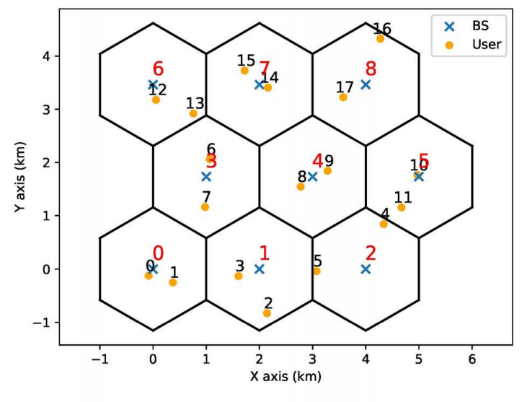
power allocation in multi-user cellular networks with deep q learning approach

cited by 23 times, [ICC 2019 - 2019 IEEE International Conference on Communications (ICC)](https://ieeexplore.ieee.org/xpl/conhome/8753818/proceeding)

1. Environment

Distributed dynamic downlink power allocation with multiple users and an interfering multiple-access channel(IMAC)

Contribution

1. Transfer learning(off-line train -> on-line train)

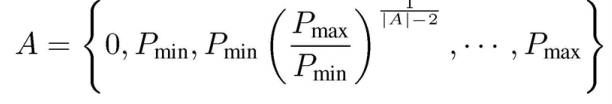
2. No future rewards

3. After centralized training, tested by distributed execution

1. State : optimal p를 current state의 CSI g\_t로만 찾기는 힘들어서 c, p라는 보충 요소 등장



1. Action : DQN의 action은 discrete해야하기 때문에 A-1 level로 나누어줌



1. Reward : 정교하게 디자인해도 대부분이 suboptimal로 수렴하기 때문에 그냥 downlink를 바로 reward.



1. Deep neural network

* 4 layer feed-forward
* The number of neurons of 2 hidden layers = 128, 64
* Activation function of output = linear, of 2 hidden = ReLU

n : BS, k : user

Independent channel gain(CSI information)

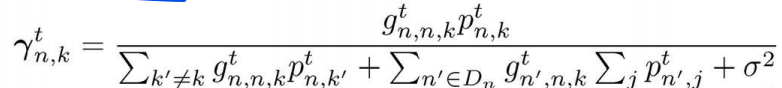
Small scale complex fading element

Large scale fading component

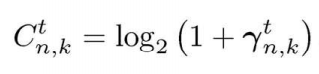
first-order complex Gauss-Markov process

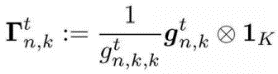


, where J0 = first kind zero-order Bessel function, f\_d = maximum doppler frequency, T\_s = time interval



D\_n = set of interference cells around the n-th cell. P = emitting power of BS, sigma = noise power

downlink rate of this link.



agent는 perfect CSI information가지고 있다고 가정 후, logarithmic normalized interferer set 정의, 1k is a vector filled with K ones.

Multiple Channel Access using Deep Reinforcement Learning for congested vehicular Networks

Cited by 2 times, [2020 IEEE 91st Vehicular Technology Conference (VTC2020-Spring)](https://ieeexplore.ieee.org/xpl/conhome/9121635/proceeding)

1. Environment :

* self-experience-based CW adaptation algorithm employing DRL
* vehicle(agents) broadcast the safety packet using V2V communication and receive transmission results from a VANET. Consequently, vehicles learn to adjust the optimum CW
* DSRC(Dedicated Short Range Communication)의 multi-channel operation : CCH(control-channel), SCH(service-channel)으로 나누어지는데 각 interval = 50ms. 그래서 모든 vehicles은 100ms마다 safety packet 전송(in CCH interval ; CCHI), SCH interval ; SCHI에는 가장 가까운(지정된) vehicle(node)에 unicast ACK 전송. SCHI 때 target vehicle 선택 & transmit ACK

