

# Generating Network Performance Management Data with Various Network Failures

夏 飛<sup>†</sup> 沢辺 亜南<sup>‡</sup> 福元 徳広<sup>†</sup> 杜 平<sup>†</sup> 中尾 彰宏<sup>†</sup>

<sup>†</sup> 東京大学工学部 〒113-0033 東京都文京区本郷 7-3-1

<sup>‡</sup> 日本電気(株) 〒108-8001 東京都港区芝 5-7-1

E-mail: <sup>†</sup> xiafei571@gmail.com, {fukumotonorihiro, duping}@g.ecc.u-tokyo.ac.jp, nakao@nakao-lab.org

<sup>‡</sup> anan.sawabe@gmail.com

**あらまし** ネットワークや機器の障害を自動的かつ迅速に検出することは、ネットワーク運用者にとって必須の課題となっている。障害に関するデータの収集の難しさに対処するために、データの生成手法が研究者の注目を集めている。データ生成手法の中でも、Generative Adversarial Network (GAN)とその派生手法は、その優れた性能から広く応用されている。そこで本研究では、KDDI が提供するネットワークトラフィックデータを用いて、データ生成のためのワークフローを提案し、実験を行った。

まず、元のデータセットを小さなサブデータセットに分割する。次に、ネットワークトラフィックデータの時系列特性を捉えるために、従来の GAN モデルの代わりに TimeGAN モデルを採用し、各サブデータセットに基づいてデータの生成を行う。そして、生成されたすべてのデータセットを統合し、元のデータを用いて学習されたトラフィックの障害タイプを確実に認識することができるランダムフォレストモデルを用いて検証を行う。実験の結果、timeGAN によって生成されたデータは、92%の再現率を達成できることが証明された。

**キーワード** GAN, 機械学習, ネットワーク, ネットワーク障害

# Generating Network Performance Management Data with Various Network Failures

XIA FEI<sup>†</sup> Anan Sawabe<sup>‡</sup> Norihiro Fukumoto<sup>†</sup> Ping Du<sup>†</sup> and Akihiro Nakao<sup>†</sup>

<sup>†</sup> Faculty of Engineering, University of Tokyo 7-3-1 Hongo, Bunkyo-ku, Tokyo 113-0033, Japan

<sup>‡</sup> NEC Headquarters Building, 5 Chome-7-1 Shiba, Minato City, Tokyo 108-8001 Japan

E-mail: <sup>†</sup> xiafei571@gmail.com, {fukumotonorihiro, duping}@g.ecc.u-tokyo.ac.jp, nakao@nakao-lab.org

<sup>‡</sup> anan.sawabe@gmail.com

**Abstract** Automatically and quickly detecting network and device failures has become an essential task for network operators. To handle the struggles with data gathering, synthetic data generation has grabbed the attention of researchers. Among synthetic data generation methods, generative adversarial network (GAN) and its variants have been widely applied due to their outstanding performances. Following this aspect, we propose a workflow and conduct an experiment based on network traffic data provided by KDDI.

First, we split the original dataset into smaller sub-datasets. Second, to capture the time series characteristic in the network traffic data, we employ a TimeGAN model, instead of the conventional GAN model, to perform synthetic data generation based on each sub-dataset. Then we merge all of the generated datasets together and verify the result through a random forest model, which has been trained by using the original data and recognize the traffic failure type for certain. The experiment results prove that the data generated by timeGAN can achieve a 92% recall rate.

**Keywords** GAN, Machine Learning, Network, Network Failure