社団法人　電子情報通信学会
THE INSTITUTE OF ELECTRONICS,
INFORMATION AND COMMUNICATION ENGINEERS

信学技報
TECHNICAL REPORT OF IEICE.

[ポスター講演] 敵対的生成ネットワークを利用した疑似トラヒック生成に関する一考察

山際　哲哉　　渡部　康平　　中川　健治

† 長岡技術科学大学　大学院工学研究科　〒940-2188 新潟県長岡市上富岡町 1603-1
E-mail: tyamagiwa@kashiwa.nagaokaut.ac.jp,{k_watabe,nakagawa}@nagaokaut.ac.jp

あらまし　ネットワークを構築する際には，実際にトラヒックを生成するトラヒックジェネレータを使用し，サーバーなどのネットワーク機器の負荷テストを行うことは重要である．しかし，トラヒックジェネレータで実際のトラヒックの特性を多面的に再現した疑似トラヒックを生成することは困難である．一方で，敵対的生成ネットワーク (Generative Adversarial Networks;GAN) という画像などのデータを模倣して生成する機械学習の研究が発展してきている．本稿では，GAN と次元圧縮手法である Encoder を用いて疑似トラヒックデータ生成手法を提案する．提案手法は，任意の長さのトラヒックデータを学習させ，疑似トラヒックデータを生成する．評価指標として，二つの確率分布が一致していないかを検定するコルモゴロフ-スミルノフ検定を用いて，疑似トラヒックデータの妥当性を検証する．そして，提案した疑似トラヒックデータが実際のトラヒックデータの分布とかけ離れた分布ではないことを確認した．
キーワード　トラヒックジェネレータ，機械学習，敵対的生成ネットワーク

A Study on Pseudo Traffic Generation Using Machine Learning

Tetsuya YAMAGIWA, Kohei WATABE, and Kenji NAKAGAWA

† Graduate School of Electrical Engineering, Nagaoka University of Technology　Kamitomiokamachi 1603-1, Nagaoka, Niigata 940-2188 Japan
E-mail: tyamagiwa@kashiwa.nagaokaut.ac.jp,{k_watabe,nakagawa}@nagaokaut.ac.jp

Abstract　When we constructing a network, it is important to use a traffic generator that generates realistic traffic and perform load tests on network devices such servers. However, it is difficult for traffic generators to generate pseudo traffic that reproduces the multi-aspect characteristics of actual traffic. On the other hand, research on machine learning in the field of image processing has been developed. In this paper, we propose a traffic generator with machine learning technology which utilizes a dimensional compression method. The proposed method learns traffic data of any length and generates traffic. The validity of the generated traffic is verified using Kolmogorov-Smirnov test as a performance measure, which tests the equality of two probability distributions. We confirmed that a distribution of generated traffic was not far from that of the actual traffic data.
Key words　Traffic generator,Machine learning, Generative adversarial networks

1. はじめに

ネットワークを構築する際には，実際にトラヒックを生成するトラヒックジェネレータを使用し，サーバーなどのネットワーク機器の負荷テストを行うことは重要である．日々トラヒックが増加しており，データ量が膨張する Web サイトでは，システムダウンを起こさずに，トラヒック変動に対応しなければならない．システムダウンを起こさないためには，キャパシティプランニングを行うことが有効である．キャパシティプランニングは，システムを運用する前にあるトラヒックを処理するのに，どの程度増強すれば良いかということを計画，検討するものである．

しかし，トラヒックジェネレータで実際のトラヒックの特性を多面的に再現した疑似トラヒックを生成することは難しい．疑似トラヒックの生成は，実際のトラヒックの複雑さを再現するために統計学的知識が必要であり，容易ではない．そこで本稿では，機械学習の一種である敵対的生成ネットワーク (Generative Adversarial Networks;GAN)[1] を用いて，疑似トラヒックデータを生成する方法を提案し，生成方法の妥当性を検証する．

