EDGE AIBENCH:
TOWARDS COMPREHENSIVE END-TO-END EDGE COMPUTING BENCHMARKING

EDITED BY

TIANSHU HAO
YUNYOU HUANG
XU WEN
WANLING GAO
FAN ZHANG
CHEN ZHENG
LEI WANG
HAINAN YE
KAI HWANG
ZUJIE REN
JIANFENG ZHAN

BenchCouncil: International Open Benchmarking Council
Chinese Academy of Sciences
Beijing, China
http://www.benchcouncil.org/EdgeAIBench/index.html

TECHNICAL REPORT NO. BENCHCOUNCIL-EDGEAIBENCH-2019
AUG 12, 2019

Edge AIBench: Towards Comprehensive End-to-end Edge Computing Benchmarking

Tianshu Hao[1,2], Yunyou Huang[1,2], Xu Wen[1,2], Wanling Gao[1,3], Fan Zhang[1], Chen Zheng[1,3], Lei Wang[1,3], Hainan Ye[3,4], Kai Hwang[5], Zujie Ren[6], and Jianfeng Zhan*[1,2,3]

[1]State Key Laboratory of Computer Architecture, Institute of Computing Technology, Chinese Academy of Sciences , {haotianshu, huangyunyou, wenxu, gaowanling, zhangfan, wanglei_2011, zhanjianfeng}@ict.ac.cn
[2]University of Chinese Academy of Sciences
[3]BenchCouncil (International Open Benchmarking Council)
[4]Beijing Academy of Frontier Sciences and Technology, {yehainan}@mail.bafst.com
[5]Chinese University of Hongkong at Shenzhen, {hwangkai}@cuhk.edu.cn
[6]Zhejiang Lab

Aug 12, 2019

Abstract

In edge computing scenarios, the distribution of data and collaboration of workloads on different layers are serious concerns for performance, privacy, and security issues. So for edge computing benchmarking, we must take an end-to-end view, considering all three layers: client-side devices, edge computing layer, and cloud servers. Unfortunately, the previous work ignores this most important point. This paper presents the BenchCouncil's coordinated effort on edge AI benchmarks, named Edge AIBench. In total, Edge AIBench models four typical application scenarios: ICU Patient Monitor, Surveillance Camera, Smart Home, and Autonomous Vehicle with the focus on data distribution and workload collaboration on three layers. Edge AIBench is publicly available from `http://www.benchcouncil.org/EdgeAIBench/index.html`. We also build an edge computing testbed with a federated learning framework to resolve performance, privacy, and security issues.

*Jianfeng Zhan is the corresponding author.

1 Introduction

Cloud computing is a mature model to share computing resources by providing network access to users [1]. In cloud computing models, users communicate with the data center to get hardware, software and other computing resources and store data. However, In recent years, the number of client-side devices (e.g. smart devices and monitors) grows rapidly. IoT Analytics [2] has reported the number of connected devices reached 17 billion in 2018 and Gartner says the IoT devices will install 26 billion units by 2020 [3]. These client-side devices produce a large amount of data to process. The overhead of data transmission and data encryption among devices and data centers becomes significant bottlenecks for many IoT scenarios, and hence it raises a daunting challenge for throughput, latency, and security guarantee.

Edge computing emerges as a promising technical framework to overcome the challenges in cloud computing. The edge computing framework adds a new layer, named the edge computing layer, on the basis of the traditional cloud computing framework. In the edge computing framework, only the real-time data processing is transferred to the edge computing layer, while other complicated data processing is still executed on the cloud server. Figure 1 shows a general edge computing framework, which includes three layers: cloud server, edge computing layer, and client-side devices.

In the edge computing scenarios, the distribution of data and collaboration of workloads on different layers are serious concerns for performance, security, and privacy issues. So for benchmarking, designing, and implementing edge computing systems or applications, we shall take an end-to-end view, considering all three layers. Unfortunately, the previous work, especially the previous benchmarking efforts [4, 6, 7, 14] ignore this most important point.

