

Analysis on Underwater Channel by Using Shapley Additive Explanations

Jongseok Kim*, Ho-Shin Cho*, Ohyun Jo^o

ABSTRACT

This study explores the limitations of relying solely on Signal-to-Noise Ratio (SNR) for Bit Error Rate (BER) prediction in underwater communication environments and underscores the critical role of eXplainable Artificial Intelligence (XAI). By employing SHapley Additive exPlanations (SHAP), the relationship between SNR and BER is thoroughly analyzed, highlighting the inadequacies of SNR as the sole predictive feature. To address these challenges, SHAP-based feature selection is utilized to identify key factors, which are subsequently employed to train machine learning models. The results demonstrate a marked improvement in prediction accuracy over traditional methods, affirming that the integration of SHAP-driven feature selection significantly enhances model performance.

Key Words : Underwater Communication, XAI, SHapley Additive exPlanations, Feature Selection

I. Introduction

Underwater communication systems encounter numerous challenges due to the distinct characteristics of underwater environments. The propagation speed of acoustic waves, approximately 1500 m/s, is considered notably slow, making the measurement and track-

ing of channel conditions particularly difficult. Additionally, various environmental factors, including noise, multi-path propagation, and signal attenuation, significantly influence communication quality. Accurately predicting the Bit Error Rate (BER) is crucial, as it directly impacts the efficiency and reliability of underwater communication systems. In conventional outdoor communication systems, SNR has been the most representative metric for indicating channel conditions. However, in underwater communication environments, the aforementioned challenges hinder SNR from accurately reflecting the true state of the channel^[1,2].

To overcome the limitation, machine learning-based approaches have gained attention^[3]. Machine learning models are powerful tools capable of learning complex patterns from high-dimensional data. These models can incorporate not only SNR but also a variety of environmental metrics, enabling more accurate BER prediction. Considering the multiple factors representing channel characteristic in underwater environments, integrating these features into prediction models is essential^[4,5]. Moreover, advancing the interpretability of these models to ensure both higher prediction accuracy and trust in their results remains a pivotal challenge.

We introduce Shapley Additive exPlanations (SHAP) to analyze underwater channel and to predict BER using machine learning models. SHAP might be a useful tool for improving model interpretability. It quantifies the contribution of individual features and facilitates the selection of meaningful features, ultimately leading to enhanced model performance.

II. SHAP : SHapley Additive exPlanations

SHAP is an eXplainable AI (XAI) technique that

※ This work was supported by Korea Research Institute for defense Technology planning and advancement(KRIT) - Grant funded by Defense Acquisition Program Administration(DAPA) (KRIT-CT-23-026).

♦ First Author : (0000-0002-1569-1743) Department of Computer Science, Chungbuk National University, kjseok@chungbuk.ac.kr, 학생회원

^o Corresponding Author : (0000-0001-8444-2786) Department of Computer Science, Chungbuk National University, ohyunjo@chungbuk.ac.kr, 정회원

* (0000-0002-6949-0904) School of Electronic and Electrical Engineering, Kyungpook National University, hscho@ee.knu.ac.kr, 종신회원

논문번호 : 202501-001-A-LU, Received December 30, 2024; Revised February 3, 2025; Accepted February 3, 2025

quantitatively evaluates the contribution of each feature to a machine learning model's prediction^[6]. Derived from game theory, SHAP focuses on fairly calculating and distributing feature contributions. By transforming complex model structures into interpretable formats, it enhances the reliability of predictions and plays a crucial role in optimizing model performance. The core of SHAP lies in generating subsets and calculating prediction changes. The contribution of a specific feature i is defined as follows:

$$\phi_i = \frac{1}{|N|!} \sum_{S \subseteq N \setminus \{i\}} w(S) \cdot (f(S \cup \{i\}) - f(S)) \quad (1)$$

where ϕ_i represents the SHAP value for feature i , S is a subset of N excluding i , and $f(S)$ denotes the model's prediction based on subset S . $w(s)$ represents the weight assigned to the feature combinations, which is calculated as:

$$w(S) = \frac{|S|! (|N| - |S| - 1)!}{|N|!} \quad (2)$$

The SHAP values computed through this formula satisfy the property that the sum of all feature contributions equals the model's final prediction. This enables a precise quantitative evaluation of feature contributions.

