

Energy-Efficient Sensor Calibration Based on Deep Reinforcement Learning

Akm Ashiquzzaman¹, Tai-Won Um², Jinsul Kim³

^{1,3} School of Electronics and Computer Engineering, Chonnam National University,
Gwangju, South Korea,

² Department of Information and Communication Engineering, Chosun University,
Gwangju, South Korea,

¹ zamanashiq3@gmail.com, ² twum@chosun.ac.kr, ³ jsworld@jnu.ac.kr

Abstract

The current development of IoT based sensor networks opens up a new era of network communication with a wide range and diverse usage. Sensor calibration to reduce power usage is essential to reduce energy consumption in sensors as well as improve the efficiency of devices. Reinforcement learning (RL) has been received much attention from researchers and now widely applied in many study fields to achieve intelligent automation. Though various types of sensors have been widely used in the field of IoT, rare researches were conducted in resource optimizing. In this novel research, a new style of power conservation has been explored with the help of RL to make a new generation of IoT devices with calibrated power sources to maximize resource utilization. Our proposed model using Deep Q learning (DQN) enables IoT sensors to maximize its resource utilization. This research focuses solely on the energy-efficient sensor calibration and simulation results show that the performance of the proposed method. This proposed model achieve a new state of the arts 96% accuracy for predicting and learning the game and give a novel solution for efficient sensor calibration.

Keywords: Deep learning, Optimization, Reinforcement Learning, Simulation.

1. Introduction

Internet of Things (IoT) has wide applications and crucial in our daily life in every sector. Nowadays, IoT devices can be used to monitor and detect activities in real-time. IoT based sensor networks are distributed in which each sensor does its own tasks to sense and obtain information from the environment. However, sensors have limited functions such as short in battery lifetime, low range communication. Rapid growth in the demand for energy in every sector such as communication, industrial production, domestic usage, etc. makes now the efficient usage of energy resources a very demanding research interest. IoT based on small scale machines with limited power resources. The idea of saving energy during operation IoT devices is one of the top priorities. The energy consumption of under-utilized resources, particularly in a remote IoT based testbed, measures for a significant amount of real energy use. So, as to utilize this usage will both make the testbed more robust and very longer lasting than usual. IoT devices are expected to be low-powered devices but the sheer number of IoT devices

Article history:

Received (April 13, 2019), Review Result (May 26, 2019), Accepted (September 2, 2019)

that Cisco predicts will be 50 billion by 2020, is an order of magnitude larger than the number of smartphones and tablets in use today [1]. According to Mohammadi et al. [2], the annual estimated market growth or economic impact of IoT Devices will be over \$2.7 trillion dollars. Rise of IoT usage trending will eventually render into many sectors of industries to adapt to such a system, resulting in more energy surge in the next decade. As a result, the research into making more energy effective solutions are on the rise.

Reinforcement Learning (RL) algorithm is not a modern-day invention. The whole idea of these algorithms hovers around natural human learning mechanisms. The main idea is to encourage positive action through some positive value, or reward and vice versa. This gives the agent or the actor/model to adapt to the real-life model scenario easily by learning rapidly. Previously, the RL algorithms were notoriously slow and not easily adapted to the environment. However, after the rise of Deep Learning Neural Networks (DNN) and backpropagation gives the DNN based agent to perform well RL based learning. The agent, which essentially, a DNN with learning capacity gets weight update based on the RL policy can rapidly learn the mechanism defined in the environment and suppress any human in the same situation. Google Inc. developed the deep learning-based Reinforcement model which mastered the ancient game of Alphago and beat the human professional player [3]. This paves a way to apply RL in many sectors of the problem that cannot be optimized only by the classical machine learning models. Although much research has been focused to solve complex problems with reinforcement learning, the energy-saving mechanism of the IoT devices with the reinforcement model has not been properly explored. The idea of resource utilization in IoT devices is crucial for any real-time sensor field. The resource scarcity will make this reinforcement learning very effective to the IoT based model to adapt to longer-lasting extreme scenarios.

Deep reinforcement learning (deep RL) has emerged as an effective solution for learning how to mimic decision flows of very complex real models [4]. Most of the applications of such models are mainly focused on game-based problem-solving. The energy utilizing a section for IoT sensors with deep reinforcement learning is a promising sector for such a domain to apply reinforcement learning solutions. Although much research focused on visual puzzle solving, using the deep reinforcement learning in the IoT sensors to utilize the energy resources is not properly explored. In this research, the idea of the kind of power consumption minimization with deep reinforcement learning for IoT sensors is focused. Minimizing battery usage with a reinforcement learning mechanism to maximize the battery output is mainly focused.

