ISSN: 2664-9640

Research and Development of Edge Intelligence

Bin Liu

[1]Artificial Intelligence Key Laboratory of Sichuan Province, Zigong 643000, China;

[2]School of Automation & Information Engineering, Sichuan University of Science & Engineering, Zigong 643000, China.

* Corresponding Author

Abstract

In recent years, the topic of the combination of edge computing and artificial intelligence has been emerging, including concepts and applications that are constantly being refreshed. With the advent of the Internet of Everything era will produce unimaginable huge data and intelligent computing needs, which will make the network of edge device data will explode, the bandwidth requirements of data transmission increased, and this also put forward higher requirements for the real-time processing of data devices. For the traditional cloud computing framework, because of the deployment on the cloud far from the terminal, there are high latency, low bandwidth, privacy disclosure and other issues, can not meet the new technology low latency, high bandwidth, privacy protection and other needs. Edge computing is a natural way to run the data processed by edge devices close to the data source, which can reduce system latency and bandwidth, and most importantly, protect user data privacy and security to a great extent, and reduce the pressure of cloud computing center. Therefore, relying on the integration of artificial intelligence and edge computing and other technologies constitutes a new architecture of intelligent edge. In addition, blockchain as the underlying support technology of Bitcoin, because of security, information is not easy to tamper with and so on has been more and more attention. Although edge intelligence is in the period of technological reserve and development, it provides the direction and impetus for the development of science and technology in the future.

Keywords

Edge computing, artificial intelligence,edge intelligence,blockchain,deep learning.

1. Introduction

There is no clear definition of the concept of edge computing, and there are several mainstream views: First, Alex Reznik, director of etsi MEC ISG Standards Committee, gives a broad definition that anything that is not a traditional data center can be someone's edge node.Second, The State of the Edge report defines a server that is limited to a network close to the last mile. Third, in Edge Computing: A Primer, edge computing is defined as any computing and network resource between a data source and a cloud data center[1].Seeing that people have different views on edge computing in different industries and technology backgrounds, however in general edge computing is a paradigm of distributed computing, just as cloud computing is also a paradigm of distributed computing, which can complement cloud computing and is a "power-saving bridge" between end devices and the cloud.

Big data was moved up to a national strategy in 2015, and after several years of rapid growth, considerable progress has been made and it is foreseeable that the volume of data will continue to increase. According to International Data Corporation, nearly 90% of the world's data will be generated in the coming years, and the amount of data is expected to reach 163ZB (1ZB = 10

ISSN: 2664-9640

terabytes) by 2025, of which 75% will come from edge and end devices, while 50% will need to be processed at the edge, making the application of edge computing even more urgent[1]. Besides, with the continuous heat of the topic of deep learning and the application of AR, the primary condition for the processing of relevant edge data is real-time. The traditional cloud computing model is characterized by data being processed after remote transmission, there are transmission delays and network fluctuations, which can not meet the real-time needs of edge applications.

2. The Development and Present Situation of Edge Intelligence

Before we talk about edge intelligence, let's talk about the development of artificial intelligence, which has experienced more than 60 years of development from concept to popularization of related applications, but whose resume shines brightly. The concept of artificial intelligence was born in 1956 at a conference in Dartmouth, USA, and its enlightenment began at this time; By 1973, after nearly two decades of lack of direction in research into artificial intelligence, Britain and the United States stopped funding AI; It wasn't until 1997 that Dark Blue (American Chess Computer) defeated the then chess world champion, and during the same period, Brian Noble scientists demonstrated how mobile technology could be used to achieve speech recognition in edge computing, and people saw the "dawn" of artificial intelligence; In 2006, Hintion scientists proposed a deep confidence network, and in 2012 ImageNet, its team beat the rest of the team, setting off a wave of deep learning, and in the same year, Cisco will upgrade the IoT's scalable distributed computing facilities, and the development of artificial intelligence is becoming more and more "in full bloom."