In edge computing scenarios, AI techniques are widely used to augment device, edge and cloud intelligence, and they are most demanding in terms of computing power, data storage, and network. Typical application scenarios include smart city, smart home, autonomous vehicle, surveillance camera, smart medical, wearable devices and so on. These scenarios are complicated because of different kinds of client-side devices, a large quantity of heterogeneous data, privacy and security issues. Most of these scenarios have a high requirement for latency and network bandwidth. However, edge computing is in the initial stage and doesn't have a uniform standard for these scenarios. Therefore, a comprehensive end-to-end edge computing benchmark suite is needed to measure and optimize the systems and applications.

Meanwhile, edge computing is still in the initial stage with a lack of testbed. Because of the privacy issue, there is no incentive to share data. Because of the complexity, there is no end-to-end application scenario to validate the architectures, systems, or specific algorithms in certain settings.

Above all, it's necessary to develop a benchmark suite and testbed for edge computing. This paper reports the BenchCouncil's coordinated effort on edge AI benchmarks, named Edge AIBench, which is publicly available from`http://www.benchcouncil.org/EdgeAIBench/index.html`. Edge AIBench includes four typical application scenarios: ICU Patient Monitor, Surveillance Camera, Smart Home, and Autonomous Vehicle, which consider the complexity of all edge computing AI scenarios. Coordinated by BenchCouncil (http://www.benchcouncil.org), we are also building an edge computing testbed with a federated learning framework to resolve security and privacy issue, which can be accessed from`http://www.benchcouncil.org/testbed/index.php`. BenchCouncil also release datacenter AI benchmarks [8, 9], HPC AI benchmarks [10], IoT AI benchmarks [11], and big data benchmarks [12, 13], publicly available from the BenchCouncil website.

2 Related Work

Since the edge computing AI applications have become more and more popular these years, benchmarks are needed to measure and optimize the systems and applications. There are several related benchmark suites. We summarize the state-of-the-art and state-of-the-practice work on edge AI benchmarking.

MLPerf [4] is a benchmark suite focusing on measuring Machine Learning(ML) performance. It provides the edge inference benchmarks, including eight ML tasks: image classification, object detection and so on [5]. But this benchmark suite just evaluates the edge computing layer with the lack of an

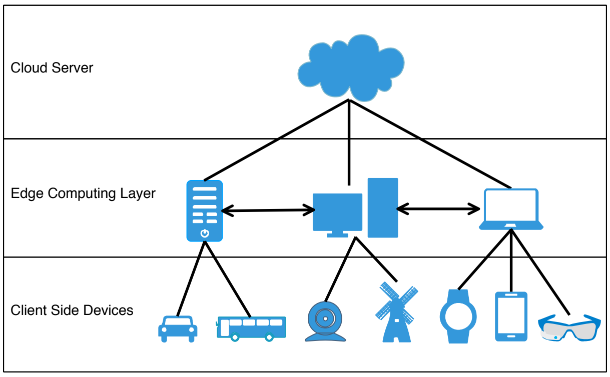

Figure 1: A General Edge Computing Framework.

Table 1: Comparison among Edge Computing AI benchmarks

Benchmark Name	End-to-end Application Scenarios	Components on Cloud Server	Components on Edge Computing Layer	Components on Client-side Devices	Open-Source
Edge AIBench	ICU Patient Monitor Surveillance Camera Smart Home Autonomous Vehicle	√	√	√	√
MLPerf	N/A	×	√	×	√
EEMBC MLMark	Not Clear	Not Clear	Not Clear	Not Clear	×
EdgeBench	N/A	√	√	√	√
AI Benchmark	N/A	×	×	√	√

end-to-end view.

EEMBC [6] develops an ML benchmark suite, named MLMark on embedded edge computing platforms. MLMark includes four AI applications: image classification, object detection, language translation, and speech recognition. However, only the licensees and members of EEMBC have the right to access these benchmarks and this benchmark suite is still in "beta" state now.

EdgeBench [14] compares two edge computing platforms–Amazon AWS Greengrass and Microsoft Azure IoT Edge. And it includes two AI applications: speech-to-text and image recognition. EdgeBench fails to provide an end-to-end application benchmarking framework.

