

고차 큐물런트를 이용한 변조 방식 이상치 탐지

*최윤철, *장민규, *윤동원[†]
 *한양대학교 융합전자공학과
 *dwyoona@hanyang.ac.kr[†]

Anomaly Detection of Modulation Scheme Using Higher-order Cumulants

*Yooncheol Choi, *Mingyu Jang, *Dongweon Yoon[†]
 *Hanyang University

요 약

비협력 상황에서 수신된 신호로부터 정보를 복원하기 위해 다양한 통신 제원의 추정이 필요하다. 이에 신호의 변조 방식을 추정하는 자동 변조 분류(automatic modulation classification, AMC) 연구가 활발히 진행되고 있으며 머신러닝을 활용한 AMC가 연구되고 있다. 그러나 송신단에서 기존에 알려지지 않은 변조 방식을 사용하는 경우 머신러닝을 이용한 AMC의 성능이 급감하는 문제가 발생한다. 따라서 비지도 학습을 통한 변조 방식의 이상치 탐지 연구가 필요하다. 본 논문에서는 수신 신호의 고차 큐물런트를 입력으로 하는 isolation forest 기반 이상치 탐지 기법을 제안하고 모의실험을 통해 성능을 분석한다.

I. 서 론

협력 통신 상황에서 수신단은 송신단에서 사용한 통신 제원 등 신호 복원에 필요한 정보를 가지고 있다. 그러나 비협력 통신 상황에서는 신호 복원을 위한 정보가 주어지지 않기 때문에 통신 제원의 추정이 필요하다. 특히 수신 신호로부터 심볼을 추출하기 위해선 변조 방식의 추정이 반드시 이루어져야 하는데 이를 위해 수신 신호만을 이용하여 변조 방식을 식별하는 자동 변조 분류(automatic modulation classification, AMC)가 활발히 연구되고 있으며 머신러닝을 활용한 AMC가 연구되고 있다.

송신단에서 기존에 알려지지 않은 변조 방식을 통해 신호를 생성하게 되면 머신러닝을 이용한 AMC 성능이 감소하게 되는 문제가 발생한다. 따라서 이를 식별할 수 있는 비지도 학습 기반 이상치 탐지 기법의 연구가 필요하다.

본 논문에서는 단일 반송파 신호의 고차 큐물런트를 입력으로 하는 isolation forest 기반 이상치 탐지 기법을 제안하고 모의실험을 통해 성능을 분석한다.

II. 고차 큐물런트 특징 인자 추출

본 논문에서는 가산성 백색 가우시안 잡음(additive white Gaussian noise, AWGN) 채널을 통과한 신호로부터 심볼을 수신하였다고 가정한다. 이때 수신 심볼은 아래와 같이 표현할 수 있다.

$$r[k] = s[k] + n[k] \quad (1)$$

여기서 $r[k]$ 는 수신 심볼, $s[k]$ 는 송신 심볼, $n[k]$ 는 AWGN을 나타낸다.

송신단에서 선택한 변조 방식에 따라 $r[k]$ 는 다양한 정상도와 통계적 특징을 가지게 된다. 이러한 특성을 잘 표현할 수 있는 특징값으로 고차 큐물런트가 있다[1].

$$C_{20} = E[r^2[k]] \quad (2)$$

$$C_{40} = \text{cum}[r[k], r[k], r[k], r[k]] \quad (3)$$

$$C_{42} = \text{cum}[r[k], r[k], r^*[k], r^*[k]] \quad (4)$$

$$C_{40I} = \text{cum}[r_I[k], r_I[k], r_I[k], r_I[k]] \quad (5)$$

여기서 $E[\bullet]$ 는 기댓값 연산자, $\text{cum}[\bullet]$ 는 큐물런트 연산자, $r^*[k]$ 는 $r[k]$ 의 켤레 복소수, $r_I[k]$ 는 $r[k]$ 의 inphase 성분이다. 평균이 0인 랜덤 변수 a, b, c, d 에 대해 4차 큐물런트는 아래와 같이 계산할 수 있다[2].

$$\text{cum}[a, b, c, d] = E[abcd] - E[ab]E[cd] - E[ac]E[bd] - E[ad]E[bc] \quad (6)$$

표1은 변조 방식 BPSK, QPSK, 8PSK, 16QAM, 64QAM에 대한 고차 큐물런트의 이론값을 나타낸다.

