

기계학습 기반의 클라우드를 위한 센서 데이터 수집 및 정제 시스템

황치곤¹ · 윤창표^{2*}

Sensor Data Collection & Refining System for Machine Learning-Based Cloud

Chi-Gon Hwang¹ · Chang-Pyo Yoon^{2*}

¹Invited Professor, Dept. of Computer Engineering, IIT, Kwangwoon University, Seoul, 01897 Korea

^{2*}Assistant Professor, Dept Of Computer & Mobile Convergence, GyeongGi University of Science and Technology, Siheung-si, 15073 Korea

요 약

기계학습은 최근 대부분의 분야에서 적용하여 연구를 하고 있다. 이것은 기계학습의 결과가 결정된 것이 아니라 입력데이터의 학습으로 목적함수를 생성하고, 이를 통해 통하여 새로운 데이터에 대한 판단이 가능하기 때문이다. 또한, 축적된 데이터의 증가는 기계학습 결과의 정확도에 영향을 미친다. 이에 수집된 데이터는 기계학습에 중요한 요인이다. 제안하는 본 시스템은 서비스 제공을 위한 클라우드 시스템과 지역의 포그 시스템의 융합 시스템이다. 이에 클라우드 시스템은 서비스를 위한 머신러닝과 기반 구조를 제공하고, 포그 시스템은 클라우드와 사용자의 중간에 위치하여 데이터 수집 및 정제를 수행한다. 이를 적용하기 위한 데이터는 스마트기기에서 발생하는 센서 데이터로 한다. 이에 적용된 기계학습 기법은 분류를 위한 SVM 알고리즘, 상태 인지를 위한 RNN 알고리즘을 이용한다.

ABSTRACT

Machine learning has recently been applied to research in most areas. This is because the results of machine learning are not determined, but the learning of input data creates the objective function, which enables the determination of new data. In addition, the increase in accumulated data affects the accuracy of machine learning results. The data collected here is an important factor in machine learning. The proposed system is a convergence system of cloud systems and local fog systems for service delivery. Thus, the cloud system provides machine learning and infrastructure for services, while the fog system is located in the middle of the cloud and the user to collect and refine data. The data for this application shall be based on the Sensitive data generated by smart devices. The machine learning technique applied to this system uses SVM algorithm for classification and RNN algorithm for status recognition.

키워드 : 클라우드 시스템, 포그 시스템, 기계학습, SVM, RNN

Keywords : Cloud system, Fog system, Machine learning, Support vector machine, Recurrent neural network

Received 29 May 2020, Revised 23 June 2020, Accepted 29 June 2020

* Corresponding Author Chang-Pyo Yoon(E-mail: cpyoon@gtec.ac.kr, Tel:+82-31-496-6410)

Assistant Professor, Dept Of Computer & Mobile Convergence, GyeongGi University of Science and Technology, Siheung-si, 15073 Korea

Open Access <http://doi.org/10.6109/jkiice.2021.25.2.165>

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. 서론

기계학습이란 학습 데이터를 바탕으로 최적의 해를 구하기 위한 목적함수를 구함으로써 입력데이터에 대한 분류, 추측, 예상하는 것이다[1]. 이러한 기계학습은 학습 데이터 집합의 유형에 따라 지도학습, 비지도 학습으로 분류된다. 그 외에도 강화학습과 반지도 학습 등이 있다. 그 외에 기계학습은 다양한 방식으로 발전하고 있다. 이 분야는 최근 들어 정보기술 분야뿐만 아니라 많은 분야에서 관심을 가지고 연구가 진행되고 있다.

기계학습은 대부분 학습 데이터의 크기가 크면 클수록 학습된 결과의 정확도는 상승한다. 그러나 학습 데이터의 크기는 학습시간에 따른 성능의 문제와 과적합의 문제가 발생하기 때문에 무조건 학습 데이터의 크기를 크게 할 수는 없다[2]. 이러한 기법은 분류문제, 회귀문제들을 해결할 수 있고, 영상이나 음성 같은 시간의 변화에 따른 특성의 변화를 가지는 시간적 데이터에도 적용되고 있다. 주로 분류와 회귀분석을 위해 사용하는 SVM이라는 기법이 있다[3]. 그리고 시간적 특성을 고려한 RNN이 있으며, 이 알고리즘의 단점을 보완한 LSTM 기법이 있다[4].

