

# 深層学習を用いた有線通信におけるネットワークトラフィック変動の予測手法

明石 季利子†      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

突然発生する通信障害は、大規模災害時による通信過多やDDoS攻撃、同時に起こるOSアップデートなど、さまざまな原因で引き起こされる。これらの通信障害が起こってからでは対応が手遅れである場合が多い。また、近年深層学習による時系列データの解析技術が急速に発展しており、株価予測や音声認識といった自然言語処理に多く活用されている。そこで本研究では、深層学習モデルの長短期記憶(LSTM)ネットワークを用いてネットワークパラメータを特徴量とした時系列データの予測を行った。有線通信時のトラフィック異常の情報から、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

深層学習とは、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。これを用いることにより、データの特徴をより深いレベルで学習することができる。LSTMとは、長期記憶を可能にした再帰型ニューラルネットワークの一つであり、時系列データの学習に用いられる。本研究で扱うパケットデータは時系列データであり、ある程度長いパターンの学習を可能にするLSTMを用いて機械学習を行う。

本研究では、Preferred Networks社が開発したChainer[1]を使用している。

## 3 実験

### 3.1 実験環境

実験用マシンを3台用意し、1台をサーバ、残りの2台をクライアントとしてiPerfを用いてTCPで通信し、送受信されたパケットをサーバ側でキャプチャする。サーバ・クライアント間にダミーネットを挟み、遅延

を100ms、サーバ・ダミーネット間の帯域を70mbpsとした(図1)。その際カーネルモニタ[2]を用いて通信時のクライアントそれぞれのcwndを取得する。TCP輻輳制御アルゴリズムはRenoを用いた。

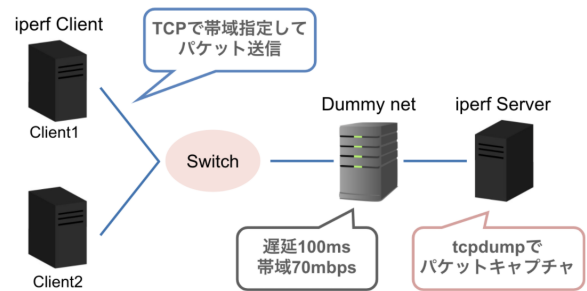


図1: 実験環境

入力データにはt-9秒~t秒の10秒間に送受信したパケットから1秒おきに6つの特徴量を導出したものとする。

- パケット送受信時刻t
- パケットサイズの平均
- 時刻tに送受信したパケット数(移動平均3.5秒)
- Client1・2のcwnd値(含む場合)
- 増加ラベル(含む場合)

増加ラベルとは、1秒前のパケット数と比べて増加したかどうかを0,1で表したものである。正解データは時刻t+1秒に届いたパケット数を0~1に正規化した値である。

### 3.2 学習モデルによる帯域予測

学習に使用したパケット数は、3秒と5秒での移動平均をとり、cwndを学習の特徴量に含める場合と含めない場合、増加ラベルを含める場合の3つの場合で合計6種類の学習モデルを作成した。それぞれの学習モデルを用いて、変動パターンが異なる3種類のバリエーションデータの帯域予測(図2)を行ったところ、目視においてはおおまかな傾向を予測できることがわかる。

図3は、従来の回帰予測に適した評価指標として、平均平方二乗誤差(RMSE: Root Mean Square Error)を求め、正解値と予測値のずれを計算した結果である。

A Method for Predicting Network Traffic Variability in Wired Communications Using Deep Learning

