https://doi.org/10.7840/kics.2024.49.10.1368

강화학습을 이용한 간섭 관리 및 자원 할당 최적화 연구

박 건 아◆, 최 계 원°

A Study on Interference Management and Resource Allocation Optimization Using Reinforcement Learning

Geon-a Park◆, Kae-won Choi°

요 약

무선 통신 환경에서 사용자의 증가하는 요구를 충족시키기 위해 초고밀도 네트워크(Ultra-Dense Network, UDN)의 도입이 필수적이다. UDN은 높은 기지국 밀도를 특징으로 하며, 이를 통해 네트워크 용량을 증가시키고 신호의 강도 및 신뢰성을 개선하여 서비스 품질을 향상시킨다. 그러나 이러한 고밀도화는 기지국 간의 간섭을 유발하여 네트워크 성능에 부정적인 영향을 미친다. 이에 따라, 효과적인 간섭 관리는 자원 할당을 최적화하는 데 있어 필수적이다. 본 논문에서는 실제 네트워크 환경을 모사하여 간섭 관리와 자원 할당 최적화 문제를 해결하기 위해 Proximal Policy Optimization(PPO) 알고리즘을 적용하는 방법을 제안한다. 제안된 방법은 채널 상태와 네트워크 공정성을 고려한 자원 할당 최적화를 통해 무선 통신 시스템의 전반적인 효율성과 성능을 향상시킬 수 있는 가능성을 보여준다. 이 연구는 UDN 환경에서의 자원 할당 및 간섭 관리 전략에 대한 실질적인 해결책을 제공함으로써 무선 통신 시스템의 성능 개선에 기여할 것으로 기대된다.

키워드 : 무선 통신, 자원 할당, 강화 학습, PPO 알고리즘
Key Words : Wireless Communication, Resource Allocation, Reinforcement Learning, PPO Algorithm

ABSTRACT

In the wireless communication environment, the introduction of Ultra-Dense Networks (UDN) is essential to meet the increasing demands of users. UDNs are characterized by a high density of base stations, which enhance network capacity and improve signal strength and reliability, thereby improving service quality. However, such densification leads to inter-base station interference that adversely affects network performance. Consequently, effective interference management is crucial for optimizing resource allocation. This paper proposes the application of the Proximal Policy Optimization (PPO) algorithm to address interference management and resource allocation optimization challenges by simulating real network environments. The proposed method demonstrates the potential to enhance the overall efficiency and performance of wireless communication systems through resource allocation optimization that considers channel conditions and network fairness. This research is expected to provide practical solutions for resource allocation and interference management strategies in UDN environments, contributing to the enhancement of wireless communication system performance.

◆ First Author : Sungkyunkwan University, Department of Digital Media and Communication Engineering, geona@g.skku.edu, 정회원
° Corresponding Author : Sungkyunkwan University, department of Electrical and Computer Engineering, kaewonchoi@skku.edu, 종신회원
논문번호 : 202404-074-B-RN, Received April 23, 2024; Revised June 22, 2024; Accepted June 26, 2024

Ⅰ. 서 론

무선 통신 기술은 고속 데이터 전송과 모바일 인터넷 접근성의 혁신을 주도하며 초연결성, 초고속 데이터 전송을 가능하게 한다. 이러한 기술의 발전과 함께 사용자 수와 데이터 사용량의 급증하였고, 이에 맞게 5G 네트워크에서는 사용자 밀집 지역에서의 높은 서비스 품질을 유지하기 위해 고밀도 네트워크(Ultra-Dense Network, UDN) 환경이 필수적이게 된다[1]. UDN은 매우 높은 기지국 밀도를 통해 네트워크 용량을 증가시키고, 신호의 강도와 신뢰성을 향상시켜 사용자 서비스 품질 경험을 개선한다. 그러나 기지국의 고밀도 배치는 기지국 간의 간섭을 증가시키는 주된 원인으로 네트워크 성능에 부정적인 영향을 주게 된다[2]. 따라서 효과적인 간섭 제어 기술은 이러한 UDN 환경에서 자원 할당과 네트워크 관리의 효율성을 극대화하기 위해 필수적이다. 간섭을 적절하게 관리하고 자원을 효율적으로 할당함으로써, 네트워크는 사용자 요구를 충족시킬 수 있게 된다.

