JKIICE Journal of the Korea Institute of Information and Communication Engineering

한국정보통신학회논문지 Vol. 23, No. 8: 1011~1017, Aug. 2019

무선 애드혹 네트워크에서 노드분리 경로문제를 위한 강화학습

장길웅*

Reinforcement Learning for Node-disjoint Path Problem in Wireless Ad-hoc Networks

Kil-woong Jang*

*Professor, Department of Data Informatics, Korea Maritime and Ocean University, Busan 49112, Korea

요 약

본 논문은 무선 애드혹 네트워크에서 신뢰성이 보장되는 데이터 전송을 위해 다중 경로를 설정하는 노드분리 경로문제를 해결하기 위한 강화학습을 제안한다. 노드분리 경로문제는 소스와 목적지사이에 중간 노드가 중복되지 않게 다수의 경로를 결정하는 문제이다. 본 논문에서는 기계학습 중 하나인 강화학습에서 Q-러닝을 사용하여 노드의 수가 많은 대규모의 무선 애드혹 네트워크에서 전송거리를 고려한 최적화 방법을 제안한다. 특히 대규모의 무선 애드혹 네트워크에서 노드분리 경로 문제를 해결하기 위해서는 많은 계산량이 요구되지만 제안된 강화학습은 효율적으로 경로를 학습함으로써 적절한 결과를 도출한다. 제안된 강화학습의 성능은 2개의 노드분리경로를 설정하기 위한 전송거리 관점에서 평가되었으며, 평가 결과에서 기존에 제안된 시뮬레이티드 어닐링과 비교평가하여 전송거리 면에서 더 좋은 성능을 보였다.

ABSTRACT

This paper proposes reinforcement learning to solve the node-disjoint path problem which establishes multipath for reliable data transmission in wireless ad-hoc networks. The node-disjoint path problem is a problem of determining a plurality of paths so that the intermediate nodes do not overlap between the source and the destination. In this paper, we propose an optimization method considering transmission distance in a large-scale wireless ad-hoc network using Q-learning in reinforcement learning, one of machine learning. Especially, in order to solve the node-disjoint path problem in a large-scale wireless ad-hoc network, a large amount of computation is required, but the proposed reinforcement learning efficiently obtains appropriate results by learning the path. The performance of the proposed reinforcement learning is evaluated from the viewpoint of transmission distance to establish two node-disjoint paths. From the evaluation results, it showed better performance in the transmission distance compared with the conventional simulated annealing.

키워드 : 노드분리 경로문제, 강화학습, Q-러닝, 무선 애드혹 네트워크

Key word : Node-disjoint path problem, reinforcement learning, Q-learning, wireless ad-hoc networks

Received 9 May 2019, **Revised** 10 May 2019, **Accepted** 3 July 2019

*** Corresponding Author** Kil-woong Jang(E-mail:jangkw@kmou.ac.kr, Tel:+82-51-410-4375)
Professor, Department of Data Informatics, Korea Maritime and Ocean University, Busan 49112, Korea

Open Access http://doi.org/10.6109/jkiice.2019.23.8.1011 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/li-censes/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. 서 론

실시간 서비스를 요구하는 다양한 멀티미디어 응용에서 네트워크 관리의 신뢰성과 견고성은 중요한 요소가 된다. 5G와 같은 초고속 네트워크나 대량의 노드가 배치된 네트워크에서 트래픽 경로의 소실이 발생하면 데이터 전송의 품질이 급속하게 저하되는 원인이 된다. 경로 소실로 인한 데이터 전송 품질을 복원하는 한 가지 방법으로 다수의 경로를 분리하여 설정하는 것이다[1, 2].

네트워크상에서 경로를 분리하는 문제는 회로 구조 설계, 네트워크 구조 설계 등 여러 응용분야에서 적용되고 있다. 네트워크상에서 소스에서 목적지로 가는 다중의 경로에서 중간 노드가 겹치지 않게 설정하는 것을 노드분리 경로문제(node-disjoint path problem)라고 한다[3]. 무선 애드혹 네트워크에서 노드가 겹치지 않게 다중 경로를 설정하는 것은 네트워크의 신뢰성을 높일 수 있다. 네트워크에서 노드가 분리된 경로를 2개 이상 설정한다면, 첫 번째 경로는 연결 및 데이터전송을 담당하고, 나머지는 첫 번째 경로에 문제가 발생했을 때 백업용으로 사용되거나, 혼잡 제어(congestion control), 부하균형(load balancing) 등과 같은 경우로 사용될 수 있다.