표 1. 변조 방식 별 고차 큐물런트의 이론값

	$ C_{20} $	$ C_{40} $	$ C_{42} $	$ C_{40I} $
BPSK	1	2	2	2
QPSK	0	1	1	0.5
8PSK	0	0	1	0.375
16QAM	0	0.68	0.68	0.34
64QAM	0	0.619	0.619	0.3095

충분히 많은 심볼을 수신하여 $r[k]$ 의 큐물런트를 계산하면 표1에 나타난 이론값을 중심으로 분포하게 된다.

III. Isolation forest 기반 이상치 탐지

송신단에서 채택한 변조 방식이 수신단에서 상정하지 않은 것일 수 있으므로 비지도 학습을 통한 이상치 탐지 기법을 사용하는 것이 바람직하다. 본 논문에서는 비지도 학습 기반 이상치 탐지 기법으로써 isolation forest를 이용한다[3]~[4]. Isolation forest는 tree 기반 이상치 탐지 알고리즘으로 정상 데이터의 수가 이상치 데이터보다 많다는 것과 이상치 데이터는 정상 데이터들과 상이한 값을 가지고 있다는 것을 전제로 한다. 주어진 데이터를 랜덤하게 분할하여 모든 데이터를 고립시키는 것으로 iTree를 생성할 수 있다. 판정하고자 하는 데이터가 정상 데이터라면 iTree의 단말 노드와 근접한 노드에서 고립될 확률이 높고 이상치 데이터라면 뿌리 노드와 근접한 노드에서 고립될 확률이 높다. 즉 정상 데이터라면 고립시키기 위해 필요한 분할 시행의 횟수가 많고 이상치 데이터라면 적을 것으로 기대할 수 있다. 따라서 iTree를 여러 개 생성하여 각 데이터가 고립되기 위해 필요한 분할 횟수의 평균을 이상치 점수로 사용할 수 있다.

IV. 모의 실험 및 성능 분석

본 논문에서는 수신 신호의 고차 큐물런트를 입력으로 하는 isolation forest 기반 이상치 탐지 기법을 제안한다. 제안한 이상치 탐지 기법의 성능 분석을 위해 BPSK, QPSK, 8PSK 신호를 정상 데이터로, 16QAM, 64QAM 신호를 이상치 데이터로 정의하였다. 신호의 E_b/N_0 는 -5dB 부터 10dB 까지 1dB 간격을 고려하였으며 정상 데이터로 정의한 변조방식에서 각 E_b/N_0 마다 5000개의 심볼을 포함하는 샘플 1000개를 생성하였다. 생성된 샘플마다 $|C_{20}|$, $|C_{40}|$, $|C_{42}|$, $|C_{40}|$ 를 원소로 가지는 4차원 벡터를 계산하였다. 또한 모델의 성능 분석을 위해 각 변조 방식의 E_b/N_0 마다 200개의 벡터를 위와 같은 방식으로 생성하였다. 앞서 정상 데이터로 생성한 48000개의 벡터에 이상치 여부를 판정하고자 하는 데이터를 1개씩 추가하여 isolation forest 모델의 학습을 진행하였다.