The rest of the paper is organized as follows, Section II discusses the related works and proper background of the theory, Section III explains the setup of the theorized experiment, Section IV discussed the findings and final Section V is the conclusion. A state is a concrete and immediate situation in which the agent finds itself.

2. RELATED WORK

Power consumption minimization and intelligent resource management had seen much research. Most of the modern research in this field specifically focused on resource utilization with most various machine learning techniques. The research mostly focused on the implementation of workload in various sectors such as data center load distribution to reduce workloads, electric power station intelligent switching to reduce power consumption and energy conservation. To our best knowledge, the novel research for utilizing IoT sensors and power conservation with reinforcement learning has not yet been thoroughly explored.

In the research conducted by Dong et al. [5], the main research focused on applying the support vector machine (SVM), which is a well know supervised machine learning algorithm

for classifying. The SVM was applied to predict the energy consumption of the building in the tropical regions. The research collected the weather data of the said regions and buildings then trained the SVM to predict proper energy consumption for the given weather in the future. The research reported y, all prediction results are found to have coefficients of variance (CV) less than 3%.

According to Shroud et al. [6], the cost of energy is one of the important factors associated with increased production costs at manufacturing facilities. This enables the decision-maker to approach this problem with at-most importance. The research proposes a mathematical model to minimize energy consumption by using machine learning methods to intelligently "Turn on" or "Turn off" the machine. This method will eventually reduce the machine energy consumption drastically. The research reports that significant reductions in energy costs can be achieved by avoiding high-energy price periods. This minimization process also has a positive environmental effect by reducing energy consumption during peak periods, which increases the possibility of reducing CO₂ emissions from power generator sites.

Mocanu et al. [7] proposed a new reinforcement learning, especially Deep Q learning-based online building energy optimization technique which energy scheduling strategies could be used to provide real-time feedback to consumers to encourage more efficient use of electricity. However, this research completely focused on the learning mechanisms of the energy optimization of the electric grids and do not give any insights on battery-based energy optimization.

All the research currently going on the field of energy conversion, and IoT based models energy optimization and implementation of the energy-saving method is currently not properly explored. The IoT based models need to be highly optimized for minimum energy consumption and the proposed method for this power consumption minimization will surely give a new frontier in the IoT based device power optimization with reinforcement learning.

3. PROPOSED METHODS

3.1. Reinforcement Learning and Neural Network

The main idea of reinforcement learning first coined in computer science by Sutton in 1998 [8]. It is separated into some sectors of definitions. RL can be properly explained using the concepts of agents, environments, states, actions, and rewards. An agent takes actions; for example, a drone making a delivery, or Super Mario navigating a video game. Action is the set of all possible moves the agent can make. The action is almost self-explanatory, but it should be noted that agents choose among a list of possible actions. The world through which the agent moves is called the environment. The reward is the feedback by which we measure the success or failure of an agent's actions. The discount factor is multiplied by future rewards as discovered by the agent in order to dampen these rewards effect on the agent's possible choice of action sets. It is designed to make future rewards worth less or more, depending on the settings, than immediate rewards. So basically, it enforces a kind of short-term hedonism or long term feedback or rewarding in the agents. The main interaction is done by the agent in the environment.

Based on the agent's action it will receive some reward. This whole action taking in the single unit is known as the state, as it moves into the next state, it gets to repeat the same action until the objective for the agent is done. Deep learning can be classified as a subset of machine learning. Usually, the term deep learning is synonymous with deep artificial neural networks. Deep artificial neural networks are a set of algorithms that

have set new records in accuracy for many important problems, such as image recognition, sound recognition, recommender systems, etc. [2][9][10]. Deep is a machine learning terminology. It refers to the number of layers in a neural network. A shallow network has one so-called hidden layer, and a deep network has more than one. Multiple hidden layers allow deep neural networks to learn the features of the data.