기계학습(machine learning)은 1950년대 후반에 인공지능 기술로 도입되었다. 시간이 지남에 따라, 기계학습은 진화되고 더 계산적으로 실행 가능한 알고리즘으로 전환되었다. 지난 10년 동안 기계학습은 생물 정보학, 음성 인식, 컴퓨터 비전과 같은 다양한 응용 분야에서 분류, 회귀 및 밀도 추정을 포함한 광범위한 작업에 사용되고 있다. 강화학습(reinforcemnet learning: RL)은 기계학습의 한 가지 패러다임으로서, 에이전트(agent)가 환경(environment)과의 시행착오에서 가장 큰 보상을 산출하는 전략으로 학습한다. 가장 대표적인 강화학습 알고리즘은 Q-러닝 알고리즘이다. Q-러닝은 Q-함수를 통해 에이전트의 상태와 동작을 나타내는 Q-값을 갱신함으로써 적절한 정책을 학습하게 되고 좋은 결과를 생성한다[4].

본 논문에서는 무선 애드혹 네트워크에서 전송거리를 고려한 노드분리 경로문제를 해결하기 위한 강화학습을 제안하고 성능을 평가한다. 전송거리를 최소화하는 노드분리 경로문제는 조합 최적화문제이며 NP-complete로 증명되어있다[5]. 일반적으로 NP 문제는 최적해를 구하는 데 방대한 계산량과 시간을 요구된다. 이를 해결하기 위해 기존에 많은 연구에서는 휴리스틱 방법을 이용하여 해결하였다[5, 6]. 본 논문에서는 무선 애드혹 네트워크에서 전송거리를 최소화하기 위한 노드분리 경로문제에 대하여 Q-러닝 알고리즘을 이용한 강화학습을 제안한다. 본 논문에서는 제안된 강화학습에 대한 구조와 동작방법에 대해 기술하고, 다양한 실험조건에서 경로의 전송거리측면에서 기존에 제안된 메타휴리스틱 알고리즘인 시뮬레이티드 어닐링과 비교 평가한다.

Ⅱ. 관련연구

네트워크상에서 데이터 전송의 신뢰성을 높이기 위한 방법은 다양하게 제안되었다. Xiong *et al*. [5]은 경로를 증가시키는 방법을 사용하여 최단경로를 가진 다수의 노드분리경로를 구하는 방법을 제안하였다. 제안된 방법은 한 개의 소스에서 여러 목적지로 최단거리 알고리즘을 사용하여 노드분리경로를 찾는다. Kim [6]은 시뮬레이티드 어닐링을 이용한 메타 휴리스틱 알고리즘을 제안하였으며, 이동 노드를 가진 무선 애드혹 네트워크에서 다중경로를 찾는 방법을 제안하였다. Lin *et al*. [7]은 전송지연과 링크효율 관점에서 노드분리경로를 찾는 방법을 제안하였다. 제안된 방법은 전송 에너지와 네트워크 장애 간에 트래픽 제어를 통해 균형을 맞추고 있다. Hsu *et al*. [8]은 링크가 겹치지 않는 링크분리경로를 최대화하는 유전 알고리즘을 제안하였다. 제안된 알고리즘은 그리디 알고리즘 및 개미 군집 알고리즘과 비교하여 알고리즘의 우수성을 비교하였다. Fisher *et al*. [9]은 다수의 케이블을 논리적으로 하나의 링크로 결합하여 경로를 결정하는 방법을 제안하였다. Kurt *et al*. [10]은 그리디 알고리즘을 적용한 노드분리 경로 알고리즘을 제안하였다. 인접한 노드 중에 가장 가까운 거리를 우선 선택하고 이미 선택된 노드는 경로에서 제외하는 방법을 사용하였다.

