應用資料擴增技術於異常浪機率預警系統之研究

A Study on the Application of Data Augmentation Techniques to the Freak Wave Warning System

陳盈智[1] (Chen Y.-C.)　陳威成[2](Chen W.-C.)　林芳如[3](Lin F.-J.)　潘琦[3](Pan C.)
蔡政翰[4](Tsai C.-H.)　董東璟[12](Doong D.-J.)

[1]國立成功大學近海水文中心　[2]國立成功大學水利與海洋工程學系　[3]中央氣象署海象氣候組
[4]國立臺灣海洋大學海洋環境資訊系

[1]Coastal Ocean Monitoring Center, National Cheng Kung University
[2]Department of Hydraulic and Ocean Engineering, National Cheng Kung University
[3]Marine Meteorology and Climate Division, Central Weather Administration
[4]Department of Marine, Environmental and Informatics, National Taiwan Ocean University

摘　要

異常浪機率預警系統係以AI方法建立，然而如可應用於訓練與驗證的實際案例數量稀少，則難以進行。為解決此問題，本研究嘗試應用資料擴增方法，當資料案例仍不充足時，仍能完成系統建置。本研究採用三種資料擴增技術進行比較與應用，包含噪聲注入、SMOTE和Mixup等方法。噪聲注入法是透過加入高斯隨機噪聲，模擬真實環境變化，有助於模型提升泛化能力。SMOTE方法透過分析少數類別樣本，利用K近鄰演算法在特徵空間內合成新樣本，達成樣本平衡並強化模型訓練成效。Mixup方法則是將兩個樣本依比例混合，對資料與標籤進行線性插值，產生新的虛擬樣本，有效擴展訓練資料的分布範圍。初步測試結果顯示，三種方法各具優勢，皆能在資料量不足與不平衡情況下，有效提升模型效能與泛化能力。本研究期望透過比較不同擴增方法的成效，尋找最適合瘋狗浪預測模型訓練之技術，作為未來資料建模與實務應用之參考依據。

關鍵字：資料擴增、噪聲注入、SMOTE、Mixup

Abstract

The freak wave warning system is developed using AI methods; however, the freak wave cases available for training and validation poses a challenge to system development. To address this issue, this study explores the application of data augmentation techniques to enable system construction even when case data remain insufficient. Three commonly used data augmentation methods were selected for comparison and implementation: Noise Injection, SMOTE, and Mixup. Noise Injection involves adding Gaussian random noise to simulate variations in real environments, thereby enhancing the model's generalization ability. SMOTE analyzes minority class samples and synthesizes new samples in the feature space using the K-Nearest Neighbors algorithm to achieve class balance and improve model training effectiveness. Mixup generates virtual samples by linearly interpolating between two existing samples and their corresponding labels, effectively expanding the distribution of the training dataset. Preliminary test results show that all three methods offer distinct advantages and can significantly improve model performance and generalization under conditions of limited and imbalanced data. This study aims to identify the most suitable technique for training rogue wave prediction models, providing a methodological foundation for future data modeling and practical applications.

Key words : Data Augmentation、Noise Injection、SMOTE、Mixup