The development course of edge intelligence is to empower the edge side on the basis of using artificial intelligence technology in recent years, and then push the development of artificial intelligence to a higher appoint, which is also a new form of application and expression of artificial intelligence technology. The definition of edge intelligence can be summed up in this way: refers to the application mode of artificial intelligence algorithms, technologies, and products near the edge of the data occurrence. The interweaving of edge computing and artificial intelligence has made the birth of edge intelligence. Edge Intelligence can cover many new applications, such as the technical requirements of real-time video analytics in Figure 1, which are critical for areas such as Autonomous Driving[4], Smart Factory, Smart City and Smart Medicine [8.

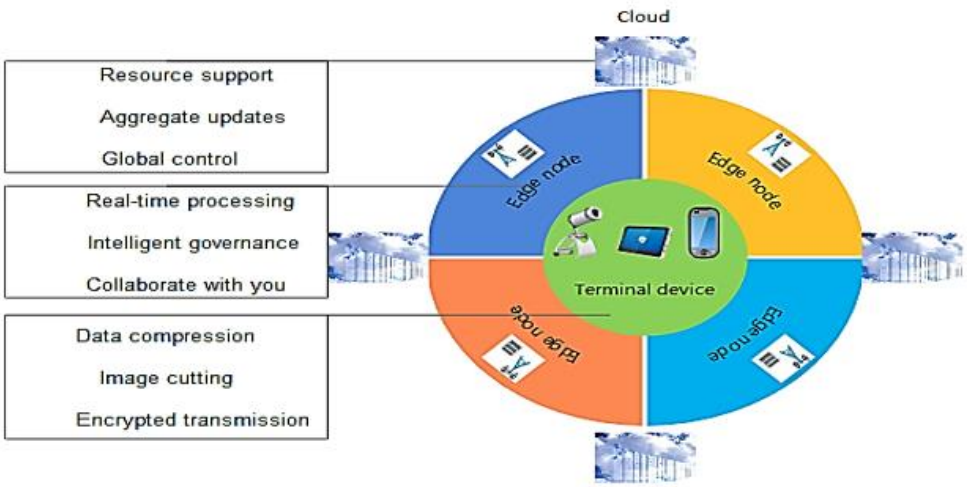

Figure 1 Real-time video analytics based on edge intelligence

Professor Shi Yusong was one of the early proposers of the concept of edge computing in China. In 2015, a network company represented by Cisco conducted a rapid technical study of edge computing, and in September of the same year, in mobile edge computing: a key technology leading to 5G published by the European Telecommunications Standards Association, a prototype of edge intelligence was proposed; In 2016, IEEE/ACM Symposium on Edge Computing was established, and in 2017, new opportunities for smart cloud and smart edge computing were presented at the Microsoft Conference; In 2018, the IoT Connectivity Alliance alliance, a coalition of IoT industry chain partners, released the "Edge Intelligence White Paper", and soon after, the ECC held a Fringe Computing Industry Summit in Beijing on the theme of "Edge Intelligence, Edge Cloud Synergy"; In 2019, the China-Tech Tech Edge Intelligence Conference was also announced in Beijing, focusing on the development trend and commercial application of popular technologies such as Edge Intelligence, AIoT, 5G, etc. At present, the research direction of edge intelligence at home and abroad is mainly in task offload and resource allocation[9,10], content caching[11,12], model training, privacy protection[14,15], etc. We can see that the promise of edge intelligence is being realized step by step, and we believe that as technology is constantly optimized, more areas will move toward edge intelligence in the future.

3. Contradictions and Challenges Faced by Edge Intelligence

Academics and industry pay more and more attention to the development of edge intelligence, and make detailed research and combing on the core technologies and application scenarios such as network architecture framework, task offload and allocation, network cache, distributed training, end-user privacy, etc. of edge intelligence, but the overall research of edge intelligence is still in the stage of technology reserve and development. However, with the upgrading of the country's strategic direction and domestic and foreign related Internet companies, communication equipment manufacturers, chip manufacturers and other departments to cooperate with each other to jointly carry out edge computing and edge intelligence research layout, will certainly break the bottleneck, usher in the spring of technological reform.