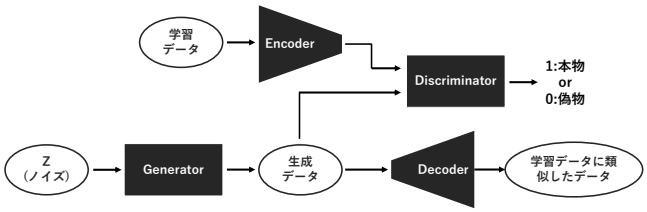

図 1　疑似トラヒック生成のための GAN

2. GAN

GAN とは，画像などのデータを生成する生成器 (Generator) G とデータが本物か生成物かを識別する識別器 (Discriminator) D の二つのニューラルネットワークからなる機械学習の一種である．応用例として，手書き文字の画像や動物の写真，顔写真などを学習させることで，この世に存在しない類似した画像を生成することが報告されている．GAN は，本物のデータの分布と生成データの分布が一致するように学習していく生成モデルである．

Discriminator は本物のデータと Generator から生成されたデータを識別するように学習する．Discriminator は入力されたデータが本物のデータ X なのかノイズ Z からの Generator の生成データ $G(Z)$ なのかを識別できるように学習する．学習の際には，入力データが本物であれば 1 を，生成データであれば 0 を出力するようにすることで，Discriminator の出力は本物のデータである確率を表している．

Generator は Discriminator を騙せるほどの本物のデータに類似したデータを生成するように学習する．Generator は生成データ $G(Z)$ を Discriminator に入力し，Discriminator の出力 $D(G(Z))$ が本物のデータである確率に近づけるように学習する．以上のように，Discriminator と Generator が敵対し交互に学習し合うことで，Generator は本物のデータに類似したデータを生成できるようになる．

3. 疑似トラヒック生成のための GAN

本研究で扱うトラヒックデータは，計測期間について一定時間間隔毎に送信されたバイト数の時系列であり，このトラヒックデータを使用し，疑似トラヒックデータを生成する．提案法は，時間間隔毎のバイト数を成分に持つベクトルを入力とし，疑似トラヒックデータを表す同様のベクトルを出力する．

提案する疑似トラヒック生成手法では次元圧縮を行う手法である Encoder を用いて，トラヒックデータの特徴を捉える．提案する手法は，Encoder で特徴を捉えることで GAN の学習及び生成を容易にし，疑似トラヒックデータの生成を可能とする．疑似トラヒック生成のための GAN の構成を図 1 に示す．

学習データを十分に学習させた Encoder 及び Decoder と GAN を組み合わせることで，疑似トラヒックを生成する．まず，学習データを Encoder 及び Decoder に学習させ，Decoder が学習データを十分復元できるまで学習を繰り返す．そして，学習し終えた Encoder と GAN を組み合わせる．学習の際には，Encoder を固定し GAN のみを学習させる．全ての学習終了後，Generator からの生成データを Decoder に入力することで疑似トラヒックデータを生成する．

表 1　Encoder,Decoder の構成

Encoder			Decoder		
Layer	Units	Act.	Layer	Units	Act.
Input	288	-	Input	2	-
Hidden	400	LReLU	Hidden	50	LReLU
Hidden	200	LReLU	Hidden	100	LReLU
Hidden	100	LReLU	Hidden	200	LReLU
Hidden	50	LReLU	Hidden	400	LReLU
Output	2	Tanh	Output	288	-

表 2　GAN の構成

Generator			Discriminator		
Layer	Units	Act.	Layer	Units	Act.
Input	2	-	Input	2	-
Hidden	10	LReLU	Hidden	50	LReLU
Hidden	30	LReLU	Hidden	40	LReLU
Hidden	40	LReLU	Hidden	30	LReLU
Hidden	45	LReLU	Hidden	10	LReLU
Output	2	-	Output	1	Sigmoid

