

APPLICATION OF AI TO MOBILE NETWORK OPERATION

Tomoyuki Otani¹, Hideki Toubé², Tatsuya Kimura¹, Masanori Furutani¹
¹DOCOMO Technology INC., ²NTT DOCOMO INC., Japan

Abstract – With the introduction of network virtualization and the implementation of 5G/IoT, mobile networks will offer more diversified services and be more complex. This raises a concern about a significant rise in network operation workload. Meanwhile, artificial intelligence (AI) technology is making remarkable progress and is expected to solve human resource shortages in various fields. Likewise, the mobile industry is gaining momentum toward the application of AI to network operation to improve the efficiency of mobile network operation [1][2]. This paper will discuss the possibility of applying AI technology to network operation and presents some use cases to show good prospects for AI-driven network operation.

Keywords – Artificial intelligence (AI), mobile network, network operation

1. INTRODUCTION

1.1. Characteristics of artificial intelligence (AI)

Businesses today are faced with a great deal of information and a shortage of human resources. To solve these issues, they are introducing AI/big data technologies at an accelerated pace to improve the efficiency of business operation.

As the world today is experiencing the third AI boom, the practical use of "machine learning" helps us to automatically identify and learn the patterns and rules from large amounts of data (the so-called "big data") based on specific criteria. This technology will enable us to set more appropriate rules based on learning from data and make more accurate judgments once evaluation criteria are given. Following this, "deep learning" has emerged, in which AI learns by itself and accumulates the knowledge of patterns and rules with no specific criteria given.

According to the 2016 White Paper on Information and Communications in Japan published by the Ministry of Internal Affairs and Communications, there are three major types of functions that AI plays in actual services: "identification", "prediction" and "execution" [3]. These functions can be utilized and applied across all industries. Standard usages of each function are as follows:

- "Identify" the current situation (characteristics) from a large amount of data (big data).

- Analyze the time characteristics of the data and "predict" future tendencies.
- Make and "execute" an optimum plan which is based on the "identified"/"predicted" data.

As shown in the white paper, AI is expected to be utilized in various business fields and perform highly advanced analysis (for improving operational efficiency), using big data in a short period of time without human manual operation

1.2. Trends of mobile network

Today, the application of virtualized network functions (NFV) to mobile core networks is in progress [4]. For years, various network functions, such as the conventional Evolved Packet Core (EPC), have been provided in their dedicated hardware (HW) such as Advanced Telecom Computing Architecture (ATCA) hardware. With the introduction of NFV, software will be able to run on a virtualized operating system (OS) of generic Intel architecture (IA) servers and be provided separately from hardware [3]. Furthermore, the NFV architecture enables integrated management and control (orchestration) of network services and resources, interworking with Management and Orchestration (MANO) and a software-defined network (SDN).

These technologies will enable most of the construction and increase/decrease work of network elements (NEs) to be performed remotely by software control without manual operation. In

addition, once software and hardware are separated, they can be constructed and increased or reduced independently at different timings. For instance, once the hardware, e.g. the IA server, is prepared, software can be readily increased/reduced on the server. Resources can be allocated whenever they are needed to accommodate the traffic, reducing the surplus network facilities (Fig. 1.1). In the conventional method, network facilities (blue line) are established in advance based on the specific network traffic volume predicted (red line). In contrast, in the virtualized network, lots of equipment units can be flexibly arranged as virtual machines (VMs) on the servers/cloud and such VMs can be allocated dynamically to allow dynamic changes in the network resources allocated.

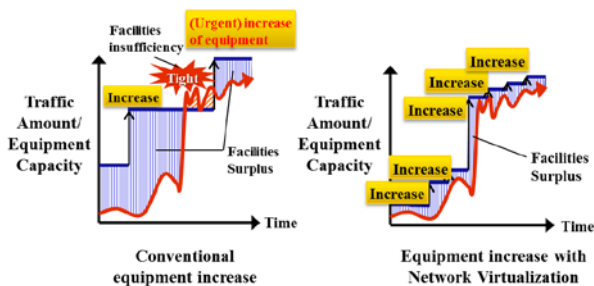


Fig. 1.1. Network resource allocation with network virtualization

Moreover, the introduction of 5G scheduled for 2020 will bring with it a variety of network services (such as IoT) with traffic characteristics and network requirements different from those of existing smartphones [5].

