INTELLIGENT PROCESS SCHEDULING

Optimizing System Resources Utilization with AI

Arwa Fawzy
Supervised by: Dr. Ahmed Baioumy Zaky

May, 2023

CHAPTER 10: INTELLIGENT PROCESS SCHEDULING

The benefits of using AI algorithms in the operating system to optimize scheduling are clear. By improving system performance, adapting to changing workload patterns, and reducing energy consumption, these algorithms can help to ensure that the system is running at optimal efficiency while minimizing operating costs. There are a few papers demonstrated that AI can enhance the performance of process scheduling, tasks operations and system resources utilization. The first paper is titled with "Comparative Analysis of Process Scheduling Algorithm using AI models" **[3]**. It uses Artificial Intelligence models such as Artificial Neural Networks (ANN), Decision Trees (DT), and Random Forests (RF) to evaluate the performance of different process scheduling algorithms. Primarily, the ANN model was implemented by first collecting a dataset of process scheduling simulations. The dataset included input parameters such as process arrival time, burst time, priority, and quantum time, in addition to output parameters such as waiting time, turnaround time, and CPU utilization. The ANN model was trained on the dataset using a backpropagation algorithm to learn the relationship between the input parameters and output parameters. Once the ANN model was trained, it was used to predict the best scheduling algorithm based on the input parameters. Succeeding implementation was a DT model that was applied by collecting a dataset that included input parameters such as process arrival time, burst time, priority, and quantum time, and output parameters such as waiting time, turnaround time, and CPU utilization. It was trained on through an algorithm that recursively split the data based on the most significant attribute or feature until a stopping criterion was met. The resulting tree was then used to predict the best scheduling algorithm based on the input parameters. Eventually, the RF model was implemented by a set of simulations were conducted to schedule processes, and data was gathered including same input parameters of ANN and DT models. The RF model was trained using an ensemble learning algorithm that combined multiple decision trees, where each tree was trained on a randomly selected subset of the data. The predictions of all decision trees were aggregated to make a final prediction. The trained RF model was subsequently used to identify the optimal scheduling algorithm based on input parameters. The paper provides a thorough analysis of various process scheduling algorithms using AI models. The dataset used for training and testing the AI models is sufficiently large and includes relevant input and output parameters. The ensemble learning technique used to train the Random Forest model is appropriate and helps to reduce overfitting. The paper also includes a comparison of the results obtained from different AI models, which adds to the ultimate rigor of the study. It showcases a strong methodology and contributes to the advancement of the field of process scheduling. To illustrate, the ANN model recorded an accuracy of 95.5% and the DT model achieved an accuracy of 89.2% in predicting the best scheduling algorithm for a given set of input parameters.

An additional scholarly article has been published under the title "A Multi-Optimization Technique for Improvement of Hadoop Performance with a Dynamic Job Execution Method Based on Artificial Neural Network **[1].**" It describes a novel method to improve the performance of Hadoop, a popular open-source distributed computing framework. Hadoop is used to store and process large sets of data across a cluster of computers. It is designed to handle big data, which refers to data sets that are too large and complex to be processed using traditional data processing systems and it supports the parallel processing. The proposed approach from the article utilizes a dynamic job execution method based on ANN to optimize multiple performance parameters simultaneously, including resource utilization, throughput, and response time. The study begins with an overview of the challenges associated with Hadoop's performance optimization, including the complex and dynamic nature of distributed computing environments. Then the authors presented the proposed multi-optimization technique, which uses an ANN-based approach to dynamically allocate resources and optimize job execution strategies. The technique involves training an ANN using historical data and using the trained model to predict the optimal job execution strategy in real-time. To evaluate the effectiveness of the proposed approach, the authors conducted a series of experiments using a Hadoop cluster. The experiments involved running various workloads with different job configurations and comparing the performance of the proposed approach with existing techniques. The results of the experiments demonstrate that the proposed approach outperforms existing methods in terms of performance optimization, achieving significant improvements in resource utilization, throughput, and response time. Therefore, this paper offers a valuable contribution to the field of distributed computing, particularly in the context of performance optimization in Hadoop-based systems, and highlights the potential of ANNs for real-time optimization of distributed computing workloads. Specifically, the proposed techniques achieved an average reduction of 28% in job completion time and an average improvement of 14% in resource utilization compared to the traditional scheduler. The ANNs achieved an accuracy of 96.2% in predicting the best execution node for a given job. The ANNs attained an average prediction time of 0.036 seconds per job, which is suitable for real-time job scheduling in Hadoop clusters.

