

Multi-Armed Jamming Bandit선택 기법의 성능 비교 분석

이 명 로*, 신 요 안°

Comparative Performance Analysis of Multi-Armed Jamming Bandit Selection Schemes

Myeong-Ro Lee*, Yoan Shin°

요 약

본 논문에서는 적군 통신 시스템에 따라 적응적으로 재밍을 수행하는 Multi-Armed Jamming Bandit 기법에서 자동 변조 분류를 적용하여 최적 재밍 전략을 도출할 때, 고려되는 경우의 수를 줄여 최적 재밍 전략 도출 속도가 향상되는 방법을 제시하고 Arm Choosing 기법 별 탐색과 이용의 절충에 대해 분석한다.

Key Words : Adaptive Jamming, Multi-Armed Jamming Bandit, Arm Choosing, Exploration/ Exploitation Trade-off

ABSTRACT

In this paper, we present how automatic modulation classification is applied to the multi-armed jamming bandit that adaptively selects the jamming strategy according to enemy communication systems, in order to reduce the number of cases considered to derive optimal jamming strategies. We also analyze the exploration/exploitation trade-offs of the arm choosing techniques.

I. 서 론

전자전 (Electronic Warfare) 중 적군 통신 정보를 알기 어려운 상황에서, 적응적으로 적군 통신 전략에 최적인 재밍 전략을 도출하는 Multi-Armed Jamming Bandit 기법이 최근 제시되었다^[1]. 또한, Jamming Bandit 기법에서 최적 재밍 전략을 도출하는 속도를 향상시키기 위해, 재밍 전략 집합의 중요 요소 중 하나인 적군 통신 신호 변조 방식을 자동 변조 분류 (Automatic Modulation Classification; AMC)를 통해 언어 전략 탐색 상태의 수를 줄이는 방법이 제시되었다^[2]. 본 논문에서는 AMC 기반 Jamming Bandit을 통해 내부 Arm (즉, 재밍 전략) 들을 탐색하여 최적 재밍 전략을 도출할 때, Arm Choosing 기법의 탐색과 이용의 절충 (Exploration/Exploitation Trade-off)에 대해 분석한다.

II. 본 론

2.1 AMC 기반 Multi-Armed Jamming Bandit 구성

AMC 기반 Multi-Armed Jamming Bandit은 적군 통신 전략을 모르는 상황에서, 구성된 재밍 전략 가운데 적군 수신기의 심벌오율 (Symbol Error Rate; SER) 등을 성능 지표로 하여 적응적으로 최적 재밍 전략 (즉, 최적의 Arm)을 도출하는 방법이다^[2].

기존에 제시된 Jamming Bandit 방법은 식 (1)과 (2)와 같은 재밍 전략 (즉, Arm)으로 구성된다^[1].

$$Arm = \{J, S\}. \tag{1}$$

$$S = \{P_j, \rho\}. \tag{2}$$

윗 식 (1)에서 J 는 적군에 대한 재밍 신호의 변조 방식, 식 (2)의 P_j 는 평균 재밍 전력 (dB), ρ 는 Duty Cycle 을 의미한다. 특히, 딥러닝 기반 AMC^[3] 등을 통해 예측된 적군 신호 변조 정보에 대응하여, 참고문헌 [4]에서 제시한 적군 신호 변조에 따른 최적의 아군 재밍 신호 변조 방식을 선택해 J 로 사용한다고 가정한다^[2].

* 본 연구는 방위사업청과 국방과학연구소가 지원하는 미래전투체계 네트워크기술 특화연구센터 사업의 일환으로 수행되었습니다 (UD190033ED).

• First Author : (OCRCID:0000-0002-8982-8984) School of Electronic Engineering, Soongsil University, mrlee1102@soongsil.ac.kr, 학생(석사과정), 학생회원

° Corresponding Author : (OCRCID:0000-0002-4722-6387) School of Electronic Engineering, Soongsil University, yashin@ssu.ac.kr, 정교수, 종신회원

논문번호 : 202208-193-A-LU, Received August 31, 2022; Revised September 4, 2022; Accepted September 4, 2022

2.2 Jamming Bandit의 Arm Choosing 기법

Arm Choosing 기법은 Multi-Armed Jamming Bandit 내부에 구성된 재밍 전략 중 적군 수신기에 재밍 영향이 가장 높은 (즉, 적군 통신 성능을 가장 악화시키는) 최적 재밍 전략을 탐색하는 방법이다. 본 논문에서는 Arm Choosing 기법으로서 Decayed ϵ -Greedy 기법과 신뢰 상한 (Upper Confidence Bound; UCB) 기법을 고려한다.