For example, a factory equipment monitoring system, in which mass data from sensors are sent to the administrator periodically, will require a highly secure and reliable network (no data loss). On the other hand, a traffic system, in which inter-vehicle communication prevents traffic congestion, will require a low-latency and highly secure network. Although the networks with different requirements need to be separated, it is inefficient to do so physically. Instead, we need to utilize the network slicing technology to virtually construct separate networks with different requirements on the same physical network.

With new technologies such as NFV and the network slicing technology mentioned above, the physical hardware configuration will remain unchanged or become simpler. However, this will bring about more issues for manual network operation, as the logical network configuration used for service provision will become complicated due to the use of multiple virtualized logical resources.

2. ISSUES OF MOBILE NETWORK OPERATION

Network operation in general can be illustrated as a cycle of activities that consist of planning, construction and maintenance, as shown in Fig. 2.1.



Fig. 2.1. Network operation cycle

In light of conventional mobile network trends, whether to achieve the following objectives can have an impact on the network operation:

- 1) accelerate the network operation cycle to provide services quickly; and
- 2) establish an analysis and operation method for advanced and complicated networks.

The issues towards achieving the above objectives are described below.

Planning is the step to analyze the network traffic and formulate future plans for network facility resources. The main activity of the conventional planning process is the analytic work comparable "prediction," which is performed to analyze the past network traffic and calculate the amounts of long-term (yearly) network resources (such as bandwidths). However, when network virtualization and various 5G/IoT services come into play and different services are superimposed on the same physical network/ hardware, it will be necessary to factor in the traffic of all services in traffic volume prediction for a single network/hardware unit. This will make the analysis very difficult and time consuming if it is done manually by humans or based on the expertise of specialists as before. Specific issues for accelerating the planning process are described in clause 2.1.

Generally speaking, the construction process involves designing, building, and testing the network equipment. When network virtualization is applied, it will be possible to increase or reduce the number of VMs used as NE equipment at any time if IA servers are in place as hardware. The network resources can be created by software operation only.

As construction can be completed simply by setting VM servers, the time for the construction will be shortened remarkably, accelerating the construction process.

The maintenance process is the step taken to analyze the impact of problems on customer services based on alarm information from the network, recover services in a way appropriate to the details of impact, and identify the faulty equipment and fix the problem using the alarm information. With the progress of NFV and network slicing technology, the network configuration for customer service provision is divided into two: the logical network configuration and the hardware configuration that configures the logical network. These network configurations will make it difficult to quickly and accurately analyze/identify the impact on customer services at the time of hardware failure.

The application of network virtualization to the construction process will accelerate the cycle of network operation. On the other hand, the planning and maintenance processes are facing some issues, which will be discussed in more specific terms in the following subclause.

2.1. Issues of planning process

For conventional traffic prediction, it has been common to predict the volume of traffic based on the time-series analysis of the measured traffic volume. However, predicting the traffic volume accurately through the time-series analysis is becoming difficult due to a dynamic variation of factors contributing to the generation of traffic, such as the emergence of new applications/contents and temporary concentration of user population associated with events.

For example, traffic prediction based on time-series analysis cannot cope with such disturbing factors as temporary traffic surges associated with events. Therefore, we need traffic tendency analysis that has eliminated those disturbing factors and other anomalous aspects responsible for the significant deviation between prediction and reality. In addition, we also need to analyze short-term traffic tendencies such as temporary traffic spikes in order to realize the network resource allocation in line with the volume of communication traffic as shown in Fig.1.1. However, accurate analysis as mentioned above requires not only an enormous analysis workload but also different sets of special skills for developing long/short-term models. Furthermore, an assumption of various services such as those of 5G/IoT presents the limitation of human analysis regarding anomalous traffic tendencies for all services.

2.2. Issues of maintenance process

The conventional network has a fixed mapping relation between the logical network configuration for service provision and the hardware that configures the logical network. When hardware fails, this fixed relation has allowed maintenance staff to analyze/identify the affected services by looking at the service topology information consisting of hardware and network configuration information as well as the equipment alarm information sent from the hardware.

With the progress of NFV and network slicing technology, the logical network configuration for service provision will include multiple virtualized logical resources. There will be a various way of service provision. In one case, the same service will be configured on different hardware every day. In another case, the same hardware may provide a different service every day. As a result, it will be difficult to achieve quick and accurate analysis with the conventional method, which analyzes/identifies the impact on customer services manually by humans based on the alarms from various equipment units.