본 연구는 실제와 유사한 무선 환경을 구현하여 해당 환경에서의 자원 할당과 간섭 관리 문제에 초점을 맞추고, 해결 방안으로 모델 프리(Model-Free) 강화학습[3] 방식 중 하나인 Proximal Policy Optimization(PPO) 알고리즘[4]을 적용한다. 강화학습은 에이전트가 환경과의 상호작용을 통해 최적의 행동 전략을 학습하는 인공 지능의 한 분야로, 고밀도 5G 환경에서 네트워크 간섭을 효율적으로 관리하고 자원 할당을 최적화하는 데 중요한 기법으로 평가된다. 본 연구는 Proportional Fairness(PF) 스케줄링 기법[5]과의 비교를 통해 제안된 접근법의 효과성을 검증한다. 이러한 분석을 통해 고밀도 무선 통신 환경에서의 간섭 관리 및 자원 할당 문제의 해결 방법을 제시하고, 무선 통신 시스템의 성능을 개선할 수 있는 강화학습 방식의 적용을 탐구한다.

Ⅱ. 연구 제안

본 연구의 핵심 목표 중 하나는 고밀도 네트워크 환경과 유사한 조건을 정밀하게 구현하여 자원 할당과 간섭 관리 전략의 실질적 효과와 적용 가능성을 평가하는 것이다. 이를 위해 복잡한 실제 네트워크 조건을 정밀하게 재현하는 세 가지 핵심 구성요소인 기지국, 사용자 단말, OFDM 채널을 활용하여 복잡한 실제 네트워크 조건을 상세히 재현하고, 강화학습을 이용해 간섭 관리 및 자원할당 전략을 학습한다.

2.1 기지국

실제 도시 환경을 정밀하게 모사하기 위해 3D 모델링 도구를 사용하여 상세 모델을 생성한다. 그림1.은 성균관 대학교 캠퍼스의 지도를 기반으로 사용자의 밀집 가능성이 있는, 간섭이 예상되는 지역을 중심으로 복잡한 건축 구조와 지형을 정밀하게 모델링한 결과를 보여준다. 건물의 위치와 높이 정보를 이용하여 기지국 위치를 결정하고, 기지국의 방향성과 경사각을 같이 고려하였다.

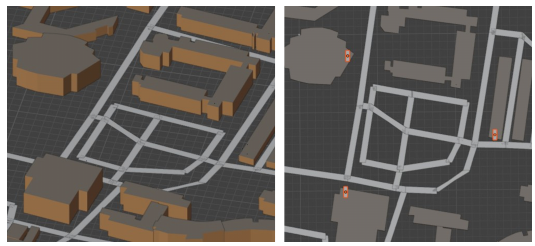

그림 1. 도시 환경 지도에서의 3D 모델링과 기지국 위치
Fig. 1. 3D modeling and base station location on urban environment maps

2.2 사용자 단말(User Equipment, UE)

고밀도 네트워크 환경에서는 단말의 고밀도 분포와 고속 이동성이 네트워크 성능에 중대한 영향을 미친다. 이에 따라 단말의 이동성을 정확하게 모델링하고 예측하는 것은 네트워크 스케줄링 전략 최적화에 필수적이다. 도시 규모의 교통 흐름과 이동성을 시뮬레이션 하는 SUMO를 활용한다. 그림 2. 에서는 그림 1.에서 구축한 도시 모델링 환경을 바탕으로 실제 도시 환경 내에서 단말의 위치, 속도 및 이동 경로를 시간에 따라 추적한다. 추적된 데이터는 단말의 이동 패턴을 정확하게 파악하여 네트워크 자원 할당 및 관리 전략을 최적화하는

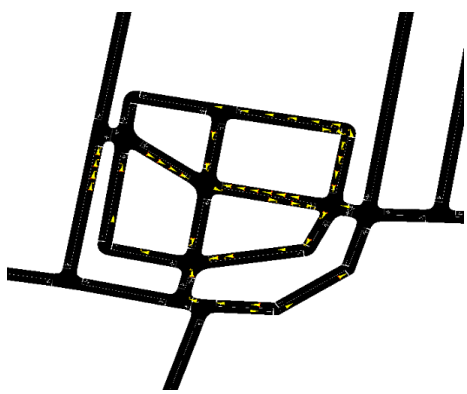

그림 2. 단말의 이동성 시뮬레이션
Fig. 2. Mobility simulation of UE

데 필수적인 정보를 제공한다[6].

