

BiLSTM-Attention과 LRP를 활용한 시계열 LPI 신호 분류 및 중요도 분석

박기완*, 남해운^o

Time-Series LPI Signal Classification and Relevance Analysis Using BiLSTM-Attention with LRP

Kiwan Park*, Haewoon Nam^o

요약

본 연구는 BiLSTM-Attention 모델과 Layer-wise Relevance Propagation (LRP)를 적용하여 저피탐지성(LPI) 신호의 분류와 중요도 분석을 수행하였다. 본 연구의 주요 목적은 시계열 데이터로 학습된 모델의 예측 근거를 LRP를 통해 해석하고, 이를 통해 모델의 예측 과정에서 의미 있는 입력 특성을 효과적으로 식별하는 것이다. LRP를 활용한 해석 결과, 모델이 시계열 도메인에서 학습된 상태임에도 불구하고, 예측 근거가 FFT(고속 푸리에 변환)로 변환된 주파수 도메인에서도 높은 일관성을 유지하며 주요 주파수 성분을 정확하게 탐지할 수 있음을 확인하였다. 다양한 SNR(신호 대 잡음비) 환경에서의 실험을 통해 본 모델은 신뢰도 높은 분류 성능을 유지할 뿐만 아니라, LRP 기반 해석을 통한 중요 지점 탐지 성능 역시 안정적임을 확인하였다.

Key Words : LPI Signal Classification, Layer wise Relevance Propagation, BiLSTM, Self Attention, Electronic Warfare (EW)

ABSTRACT

This study applies a BiLSTM-Attention model and Layer-wise Relevance Propagation (LRP) to classify and analyze the importance of low probability of intercept (LPI) signals. The goal is to interpret the predictions of a time-series trained model using LRP and effectively identify meaningful input features. The analysis shows that the model maintains high consistency in its prediction rationale even in the frequency domain, transformed through Fast Fourier Transform (FFT). Experiments across various Signal-to-Noise Ratio (SNR) conditions confirm that the model delivers reliable classification performance while ensuring stable detection of key features through LRP-based interpretation.

I. 서론

저피탐지성(LPI) 신호는 전자전(EW) 환경에서 탐지가 어렵도록 설계된 신호로, 주로 군사 레이더 시스템이나 통신 시스템에서 사용된다. 이러한 신호는 기존의 탐지 기법으로는 포착이 어려워 신호 분류와 분석에 있어 도전적인 문제를 제기한다.^[1] 특히, LPI 신호의 다양한 변종과 SNR(신호 대 잡음비) 조건에 따라 정확한 분류를 수행하는 것은 신호 처리 분야에서 중요한 연구 주제이다. 최근 딥러닝 기반 모델은 시계열 데이터의 분류 문제에서 좋은 성능을 보이며, LPI 신호 분류에도 효과적으로 적용되고 있다. 그러나 딥러닝 모델의 예측 과정이 블랙박스 형태로 이루어지기 때문에, 사용자는 모델이 어떤 근거로 특정 예측을 수행했는지 알기 어렵다. 이러한 문제를 해결하기 위해 Layer-wise Relevance Propagation(LRP)는 입력 데이터의 각 부분이 예측에 얼마나 기여했는지를 가시화해 모델의 해석력을 높이는 데 효과적이다.^[2] 본 연구에서는 BiLSTM-Attention 모델에 LRP 기법을 적용하여 LPI 신호 분류의 해석 가능성을 높이하고자 한다. 모델이 시계열 도메인에서 학습된 상태에서도 FFT(고속 푸리에 변

* 본 기술은 과학기술정보통신부 및 정보통신기획평가원의 대학ICT연구센터사업의 연구결과로 개발한 결과물입니다(IITP-2024-RS-2023-00258639).

• First Author : (ORCID:0009-0007-4557-391X) Hanyang University Department of Information Security, rldhks6899@hanyang.ac.kr, 학생(석사), 학생회원

o Corresponding Author : (ORCID:0000-0001-9847-7023) Hayang University Department of Electronic Engineering, hnam@hanyang.ac.kr, 정교수, 정회원

논문번호 : 202410-240-A-LU, Received October 14, 2024; Revised November 2, 2024; Accepted November 2, 2024

환)로 변환된 주파수 도메인에서 LRP를 통한 중요 지점 탐지가 잘 이루어지는지 검증한다. 이를 위해 본 연구는 12가지의 LPI 클래스에 대해 다양한 SNR 환경에서 실험을 수행하여, 모델의 분류 성능과 LRP 기반 해석 성능이 안정적으로 유지됨을 확인한다.