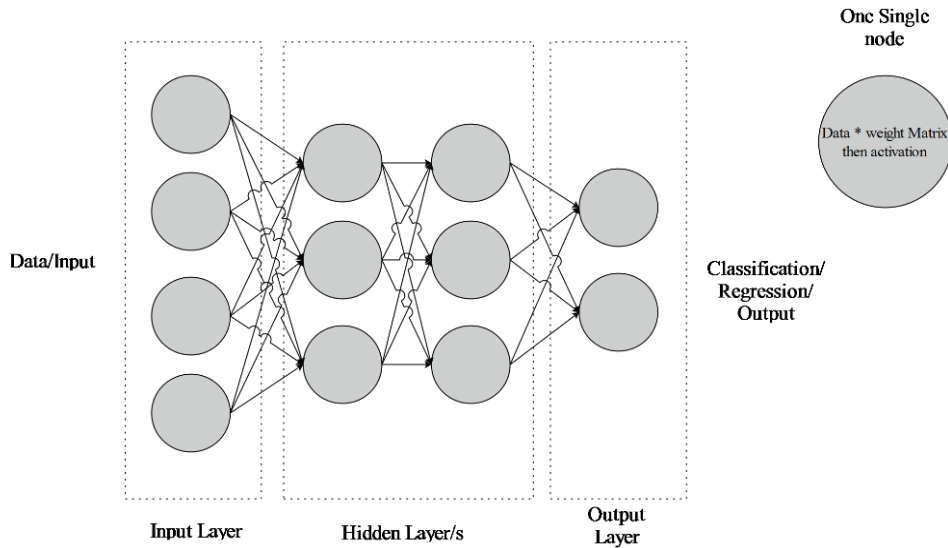


Figure. 1. Deep Learning Neural Network.

As shown in Figure 1, the main idea of the deep learning neural networks is in the multi-layered, actually more than one hidden layer in the model. These multi-layered networks are often configured based on the needs of classification and regression. The detailed model of our proposed DNN was discussed in the later experiment section in detail. The other action of the deep learning works as same as the Multilayer Perceptron (MLP). The main power of DNN came from the rise of the backpropagation algorithm [11] and GPU based training of the model in huge datasets [12].

3.2. Deep Q learning

Q-learning is a reinforcement learning technique used in machine learning. The goal of Q-Learning is to learn a policy, which tells an agent what action to take under what circumstances. It does not require a model of the environment and can handle problems with stochastic transitions and rewards, without requiring adaptations. For any finite Markov decision process (FMDP), Q-learning finds a policy that is optimal in the sense that it maximizes the expected value of the total reward overall successive steps, starting from the current state. Q names the function that returns the reward used to provide the reinforcement and can be said to stand for the quality of an action taken in a given state. In value-based model-free reinforcement learning methods, the action-value function is represented using a function approximation, such as a neural network. In one-step Q-learning, the parameters of the action-value function are learned by iterative minimization of a sequence of loss functions, which is essentially the backbone of any DNN.

Algorithm 1. Deep Q learning Algorithm.

Q -learning: Learn function $Q : \mathcal{X} \times \mathcal{A} \rightarrow \mathbb{R}$

Require:
 States $\mathcal{X} = \{1, \dots, n_x\}$
 Actions $\mathcal{A} = \{1, \dots, n_a\}$, $A : \mathcal{X} \Rightarrow \mathcal{A}$
 Reward function $R : \mathcal{X} \times \mathcal{A} \rightarrow \mathbb{R}$
 Black-box (probabilistic) transition function $T : \mathcal{X} \times \mathcal{A} \rightarrow \mathcal{X}$
 Learning rate $\alpha \in [0, 1]$, typically $\alpha = 0.1$
 Discounting factor $\gamma \in [0, 1]$

procedure QLEARNING($\mathcal{X}, A, R, T, \alpha, \gamma$)
 Initialize $Q : \mathcal{X} \times \mathcal{A} \rightarrow \mathbb{R}$ arbitrarily
while Q is not converged **do**
 Start in state $s \in \mathcal{X}$
 while s is not terminal **do**
 Calculate π according to Q and exploration strategy (e.g. $\pi(x) \leftarrow \arg \max_a Q(x, a)$)
 $a \leftarrow \pi(s)$
 $r \leftarrow R(s, a)$ ▷ Receive the reward
 $s' \leftarrow T(s, a)$ ▷ Receive the new state
 $Q(s', a) \leftarrow (1 - \alpha) \cdot Q(s, a) + \alpha \cdot (r + \gamma \cdot \max_{a'} Q(s', a'))$
 $s \leftarrow s'$
return Q

