

SDN/NFV 자동화를 위한 머신러닝기술 연구 동향

조선우*, 정다은*, 이수환**, 신명기**, 박형곤°

Survey on Machine Learning Algorithms for SDN/NFV Automation

Sunwoo Cho*, Daeun Jung*, Soohwan Lee**, Myung-Ki Shin**, Hyunggon Park°

요약

네트워크의 어플리케이션이 다양해지고 단말기의 수가 증가함에 따라 현재의 네트워크는 기존의 전통적 네트워크에 비하여 더 복잡하고 다양한 수요를 반영해야 한다. 특히 초연결, 초고속, 초저지연을 목표로 하는 5G와 같은 네트워크는 기존의 하드웨어를 기반으로 한 구조에서 발전하는 데 한계가 있어, 이러한 한계점을 극복하기 위하여 소프트웨어적으로 프로그래밍이 가능한 SDN과 NFV를 적용하려는 노력이 있다. SDN과 NFV를 적용함에 따라 이들을 자동으로 제어하기 위하여 머신러닝 기술을 기반으로 이들의 지능화를 시도하고 있으며 SDN과 NFV 환경에 머신러닝을 적용하는 연구가 활발하게 이루어지고 있다. 본 논문에서는 SDN/NFV 환경에 머신러닝을 적용한 네트워크 관리 및 오케스트레이션에 대한 선행 연구를 요약 및 분석하고 선행 연구의 한계점과 한계점 극복을 위한 향후 연구 방향을 제시하였다.

Key Words : Software Defined Networking (SDN), Network Function Virtualization (NFV), Machine Learning, Network intelligence, Network automation

ABSTRACT

As applications of network become popular and the number of the end devices increases, modern network needs to cover more intensive demands compared to traditional networks. Since networks such as 5G that aim to guarantee massive connectivity, high data rate, and ultra-low latency, have a limitation based on hardware-based architectures, it has been proposed to deploy software-based programmable SDN and NFV architectures. By deploying SDN and NFV, it has been proposed to adopt machine learning algorithms to automatically control SDN and NFV, leading to intelligent networks. In this paper, we extensively review and summarize prior works on machine learning based SDN/NFV network management and orchestration. Also, we discuss limitation of prior works and direction for future research.

※ 본 연구는 과학기술정보통신부 및 정보통신기술진흥센터의 정보통신-방송 연구개발 사업(2018-0-00701)의 연구결과로 수행되었으며 2017년도 정부(과학기술정보통신부)의 재원으로 한국연구재단(No. NRF-2017R1A2B4005041)의 지원을 받아 수행된 연구임.

• First Author : (ORCID:0000-0003-2193-5246)Department of Electronic and Electrical Engineering, Ewha Womans University, sunwoo.cho@ewhain.net, 학생회원

° Corresponding Author : (ORCID:0000-0002-5079-1504)Department of Electronic and Electrical Engineering, Ewha Womans University, hyunggon.park@ewha.ac.kr, 종신회원

* (ORCID:0000-0001-6278-0952)Department of Electronic and Electrical Engineering, Ewha Womans University, daeun0710@ewhain.net, 학생회원

** Electronics and Telecommunications Research Institute, soohwan@etri.re.kr, mkshin@etri.re.kr, 정회원

논문번호 : 201810-343-B-RN, Received October 25, 2018; Revised November 12, 2018; Accepted December 18, 2018

I. 서 론

통신 네트워크는 전화선을 이용한 초고속 인터넷 ADSL(Asymmetric Digital Subscriber Line)과 VDSL(Very high-data rate Digital Subscriber Line)에서부터 아날로그 통신 기반이었던 1세대 무선통신, CDMA(Code Division Multiple Access)를 기반으로 한 2세대 무선통신, WCDMA(Wideband Code Division Multiple Access)를 이용한 3세대 무선통신과 스마트폰 대중화를 이루고 있는 LTE(Long Term Evolution)기반 4세대 이동통신에 이어 5세대 이동통신(5G) 네트워크로 발전하고 있다¹⁾. 5G는 사물인터넷, 자율주행 자동차 등의 발전으로 인한 다양한 네트워크 요구를 만족시키기 위한 초연결, 초고속, 초저지연을 목표로 등장하게 되었고, 2019년 상용화를 앞두고 있다.

5G 시대의 핵심 기술 중 하나인 SDN(Software Defined Networking)^{2,3)}과 NFV(Network Function Virtualization)^{4,5)}는 기존의 구조, 기능 및 전략 등 한번 결정되면 변경하기 어려운 경직된 네트워크 형태에서 소프트웨어 기반의 유동적이고 변화가 자유로운 네트워크 구조를 지향한다. 하드웨어기반의 전통적 네트워크와는 달리 네트워크 구조 및 기능을 소프트웨어적으로 제어 할 수 있기 때문에 SDN, NFV의 등장과 함께 네트워크 자동화를 위한 노력이 지속되고 있다. 하지만, 현재까지 많은 부분이 수동적으로 관리 및 조절되고 있으며 자동화가 시도되었다고 해도 부분적인 자동화에 그치는 수준이다.

SDN/NFV에 네트워크 자동화를 위하여 인공지능 기술을 활용하기 위한 연구 개발의 대표적인 프로젝트로 유럽 표준기구인 5G-PPP(5G Public-Private Partnership Association)의 CogNet과 SELFNET을 들 수 있다. CogNet과 SELFNET은 머신러닝을 중심으로 5G 환경에 적합한 네트워크 관리 자동화 시스템을 연구하였다. 또한, 유럽의 통신 네트워크 표준화기구인 ETSI(European Telecommunications Standards Institute)는 정보통신기술 표준화를 통해 2021년까지 우리 주변의 모든 환경을 자동화 하는 것을 목표로 하고 있다⁶⁾.

인공지능을 구현하기 위하여 대표적으로 활용되는 머신러닝은 입력 데이터에 대한 결과 값을 일일이 지정해 주지 않아도 기계가 스스로 학습하고 데이터의 경향을 파악하여 새로운 입력 데이터가 주어졌을 때 이 입력이 어떠한 출력을 나타낼지 예측하게 하는 알고리즘을 의미한다. 머신러닝을 이용하면 방대한 양의

데이터에 대해 관리자가 입력 데이터에 따른 출력을 일일이 지정해 주지 않아도 기계가 데이터를 이용하여 스스로 입력에 대한 출력을 예측할 수 있어 방대한 입력 데이터를 빠르게 처리할 수 있고 인간의 개입 및 판단에 의한 오류를 줄일 수 있으며 더욱 정확하고 빠른 결정을 내릴 수 있다는 장점이 있다. 네트워크 관리의 측면에서 볼 때, 네트워크는 다량의 데이터가 존재하여 학습 시키는 데에 유리하고 데이터의 경향성을 예측할 수 있게 되면 수동적인 관리에 비해 더 빠르고 정확한 대응을 할 수 있게 된다. 따라서 네트워크 관리 자동화를 위하여 머신러닝을 적용하는 것은 적절한 접근 방법이라 할 수 있다.

본 논문에서는 SDN과 NFV 환경에서 머신러닝을 이용한 네트워크 관리 및 오케스트레이션 자동화에 대한 선행 연구를 요약하고 분석한다. II장에서는 SDN과 NFV 구조 및 특징에 대해 설명하고, III장에서는 머신러닝의 정의와 머신러닝 알고리즘의 대표적인 유형인 지도학습과 비지도학습, 강화학습, 그리고 심층학습에 대해 정리한다. IV장에서는 기존 선행 연구 논문들을 머신러닝을 적용한 목적에 따라 선행 연구를 나누어 요약하고 V장에서 선행 연구의 한계점을 논하고 향후 연구 방향을 제시한 뒤 VI장에서 결론을 맺는다.