3. NETWORK OPERATION WITH AI

The application of AI will enable us to respond to the above-mentioned problems in the planning and maintenance processes quickly and efficiently even when sufficient human resources, experience and special skills are not available. So, we are aiming to apply AI to the planning/maintenance processes in order to conduct more efficient and advanced analysis work for the planning/ maintenance tasks. This clause qualitatively explains how the application of AI will make network operation effective.

3.1. Approach to applying AI to planning process

This section explains how AI is applied to traffic demand prediction during the planning process.

We use AI to predict and analyze traffic demand. Basically, traffic tendencies can be divided into two types: short-term tendencies, such as temporary traffic increases during events; and long-term tendencies, from which anomalous tendencies such as temporary traffic increases during events have been removed. In AI-driven traffic prediction, we make AI learn the short-/long-term traffic tendencies that have respectively totally different factors causing traffic fluctuations and mechanisms in order

to predict both short-/long-term traffic demands. An example of AI-driven traffic demand prediction is shown in Fig. 3.1.

- 1) Input traffic information and date and time information of events in each area as the data for AI learning.
- 2) Have AI learn long-term traffic increase tendencies excluding temporary traffic surges such as those during events (for generation of long-term traffic demand prediction model). Furthermore, have AI learn the correlation between the time of events and temporary traffic surges mentioned above to learn the occurrence tendencies (timing, increment) of short-term traffic increases (due to events, etc.) (for generation of temporary traffic demand prediction model).
- 3) Have AI output long-/short-term traffic demand prediction models for each area.

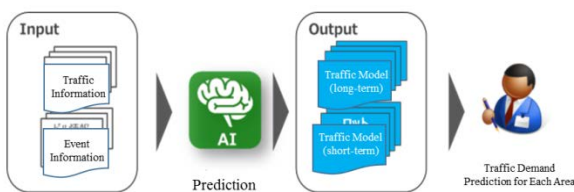


Fig. 3.1. Process of traffic demand prediction with AI

3.2. Approach to applying AI to maintenance process

This section explains the application of AI to network monitoring (network abnormality detection).

3.2.1. Application of AI to network monitoring

3.2.1.1. Necessity of service monitoring

As described in clause 2.2, the network maintenance process will become complicated. Above all, it will be much more difficult to analyze and grasp the impact on customer services accurately in the network monitoring process, in which speed is the key.

With the conventional method of identifying the impact on customer services from equipment alarm information, it is difficult to identify the impact when no alarm is given and information is insufficient, especially in the case of silent alarms.

In addition, as the conventional method often relies on the skill and expertise of maintenance staff, sometimes the impact on service cannot be reasonably grasped from the standpoint of customer experience.

Under such circumstances, it is necessary to realize a service monitoring system that is able to detect the quality close to what customers are actually experiencing, estimating it from the network data, independent of alarm information. Image of Service monitoring is shown in Fig. 3.2.

This can be achieved by collecting various data from the network, including equipment alarm information and integrating them into the big data for multi-angle analysis (AI analysis).

If service monitoring is realized, the maintenance process will be innovated from facility maintenance centered on device alarm to network quality maintenance based on customer experience. Taking the initial reaction amount, for example in the conventional method, it corresponds to the number of alarms notified from the network. While in the service monitoring method, it corresponds to the number of influences on customer service caused by the equipment failure, which is expected to be reduced drastically.

Regarding the workload of fixing the failure from the maintenance viewpoint, since the necessity of repairing the equipment is irrelevant to the service effect occurrence, there will not be a great difference in the maintenance workload. In the maintenance process, there is also a possibility to arrange the maintenance automatically by converting the device alarm information and work log into big data for machine learning.

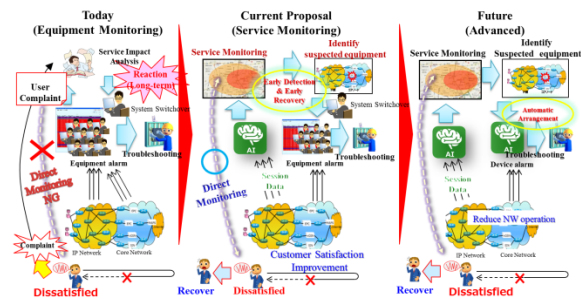


Fig. 3.2. Image of Service monitoring method

3.2.1.2. Application of AI to service monitoring

In order to realize the network service monitoring based on customer experience independent of equipment alarm information, a method of analysis that consists of the following two stages can be used:

- (A) Have AI analyze the network data and estimate the index which can express the quality of customer experience (QoE) for each service [6];
- (B) Collect the estimated QoE for each service provision area (cell/eNB area) and have AI

learn the feature value of distribution. Have AI detect any unusual state (deterioration of network quality) based on the feature value and make a judgement on the distribution state to decide whether it is "different from the usual state or not" [1], [7].