본 논문에서는 클라우드 시스템과 포그 시스템을 이용한 사용자의 센서 데이터를 수집하여 클라우드에서 머신러닝 기법을 이용하여 목적함수를 도출하고 이를 통해서 포그에서 수집된 데이터를 정제하여 상태를 인지하는 시스템을 제안하고자 한다.

본 논문의 구성은 2장에서 클라우드 시스템과 기계 학습에 대해 살펴보고, 3장에서 본 시스템의 구성과 알고리즘에 관해서 기술하고, 4에서 평가를 기술한다. 그리고 5장은 결론을 기술한다.

II. 관련 연구

2.1. 클라우드 컴퓨팅(Cloud Computing)

포그 컴퓨팅은 기존의 클라우드 컴퓨팅 기법을 엣지 네트워크로 확장한 모델이다[5]. 클라우드 컴퓨팅은 인터넷을 클라우드로 인식하고, 클라우드 내부에 가상화된 컴퓨터의 시스템 자원을 요구에 따라 제공하는 온디맨드(on-demand) 방식의 컴퓨팅 기술로, 인터넷 기반 컴퓨팅의 일종으로 정보를 자신의 컴퓨터에 있는 자원

이 아닌 클라우드 상에 연결된 다른 컴퓨터로 처리하는 기술이다. 그러나 점차 다양하고 대용량인 데이터를 클라우드 시스템으로 저장하고, 처리함으로써 많은 사용자에 대한 연결 대기시간 증가와 서버 과부하 문제가 발생하게 되었다[6]. 이를 해결하기 위해 포그 컴퓨팅이 제안되었으며, 포그 컴퓨팅은 기본적으로 해당 지역 또는 연결된 영역 내의 데이터 수집, 처리, 필터링, 분석 등의 작업을 부분적으로 처리하여 일반 사용자와 클라우드 컴퓨팅 시스템 사이에 존재하여, 분산처리와 클라우드 서버의 서버 과부하 문제와 연결 대기시간 문제를 해결하였다. 본 논문에서 클라우드와 포그 시스템은 적용 환경이 된다.

2.2. 기계학습(Machine Learning)

기계학습은 기계가 스스로 학습 데이터를 학습하여 새롭게 수집하는 데이터에 대해 적절한 서비스를 제공하는 기술 중 하나이다[1]. 이러한 기계학습도 다양한 기법들이 연구되고 있으며, 이중 시한성과 관계된 기법들은 분류와 회귀분석을 위한 SVM[3], 시한성에 따른 상태의 변화를 측정하기 위한 RNN, LSTM 같은 기법이 있다[4].

Vapnik에 의해 제안된 서포트 벡터 머신(SVM)은 패턴 인식 및 자료 분석에 활용되는 지도 학습(supervised learning) 모델이며, 선형이나 비선형 분류, 회귀, 이상치 탐색에도 사용할 수 있는 기계학습 모델이다. 두 카테고리 중 어느 하나에 속한 데이터의 집합이 주어졌을 때, SVM 알고리즘은 주어진 데이터 집합을 바탕으로 하여 새로운 데이터가 어느 카테고리에 속할지 판단하는 비확률적 이진 선형 분류 모델을 만든다. 만들어진 분류 모델은 데이터가 사상된 공간에서 경계로 표현되는데 SVM 알고리즘은 그 중의 가장 큰 폭을 가진 경계를 찾는 알고리즘이다. 새로운 실험 데이터의 추가에 대하여 소속될 집단을 결정하기 위해 분류 모델을 만들고, 이 분류 모델은 지지 벡터 또는 지지 평면이라고 하며, 이 모델은 소프트마진(soft margin)을 최대화하는 모델이다[7-9]. 기본적인 아이디어는 그림 1과 같다. 이에 따른 선형 SVM 예측을 위한 목적함수 $d(x)$ 를 기반으로 라그랑주 승수를 이용한 비선형 SVM 예측 알고리즘은 수식 2와 같고, 이를 통해 학습 데이터를 분류한다[10].

$$d(x) = \sum_{\alpha_k \neq 0} \alpha_k y_k K(x_k, x) + b \quad (1)$$

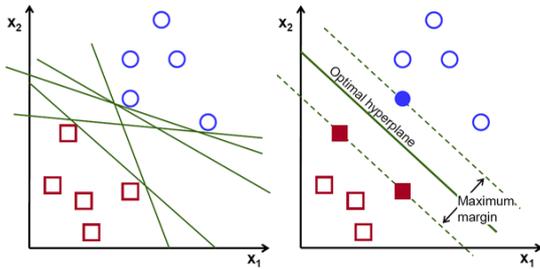


Fig. 1 linear SVM model.

