

深層学習を用いたネットワークトラフィック予測 Network Traffic Prediction with Deep Learning

徳山 雄丈[†] 福島 行信[‡] 樽谷 優弥[†] 横平 徳美[†]

Yusuke Tokuyama Yukinobu Fukushima Yuya Tarutani Tokumi Yokohira

1. はじめに

クラウドサービスや動画配信サービスの普及により、世界中でインターネットトラフィックは増加傾向にある。シスコは 2019 年のインターネットトラフィックが 2014 年のものと比べておよそ 3 倍になると予測している[1]。このように増加するトラフィックを収容するために、ネットワーク事業者はネットワーク設備の増強やトラフィックエンジニアリングを実施しなければならないが、これらの対策を適切に実施するためには将来のネットワークトラフィックを正確に予測することが有用である。

ネットワークトラフィックの予測に関する従来研究[2]では、リカレントニューラルネットワーク (RNN) を用いた予測手法 (RNN 手法) が最良の予測精度を示すことが報告されている。RNN とは内部に再帰的構造を持つことにより時系列データの処理を得意とするニューラルネットワークである。従来の RNN 手法では、トラフィック予測の際にトラフィック量のみを入力として用いており、トラフィックに関するその他の属性情報 (時刻や曜日) を用いていない。しかし、トラフィック量は時刻と曜日に応じて変化する傾向があるため、これらの属性情報を予測手法が入力として用いることでその予測精度が向上する可能性がある。

本研究では、トラフィック量に加えてトラフィックの属性情報を入力として用いる RNN 手法を 2 つ提案し、それらの有効性を検証する。さらに、RNN 手法で用いられる RNN モデルのハイパーパラメータが予測精度に及ぼす影響についても検討する。

2. ネットワークトラフィックの予測

2.1 ネットワークトラフィック予測問題

本研究で取り組むネットワークトラフィック予測問題を図 1 に示す。この問題では、着目するリンクにおいて、過去のトラフィック履歴に基づいて、将来のトラフィック量の推移を予測する。具体的には、現在時刻 t_c において、 t_{c-m+1} から t_c までの m 個の時点のトラフィック量が与えられた時に、将来の t_{c+1} から t_{c+n} までの n 個の時点のトラフィック量を予測する。

2.2 予測手法

トラフィック予測の際に、トラフィックの属性情報を用いることで予測精度が向上する可能性がある。文献[3]では、バックボーンリンクのトラフィック量は 6:00 から 12:00 にかけておよそ 3 倍に増加し、また週末には平日と比べてトラフィック量がおよそ 25%低下することが報告されている。

本研究では、トラフィック予測において時刻情報と曜日情報を用いることの有効性を検証するために 2 つの RNN 予測手法: RNN-VT と RNN-VTD を提案する。RNN-VT は

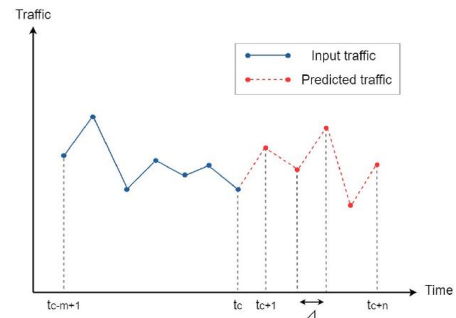


図 1. ネットワークトラフィック予測問題

トラフィック量に加えて時刻情報を入力として用いる RNN 手法であり、RNN-VTD はトラフィック量に加えて時刻情報と曜日情報を入力として用いる RNN 手法である。

RNN-VT で用いられる RNN モデルを図 2 に示す。このモデルには 2 つの入力層が含まれる。第 1 入力層はトラフィック量を入力層、第 2 入力層は時刻情報を入力層である。第 1 入力層の出力は GRU 層で構成される第 1 隠れ層を通過後、時刻情報と結合され、通常層で構成される第 2 隠れ層を通過した後に将来のトラフィック量として出力される。

