

# Improving Network Management and Operation with Machine Learning and Data Analytics

*Keisuke Ishibashi, Takanori Hayashi,  
and Kohei Shiimoto*

## Abstract

In this article, we describe the current and future direction of our research on the use of machine learning and data analytics to solve network management and operation problems. Specifically, we introduce ways to predict or detect network failures using social networking services or network logs, ways to extract workflows using operator logs, and methods to predict mobile traffic by investigating traffic generation factors.

*Keywords: network failure, self-operation, traffic analysis*

## 1. Introduction

We are conducting research on methods to solve network management and operation problems using machine learning and data analytics. Machine learning can extract latent rules and generation models of network behavior through big data analytics, and those rules and models can be used for predicting and optimizing network management and operation, which improves network service and reduces costs. In this article, we introduce methods to predict and detect network failures using a social networking service (SNS) or network logs, extract workflows from operator logs, and predict mobile traffic from traffic generation factors.

## 2. Predicting and detecting network failures

To minimize the impact on services caused by a network failure, we must detect a failure, or a fault that leads to a failure, before it occurs or as soon as possible after it occurs. Current failure detection methods are rule based, where rules between network logs and failure modes are set manually in advance. By trapping a log, we can detect a corresponding

network failure. However, much progress has been made recently in software-based network function virtualization, which means that the network configuration is dynamically changing. Consequently, building those rules and updating them are difficult and time-consuming tasks. In addition, because of the increasing network complexity and the growing number of network roles, the monitoring information (network logs) obtained using existing methods is insufficient for monitoring the network status. To solve this problem, we are working on enhancing network monitoring and improving the accuracy of failure detection.

### 2.1 Enhancing monitoring objects

Current network monitoring techniques are based on the use of internal network data such as network logs and external network data such as service monitoring and user claims. To improve the coverage and agility of monitoring, we are implementing failure detection based on SNS data. However, because SNS data consist of free-format text messages and include a huge number of messages other than those related to network outages, we have developed a technique to accurately extract messages that correspond to

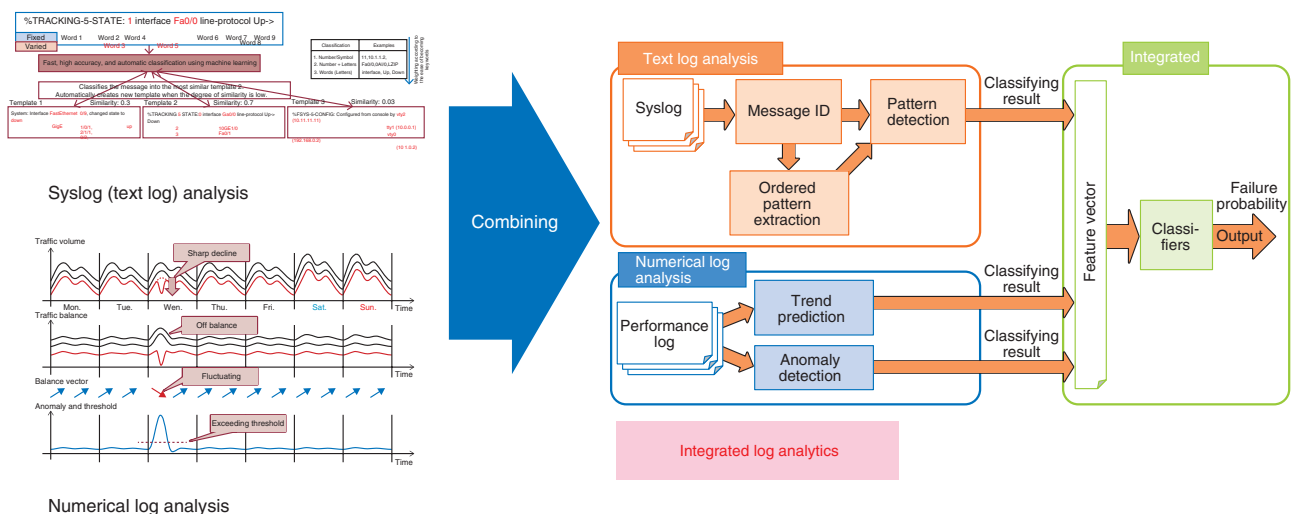


Fig. 1. Log analysis technique.

network failures. In addition, by estimating the locations of such messages, we can estimate the location and the impact of the network failure.