The illustration of the proposed method is shown in Fig. 3.3.

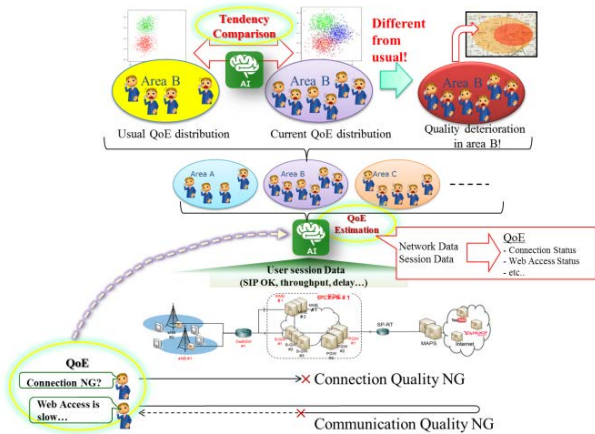


Fig. 3.3. Apply AI to service monitoring method

The quality of customer experience (QoE) estimation and service abnormality determination methods are described below.

(A) (QoE) estimation

QoE is an index of customers’ subjective feeling (comfortable, clean, etc.), which is difficult to measure directly on the network. Therefore, we use another quality index, for example an application quality of service (QoS) such as webpage loading time as an alternative for QoE. As it is also difficult to measure the application QoS on the network directly, we estimate the application QoS from the network QoS (TCP throughput, etc.), which can be measured on the network, and use the estimated application QoS as a substitute of QoE (Fig. 3.4).

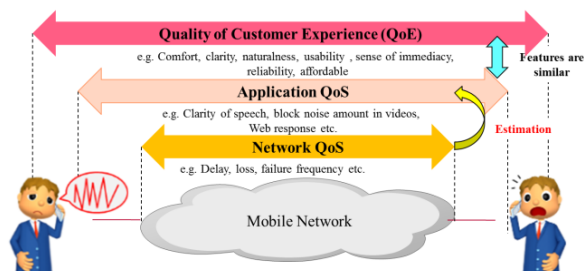


Fig. 3.4. Relation between QoS and QoE

The following steps explain the method of using AI analysis to estimate the application QoS from the network QoS obtainable from the network, using the loading time of web access as an example. (Hereinafter, QoE means application QoS.)

- (1) Measure and acquire information on the actual webpage loading time by the test terminal, etc.as training data.
- (2) Collect communication quality information at network level observable on the network (throughput, etc.) or equipment traffic information (statistical values such as the number of established connections for each equipment unit). Have AI learn the information’s correlation with the webpage loading time mentioned above in (1) and detect highly-correlated information from the web access session information and equipment traffic information. Then, have AI generate an analysis model (AI model) for calculating an estimated value of webpage loading time, using high-correlated information (in other words, information which can be acquired on the network side).
- (3) Using the generated AI model, estimate the webpage loading time from the communication quality information at a network level (throughput, etc.) or equipment traffic information.

The explanation above uses an example of communication quality in which QoE is difficult to measure. In contrast, when the connection quality is determined simply by whether the connection is OK (successful) or NG (unsuccessful), such network information (disconnection information) can be directly applied as the QoE index. In this case, AI analysis is not necessarily needed for estimating QoE.

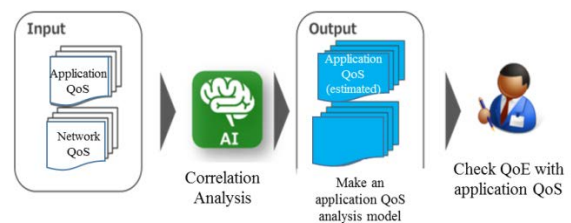


Fig. 3.5. Process of QoE estimation with AI

(B) QoE anomaly analysis

Using the following steps, we explain service quality abnormality analysis in which quality degradation caused by the network is detected based on the QoE index output in (A).

