

ネットワークのトラフィック変動時の深層学習による輻輳の解析

小山内 遥香 (指導教員: 小口 正人)

1 はじめに

近年, 各地で大規模な災害が多発している. 大規模災害時には, 被災状況や安否確認, 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非常に重要である. しかし, このような災害時には, 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し輻輳状態になってしまうため, 通信がほとんどできなくなる可能性がある. そのため, 通信障害を早期に検知し, 効率的に対処する必要がある.

また, 近年深層学習の技術が発達してきている. 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, 非常に高い解析性能により社会に多大な利益をもたらしている.

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状の情報を抽出することで, ネットワークトラフィックが大きく変動するときの兆候を掴む手法を提案する.

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り, 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.

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ある. 入力層, 隠れ層, 出力層を持ち, 各層は複数のノードがエッジで結ばれる構造となっており, 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.

2.1 RNN

RNN は, 時系列データの学習に用いられるネットワークで, 以前に計算された情報を記憶することができる (図 1). しかし, 系列が長くなり深いネットワークになると, 実際には 2, 3 ステップ前までの記憶しか保持することができない.

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) であり, LSTM は RNN の一種である. 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 では扱うことができなかった長期依存を扱えるようになっている (図 2).

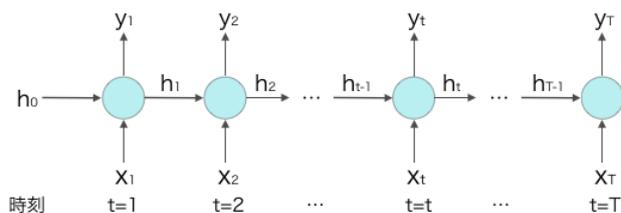


図 1: RNN のネットワーク図

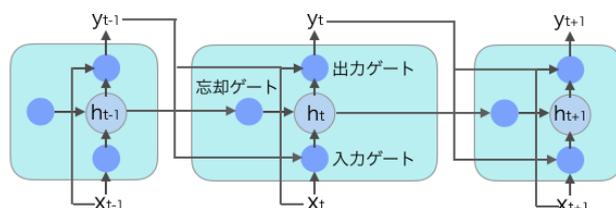


図 2: LSTM のネットワーク図

3 実験

3.1 実験概要

本研究では, LSTM を用いて 1 秒毎のパケット数を予測する実験を行う. 使用したフレームワークは PFN 社の Chainer[1] である.

学習を行う際の入力データと正解データの作成方法について説明する. 正解データは時刻 $t+1$ 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~ $t+6$ 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.

入力データにはパケット情報から導出した 18 種の特徴量を用いており, 詳細は以下である.

- パケット送受信時刻
- 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 10 個
 - res, ns, cwr, ecn, urg, ack, push, reset, syn, fin
- ウィンドウサイズ

$t-9 \sim t$ 秒の 10 秒間に届いたパケットそれぞれの上記 18 種の特徴量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 学習データ

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、研究室 LAN 内においてブラウザで動画サイトをひとつだけ開き、異なる 30 秒間の動画 4 つを自動で連続再生したときの packets をキャプチャした。このデータを用いて学習を行い予測モデルを作成した。学習時の epoch 数は 700、隠れ層の数は 49 である。

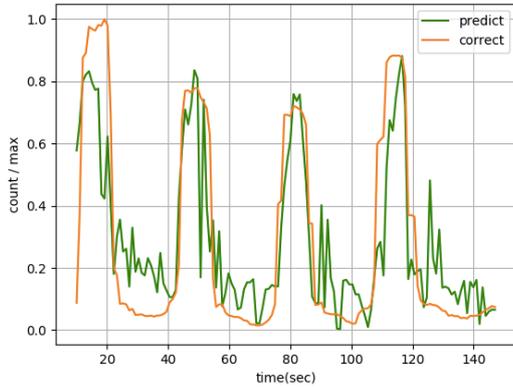


図 3: 学習データを用いた予測

予測モデルに学習データを入力し予測した結果、学習データの特徴をうまく学習できていることがわかる(図 3)。

3.3 性能評価

性能評価では、2 つのテストを行なった。

- テスト 1: 学習データを取得したときと同じ環境で、学習データとは異なる 4 つの 30 秒間の動画を自動で連続再生し packets を取得
- テスト 2: 動画サイトを変更し、30 秒間の動画 4 つを自動で連続再生し packets を取得

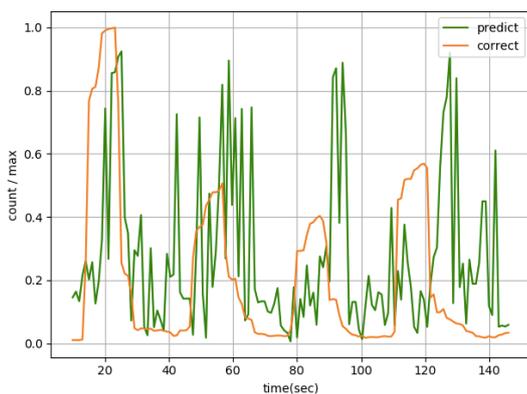


図 4: テスト 1 の予測結果

図 4 はテスト 1 の予測結果である。学習データと比較すると精度が下がるが、パケット数の増加や減少が起こる箇所は概ね予測できていることがわかる。

図 5 はテスト 2 の予測結果である。パケット数の増加、減少のタイミングがずれているが、増加や減少の差が明確に予測できている。

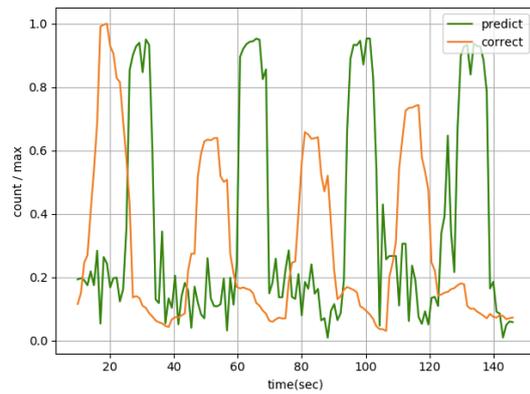


図 5: テスト 2 の予測結果

それぞれの予測結果の平均二乗誤差を表 2 に示す。テスト 2 の誤差が非常に大きくなっているのは正解データと予測にずれがあるためであり、正解データと予測のずれが解消できれば誤差が減少すると考えられる。

表 2: 平均二乗誤差

学習データ	テスト 1	テスト 2
0.03169	0.10636	0.24920

4 まとめと今後の課題

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、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が、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。

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。また、テスト 2 の予測結果でパケット数増減のタイミングがずれてしまう原因について考えたい。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>