4ZC-07

有線通信におけるネットワークの輻輳予測に用いる特徴量に関する検討

明石 季利子 †　　中尾 彰宏 ††　　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，様々な原因で引き起こされる．TCP によるデータ転送において，既に輻輳制御アルゴリズムが実装されている．しかし，輻輳によるスループットの低下はネットワーク機器依存の要因の他に，ユーザの行動が原因となるようなものも含まれ，それに対処するには従来の方法だけでは難しい．本研究では，ネットワークのトラフィックデータを用いて深層学習 LSTM モデルによる輻輳の時系列予測を行うことで，なるべく早く輻輳の検出をすることを目指す．加えて，有線通信時のパケット情報から，輻輳が起きることを高確率で予測できる特徴量を検討する．

2　実験

2.1　実験環境

実験用マシン 1 台をサーバ，5 台をクライアントとして iPerf を用いた TCP 通信をし，送受信されたパケットを各マシンでキャプチャする（図 1）．その際カーネルモニタ [1] を用いて通信時の cwnd を取得する．TCP 輻輳制御アルゴリズムは Reno を用いた．

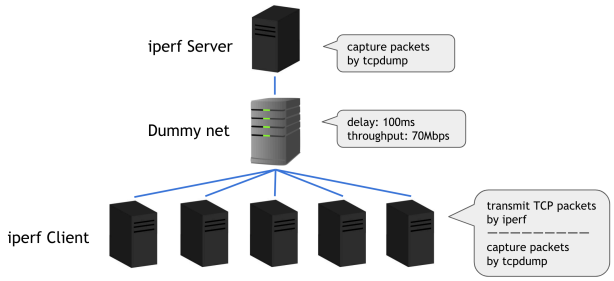

図 1: 実験環境

A Study on Features for Predicting Network Congestion in Wired Communications
†Kiriko Akashi　††Akihiro Nakao　†††Saneyasu Yamaguchi
†Masato Oguchi
†Ochanomizu University
††University of Tokyo
†††Kogakuin University

2.2　学習データとモデル作成

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 3 つの特徴量（パケット数，rwnd の値，cwnd の値）を導出したものとする．正解データは，時刻 t+1 秒の輻輳状態を 0,1 で表した値とする．本実験の輻輳状態とは，重複 ACK・高速再転送・重複受信・ACK ロスによる再送パケットをマークしたものとする．また，データの取得元をサーバ・クライアント・その両端に場合分けをして学習モデルを作成した．

3　性能評価

3.1　モデルの評価

パケットデータの取得元サーバ 3 種類，特徴量 4 種類のデータを用いて，モデルは全 12 種類作成した．学習モデルの正解率と損失の時系列グラフを図 2 に示す．

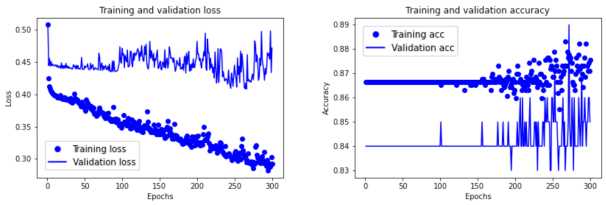

図 2: 両端のサーバで取得した cwnd を特徴量に用いた学習モデル（Loss と Accuracy）

3.2　学習モデルによる輻輳予測

全 12 種類のモデルを 100 秒間のバリデーションデータを用いて輻輳予測を行なった結果を図 3 に示す．

本実験では，学習モデルの評価のため混合行列を使用した．混合行列は，実際のクラスを縦軸，モデルが予測したクラスを横軸として表した行列である．輻輳状態である「Positive」もしくはそうでない「Negative」の 2 クラスに分類する場合の混合行列の各要素を表 1 に示す．

正解値と予測値の混合行列を基に，正解率・適合率（陽性と予測されたサンプルのうち正解したサンプルの割合）・再現率（実際に陽性のサンプルのうち正解したサンプルの割合）・F 値（適合率と再現率の調和平

Copyright ©2022 Information Processing Society of Japan.
All Rights Reserved.