II. SDN과 NFV 구조

본 장에서는 SDN과 NFV의 개념과 구조에 대해 알아보고, SDN과 NFV의 특징과 머신러닝 적용 가능성에 대해 논의한다.

하드웨어 기반으로 설계되어 유동성이 제한되었던 기존의 네트워크는 다양한 응용기술에 대한 요구 및 빠르게 변화하는 5G 상황에 적절히 대응하기에 한계가 있었다. 이러한 요구의 변화에 따라 유동적으로 변화할 수 있는 소프트웨어 기반의 프로그래머블(programmable) 네트워크의 필요성이 제기되었고, SDN과 NFV의 개념이 생겨났다.

2.1 SDN

일반적으로 네트워크에서 패킷을 송신부로부터 수신부까지 전달할 때, 패킷을 어느 경로로 전송하는 것이 적절한지 판단을 내린 뒤 실제로 패킷을 전달하는 과정을 거치게 된다. 이 때 패킷 전달 경로를 결정하는 과정을 담당하는 부분을 제어 평면, 실제 패킷 전송을 담당하는 부분을 데이터 평면이라고 부른다. 기존의 라우터, 스위치 등 네트워크 장비는 개개의 장비

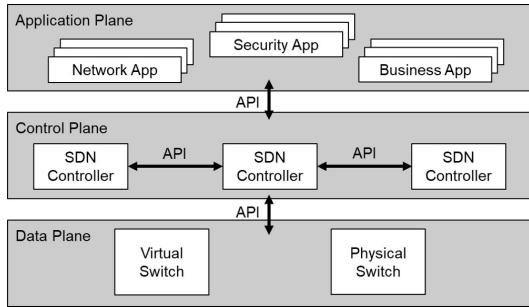


그림 1. SDN 구조[7]
Fig. 1. SDN architecture[7]

가 데이터 평면과 제어 평면을 모두 가지고 있어 하나의 네트워크 장비가 패킷의 전달경로를 결정하고 패킷을 전송하기까지 하는 복잡도가 높은 과정을 담당해야 했다. 또한 네트워크 장비 회사마다 자사 고유의 소프트웨어를 사용하여 타사의 장비와 호환하기에 어려움이 있었다.

이러한 기존 네트워크의 단점을 개선하기 위해 SDN이 등장하였다. SDN은 데이터 평면과 제어 평면을 분리하여 네트워크 장비에는 데이터 평면만 남겨 패킷의 전달에만 집중하도록 하고, 제어 평면은 외부 서버와 같은 중앙 컨트롤러로 옮겨 패킷의 경로 결정만 담당하도록 하는 기술이다^[2,3]. 이 때 분리된 평면 간의 통신을 위해 오픈소스 API(Application Programming Interface)를 사용한다. 이러한 구조를 그림으로 나타내면 그림 1과 같다.

SDN은 네트워크 장비에서 경로 결정 기능을 제거하였기 때문에 각 네트워크 장비의 부하가 감소하게 되고, 중앙 컨트롤러가 전체 네트워크의 정책을 결정하기 때문에 전체 네트워크에 동시에 같은 정책을 적용할 수 있어 네트워크 제어가 간편해진다는 장점이 있다. 또한 오픈소스 API의 사용으로 네트워크 장비 회사에 대한 의존성이 줄어들고 다양한 네트워크 소프트웨어 개발 및 이용이 가능하게 된다.

2.2 NFV

네트워크는 다양한 역할을 담당하는 네트워크 기능들로 구성되어 있다. 네트워크 기능에는 방화벽, 로드 밸런서(Load balancer), 침입 탐지, NAT(Network Address Translation) 등이 있는데, 기존의 전통적 네트워크에서는 네트워크 기능들은 각각의 기능을 담당하는 전용 하드웨어 플랫폼을 통해 구현되었다. 따라서 특정 네트워크 기능을 네트워크에 포함하기 위해서는 해당 기능을 위한 하드웨어 장비를 구입해야 하고, 네트워크 기능이 필요하지 않아져도 하드웨어는

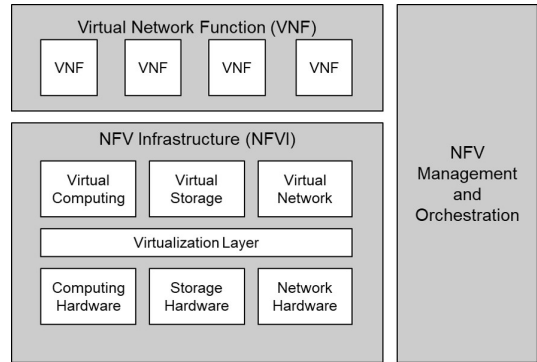


그림 2. NFV 구조[7]
Fig. 2. NFV architecture[7]

여전히 존재하여 자원이 낭비된다.

NFV는 그림 2와 같이 네트워크 기능을 클라우드 등의 VM(Virtual Machine)에 소프트웨어적으로 구현하는 기술을 말한다^[4,5]. NFV는 전통적 네트워크와 달리 하드웨어의 구입 없이 사용자가 원하는 때에 원하는 네트워크 기능을 만들어서 사용할 수 있기 때문에 네트워크 상황이 변화하여 새로운 기능이 요구되어도 유연하게 대응할 수 있다. 또한 VM 위에 구현한 네트워크 기능은 여러 사용자들이 공유하여 사용할 수 있고, 필요하지 않게 되면 언제든지 삭제가 가능하기 때문에 자원의 활용 측면에서 훨씬 효율적이다.

이와 같이 SDN과 NFV는 하드웨어 기반이었던 전통적 네트워크를 소프트웨어적으로 제어할 수 있는 기술적인 방법을 제공하고 있어 머신러닝을 이용하여 네트워크에서 발생하는 방대하고 다양한 데이터를 자동적으로 수집하고 분석하여 적절한 최적 경로 설정 및 네트워크 기능을 설정할 수 있는 네트워크 자동화에 활용할 수 있다.

III. 머신러닝

머신러닝이란 데이터를 기반으로 컴퓨터를 학습시켜 컴퓨터 스스로가 데이터의 경향성을 파악하게 하고, 새로운 입력 데이터의 출력을 예측하도록 하는 알고리즘이다^[8]. 머신러닝은 크게 지도학습, 비지도학습, 강화학습, 심층학습으로 분류할 수 있다.

지도학습은 입력 데이터와 그에 대한 출력 데이터를 포함하는 레이블이 있는 데이터(labeled data)를 사용하여 학습을 진행하는 방법으로, 학습 데이터의 입력 데이터와 출력 데이터로 해당 시스템의 함수를 추론하여 새로운 입력이 들어왔을 때 해당 함수를 이용하여 대응하는 출력을 예측한다. 지도학습은 연속적인

출력 데이터에 대한 예측인 회귀 분석(regression)과 불연속적인 출력 데이터에 대한 예측인 분류(classification)로 나뉜다. 회귀 분석은 연속적인 입력 값에 대해 연속적인 출력 값을 추론하는 방식이고, 분류는 입력 데이터가 이산적인 그룹들 중 어떤 그룹에 속하는지 소속 그룹을 의미하는 이산적인 출력을 추론하는 방식이다. 대표적인 회귀 분석 알고리즘으로는 선형 회귀 알고리즘이 있고, 분류 알고리즘으로는 결정 트리(decision tree)^[9], 랜덤 포레스트(random forest)^[10], SVM(Support Vector Machine)^[11] 등이 있다.