순환신경망(RNN)[11,12]은 시간에 따라 상태가 동적으로 변동되는 가변적인 길이를 가지는 데이터를 효과적으로 처리하기 위한 기법이다. 이러한 특성으로 인해 시계열 데이터와 같은 순차성을 갖는 데이터 처리에 효과적이고, 데이터의 시간성, 가변 길이, 문맥 의존성이라는 특징을 가진다.

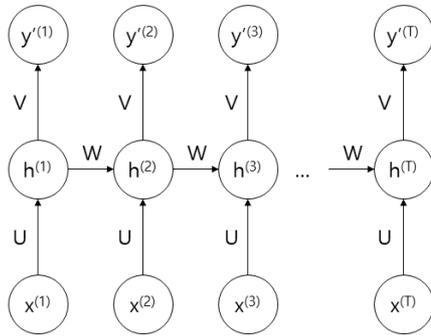


Fig. 2 structure of RNN

그림 2는 RNN의 구조로 $t(1, 2, 3, \dots, T)$ 의 순간에 입력된 $x(t)$ 와 직전 순간에 은닉층 값 $h(t-1)$ 에 따라 결정되며, T 번 반복했음을 의미한다. 이에 사용되는 매개변수

Θ 는 U, V, W 에 해당한다. U 는 입력층의 데이터를 은닉층으로 입력하기 위한 가중치, W 는 은닉층과 출력층을 연결하는 가중치, V 는 은닉층과 출력층을 연결하는 가중치이다. 이때 순간 t 의 은닉층 값 $h(t)$ 는 입력값 $x(t)$ 와 직전 순간의 은닉층 값 $h(t-1)$ 에 따라 결정된다. 이에 따라 t 시간의 은닉층을 구하는 함수는 수식 2과 같다.

$$h^{(t)} = f(h^{(t-1)}, x^{(t)}; \Theta) \quad (2)$$

그러나, RNN은 멀리 떨어져 있는 데이터에 대한 장기 의존성이 떨어져 거리가 먼 과거 데이터에 대한 영향력이 점차 줄어들거나 소실되는 문제를 가진다. 이 문제를 극복하기 위한 가장 효과적인 해결책이 LSTM(Long Short-Term Memory)이다[4]. LSTM은 기울기 소실 문제를 해결하기 위해 개발되었다. 그러나, 제안하고자 하는 시스템에서는 시간성을 가지는 데이터, 센서로부터 수집된 데이터를 통해서 보행자의 전복 여부를 확인하고자 하는 경우는 시간의 변화에 따른 특성을 이용해야 한다. 장기간의 데이터가 아니라 직전의 상태와 현재의 상태에 따른 변화로 상태를 결정하는 것이기 때문에 RNN 기법을 적용하는 것이 적합하다.

III. 제안시스템

본 논문에서 제안하는 시스템은 클라우드 시스템과 포그 시스템을 이용한 하이브리드 시스템을 기반으로 대상자들의 스마트 기기에 장착된 센서에서 발생하는 데이터를 수집하여 사용자들의 상태를 기계학습으로 수행된 결과를 통지하는 시스템이다.