According to a new report from Grand View Research, the global edge computing market is expected to reach $15.4 billion by 2027. In industry, foreign search engine companies, represented by Google's edge network, chipmaker Intel's Edge Virtualization, Microsoft and other technology giants, as well as the world's leading cloud giants such as Huawei, Internet companies Alibaba and Tencent, have validated the intelligence perception and computing advantages of the new edge computing model. Edge intelligence will become the next field of the rise of the strong army, but also will quickly drive the development of such as intelligent manufacturing and other industries.

3.1. Task offload and resource management

In MEC systems, task offloading and resource management have been hot topics because they involve saving energy and reducing latency across the edge intelligence system. Task offloading is a process that includes transport, execution, and result backhaul. The computational offload scheme for the multi-user-multi-mobile edge server MEC system in literature improves the performance of traditional MEC systems by using multiple mobile edge servers for joint optimization (two-tier optimization), minimizing the overall energy consumption of the systems in the region. However, the neural network calculation and unloading scheme of adaptive differential algorithm is adopted for the change of user position. The literature studies the key problems of resource allocation and task scheduling in edge intelligence, and puts forward dynamic adaptive bandwidth allocation strategy based on intensive learning in view of the fairness of multi-user service quality in edge video transmission; In view of the cost

optimization of large-scale reasoning services in edge reasoning, the adaptive reasoning mechanism of multi-version model and multi-data model is proposed. The paper maximizes the system training task throughput rate under the resource limit of edge node and the completion time limit of the task, designs the RDJOA algorithm for model training task offload and resource allocation, and can improve the system training task throughput by 56% and the average edge node resource utilization rate by 53% compared to the benchmark algorithm.

3.2. Model training and deployment

As edge intelligence grows in size, the edge side will take on more and more intelligent services, including model training and deployment can be handled on the edge side. Because of the large number of devices on the edge and geographical dispersion, it has become a popular technology trend to train the distributed models by presenting a large amount of data with geo-distributed characteristics. Model training is also a key step in the business development based on technologies such as deep learning. The model training task is a resource-intensive process that involves high-intensity parameter calculation and updating, and after the structure of the model is built, one to more rounds of training on the data set are required. In order to cope with a wide variety of complex tasks, more and more complex models have appeared in the field of machine learning, but also the emergence of larger and larger data sets. Although these large-scale machine learning resource pressures can be alleviated to a certain extent by improving stand-alone processing power, many model training tasks are still far beyond acceptable time for traditional training on a stand-alone machine, making model training parallel to a natural evolutionary direction[16]. In many parallelization methods, the parallelization of training methods based on gradient descent methods has attracted a lot of attention and has been widely used. In a typical model training system based on distributed random gradient decline, there are several working nodes responsible for the training of the model, and the training results of these nodes are combined in some way to form the final trained model.

3.3. Move edge network cache

Mobile Edge Network Cache is considered to be an extremely effective technique for balancing backhaul link loads, reducing network latency, and improving user experience. Academics and industry have also done a lot of research and exploration in the field of cache system algorithms, and accurate prediction of the future prevalence of cached content files during the deployment phase of cached content files is an important breakthrough to further improve the performance of mobile edge caching systems[19]. But most of the previous research work was limited to the popularity of content files themselves. In order to deal with the shortcomings of the current research work on mobile edge network caching, and inspired by the new ideas of data analysis brought about by deep learning.

3.4. Privacy protection

Because a large number of devices are often required to work together in an edge computing platform, there is inevitably a large amount of data transmission between different devices, which not only causes a huge traffic load on the communication network, but also makes the data lack privacy protection. To address data privacy issues, a framework of cloud-side collaborative computing that meets data privacy, security, and performance requirements is required. Therefore, federal learning is further proposed, which enables the artificial intelligence model to be distributed training in the edge computing architecture, and does not need to upload sample data, only after the training parameters are updated and uploaded, and then by the edge node aggregation parameters update and parameters are sent. The main areas of research currently under federal study include optimizing communication mechanisms, data privacy security, business scenario applications, and so on [22].