The basic algorithm is described in Algorithm 1. above. Most of the implementation of the DQN, despite the programming specification, follow the algorithm. For any IoT devices, the main part of the computation unit is small and resource-intensive. So the actor, which is a deep learning model needs to be much utilized to use less power and computation during training. So, in the simulated scenario, the environment was proposed as supposed to ensure proper learning. The main component of the environment of the Action detection sensor, in the proposed method, such a sensor was imagined as the action detection or motion detection sensor, for both Ultrasonic and Ultraviolet light sensors. The IoT devices or the actor in this model reads the data of the sensors in the time-step and action control of the model, as to figure the given probability of its active state. Based on the time, the model gives the estimation for turning off the sensors based on the values on sensors and then it compares with the actual value of success or not. Based on the success of the model, the actor model gets a reward and thus learns to save the utilities on how much it conserves the energy. Later the battery data reading compared also add the reward policy to the basic learning as a discount factory randomly to tighten the immediate reward.

To train the DQN models, a starting point an off-line database is used at first. A simple simulation database with data was created to build the environment as a game. In this game for each step, a sample action space was given for the agent to take in a Gaussian distribution, followed by the battery charge level. We gave a positive reward to it if it was close to our optimization goal. If not, we assigned to that possibility a negative reward. At the begin of the learning there are a lot of random choices, but in time (many iterations), the reinforcement learning model converges and will learn to choose just the possibilities which are close to the optimization goal. The training of the actor was properly done in the simulated environment, but the data learned from this simulation is fully optimized for real deployment as the learned model act indifferent to the real-life because of the data invariant. The main idea of the agent is to learn from the sensor's electric usage and power consumption for battery discharge rates.

4. EXPERIMENTS

The simulation and training were done in the open-source, python-based simulator. The proposed Deep Q learning Agent (DQN) agent to learn to optimize the simple task of conserving energy by guessing the "turn off" states. In the generated result, the data is composed of a constant and repetitive profile of realistic data for any given sensors energy consumption, combined with a perfect forecast for a later time. The agent has both the ability to memorize this profile and lives in a near Markov environment. This results from an equilibrium of proper learning environments.

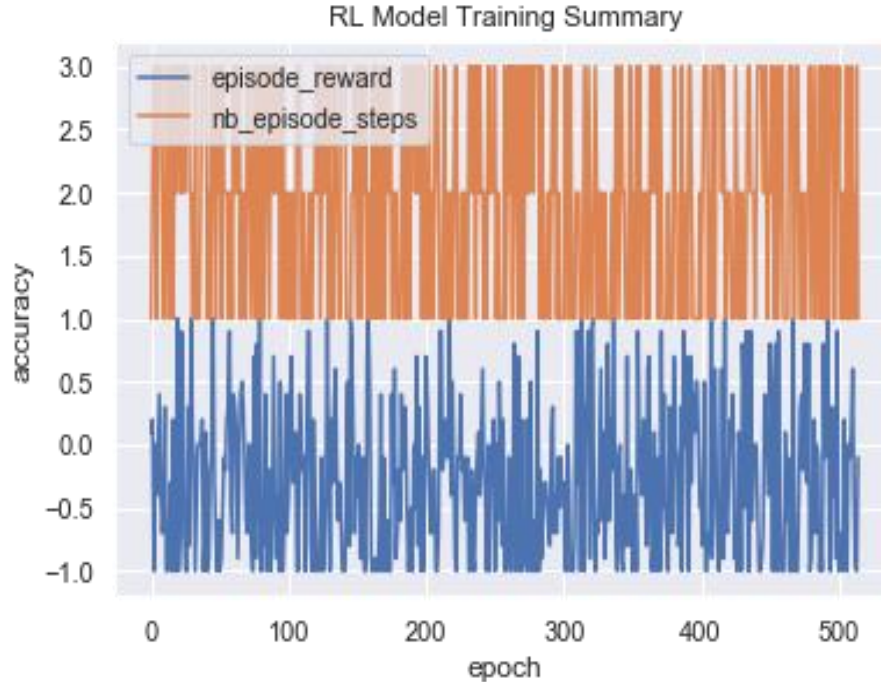


Figure. 2. Reinforcement Learning agent learning and given reward in each step taken by the agent.