2.3 OFDM 채널

무선 통신 시뮬레이션인 Sionna Ray Tracing을 사용하여 기지국과 단말 간의 전파 경로를 계산하고[7], 이를 활용하여 채널 정보를 생성한다. Ray Tracing은 무선 신호가 환경 내에서 어떻게 반사, 굴절, 회절 되는 지를 계산하여 빌딩과 같은 다른 장애물에 의해 어떻게 영향을 받는지 정밀하게 예측할 수 있다. 시뮬레이션을 통해 시간 도메인에서의 무선 채널 특성을 표현하는 채널 임펄스 응답(Channel Impulse Response, CIR) 데이터를 수집할 수 있다[8]. CIR 데이터는 시간 영역에서의 채널의 특성을 나타내며, 이를 푸리에 변환을 통해 주파수 영역으로 변환하여 주파수 도메인에서의 채널 특성을 파악할 수 있다. 푸리에 변환을 통해 주파수 영역의 표현에서는 각 주파수 별 주요한 채널 특성인 채널 이득 데이터를 얻게 된다.

그림 3.에서는 기지국에서 단말로 전달되는 신호를 직접 경로, 반사, 굴절 회절 경로와 같은 다양한 전파 경로를 시각적으로 확인할 수 있다.

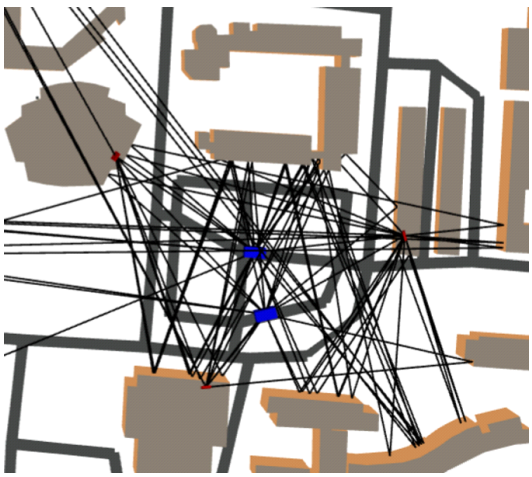

그림 3. Ray tracing 기반 OFDM 채널 시뮬레이션
Fig. 3. Ray tracing based OFDM channel simulation

2.4 강화학습

실제 환경을 모사하여 얻은 채널 이득 데이터를 이용하여 기지국-단말 연결 및 자원 할당 전략을 OpenAI Gym[9]을 사용하여 개발한다. 기지국-단말 간의 최적 연결을 결정하고, 서브캐리어 단위에서 자원 블록(Resource Block, RB)단위로의 재구성의 전처리 단계를 통해 실험에서 사용되는 데이터는 무선 통신 시스템의 실제 운영 환경을 반영하여 에이전트가 관측할 수 있는 환경을 형성하는데 필수적인 정보를 제공한다.

에이전트에게 전달되는 관측치(Observation)은 변환된 채널 파워 데이터와 기지국-단말 간의 연결 정보, 현재까지의 단말들의 평균 처리량(Average Throughput) 데이터를 포함한다. 이러한 관측치는 에이전트가 현재 네트워크 상태를 이해하고 최적의 스케줄링 전략을 결정하는데 중요한 기초를 마련한다.

행동(Action) 선택 과정에서는 에이전트가 확률을 기반으로 스케줄링 할 사용자 단말을 선택한다. 이는 각 기지국 별 RB에 대한 할당 여부를 결정하는데 사용되며, 간섭을 제어하기 위해 특정 기지국에서는 특정 RB에 대해 아무것도 할당하지 않는 선택지 또한 포함된다. 이 과정을 통해 에이전트는 네트워크의 현재 상태와 전략적 목표에 기반하여 최적의 액션을 도출할 수 있다.

보상(Reward) 메커니즘은 스케줄링된 사용자 단말들의 신호 대 잡음 비(SNR)가 설정된 임계 값(SNR threshold) 이상인지를 평가하며, SNR이 임계 값보다 낮은 경우 할당에 실패하였음을 나타낸다. 추가적으로 현재 사용자 단말들의 처리량을 기반으로 한 유틸리티 함수를 통해 네트워크의 공정성(Fairness)도 평가한다. 이러한 보상 구조는 에이전트가 네트워크의 전반적인 성능 및 공정성을 동시에 개선할 수 있는 전략을 학습하도록 한다.