3.5. Blockchain and edge intelligence

In the scenario of edge intelligent computing, there are more device-to-device collaborations, and multiple devices running AI algorithms are scattered across the edge network, which require frequent communication between these devices in order to collaborate on AI computing tasks or joint intelligent group decisions, which poses a great threat to network security and privacy. With the popularity of the Bitcoin concept, as the underlying support technology of Bitcoin can be built in a secure and verifiable way in a pan-central environment, blockchain has attracted more and more attention due to its security and information is not easy to tamper with. In the blockchain network of any node can be transaction verification and forwarding, nodes together maintain the transaction containing the block ordered link ledger, thus building an undone, network-wide consistent state record. It is therefore widely used in the financial economy, the Internet of Things, big data, cloud computing and edge computing.

4. Edge Smart Scenarios

4.1. Internet of vehicle

At present, the use of vehicles and people's lives are closely related, but the traffic accident rate is still high, the emergence of car networking will improve the reliability of traffic safety and traffic efficiency. In the Internet of vehicle scenario, the intelligent vehicle is limited in the body perception range is not wide and the computing resources are scarce, to achieve extremely safe mode of autonomous driving is an arduous task. Therefore, in response to the above-mentioned problems,the edge intelligence of multi-source data processing to deal with the car network autonomous driving. The former consists mainly of on-board sensors for data collection, while the latter is the use of V2V technology between vehicles for direct communication without passing through the road-side unit. In order to improve the reliability of the autonomous driving system, the author uses the federal learning method described above to improve the model performance, that is, the vehicle trains a local model on the local data set and uploads it to the server for global model aggregation; Or by the data distribution strategy, the vehicle uses the massive multi-source data in the network to verify and repair the locally perceived data and improve the accuracy of neural network inference. The advantages of the system are as follows: In the first place, by exploring the idle resources of nodes in the network, improve the throughput of the system. In another, use the distributed data in the network to improve the accuracy of neural network inference. What's more, the original vehicle nodes are mapped to a virtual node collection, the data transmission of the nodes is decoupled, so that the original problem is transformed into the minimum cost of the maximum flow problem to solve, to achieve the autonomous driving network computing, communication resources efficient allocation.

4.2. Smart home

According to Machina Research data projections, the Internet of Things will generate more than 2 bytes of data traffic by 2025, mainly from consumer electronics devices. With the improvement of people's living standard, the comfort and energy needs of living conditions are more urgent, as a feasible solution to this problem, smart home emerged. Smart home is based on family home, security monitoring technology, network information technology, multimedia technology as the technical basis to enhance the home environment safety, convenience, comfort as the goal of smart buildings. First, the data processing part of smart home: is the use of distributed edge computing processing, for real-time business, data screening, network bandwidth and heterogeneous system interconnection and security confidentiality to provide a good solution; Second, smart home data transactions and security parts: the use of blockchain technology and IPFS, storage to encrypt the data, to further ensure data security and increase

ISSN: 2664-9640

the space for local storage; Third, smart home system through the blockchain sharing platform to publish data. By applying blockchain to smart home data sharing system, the reliability of information can be effectively guaranteed, and due to the decentralization of blockchain, poor quality and false data are exposed on the chain, which helps to create a good smart home data sharing environment.

