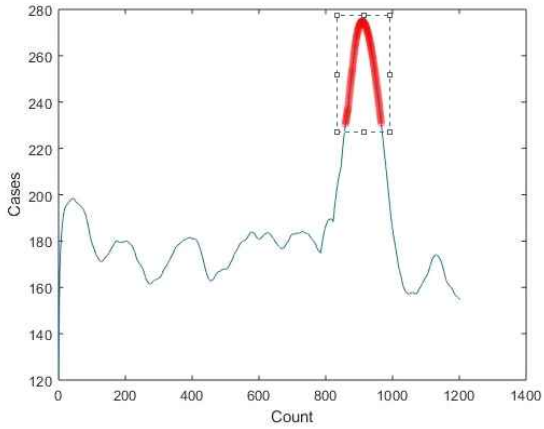

그림 7. 온도, 사운드를 퓨전한 데이터
Fig. 7. Fused State of Temperature & Sound

LTSM 학습은 반복 학습 250회, 초기 학습율 0.005, 학습율을 매 125회 마다 0.2씩 감소시키도록 설정하였다. 그림 8은 LTSM 학습을 통해 기존 측정데이터에 기반을 두어 예측한 값을 표시하였다. 그림 9에서 퓨전 데이터와 분봉 예측 데이터 간의 RMSE가 6.1558로 매우 양호한 수준의 에러율을 표시하고 있다.

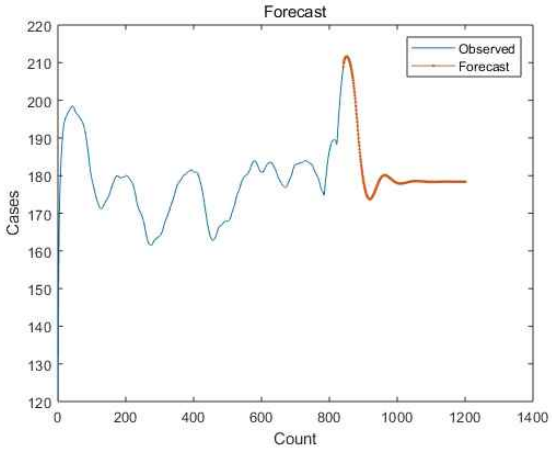

그림 8. 퓨전 데이터의 분봉 예측
Fig. 8. Forecast of Swarming for Fused Data

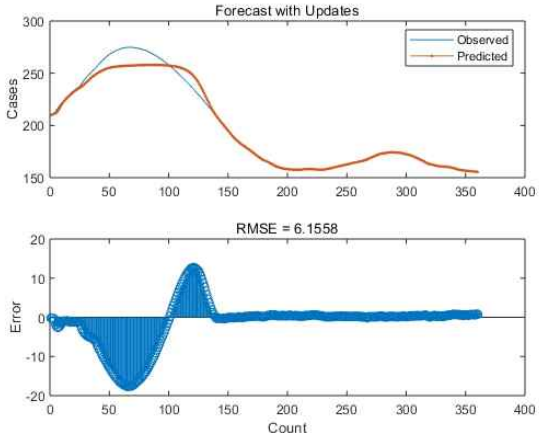

그림 9. 퓨전 데이터의 예측 에러(RMSE)
Fig. 9. RMSE of Fused Data

Ⅴ. 결론 및 평가

양봉 관리 및 분봉 발생 탐지를 위해 온습도, 사운드, 무게 센서를 탑재한 IoT 양봉 상자를 구성하였으며, 센서들 중에서 온도센서가 가장 의미가 크고, 사운드 센서가 보조적으로 사용될 수 있음을 확인하였다. 온도와 사운드 센서 값은 선형 모델로 볼 수 있으므로 칼만 필터를 통해 퓨전하였다.

온도와 사운드 센서 값을 이용하여 퓨전된 데이터를 이용하여 LSTM 네트워크로 학습시켜 분봉 발생 탐지가 가능함을 보였으며, 퓨전 데이터와 분봉 예측 데이터 간의 에러율이 낮은 수준을 보였다.

이번 연구를 통해 선형 모델로 볼 수 있는 멀티 센서의 값을 데이터 퓨전을 통해 합성하여 유용한 데이터를 생성하고 신뢰성이나 판별력 개선이 가능하며, LSTM 네트워크 학습을 통해 양봉 분봉 예측에 사용이 가능함을 제시하였다.

감사의 글

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

참고문헌

[1] C. W. Nho, H. Y. Kim: Convergence Research Review, Convergence Research Center, *Technical Report*, Vol4. No.3, March, 2018.

[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] J. R. Raol, *Multi-Sensor Data Fusion with MATLAB,* Baton Rouge: CRC Press, 2010.

[9] 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.

[10] L. Donghyun, “LSTM RNN-based Korean Speech Recogniti

on System Using CTC," *Journal of Digital Contents Society*, Vol. 18, No. 1, pp. 93-99, 2017.
[11] N. Van Quan, "LSTM-based Anomaly Detection on Big Data for Smart Factory Monitoring," *Journal of Digital Contents Society*, Vol. 19, No. 4, pp. 789-799, 2018.

권기현(Ki-Hyeon Kwon)

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

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

김진수(Jin-Su Kim)

2017년 : 강원대학교 전자정보통신공학부 정보통신공학전공(학사)
2017년 ~ 현 재: 강원대학교 전자정보통신공학과(석사과정)

※관심분야 : 패턴 인식 (Pattern Recognition), 정보보안, LSTM

이형봉(Hyung-Bong Lee)

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

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