Network logs that are used for network monitoring are divided into numerical logs such as CPU (central processing unit) and interface loads, and messages in text format such as syslog messages. Numerical logs are used for detecting network failures by applying predefined thresholds. However, defining thresholds for a huge number of logs is difficult. We apply statistical outlier detection based on a non-supervised machine learning method. In addition, we apply this method to obtain the time series of each log as well as the correlations among logs.

Text logs are used for failure detection by monitoring keywords. However, this involves looking at the relations between keywords. The network status is not directly monitored, and therefore, the accuracy needs to be improved. We adopt a machine learning technique such as clustering to monitor text logs that are not only keyword based but also message based and that generate patterns. We also try to merge both numerical logs and text logs to detect network failures that are not detected for them individually. (Fig. 1).

## 2.2 Improving detection accuracy

Adopting a statistical method for detection as described in the previous section can potentially result in a *miss* detection (false positive) or detection error (false negative). In addition, the current supervised machine learning method requires both positive

(failure mode) and negative (normal mode) samples, but network failures seldom occur and positive samples are not easily obtained. Here, we focus on the partial area under ROC (receiver operating characteristic) curve (pAUC), which is the index for a balance between false positives and negatives, and we try to adopt a technique that directly optimizes pAUC [1].

## 3. Extracting operation workflow

When network operators resolve network failures, if the resolution processes are not fixed and no manuals are available, operators must take action based on their knowledge, which of course depends on their experience. This increases the time needed to resolve the failure (the time-to-fix), especially for non-skilled operators. This in turn increases the need for Runbook Automation (RBA), which enables auto-operation in the event of network failures. However, building a workflow (scenario) to be used for RBA is a time-consuming task.

To solve this problem, we take two approaches to extract and visualize workflows in resolving network failures.

First, we develop a technique to extract workflows using a trouble ticket log where operators manually record the processes they carry out from the time a failure starts to when the problem has been resolved. Within a trouble ticket, operators record resolution processes, which provide useful information for building workflows for the processes. However, these records consist of free-format text data, which means

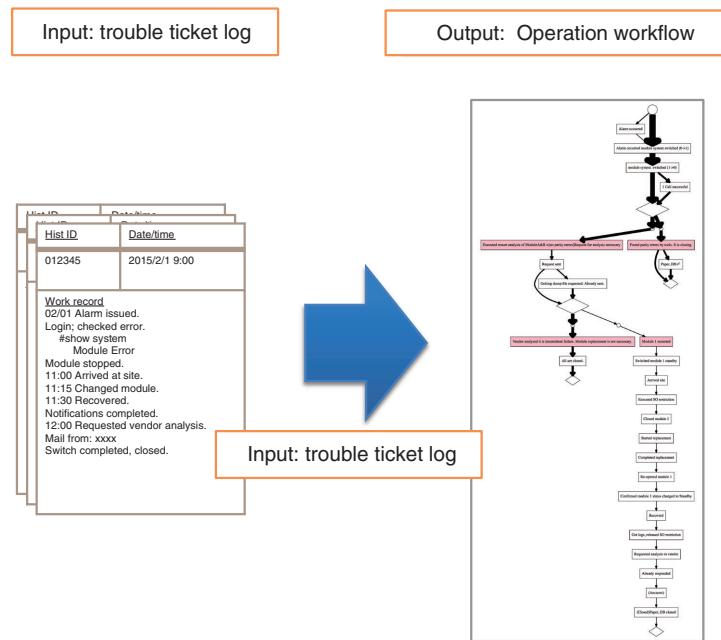


Fig. 2. Extracting operation workflow using trouble ticket log.