Fig. 2 shows the graphical representation of the data generated for the RL training in the simulator. The trained model was learn based on the reward policy generated in each episode of the models. The total steps for the simulation were set at 1,000,000. The agent learning configuration was set as following, the discount factor or γ was set to 0.97 to look into a future reward for maximizing the conservation. Fig. 3 shows the total episode rewards in the total simulation run-time for an agent. Based on the saved energy and utilization, the model receives a reward that is shown in Fig. 3. A real-world application of using reinforcement learning to control a battery would have to deal with both a variable price profile. The main advantage of (deep) reinforcement learning is that after the model is trained completely on such an off-line database, it can be exploited on-line in a real environment. If this online environment has slightly changed from the trained environment, the reinforcement learning model can learn by default these changes and adapt dynamically its behavior to achieve the best possible performance.

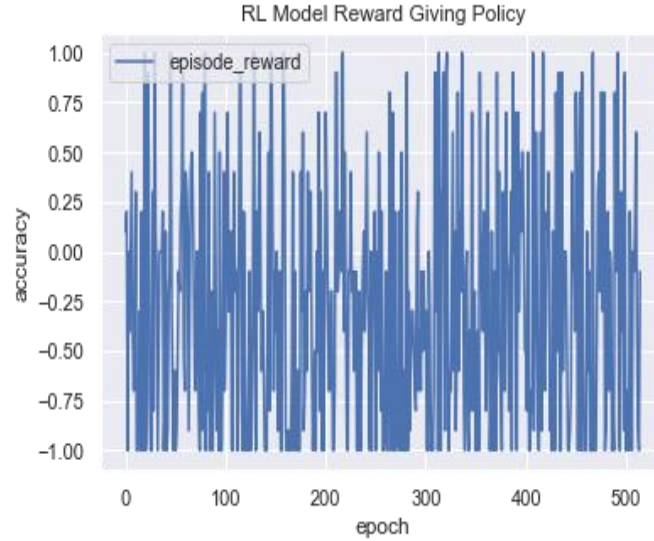


Figure. 3. Reward Policy for the agent in the simulator.

Table. 1. Learning accuracy comparison and baseline for Neural network models.

Method for the DQN Agent	Accuracy achieved	Remarks
Baseline (Just LSTM)	83.35%	Not setup with DQN, for baseline research
DNN DQN Agent	90%	Not adaptive to change in real-time
LSTM DQN Agent	96%	Has an adaptive scale for adding the time series learning.

The code of the DNN was modeled by was focused on a variety of combinations to ensure proper results. The neural network gets a basic 4 layers hidden configuration with 32,32,16,16 nodes in each layer. All of the layers had ReLu (Rectified Linear Unit). The learning rates for the neural network were fixed at 0.00001 for proper learning. The main general model was a simple neural network with 3 hidden layers, this layer started getting the changes and the model updating for recognizing the correct actions for a massive payout, or maximizing reward was excellent. This receives an over 90% accuracy for learning command sequences for maximizing reward. But the LSTM (Long Short Term Memory) model in DQN gets around 96% accuracy for the data system and very adaptive to the change in the rapid environment as shown in the Table 1. This is more important for the experiment and future application as the real-life data is also ever-changing in IoT based module.

5. CONCLUSION

Sensor utilization for power conservation is an important task to ensure the reliability and availability in controlling and monitoring a sensor network since they have limitations on battery and computational capacity. Though the work on sensor data for the new sensor network

generation is still cumbersome. For the best of our knowledge, our proposed method on sensor utilization employing RL algorithms is the first attempt to explore the application of machine learning in information utility-based sensor collaboration. In this research, we have generalized the estimation of an information utility for RL algorithms. Later we have designed a sensor and proposed a new model to utilize the battery or power consumption of the IoT node sensors. The whole simulation was successful indicative of the real-life RL deployment in the real-time IoT based devices later to enable RL based optimization later. The promising result opens up new research frontier for RL based IoT energy conservation and collaboration in real-time in future.

Acknowledgements

This work was supported by research fund from Chosun University and the National Research Foundation of Korea (NRF) grant funded by the Korea government (MSIT) (No.2018R1A2B2003774) and this research is also supported by the MSIT (Ministry of Science and ICT), Korea, under the ITRC (Information Technology Research Center) support program (IITP-2018-2016-0-00314) supervised by the IITP (Institute for Information & communications Technology Promotion).