강화학습을 노드분리 경로문제에 적용한 연구는 없지만 다양한 연구에서 적용되고 있다. Forster *et al*. [11]은 다중의 싱크를 가진 무선 센서 네트워크에서 경로 최적화를 위해 강화학습을 제안하였다. Li *et al*. [12]은 무선 센서 네트워크에서 무선 비콘을 이용하여 노드의 위치를 파악하기 위한 강화학습을 제안하였다. Yu. *et al*.

[13]은 이종의 무선 네트워크에서 매체접근제어를 위해 딥러닝을 이용한 강화학습을 제안하였다.

Ⅲ. 노드분리 경로문제를 위한 강화학습

3.1. 시스템 모델

무선 애드혹 네트워크에서 노드분리 경로문제에 적용할 네트워크 모델과 제약조건은 다음과 같다. 본 논문에서 사용되는 네트워크는 n개의 노드를 가지며, 노드 사이의 링크는 노드 간의 거리를 가중치로 가진다. 이때 노드 집합을 V, 가중치를 가진 링크 집합을 E로 나타내며, 네트워크는 가중치가 가진 비방향성 그래프 $G(V, E)$로 표현한다. 모든 노드는 최대전송반경 R_{max}을 가지며, 본 논문에서는 모두 동일하다고 가정한다. 각 링크의 거리는 최대전송반경보다 작거나 같으며, 유클리드 함수에 의해 계산된다.

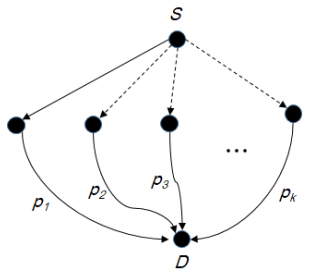


Fig. 1 Example of k node-disjoint paths

그림 1은 소스 S에서 목적지 D로 가는 k개의 노드분리 경로를 나타낸 것이다. 설정된 경로들의 집합은 $P = \{p_1, p_2, p_3,..., p_k\}$로 표현되며, 노드분리 경로문제에서 발생하는 전송거리 d_P는 다음과 같다.

$$\delta_P = \sum_{i=1}^{k} \delta(p_i) \tag{1}$$

본 논문에서 무선 애드혹 네트워크에서 노드분리 경로문제에 적용되는 목적함수는 전송거리를 최소화하는 것이다. 따라서 앞서 제시한 네트워크 모델에서 노드분리 경로문제의 전송거리를 최소화하는 것은 조합 최적화 문제로 정식화할 수 있으며 다음과 같다.

최소화

$$\delta_P = \delta_S + \sum_{x \in P, x \neq S} \delta_x \tag{2}$$

여기서 d_S는 소스 노드가 데이터를 전송할 때 필요한 전송거리를 나타내며, d_x는 결정된 경로의 중간 노드인 x의 전송거리를 나타낸다. 그림 1에서 소스 노드는 k개의 출력 링크를 가지며, 목적 노드를 제외한 경로 상에 존재하는 중간 노드는 다음 노드로 데이터를 전송하기 위해 출력 링크를 한 개씩 가진다. 여기서 소스 노드는 무선 네트워크의 브로드캐스팅 방법을 이용하여 최대 전송거리 내에 있는 인접한 노드로 데이터 전송함으로써 전체 전송거리를 줄일 수 있는 효과가 있다.

3.2. 강화학습

강화학습은 주어진 환경에서 에이전트가 학습을 통해 기대하는 보상을 최대로 받기 위해 상황별로 행동을 결정하는 것이다. 그림 2와 같이 특정 시간 t에서 에이전트는 환경 상태를 관찰하고 상태 s_t의 환경에서 행동 a_t를 취하면, 환경으로부터 보상 r_t를 받고, 환경은 새로운 상태 s_{t+1}로 변하는 과정을 반복하면서 강화학습이 수행된다.

