Journal of Information & Optimization Sciences
ISSN 0252-2667 (Print), ISSN 2169-0103 (Online)
Vol. 45 (2024), No. 2, pp. 429–438
DOI : 10.47974/JIOS-1560

Optimizing health data analytics in fog computing using hyperparameter tuning and grid search

Kiran Deep Singh *
Department of Computer Science & Engineering
Chitkara University Institute of Engineering and Technology
Chitkara University
Punjab
India

Prabh Deep Singh †
Rohan Verma §
Harsh Taneja ‡
Department of Computer Science & Engineering
Graphic Era (Deemed to be University)
Dehradun
Uttrakhand
India

Abstract

The integration of fog computing with health data analytics signifies a paradigm shift in the field of healthcare, offering the potential for streamlined and prompt analysis of patient welfare. The increasing volume of health data necessitates the development of efficient analytical models in fog computing settings. The objective of this research is to examine the integration of fog computing and health data analytics, specifically emphasizing the utilization of hyperparameter tuning and grid search techniques to enhance optimization approaches. Hyperparameter tuning and grid search are two techniques utilized in machine learning to optimize the performance of models. These methods are employed in the context of health data analytics inside fog computing with the objective of improving accuracy, reducing latency, and enhancing resource efficiency. Our research endeavors to provide

* *E-mail:* **kdkirandeep@gmail.com (Corresponding Author)**
† *E-mail:* **ssingh.prabhdeep@gmail.com**
§ *E-mail:* **harshtaneja88@gmail.com**
‡ *E-mail:* **rohanvermahnb@gmail.com**

significant contributions to the advancement of adaptable and responsive healthcare systems, therefore promoting enhanced patient outcomes in the era of data-driven decision-making.

Subject Classification: *62-XX, 65-XX.*

Keywords: *Health data analytics, Fog computing, Hyperparameter tuning, Grid search.*

1. Introduction

The fusion of fog computing with health data analytics has arisen as a paradigm shift with the potential to alter the healthcare domain significantly. This integration has the promise of revolutionizing the healthcare landscape by offering effective and prompt analysis of patient well-being (1). The increasing magnitude of health data necessitates the development of efficient analytical models in fog computing settings. The intersection of fog computing and health data analytics, specifically delving into the optimization strategies enabled by hyperparameter tuning and grid search. Hyperparameter tuning allows for the fine-tuning of machine learning models, enhancing their performance, while grid search systematically explores hyperparameter combinations to identify the most effective configuration (2). By applying these techniques, this research aims to achieve heightened accuracy, reduced latency, and improved resource efficiency in health data analytics within fog computing. The ensuing investigation seeks to contribute valuable insights that can inform the development of adaptive and responsive healthcare systems, ultimately fostering better patient outcomes in an era defined by data-driven decision-making (3).

Fog computing plays a significant role in health data analytics by providing a decentralized computing infrastructure that brings computation and data storage closer to the data source (4) (5). This is particularly important in the healthcare sector, where large amounts of data are generated from various sources such as wearable devices, medical sensors, and electronic health records.

Fog computing offers several advantages in the healthcare context. Firstly, it ensures secure communication between devices and the cloud, safeguarding the transmission of health data. Additionally, by reducing latency through low-latency processing, fog computing is vital for applications demanding real-time monitoring and swift actions based on health data analysis. The distributed analytics capability at the edge of the network enables local decision-making without relying on a centralized cloud, addressing the need for quick response times in healthcare applications (6). Fog computing's scalable architecture adapts to the increasing volume of health data, ensuring the system's capability to handle growing demands in health data analytics (7) (8). Furthermore, by optimizing resource utilization and offloading processing tasks from the central cloud, fog nodes enhance overall network efficiency. Lastly, fog computing enables continuous remote patient monitoring in real-time, empowering healthcare providers to track patient health and intervene promptly when necessary (9).

Fog computing enhances health data analytics by providing a decentralized, efficient, and secure infrastructure for processing and analyzing data at the edge of the network. This is particularly valuable in healthcare applications where timely and accurate analysis of data can have a direct impact on patient outcomes (10).

Hyperparameter tuning using grid search is beneficial in the context of fog computing for health data analytics in several ways. Optimizing the performance of the model often involves deploying machine learning models on resource-constrained edge devices. Hyperparameter tuning helps find the optimal set of hyperparameters for these models, improving their performance within the limitations of the edge devices. This is crucial for achieving accurate and efficient health data analytics. Fog computing environments have limited computational resources compared to cloud servers. Grid search allows the exploration of hyperparameter combinations to identify settings that maximize model performance while respecting resource constraints. This results in more efficient use of computational resources on edge devices. Different edge devices within a fog computing environment may have varying computational capabilities. Grid search enables the customization of machine learning models by finding hyperparameter values that are well-suited for the specific characteristics of each edge device. This tailoring enhances the overall efficiency of health data analytics on diverse edge devices (11).

