Research on Synchronization Mechanism of Artificial Intelligence Model Based on Belief Measure and Its Application in Pumped Storage Power Station

Jinliang Dong\textsuperscript{a}, Xu Zhang\textsuperscript{b}, Haijiang Li\textsuperscript{c}, Wenzhi Song\textsuperscript{d}, Jinglin Guo\textsuperscript{e}\textsuperscript{*}

\textsuperscript{a}Xinjiang Hami pumped storage Co., Ltd., Hami 839099, China
\textsuperscript{b}Beijing Guodiantong Network Technology Co., Ltd., Beijing 100000, China
\textsuperscript{c}137265241@qq.com, \textsuperscript{b}94698198@qq.com, \textsuperscript{c}2206496336@qq.com, \textsuperscript{d}songwenzhi@sgitg.sgcc.com.cn, \textsuperscript{e}478817379@qq.com

Abstract—For the security monitoring of pumped storage power station, a model synchronization mechanism for cloud edge cooperation framework is proposed. The method uses the belief function to describe the threshold and uses the ping-pong operation strategy to update the model alternately, which solves the problem of artificial intelligence model synchronization and update of edge equipment. The cloud is based on Baidu BML platform, the edge uses customized servers, and the average model update cycle is about three months.

1. Introduction
With the implementation of the double carbon policy, the construction of energy Internet has accelerated. Pumped storage power station (PSPS) is an important part of energy Internet. In China, about more than 60 PSPSs have been built or are under construction. "Few people on duty" is a feature of PSPS. Therefore, artificial intelligence technology has a wide application prospect in PSPS. This work proposes a model synchronization mechanism for cloud edge cooperation framework, which solves the problem of artificial intelligence model synchronization and update of edge devices.

2. System overview
2.1 Basic ideas
The artificial intelligence models used in the security monitoring of PSPS can be divided into two categories: (1) applications less related to the environmental background, such as not wearing safety helmets and smoking, etc.; (2) Applications related to environmental background, such as plant scaffold construction, etc. For (1), all PSPSs adopt the general model; for (2), personalized model is adopted for different PSPS. In the cloud, the basic sample warehouse and model warehouse are realized. At the same time, personalized samples are saved for different PSPS, and the corresponding models are trained. At the Edge, the customized equipment is used to realize the real-time image detection, and the model synchronization and update mechanism and sample return mechanism are designed and implemented.
2.2 Introduction to system architecture
Edge AI has the potential to benefit both the data and the network infrastructure itself. At a network level, it could be used to analyze the flow of data for network prediction and network function management, while enabling edge AI to make decisions over the data itself offers significantly reduced backhaul to the cloud, negligible latency and improved security, reliability and efficiency across the board. Another key function of edge AI is sensor fusion: combining the data from multiple sensors to create complex pictures of a process, environment or situation. Consider an edge AI device in an industrial application, tasked with combining data from multiple sensors within a factory to predict when mechanical failure might occur. This edge AI device must learn the interplay between each sensor and how one might affect the other and apply this learning in real-time.

In this paper, Baidu BML platform is used as the bottom support [1], and customized server is used to realize the edge computing support.

In Figure 2, M: N means the comparison of basic samples and personalized samples in the training set.

2.3 Near edge device
In the implementation of Near Edge device, the heterogeneous computing [2] mode is adopted, the main FPGA accelerates the common model, and the h.265 decoding card is used to improve the speed of video coding and decoding, and GPU is used to realize the image recognition based on deep learning. Furthermore, in the airborne equipment of the project, the Yolo V4 [3] framework is adopted. A large
number of calculations are performed at the edge, and the cloud is mainly responsible for training the model.

2.3.1 PCI-E communication module
PCI-E is a high-speed serial point-to-point dual channel high bandwidth transmission. In many applications, PCI-E switching is the main implementation way in heterogeneous systems. Its main advantage is high data transmission rate. The highest 16X2 version now achieves 10GB/s. In fact, in general applications, high speed Ethernet technology can be used in data communication, and the advantage of using PCI-E is that the real time is guaranteed.

