JKIICE

Short Paper

한국정보통신학회논문지 Vol. 25, No. 12: 1960~1963, Dec. 2021

다중경로 통신 시스템에서 톰슨 샘플링을 이용한 경로 선택 기법

Thompson sampling based path selection algorithm in multipath communication system

Byung Chang Chung*

*Assistant Professor, Department of Information and Communication Engineering, Gyeongsang National University, Tongyoung, 53064 Korea

ABSTRACT

In this paper, we propose a multiplay Thompson sampling algorithm in multipath communication system. Multipath communication system has advantages on communication capacity, robustness, survivability, and so on. It is important to select appropriate network path according to the status of individual path. However, it is hard to obtain the information of path quality simultaneously. To solve this issue, we propose Thompson sampling which is popular in machine learning area. We find some issues when the algorithm is applied directly in the proposal system and suggested some modifications. Through simulation, we verified the proposed algorithm can utilize the entire network paths. In summary, our proposed algorithm can be applied as a path allocation in multipath-based communications system.

Keywords : Multi-armed bandit, Machine learning in communications, Multipath TCP, Reinforcement learning

Ⅰ. 서 론

최근 통신 시스템의 요구사항은 빠른 속도에만 그치지 않고, 생존성, 신뢰성 등 다양한 요구사항들이 등장하는 추세이다. 이러한 다양한 요구사항을 만족할 수 있도록 다중경로를 이용한 통신 시스템이 제안되었다. 다중경로를 이용하는 방식은 통신의 어느 계층에서 수행하느냐에 따라 그 이름에 조금씩 차이를 보인다. LTE 접속 기술과 WiFi접속 기술이 다중경로로 동시에 이용되는 경우를 가정해보자. 먼저 LTE와 WiFi가 링크 계층에서 결합하는 경우는 3GPP의 LAA(licensed assisted access), LWA(LTE-WiFi link aggregation) 등으로 표준화되어 있다[1]. 둘째로 LTE와 WiFi가 전달 계층에서 결합하는 경우, IETF에서 MPTCP(multipath TCP) 기술로 표준화되어있다[2].

이러한 다중경로 통신 시스템에서는 서비스별로 어떤 경로를 할당해주는지가 매우 중요하다[3-4]. 그림 1은 이러한 다중경로 전송의 예시를 보여준다. 각각의 경로는 경로의 물리적 특성, 대역폭, 기존 트래픽 양 등의 이유로 서로 다른 품질을 가진다. 이러한 환경에서의 경로 할당 문제는 통신망에서의 자원 할당 방안과 유사한 방식으로 해결이 가능하고, 최근에는 기계학습 등을 이용하여 해결한 연구도 존재한다[3]. 또한, 다양한 망이 공존하므로 망 선택의 결과에 따라 특정 망에 포화가 발생할 수 있으므로 이를 관리하기 위한 혼잡 제어 방안 또한 제안되었다[4].

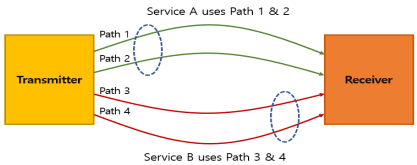

Fig. 1 System model of Multi-path system

본 논문에서는 기계학습의 방안 중 하나인 다중 톰슨 샘플링 기술을 이용하여 다중경로 환경에서의 경로 선택 문제를 풀고자 한다. 다중 톰슨 샘플링 기술의 가장 큰 장점은 기존의 강화학습이 행동을 한 가지 선택하던 부분을 여러 가지 선택으로 쉽게 확장할 수 있는 점이다. 이러한 특징이 논문의 환경인 다중경로 통신 시스템과 가장 유사하다고 판단되어 해당 알고리즘을 선택하였고, 해당 알고리즘 적용 시 발생하는 문제점을 알고리즘 수정을 통해 해결하였다.

Received 7 October 2021, **Revised** 13 October 2021, **Accepted** 17 November 2021

*** Corresponding Author** Byung Chang Chung (E-mail:bcchung@gnu.ac.kr, Tel:+82-55-772-9175)
Assistant Professor, Department of Information and Communication Engineering, Gyeongsang National University, Tongyoung, 53064 Korea

Open Access http://doi.org/10.6109/jkiice.2021.25.12.1960 print ISSN: 2234-4772 online ISSN: 2288-4165

Ⅱ. 톰슨 샘플링 기법

본 장에서는 강화학습 기법 중 하나인 톰슨 샘플링 기법에 대해 설명할 것이다. 톰슨 샘플링을 적용할 수 있는 환경인 multi-armed bandit 환경에 대해 간단히 소개 후 톰슨 샘플링 기법에 대해 소개하겠다.