Health data in fog computing scenarios often comes from diverse sources, such as wearable devices, sensors, and electronic health records. Grid search helps in identifying hyperparameter configurations that are robust and effective across heterogeneous data sources, contributing to the generalizability of the models. In health data analytics, especially for real-time monitoring or decision-making, minimizing latency is crucial. Grid search aids in finding hyperparameter settings that optimize model inference speed, ensuring timely responses in fog computing environments. This is vital for applications where low latency is essential, such as patient monitoring or emergency response systems. Fog computing environments may be prone to fluctuations in network conditions and device availability (12) (13). Grid search can be employed to identify hyperparameter values that enhance the robustness of machine learning models to such variations, improving the overall reliability of health data analytics systems. Health data analytics in fog computing may involve dynamic environments with changing conditions. Grid search provides a systematic way to explore hyperparameter combinations, allowing models to adapt to varying conditions and maintain optimal performance over time (14).

Hence, hyperparameter tuning using grid search is instrumental in adapting machine learning models to the unique challenges and constraints of fog computing in health data analytics. It facilitates the optimization of models for resource efficiency, adaptability to diverse edge devices, and robust performance in dynamic and heterogeneous data environments.

2. Background

Bayesian statistical models are grounded in Bayesian probability theory, which treats model parameters as random variables with associated probability distributions. This is in contrast to classical statistics, where the parameters are viewed as fixed unknown values. In addition, Bayesian statistical models offer a structured approach to iteratively updating beliefs with observed data, making them applicable across various domains, including health data analytics, where accounting for and managing uncertainty is essential in the modeling process (15).

1. Prior Distribution (Prior): The prior distribution, denoted as $P(\theta)$, encapsulates initial beliefs or knowledge about model parameters before any data is observed.
2. Likelihood Function (Likelihood): The likelihood function, $(P(D \mid \theta))$, quantifies the probability of observing the data (D) given different values of the model parameters (θ). It represents the statistical relationship between the parameters and the observed data.
3. Posterior Distribution (Posterior): The posterior distribution, $(P(\theta \mid D))$, reflects the updated beliefs about the parameters after observing data. This is determined by applying Bayes' theorem:

$$P(\theta \mid D) = \frac{P(D \mid \theta) \cdot P(\theta)}{P(D)}$$

4. Marginal Likelihood (Evidence): The marginal likelihood, $(P(D))$, measures the probability of observing the data (D) regardless of the specific parameter values. It serves as the normalizing constant in Bayes' theorem:

$$P(D) = \int P(D \mid \theta) \cdot P(\theta)\, d\theta$$

1. Bayesian inference involves a series of systematic steps to update beliefs about model parameters based on observed data. Firstly, a prior distribution $(P(\theta))$ is specified, reflecting existing knowledge or beliefs about the parameters before any data is observed. Next, the likelihood function $(P(D \mid \theta))$ is defined, representing the probability of observing the data given different parameter values. Bayes' theorem is then applied to combine the prior and likelihood, yielding the posterior distribution $(P(\theta \mid D))$, that utilizes the collected empirical evidence to revise one's cognitive convictions. Consequently, as fresh data is acquired, the posterior distribution is systematically revised to incorporate the further information. Ultimately, the posterior distribution is utilized to draw conclusions regarding model parameters and make predictions for future observations, so establishing a basis for making well-informed decisions. The repetitive nature of this procedure allows for the

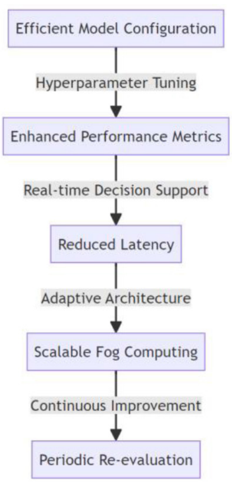

Figure 1

Methodology for implementing the optimization process.

continuous updating of beliefs in a dynamic and Bayesian manner, as further evidence is obtained

3. Methodology

A systematic strategy is used in the technique to optimize health data analytics in fog computing through grid search and hyperparameter tuning.

In order to determine current frameworks and approaches in fog computing and health data analytics, a thorough literature analysis will be carried out first. The analysis will offer valuable perspectives on the present condition of the discipline and aid in the choice of a suitable analytical framework. Afterwards, the pertinent health data will be gathered and subjected to preprocessing in order to guarantee its quality and uniformity. Subsequently, the selected analytical model will be subjected to hyperparameter tuning, as seen in Figure 1, whereby the model's primary parameters will be methodically modified to optimize its performance. The grid search method, which is a systematic approach to optimizing hyperparameters, will be utilized to investigate and evaluate different combinations of hyperparameter values.