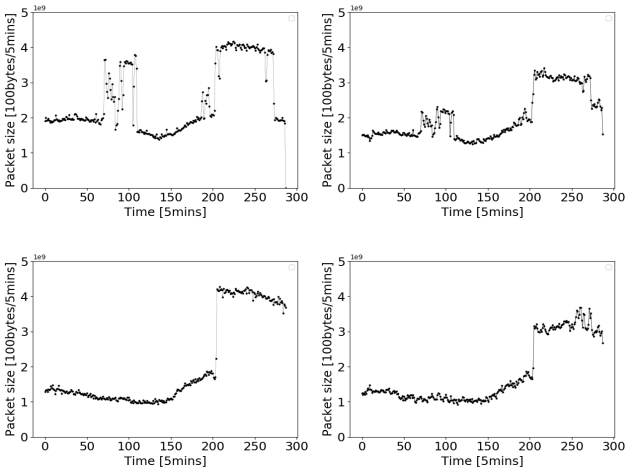

図 2　学習データ (左) と生成データ (右) の例

4. トラヒックデータ生成実験

公開されているトラヒックデータ [2] を学習データとし，疑似トラヒック生成のための GAN を用いてトラヒックデータの生成を行った．ここで，トラヒックデータは 5 分間毎のバイト数を表しており，1 日分のデータを 1 つの学習データとし，全部で 168 日分ある．Encoder，Decoder 及び GAN のパラメータ構成は表 1,2 とした．実際のトラヒックデータと生成した疑似トラヒックデータの例を図 2 に示す．

評価指標として，二つの確率分布が一致していないかを検定するコルモゴロフ-スミルノフ検定 (以下 KS 検定) を用いる．KS 検定は，KS 統計量が設定した有意水準以上の場合，棄却され分布が一致していないと結論付ける．本研究では，5 分間毎のバイト数における実際のトラヒックデータと生成した疑似

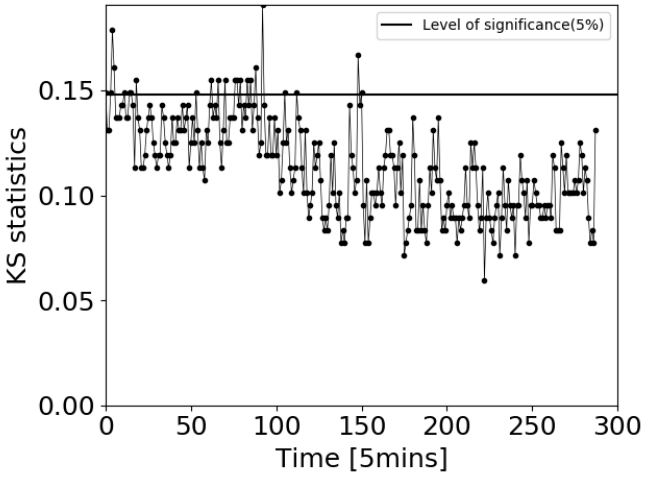

図 3　KS 検定の結果

トラヒックデータをそれぞれ 168 サンプルずつで KS 検定を行い，評価する．

実際のトラヒックデータと GAN を利用して生成した疑似トラヒックデータの KS 検定の結果を図 3 に示す．結果の図から，多数の点が KS 検定で棄却されていないことを確認できる．つまり，生成した疑似トラヒックデータは実際のトラヒックデータの分布とかけ離れた分布ではないことが言える．GAN は，ノイズ Z によってランダムな生成になるため，必ずしも KS 検定で棄却されない訳ではない．そこで，GAN で何度も生成しサンプル数を増やすことで，KS 検定で棄却される割合を確認した．500 サンプルまで生成した結果，KS 検定で棄却される割合は 2.3% に収束していくことが確認できた．本研究では有意水準を 5% と設定しており，本物のトラヒックデータで検定した場合 95% 棄却されないが，生成した疑似トラヒックデータで検定した場合 97.7% 棄却されていない．つまり，GAN を用いて生成した疑似トラヒックデータは本物のトラヒックデータとほぼ同等であることを確認した．

5. おわりに

機械学習の一種である GAN を用いて，実際のトラヒックに近いデータを生成する方法を提案し，生成方法の妥当性を検証した．今後は，GAN の精度向上及び生成されるトラヒックに多様性を持たせられるか確認する．

謝　辞

本研究の一部は，JSPS 科研費 JP17K00008，および JP18K18035 の助成を受けたものである．

文　献

[1] I. Goodfellow *et al.*, “Generative Adversarial Nets,” Advances in Neural Information Processing Systems 27, pp. 2672-2680, 2014

[2] Abilene Topology and Traffic Dataset. http://www.cs.utexas.edu/~yzhang/research/AbileneTM/