References

- [1] J. Dofe, J. Frey, and Q. Yu, "Hardware security assurance in emerging iot applications," in 2016 IEEE International Symposium on Circuits and Systems (ISCAS), pp.2050-2053, IEEE, (2016). DOI:10.1109/ISCAS.2016.7538981
- [2] M. Mohammadi, A. Al-Fuqaha, S. Sorour, and M. Guizani, "Deep learning for iot big data and streaming analytics: A survey," IEEE Communications Surveys & Tutorials, (2018).
- [3] E. Gibney, "Google ai algorithm masters ancient game of go," Nature News, vol. 529, no. 7587, p. 445, (2016). DOI: 10.1038/529445a
- [4] V. Mnih, K. Kavukcuoglu, D. Silver, A. Graves, I. Antonoglou, D. Wierstra, and M. Riedmiller, "Playing atari with deep reinforcement learning", arXiv preprint arXiv:1312.5602, (2013).
- [5] C. Dong, M. C. Leu, and Z. Yin, "American sign language alphabet recognition using microsoft kinect," in Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops, pp. 44-52, (2015).DOI: 10.1109/CVPRW.2015.7301347
- [6] F. Shrouf, J. Ordieres-Mere, A. Garc ´ıa-Sanchez, and M. Ortega-Mier, "Optimizing the production scheduling of a single machine to minimize total energy consumption costs," Journal of Cleaner Production, vol. 67, pp. 197-207, (2014). DOI: 10.1016/j.jclepro.2013.12.024
- [7] E. Mocanu, D. C. Mocanu, P. H. Nguyen, A. Liotta, M. E. Webber, M. Gibescu, and J. G. Slootweg, "On-line building energy optimization using deep reinforcement learning," IEEE Transactions on Smart Grid, (2018). DOI: 10.1109/TSG.2018.2834219
- [8] R. S. Sutton, A. G. Barto, et al., Introduction to reinforcement learning, vol. 135. MIT press Cambridge, (1998).
- [9] A. Rosebrock, "Imagenet: Vggnet, resnet, inception, and xception with keras" Mars, (2017).
- [10] A. Ashiquzzaman, A. K. Tushar, M. R. Islam, D. Shon, K. Im, J.-H. Park, D.-S. Lim, and J. Kim, "Reduction of overfitting in diabetes prediction using deep learning neural network," in International Conference on Information Theoretic Security, pp. 35-43, Springer, Singapore, (2017).DOI: 10.1007/978-981-10-6451-7_5
- [11] Y. Hirose, K. Yamashita, and S. Hijiya, "Back-propagation algorithm which varies the number of hidden units," Neural Networks, vol. 4, no. 1, pp. 61-66, (1991). DOI: 10.1016/0893-6080(91)90032-z
- [12] G. E. Hinton, N. Srivastava, A. Krizhevsky, I. Sutskever, and R. R. Salakhutdinov, "Improving neural networks by preventing co-adaptation of feature detectors," arXiv preprint arXiv:1207.0580, (2012)

Authors



Akm Ashiquzzaman

Akm Ashiquzzaman received his B.Sc. (Hons.) Degree in Computer Science and Engineering from University of Asia Pacific, Dhaka Bangladesh in 2017. Currently he is a Masters student at the Smart Mobile and Media Computing Laboratory, School of Electronics and Engineering, Chonnam National University, South Korea.



Tai-Won Um

Tai-Won Um received the B.S. degree in electronic and electrical engineering from Hongik University, Seoul, South Korea, in 1999, and the M.S. and Ph.D. degrees from the Korea Advanced Institute of Science and Technology (KAIST), Daejeon, South Korea, in 2000 and 2006, respectively. He is currently an Associate Professor with Chosun University, Gwangju, South Korea. He has been actively participating in standardization meetings including ITU-T SG 13 (future networks, including mobile, cloud computing, and NGN).



JinSul Kim

JinSul Kim received the B.S. Degree in computer science from the University of Utah, Salt Lake City, Utah, USA in 2001, and the M.S. and Ph.D degrees in digital media engineering, Department of Information and Communications of the Korea Advanced Institute of Science and Technology (KAIST), Daejeon, South Korea in 2005 and 2008. He worked as a researcher in the IPTV Infrastructure Technology Research Laboratory, Broadcasting/Telecommunications Convergence Research Division, Electronics and Telecommunications Research Institute (ETRI), Daejeon, Korea from 2005 to 2008. He worked as a professor in the Korea Nazarene University, Chon-an, Korea from 2009 to 2011. Currently, he is a professor in Chonnam National University, Gwangju, Korea. His research interests include QoS/QoE, Measurement/ Management, IPTV, Mobile IPTV, Smart TV, Multimedia Communication and Digital Media Arts