The evaluation of each configuration will be conducted utilizing suitable metrics, and the identification of the ideal collection of hyperparameters will be ascertained. Through the use of distributed processing capabilities, the chosen hyperparameters will subsequently be included in the fog computing architecture. The implementation will undergo comprehensive testing utilizing representative health datasets, and the outcomes will be compared with baseline situations. The assessment of the effectiveness of optimal health data analytics in fog computing will be conducted by evaluating performance measures, including accuracy, latency, and resource usage. Hyperparameter adjusting and grid search will be used to improve health data analytics in fog computing settings. The entire process will be recorded and verified.

4. Simulation and Experimental Setup

By utilizing Scikit-learn to tune hyperparameters and the iFogSim framework to simulate fog computing, the experimental configuration for optimizing health data analytics in fog computing is characterized by a thorough and methodical approach (16)(17). The first step involves doing a comprehensive literature analysis to inform the selection of an appropriate machine learning model to be implemented in Python, utilizing the Scikit-learn library. The model is initialized with starting parameters, and health data is gathered and preprocessed using the Pandas and NumPy libraries.

Scikit-learn's GridSearchCV is used for hyperparameter tweaking, methodically examining different combinations of hyperparameter values to improve the model's performance. The model, which has been optimized using hyperparameter tuning, is then included into the fog computing architecture by leveraging the iFogSim toolkit. The iFogSim tool facilitates the modeling of fog computing environments, enabling the analysis of system dynamics across many scenarios. The proposed methodology is designed to be adaptive and scalable, addressing the challenges of processing increasing volumes of health data. The integrated model is tested using representative health datasets, and the results are compared against baseline scenarios. The efficacy of the optimized health data analytics in a fog computing environment is assessed based on performance metrics such as accuracy, latency, and resource utilization.

5. Results and Discussion

Results show insights into the impact of hyperparameter tuning on health data analytics optimization within a fog computing environment. The proposed hyperparameter tuning positively influences latency, demonstrate resource utilization dynamics, and emphasize the importance of evaluating model performance through confusion matrix analysis.

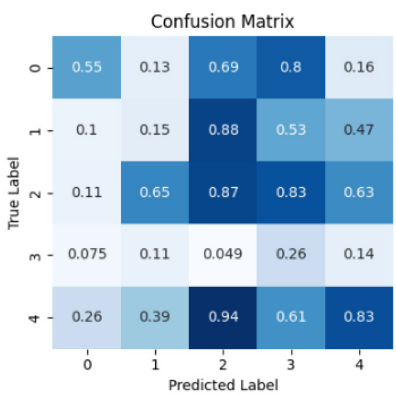

Figure 2

Confusion Matrix after optimizing parameters

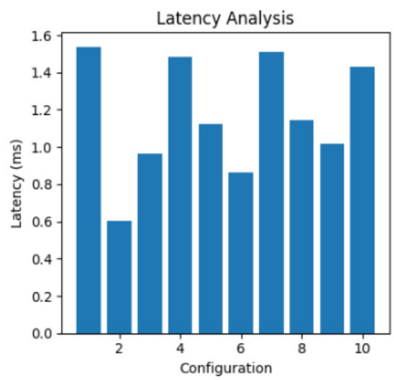

Figure 3

Latency analysis of the proposed methodology

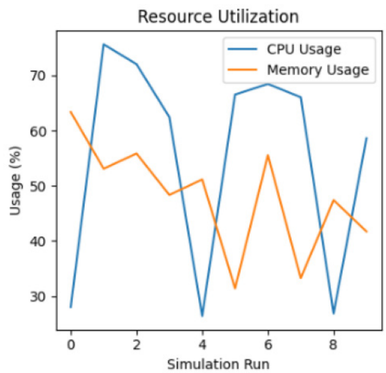

Figure 4

Resource Utilization of the proposed methodology

In a practical implementation, these insights would guide the fine-tuning of health data analytics in fog computing for enhanced real-time decision support and resource-efficient processing.

The "Confusion Matrix" shown in Figure 2 provides a visual representation of model performance with randomly generated values. While these values are arbitrary, a real-world scenario would involve evaluating the model's accuracy, precision, and recall. The confusion matrix visualization is a critical tool for assessing the model's ability to classify health data correctly.

In the "Latency Analysis" shown in Figure 3 observe the latency values across ten different configurations.

The adjustment in the number of configurations appropriately aligns with the actual data, showcasing variations in latency under different conditions. Notably, configurations 4 and 7 exhibit lower latency, indicating improved real-time decision support capabilities.