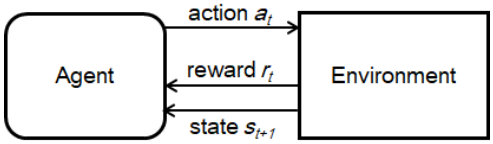


Fig. 2 Reinforcement learning process

강화학습이 적용되는 문제는 마코프 결정과정으로 표현할 수 있으며, 마코프 결정과정으로 강화학습을 기술할 때 명시되어야 할 요소는 다음과 같다. 환경에서 나타날 수 있는 상태의 집합 $S = \{s_1, s_2, ..., s_n\}$, 에이전트가 할 수 있는 행동의 집합 $A = \{a_1, a_2, ..., a_m\}$, 특정 시점 t의 환경 상태 s_t에서 행동 a_t가 수행될 때 다음 상태 s_{t+1}로 상태전이를 결정하는 규칙은 (s_t, a_t, s_{t+1})로 표현되며, 각 상태에서 취할 행동을 결정해 놓은 것을 정책 p라고 한다. 상태전이 (s_t, a_t, s_{t+1})에 의해서 상태 s_{t+1}이 될 때 에이전트에게 주어지는 보상 값을 결정하는 함수 R_t는 다음과 같다.

$$R_t = \sum_{k=t}^{\infty} \gamma^{k-t} r_{k+1} \tag{3}$$

여기서 γ은 할인율(discount rate)을 나타내며, 상태 s에서 행동 a를 취한 다음 정책 π에 따라 행동하게 될 때 얻게 되는 기대보상값을 반환하는 함수를 행동 가치함수 $Q^{\pi}(s, a)$라고 하며 다음과 같다.

$$Q^{\pi}(s, a) = E[R_t | s_t = s, a_t = a, \pi] \tag{4}$$

행동 가치함수 $Q^{\pi}(s, a)$를 Q-함수라고도 하며, Q-함수는 어떤 상태 s에서 행동 a를 하는 것이 얼마나 좋은지 계산한다. 최적 행동 가치함수 $Q^{*}(s, a)$는 가장 좋은 행동 가치함수 값을 주는 정책을 따르는 경우의 행동 가치함수이며, 다음과 같이 벨만 최적 방정식을 따른다.

$$\begin{aligned} Q^{*}(s, a) &= \max_{\pi} Q^{\pi}(s_{t+1}, a_{t+1}) \\ &= E_{s'}[r_{t+1} + \gamma \max Q^{*}(s_{t+1}, a_{t+1}) | s_t = s, a_t = a, \pi] \end{aligned} \tag{5}$$

여기서 r_{t+1}은 보상값을 나타낸다. 최적 행동 가치함수가 구해지면, 최적 정책 π는 다음과 같이 구해진다.

$$\pi^{*}(s) = \mathrm{argmax}(Q^{*}(s, a)) \tag{6}$$

여기서 *argmax*()는 현재 상태 s에 대해서 가장 큰 행동 가치함수 값을 주는 행동 a를 선택하는 함수이다.

3.3. 노드분리 경로문제

앞서 기술한 일반적인 강화학습을 바탕으로 본 논문에서 해결하고자 하는 노드분리 경로문제를 위한 제안된 강화학습을 기술한다.

에이전트 상태와 행동 : 본 논문에서는 노드분리 경로문제에 대해서 에이전트 상태는 목적 노드의 ID와 현재 노드에 인접한 노드에 대한 정보로 정의한다. 에이전트 상태에 따라 가능한 행동이 결정된다. 에이전트 행동은 현재 노드에서 인접 노드로 하나의 가능한 경로를 결정하는 것이다. 앞서 기술한 바와 같이 행동의 집합 $A = \{a_1, a_2, ..., a_m\}$은 각 노드마다 다르게 결정된다. 각 노드는 서로 다른 인접 노드를 가지기 때문에 에이전트는 현재 노드의 인접 노드의 수에 따라 행동의 결정은 달라진다.