Ⅲ. 실험 결과

3.1 실험 시나리오 및 학습 매개변수

본 실험에서는 실제와 유사한 네트워크 환경을 모사하기 위해, 사용된 채널 이득 데이터를 기반으로 네트워크의 핵심 구성 요소인 기지국 수, 단말 수, 서브캐리어 수를 각각 정의하였고, 표 1에서 확인할 수 있다.

두 서브캐리어 간의 주파수 간격, 즉 서브캐리어 스페이싱은 120kHz로 설정하여 무선 통신 시스템에서 흔

표 1. 네트워크 시뮬레이션을 위한 실험 파라미터
Table 1. Experimental Parameters for Network Simulation.

Parameter	Value
Base station	3
UE	50
RB	10
SNR threshold	15dB
Subcarrier	120
Subcarrier spacing	120kHz
Noise	-174dBm/Hz

히 보이는 고밀도 네트워크 환경을 정밀하게 재현하였다. 또한 120개의 서브캐리어를 활용하여 무선 통신 시스템의 RB를 정의하였다. 무선 통신 시스템의 효율적인 자원 관리와 할당을 위해 12개의 서브캐리어를 하나의 묶음으로 구성하여 총 10개의 RB를 할당하도록 하였다. 소음 전력 밀도(N0)의 값은 -174dBM/Hz로 설정하고, 이 값은 에이전트에 전달되는 관측치에서 신호대 잡음비(SNR) 계산에 사용하였다.

다음은 실험에서 사용된 강화학습 알고리즘의 hyper-parameter에 대해 설명한다. 채널 이득 데이터는 0.1초 간격으로 수집되며 총 200초 동안 2000개의 타임 스텝을 포함한다. 실험은 총 100번의 에피소드로 구성되어 있으며, 각 에피소드는 무선 통신 네트워크의 한 시나리오를 시뮬레이션 한다. 각 에피소드 동안 발생하는 모든 타임 스텝에서의 데이터는 experience buffer에 저장된다. 에피소드가 종료될 때마다 저장된 데이터를 사용하여 학습이 진행된다. 학습 과정에서의 배치 크기는 128로 설정하였다. PPO 알고리즘의 클립 매개변수는 0.2로 설정하여 정책 업데이트 시 너무 큰 변화를 방지하였다. Learning rate는 0.0005로 설정하여 학습 과정의 안정성과 효율성을 주었다. 보상의 할인된 누적 계산에 영향을 주는 GAE에서의 Lambda 값은 0.95로 설정하였고, 미래 보상에 대한 현재 가치를 결정하는 Discount factor 또한 0.95로 설정하였다.

표 2. 강화학습 시뮬레이션을 위한 파라미터
Table 2. Parameters for Reinforcement Learning Simulation.

Parameter	Value
Episode	100
Time step	2000
Batch size	128
Learning rate	0.0005
Lambda	0.95
Discount factor	0.95
Clip	0.2

3.2 학습 및 시뮬레이션 결과

본 연구는 실제 채널을 모사하여 얻은 채널 데이터로 PPO 기반 강화학습 알고리즘을 적용하여 네트워크 스케줄링 최적화를 진행하였다. 실험은 Jain's Fairness Index[10]와 Average Throughput의 변화를 관찰하여 학습 에피소드가 진행됨에 따라 네트워크 성능이 어떻게 개선되는 지를 평가한다. 또한 마지막 에피소드에서의 전체 사용자 장비의 평균 처리량 변화를 통해 PPO 기반 스케줄링 성능을 사용자의 평균 데이터 전송률을 기준으로 자원을 할당하는 알고리즘인 PF스케줄링과 비교하여 분석하였다.

3.2.1 Jain's Fairness Index

Jain's Fairness Index를 사용하여 네트워크의 공정성을 측정하였다. 학습 에피소드가 진행됨에 따라 Jain's Fairness Index는 그림 4.와 같이 점진적으로 증가하는 경향을 보였다. 이는 PPO 알고리즘이 네트워크 자원을 공정하게 분배하는 방식을 학습하며, 각 단말에 대한 서비스 품질을 균등하게 개선하고 있음을 의미한다. 비교군으로 설정된 PF 스케줄링과의 성능 비교 또한 그림 4.에서 확인할 수 있듯이 PF 스케줄러 대비 PPO 알고리즘이 더 공정한 스케줄링을 하고 있음을 확인할 수 있었다.

