

RNNを用いたネットワークトラフィックの変動予測

小山内 遥香[†]中尾 彰宏[‡]山本 周[‡]山口 実靖[§]小口 正人[†][†]お茶の水女子大学[‡]東京大学[§]工学院大学

1. はじめに

近年、各地で大規模災害が多発しており、そのような災害時には被災状況や安否の確認、避難経路等の情報を得るために、通話やメールなどの通信による情報伝達が可能であることが重要である。しかし、被災地域では被災地内外から通信ネットワークの処理能力を大幅に超える通信が集中する。そのため、輻輳状態になってしまい通信がほとんどできなくなる可能性がある。このような状況に陥るのを防ぐため、通信障害を早期に検知し、効率的に対処する必要がある。

本研究では、深層学習のモデルの一種である Recurrent Neural Network(RNN) を用いてトラフィック異常の情報を抽出し、トラフィック変動の兆候を掴むための手法を提案する。

2. 深層学習

深層学習は、機械学習の一種であるニューラルネットワークの階層を深めたアルゴリズムである。ニューラルネットワークとは、生物の脳の神経回路をモデルとしたアルゴリズムで、入力層、隠れ層、出力層を持つ。各層は複数のノードがエッジで結ばれる構造であり、各エッジはそれぞれ重みを持っている。出力層からの出力と正解データの誤差を計算し、それを出力層から入力層へ逆に伝播させることにより重みを更新して学習を行う。深層学習は、画像や音声の認識、自動車の自動運転など様々な分野に応用され、社会に大きな利益をもたらしており注目されている。

2.1 RNN

RNN は、時系列データの学習に適したネットワークで、以前に計算された情報を記憶することができる。しかし、系列が長く深いネットワークになると、誤差逆伝播のアルゴリズムでは勾配の消失、発散などの問題が生じてしまう。

この問題を解決するために RNN を改良したモデルが Long Short-Term Memory(LSTM) である。LSTM は、RNN に入力ゲート、出力ゲート、忘却ゲートを導入することにより、RNN では扱えなかった長期依存を扱えるようになっている。

3. 実験

本研究では、LSTM を用いて 1 秒毎の packets 数を予測する実験を行う。使用した深層学習のフレームワークは PFN 社の Chainer[1] である。

3.1 実験概要

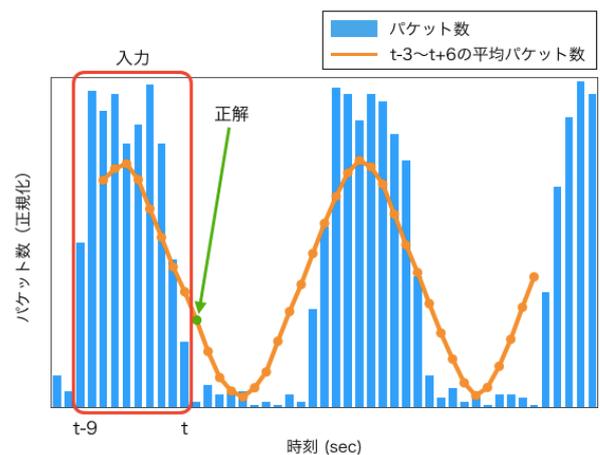


図 1: 入力データと正解データのイメージ図

学習を行う際の入力データと正解データの作成方法を図 1 に示す。

正解データは時刻 t 秒の前 4 秒と後 5 秒の間に届いたパケット数の平均を正規化したものである。つまり、時刻 $t-3$ 秒 \sim $t+6$ 秒の 10 秒間に届いたパケット数の平均を正規化した値であり、図 1 の緑色の点が表示している。10 秒間のパケット数の平均とした理由としては、パケット数の詳細な予測を行うことは困難であるが、ネットワークトラフィックのおおまかな変動が予測できれば十分役に立つためである。また、その値を正規化したものを正解データとしたのは、学習効率を上げるためである。

入力データには以下の 16 個のパケット情報を用いている。

- パケットが送受信された時間
 - プロトコル名 (TCP, UDP, ARP など)
 - パケットサイズ
 - 送信元 IP アドレス
 - 宛先 IP アドレス
 - 送信元ポート番号
 - 宛先ポート番号
 - TCP ヘッダのフラグ 9 個
- res, ns, cwr, ecn, urg, ack, push, reset, syn

Prediction of Variation in Network Traffic by RNN

Haruka Osanai[†]

Akihiro Nakao[‡]

Shu Yamamoto[‡]

Saneyasu Yamaguchi[§]

Masato Oguchi[†]