Q-값과 갱신 : Q-값은 에이전트의 행동이 각 노드에서 적절한지를 나타내는 것이다. 즉 소스 노드에서 목적 노드로의 경로가 적절한지를 나타내는 지표이다. 본 논문에서는 제안된 강화학습의 Q-값은 목적 노드에 도착하기 위한 경로 상의 전송거리를 사용한다. 수식 (5)를 이용하여 Q-값을 결정한다. 시간 t의 (s_t, a_t) 상태에서 보상이 r_{t+1}일 경우, Q-러닝은 다음과 같이 갱신된다.

$$Q(s_t, a_t) \leftarrow Q(s_t, a_t) + \alpha[R(a_{t+1}) - Q(s_t, a_t)] \tag{7}$$

여기서 α는 학습율(learning rate)을 나타내며, $\alpha \in (0, 1)$이다. 본 논문에서 $\alpha = 0.1$로 설정하였다. 에이전트는 학습하는 동안 에이전트의 행동에 따라 환경으로부터 보상값을 받아야 한다. 보상값은 일반적으로 수식 (5)처럼 Q-값 중에 최대값을 취하는 방식을 취하지만 본 논문에서는 다루는 노드분리 경로문제는 전송거리를 Q-값으로 사용하기 때문에 Q-값의 최대값 대신에 최소값으로 수정하여 다음과 같이 사용한다.

$$R(a_{t+1}) = r_{t+1} + \gamma \min_a Q(s_{t+1}, a) \tag{8}$$

여기서 할인율 $\gamma = 1.1$로 설정하였다.

정책 : Q-러닝은 각각의 상태와 행동 (s, a)에서 반복적으로 $Q^{*}(s, a)$를 추측함으로써 정책을 결정한다. 본 논문에서의 정책은 다음 3가지 방법을 사용한다. 첫 번째 방법은 ε-그리디(ε-greedy) 방법으로 에이전트가 1 - ε 확률로 식 (6)과 같은 행동을 결정하거나 ε 확률로 랜덤한 행동을 결정하는 방법이다. 랜덤한 행동을 선택하는 이유는 최적 행동 가치함수 $Q^{*}(s, a)$에 수렴하지 않은 상태에서 항상 같은 행동을 하는 것을 막고 새로운 영역을 검색하기 위해서이다. 이때 식 (6)과 같이 기존에 사용된 경로를 사용하는 것을 활용(exploit)이라고 하며, 랜덤하게 행동을 선택하는 것을 탐험(exploration)이라고 한다. 두 번째 방법은 퇴색 ε-그리디(decay ε-greedy) 방법으로 초기 학습에는 ε 확률을 높게 하여 주로 랜덤하게 행동을 선택하다가 학습 횟수가 증가하면 학습 횟수

에 비례하게 ε 확률을 낮춤으로써 기존의 경로를 주로 학습하게 하는 방법이다. 세 번째 방법은 추가 랜덤 노이즈(add random noise) 방법으로 현재 위치의 $Q(s, a)$에 랜덤하게 노이즈를 추가하여 이 중 가장 작은 Q-값을 가진 경로를 선택하는 방법이다.

위와 같은 정책을 사용하여 소스 노드에서 목적 노드까지 경로를 학습시킨다. 학습과정은 에이전트가 정책에 따라 현재의 노드에서 이웃한 노드를 선택하여 이동한다. 각 노드는 인접한 노드와 연결된 링크에 대하여 Q-값을 가지며 보상을 받을 때마다 Q-값을 갱신한다. 이러한 과정을 반복함으로써 최종적으로 Q-값이 작은 경로를 선택하여 최적의 경로를 결정한다.

Ⅳ. 성능평가

이번 장에서는 무선 애드혹 네트워크에서 노드분리 경로문제를 해결하기 위해 제안된 강화학습의 성능을 평가한다. 성능평가를 위해 컴퓨터 시뮬레이션을 수행하였으며, 메모리 4GB와 3.6GHz 인텔 CPU 프로세서로 이루어진 윈도우 기반의 운영체제 하에서 컴퓨터 시뮬레이션이 수행되었다. 제안된 알고리즘과 평가 비교된 각 알고리즘은 파이썬으로 구현하여 성능평가가 이루어졌다. 제안된 알고리즘은 노드가 분리된 경로 상의 전송거리 측면에서 기존에 제안된 시뮬레이티드 어닐링 [7]과 비교 평가하였다.