그림 4. 에피소드 별 Jain's Fairness 변화
Fig. 4. Episode-wise Variation in Jain's Fairness Index

3.2.2 평균 처리량(Average Throughput)

Average Throughput 역시 학습 에피소드를 거치며 증가하는 것을 확인할 수 있었다. 이는 그림 5.의 그래프와 같이 네트워크 총 처리량을 극대화하는 방향으로 학습하고 있음을 알 수 있다.

학습이 끝난 후 마지막 에피소드에서의 타임 스텝별로 관찰한 전체 단말의 평균 처리량은 PPO기반 스케줄링이 네트워크 자원을 효율적으로 활용하여 단말들의 데이터 전송률을 상당히 향상시킬 수 있음을 보여준다. 시간 흐름에 따른 그림 6.의 평균 처리량 비교 그래프에서 PPO 기반 강화학습 알고리즘이 대체적으로 더 높은 평균 처리량을 제공하는 것을 확인할 수 있었다. 그러나 특정 타임 스텝에서 PF 스케줄러가 더 높은 평균 처리량을 나타내는 부분 또한 관찰되었다. 이는 PF 스케줄러가 특정 조건 하에서는 사용자 단말 간의 공정성을 우선시하면서도 효율적인 자원 할당을 통해 높은 처리량을 달성할 수 있음을 보여준다. 이러한 현상은 네트워

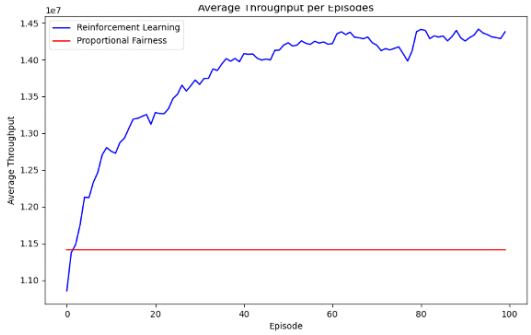

그림 5. 에피소드 별 Average Throughput 변화
Fig. 5. Episode-wise Variation in Average Throughput

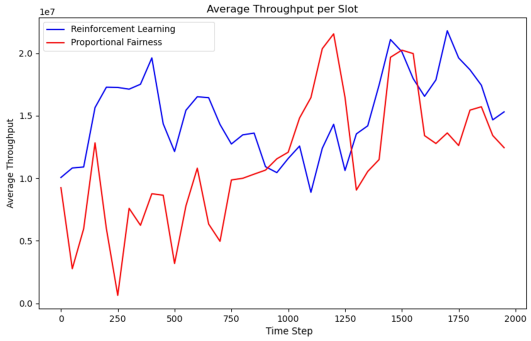

그림 6. Time step 별 강화학습과 PF 스케줄러의 Average throughput 비교
Fig. 6. Comparison of Average Throughput Between Reinforcement Learning and PF Scheduler Across Time Steps

크 상태의 변화나 사용자 단말의 특성 등에 따라 최적의 스케줄링 전략이 변할 수 있음을 나타낸다.

Ⅳ. 결 론

본 연구에서는 PPO 기반 강화학습 알고리즘을 통해 무선 통신 네트워크 내에서의 최적 스케줄링 및 자원 할당 전략을 탐색하였다. 실제 환경을 모사한 데이터를 가지고 얻은 실험 결과는 강화학습 알고리즘의 적용이 네트워크의 평균 처리량과 공정성을 동시에 개선할 수 있음을 시사한다. Jain's Fairness Index와 Average Throughput의 측정을 통해, 학습 에피소드가 진행됨에 따라 네트워크 성능의 점진적인 향상이 관찰되었다. 이는 PPO 알고리즘이 네트워크 상태와 사용자 요구를 정확히 반영하여 최적의 결정을 내릴 수 있는 능력을 갖추었음을 나타낸다.