AI Benchmark [7] is a benchmark suite for AI applications on smartphones, and it includes nine AI applications. It's an IoT benchmark suite and only focuses on the client-side devices (smartphones)' performance.

Table 2 compares the state-of-the-art and state-of-the-practice edge computing AI benchmarks. It shows many of them only focus on the edge computing layer instead of the whole edge computing framework. Our benchmark suite Edge AIBench provides an end-to-end application benchmarking framework, including train, validate and inference stages. Moreover, Edge AIBench includes four typical edge computing AI scenarios and measures the whole three-layer edge computing framework.

3 The summary of Edge AIBench

Edge AIBench includes four typical scenarios: Intensive care unit(ICU) patient monitor, surveillance camera, smart home, and autonomous vehicle. These four AI scenarios can present the complexity of edge computing AI scenarios from different perspectives.

3.1 ICU Patient Monitor

ICU is the treatment place for critical patients. Therefore immediacy is significant for ICU patient monitor scenario to notify doctors of the patients' status as soon as possible. The dataset we use is MIMIC-III [15]. MIMIC-III provides many kinds of patients data such as vital signs, fluid balance and so on. Moreover, we choose heart failure prediction [16] and endpoint prediction [17] as the AI benchmarks.

Heart failure prediction uses the MIMIC-III dataset and a two-level neural attention model. It collects the patients' data on the virtual client-side devices, trains on the cloud server (the data will be sent from the edge) and predicts the heart failure on the edge computing layer.

Endpoint prediction benchmark also uses the MIMIC-III dataset, and it uses an LSTM model. This benchmark collects patients' data on the virtual patient device generator and then transmit it to the edge to make the inference. Then the data will be sent to the cloud server to do more training.

3.2 Surveillance Camera

There are many surveillance cameras all over the world nowadays, and these cameras will produce a large quantity of video data at all times. If we transmit all of the data to cloud servers, the network transmission bandwidth will be very high. Therefore, this scenario focus on edge data preprocesses and data compression.

We choose the person re-identification application as the component benchmark. It collects data from the virtual camera devices and pre-process and infer these video data on the edge computing layer. Then the edge computing layer will send the compressed data to the cloud server. Moreover, the decompression and training process are on the cloud server.

3.3 Smart Home

Smart home includes a lot of smart home devices such as automatic controller, alarm system, audio equipment and so on. Thus, the uniqueness of the smart home includes different kinds of edge devices and heterogeneous data. We will choose two AI applications as the component benchmarks: speech recognition and face recognition. These two components have heterogeneous data and different collecting devices. These two component benchmarks both collect data on the client side devices(e.g. camera and smartphone), infer on the edge computing layer and train on the cloud server.

Speech recognition uses the DeepSpeech2 [18] model and the LibriSpeech dataset [19].

Face recognition uses the FaceNet [20] model and uses the LFW(Labeled Faces in the Wild) [21] dataset.

3.4 Autonomous Vehicle

The uniqueness of the autonomous vehicle scenario is that the high demand for validity. That is to say, it takes absolute correct action even without human intervention. This feature represents the demand of some edge computing AI scenarios. The automatic control system will analyze the current road conditions and make a corresponding reaction at once. We will choose the road sign recognition as the component benchmark.

The road sign recognition will collect the road signs data from the camera, train these data on the cloud and infer on the edge computing layer.

Table 2 shows the component benchmarks of Edge AIBench. Edge AIBench provides an end-to-end

Table 2: The Summary of Edge AIBench

End-to-end Application Scenarios	AI Component Benchmarks	Cloud Server	Edge Computing Layer	Client Side Device
ICU Patient Monitor	Heart Failure Prediction	Train	Infer Send Alarm	Generate Data
ICU Patient Monitor	Endpoint Prediction	Train	Infer	Generate Data
Surveillance Camera	Person Re-Identification	Decompress Data Train	Compress Data Infer	Generate Data
Smart Home	Speech Recognition	Train	Infer	Generate Data
Smart Home	Face Recognition	Train	Infer	Generate Data
Autonomous Vehicle	Road Sign Recognition	Train	Infer	Generate Data

application benchmarking, consisting of train, inference, data collection and other parts using a general three-layer edge computing framework.