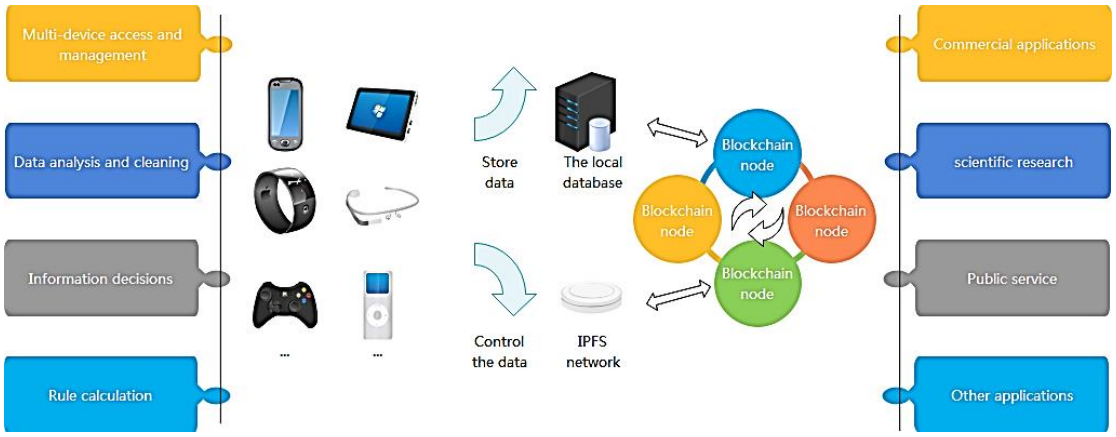

Figure 2 Smart home model that blends blockchain and edge computing

4.3. Smart manufacturing

With the continuous upgrading of the industrial industry, the increase of hardware and software equipment facilities and sensors in industry is very challenging to dispatch and distribute, and the high real-time and certainty of the network is an increasing number demanding. The use of edge intelligence can enhance the realization of intelligent manufacturing, fully complete the industrial hardware and software equipment facilities and sensor equipment between the "things" interconnection operation, to achieve all aspects of industry: intelligent production, equipment, products, logistics digital information, so that all equipment can be quickly and flexibly replaced, rapid adjustment and change of production plans, to promote the rapid deployment of new processes and new processes.

4.4. Smart medicine

Edge intelligence can also be integrated into the medical field, in this new coronary pneumonia outbreak, edge intelligence to achieve medical health information such as health code, medical equipment, public health security intelligent supervision and management, greatly improve the solution of medical service level, safety and production hidden dangers. The terminal layer includes 5G data terminals and smart devices, which provide networking functions in various forms according to the type of intelligent medical device, and the network layer can include 5G, 4G, optical fiber, etc. The edge layer includes the MEC platform and user face function (UPF) shunt gateway to realize the functions of local shunt, ultra-low latency network, network capability open near-end computing, and the cloud platform layer includes the central cloud platform such as the medical blockchain platform and the 5G smart medical application platform, which is deployed in the client room or the carrier Internet data center (IDC) room, and the business layer is mainly to realize the hospital's medical education, medical treatment, medical research and other intelligent medical applications. Based on the medical blockchain platform, the valuable data of the hospital can realize the safe sharing of medical data through the chain storage certificate on the 5G network.

5. Conclusion

This paper introduces the fusion of cutting-edge technologies such as artificial intelligence and edge computing, and discusses the key issues such as network security protection, mobile wireless network computing and communication resource management. In view of the progress of applied research on edge intelligence, this paper expounds the application situation of edge intelligence in data security, privacy protection, authentication, access control, computing migration and network management, such as mobile edge computing scenarios such as car networking, smart home, smart manufacturing, smart medical and so on. The application of edge intelligence in these fields can effectively improve the efficiency and level of public security management, greatly reduce the cost of human and material resources, has great value to urban management and people's livelihood improvement, has broad market space, and the basic conditions for technology application are being solidified. Computing power will become the key to the development of edge computing and edge intelligence, and the characteristics of real-time and locality of edge computing determine that local computing power plays a vital role in the era of edge intelligence. Whether the intelligence of the edge side can be realized depends on whether the algorithm model on the edge side is applicable or not, and on the other hand, whether the edge side can provide sufficient acuity guarantee.

References

[1] Jie Cao,Quan Zhang,Weisong Shi.EdgeComputing:APrimer[M].Springer,Cham:2018-01-01.

[2] The Edge Computing Industry Alliance (ECC) and the Industrial Internet Industry Alliance (AII). The Edge Computing and Cloud Computing Collaboration White Paper(2018) [R/OL].