2.1. Multi-armed bandit

Multi-armed bandit을 이해하기 위해서 가장 많이 언급되는 내용이 슬롯머신이다. 확률분포에 따라 보상을 지급하는 여러 개의 슬롯머신이 존재하는 환경에서 한 타임슬롯에 한 슬롯머신만 조작 가능한 플레이어를 가정하자. Multi-armed bandit 문제는 플레이어가 T번의 타임슬롯 동안 슬롯머신들을 조작했을 때 취득하는 보상의 총합을 최대화하는 문제이다[5]. 이 문제를 다시 정리하면, 플레이어는 매 시간 t마다 K개의 슬롯머신 중 하나를 선택하여 확률 분포에 따라 보상 r(t)을 얻는다. 이 상황에서 문제는 보상 r의 시간 누적 합 최대화 문제가 된다.

2.2. Thompson sampling

앞서 언급한 multi-armed bandit 문제를 해결하기 위한 방안 중 하나로, 톰슨 샘플링이라는 알고리즘이 존재한다[6]. 톰슨 샘플링은 구글, 넷플릭스 등 커머셜 분야에서도 응용되는 알고리즘으로, 해당 문제에서 높은 성능을 보여주고 있다.

톰슨 샘플링에서는 각 순간의 행동과 보상 조합과 확률 분포 파라미터를 이용해 likelihood 함수를 설계하고, prior를 가정하여 솔루션을 구하는 것이다. 이 때 각각의 arm이 Bernoulli 분포를 따른다고 가정하면, prior는 beta 분포가 되고 아래와 같이 알고리즘을 구성할 수 있다.

<톰슨 샘플링 알고리즘>

1: for t=1, ..., T do
2: for k=1, ..., K do
3: Draw θ_i with beta(s_i, f_i)
4: end
5: Draw arm $i^* = \arg\max_i \theta_i$
6: Observe reward r(t)
7: if r(t)=1 then
8: $s_i = s_i + 1$
9: else
10: $f_i = f_i + 1$
11: end if
12: end for

위 알고리즘은 베르누이 분포를 가지는 bandit에서의 톰슨 샘플링 알고리즘이다. 이 알고리즘은 기존의 알고리즘 등과 비교하였을 때 상대적으로 낮은 복잡도와 높은 성능을 가지는 것으로 알려져있다. 한 가지 장점을 추가적으로 이야기하면 톰슨 샘플링은 multiplay 환경에서도 유리하다[7]. Multiplay 상황이란, 한 번에 하나의 arm에 대해서만 보상을 받는 것이 아닌 여러 개의 arm을 플레이할 수 있는 상황을 뜻한다. Multiplay 환경에서 톰슨 샘플링을 적용할 수 있는 여러 가지 방법이 존재하나, 여기서는 알고리즘 상 5번째 줄에서 arm을 한 개 고르는 것이 아닌 최대화 순서대로 여러 개 고르는 방법을 가정하도록 하겠다.

Ⅲ. 다중경로에서의 톰슨 샘플링

톰슨 샘플링은 그 방법의 간단함과 multiplay의 적용 확장성이 높으므로, 다중경로 선택 알고리즘의 후보로 매우 적절하다고 볼 수 있다. 하지만 후술할 문제들로 인하여 다중경로 시스템에서 톰슨 샘플링을 직접 적용하기 어렵다. 본 장에서는 다중경로 시스템에 톰슨 샘플링을 직접 적용했을 때의 문제를 알아보고, 이를 해결하는 개선방안을 제안한다.

3.1. 다중경로 시스템에서의 톰슨 샘플링 문제

다중경로 시스템에서 톰슨 샘플링을 그대로 적용하기에는 두 가지 문제가 있다. 첫째는 우리의 시스템에서 플레이어가 가질 수 있는 행동의 경우의 수가 상당히 많다는 점이다. 통신 시스템의 전송 단말 입장에서 바라보면, 다양한 서비스들이 고유의 제약 조건을 가지고 존재한다. 이것을 하나의 행동으로 모델링하기에는 행동의 경우의 수가 너무 높아져 톰슨 샘플링이 수렴하는데 너무 많은 시간이 걸린다. 각각의 서비스를 개별 플레이어로 보고 자체적으로 톰슨 샘플링을 수행하게 되면, 행동 경우의 수는 multiplay 톰슨 샘플링을 적용하기 알맞게 된다. 그리고 각 서비스 고유의 제약 조건을 만족하는가의 여부에 따라 성공과 실패를 누적하여 플레이를 반복

하도록 하면 첫 번째 문제를 해결 가능하다.