In a mobile network, where resource utilization changes greatly according to the mobility of user population, it is highly likely that QoE will also fluctuate by the time and place. Taking this into account, we are studying an analysis method which consists of the steps as described below.

- (1) Classify the collected QoE data into macroscopic observation units, which are in this case communication areas. Have the AI learn the distribution of feature amounts in the QoE data observed in each area in time series for both cases in which the network is in normal and abnormal states. Generate the AI analysis model with the QoE distribution (steady state) for each area when the network is normal. (In the distribution learning of QoE, there will be a large amount of analysis if every individual QoE value is used. Therefore, it will be more effective to learn and determine with the statistical values generated in advance, such as the maximum, minimum, average and dispersion of QoE for each area, which represents a macroscopically aggregated unit.)
- (2) With the AI analysis model, create a clustering of normal and abnormal states from the QoE information collected regularly. Determine the service quality as abnormal when detecting a deviation tendency from the steady state.

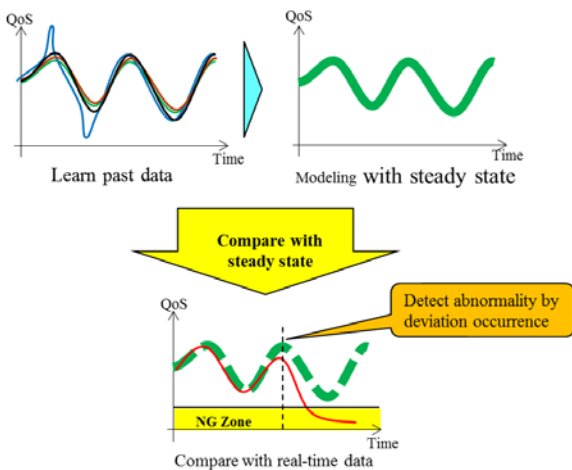


Fig. 3.6. Abnormality determination in comparison with steady state

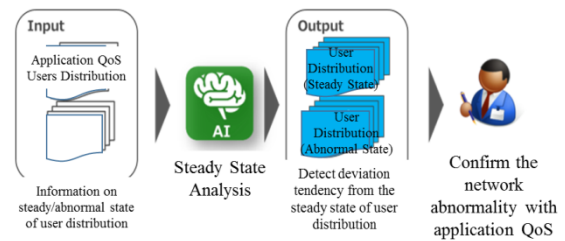


Fig. 3.7. Process of service quality abnormality determination with AI

Moreover, in this method, in which the presence of users at the edge of a service area is also taken into consideration; any case that deviates from the normal network state is considered to be under the influence of network abnormality (abnormality of the network itself or another network) and only service quality degradation caused by network abnormality can be detected.

4. CONCLUSION

In this article, we described the possibility of applying AI technology to mobile network operation and presented some use cases to show significant benefit from AI-driven network operation.

The study on the application of AI to mobile network operation in the telecommunications field is still in its infancy and there has been no report of a commercial network which has actually introduced AI to its operation. Based on the concept described in clause 3, we plan to verify technical aspects of AI application in a trial environment in which commercial network data is available and aim to realize network operation taking advantage of AI.

REFERENCES

[1] “Development of NTT Group AI technology “corevo™” for network operation, the AI technology able to autonomously derive the correlation between the “causes of faults and alarms” for highly accurate detection of network faults.

<http://www.ntt.co.jp/news2017/1701/170130a.html>

- [2] “KDDI R&D Laboratories completes World’s first successful AI-assisted automated network operation system PoC” ., News release on Feb. 22 2016.
<https://wirelesswire.jp/2016/02/50505/>
- [3] “Present and Future of Artificial Intelligence (AI),” White Paper on Information and Communications in Japan 2016, Chapter 4, Section 2.
- [4] Hiroyuki Oto, “Application of Virtualization Technology to Communications Networks,” NTT DOCOMO Technical Journal Vol. 18 No. 1 pp. 4-10.
- [5] Takuya Shimojo., “Future Core Network for the 5G Era,” NTT DOCOMO Technical Journal Vol. 17 No. 4 pp. 50-59.
- [6] Takanori Hayashi, “QoE-centric Operation for Optimizing User Quality of Experience.”
<https://www.ntt-review.jp/archive/ntttechnical.php?contents=ntr201509fa3.html>
- [7] “MasterScope: Features and Experimental Applications of System Invariant Analysis Technology” (NEC Technical Journal Vol. 65 No. 2/2012).