[3] Alibaba Cloud Computing & China Institute of Electronic Technology Standardizati on.Edge cloud computing technology and standardized white in 2018 Book of Books (M/OL). (2018-12-12) [2018-12-14].

[4] Shi W, Cao J, Zhang Q, et al. Edge computing: Vision and challenges[J]. IEEE internet of things journal, 2016, 3(5): 637-646.

[5] ECC. Edge computing consortium[EB/OL]. [2018-11-03].

[6] Shi Yusong, Sun Hui, Cao Jie, etc. Edge Computing: New Computing Models in the Age of the Internet of Everything [J] Computer Research and Development, 2017, 54 (5): 907.

[7] Sapienza M, Guardo E, Cavallo M, et al. Solving critical events through mobile edge computing:An approach for smart cities[C]//2016 IEEE International Conference on Smart Computing (SMARTCOMP). IEEE, 2016: 1-5.

[8] Salahuddin M A, Al-Fuqaha A, Guizani M, et al. Softwarization of internet of things infrastructure for secure and smart healthcare[J]. arXiv preprint arXiv:1805.11011, 2018.

[9] Zhang Menglin. Study on user task offloading and resource allocation in Edge Intelligence [D]. Hunan Normal University, 2020.

[10] Jiang Yu. Research on resource allocation and task scheduling for edge intelligence[D]. Jilin University,2020.

[11] Venugopal S, Gazzetti M, Gkoufas Y, et al. Shadow puppets: Cloud-level accurate {AI} inference at the speed and economy of edge[C]//{USENIX} Workshop on Hot Topics in Edge Computing (HotEdge 18). 2018.

[12] Song Xuming, Shen Yifei, Shi Yuanming. Intelligent Mobile Edge Network Cache Based on Deep Learning [J] Journal of the University of the Chinese Academy of Sciences, 2020, 37 (01): 128-135.

[13] Wang Haichuan. Model Training Service Deployment and Task Offloading Research for Edge Intelligence[D]. China University of Science and Technology, 2020.

[14] Liu L, Zhang X, Qiao M, et al. SafeShareRide: Edge-based attack detection in ridesharing services[C] //Proc of the 3rd IEEE/ACM Symp on Edge Computing (SEC). Piscataway, NJ: IEEE, 2018: 17-29

[15] Roman R, Lopez J, Mambo M. Mobile edge computing, fog et al.: A survey and analysis of security threats and challenges[J]. Future Generation Computer Systems, 2018, 78: 680-698.

[16] Qingshan Liu,Jun Wang,Zhigang Zeng. Advances in Neural Networks, Intelligent Control and Information Processing[J]. Neurocomputing,2016,198.

[17] Liu Wei, Yang Chenyang, Soshiqiang, Huang Yuanfang. Edge Intelligence in Wireless Communications[J]. Signal Processing, 2020, 36 (11): 1789-1803.

[18] Yu Jiang, Qiu Ling. Cache content decisions and user attribution based on collaborative filtering under dense small station networks[J]. Journal of the University of the Chinese Academy of Sciences, 2016, 33 (06): 802-807.

[19] Azimi S M, Simeone O, Sengupta A, et al. Online edge caching and wireless delivery in fog-aided networks with dynamic content popularity[J]. IEEE Journal on Selected Areas in Communications, 2018, 36(6): 1189-1202.

[20] Somuyiwa S O, György A, Gündüz D. A reinforcement-learning approach to proactive caching in wireless networks[J]. IEEE Journal on Selected Areas in Communications, 2018, 36(6): 1331-1344.

[21] Li Dongdong. Study on edge caching strategy based on energy consumption in mobile edge networks[D]. Nanjing University of Posts and Telecommunications, 2020.

[22] Li T, Sahu A K, Talwalkar A, et al. Federated learning: Challenges, methods, and future directions[J]. IEEE Signal Processing Magazine, 2020, 37(3): 50-60.

[23] Bitcoin:a peer-to-peer electronic cash system [EB/OL].[2019-11-16]. https:// org/en/bitcoin-paper.