두 번째 문제는 각각 플레이어로 나누어 톰슨 샘플링을 적용하였을 때 발생하는 문제다. 플레이어의 수가 증가하면서, 제약 조건에 따라 최적의 arm을 찾는 속도가 달라진다. 예를 들어 앞서 언급한 알고리즘을 곧바로 적용했을 경우, 제약 조건이 낮은 플레이어들이 빠르게 최고의 보상을 주는 경로를 찾고 이 경로를 미리 선점하게 된다. 각각의 경로는 대역폭이 유한하므로, 요구조건이 낮은 플레이어들이 미리 자리를 차지하게 되면 요구조건이 높은 플레이어들은 계속 실패를 반복한다. 결과적으로 각각 플레이어들의 수렴 속도 차이가 발생하면서, 요구조건이 높은 플레이어들이 성공할 수 있는 적절한 행동을 찾을 수 없게 된다. 이를 해결하기 위해서는 성공과 실패를 예측하는 Bernoulli 확률분포를 시스템에 맞도록 개선하거나, 성공과 실패를 산출하는 조건을 엄격하게 조절함으로서 해결할 수 있다.

3.2. 다중경로 시스템에서의 톰슨 샘플링 개선방안

플레이어 간 제약 조건 격차에 따른 수렴 문제를 해결하기 위해, 본 논문에서는 잉여 자원 발생에 대한 페널티를 주어 성공과 실패를 산출하는 조건을 조절하였다.

< Reward decision 알고리즘 >
Input: # of services S,
result capacity of services c_s,
capacity requirement of services r_s,
surplus threshold $T > 1$
1: for s=1, ...,S do
2: if $c_s > r_s$ && $c_s < T \times r_s$ then
3: $s_{s,i} = s_{s,i} + 1$
4: else
5: $f_{s,i} = f_{s,i} + 1$
6: end if
7: end for

위 알고리즘은 서비스의 선택 결과에 따라 잉여 전송량이 과하게 발생하는 경우, 전체 네트워크를 최적화하려는 목적에 반하게 되므로 페널티를 준다. 이렇게 페널티를 주게 되면, 제약 조건이 크지 않은 서비스들이 제약 조건 수준에 따라 알맞게 경로를 선택해서 앞서 이야기한 수렴 속도 차이에 따른 문제가 발생하지 않게 된다.

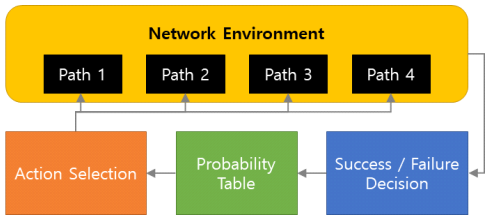

Fig. 2 Diagram of proposed algorithm

그림 2는 제안 방안의 다이어그램을 보여준다. 제안 방안에서는 prior에 따라 행동을 선택하고, 그 결과를 바탕으로 성공/실패를 결정하고 확률 분포를 업데이트한다. 여기서 기존 톰슨 샘플링과의 차이는 성공/실패를 판별하는 기준에서 차이가 발생하게 된다.

Ⅳ. 성능 분석 결과

제안하는 알고리즘의 성능 분석을 위해, MATLAB을 이용하여 시뮬레이션을 수행하였다. 시뮬레이션을 통해 측정한 메트릭은 speed satisfaction rate(SSR)로, 정의는 다음과 같다.

$$ssr = \sum_{s=1}^{S} \frac{c_s}{r_s} \tag{1}$$

위 식을 정리하면, 각각 서비스가 실질적으로 얻는 통신 속도 합과 각각 서비스의 요구 통신 속도 합의 비율로, 1에 가까울수록 각각 서비스가 요구하는 통신 속도를 만족한다고 볼 수 있다. 또한 각각 서비스가 잉여 주파수 자원을 얻더라도 요구 통신 속도를 달성한 상황을 가정하자. 이 요구 통신 속도는 사용자 트래픽을 고려하기 때문에 트래픽이 무한하지 않은 환경에서는 통신 속도는 요구 통신 속도를 넘을 수 없다. 결과적으로 SSR은 1보다 클 수 없다.

시뮬레이션을 위해 각각의 경로는 5개의 노드가 서로 이어지는 경로를 가정하였고, 각각의 노드에 Poisson 분포로 background traffic을 발생시켰다. 이 background traffic의 차이로 인해 각 경로 간 가용 대역폭 차이가 발생한다. 덧붙여 각각의 노드를 잇는 링크마다 주파수 효율(spectral efficiency) 차이를 두어 경로 별 격차가 두드러지는 환경을 가정하였다.