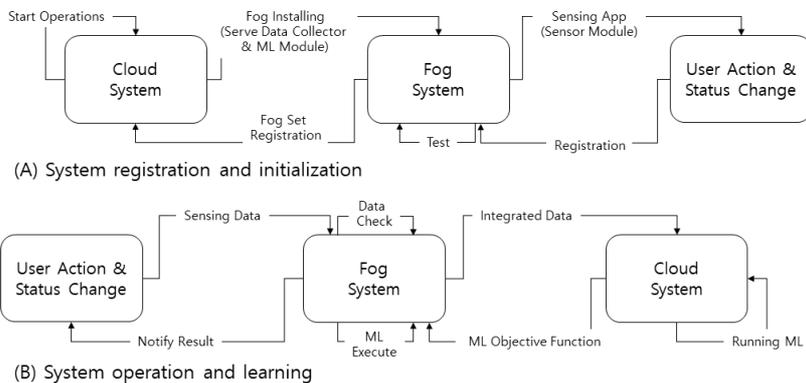


Fig. 3 System Initialization and Operation Scenario

그림 3은 제안하는 시스템을 위한 세부적인 시나리오로써 시스템 등록(그림 3의 (A))과 초기화하는 부분과 시스템 운용을 위한 부분(그림 3의 (B))으로 구성되어 있다. (A)는 클라우드 시스템으로부터 시작되며, 포그에 시스템의 시작을 위한 설정, 사용자 등록, 사용자 상태 데이터의 수집 및 제공, 클라우드 시스템의 대리 역할을 위한 모듈을 수행하고, 수행된 결과를 사용자에게 통지한다. (B)는 운영과정의 시나리오로 클라우드는 학습 데이터를 통해 목적함수를 생성하고, 생성된 목적함수는 포그에 설치되며, 포그는 수신된 사용자의 상태에 따른 결과를 생성하여 사용자에게 통지하고, 상태 데이터와 결과는 클라우드에 축적하여 기계학습을 위해 활용한다.

3.1. 포그 시스템

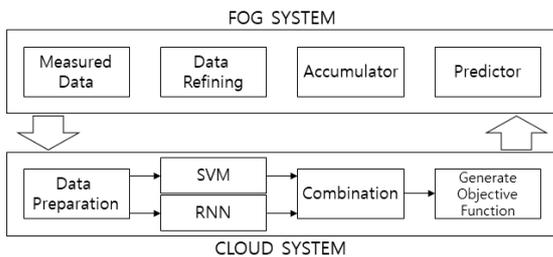


Fig. 4 Relationship between Fog System and Cloud System

클라우드 시스템과 포그 시스템 간의 데이터 교환은 그림 4와 같이 수행되고, 포그 시스템은 사용자가 가진 스마트 기기의 가속도 센서와 자이로 센서를 통해 데이터를 수집, 정제 및 축적을 통해 클라우드 시스템에 제공한다. 그리고 클라우드 시스템에서 제공하는 목적함수를 기반으로 발생하는 신규 데이터에 대하여 사용자의 상태를 인지하고 이 데이터도 클라우드에 제공하여 학습을 위한 기반 데이터가 된다. 이에 포그 시스템의 각 구성요소는 다음과 같다.

- **Measured Data:** 사용자 스마트 기기로부터 검출된 정보를 수집한다.
- **Data Refining:** 수집된 데이터를 정제한다. 데이터 규모의 차이에 따른 정규화, 명칭값에 대한 One-hot 코딩에 대한 적용으로 수행한다.
- **Accumulator:** 수집하여 정제된 데이터, 목적함수를

통해 인지된 결과를 클라우드 시스템에 제공하기 위하여 축적한다.

- **Predictor:** 클라우드 시스템에서 제공된 목적함수를 통하여 사용자의 상태를 인지하여 통지하는 역할을 한다.

Table. 1 example of training data

| timestamp | accX | accY | accZ | gyX | gyY | gyZ | y |
|----------------------------|---------|---------|--------|---------|---------|---------|----|
| ... | | | | | | | |
| 2020-05-08 19:41:28:133 | 9.9958 | 19.6127 | 5.3672 | 0.4293 | 1.1931 | -0.3334 | 01 |
| 2020-05-08 19:41:28:153 | 3.3477 | 7.6034 | 5.1547 | 1.1814 | -0.3526 | -1.3486 | 01 |
| 2020-05-08 19:41:28:172 | 3.3477 | 7.6034 | 5.1547 | 1.1814 | -0.3526 | -1.3486 | 01 |
| ... | | | | | | | |
| 2020-05-08 19:41:34:733 | -2.5199 | 2.3128 | 2.3691 | -0.2503 | 0.7329 | -0.0234 | 01 |
| 2020-05-08 19:41:34:752 | -0.5770 | 5.6581 | 5.6814 | -0.2503 | 0.7329 | -0.0234 | 10 |
| 2020-05-08 19:41:34:772 | -0.5770 | 5.6581 | 5.6814 | 2.3255 | -0.9129 | 0.3728 | 10 |
| ... | | | | | | | |

이 클라우드 시스템에서 기계학습의 수행을 위한 원시 데이터는 스마트 기기의 가속도 센서와 자이로 센서를 이용하여 데이터를 추출하며, 추출한 데이터의 형태는 table. 1과 같다. 이 데이터들은 상태의 분류를 위한 SVM의 목적함수를 위한 입력데이터이다.