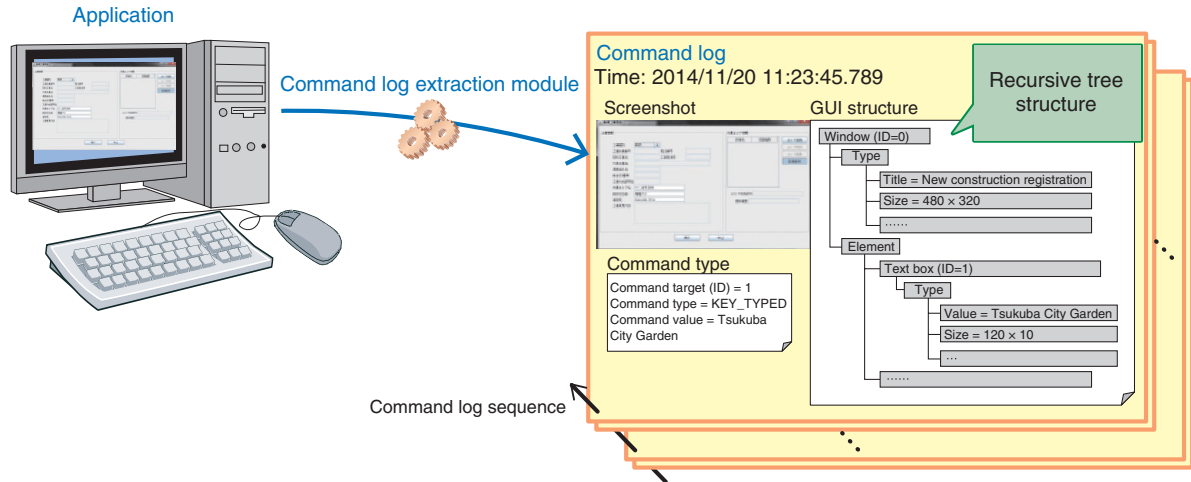


Fig. 3. Extracting operation workflow using GUI.

that the same process can be recorded using different words, and some processes may not be recorded. Therefore, we adopt a sequence alignment technique to adjust and complement these records (Fig. 2).

However, some processes, specifically the initial processes implemented for critical failures, tend not to be recorded because the actions are taken prior to recording them. To extract workflows for such pro-

cesses, we should not rely on trouble ticket logs, but rather use command logs. To extract command logs for graphical user interface (GUI) applications, we have developed a technique to independently build GUI command log sequences on applications (Fig. 3).

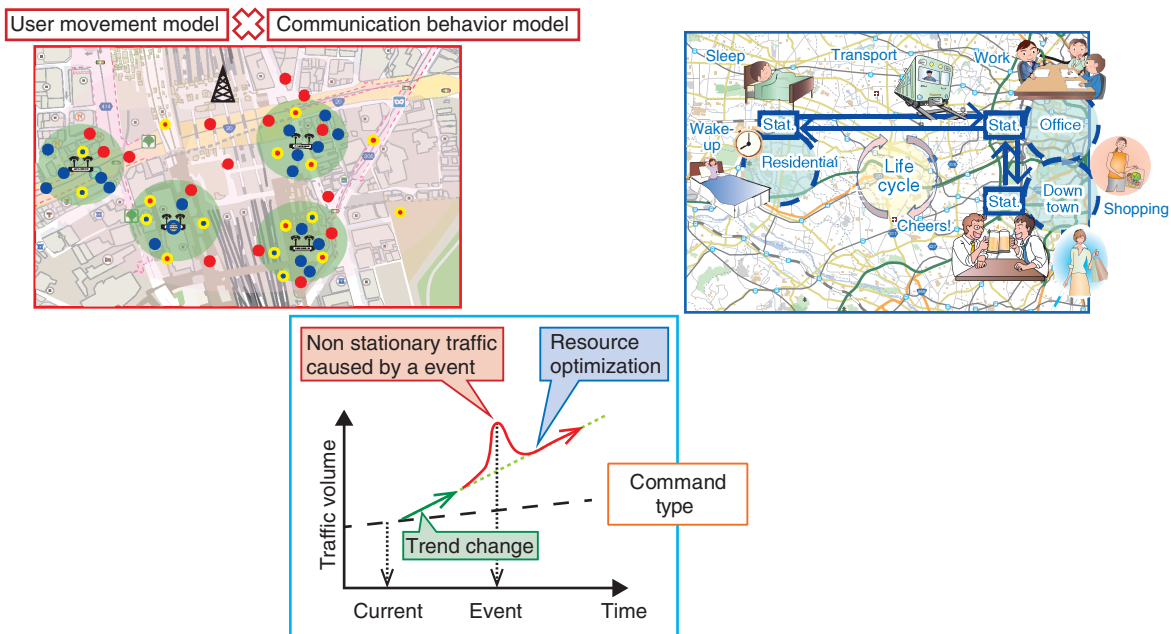


Fig. 4. High accuracy traffic prediction.