학습 데이터가 입력과 출력 데이터를 모두 포함하는 지도학습과 달리, 비지도학습은 입력 데이터만을 가지는 레이블이 없는 데이터(unlabeled data)를 이용하여 학습한다. 입력 데이터에 대해 어떠한 출력 데이터가 존재해야 하는지에 대한 학습이 불가능하기 때문에, 입력 데이터의 특성을 이용하여 비슷한 특성을 가진 데이터들끼리 무리 짓는 K-means^[12] 등과 같은 군집화(clustering) 알고리즘이 대표적이다.

강화학습은 주어진 상황에서 어떤 행동을 취했을 때 받을 수 있는 보상이 최대가 되는지를 계산하여 취해야 하는 최적의 행동을 에이전트가 스스로 학습해 나가는 방식이다. 행동에 대한 보상을 이용하여 미래에 받을 수 있는 보상의 평균을 최대화 하는 정책을 만드는 것이 강화학습의 목표가 된다. 강화학습의 대표적인 예시로는 Q-러닝^[13]을 들 수 있다.

심층학습은 인간의 두뇌를 모방하여 만든 인공신경망 기반으로 다량의 데이터 속에서 특징을 추출하는 머신러닝 기법이다. 인공신경망은 입력계층과 출력계층을 그 사이의 은닉계층이 연결하는 형태를 이루는데, 여러 개의 은닉계층을 두어 결과를 도출하는 것이 심층학습이다. 심층학습은 지도학습과 비지도학습에 모두 이용될 수 있으며, 대표적인 심층학습 알고리즘으로는 DBN(Deep Belief Network)^[14], CNN(Convolutional Neural Network)^[15], LSTM(Long Short-Term Memory)^[16] 등이 있다.

SDN/NFV 기반 네트워크 환경에서는 패킷을 일정 기준으로 나누는 것이 학습의 목적이 되는 경우가 많기 때문에 네트워크는 분류 또는 군집화 알고리즘이 많이 사용된다. 네트워크에서 얻을 수 있는 데이터의 특성에 따라 지도학습이 가능한 경우에는 분류 알고리즘이, 지도학습이 불가능한 경우에는 군집화 알고리즘이 사용된다. 하지만 네트워크 관리 및 오케스트레이션은 목적이 다양하여 자원 할당과 같이 연속적인

값을 추론해야 하는 목적도 존재하기 때문에 회귀 분석 알고리즘 역시 네트워크에 적용될 수 있다.

다음 장에서는 머신러닝이 SDN/NFV 환경에 적용된 다양한 연구 사례에 대해 살펴본다.

IV. 머신러닝을 적용한 SDN/NFV 제어

SDN/NFV 환경에서 네트워크를 효율적으로 관리하기 위해 현상을 예측하거나 예측한 결과를 바탕으로 네트워크 스스로 대처하도록 하기 위하여 머신러닝을 적용한다. ETSI ISG(Industry Specification Group) ENI(Experiential Network Intelligence) 표준화 그룹은 네트워크 지능 기술을 적용할 수 있는 분야를 크게 인프라 관리, 네트워크 운용, 서비스 요구사항 보장^[17]으로 나누고 있다. 인프라 관리 분야에 해당하는 구체적 목적의 예로는 트래픽 예측^[18], 네트워크 자원관리^[19-22], 네트워크 오토스케일링^[23,24], 최적의 제어 알고리즘 추적^[25]이 있고, 네트워크 운용 분야에 속하는 예로는 이상 감지^[26-29], QoS 향상 및 모니터링 비용 최소화^[30], VNF 실패 예측^[31]이 있다. 그리고 서비스 요구사항 보장 분야에 속하는 목적의 예에는 구체적으로 어플리케이션 식별^[32,33], VNF 선택 및 배열^[34,35], 콘텐츠 인기도 예측^[36]이 있다. 이번 장에서는 목적에 따라 SDN/NFV 환경에 적용된 머신러닝 알고리즘과 활용한 데이터에 대하여 알아본다.

4.1 인프라 관리

4.1.1 트래픽 예측

네트워크의 효율적인 관리를 위해서는 트래픽이 언제, 어디에 몰리는지 예측하는 것이 중요하다. 트래픽 예측은 자원의 할당이나 네트워크 오토스케일링, 네트워크 이상 감지 등 다양한 문제의 해결에 기본 과정으로 활용 가능하며, 네트워크 문제 상황에 보다 빠르게 대처하는 것이 가능하게 한다.

[18]은 SELFNET 프로젝트의 기본 목표인 네트워크의 자가 치유, 자가 보호 및 자가 최적화를 목적으로 하고 있다. 특히 트래픽 혼잡을 방지하기 위해 분류 알고리즘을 이용한 트래픽 예측에 대해 연구하였고 이를 위해 5G 네트워크 테스트베드를 실제로 구현하여 네트워크 내 서버들 및 PC에서 네트워크 관리감독 시스템인 Zabbix client로 트래픽 데이터를 직접 관찰하여 학습 데이터로 사용하였다. 이 때, 트리 구조를 이용하여 정보 이득이 가장 큰 특성을 기준으로 분류해 나가는 결정 트리, 데이터를 낮은 차원으로 사

영한 뒤 적절한 결정 경계를 구하여 분류하는 선형 판별 분석(linear discriminant), 각 분류 집단의 데이터와 가장 큰 거리를 가지는 초평면을 구하여 분류하는 SVM, 그리고 해당 데이터와의 거리를 기준으로 가장 가까운 학습 데이터의 분류 집단으로 분류하는 k-NN의 4가지 분류 알고리즘을 사용하여 각각의 알고리즘에 따른 성능을 비교하였다. 시뮬레이션 결과 결정 트리와 k-NN은 특정 데이터 특징에 대해서 100%의 정확도를 나타내며 좋은 성능을 보였다.

4.1.2 네트워크 자원관리

네트워크에 투입하는 자원은 연속적이기 때문에, 네트워크 자원 관리는 다른 네트워크 관리 분야에 비하여 머신러닝 알고리즘 중 회귀 분석이 많이 사용되는 특징이 있다.

[19]는 SDN 환경 하에서 트래픽을 예측하고 그에 따라 자원을 어떻게 할당하는 것이 효율적인지에 대해 연구하였다. 트래픽 예측을 위해서 신경망 알고리즘의 일종으로 오차를 줄이기 위해 신경망 출력의 오차를 입력 계층 쪽으로 전파시켜 오차의 원인을 찾아 줄여 나가는 BP(Back Propagation) 알고리즘과 SVM의 접근 방법을 회귀 분석에 적용한 SVR(Support Vector Regression) 알고리즘을 사용하였다. 예측한 트래픽에 따른 자원 할당 정책은 최적화 알고리즘 중 하나인 유전 알고리즘을 이용하였다.

[20]에서는 KDN(Knowledge-defined Networking)에 관련된 기존 연구^[37]에서 사용하였던 데이터셋을 이용하여 네트워크에 자원을 얼마나 투입해야 하는지에 대해 연구하였다. 사용한 머신러닝 알고리즘은 SVR, ANN(Artificial Neural Networks)이고, 각각의 알고리즘에 데이터의 좌표계를 변환하여 데이터 성분 중 중요한 성분만 남기는 PCA(Principal Component Analysis)를 추가적으로 적용하여 PCA의 적용 여부에 따라 성능에 변화가 있는지 살펴보았다. 실험 결과 예측의 오차는 SVR을 단독으로 적용하였을 때가 가장 작았으나, SVR을 단독으로 사용하였을 때보다 SVR에 PCA를 추가적으로 적용하였을 때 평균 실행 시간이 1/5정도로 크게 개선되는 것을 확인할 수 있다. SVR에 PCA를 추가한 것과 SVR만 사용한 것이 정확도 측면에서는 큰 차이가 없었으므로 SVR과 PCA를 함께 사용하는 것이 종합적으로 가장 좋은 성능을 나타낸다고 할 수 있다.