먼저, ϵ -Greedy 기법은 식 (3)과 같이 표현될 수 있다.

$$A_t = \begin{cases} \arg \max_a Q_t(a), & \epsilon < \delta \\ rand(a), & \epsilon \geq \delta \end{cases} \quad (3)$$

여기서, A_t 는 시간 t 번째 단계에서 수행할 재밍 전략, $Q_t(a)$ 는 시간 t 번째 단계에서 재밍 전략 a 를 수행할 시 얻어지는 보상 (즉, 적군 수신기 SER), ϵ 은 재밍 전략의 탐색 영향을 조절하는 매개변수, δ 는 균등 분포를 따르는 값이며 임계값을 의미한다. 따라서, ϵ 값이 임계값 δ 보다 작을 때는 보상 값이 큰 Greedy한 재밍 전략을 선택하고, 그렇지 않을 때는 임의의 재밍 전략을 선택한다.

ϵ -Greedy 기법에 감쇄 지수 (Decay Factor)를 반영한 Decayed ϵ -Greedy 기법의 경우, 식 (3)의 ϵ 값을 초기에 큰 값으로 설정 후 매 시간 스텝마다 ϵ 값에 감쇄 지수를 반영하여 값을 점점 줄여 간다. 따라서, 초기 시간 스텝에서는 다양한 재밍 전략을 탐색한 후, 시간이 지날수록 탐색 빈도를 줄여 탐색된 전략 중 최적 재밍 전략을 선택하는 빈도를 늘리는 기법이다⁵⁾.

본 논문에서 고려하는 다른 형태의 Arm Choosing 기법인 UCB 기법은 식 (4)와 같이 표현된다⁵⁾.

$$A_t = \arg \max_a \left[Q_t(a) + c \sqrt{\frac{\ln t}{N_t(a)}} \right] \quad (4)$$

여기서 $N_t(a)$ 는 시간 t 번째 단계에서 재밍 전략 a 가 선택된 횟수, c 는 신뢰구간의 영향을 조절하는 가중치이다. 이 방식은 식 (3)의 ϵ -Greedy 기법과 유사하나, 루트 항 내에 신뢰구간을 더해 많이 선택된 재밍 전략 일수록 신뢰구간이 작아지게 된다. 따라서, 재밍 전략들 중 UCB 과정의 신뢰구간 값이 매우 작음에도 다른 재밍 전략보다 얻어지는 보상이 상대적으로 클 경

우, 주어진 재밍 전략 중 최적의 재밍 전략이라고 할 수 있다.

III. 모의실험 결과 및 결론

참고문헌 [6]에서는 기본적인 환경에서 Arm Choosing 기법들의 성능을 일부 평가하였으며, 본 논문에서는 이와 동일한 모의실험 설정에서 Arm Choosing 기법의 매개변수를 표 1과 같이 조정하여 성능을 비교 분석한다. 특히, 덤러닝 기반 AMC³⁾를 통해 추정된 적군 변조 방식이 BPSK라 가정하고, 참고문헌 [5]에서 제시된 바와 같이 이에 대한 최적의 재밍 신호 변조 방식인 BPSK를 사용한다고 가정한다. 또한, 다양한 적군의 신호대잡음비 (Signal-to-Noise Ratio; SNR) 환경에서, 각 기법 별로 도출되는 최적 재밍 전략 결과를 도출한다.

그림 1은 매 시간 단계에서 UCB 기법과 Decayed ϵ -Greedy 기법을 통해 실시간으로 도출되는 보상을 도시하며, 적절한 매개변수 조정을 통해 최대 보상의 도출 횟수를 증가시킬 수 있음을 확인할 수 있다. 또한, 이 그림에서 $\epsilon=0.3$ 과 $\epsilon=0.4$ 인 경우를 비교하면, 탐색 영향 ϵ 의 차이에 따라 보상의 최대값이 도출되는 (즉, 최적 재밍 전략이 탐색되는) 시간적 차이가 있다. 따라서, Arm Choosing 기법의 탐색 영향을 적절히 조절한다면 탐색과 이용의 절충을 완화하여, 동일한 모의실험 횟수 동안 빠른 최적 재밍 탐색 및 최적 재밍 수행 횟수의 증가가 가능함을 알 수 있다 (즉, 성능이 개선되었음을 의미).