입력 인자로 이용되는 것은 측정된 시간을 의미하는 timestamp, 가속도 센서의 값(accX, accY, accZ), 자이로 센서값(gyX, gyY, gyZ), 측정 당시의 상태인 y가 있다. 상태 값 y는 정지상태, 보행상태, 넘어진 상태 및 누운 상태로 구분하였고, 이를 One-hot 코딩 기법으로 각각 00, 01, 10, 11로 변환하였다.

3.2. 클라우드 시스템

그림 4의 클라우드 시스템은 포그 시스템에서 수집된 데이터를 기반으로 기계학습을 통해 목적함수를 생성하고, 생성된 목적함수를 포그 시스템에 제공하는 것이 클라우드 시스템의 역할이다. 클라우드 시스템의 각 구성요소는 다음과 같다.

- **Data Preparation:** 포그 시스템에서 전달되는 정제된

데이터를 머신러닝을 위해 준비 및 축적한다.

- SVM & RNN: 준비된 데이터를 기반으로 머신러닝을 수행하여 결과를 산출한다.
- Combination: 이전의 머신러닝 결과를 기반으로 결과 생성 및 목적함수를 위해 조합한다.
- Generate Objective Function: 목적 함수를 생성한다.

이러한 학습 데이터를 이용하여 클라우드 시스템에서 기계학습을 수행한다. 수행하는 알고리즘은 비선형 SVM과 RNN 기법을 이용한다. 두 가지 알고리즘을 이용하는 이유는 다음과 같다. SVM은 특정 상태에 대한 분류에 장점이 있어 시간성에 대한 인지가 부족하다. RNN은 이에 비해 시간성, 가변 길이, 문맥 의존성을 가지는 순차 데이터의 예측이나 인지에 효율적이다. 이런 이유로 두 가지 머신러닝 알고리즘을 이용했고, 두 알고리즘의 결과를 조합하여 사용자의 상태를 인지한다. 두 번째 알고리즘은 사용자의 상태 변화에 따른 상태를 인지하기 위하여 RNN을 적용한다. 사람의 행위는 멈춰있는 상태가 아니라 지속적으로 변한다. 즉, 사람은 이전 상태와 현재 상태에 따라서 현재 어떤 상태인지 인지할 수 있다. 이 두 개의 결과는 Combination 요소를 이용하여 조합을 통하여 사용자의 상태를 시간적, 상태적 인지를 가능하게 한다.

IV. 평가

본 논문은 클라우드 시스템과 포그 시스템의 상호운용을 위해 기계학습 기법을 이용하여 사용자의 상태를 파악하기 위한 시스템을 제안하였다. 이에 대한 적용 사례로 사용자의 보행상태를 학습 데이터로 이용하였다. 이러한 사례를 기반으로 적용한 것을 실제 상황과 인지한 상황에 대한 테스트 데이터를 기반으로 비교를 수행하였다. 실제 행위와 행위에 대한 인지 여부를 시간의 순서에 따라 측정된 데이터를 기반으로 수행한 결과 그림 5와 같다.

그림에서 실선은 실제 측정된 데이터를 기반으로 구성되었으며 실제의 상황을 레이블로 표기하였다. 점선은 측정된 데이터를 기반으로 제안한 방식으로 운영하였을 때 산출된 상태이다. 상태값은 0, 1, 2, 3이며, 0은 정지상태, 1은 보행상태, 2는 넘어진 상태, 3은 누운 상

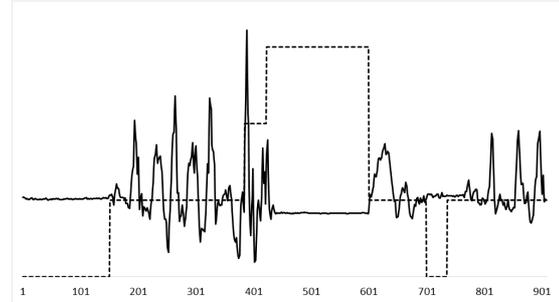


Fig. 5 Comparison of the relationship between measurement data and cognition

태이다. 이렇게 측정하였을 때, 실제의 상태와 인지된 상태 사이의 간격이 전체 측정 공간에 비해 1.6% 이내가 나왔다. 이 오류는 상태의 대상인 누운 상태에서 일어나는 상태까지의 간격이며, 이러한 부분은 더 상세하게 분류할 필요가 있다.