[21]에서는 MongoDB^[38]라는 데이터베이스에서 제공하는 데이터를 이용하여 K-means 알고리즘과 선형 회귀 분석(linear regression)을 단계적으로 사용하여

NFV 네트워크 환경에서 최적의 자원 할당 알고리즘에 대해 연구하였다. K-means 알고리즘은 군집화 알고리즘의 일종으로 K개의 군집이 있다고 가정할 때 각 군집의 중심점을 잡고 그 중심점에서 거리가 가까운 데이터들을 같은 군집으로 묶는 방법을 말하고, 선형 회귀 분석은 입력 데이터와 출력 데이터 간의 관계를 선형 함수로 모델링하여 연속적인 출력 값을 추론하는 방법을 말한다. [21]에서는 K-means 알고리즘을 이용하여 현재 네트워크 상태에 대하여 판단하고, 그 결과에 따라 선형 회귀 분석을 이용하여 자원을 할당하였다. 실험을 통해 본 알고리즘을 이용하여 예상한 자원이 실제 네트워크가 필요로 하는 자원과 비슷한 값을 나타낸다는 것을 확인하였다.

[22]에서는 스마트시티 구현에 적용할 수 있도록 네트워킹, 캐싱 및 자원 계산을 동적 오케스트레이션할 수 있는 통합 프레임워크를 제안하였다. 각 기능을 동시에 모두 고려하여 최적의 자원 할당을 하기 위해 심층학습 구조를 이용하여 Q-러닝의 Q 함수를 구하는 알고리즘인 심층 Q-네트워크(DQN)를 이용하였고, 기존의 정적인 방법 대비 제안된 방법이 더 좋은 성능을 얻은 것을 확인하였다.

4.1.3 네트워크 오토스케일링

네트워크에 트래픽이 갑자기 증가하게 되면, 한정된 네트워크 자원이 트래픽을 감당하지 못하고 다운되어 버리는 문제가 발생하게 된다. 이러한 상황을 방지하기 위하여 네트워크 오토스케일링을 적용할 수 있는데, 네트워크 수요가 갑자기 많아지면 서버 등의 자원을 늘려 트래픽 요구량을 감당할 수 있게 하고 반대로 수요가 적어지면 자원을 줄여 낭비되는 자원이 없도록 하는 방법이다. NFV는 원하는 때에 네트워크의 기능과 자원을 클라우드 등 가상공간에 만들고 없애는 것이 자유롭기 때문에 네트워크 오토스케일링에 적합한 환경이라고 할 수 있다.

[23]에서는 대서양 링크의 ISP에서 얻은 실제 트래픽 로드 데이터^[39]를 머신러닝 소프트웨어인 WEKA 환경에서 실험하기에 적절한 7개의 분류 알고리즘 - 랜덤 트리(random tree), J48, 랜덤 포레스트, multi-layer perceptron, 베이지안 네트워크(Bayesian network) 등 - 에 대하여 학습시켜 각 알고리즘에 따른 오토스케일링의 정확도를 비교하였다. 랜덤 트리는 데이터셋에서 데이터를 무작위로 추출하여 결정 트리를 만드는 방법이고, J48은 결정 트리의 일종이며, 랜덤 포레스트는 랜덤 트리를 여러 개 만들어 그 결과들을 종합하여 판단을 내리는 알고리즘이다. Multi-layer

perceptron은 심층학습 네트워크의 가장 간단한 구조를 말하고, 베이지안 네트워크는 랜덤 변수들 간 조건부 종속성을 이용하여 학습하는 방법이다. 실험 결과 랜덤 포레스트의 정확도가 가장 높았고 베이지안 네트워크의 정확도가 가장 낮았다.

[24]는 네트워크 오토스케일링을 위해 강화학습 알고리즘을 사용하였다. 실제 모바일 네트워크 오퍼레이터의 베타 사이트에 존재하는 데이터를 이용하였는데, 이 데이터셋은 일상적인 트래픽의 시나리오와 폭발적으로 트래픽이 증가하는 시나리오를 모두 포함하고 있다. 강화학습의 일종으로 에이전트가 특정 상황에서 어떠한 행동을 취하였을 때 얻을 수 있는 보상의 최댓값을 출력하는 함수인 Q 함수를 구하는 것을 목적으로 하는 Q-러닝을 사용하여 네트워크를 오토스케일링하는 알고리즘을 제안하였고, 기존의 방법에 비해 안정적으로 SLA(Service Level Agreement)를 보장하는 것을 보였다.

4.1.4 최적의 접근 제어 알고리즘 추적

SDN의 도입으로 데이터평면과 제어평면이 분리되면서 네트워크로 들어오는 연결 요청에 대한 접근 제어(admission control)가 가능해졌다. 접근 제어 알고리즘을 이용하면 SDN이 현재 네트워크의 자원 양에 따라 요청을 수락하거나 거절할 수 있기 때문에 네트워크 관리에 효과적이다.

[25]는 트래픽 상황을 알 수 없을 때 최적의 접근 제어 알고리즘이 무엇인지를 추적하기 위해 유럽 지역의 연구와 교육을 위한 높은 대역폭의 백본인 GEANT으로부터 네트워크 트래픽 분석 프로토콜인 NetFlow를 이용하여 얻은 데이터^[40]와 서로 다른 차치 시스템 간의 라우팅 프로토콜인 BGP(Border Gateway Protocol) 라우팅 정보를 데이터로 이용하였고, 알고리즘으로는 보상을 최대화하기 위한 전략을 과거의 행동 정보를 이용하여 시간에 따라 결정하는 SEA(Strategic Expert meta-Algorithm)를 적용하였다. 시뮬레이션 결과 기존의 FLA(Follow-the-Leader meta-Algorithm)에 비해 제안된 알고리즘이 더 좋은 성능을 가진다는 것을 보였다.

4.2 네트워크 운용

4.2.1 이상 감지

정상 데이터와 비정상 데이터를 지도학습으로 학습시켜 이상 상황을 판단하게 하는 방법은 네트워크뿐만 아니라 다른 여러 어플리케이션에서 사용되고 있

다. 네트워크 환경에서 이상 감지는 보안과 관련 지어 정상 패킷과 공격 패킷을 구별하는 용도로 사용된다.

[26]은 네트워크에서 정상 패킷과 공격 패킷을 구분하기 위해 분류의 방식을 이용하였다. 데이터로는 기존 다른 연구에서 사용했던 다양한 타입의 공격 패킷과 정상 패킷이 섞여 있는 덤프 데이터^[41]를 사용하였다. 분류 알고리즘으로는 SVM, 결정 트리, bagged tree, 랜덤 포레스트, k-NN(k-Nearest Neighbors) 알고리즘들을 사용하여 각각의 성능을 비교하였다. 이때 bagged tree는 랜덤 트리와 같은 의미를 갖는다. 정확도 측면에서 볼 때 랜덤 포레스트의 성능이 가장 좋고, 시간 측면에서 볼 때에는 결정 트리가 평균 학습 시간과 평균 검증(validation) 시간 모두 최소를 기록하여 가장 높은 시간 성능을 가짐을 확인할 수 있다.