3.5 A Federated Learning Framework Testbed

We have developed an edge computing AI testbed to provide support for researchers and common users, which is publicly available from `http://www.benchcouncil.org/testbed.html`. Security and privacy issues become significant focuses in the age of big data, as well as edge computing. Federated learning is a distributed collaborative machine learning technology whose main target is to preserve the privacy [22]. Our testbed system will combine the federated learning framework.

At present, we are implementing the ICU scenario on the testbed. We are developing a "virtual patient" data generator and a federated machine learning training model. Doctors can train the model on the local server and transmit the encrypting parameter to the cloud server. Then the cloud server computes the overall parameter on the basis of these encrypting parameter from different hospitals. After all, the cloud server will send the overall parameter to the local server and the local server will decrypt it to update their models. Figure 2 shows our federated learning testbed framework.

4 Conclusion

This paper presents an edge computing AI benchmark, named Edge AIBench, which consists of four end-to-end application benchmarking framework and six component benchmarks. These scenarios we choose can present the complexity of edge computing scenarios from different perspectives. Also, we build an edge computing AI testbed with a federated learning framework.

5 Acknowledgment

This work is supported by the Standardization Research Project of Chinese Academy of Sciences No.BZ201800001.

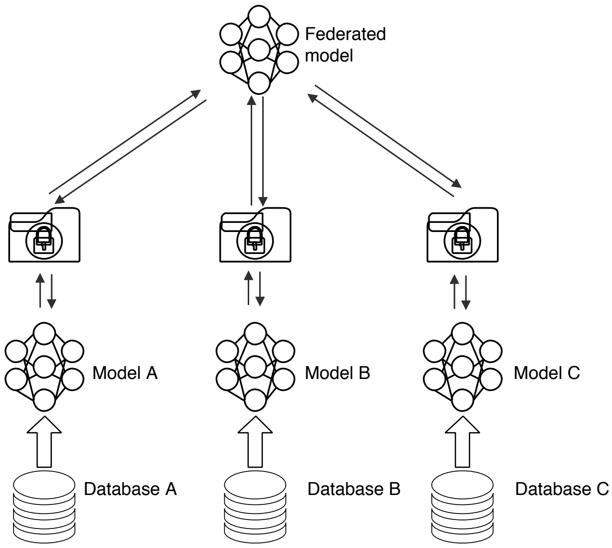

Figure 2: An Edge Computing AI Testbed with a Federated Learning Framework

References

[1] Peter M, Timothy G.: "The NIST definition of cloud computing, recommendations of the National Institute of Standards and Technology[J]," In National Institute of Standards and Technology (NIST) Special Publication 800-145, Tech. Rep. (2011).

[2] IoT Analytics, `https://iot-analytics.com/state-of-the-iot-update-q1-q2-2018-number-of-iot-devices-now-7b/`.

[3] Gartner Says the Internet of Things Will Transform the Data Centre, `https://prwire.com.au/pr/42679/gartner-says-the-internet-of-things-will-transform-the-data-centre`.

[4] MLPerf, `https://mlperf.org/`.

[5] Vijay J R.: "An ML Benchmark Suite for ML Software Frameworks and ML Hardware Accelerators in ML Cloud and Edge Computing Platforms," Report in BenchCouncil International Symposium on Benchmarking, Measuring and Optimizing, 2018.

[6] EEMBC, `https://www.eembc.org/`.

[7] Ignatov A, Timofte R, Chou W, Wang K, Wu M, Hartley T, Van Gool L.: Ai benchmark: Running deep neural networks on android smartphones. In Proceedings of the European Conference on Computer Vision (ECCV) 2018 (pp. 0-0).

[8] Gao W, Tang F, Wang L, Zhan J, Lan C, Luo C, Huang Y, Zheng C, Dai J, Cao Z, Zheng D, Tang H, Zhan K, Wang B, Kong D, Wu T, Yu M, Tan C, Li H, Tian X, Li Y, Lu G, Shao J, Wang Z, Wang X, and Ye H.: AIBench: An Industry Standard Internet Service AI Benchmark Suite. Technical Report 2019.