2.3.2 FPGA module
FPGA is more efficient than CPU even GPU, mainly due to its architecture without instruction and no memory sharing. There are two functions of memory in the von Neumann structure. One is the save state, and the two is the inter cell communication. Because memory is shared, we need to do access arbitration. In order to make use of the locality of access, every executive cell has a private cache, which means we need to maintain consistency between components. For the needs of saving state, registers and on chip memory (BRAM) in FPGA belong to their own control logic, without unnecessary arbitration and caching. For communication needs, the connection between each logical unit of FPGA and surrounding logic units has been determined during reprogramming, and no communication is required through shared memory. In this work, Xilinx Virtex-6 is used.

2.3.3 Performance optimization
In order to improve the performance, in the case of engineering application precision, this paper made three changes during the implementation process.

The decode image is completed by using the H.265 decoding card.
2.3.4 Communication in the microservice

In the near edge server, it is a better choice to support endpoint devices with microservice architecture. In the microservice architecture, the patterns of communication between clients and the application, as well as between application components, are different than in a monolithic application. There are two main approaches to inter-process communication in a microservice architecture. One option is to use a synchronous HTTP-based mechanism such as REST (Representational State Transfer) or SOAP (Simple Object Access Protocol). This is a simple and familiar IPC (Inter-Process Communication) mechanism. It’s firewall friendly so it works across the Internet and implementing the request/reply style of communication is easy. The downside of HTTP is that it doesn’t support other patterns of communication such as publish-subscribe.

In this work, we use one FPGA and six GPUs (each about 20Tflops), which can support about 128 channels of 1080p video monitoring at the same time.

3. Ping-pong operation

Ping Pong operation is a data buffer optimization design technology in FPGA development[4], which can be regarded as another form of pipeline technology. This paper uses its design idea for reference, combines it with reliability function and applies it to model updating.
Strategy of Ping Pong operation resolver

1) Define the belief function BF (which can be replaced by Bayesian formula). In case of simplification, F1 score can be used instead of this function [5].

2) For the same target problem, two yolov4 models are trained with different sample sets without losing generality, model a and model b.

3) According to the performance of model a and model b in the test set, the priority is defined.

4) Model a and model b are deployed on edge devices.

5) In the process of edge target recognition, when the results of model a and model b are different, the difference image is uploaded to the cloud for manual confirmation, and the confirmed results are used as the evaluation of model a and model b and accumulated.

6) When the difference reliability function is greater than the threshold, the model with low evaluation score is updated.

4. Conclusion

Based on the results and discussions presented above, the conclusions are obtained as below:

1) Ping Pong operation strategy can effectively improve the accuracy of artificial intelligence model.

2) The performance of customized server is better than general server in cloud side collaborative application.

References

[1] Zhong Z , Zheng L , Luo Z , et al. Invariance Matters: Exemplar Memory for Domain Adaptive Person Re-identification[J]. 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), 2019.

[2] Shuai C , Boyer M , Meng J , et al. Rodinia: A benchmark suite for heterogeneous computing[C]// Proceedings of the 2009 IEEE International Symposium on Workload Characterization, IISWC 2009, October 4-6, 2009, Austin, TX, USA. IEEE, 2009.

[3] Bochkovskiy A , Wang C Y , Liao H . YOLOv4: Optimal Speed and Accuracy of Object Detection[J]. 2020.

[4] Zhang Z H , Chu W , Huang S Y . Ping-Pong Methodology for Boosting the Resilience of Cell-Based Delay-Locked Loop[J]. IEEE Access, 2019, PP(99):1-1.

[5] Huang H , Xu H , Wang X , et al. Maximum F1-Score Discriminative Training Criterion for Automatic Mispronunciation Detection[J]. IEEE/ACM Transactions on Audio, Speech, and Language Processing, 2015, 23(4):787-797.