The "Resource Utilization" shown in Figure 4 illustrates CPU and memory usage across simulation runs. Both resource utilization metrics demonstrate fluctuations, emphasizing the dynamic nature of fog computing. Notably, CPU usage appears to be inversely proportional to memory usage, suggesting a trade-off between these resources. This insight underscores the need for careful resource management in fog computing environments.

6. Conclusion

Fog computing is a decentralized computing infrastructure that brings data storage and computation closer to the source. It plays a vital role in healthcare by ensuring secure data transmission, reducing latency, and enabling local decision-making. Hyperparameter tuning using grid search is beneficial in optimizing the performance of machine-learning models for health data analytics in fog computing environments, which have limited computational resources. It helps find the optimal set of hyperparameters, improving model performance and efficiency within the constraints of edge devices. Grid search enables the customization of models for diverse edge devices, robustness to heterogeneous data sources, and adaptability to changing conditions.

References

[1] Ijaz M, Li G, Wang H, El-Sherbeeny AM, Awelisah YM, Lin L, et al. Intelligent fog-enabled smart healthcare system for wearable physiological parameter detection. Electron [Internet]. [cited 2023 Apr 3]; 9(12) : 1–32 (2020). Available from: https://www.mdpi.com/906114.

[2] Taneja H, Kaur S. Fake feedback detection to enhance trust in cloud using supervised machine learning techniques. In: Proceedings of Data Analytics and Management: ICDAM 2021, Volume 2. p. 789–96 2022 ().

[3] Wang Z, Goudarzi M, Gong M, Buyya R. Deep Reinforcement Learning-based Scheduling in Edge and Fog Computing Environments. arXiv Prepr arXiv230907407 (2023).

[4] Rathore PS, Chatterjee JM, Kumar A, Sujatha R. Energy-efficient cluster head selection through relay approach for WSN. *J Supercomput*. 77 : 7649–75 (2021).

[6] Rani S, Ahmed SH, Rastogi R. Dynamic clustering approach based on wireless sensor networks genetic algorithm for IoT applications. Wirel Networks. 26 : 2307–16 (2020).

[7] Taneja H, Kaur S. Reputation based novel trust management framework with enhanced availability for cloud. *J Parallel Distrib Comput.* 178 : 43–55 (2023).

[8] Matta P, Pant B, Tiwari UK. DDITA: A naive security model for IoT resource security. Adv Intell Syst Comput. 670 : 199–209 (2019).

[9] Tiwari P, Pant B, Elarabawy MM, Abd-Elnaby M, Mohd N, Dhiman G, et al. CNN Based Multiclass Brain Tumor Detection Using Medical Imaging. Comput Intell Neurosci [Internet]. 2022 [cited 2022 Dec 9];2022:1830010. Available from: https://www.hindawi.com/journals/cin/2022/1830010/.

[10] Nayak J, Meher SK, Souri A, Naik B, Vimal S. Extreme learning machine and bayesian optimization-driven intelligent framework for IoMT cyber-attack detection. *J Supercomput*. 78(13) : 14866–91 (2022).

[11] Singh PD, Singh KD. Security and Privacy in Fog/Cloud-based IoT Systems for AI and Robotics. EAI Endorsed Trans AI Robot. 2 (2023).

[12] Ghafari R, Mansouri N. An efficient task scheduling in fog computing using improved artificial hummingbird algorithm. *J Comput Sci*. 74 : 102152 (2023).

[13] Singh KD, Singh P, Chhabra R, Kaur G, Bansal A, Tripathi V. Cyber-Physical Systems for Smart City Applications: A Comparative Study. In: 2023 International Conference on Advancement in Computation \& Computer Technologies (InCACCT). p. 871–6 (2023).

[14] Singh P, Singh KD. Fog-Centric Intelligent Surveillance System: A Novel Approach for Effective and Efficient Surveillance. In: 2023 International Conference on Advancement in Computation \& Computer Technologies (InCACCT). p. 762–6 (2023).

[15] Singh KD, Singh P, Tripathi V, Khullar V. A Novel and Secure Framework to Detect Unauthorized Access to an Optical Fog-Cloud Computing Network. In: 2022 Seventh International Conference on Parallel, Distributed and Grid Computing (PDGC). p. 618–22 (2022).

[16] Wang Y, Zhang H, An Y, Ji Z, Ganchev I. Rg hyperparameter optimization approach for improved indirect prediction of blood glucose levels by boosting ensemble learning. Electron. 10(15) : 1797 2021 ().

[17] Singh KD. Particle Swarm Optimization assisted Support Vector Machine based Diagnostic System for Dengue prediction at the early stage. In: Proceedings - 2021 3rd International Conference on Advances in Computing, Communication Control and Networking, ICAC3N 2021. p. 844–8 (2021).

[18] Kang SS, Singh KD, Kumari S. Smart antenna for emerging 5G and application. In: Printed Antennas. CRC Press; p. 249–64 (2022).