[9] Gao W, Luo C, Wang L, Xiong X, Chen J, Hao T, Jiang Z, Fan F, Du M, Huang Y, Zhang F, Wen X, Zheng C, He X, Dai J, Ye H, Cao Z, Jia Z, Zhan K, Tang H, Zheng D, Xie B, Li W, Wang X, Zhan J.: AIBench: Towards Scalable and Comprehensive Datacenter AI Benchmarking. In BenchCouncil International Symposium on Benchmarking, Measuring and Optimizing (Bench18), 2018.

[10] Jiang Z, Gao W, Wang L, Xiong X, Zhang Y, Wen X, Luo C, Ye H, Lu X, Xu W, Zhang Y, Feng S, Li K, Zhan J.: HPC AI500: A Benchmark Suite for HPC AI Systems. In BenchCouncil International Symposium on Benchmarking, Measuring and Optimizing (Bench18), 2018.

[11] Luo C, Zhan J, Zhang F, Huang C, Xiong X, Chen J, Wang L, Gao W, Ye H.: AIoT Bench: Towards Comprehensive Benchmarking Mobile and Embedded device Intelligence. In BenchCouncil International Symposium on Benchmarking, Measuring and Optimizing (Bench18), 2018.

[12] Wang L, Zhan J, Luo C, Zhu Y, Yang Q, He Y, Gao W, Jia Z, Shi Y, Zhang S, Zheng C.: Bigdatabench: A big data benchmark suite from internet services. In2014 IEEE 20th International Symposium on High Performance Computer Architecture (HPCA) 2014 Feb 15 (pp. 488-499). IEEE.

[13] Jia Z, Wang L, Zhan J, Zhang L, Luo C.: Characterizing data analysis workloads in data centers. In 2013 IEEE International Symposium on Workload Characterization (IISWC) 2013 Sep 22 (pp. 66-76). IEEE.

[14] Das A, Patterson S, Wittie M.: EdgeBench: Benchmarking Edge Computing Platforms. In 2018 IEEE/ACM International Conference on Utility and Cloud Computing Companion (UCC Companion) 2018 Dec 17 (pp. 175-180). IEEE.

[15] Johnson AEW, Pollard TJ, Shen L, Lehman L, Feng M, Ghassemi M, Moody B, Szolovits P, Celi LA, and Mark RG.: MIMIC-III, a freely accessible critical care database. Scientific data 3 (2016): 160035.

[16] Choi E, Bahadori MT, Sun J, Kulas J, Schuetz A, Stewart W.: Retain: An interpretable predictive model for healthcare using reverse time attention mechanism. In Advances in Neural Information Processing Systems 2016 (pp. 3504-3512).

[17] Liu L, Shen J, Zhang M, Wang Z, Tang J.: Learning the joint representation of heterogeneous temporal events for clinical endpoint prediction. In Thirty-Second AAAI Conference on Artificial Intelligence 2018 Apr 25.

[18] Amodei D, Ananthanarayanan S, Anubhai R, Bai J, Battenberg E, Case C, Casper J, Catanzaro B, Cheng Q, Chen G, Chen J. Deep speech 2: End-to-end speech recognition in english and mandarin. In International conference on machine learning 2016 Jun 11 (pp. 173-182).

[19] Panayotov V, Chen G, Povey D, Khudanpur S. Librispeech: an ASR corpus based on public domain audio books. In 2015 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP) 2015 Apr 19 (pp. 5206-5210). IEEE.

[20] Schroff F, Kalenichenko D, Philbin J. Facenet: A unified embedding for face recognition and clustering. In Proceedings of the IEEE conference on computer vision and pattern recognition 2015 (pp. 815-823).

[21] Huang GB, Mattar M, Berg T, Learned-Miller E. Labeled faces in the wild: A database forstudying face recognition in unconstrained environments. In Workshop on faces in'Real-Life'Images: detection, alignment, and recognition 2008 Oct.

[22] Yang Q, Liu Y, Chen T, Tong Y.: Federated Machine Learning: Concept and Applications. ACM Transactions on Intelligent Systems and Technology (TIST). 2019 Jan 28;10(2):12.