#### 4. Traffic prediction through its generation process

Traffic prediction is based on using past values to extrapolate future values. However, the drawback of this method is that it cannot adapt to changes in the traffic generation mechanism such as application usage or popular content. Specifically, mobile traffic is critically affected by changes in human movement such as those occurring during sports or music events. We are developing a technique to predict future traffic based on not only past traffic values but also traffic generation mechanisms such as human movement patterns and application usages (**Fig. 4**). We also adopt a method for long-term future traffic prediction

by analyzing and feeding back prediction errors.

#### 5. Future direction

In this article, we described our research that focuses on solving problems related to network management and operation by applying machine learning and data analytics. We plan to move forward with this research from analyzing and predicting the network status to optimizing it.

#### Reference

- [1] O. Komori and S. Eguchi, "A Boosting Method for Maximizing the Partial Area under the ROC Curve," BMC Bioinformatics, Vol. 11, pp. 314–330, 2010.



### Keisuke Ishibashi

Senior Research Engineer, Supervisor, NTT Network Technology Laboratories.

He received a B.S. and M.S. in mathematics from Tohoku University, Miyagi, in 1993 and 1995, and a Ph.D. in information science and technology from the University of Tokyo in 2005. Since joining NTT in 1995, he has been researching traffic issues in computer communication networks. He received the Young Researcher's Award from the Institute of Electronics, Information and Communication Engineers (IEICE) in 2002, the Information Network Research Award in 2002 and 2010, and the Internet Architecture Research Award in 2009. He is a member of the Institute of Electrical and Electronics Engineers (IEEE), IEICE, and the Operations Research Society of Japan.



### Takanori Hayashi

Senior Research Engineer, Communication Traffic & Service Quality Project, NTT Service Integration Laboratories.

He received his B.E., M.E., and Ph.D. in engineering from the University of Tsukuba, Ibaraki, in 1988, 1990, and 2007. Since joining NTT in 1990, he has been conducting research on subjective quality assessment of multimedia telecommunications and network performance measurement methods. He is currently working on a multimodal quality assessment method over IP networks. He is a member of IEICE.



### Kohei Shiomoto

Senior Manager of Communication & Traffic Service Quality Project, NTT Network Technology Laboratories.

He received his B.E., M.E., and Ph.D. in information and computer sciences from Osaka University in 1987, 1989, and 1998. He joined NTT in 1989 and began researching asynchronous transfer mode (ATM) traffic control and ATM switching system architecture design. During 1996–1997, he was a Visiting Scholar at Washington University in St. Louis, MO, USA. During 1997–2001, he directed architecture design for the high-speed IP/MPLS (multiprotocol label switching) label switching router research project at NTT Network Service Systems Laboratories. He also conducted research on photonic IP router design and routing algorithms, and Generalized MPLS (GMPLS) routing and signaling standardization at NTT Network Innovation Laboratories and then at NTT Network Service Systems Laboratories, and he was involved in GMPLS standardization in the Internet Engineering Task Force. He led the IP Optical Networking Research Group in NTT Network Service Systems Laboratories from April 2006 to June 2011 and the traffic engineering research group at NTT Service Integration Laboratories from July 2011 to June 2012. He has been in his present position since July 2012. He has chaired various committees of the Asia-Pacific Board of the IEEE Communications Society and served as Secretary for International Relations of the Communications Society of IEICE. He has also been involved in organizing several international conferences including MPLS, iPOP, and WTC. He received the Young Engineer Award from IEICE in 1995 and the Switching System Research Award from IEICE in 1995 and 2000. He is a fellow of IEICE and a member of IEEE and the Association for Computing Machinery.