SHAP analysis precisely measures the impact of each factor on BER (Bit Error Rate) prediction, allowing for the selection of key features essential for model training. For instance, selecting the k most important features based on SHAP values and including them in the training data can improve model performance. This process can be expressed as:

$$F_{selected} = \{x_i \mid \phi_i \in Top_k(\Phi)\} \quad (3)$$

where F represents the selected feature set, x_i denotes individual features, and ϕ_i is the set of all SHAP values. The condition $Top_k(\Phi)$ indicates that only the features with SHAP values in the top k are selected.

This study utilizes SHAP analysis to identify the features most significantly affecting BER prediction in underwater environments. Based on the analysis,

only the top k features with the highest SHAP values were selected for model training. This approach gradually improved both the efficiency and predictive accuracy of the model. Such a methodology enhances the reliability of the underwater communications.

III. Results and Discussion

Real-world underwater data was collected from the Gulf of Incheon for experimental purposes^[1]. The collected data consists of Data Rate, Mean Excess Delay (MED), Root Mean Square Delay (RMS), Coherence Bandwidth (CB), Doppler Spread, Frequency Shift, and SNR. While BER is the dominant metric to represent the channel, underwater environments require consideration of a wider range of conditions and characteristics. This can be intuitively analyzed by visualizing the correlation between SNR and BER in underwater environments. Figure 1 illustrates the correlation between SNR and BER, while simultaneously visualizing the problem statement addressed.

Based on the observations from the figure, it was confirmed that an increase in SNR does not inherently result in improved BER performance. This highlights the challenges of predicting BER in underwater environments based solely on SNR characteristics. Furthermore, the importance of selecting features beyond SNR for BER prediction through machine learning models is emphasized. SHAP evaluates feature importance to deliver more accurate analyses in this context. Figure 2 illustrates the results of analyzing the relationship between Data Rate and BER using SHAP.

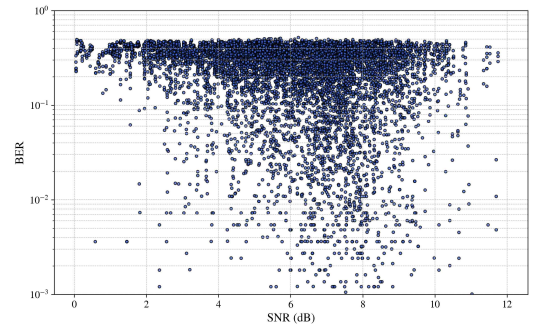


Fig. 1. Correlation Between SNR and BER in Underwater Channel

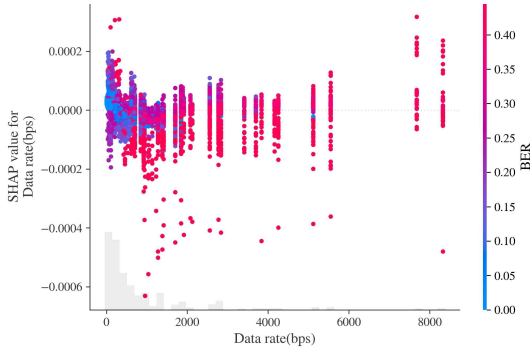


Fig. 2. SHAP Value Analysis for Data Rate in BER Prediction

In Figure 2, the gray bars represent the data distribution according to the data rate, while the color of the dots indicates the corresponding BER values. The yaxis, labeled as SHAP Value, quantifies the contribution of each feature to predicting the BER. A larger absolute value of the SHAP Value, regardless of its sign, implies a higher contribution, whereas values closer to zero indicate a minimal role in the prediction.

Notably, as the data rate increases, the SHAP Value also shows an increasing trend. Specifically, in regions where a higher data rate correlates with higher actual BER values, the SHAP Value reveals a strong relationship between the two features. However, not all features exhibit such distributions of SHAP Values. Figure 3 illustrates the SHAP Value distribution for SNR and its relationship with BER.