II. BiLSTM-Attention with LRP

본 연구에서는 BiLSTM-Attention 모델에 LRP를 적용하여, 예측 과정에서 입력 데이터가 어떻게 기여했는지를 해석한다. LRP는 모델의 출력 결과를 입력 데이터로 거슬러 올라가며(Relevance 역전파) 각 입력 시점이 모델 예측에 얼마나 기여했는지를 정량적으로 측정하는 기법이다. BiLSTM-Attention 모델은 LSTM과 Self-Attention 메커니즘을 결합하여 시계열 신호의 시간적 의존성을 효과적으로 학습한다. BiLSTM은 신호를 순방향과 역방향 모두에서 처리해 각 시점의 hidden state를 생성하며, Self-Attention은 중요한 시간 스텝에 더 큰 가중치를 부여해 예측 성능을 향상시킨다. LRP를 모델에 적용하는 과정은 먼저 Fully Connected(FC) 계층에서의 relevance 역전파로 시작한다. 모델이 예측한 타겟 클래스에 해당하는 중요도(relevance)만 남기고, 나머지 클래스에 대한 중요도는 0으로 설정한다. 이렇게 초기화된 relevance는 FC 계층의 가중치를 따라 이전 계층으로 전파된다. 구체적으로, FC 계층의 가중치 W_{fc} 와 출력 클래스에 대한 초기 relevance R_y 사용해 Context Vector에 대한 relevance R_C 를 다음과 같이 계산한다.

$$R_C = \sum_j \frac{cls}{\epsilon + |W_{fc_j}|} \cdot C \cdot W_{fc_j} \cdot R_{y_j} \quad (1)$$

Context Vector의 relevance는 다음으로 Attention 가중치로 전파된다. Self-Attention 계층에서 각 시간 스텝에 부여된 가중치 α_t 는 그 시점의 hidden state가 예측에 얼마나 기여했는지를 반영한다. LRP는 Context Vector에서 relevance를 Attention 가중치로 전파하여 각 시간 스텝 t 에 대한 relevance R_{c_t} 를 구할 수 있다. 이는 다시 LSTM 계층의 hidden state로 전달된다. LSTM 계층에서는 양방향 hidden state에 대해 relevance를 나누어 계산하며, 각 시점의 relevance가 이전 시점으로 전파된다. 특히, LSTM의 Forget·Input·Output Gate 연산을 통해 각 시점의 hidden state와 cell state로 relevance가 분배된다. 이때 각 시점의 두 state

표 1. BiLSTM-Attention 모델의 분류 성능
Table 1. Classification Performance of the BiLSTM-Attention Model

SNR (dB)	Accuracy (%)
0dB	98.36
-2dB	96.92
-4dB	95.53
-6dB	91.64
-8dB	88.17
-10dB	81.75
-12dB	74.50
-14dB	63.58
-16dB	47.94

에 대한 relevance는 다음 수식을 통해 계산된다.

$$R_{h_t} = \alpha_t \cdot R_c \quad (2)$$

$$R_{c_t} = f_t \cdot R_{c_{t-1}} + R_{h_t} \cdot o_t \cdot (1 - \tanh^2(c_{l_t})) \quad (3)$$

cell state에서 계산된 relevance는 각 입력 시점으로 매핑되어 해석되어 시점에서 입력이 예측 결과에 미친 기여도를 파악할 수 있다. 입력 시점에 대한 relevance는 다음과 같은 수식을 통해 계산된다:

$$R_{x_t} = W_{ih}^o(R_{c_t} \cdot o_t) + W_{ih}^c(R_{c_t} \cdot \tanh(c_{l_t})) \quad (4)$$

$$R_{x_t} = R_{x_t}^{fw} + R_{x_t}^{bw} \quad (5)$$

여기서 W_{ih}^o 와 W_{ih}^c 은 각각 Output gate와 Cell state의 입력 가중치를 의미한다.

III. 실험

본 연구는 제안된 모델과 LRP 기법을 다양한 SNR 환경에서 평가하였다. 12가지 LPI 신호 클래스를 대상으로 시간 시계열 데이터의 분류 성능을 측정하였으며, 각 SNR별 1000개씩 총 108,000개의 샘플을 사용하였다. 신호 길이와 진폭이 무작위로 설정되었기 때문에, 모든 신호에 zero-padding을 적용해 동일한 길이로 맞추었다. 이는 모델이 가변적인 입력과 잡음 환경에서도 유효한 정보만을 처리하도록 설계된 것이다.

3.1 Classification Performance

표 1에 따르면, 모델은 SNR이 높은 환경에서는 98% 이상의 높은 정확도를 기록하였고, SNR이 낮아짐에 따라 성능이 점진적으로 감소하였으나 -8 dB 이하의 저잡음 환경에서도 일정 수준의 성능을 유지하며, -16 dB 환경에서 47.94%의 정확도를 기록하여 심한 잡음 속에서도 일부 신호를 분류할 수 있음을 확인하였다.