RNN-VTD で用いられる RNN モデルを図 3 に示す。このモデルには 3 つの入力層が含まれる。第 1 入力層と第 2 入力層は RNN-VT と同じくトラフィック量と時刻情報を入力層である。第 3 入力層は曜日情報を入力層である。RNN-VT との違いは、第 2 隠れ層の出力がさらに曜日情報と結合された上で、第 3 隠れ層を通過した後に将来のトラフィック量として出力される点である。

3. 性能評価

3.1 実験環境

本研究では Amazon Web Service (AWS) [4] にて GPU コンピューティングインスタンスを使用して実験を行った。言語は Python を使用し、深層学習用のライブラリは Keras[5] を使用した。バックエンドには Google 社の TensorFlow[6] を指定した。

3.2 データセット

データセットとしては、WIDE プロジェクト[7]の MAWI ワーキンググループが提供しているトラフィックデータを用いる。このデータは 2014 年から 2017 年のおよそ 4 年間にバックボーンリンクで観測されたトラフィックデータを含む (観測間隔は 10 分)。最初の 47 ヶ月分を訓練データ、最後の 1 ヶ月分をテストデータとして利用する。予測手法

[†] 岡山大学大学院 ヘルスシステム統合科学研究科 The Graduate School of Interdisciplinary Science and Engineering in Health Systems, Okayama University

[‡] 岡山大学 自然科学研究科 The Graduate School of Natural Science and Technology, Okayama University

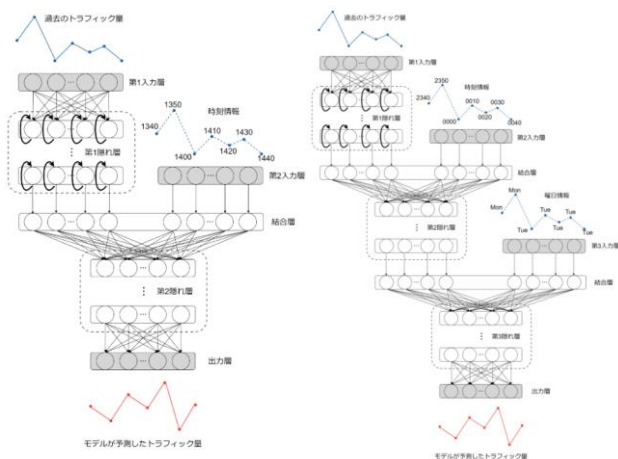


図 2. RNN-VT

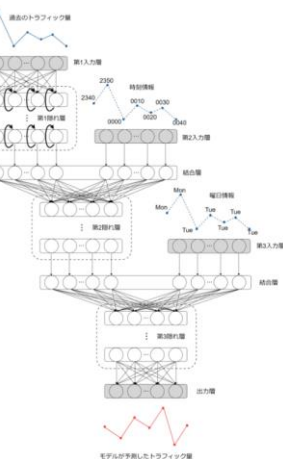


図 3. RNN-VTD

にトラフィックデータを与える際には、(1) トラフィック量は平均 0、分散 1 となるように標準化し、(2) 時刻は 4 つの整数(24 時間表記)で表し、(3) 曜日は 0 から 6 までの 7 通りの整数で表す。

3.3 ハイパーパラメータの設定

表 1 はハイパーパラメータの設定一覧である。入力トラフィックの長さは 1 日 (1440 分)、出力トラフィック (予測されるトラフィック) の長さは半日 (720 分) とした。これに伴い、入力層のニューロン数は 144、出力層のニューロン数は 72 とした。

3.4 評価指標

評価指標としては平均二乗誤差 (MSE) を用いる。1 回の予測における MSE は以下のように表される。

$$MSE = \frac{\sum_{i=1}^n \{\hat{v}(t_{c+i}) - v(t_{c+i})\}^2}{n}$$

ここで、 $\hat{v}(t)$ は時刻 t におけるトラフィック量の予測値、 $v(t)$ は時刻 t における実際のトラフィック量、そして n は出力トラフィックの長さである。テストデータに含まれる各入力トラフィックに対する予測の MSE を算出し、それらの平均値を予測手法の予測精度とする。

