World Academy of Science, Engineering and Technology
International Journal of Computer and Systems Engineering
Vol:17, No:07, 2023

Hard Disk Failure Predictions in Supercomputing System Based on CNN-LSTM and Oversampling Technique

Authors : Yingkun Huang, Li Guo, Zekang Lan, Kai Tian

Abstract : Hard disk drives (HDD) failure of the exascale supercomputing system may lead to service interruption and invalidate previous calculations, and it will cause permanent data loss. Therefore, initiating corrective actions before hard drive failures materialize is critical to the continued operation of jobs. In this paper, a highly accurate analysis model based on CNN-LSTM and oversampling technique was proposed, which can correctly predict the necessity of a disk replacement even ten days in advance. Generally, the learning-based method performs poorly on a training dataset with long-tail distribution, especially fault prediction is a very classic situation as the scarcity of failure data. To overcome the puzzle, a new oversampling was employed to augment the data, and then, an improved CNN-LSTM with the shortcut was built to learn more effective features. The shortcut transmits the results of the previous layer of CNN and is used as the input of the LSTM model after weighted fusion with the output of the next layer. Finally, a detailed, empirical comparison of 6 prediction methods is presented and discussed on a public dataset for evaluation. The experiments indicate that the proposed method predicts disk failure with 0.91 Precision, 0.91 Recall, 0.91 F-measure, and 0.90 MCC for 10 days prediction horizon. Thus, the proposed algorithm is an efficient algorithm for predicting HDD failure in supercomputing.

Keywords : HDD replacement, failure, CNN-LSTM, oversampling, prediction
Conference Title : ICCC 2023 : International Conference on Cluster Computing
Conference Location : Tokyo, Japan
Conference Dates : July 17-18, 2023