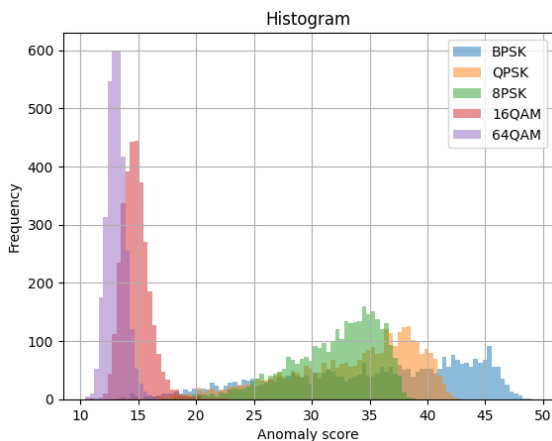


그림 1. 변조 방식 별 이상치 점수 분포

그림 1에 각 변조 방식에 따른 이상치 점수의 분포를 나타내었다. 정상 데이터로 정의한 BPSK, QPSK, 8PSK의 경우 이상치 데이터로 정의한 16QAM, 64QAM에 비해 이상치 점수가 높은 것을 확인할 수 있다. 따라서 특정한 임계값을 선택하여 입력된 데이터의 이상치 점수가 임계값보다 높다면 정상 데이터로, 낮다면 이상치 데이터로 분류할 수 있다.

이러한 이진 분류의 경우 임계값에 따라 성능이 변화

한다. 따라서 모델 성능 분석을 위해 임계값에 영향을 받지 않는 수신자 조작 특성 그래프의 아래 면적(area under receiver operating characteristic curve, AUROC)를 성능 지표로 도입하였다. 이때 AUROC의 값이 1에 가까울수록 성능이 좋은 모델이라고 평가할 수 있다.

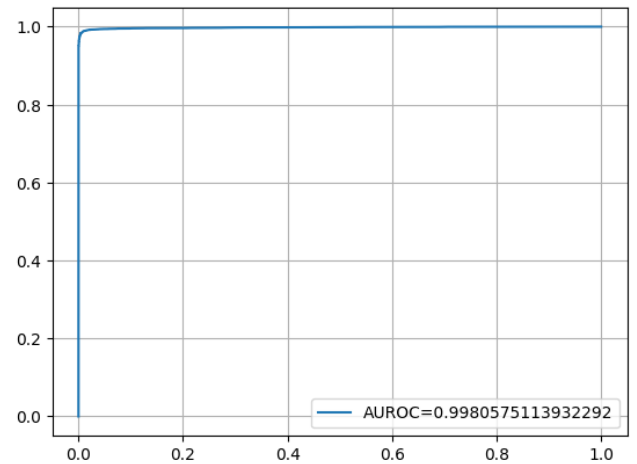


그림 2. ROC 그래프

그림 2에 모델의 ROC 그래프와 AUROC 값을 함께 도시하였다. 그래프의 형태가 좌측 상단에 근접하고 AUROC가 1에 가까운 것으로 제안한 모델의 성능이 매우 우수한 것을 확인할 수 있다.

V. 결론

본 논문에서는 고차 큐물런트를 입력으로 하는 isolation forest 기반 이상치 탐지 기법을 제안하고 모의실험을 통해 제안한 기법의 이상치 탐지 성능을 확인하였다. 추후 연구로는 isolation forest의 단점을 일부 극복한 rotated isolation forest와 extended isolation forest를 적용하는 연구가 진행되어야 할 것이다.

ACKNOWLEDGMENT

이 논문은 2021년도 정부(과학기술정보통신부)의 재원으로 정보통신기획평가원의 지원을 받아 수행된 연구임 (2019-0-00964, 스펙트럼 챌린지를 통한 기존 무선국 보호 및 주파수 공유기술 개발)

참고 문헌

- [1] J. Lee, S. Ahn, J. Choi, and D. Yoon, "Phase mapping recognition and modulation classification algorithm using new higher-order cumulants," *Journal of Korean Institute Inf. Technol.*, vol. 15, no. 2, pp. 153-160, Feb. 2017
- [2] A. Swami, and B. M. Sadler, "Hierarchical digital modulation classification using cumulants," *IEEE Trans. Commun.*, vol. 48, no. 3, pp. 416-429, Mar. 2000
- [3] T. Liu, K. M. Ting, and Z. Zhou, "Isolation forest," *IEEE Int. Conf. Data Mining*, pp. 413-422, Dec. 2008
- [4] S. Hariri, M. C. Kind, and R. J. Brunner, "Extended isolation forest," *IEEE Trans. Knowledge and Data Engineering*, vol. 33, no. 4, pp. 1479-1489, Oct. 2019