[27]에서는 DDoS(Distributed Denial of Service) 공격을 탐지하기 위하여 심층학습 기반 알고리즘을 사용하였다. 다양한 심층학습 신경망 알고리즘을 이용하여 SDN의 OpenFlow 스위치에 들어오는 패킷이 정상 패킷인지 공격 패킷인지 구별하는 방식인데, 학습 데이터로 실제 네트워크에서 2일 동안 수집한 데이터로 이루어진 이상 감지용 데이터셋인 ISCX2012 데이터셋^[42]을 사용하였다. 심층학습 신경망을 이용한 알고리즘인 LSTM과 CNN에 LSTM을 접목한 CNN/LSTM, GRU(Gated Recurrent Unit), 그리고 LSTM 3개를 연결한 3LSTM을 각각 적용한 뒤 성능을 비교하였다. 여기서 LSTM은 은닉 계층이 방향성이 있는 연결을 가지는 순환 신경망(Recurrent Neural Network)에 입력을 저장할 수 있는 셀과 게이트를 추가하여 성능을 높인 알고리즘이고, CNN은 입력 데이터를 합성 곱(convolution) 처리하여 특징을 추출하고 분류할 수 있는 알고리즘이며, GRU는 LSTM의 구조를 단순화 한 알고리즘으로, 실험 결과 3LSTM의 정확도가 가장 높은 것을 확인하였다.

[28]은 NFV 환경에서 머신러닝, 특히 심층학습을 이용하여 이상을 감지하는 시스템을 2단계로 나누어 제안하였다. 첫 번째 단계는 ASD(Anomaly Symptom Detection)이고 그 다음 단계는 NAD(Network Anomaly Detection)이다. ASD에서는 확률적 이진 유닛(stochastic binary unit)을 이용하여 비지도학습으로 최대가능도(maximum likelihood)를 예측하는 RBM(Restricted Boltzmann Machine)의 다층 구조 형태인 DBN과 신경망 구조에서 비지도학습이나 특징 추출을 할 때 쓰이는 SAE(Stacked Autoencoder)를 사용하고 NAD에서는 LSTM을 이용한다. ASD는 네트워크에서 이상으로 추정되는 증상을 빠르게 찾아

내는 것이 중요하기 때문에 속도 측면에서 우수한 성능을 보이는 DBN과 SAE를, ASD에서 판단한 증상을 이용하여 이것이 실제 이상인지 아닌지에 대해 판단하는 NAD는 ASD의 처리속도가 충분히 빠르면 속도 관점에서 높은 성능을 보여야 할 필요가 없기 때문에 정확도를 더 중시하는 LSTM을 사용하여 학습한다. [28]은 네트워크의 흐름(flow)을 기록하는 프로토콜인 NetFlow를 이용하여 RAN(Radio Access Network)의 실제 패킷의 흐름 데이터를 모아 학습에 사용하였다. 배치 크기(batch size)와 DBN의 은닉계층의 개수에 따라 시간당 처리 속도가 어떻게 달라지는가에 대해 실험하였다.

[29]는 [28]에서 제안하였던 시스템을 확장하여 더 좋은 성능을 얻고자 하였다. [29]는 CTU-13^[43]이라는 붓넷 데이터셋을 사용하였는데, CTU-13은 붓넷 트래픽과 정상 트래픽을 포함하는 레이블이 포함된 데이터셋(labeled dataset)으로, 13가지의 각기 다른 붓넷 시나리오 데이터를 가지고 있다. 시뮬레이션에서는 DBN의 특징 은닉 계층 개수와 배치 크기에서 심층학습 라이브러리와 GPU/CPU에 따른 시간당 처리속도를 비교하였다.

4.2.2 QoD(Quality of Decision) 향상 및 모니터링 비용 최소화

NFV의 VNF는 소프트웨어적으로 구현되고 여러 사용자들이 공유 할 수 있도록 하는 것을 목표로 하기 때문에 일반적으로 클라우드와 같은 가상 머신 위에 만들어 진다. 이 때 이러한 클라우드를 관리하는 CMS(Cloud Management System)는 머신러닝을 활용하여 효율적으로 클라우드를 관리할 수 있다. 하지만 CMS의 QoD를 향상시키기 위해서는 네트워크를 더 자주 모니터링 해야 하기 때문에 모니터링 비용이 증가하는 단점이 존재한다.

[30]에서는 이러한 상황을 극복하기 위해 CMS의 QoD를 향상시키는 동시에 모니터링 비용을 최소화 하는 알고리즘인 z-TORCH(Zero Touch Orchestration)에 대해 연구하였다. CMS의 QoD 향상을 위해 향상된 K-means 군집화 알고리즘을 이용하였고, 모니터링 비용을 최소화하기 위해 강화학습의 일종인 Q-러닝을 이용하였다. 시뮬레이션을 위해 모바일 네트워크를 구현할 수 있는 소프트웨어인 OpenEPC를 이용하여 자체 테스트베드를 구축하여 실험하였고 그 결과 제안된 z-TORCH가 최적의 상황과 근접한 성능을 나타낸다는 것을 확인하였다.

4.2.3 VNF 실패 예측

NFV에서 목적에 따라 가상 머신 위에 만들어 사용하는 기능을 VNF(Virtual Network Function)라 한다. 이 때 VNF가 제대로 동작하지 않는 경우를 대비하여 마스터 VNF 외에 백업 VNF를 만들어 두기도 하는데, 마스터 VNF의 동작 실패 여부를 예측하여 백업 VNF의 사용 여부를 미리 판단할 수 있으면 NFV 관리에 유리하게 된다.

[31]에서는 마스터 VNF의 실패 예측을 위해 서버의 실패에 대한 공개 데이터셋^[44]을 사용하였다. 해당 데이터셋은 VNF가 실패하는 시간 간격과 온도, 습도 간의 관계에 대한 데이터셋이고, 이 데이터셋을 활용하여 SVM, SVC(Support Vector Clustering) 및 랜덤 포레스트 방식을 이용해 분류하여 VNF 실패를 예측하도록 하였다. SVC는 SVM의 접근법을 이용한 군집화 알고리즘을 말한다. 그 결과, SVM을 이용한 예측은 정확도가 50.08%인 반면, SVC와 랜덤 포레스트를 이용한 예측은 각각 99.99%, 99.98%의 정확도를 나타내며 높은 성능 차이를 보였다.

4.3 서비스 요구사항 보장

4.3.1 어플리케이션 식별

어플리케이션 식별은 네트워크에서 이동하는 패킷들이 어떤 어플리케이션을 위한 패킷인지 식별하는 것을 의미한다. 네트워크의 패킷들로부터 특정 어플리케이션에 대한 정보를 파악하여 해당 어플리케이션에 최적 자원 할당을 할 수 있고, 이상 상황이 생겼을 때 원인을 파악하기 용이하며 어플리케이션에 따라 다른 사용자의 요구에 더 잘 대응할 수 있다.

[32]에서는 각 패킷의 활성화된 네트워크 소켓, 패킷 크기 및 포트 번호 등에 대한 패킷 흐름 데이터를 이용하여 각 패킷이 어떤 어플리케이션을 위한 것인지 분류하는 'Atlas'라는 프레임워크를 제안하였다. 분류 알고리즘으로는 결정 트리 학습법을 사용하였다. 구글 플레이스토어 기준 가장 인기가 많은 어플리케이션 30개를 식별하는 시뮬레이션을 진행하였는데, 평균 정확도가 96%로 높은 정확도를 나타냄을 보였다.

[33]에서는 SDN 제어평면과 데이터평면의 통신 프로토콜인 OpenFlow에서 얻을 수 있는 트래픽 데이터를 이용하여 머신러닝 알고리즘을 학습시켰는데, 학습의 기준이 되는 특징으로 패킷의 크기, 패킷의 타임스탬프, 패킷 도착시간 간격, 맥 주소, IP 주소 등을 추출하여 사용하였다. 분류 알고리즘으로는 랜덤 포레스트

트와 앙상블 모델에서 경사 하강법(gradient descent)을 이용하여 손실 함수를 최소화하는 모델로 움직이도록 하는 경사 부스팅(gradient boosting) 알고리즘의 일종인 확률 경사 부스팅(stochastic gradient boosting)과 익스트림 경사 부스팅(extreme gradient boosting)을 이용하여 각각의 알고리즘마다 8개 어플리케이션(BitTorrent, Dropbox, Facebook 등)에 대해 분류 성능을 비교하였고, 랜덤 포레스트의 평균 정확도가 가장 높음을 확인하였다.