V. 결론

본 논문은 서비스의 효율적인 제공을 위하여 클라우드 시스템과 포그 시스템을 혼용하는 하이브리드 방식을 채택하였다. 이러한 방식은 클라우드 시스템에 집중되는 부하를 방지하기 위함이다. 이에 사용자의 상태를 효율적으로 판단하기 위해 기계학습 기법 중 SVM과 RNN을 이용한 인지 시스템을 구축하였다. 이는 스마트폰에서 측정된 센서 데이터를 이용하여 사용자의 상태를 인지하여 사용자와 이해관계자에게 알림 서비스를 수행하기 위해 연구되었으며, 센서를 기반으로 하는 클라우드 시스템을 이용하여 헬스케어 서비스의 많은 분야에 적용할 수 있다.

References

[1] B. Yoshua, "Learning deep architectures for AI," *Foundations and trends in Machine Learning*, vol. 2, no. 1, pp. 1-127, 2009.

[2] R. Sheikhpour, M. A. Sarram, S. Gharaghani, and M. A. Z. Chahooki, "A Survey on semi-supervised feature selection methods," *Pattern Recognition*, vol. 64, pp. 141-158, Apr. 2017.

- [3] A. M. Abd and S. M. Abd, "Modelling the strength of lightweight foamed concrete using support vector machine (SVM)," *Case studies in construction materials*, vol. 6, pp. 8-15, 2017.
- [4] A. Sherstinsky, "Fundamentals of Recurrent Neural Network (RNN) and Long Short-Term Memory (LSTM) network," *Physica D: Nonlinear Phenomena*, vol. 404, 2020.
- [5] M. Aazam and E. N. Huh, "Fog Computing and Smart Gateway Based Communication for Cloud of Things," *International Conference on Future Internet of Things and Cloud(FiCloud) 2014, International Conference on IEEE*, pp. 464-470, Dec. 2014.
- [6] S. Singh and I. Chana, "A survey on resource scheduling in cloud computing: Issues and challenges," *Journal of grid computing*, vol. 14, no. 2, pp. 217-264, 2016.
- [7] Y. D. Cai, P. W. Ricardo, C. H. Jen, and K. C. Chou, "Application of SVM to predict membrane protein types," *Journal of theoretical biology*, vol. 226, no. 4, pp. 373-376, 2004.
- [8] D. Zhang, "Support Vector Machine," in *Fundamentals of Image Data Mining*, Springer International Pub., ch. 8, pp. 179-205, 2019.
- [9] M. N. Murty and R. Raghava, "Kernel-based SVM," *Support vector machines and perceptrons*, Springer, pp. 57-67, 2016.
- [10] Understanding SVM[Internet]. Available: https://opencv-python-tutorials.readthedocs.io/en/latest/py_tutorials/py_ml/py_svm/py_svm_basics/py_svm_basics.html.
- [11] S. Li, W. Li, C. Cook, C. Zhu, and Y. Gao, "Independently recurrent neural network (indrnn): Building a longer and deeper rnn," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 5457-5466, 2018.
- [12] G. Y. Lim and Y. B. Cho, "Dynamic RNN-CNN malware classifier correspond with Random Dimension Input Data," *Journal of the Korea Institute of Information and Communication Engineering*, vol. 23, no. 5, pp. 533-539, May. 2019.



황치곤(Chi-Gon Hwang)

2012년 광운대학교 컴퓨터과학과 (공학박사)
2006년~2015년:(주)인찬 연구원
2016년~2018년: 경민대학교 인터넷정보과 교수
2019년~현재: 광운대학교 정보과학교육원 컴퓨터공학전공 교수
※관심분야: 모바일 클라우드, 온톨로지, 기계학습, NLP



윤창표(Chang-Pyo Yoon)

2012년 : 광운대학교 컴퓨터과학과 (공학박사)
2012년~현재: 경기과학기술대학교 컴퓨터모바일융합과 교수
※관심분야: 기계학습, 모바일 시스템, 네트워크 보안, 무선 네트워크, 온톨로지