성능평가에 사용된 네트워크는 1000×1000 m^2 크기를 가지며, 최대전송범위가 200 m인 노드를 랜덤하게 배치하여 구성하였다. 또한 다양한 노드의 밀도를 구성하기 위해 노드의 수를 200에서 1000까지 100씩 증가하며 수행하였으며, 소스 노드의 위치를 (0, 0), 목적 노드의 위치를 (1000, 1000)으로 설정하여 노드 간의 거리를 최대로 설정하였다. 제안된 강화학습은 노드 수 n을 가진 각 실험에서 노드 수와 비례하게 $n \times 10000$번을 학습시켰으며, 각 시뮬레이션은 10번씩 시도하여 평균값으로 결과를 나타내었다.

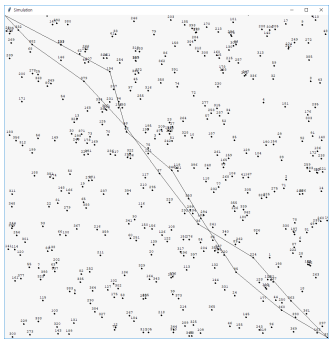
표 1은 세 가지 탐험 방식을 적용한 제안된 강화학습과 시뮬레이티드 어닐링을 적용한 결과를 나타낸 것이다. ε-그리디 방식에서는 $\varepsilon = 0.1$로 설정하였으며, 학습 반복횟수와 관계없이 일정하게 ε 확률을 유지한다. 퇴색 ε-그리디 방식은 $\varepsilon = 1.0 / ((t / n) + 1)$로 계산되며, 이때 t는 학습 횟수를 나타낸다. 이 방식은 학습이 반복될수록 탐험하는 비율은 줄어들고 활용방식이 증가하게 된다. 즉 학습이 진행될수록 이전 결과가 좋은 경로를 많이 찾아가는 방식을 취한다. 추가 랜덤 노이즈 방식은 현재 각 노드의 Q-값에 0과 R_{max} 사이의 랜덤값을 더하여 그 중 가장 낮은 값을 가진 경로를 선택하도록 하였다. 반면에 메타휴리스틱 알고리즘인 시뮬레이티드 어닐링은 랜덤하게 2개의 경로를 생성한 후 그 경로 상에 있는 노드를 인접한 노드로 변경하거나 삭제하는 이웃해 생성방식을 적용하여 새로운 경로를 만든다. 새로 생성된 경로의 전송거리와 이전에 생성된 경로들의 최소 전송거리를 비교한다. 비교 결과에서 새로 생성된 경로가 현재까지의 경로보다 전송거리가 짧을 경우 새로 생

Table. 1 Comparison of total transmission distance of proposed RL and simulated annealing

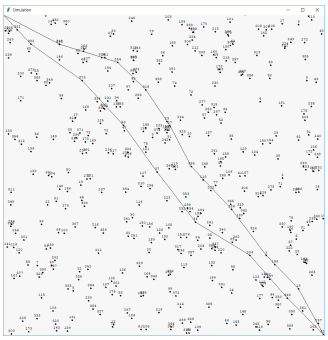
n	Proposed RL						Simulated annealing	
	ε-greedy		decaying ε-greedy		add random noise		path1	path2
	path1	path2	path1	path2	path1	path2		
200	1433	1456	1433	1496	1433	1512	1443	1506
300	1429	1438	1429	1461	1429	1442	1439	1452
400	1423	1467	1423	1474	1423	1508	1443	1501
500	1416	1420	1416	1457	1416	1426	1453	1456
600	1419	1444	1419	1488	1419	1512	1434	1494
700	1418	1428	1418	1454	1418	1429	1421	1460
800	1420	1439	1420	1439	1420	1455	1440	1459
900	1424	1428	1428	1479	1428	1441	1438	1471
1000	1417	1419	1419	1424	1419	1447	1429	1454