[†]Ochanomizu University

[‡]University of Tokyo

[§]Kogakuin University

図1の赤枠t-9~t秒の10秒間に届いたパケットそれぞれの上記16個の情報と、正解データを組にして学習を行う。

表 1: 実験で用いた計算機の性能

OS	Ubuntu 14.04.4LTS
CPU	Intel Core i7-6700K CPU @4.00GHz
GPU	GeForce GTX 1080
Memory	32Gbyte

実験で用いた計算機の性能を表1に示す。

3.2 実験結果

ネットワークアナライザソフトウェアであるWireshark[2]を用いて研究室LAN内で取得したデータを学習データとしている。学習時のepoch数は1000、隠れ層の数は32、学習率は0.03である。

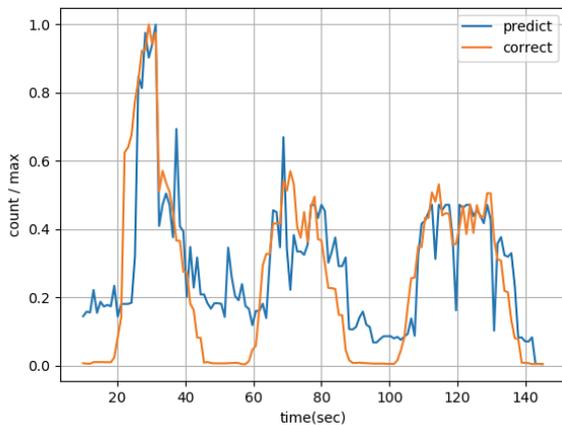


図 2: 学習データを用いた予測

学習したモデルに学習データを入力して予測を行なった結果が図2である。多少正解データとの誤差が大きい箇所があるが、概ねトラフィック変動の予測が可能であることがわかる。

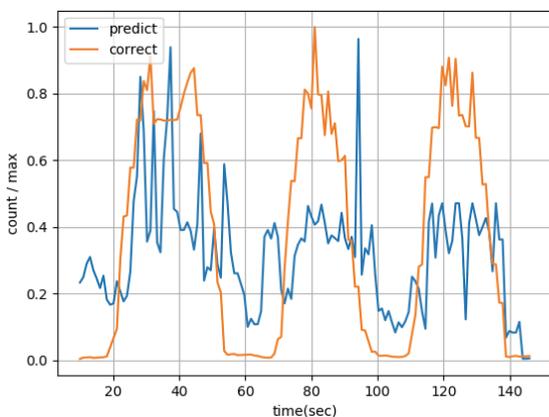


図 3: テストデータを用いた予測

表 2: 平均二乗誤差

train	test
0.02586	0.09511

次に、テストデータを用いた予測性能評価の結果を図3に示す。テストデータも、学習データ同様、研究室LAN内で取得している。パケット数の増加や現象が起こる箇所は概ね予測できているが、学習データと比較すると精度が下がることがわかる。

学習データを用いたときとテストデータを用いたときの平均二乗誤差を表2に示す。この表からも、テストデータを用いたときは、学習データを用いたときと比較して、誤差が約3.7倍になり予測精度が大幅に下がっていることがわかる。

4. まとめと今後の予定

RNNの一種であるLSTMを用いてネットワークトラフィックの変動予測を行い、性能評価を行なった。テストデータを用いて性能評価を行なった結果、パケット数の増加、減少のタイミングをおおまかに予測することができた。ただし、学習データを用いたときと比較すると精度が下がってしまった。また、学習データを用いて予測を行なった場合も、誤差が大きい箇所が見られるため、学習パラメータのチューニングや学習データ量の増加をさせる必要があると考えられる。

今後は、まずパラメータチューニングや学習データ量の増加を行い、予測性能の向上を目指す。その後、予測結果を用いた輻輳検知システムの構築や、ネットワーク制御を行う手法を検討する。また、実際のトラフィック異状時においても現在の予測システムが適用可能かを確認するため、輻輳発生時など実際のデータを用いた学習、予測の実験も行いたいと考えている。

謝辞

本研究は一部、総務省戦略的情報通信開発推進事業(SCOPE)先進的通信アプリケーション開発推進型研究開発によるものである。

参考文献

- [1] Tokui, S., Oono, K., Hido, S. and Clayton, J.: Chainer: a Next-Generation Open Source Framework for Deep Learning, In Proceedings of Workshop on Machine Learning Systems (LearningSys) in The Twenty-ninth Annual Conference on Neural Information Processing Systems (NIPS) (2015). 6 pages.
- [2] <https://www.wireshark.org>