또한, 본 연구에서는 PF 스케줄링과의 성능 비교를 통해 PPO 기반 스케줄링의 우수성을 입증하였다. 시간에 따른 평균 처리량의 변화를 분석한 결과, PPO 알고리즘은 대체적으로 더 높은 처리량을 제공하였으나, 특정 조건에서 PF 스케줄링이 우수한 성능을 나타내는 경우도 있었다. 이러한 현상은 네트워크 환경과 사용자 요구의 변화에 따른 스케줄링 전략의 유연한 적용 필요성을 강조한다.

본 연구의 결과는 무선 통신 네트워크의 자원 할당 및 스케줄링 최적화에 강화학습 알고리즘을 효과적으로 적용할 수 있음을 보여준다. PPO 알고리즘의 적용을 통해 각 사용자에게 공정하게 자원을 할당하면서도 높은 네트워크 평균 처리량을 동시에 달성할 수 있음을 확인하였다. 또한, PPO 알고리즘은 동적인 네트워크 환경에서 사용자의 서비스 품질을 극대화할 수 있는 새로운 방법을 제시한다. 이러한 접근 방식은 미래의 5G 및 차세대 무선 네트워크 설계와 운영에 있어 중요한 시사점을 제공할 것으로 기대된다.

References

[1] D. López-Pérez, M. Ding, H. Claussen, and A. H. Jafari, "Towards 1 Gbps/UE in cellular systems: Understanding ultra-dense small cell deployments," *IEEE Commun. Surv. and Tuts.*, vol. 17, no. 4, pp. 2078-2101, Oct. 2015. (https://doi.org/10.1109/COMST.2015.2439636)

[2] J. Liu, M. Sheng, L. Liu, and J. Li, "Interference management in ultra-dense networks: Challenges and approaches," *IEEE Netw.*, vol. 31, no. 6, pp. 70-77, Nov. 2017. (https://doi.org/10.1109/MNET.2017.1700052)

[3] R. S. Sutton and A. G. Barto, *Reinforcement learning: an introduction*, MIT Press, 1998.

[4] J. Schulman, F. Wolski, P. Dhariwal, A. Radford, and O. Klimov, "Proximal policy optimization algorithms," *Comput. Sci.*, Jul. 2017. (https://doi.org/10.48550/arXiv.1707.06347)

[5] C. Wengerter, J. Ohlhorst, A. Golitschek, and E. Von Elbwart, "Fairness and throughput analysis for generalized proportional fair frequency scheduling in OFDMA," *2005 IEEE 61st Veh. Technol. Conf.*, 2005.

[6] D. Krajzewicz, J. Erdmann, M. Behrisch, and L. Bieker, "Recent development and

applications of SUMO-simulation of urban mobility," *Int. J. Advances in Syst. and Measurements*, vol. 5, no. 3 and 4, 2012.

[7] J. Hoydis, et al., "Sionna: An open-source library for next-generation physical layer research," Mar. 2022. (https://doi.org/10.48550/arXiv.2203.11854)

[8] Y. J. Noh and K. W. Choi, "A study on the construction of high-precision digital twins through 3D modeling," in *Proc. KICS Int. Conf. Commun.*, pp. 828-829, 2023.

[9] G. Brockman, et al., "OpenAI Gym," Jun. 2016. (https://doi.org/10.48550/arXiv.1606.01540)

[10] R. K. Jain, D.-M. W. Chiu, and W. R. Hawe, "A quantitative measure of fairness and discrimination," *Eastern Research Laboratory*, vol. 21, p. 1, Hudson, MA, 1984.

박 건 아 (Geon-a Park)

2015년 2월 : 충북대학교 정보통신공학부 졸업

2015년 3월~현재 : 삼성전자 책임 연구원

2023년 3월~현재 : 성균관대학교 DMC 공학과 석사

<관심분야> 무선통신, 강화학습

[ORCID:0009-0003-9199-6746]

최 계 원 (Kae-won choi)

2007년 8월 : 서울대학교 전기컴퓨터공학부 박사

2010년 9월~2016년 8월 : 서울과학기술대학교 컴퓨터공학과 조교수

2016년 9월~2023년 8월 : 성균관대학교 전자전기컴퓨터공학과 부교수

2023년 9월~현재 : 성균관대학교 전자전기컴퓨터공학과 교수

<관심분야> 무선통신, 강화학습, 레이더

[ORCID:0000-0002-3680-1403]