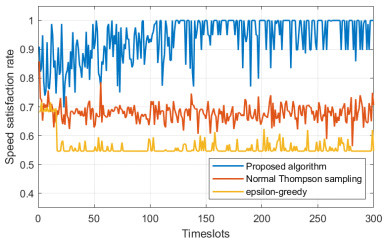

Fig. 3 Speed satisfaction rate according to timeslots

그림 3은 타임슬롯에 따른 SSR 변화를 보여준다. 성능 검증을 위해 비교한 알고리즘들은 제안 방안(proposed algorithm), 잉여 자원에 대한 페널티를 고려하지 않은 일반적인 톰슨 샘플링(normal Thompson sampling), 입실런의 확률을 제외하고는 최고의 보상을 얻을 수 있는 행동을 선택하는 입실런 그리디 방안(epsilon-greedy)을 함께 고려하였다. 제안 방안의 경우, 앞 50~100회 동안은 최적의 선택을 찾기 위해 SSR이 흔들리는 것을 보이지만, 그 뒤 안정적으로 1에 가깝게 동작을 한다. 반면, 일반적인 톰슨 샘플링과 입실런 그리디 방안의 경우 각 서비스 간 수렴속도 차이로 인해 요구조건이 낮은 서비스들이 품질이 좋은 경로를 많이 차지하게 되고, 이에 따라 요구조건이 높은 경로들이 만족도를 높이기 어렵게 된다.

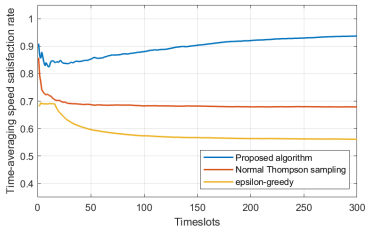

Fig. 4 Time-averaging speed satisfaction rate

이와 같은 경향은 그림 4의 시간 평균 SSR의 경우에 더 잘 볼 수 있다. 제안 방안 또한 exploration이 있기 때문에 SSR이 정확히 1에 수렴하지는 않으나, 1에 가까운 값으로 향하는 것을 볼 수 있다. 반면 나머지 두 방안의 경우 성능이 1에 도달하지 못하고 중간 어딘가에서 수렴하는 것을 확인할 수 있다.

Ⅴ. 결 론

본 논문에서는 다중경로 통신 시스템에서의 다중경로를 톰슨 샘플링을 이용하여 선택하는 방안을 제안하였다. 톰슨 샘플링을 그대로 적용하기에는 다수의 플레이어가 참가하고, 플레이어 간 제약조건에 따른 수렴 속도의 차이가 발생하여 이를 해결하기 위한 성공 보상 조건을 수정하였다. 이를 통해 플레이어 간 수렴 속도 차이를 보정할 수 있었고, 시뮬레이션을 통해 제안 방안이 주어진 다중 경로를 모두 활용하며 서비스 별 요구 조건을 달성하는 것을 확인하였다. 특히 수렴 속도 차이를 보정하는 부분은 다중 플레이어가 존재하는 환경이라면 어느 시스템에도 적용 가능한 알고리즘이 될 것이다. 해당 아이디어에 대한 수학적인 분석으로 연구를 확장시키고 실제 환경에서의 검증을 통해, 더욱 다양한 시스템에 범용적으로 적용하는 것을 기대한다.

REFERENCES

[1] M. Ali, S. Qaisar, M. Naeem, W. Ejaz, and N. Kvedaraite, "LTE-U WiFi HetNets: Enabling Spectrum Sharing for 5G/Beyond 5G Systems," *IEEE Internet of Things Magazine*, vol. 3, no. 4, pp. 60-65, Dec. 2020.

[2] Y. Xing, J. Han, K. Xue, J. Liu, M. Pan, and P. Hong, "MPTCP Meets Big Data: Customizing Transmission Strategy for Various Data Flows," *IEEE Network*, vol. 34, no. 4, pp. 35-41, Jul./Aug. 2020.

[3] B. C. Chung and H. Park, "Path selection algorithm for multi-path system based on deep Q learning," *Journal of the Korea Institute of Information and Communication Engineering*, vol. 25, no. 1, pp. 50-55, Jan. 2021.

[4] M. S. Kim, J. Y. Lee, and B. C. Kim, "Design of MPTCP congestion control based on BW measurement for wireless networks," *Journal of the Korea Institute of Information and Communication Engineering*, vol. 21, no. 6, pp. 1127-1136, Jun. 2017.

[5] R. S. Sutton and A. G. Barto, *Reinforcement Learning: An Introduction*, Cambridge, MA: The MIT Press, Mar. 1998.

[6] O. Chapelle and L. Li, "An empirical evaluation of Thompson sampling," in. *Proc. of Advances in Neural Information Processing Systems*, pp. 2249-2257, 2011.

[7] J. Komiyama, J. Honda, and H. Nakagawa, "Optimal Regret Analysis of Thompson Sampling in Stochastic Multi-armed Bandit Problem with Multiple Plays," in. *Proc. of the 32nd International Conference on Machine Learning*, pp. 1152-1161, 2015.