표 2는 다양한 적군 SNR 상황에서 UCB 기법과 Decayed ϵ -Greedy 기법을 통해 가장 많이 선택된 재밍 전략을 정리하였으며, 실제 특정 적군 SNR에 대한 최적 (즉, 적군 수신기 SER을 최대로 하는) 재밍 전략도 같이 표시하였다. UCB 기법의 경우, 다양한 적군 SNR 환경에서도 재밍 전략에 대해 충분한 탐색을 수

표 1. 모의실험을 위한 매개변수 설정
Table 1. Parameter settings for simulations

Jamming Power	[10 dB, 15 dB]
Enemy SNR	20 dB
Jammer Modulation Scheme	BPSK
Enemy Modulation Scheme	BPSK
Number of simulations	100
UCB	$c = 0.1$
Decayed ϵ -Greedy	$\epsilon=0.3, 0.4$
Decay Factor	0.999

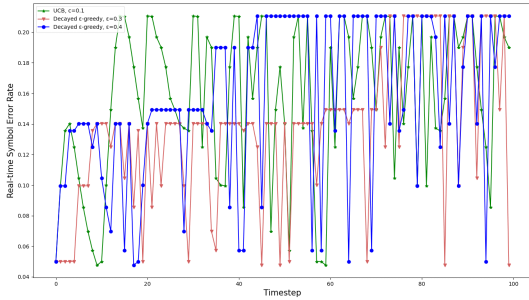


그림 1. Arm Choosing 기법의 실시간 보상 결과
Fig. 1. Real-time compensation for arm choosing schemes

표 2. 적군 SNR에 따른 가장 많이 선택된 재밍 전략의 결과
Table 2. Results of the most frequently-selected jamming strategy according to the enemy SNR

Enemy SNR	UCB ($\epsilon=0.1$)	Decayed ϵ -Greedy ($\epsilon=0.4$)	Decayed ϵ -Greedy ($\epsilon=0.3$)	Optimal Jamming Strategy
5 dB	Arm20	Arm10	Arm10	Arm20
10 dB	Arm20	Arm17	Arm8	Arm20
15 dB	Arm17	Arm17	Arm5	Arm17
20 dB	Arm15	Arm15	Arm4	Arm15
25 dB	Arm14	Arm14	Arm3	Arm14

행하여 최적 재밍이 수행 가능하다. 또한, 탐색된 재밍 전략 중 최적 재밍 전략 결과와 동일한 재밍 전략이 가장 많이 선택되었다. 반면, Decayed ϵ -Greedy 기법은 UCB 기법과 달리 일부 SNR 상황에서 최적 재밍 전략이 탐색되지 않았고, 가장 많이 선택된 재밍 전략도 최적 재밍 전략과 동일하지 않으므로 최적 재밍이 수행 불가능함을 알 수 있다. 하지만, Decayed ϵ -Greedy 기법 중 $\epsilon=0.4$ 일 때를 보면, 적군 SNR이 각각 15 dB, 20 dB, 25 dB일 경우는 재밍 전략에 대해 충분한 탐색이 되어 최적 재밍이 수행 가능한 결과를 볼 수 있다. 또한, Decayed ϵ -Greedy 기법은 UCB 기법보다 최적 재밍 전략이 선택된 횟수가 많았다. 이는, UCB 기법은 최적 재밍 전략이 탐색 되어도 시간이 커지면 지속적으로 탐색하여 최적이지 않은 재밍 전략을 수행할 수도 있기 때문이다. 따라서, 재밍 전략 탐색을 충분히 수행했다고 가정했을 때, Decayed ϵ -Greedy 기법은 동일한 시간 관점에서 UCB 기법보다 최적 재밍 수행 횟수가 많으므로 재밍 성능이 더 우수하다고 할 수 있다.

References

- [1] S. Amuru, C. Tekin, M. v. der Schaar, and R. M. Buehrer, "Jamming bandits—A novel learning method for optimal jamming," *IEEE Trans. Wireless Commun.*, vol. 15, no. 4, pp. 2792-2808, Apr. 2016. (<https://doi.org/10.1109/TWC.2015.2510643>)
- [2] M. R. Lee and Y. Shin, "Jamming bandit performance analysis based on automatic modulation classification," in *Proc. 2022 Summer Conf. Electrical & Electronic Soc. of Korea*, Jeju, Korea, Aug. 2022.
- [3] S. Lee, K.-Y. Kim, and Y. Shin, "Effective feature selection method for deep learning-based automatic modulation classification scheme using higher order statistics," *Applied Sciences*, vol. 10, pp. 1-14, Jan. 2020. (<https://doi.org/10.3390/app10020588>)
- [4] S. Amuru and R. M. Buehrer, "Optimal jamming against digital modulation," *IEEE Trans. Inform. Forensics & Secur.*, vol. 10, no. 10, pp. 2212-2224, Oct. 2015. (<https://doi.org/10.1109/TIFS.2015.2451081>)
- [5] R. S. Sutton and A. G. Barto, *Reinforcement Learning: An Introduction*, The MIT Press, 2018.
- [6] M. R. Lee and Y. Shin, "Analysis of the performance of the arm choosing method of the AMC-based jamming bandit," in *Proc. 32nd Joint Conf. Signal Proc.*, Seoul, Korea, Sep. 2022.