성된 경로가 최적의 경로로 설정되고 이 경로를 이용하여 다음 단계의 경로를 생성한다. 만약 현재까지의 경로보다 새로 생성된 경로의 전송거리가 길 경우에는 사용되는 조건 값에 따라 다음 경로를 생성하는 경로가 결정된다. 이러한 절차에 따라서 지정된 횟수만큼 새로운 경로 생성 과정을 반복한다. 시뮬레이션 결과에서 제안된 강화학습이 시뮬레이티드 어닐링과 비교했을 때 전반적으로 우수한 결과를 나타내고 있으며, 제안된 강화학습의 정책 간에는 ε-그리디 방식이 다른 방식에 비해 근소하게 좋은 결과를 나타내고 있음을 볼 수 있다. 제안된 강화학습이나 시뮬레이티드 어닐링은 노드분리 경로문제와 같은 NP-complete 문제에 대해서 부분적으로 지역해에 빠져 최적해를 구하지 못할 수 있다. 특히 시뮬레이티드 어닐링과 같은 메타 휴리스틱 알고리즘은 초기해 생성이나 이웃해 생성 방식에 따라 그 결과의 변동성이 높게 나타날 수 있다. 반면에 강화학습은 정책에 영향을 받기는 하지만 에이전트가 상태에 따라 행동을 자동적으로 취하기 때문에 더 일관적으로 최적해에 가깝게 도달하게 된다. 따라서 성능평가에서 제안된 강화학습이 시뮬레이티드 어닐링보다 더 우수한 결과를 나타내는 것으로 판단된다.

그림 3은 ε-그리디 방식을 사용한 제안된 강화학습과 시뮬레이티드 어닐링을 사용하여 나타나는 경로를 그림으로 나타낸 것이다. 네트워크에 노드가 랜덤하게 배치하고 있음을 볼 수 있으며, 좌표가 (0, 0)인 소스 노드에서 좌표가 (1000, 1000)인 목적 노드로 중간 노드가 겹치지 않게 2개의 경로가 설정된 것을 볼 수 있다. 그림에서 제안된 강화학습과 시뮬레이티드 어닐링이 서로 다른 경로를 사용하여 노드분리 경로가 설정됨을 볼 수 있다.



(a)



(b)

Fig. 3 Simulation results of ε-greedy RL and simulated annealing, where n = 400 and R_{max} = 200.
(a) ε-greedy RL (b) simulated annealing

그림 4는 노드의 수가 400일 때, 강화학습의 학습 횟수에 따른 전송거리의 값을 나타낸 것이다. 경로 2의 경우에 반복횟수가 $n \times 4000$일 때 1510을 나타내다가 학습 횟수가 증가할수록 점점 전송거리가 수렴함을 볼 수 있다. 그림에는 나타나지 않지만 학습 횟수가 $n \times 2000$ 이하 일 때는 학습이 제대로 되지 않아 경로 설정이 이루어지지 않았다.

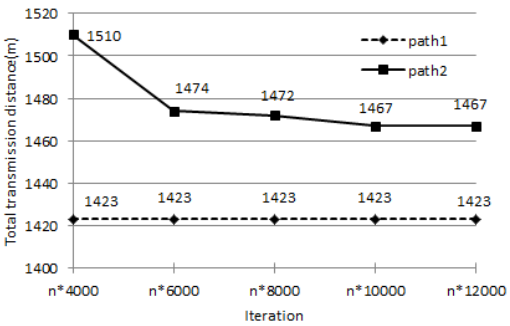


Fig. 4 Convergence speeds of ε-greedy RL according to iteration, where n = 400 and R_{max} = 200

Ⅴ. 결 론

본 논문은 무선 애드혹 네트워크에서 신뢰성 있는 데이터 전송을 위해 노드가 겹치지 않는 다중 경로를 설정하기 위하여 Q-러닝을 이용한 강화학습을 제안하였다. 정의된 노드분리 경로문제에 대하여 전송거리를 최소화하기 위한 강화학습을 제안하였으며, Q-러닝 알고리

즘으로 사용하였다. 제안된 강화학습은 주어진 네트워크 환경에서 에이전트가 각 노드의 상태에 따라 행동을 결정하고 보상받는 방법으로 2개의 경로를 설정하였다. 제안된 강화학습을 성능 평가하기 위해 소스 노드에서 목적 노드까지 설정된 경로상의 전송거리 측면에서 메타휴리스틱 알고리즘인 시뮬레이티드 어닐링과 비교 평가하였다. 성능평가 결과에서 제안된 강화학습은 기존의 방식보다 우수한 결과를 도출하였으며, 복잡한 구조를 가진 대규모의 네트워크에서도 노드분리 경로문제를 해결할 수 있으며, 신뢰성이 요구되는 네트워크에서 효과적으로 적용될 수 있을 것으로 판단된다. 향후 과제로 테이블을 이용하는 Q-러닝 방식이 아닌 딥러닝을 이용한 심층강화학습으로 노드분리 경로문제를 해결하고자 한다.