4.3.2 VNF 선택 및 배열

SDN와 NFV가 결합된 네트워크 환경에서는 다양한 VNF가 존재하고, SDN의 제어 평면이 특정 패킷이 송신부에서 수신부로 이동할 때 어떠한 VNF들을 거쳐서 이동하는 것이 적절한 지에 대해 판단을 내리게 된다. 이 때 특정 서비스를 제공하기 위해 거쳐야 하는 VNF들의 순열을 SFC(Service Function Chain)라고 한다.

[34]에서는 SFC 요청에 대해 최적의 VNF 선택과 순서 배열 만들기 위해 머신러닝을 적용하였다. 심층 학습 기반의 DBN 알고리즘을 그래프 계층화 접근법을 통해 얻은 레이블이 있는 데이터^[45,46]를 이용하여 학습시켜 속도 관점에서 성능을 분석하였다. 시뮬레이션을 통해 네트워크 규모의 증가에 따라 급격하게 실행 시간이 증가하는 기존의 방법과 달리 제안된 방법은 네트워크 규모에 증가에도 속도의 변화가 크지 않음을 보였다.

[35]는 VNF의 규모를 늘리거나 줄일 때 기존의 사후처리식(reactive) 접근방법의 단점을 보완하기 위해 사전처리식(proactive) 접근방법을 시도하였다. 이를 위해 매 시간마다 경사 하강법을 이용하여 손실함수를 줄이는 액션으로 이동하는 온라인 학습법인 온라인 경사 하강법(online gradient descent)을 적용하여 미래의 SFC 트래픽과 부하, 즉 수요를 예측하고 그 결과를 VM 확장에 이용하였다. 데이터로는 Amazon Elastic MapReduce Trace 톨을 이용하여 얻은 HTTP 요청 파일, 송신부 IP 주소, 그리고 시간 스탬프가 포함된 데이터를 사용하였고, 기존의 방식들과 비교하여 예측 오차와 VM 확장 경쟁력을 동시에 고려할 때 더 좋은 성능을 얻는 것을 확인하였다.

4.3.3 콘텐츠 인기도 예측

캐싱(caching)이란 요청되는 빈도가 높은 데이터의 경우 따로 캐시에 저장해 두어 다시 요청되었을 때 빠르게 데이터를 제공할 수 있도록 하는 개념이다. 네트

워크 내에서 콘텐츠가 얼마나 인기 있는지, 즉 요청을 많이 받는지 예측하면 캐싱에 유리하기 때문에 전체 네트워크의 효율성이 높아진다.

[36]은 콘텐츠의 인기도 예측을 위해 머신러닝을 적용하였다. ANN과 연속적인 입력을 0과 1의 불연속적 클래스로 매핑하는 함수를 이용하는 로지스틱 회귀(logistic regression) 분류 방식의 다중 클래스 방법인 Softmax를 이용한 SAE 알고리즘을 제안하고, 실험을 위해 구성된 네트워크에서 얻은 콘텐츠의 공간-시간의 결합 분포를 포함한 데이터셋으로 학습시키는 방식으로 연구를 진행하였다. 신경망만을 이용하는 알고리즘과 자동 회귀(auto regression)만을 이용하는 알고리즘을 제안된 알고리즘과 비교하여 실험한 결과 제안된 알고리즘이 가장 높은 정확도를 가진다는 것을 보였다.

위의 선행 연구들을 표로 정리하면 표 1과 같다.

V. 선행 연구의 한계점 및 향후 연구 방향

본 논문에서 기술한 것과 같이 SDN/NFV 환경에 머신러닝 알고리즘을 적용한 네트워크 자동화 연구는 활발히 진행되고 있다. 하지만 대부분의 선행 연구들은 SDN/NFV 환경에 적용할 머신러닝 알고리즘을 선택할 때에 데이터의 특성을 깊게 고려하지 않거나 일반적으로 성능이 좋다고 알려진 알고리즘을 선택하여 적용해 왔다. 데이터의 특성을 고려하지 않은 머신러닝 적용의 결과는 최적의 결과라고 확인할 수 없으며, 최적의 알고리즘을 선택하기 위해서는 가능한 모든 머신러닝 알고리즘을 적용하여 결과를 얻은 뒤 성능을 비교해야 하는 단점이 있다. 무엇보다 최적 성능을 보여주는 알고리즘에 대한 이해 및 해석을 하기 어려운 문제를 포함하고 있다. 즉, 선행 연구들은 SDN/NFV 환경에서 다양한 목적을 위해 머신러닝을 적용할 수 있다는 가능성을 보여주는 단계에 머무르고 있다.

이러한 한계점을 극복하기 위해서 SDN/NFV 네트워크 데이터 특성을 고려하고 이러한 데이터의 특성에 맞는 최적 머신러닝 알고리즘을 결정하는 방법에 대한 연구 방향이 반드시 필요하며, 데이터는 물론 네트워크 구조에 의하여 결정되는 요소들도 함께 고려되어야 할 것이다. 최종 목표인 전자동 네트워크 관리를 위하여 자동화된 머신러닝(Automated Machine Learning)등을 활용할 수 있는 연구도 함께 진행되어야 할 것이다.

표 1. SDN/NFV에 머신러닝을 적용한 선행 연구
 Table 1. Machine learning algorithms adopted in prior works for SDN/NFV

Purpose		Reference	Data for Learning	Machine Learning Algorithms	Simulation Results	
Infrastructure Management	Traffic prediction	[18]	Data from testbed network servers, collected by Zabbix client	Decision tree, linear discriminant, SVM, k-NN	Decision tree and k-NN have the best precision (100%) for a specific feature	
	Resource management	[19]	Not mentioned	Back propagation, SVR	Higher precision was obtained by increasing the number of layers of the neural networks	
		[20]	Data about CPU consumption under real traffic	SVR, ANN, PCA	When SVR and PCA used together, high precision and the shortest running time are obtained	
		[21]	Dataset from MongoDB	K-means, linear regression	The predicted bandwidth requirement by linear regression is similar to the actual bandwidth requirement	
		[22]	Not mentioned	DQN	Compared with existing static scheme, the proposed scheme has higher total utility	
	Network autoscaling	[23]	Real traffic load data from ISP in the Atlantic(time-traffic load data)	Random tree, J48, REPTree, decision table, random forest, multi-layer perception, Bayesian network	Random forest has the best precision and Bayesian network has the worst precision	
		[24]	Dataset on the real mobile network operator beta site with traffic scenarios	Q-learning	Compared with existing policies (static threshold and voting policy), the proposed policy outperforms in SLA guarantee	
	Best admission control tracking	[25]	Data from GEANT, collected by NetFlow	SEA	Compared with existing FLA algorithm, the proposed SEA algorithm outperforms	
	Network Operation	Anomaly detection	[26]	Dump data with the mixture of normal packets and attack packets	SVM, decision tree, bagged tree, random forest, k-NN	Random forest has the best precision, and decision tree has the best performance in training and validation time
			[27]	ISCX2012 intrusion detection dataset	LSTM, CNN/LSTM, GRU, 3LSTM	3LSTM has the best precision (over 98.5%)
[28]			Flow data from RAN, collected by NetFlow	LSTM, DBN, SAE	Processing speed is higher when the number of hidden layers of DBN is smaller	
[29]			CTU-13 botnet dataset	LSTM, DBN, SAE	cuBLAS has the highest processing speed when using GPU and TensorFlow has the highest processing speed when using CPU for every batch sizes	