1. state : <CW, F, S>

CW : contention window

F : frequency value

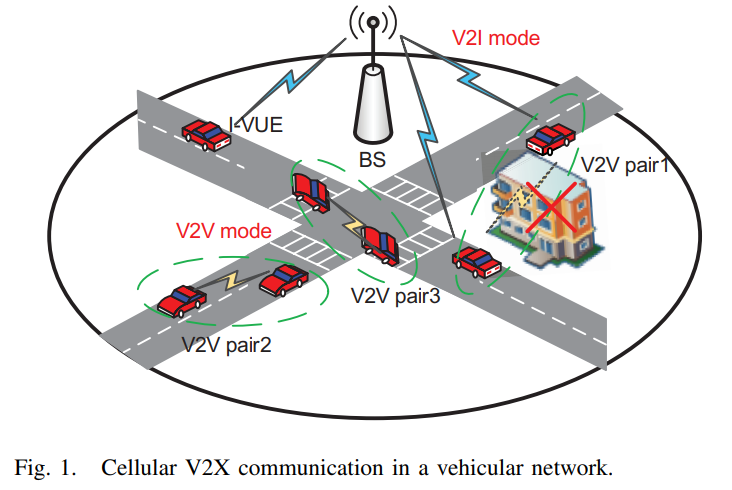
S : success rate

1. action : 3가지(keep, increase, decrease)
2. reward : broadcast 성공시 +1, 실패시 -1 -> 그래서 SCHI 때 각 agent의 broadcast성공인지 판단 가능
3. Deep neural network

* 3 hidden layers with the number of neurons 256, 128, 64 using Leaky-Relu as activation function.

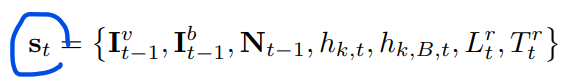
DQELR : An adaptive Deep Q-Network-based energy- and latency-aware routing protocol design for underwater acoustic sensor networks.

Cited by 33 times, [IEEE Access](https://ieeexplore.ieee.org/xpl/RecentIssue.jsp?punumber=6287639) ( Volume: 7)

1. Environment

BS : 중앙 위치, VUE : random distributed, single antenna

1. State



I\_v : the received interference power at the V2V receiver

I\_b : the received interference power at the BS

N : the number of selected neighbors

h\_k,t : the large scale channel gain from the V2V transmitter to its corresponding V2V receiver

h\_k,B,t : the large scale channel gain from the V2V transmitter to the BS

L : current load

T : remaining time to meet the latency threshold

1. Action

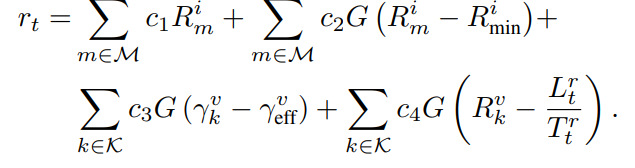


A : the RB allocation

S : communication mode selection

P : transmit power level of the V2V transmitter

1. Reward



1st : sum capacity revenue of I-VUEs

2nd : penalty of unsatisfied capacity for I-VUEs

3rd, 4th : impacts of the reliability and latency requirement

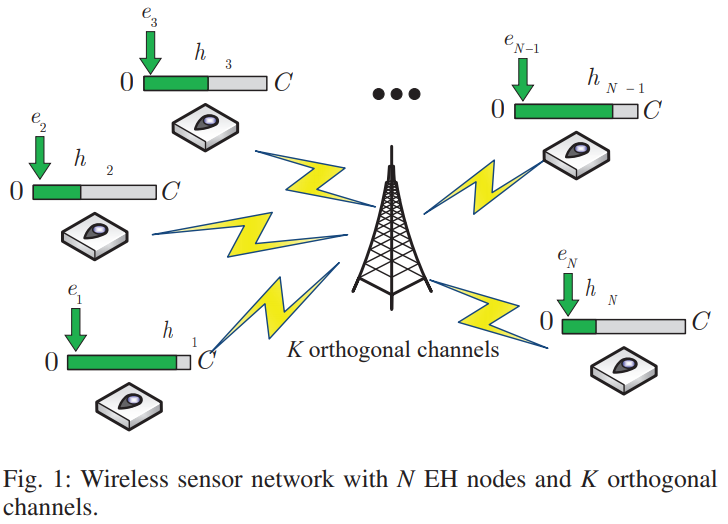
1. Deep neural network

* 1 hidden layer(256), 모두 fully connected layer,

Partially Observable Double DQN Based IoT Scheduling for Energy Harvesting

Cited by 3 times [2019 IEEE International Conference on Communications Workshops (ICC Workshops)](https://ieeexplore.ieee.org/xpl/conhome/8751668/proceeding)

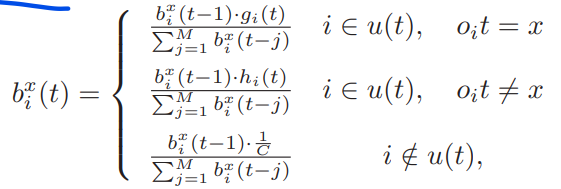
1. Environment

POMDP라 가운데 BS가 node의 상태를 완전히 알 수 없다.