In Figure 3, the distribution of SHAP values for SNR in BER prediction is observed to be more concentrated compared to Figure 2. Specifically, while the range of 0 dB to 12 dB includes data points with

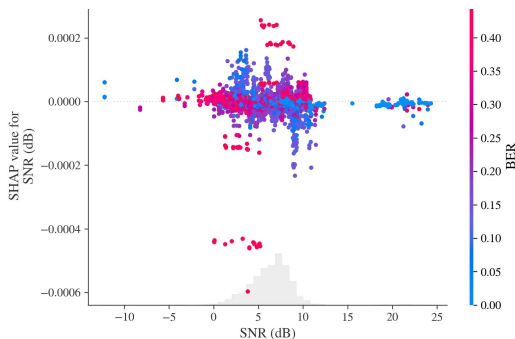


Fig. 3. SHAP Value Analysis for SNR in BER Prediction

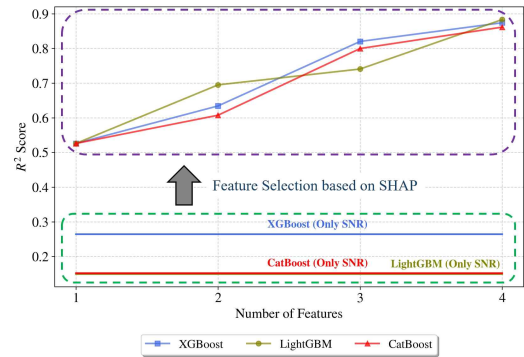


Fig. 4. Impact of Feature Selection on BER Prediction Performance

high SHAP values, the majority of SHAP values are clustered around 0. Furthermore, contrary to expectations that higher SNR would contribute to higher SHAP values in BER prediction, most SHAP values remain near 0 as SNR increases. This suggests that predicting BER based solely on SNR is challenging. Additionally, utilizing features with clearer SHAP values than SNR could lead to the improved prediction performance. To validate this, performance comparisons are conducted using typical machine learning models. Figure 4 illustrates the prediction performance based on the number of selected features.

To assess model performance based on the number of selected features, boosting models such as XG-Boost, LightGBM, and CatBoost were employed. Additionally, the R^2 score, a widely accepted metric in regression analysis, was utilized to compare overall performance. Incorporating features selected based on SHAP's feature importance resulted in approximately a twofold improvement in performance. Furthermore,

as the number of SHAP-selected features increased, all models exhibited an enhancement in performance. These findings underscore the critical role of feature selection in underwater environments, as machine learning models derive benefits from an optimal combination of relevant features. Additionally, SHAP method for feature selection was confirmed as an effective training approach, emphasizing the potential of eXplainable Artificial Intelligence (XAI) in advancing underwater communication systems.

IV. Conclusion

This study highlighted the limitation in prediction of BER by using only SNR in underwater communication environments and demonstrated the need of XAI. By utilizing SHAP-based feature selection, key features were identified, significantly improving prediction accuracy. This clearly underscored the value and importance of feature selection. Future research will focus on developing more advanced feature selection algorithms beyond SHAP to train robust models for diverse underwater environments.

References

- [1] J. Byun, Y.-H. Cho, T. Im, et al., "Iterative learning for reliable link adaptation in the internet of underwater things," *IEEE Access*, vol. 9, pp. 30408-30416, 2021.
(<https://doi.org/10.1109/Access.2021.3058981>)
- [2] T.-G. Chung, H.-L. Ko and T.-H. Im, "Development of an ocean pollution monitoring system using the underwater base station-based underwater communication," *J. KICS*, vol. 47, no. 8, pp. 1162-1169, 2022.
(<https://doi.org/10.7840/kics.2022.47.8.1162>)
- [3] H. Yang, A. Gan, H. Chen, Y. Pan, J. Tang, and J. Li, "Underwater acoustic target recognition using SVM ensemble via weighted sample and feature selection," in *2016 13th Int. Bhurban Conf. Appl. Sci. and Technol. (IBCAST)*, IEEE, pp. 522-527, 2016.
(<https://doi.org/10.1109/IBCAST.2016.7429928>)
- [4] R. Dwivedi, D. Dave, H. Naik, et al., "Explainable ai (xai): Core ideas, techniques, and solutions," *ACM Computing Surveys*, vol. 55, no. 9, pp. 1-33, 2023.
(<https://doi.org/10.1145/3561048>)
- [5] J. Li, K. Cheng, S. Wang, et al., "Feature selection: A data perspective," *ACM Computing Surveys (CSUR)*, vol. 50, no. 6, pp. 1-45, 2017.
(<https://doi.org/10.1145/3136625>)
- [6] S. Lundberg, "A unified approach to interpreting model predictions," *arXiv preprint arXiv:1705.07874*, 2017.
(<https://doi.org/10.48550/arXiv.1705.07874>)