Purpose		Reference	Data for Learning	Machine Learning Algorithms	Simulation Results
	QoS enhancement and minimizing monitoring cost	[30]	Not mentioned	K-means, Q-learning	Compared with the existing instant placement method, z-TORCH has closer QoS to the optimum and much lower monitoring load
	VNF failure forecasting	[31]	Open dataset about the relationship between server failure and the external environments	SVM, SVC, random forest	SVM has the worst precision (50.08%) and SVC has the best precision (99.99%)
Assurance	Application identification	[32]	Flow data with open network socket, packet size, port number, etc., collected by OpenFlow	Decision tree	Decision tree has 96% of average precision for identify 30 applications
		[33]	Traffic data with packet size, packet interval, MAC address, IP address, etc., collected by OpenFlow	Random forest, stochastic gradient boosting, extreme gradient boosting	Random forest has the highest precision for all applications
	Making optimal selection and chaining of the VNF instances	[34]	Labeled data collected by graph layering approach	DBN	Compared with the existing Graph Layering algorithm, the proposed algorithm has much lower execution time for the large scaled network
		[35]	Data with HTTP request file, source IP address, time stamp	Online gradient descent	Compared with the existing FTL and heuristic method, the proposed method has better performance in terms of prediction error and VNF scalability
	Content popularity prediction	[36]	Measure parameters of content being requested at some distributed nodes	SAE, Softmax	Compared with the neural network only method and auto regression method, the proposed method has the highest prediction accuracy

VI. 결 론

네트워크에 요구되는 수요가 다양해지고 빠르게 변함에 따라 네트워크를 소프트웨어 기반으로 제어할 수 있는 SDN과 NFV의 필요성이 증대되었다. 더불어 네트워크 제어의 정확성, 실시간성 및 안정성 확보를 위해 네트워크 자동화에 대한 필요성이 대두되고 있다. 이에 따라 SDN/NFV 자동화를 위한 연구가 활발히 이루어지고 있어, 본 논문에서는 SDN/NFV 네트

워크 제어 및 오케스트레이션에 머신러닝을 적용하여 자동화를 시도한 선행 연구를 조사하고 목적에 따라 분류하여 분석하였다. 또한 선행 연구의 한계점을 기술하고 한계점 극복을 위한 향후 연구 방향을 제시하였다.

References

[1] H. Kim, J. Moon, D. Kim, Y. Lee, and D.

- Choi, "A study on development of the e-government based on 5G," in *Proc. Symp. KICS*, pp. 1194-1195, Jeju, Korea, Jun. 2017.
- [2] ONF Market Education Committee, *Software-Defined Networking: The New Norm for Networks*(2012), Retrieved Oct. 21, 2018, from <http://www.opennetworking.org>.
- [3] B. Raghavan, M. Casado, T. Kopenon, S. Ratnasamy, A. Ghodsi, and S. Shenker, "Software-defined internet architecture: decoupling architecture from infrastructure," in *Proc. 11th ACM Workshop on Hot Topics in Netw.*, pp. 43-48, Washington, USA, Oct. 2012.
- [4] AT&T, BT, and other operators, *Network Functions Virtualisation - An Introduction, Benefits, Enablers, Challenges & Call for Action*(2012), Retrieved Oct. 20, 2018, from https://portal.etsi.org/NFV/NFV_White_Paper.pdf.
- [5] R. Mijumbi, J. Serrat, J. Gorricho, N. Bouten, F. De Turck, and R. Boutaba, "Network function virtualization: State-of-the-Art and research challenges," *IEEE Commun. Surveys & Tuts.*, vol. 18, no. 1, pp. 236-262, Firstquarter 2016.
- [6] ETSI, *Long-Term Strategy 2016-2021*, Retrieved Oct., 21, 2018, from https://www.etsi.org/images/files/Brochures/ETSI_LTS%20Brochure_WEB.pdf.
- [7] W. Stallings, *Foundations of modern networking: SDN, NFV, QoE, IoT, and Cloud*, Addison-Wesley Professional, 2015.
- [8] T. Hastie, R. Tibshirani, and J. Friedman, *The Elements of Statistical Learning*, Springer Series in Statistics, 2009.
- [9] J. R. Quinlan, "Induction of Decision Trees," *Machine Learning*, vol. 1, no. 1, pp. 81-106, Mar. 1986.
- [10] A. Liaw and M. Wiener, "Classification and regression by randomForest," *R news*, vol. 2, no. 3, pp. 18-22, Dec. 2002.
- [11] M. A. Hearst, S. T. Dumais, E. Osuna, J. Platt, and B. Scholkopf, "Support vector machines," *IEEE Intell. Syst. and their Appl.*, vol. 13, no. 4, pp. 18-28, Jul.-Aug. 1998.
- [12] J. A. Hartigan and M. A. Wong, "Algorithm AS 136: A k-means clustering algorithm," *J. Royal Statistical Soc. Series C (Applied Statistics)*, vol. 28, no. 1, pp. 100-108, 1979.
- [13] C. J. C. H. Watkins and P. Dayan, "Q-Learning," *Machine Learning*, vol. 8, no. 3-4, pp. 279-292, May 1992.
- [14] G. E. Hinton, S. Osindero, and Y. The, "A fast learning algorithm for deep belief nets," *Neural Computation*, vol. 18, no. 7, Jul. 2006.
- [15] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "ImageNet classification with deep convolutional neural networks," in *Proc. Advances in Neural Inf. Process. Syst. (NIPS)*, pp. 1097-1105, Nevada, USA, Dec. 2012.
- [16] S. Hochreiter and J. Schmidheber, "Long short-term memory," *Neural Computation*, vol. 9, no. 8, pp. 1735-1780, Nov. 1997.
- [17] H. Kim, M. Shin, B. Ahn, J. Lee, S. Lee, S. Lee, J. Ham, and S. Hyeon, "Network intelligence technologies," *ETRI Insight*, 2018.
- [18] W. Jiang, M. Strufe, and H. D. Schotten, "Experimental results for artificial intelligence-based self-organized 5G networks," in *Proc. IEEE Annu. Int. Symp. Pers., Indoor, and Mob. Radio Commun. (PIMRC)*, pp. 1-6, Montreal, Canada, Oct. 2017.
- [19] S. Sun, L. Gong, B. Rong, and K. Lu, "An intelligent SDN framework for 5G heterogeneous networks," *IEEE Commun. Mag.*, vol. 53, no. 11, pp. 142-147, Nov. 2015.
- [20] H. Jmila, M. Ibn Khedher, and M. A. El Yacoubi, "Estimating VNF resource requirements using machine learning techniques," in *Proc. Int. Conf. Neural Inf. Process. (ICONIP)*, Guangzhou, China, Nov. 2017.
- [21] A. Martin, J. Egaña, J. Flórez, J. Montalbán, I. G. Olaizola, M. Quartulli, R. Viola, and M. Zorrilla, "Network resource allocation system for QoE-aware delivery of media services in 5G networks," *IEEE Trans. Broadcasting*, vol.