Currently, a published academic work titled "Organization Based Intelligent Process Scheduling Algorithm (OIPSA)" is available for review **[2]**. It discusses the significance of process scheduling in distributed computing environments and the challenges associated with traditional scheduling algorithms. The authors argue that the complexity and heterogeneity of modern computing systems require new approaches that can handle dynamic and unpredictable workloads. They introduce OIPSA as an original scheduling algorithm that is designed to address these challenges using AI techniques. The algorithm consists of three main components: task clustering, organization modeling, and scheduling. The task clustering involves grouping similar tasks together based on their attributes, such as priority, resource requirements, and execution time. The authors explain how clustering can improve the efficiency of scheduling by reducing the number of scheduling decisions that need to be made. Furthermore, organization modeling involves creating a virtual organization that represents the distributed computing environment. The authors explained how the organization model captures information about the available resources, task dependencies, and performance metrics. They also describe how the organization model is used to make intelligent scheduling decisions based on the current state of the system. Scheduling involves using the organization model and task clustering to make scheduling decisions. The authors explain how the scheduling algorithm is designed to be dynamic and adaptable, adjusting to changes in the system and workload. They also discuss how the algorithm incorporates feedback mechanisms to improve performance over time. The experimental results section of the paper presented the findings of a simulation study that evaluates the performance of the OIPSA algorithm compared to traditional scheduling algorithms. The authors provide detailed results and analysis of various performance metrics, including makespan, turnaround time, and resource utilization. They demonstrate that OIPSA outperforms traditional algorithms in terms of efficiency and adaptability. This scholarly paper discussed the advantages of using AI techniques in process scheduling and suggest that OIPSA can serve as a valuable tool for improving the efficiency and effectiveness of scheduling in modern computing systems. It also suggested several directions for future research, including the development of more advanced organization models and the incorporation of additional AI techniques in the scheduling process. OIPSA accomplished an accuracy of 89.72% in predicting the best scheduling policy for a given set of process parameters. Furthermore, it completed an average execution time of 0.28 seconds per process, which is suitable for real-time scheduling applications.

AI algorithms can analyze system workload and resource utilization patterns to identify optimal scheduling policies that ensure that the available resources are used efficiently. This can help to avoid bottlenecks and reduce the likelihood of system crashes, thereby improving system reliability. In addition, AI-based scheduling algorithms can adapt to changing workload patterns and system conditions. This means that the scheduling algorithm can continuously adjust its policies based on feedback from the system to optimize resource usage in real-time. As a result, the system can be more responsive to changing workloads and able to adapt to changing conditions more quickly.

REFERENCES

[1] Alanazi, R., Alhazmi, F., Chung, H. et al. A Multi-Optimization Technique for Improvement of Hadoop Performance with a Dynamic Job Execution Method Based on Artificial Neural Network. SN COMPUT. SCI. 1, 184 (2020). https://doi.org/10.1007/s42979-020-00182-3

[2] Shah, M. A., Shahid, M. B., Zhang, S., Mustafa, S., & Hussain, M. (2017, September). Organization Based Intelligent Process Scheduling Algorithm (OIPSA). In 2017 21st International Conference on Automation and Computing (ICAC) (pp. 1-6). IEEE.

[3] M. M. A. Moni, M. Niloy, A. H. Chowdhury, F. J. Khan, M. F. -U. -A. Juboraj and A. Chakrabarty, "Comparative Analysis of Process Scheduling Algorithm using AI models," 2022 25th International Conference on Computer and Information Technology (ICCIT), Cox's Bazar, Bangladesh, 2022, pp. 587-592, doi: 10.1109/ICCIT57492.2022.10055395.