†Kiriko Akashi    ††Akihiro Nakao    †††Shu Yamamoto

†††Saneyasu Yamaguchi    †Masato Oguchi

†Ochanomizu University

††University of Tokyo

†††Kogakuin University

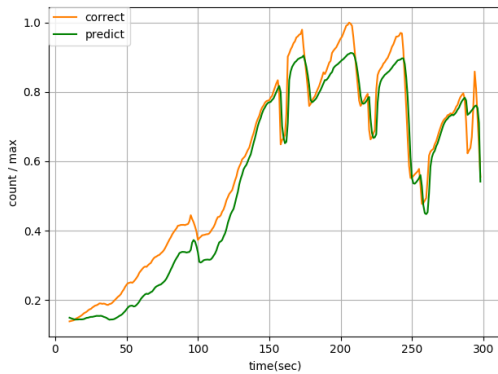


図 2: バリデーションデータの予測結果

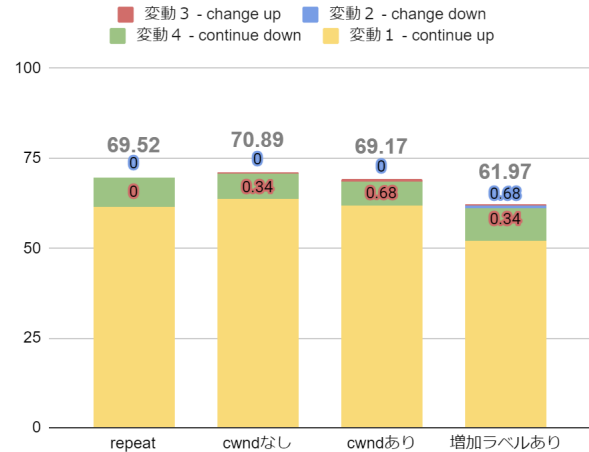


図 4: バリデーションデータを用いた各モデルの MATCH

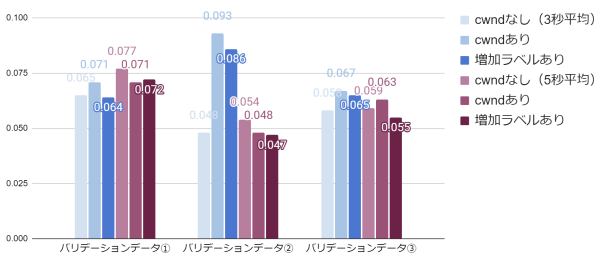


図 3: 提案モデルの RMSE の値

#### 4 提案手法の性能評価

送受信パケット数の変動などの時系列データの予測において、増加し続けるなど同じ変動が続く場合、直前の変動を繰り返すことで、ある程度高い精度で予測ができてしまうモデルを本研究では repeat と呼ぶ。表 1 は、送受信パケット数の変動の種類を表したものである。repeat は増減が変化する箇所 (変動 2, 3) を予測することができないため、その点に注目して提案手法である cwnd なし・cwnd あり・増加ラベルありの学習モデルを用いて予測との性能比較を行う。その際、時間的に分割した各区間が変動 1~4 の 4 種類のどれになるかを予測し、それがどの程度正解となったか、という評価結果 (MATCH) [3] を算出した。

表 1: 変動の種類

		現在の変動	
		増加	減少
直前の変動	増加	変動 1	変動 2
	減少	変動 3	変動 4

図 4 では、提案手法である cwnd なし, cwnd ありの MATCH の値を repeat と比較し、全体的に差異が見ら

れない結果となった。また、増加ラベルありの MATCH の値は、増減方向の変化点において 1 割程度の増加分が見られ、repeat による予測ができない帯域の予測ができることがわかった。この他 2 種のバリデーションデータを用いて MATCH を算出した結果も図 4 と同様の結果となった。

#### 5 まとめと今後の課題

LSTM をネットワークトラフィックの予測に応用し、性能評価を行なった。MATCH の観点から他モデルとの性能比較を行うことにより、他モデルでは予測できない帯域の変動も提案手法では予測が可能であることを示した。しかし、今後も repeat モデルと比較した学習モデルの精査が必要であることがわかった。学習データの増加、マシンの台数を増やすなど、より複雑なトラフィックでの実験を行いたいと考えている。

#### 参考文献

- [1] Seiya Tokui, Kenta Oono, Shohei Hido, and Justin Clayton. 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.
- [2] Kaori Miki, Saneyasu Yamaguchi, and Masato Oguchi. Kernel monitor of transport layer developed for android working on mobile phone terminals. In *Tenth International Conference on Networks (ICN)*, 2011.
- [3] 南正太郎. LSTM-RNN を用いたイベント考慮後の株価時系列予測. 第 19 回 人工知能学会 金融情報学研究会 (SIG-FIN), 2017.