敵対的生成ネットワークを利用した疑似トラヒックデータ生成手法

Pseudo Traffic Data Generation Method Using Machine Learning

山際 哲哉 Tetsuya Yamagiwa 渡部 康平 Kohei Watabe 中川 健治 Kenji Nakagawa

長岡技術科学大学 大学院工学研究科

Graduate School of Electrical Engineering, Nagaoka University of Technology

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.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.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.

2 GAN

GANとは、画像などのデータを生成する生成器 (Generator) とデータが本物か生成物かを識別する識別器 (Discriminator) 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。 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 疑似トラヒックデータ生成のための 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ことで GAN の学習及び生成を容易にする. 疑似トラヒック生成のための GAN の構成を図 1 に示す.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、GAN のみを学習させる.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.

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] を使用し、疑似トラヒック生成のための GAN を用いて疑似トラヒックデータの生成を行った。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。

評価として、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(KS検定)を用い

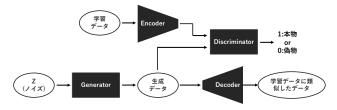


図 1 疑似トラヒック生成のための GAN

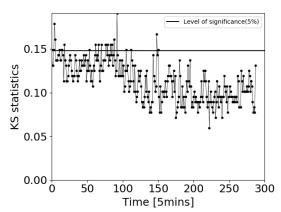


図 2 KS 検定の結果

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疑似トラヒックデータの KS 検定の結果を図 2 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.

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.

斜辞

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.

参考文献

- [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/