3.5 結果

表 2 に、従来手法である RNN-V と提案手法である RNN-VT, RNN-VTD それぞれにおいてハイパーパラメータを最適な値に設定した際に得られた最良の MSE を示す。

表 2 から、RNN-VT は、従来手法である RNN-V とほぼ同じ MSE を示すことがわかる。このことから、時刻情報は予測精度の向上においてさほど有効ではないことがわかった。一方で RNN-VTD は、従来手法である RNN-V と比較して MSE を 43% 程度低減できていることがわかる。このことから、曜日情報は予測精度の向上において有効であるといえる。

ハイパーパラメータである隠れ層内の層数と隠れ層内の各層のニューロン数が予測精度に及ぼす影響の評価結果を述べる。隠れ層内の層数については、いずれの予測手法においても 1 とした場合に予測精度が最良となった。隠れ層内の各層のニューロン数については、RNN-V と RNN-VT では 10、RNN-VTD では 50 とした場合に予測精度が最良

表 1. ハイパーパラメータの設定

ハイパーパラメータ	値
入力トラフィックの長さ	1 日 (1440 分)
予測トラフィックの長さ	半日 (720 分)
RNN の種類	GRU
隠れ層内の層数	1, 2, 3
隠れ層内の各層のニューロン数	10, 50, 100, 1000, 3000
バッチサイズ	20
最適化アルゴリズム	Adam
損失関数	MSE (Mean Squared Error)
第 1 隠れ層の活性化関数	tanh
第 2, 第 3 隠れ層の活性化関数 (第 3 隠れ層は RNN-VTD のみ)	ReLU
出力層の活性化関数	linear

表 2. 各予測手法における最良の MSE

予測手法	MSE の値
RNN-V	0.28
RNN-VT	0.27
RNN-VTD	0.16

となった。以上のことから、隠れ層内の層数や隠れ層内の各層のニューロン数を増やして RNN モデルを複雑にすることは必ずしも予測精度向上に繋がらないことがわかった。

4. むすび

本研究では、ネットワークトラフィック予測におけるトラフィック量の属性情報 (時刻, 曜日) の利用の有効性を検証するために 2 つの予測手法: RNN-VT と RNN-VTD を提案した。RNN-VT はトラフィック量に加えて時刻情報を、RNN-VTD はそれらに加えて曜日情報を入力として用いる。実験の結果、1) RNN-VT は従来手法である RNN-V とほぼ同じ予測精度を示すこと、2) RNN-VTD は従来手法である RNN-V と比べておよそ 43% 程度予測精度を改善できることがわかった。このことから、時刻情報は予測精度の向上にさほど有効ではないものの、曜日情報は予測精度の向上に有効であると言える。

謝辞

本研究で使用したクラウド資源は、平成 29 年度国立情報学研究所クラウド活用実証実験において提供されたものである。

参考文献

- [1] Cisco Visual Networking Index, https://www.cisco.com/c/ja_jp/solutions/collateral/service-provider/visual-networking-index-vni/white_paper_c11-520862.html
- [2] T. P. Oliveira, J. S. Barbar, and A. S. Soares, "Computer network traffic prediction: a comparison between traditional and deep learning neural networks," *International Journal of Big Data Intelligence*, Vol. 3, No. 1, pp. 28--37, 2016.
- [3] K. Thompson, G. J. Miller, and R. Wilder, "Wide-area Internet traffic patterns and characteristics," *IEEE Network*, Vol. 11, Iss. 6, pp. 10--23, Nov./Dec. 1997.
- [4] Amazon Web Service, <https://aws.amazon.com/jp/>
- [5] Keras Documentation, <https://keras.io/ja/>
- [6] TensorFlow, <https://www.tensorflow.org/>
- [7] WIDE プロジェクト, <http://www.wide.ad.jp/>