- 64, no. 2, pp. 561-574, Jun. 2018.
- [22] Y. He, F. R. Yu, N. Zhao, V. C. M. Leung, and H. Yin, "Software-defined networks with mobile edge computing and caching for smart cities: A big data deep reinforcement learning approach," *IEEE Commun. Mag.*, vol. 55, no. 12, pp. 31-37, Dec. 2017.
- [23] S. Rahman, T. Ahmed, M. Huynh, M. Tornatore, and B. Mukherjee, "Auto-scaling VNFs using machine learning to improve QoS and reduce cost," in *Proc. IEEE Int. Conf. Commun. (ICC)*, pp. 1-6, Kansas City, USA, May 2018.
- [24] P. Tang, F. Li, W. Zhou, W. Hu, and L. Yang, "Efficient auto-scaling approach in the telco cloud using self-learning algorithm," in *Proc. IEEE Global Commun. Conf. (GLOBECOM)*, pp. 1-6, San Diego, USA, 2015.
- [25] J. Leguay, L. Maggi, M. Draief, S. Paris, and S. Chouvardas, "Admission control with online algorithms in SDN," in *Proc. IEEE/IFIP Netw. Operations and Management Symp. (NOMS)*, pp. 718-721, Istanbul, Turkey, Apr. 2016.
- [26] G. A. Ajaeiya, N. Adalian, I. H. Elhadj, A. Kayssi, and A. Chehab, "Flow-based intrusion detection system for SDN," in *Proc. IEEE Symp. Comput. and Commun. (ISCC)*, pp. 787-793, Heraklion, Greece, Jul. 2017.
- [27] C. Li, Y. Wu, X. Yuan, Z. Sun, W. Wang, X. Li, and L. Gong, "Detection and defense of DDoS attack - based on deep learning in OpenFlow based SDN," *Int. J. Commun. Syst.*, vol. 31, no. 5, Jan. 2018.
- [28] L. Fernández Maimó, F. J. García Clemente, M. Gil Perez, and G. Martínez Perez, "On the performance of a deep learning-based anomaly detection system for 5G networks," in *Proc. IEEE 3rd Smart World Congress*, Leicester, UK, Aug. 2017.
- [29] L. Fernández Maimó, Á. L. Perales Gómez, F. J. García Clemente, M. Gil Pérez, and G. Martínez Pérez, "A self-adaptive deep learning-based system for anomaly detection in 5G networks," *IEEE Access*, vol. 6, pp. 7700-7712, 2018.
- [30] V. Sciancalepore, F. Z. Yousaf, and X. Costa-Perez, "z-TORCH: An automated NFV orchestration and monitoring solution," *IEEE Trans. Network and Service Management (Early Access)*, Aug. 2018.
- [31] H. Huang, S. Guo, and K. Wang, "Proactive failover for NFV in edge networks," *IEEE Commun. Mag.*, Jul. 2018.
- [32] Z. Ayyub Qazi, J. Lee, T. Jin, G. Bellala, M. Arndt, and G. Noubir, "Application-awareness in SDN," *ACM SIGCOMM Comput. Commun. Rev.*, vol. 43, no. 4, pp. 487-488, 2013.
- [33] P. Amaral, J. Dinis, P. Pinto, L. Bernardo, J. Tavares, and H. S. Mamede, "Machine learning in software defined networks: Data collection and traffic classification," in *Proc. IEEE Intl. Conf. Netw. Protocols. (ICNP)*, Singapore, pp. 1-5, Nov. 2016.
- [34] J. Pei, P. Hong, and D. Li, "Virtual network function selection and chaining based on deep learning in SDN and NFV-enabled networks," in *IEEE Intl. Conf. Commun. Workshops (ICC Workshops)*, pp. 1-6, Kansas City, USA, May 2018.
- [35] X. Zhang, C. Wu, Z. Li, and F. C. M. Lau, "Proactive VNF provisioning with multi-timescale cloud resources: Fusing online learning and online optimization," in *Proc. IEEE Intl. Conf. Computer Commun. (INFOCOM)*, pp. 1-9, Atlanta, USA, May 2017.
- [36] W. Liu, J. Zhang, Z. Liang, L. Peng, and J. Cai, "Content popularity prediction and caching for ICN: A deep learning approach with SDN," *IEEE Access*, vol. 6, pp. 5075-5089, Dec. 2017.
- [37] *Knowledge-Defined Networking Training Datasets*, Retrieved Oct. 21, 2018, from <http://knowledgedefinednetworking.org/>.
- [38] *MongoDB*, Retrieved Oct. 21, 2018, from <https://www.mongodb.com/>.
- [39] *Internet traffic data*, Retrieved Oct. 21, 2018, from <https://datamarket.com>.

[40] S. Uhlig, B. Quoitin, J. Lepropre and S. Balon, "Providing public intradomain traffic matrices to the research community," *ACM SIGCOMM Comput. Commun. Rev.*, vol. 36, no. 1, Jan. 2006.

[41] *Network forensics training, challenges and contests*, Retrieved Oct. 21, 2018, from <http://www.netresec.com/?page=PcapFiles>.

[42] *ISCX intrusion detection evaluation DataSet*, Retrieved Oct., 8, 2018, from <http://www.unb.ca/cic/datasets/index.html>.

[43] *CTU Dataset*, Retrieved Aug. 13, 2018, from <https://mcfp.weebly.com/the-ctu-13-dataset-a-labeled-dataset-with-botnet-normal-and-background-traffic.html>.

[44] *Small-scale trace of server machine failures*, Retrieved Oct. 21, 2018, from <https://bigml.com>.

[45] Z. Cao and M. Kodialam, et al., "Traffic steering in software defined networks: planning and online routing," *ACM SIGCOMM Comput. Commun. Rev.*, vol. 44, no. 4, pp. 65-70, 2014.

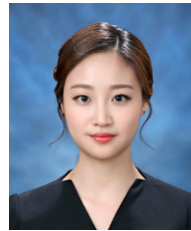
[46] D. Abhishek and T. Wolf, "Adaptive service-chain routing for virtual network functions in software-defined networks," in *Proc. 2016 workshop on Hot topics in Middleboxes and Network Function Virtualization*, pp. 32-37, 2016.

조 선 우 (Sunwoo Cho)



2018년 2월 : 이화여자대학교 전자공학과 학사
 1996년 3월~현재 : 이화여자대학교 전자전기공학과 석사과정
 <관심분야> 머신러닝, 5G네트워크, SDN/NFV, 게임이론

정 다 은 (Daeun Jung)



2014년 3월~현재 : 이화여자대학교 전자전기공학과
 <관심분야> 머신러닝, 베이지안 네트워크, 단백질 네트워크

이 수 환 (Soohwan Lee)



2009년 2월 : 경북대학교 전자전기컴퓨터학부 학사
 2011년 2월 : 한국과학기술원 전기 및 전자공학 석사
 2016년 8월 : 한국과학기술원 전기 및 전자공학 박사
 2016년 11월~현재 : 한국전자통신연구원 표준연구본부 선임연구원

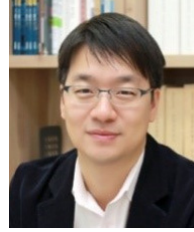
<관심분야> 이동통신 망 구조 및 주파수 관리 기법

신 명 기 (Myung-Ki Shin)



2003년 8월 : 충남대학교 대학원 컴퓨터공학과 박사
1994년 2월~현재 : 한국전자통신연구원 네트워크표준연구실 실장/책임연구원
<관심분야> 5G 네트워크, SDN, NFV, 미래인터넷

박 형 곤 (Hyunggon Park)



2004년 2월 : 포항공과대학교 전자공학과 졸업
2006년 3월 : University of California, Los Angeles (UCLA) M.S.
2008년 12월 : University of California, Los Angeles (UCLA) Ph.D.

2010년~현재 : 이화여자대학교 전자전기공학과 부교수
<관심분야> 멀티에이전트 네트워크 시스템, 머신러닝 기반 분산적 의사 결정 전략, 게임이론 기반 네트워크 분산적 자원 관리, 네트워크 코딩