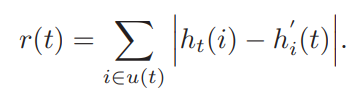
1. State

MDP라면 전체 정보를 알기에 이렇게 정의를 하겠지만,,



POMDP이기에 b = belief state 정의

1. Action : scheduling policy(broadcast to all nodes and receive the information about current power)
2. Reward

h : the power information about the scheduled node

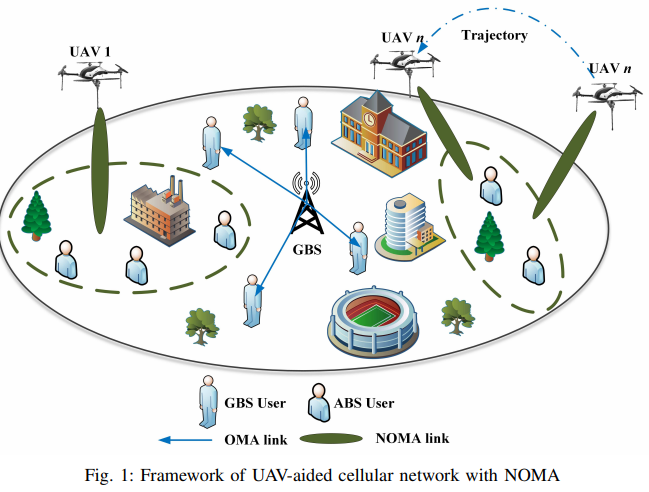
h` : the residual power to BS again after attempting to transit data

1. Deep neural network(keras in tensorflow)

3 fully connected layer(Sequential), hidden layer의 뉴론 수는 (input 뉴론 + out 뉴론) / 2, ReLU

Multi-Agent Reinforcement Learning in NOMA-aided UAV Networks for Cellular Offloading

Cited by 0 times, [arXiv.org](https://arxiv.org/)

1. Environment

야외 down-link, 유저 많음, 가운데 GBS

UAV : single 안테나, NOMA(intra-cell 간섭 영향받음)

UAV : GBS와는 다른 주파수대역씀

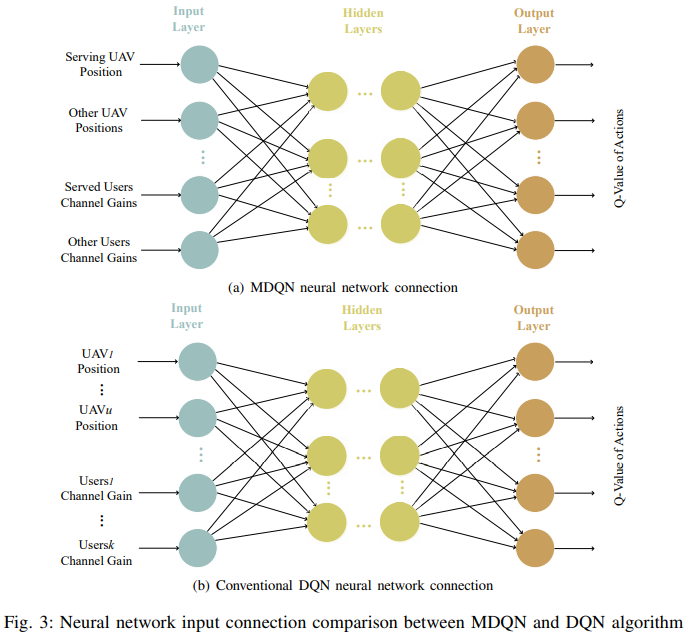
유저 : random roaming + directional walking

.Cluster : All user served, 반복 x, helpful to reduce 간섭

UAV : 유저 위치 체크, re-clustering

최적화 : user clustering + optimization for trajectory and power allocation

online  
여기서 특이한 점이, nn을 모든 agent가 공통모델로 학습하는데 1번 agent가 nn에 연결되면 나머지 agent에게는 제한을 건다.(아래 그림 참조)