REFERENCES

[1] G. Lin, S. Soh, K. Chin, and M. Lazarescu, "Efficient heuristics for energy-aware routing in networks with bundled links," *Computer Networks*, vol. 57 no. 8, pp. 1774-1788, 2013.

[2] D. Zhang, Q. Liu, L. Chen, W. Xu, and K. Wang, "Multi-layer based multi-path routing algorithm for maximizing spectrum availability," *Wireless Networks*, vol. 24 no. 3, pp. 897-909, 2018.

[3] R. Hadid, M. H. Karaata, and V. Villain, "A Stabilizing Algorithm for Finding Two Node-Disjoint Paths in Arbitrary Networks," *International Journal of Foundations of Computer Science*, vol. 28, no. 4, pp. 411-435, 2017.

[4] M. A. Alsheikh, S. Lin, D. Niyato, and H. P. Tan, "Machine learning in wireless sensor networks: algorithms, strategies, and applications", *IEEE Communications Surveys and Tutorials*, vol. 16. no. 4, pp. 1996-2018, April 2014.

[5] K. Xiong, Z. Qiu, Y. Guo, and H. Zhang, "Multi-constrained shortest disjoint paths for reliable QoS routing," *ETRI Journal*, vol. 31, no. 5, pp. 534-44, 2009.

[6] S. Kim, "Adaptive MANET multipath routing algorithm based on the simulated annealing approach," *Scientific World Journal*, vol. 2014, pp. 1-8, 2014.

[7] G. Lin, S. Soh, K. W. Chin, and M. Lazarescu, "Energy aware two disjoint paths routing," *Journal of Network and Computer Applications*, vol. 43, pp. 27-41, 2014.

[8] C. Hsu and H. Cho, "A genetic algorithm for the maximum edge-disjoint paths problem," *Neurocomputing*, vol. 148, pp. 17-22, 2015.

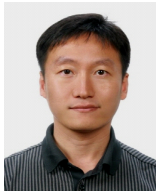
[9] W. Fisher, M. Suchara, and J. Rexford, "Greening back bone networks: reducing energy consumption by shutting off cables in bundled links," in *Proceeding of ACM SIGCOMM Workshops on Green Networking*, pp. 29-34, 2010.

[10] M. Kurt, M. Berberler, and O. Ugurlu, "A new algorithm for finding vertex-disjoint paths," *The International Arab Journal of Information Technology*, vol. 12, no. 6, pp.550 – 555, 2015.

[11] A. Forster and A. L. Murphy, "FROMS: feedback routing for optimizing multiple sinks in WSN with reinforcement learning", in *Proceeding of IEEE Intelligent Sensors, Sensor Networks and Information Processing Conference*, pp. 371-376, 2007.

[12] S. Li, X. Kong, and D. Lowe, "Dynamic Path Determination of Mobile Beacons Employing Reinforcement Learning for Wireless Sensor Localization", in *Proceeding of IEEE International Conference on Advanced Information Networking and Applications Workshops*, pp. 760-765, 2012.

[13] Y. Yu, T. Wang, and S. C. Liew, "Deep-Reinforcement Learning Multiple Access for Heterogeneous Wireless Networks", in *Proceeding of IEEE International Conference on Communications*, pp. 1- 7, 2018.



장길웅(Kil-woong Jang)

1997년 2월 : 경북대학교 컴퓨터공학과 학사
1999년 2월 : 경북대학교 컴퓨터공학과 석사
2002년 8월 : 경북대학교 컴퓨터공학과 박사
2003년 3월 ~ 현재 : 한국해양대학교 데이터정보학과 교수
※ 관심분야 : 네트워크 프로토콜, 네트워크 최적화