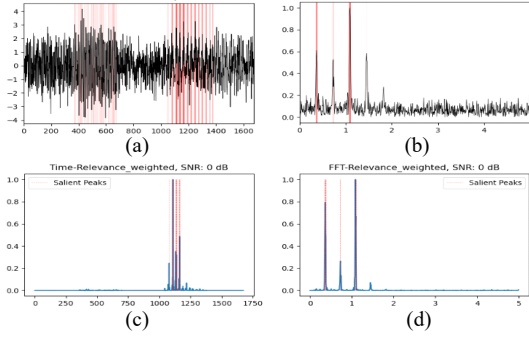


그림 1. Costas 신호에 대한 도메인 별 R_w
 Fig. 1. Domain-wise R_w analysis for Costas signal (a) Original Signal with R_x (b) FFT Signal with R_w^{FFT} .

3.2 Peak Alignment Score (PAS)

본 연구에서는 주파수 도메인에서의 주요 성분 탐지 성능을 정량적으로 평가하기 위해 Peak Alignment Score (PAS) 지표를 제안한다. 그림 1은 Costas 신호에 대한 시간 및 주파수 도메인에서의 분석 결과를 나타낸다. 그림 1(a)는 원본 시계열 데이터에 대한 relevance 값이, 그림 1(b)는 FFT 스펙트럼에 대한 relevance 값을 보여준다. FFT relevance 값은 먼저, 모델이 도출한 relevance 값을 FFT로 변환 후, 주파수 도메인에서 가중 relevance를 계산한다. 각 주파수 성분의 가중 relevance는 다음과 같은 수식으로 정의되며, :

$$R_w = A_{R_{FFT}}(f) \cdot (A_{X_{FFT}}(f))^2 \quad (6)$$

이때 데이터 진폭의 제공은 해당 주파수 성분의 강도를 강조하여, 더 중요한 성분에 더 높은 가중치를 부여하기 위해 사용된다. PAS는 FFT를 통해 변환된 relevance 값에서 탐지된 salient peaks의 기여 비율로 정의된다. 모델이 도출한 relevance 값을 FFT로 변환하고, salient peaks는 수식 7과 같이 \mathcal{X} 데이터의 평균과 표준편차로 계산되는 동적 임계값을 초과하는 성분으로 탐지된다.

$$\lambda(x) = \mu(x) + k \cdot \sigma(x) \quad (7)$$

각 도메인에서 탐지된 salient peaks는 그림(c)와 그림(d)과 같다. salient peaks의 relevance 합을 전체 relevance 합으로 정규화하여 PAS를 계산한다.

$$PAS = \frac{\sum_{f \in peaks} R_w^{FFT}(f)}{\sum R^{FFT}(f)} \quad (8)$$

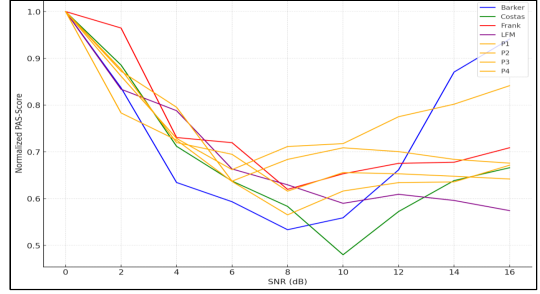


그림 2. 피크값 정렬 지표
 Fig. 2. Peak Alignment Score

PAS는 주파수 도메인에서의 주요 성분 탐지 성능을 정량화하며, 모델이 잡음 환경에서도 중요한 성분을 포착할 수 있는 능력을 평가한다. 그림 2는 SNR 수준별 PAS 변화를 보여주며, 0dB에서 -8dB까지 PAS가 점차 감소하는 것은 잡음에 따른 주요 성분 탐지 성능 저하를 의미한다. 그러나 -8 dB 이하에서는 PAS가 다시 증가하는 현상이 관찰되었다.

IV. 결론

본 연구에서는 BiLSTM-Attention 모델과 LRP 기법을 활용해 LPI 신호의 분류 성능과 해석 가능성을 평가하였다. LRP가 잡음 속에서도 특정 주파수 성분에 relevance를 집중시킨 결과로, 신호 전체를 명확히 탐지하지 못하더라도 주요 성분을 포착해 성능을 유지할 수 있음을 보여준다. 이러한 결과는 LRP가 잡음 환경에서도 의미 있는 성분을 부각시키며, 일관된 해석과 성능을 제공할 수 있음을 시사한다. 본 연구의 모델은 다양한 환경에서 예측 성능과 해석력을 충족하며, LRP가 신호 분석에 유용한 도구임을 뒷받침한다.

References

- [1] Y.-J. Kim, K. Yu, and J. Han, "Research for radar signal classification model using deep learning technique," *J. Korea Inst. Military Sci. and Technol.*, vol. 22, no. 2, pp. 172-173, 2019. (<https://doi.org/10.9766/KIMST.2019.22.2.170>)
- [2] M. Zhang, L. Liu, and M. Diao, "LPI radar waveform recognition based on time-frequency distribution," *Sensors*, vol. 16, no. 10, p. 1682, Oct. 2016. (<https://doi.org/10.3390/s16101682>)