1. State



L\_u : connecting agent(UAV)의 3차원 좌표

L\_s : 다른 UAV들 좌표(inter-cluster 간섭 원인)

g\_u : 연결된 유저들의 channel gain

g\_s : 다른 UAV와 연결된 유저들의 channel gain

1. Action

* Movement action space : 7개(수평 왼, 수평 오, 수평 앞, 수평 뒤, 수직 위, 수직 아래, 그대로)

If out of bound, action default is hover

* Power allocation action space : multiple gears

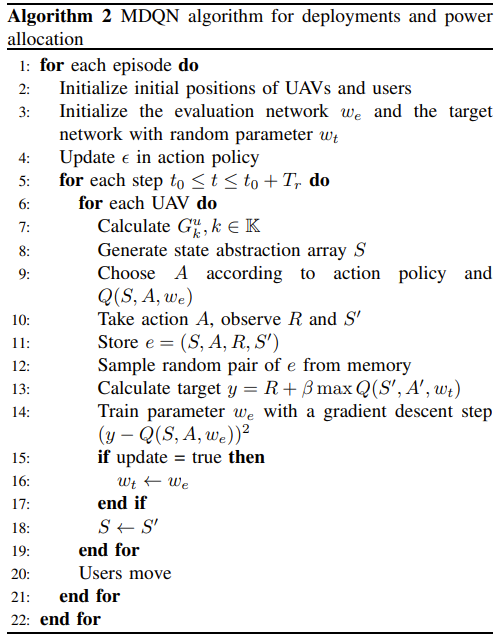
1. Reward : total throughput under constraints



R : sum data rate(multi agent니까 리워드를 total sum으로 설정!)

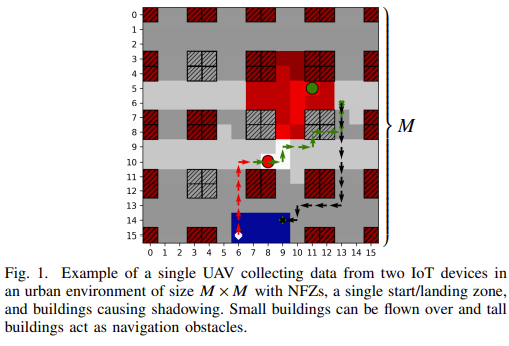
Gamma : penalty coefficient(QoS 최대한 보장키 위한 수단)

1. Deep neural network

* 3 layers (a 40 nodes hidden layer), ReLU, MSE, Adam Optimizer
* 

Multi-UAV Path Planning for Wireless Data Harvesting with Deep Reinforcement Learning

Cited by 1 times, [arXiv.org](https://arxiv.org/)

1. Environment

Square grid world = start/landing position + position UAV cannot occupy + Obstacles blocking wireless link

A team of UAV(동일한 UAV)의 path planning problem : Dec-POMDP

제약 : trajectory + battery + urban environment + wireless(random signal blocking events)

Contribution : 시나리오 파라미터에 여유를 주는 일반화가 가능한 DRL method by using “centered global-local map processing”

* Flying time의 constraint : Dec-POMDP with full reward function description
* Dec-POMDP를 Deep multi-agent RL로 해결
* Dual global-local map processing : 큰 맵과 state spaces에 대한 학습과 적응 효율성에 있어 이점 using map centering!
* Parameter generalization : the learned policy can be reused over a wide array of scenario parameters.
* UAV model
* UAV의 state = 3D position + operation status + battery level
* UAV의 action = 6가지, 1 time slot에 움직이는 거리 = cell size
* Link performance model
* Communication time slot이 mission time slot보다 작게 정의해서 활용
* IoT Device sensor has a finite amount of Data
* UAV-to-ground channel model : links with LOS/NLOS + path loss + shadow fading
* Multiple access protocol
* TDMA : a UAV to various ground user, inter-UAV interference는 없음, IoT device는 multi-band node로 작동해서 all UAV와 통신 가능(Scheduling decision은 action space에 속하지 않음
* Dec-POMDP = state space + joint action space + transition probability function + reward function + joint observation space + observation function + discount facto
* r

1. State

-environment information(Landing Zone + NFZs + Obstacles) + Agents(UAV Positions + Flying Times + Operational Status) + Devices(Device Positions + Device Data)

1. Action

* Safety controller so as to collision avoidance + NFZ + obstacle avoidance + excluding landing + evaluates(해당 agent의 action을 수용할지 안할지, 안한다면 hovering)

1. Reward

* Collective reward + individual penalty(when safety controller rejected) + individual penalty(when not landing) + constant movement penalty

1. Deep neural network