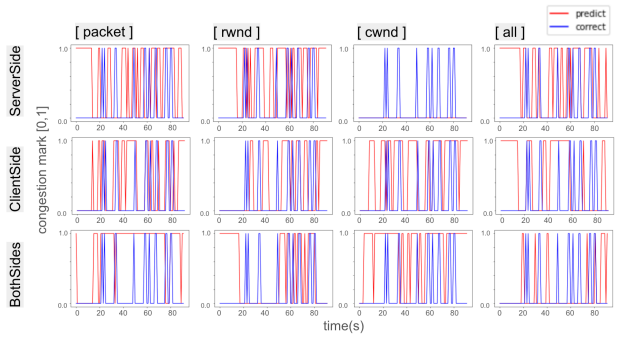

図 3: バリデーションデータの輻輳予測

表 1: 混合行列

実際＼予測	Positive	Negative
Positive	TP：真陽性	FN：偽陰性
Negative	FP：偽陽性	TN：真陰性

均）を算出する式をそれぞれ以下に示す．適合率の値が大きいほど，誤検知が少なく，再現率の値が大きいほど，検知漏れが少ないと判断できる．

- 正解率 = TP + TN / (TP + TN + FP + FN)
- 適合率 = TP / (TP + FP)
- 再現率 = TP / (TP + FN)
- F 値 = (2 ×適合率×再現率) / (適合率+再現率)

3.3 結果と考察

各モデルにおける 4 つの評価指標をグラフ化したものを図 4 に示す．

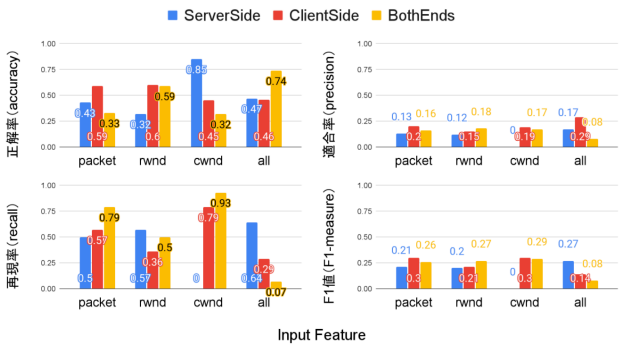

図 4: 学習モデルの性能評価

潜在的である輻輳発生を考慮すると，輻輳制御を行う通信事業者は，良性の活動が誤って輻輳が発生したとして認識される「偽陽性」よりも輻輳が検知されない「偽陰性」に対してより敏感であると考える．

サーバ側で取得した cwnd を学習したモデルは，正解率が 85%と最も高い数値であった．しかし，バリデーションデータの輻輳予測を見ると，輻輳が起きてないことを予測し続けることで正解率を高めていたことがわかる．再現率が 0 であり偽陰性が高いため輻輳予測としては良くないモデルであると言える．また，両端のサーバで取得した cwnd を用いたモデルは，再現率 93%と最も高い数値であり，輻輳予測として評価が高いモデルであると言える．しかし，輻輳が起きることを予測し続けることも予測として意味を成さないため，ある一定の正解率が担保される必要がある．

取得元サーバの比較において，再現率の観点から両端のサーバで取得したデータを用いた場合が，サーバ側やクライアント側のみのデータを用いた場合より精度が高いという結果が得られた．しかし，特徴量の比較において，両端のサーバで取得した packet や rwnd, cwnd を全て学習させたモデルは正解率は高いものの再現率が著しく低い結果となった．

4 まとめと今後の課題

4.1 結論

本研究は，輻輳の早期検知を目指し，ネットワークトラフィックデータを用いて深層学習モデル LSTM による輻輳の時系列予測を行った．輻輳が検知されない「偽陰性」を抑えることを評価軸とし，高確率で予測できる特徴量を検討した．再現率の観点から，iPerf サーバとクライアント両端で取得した cwnd の値を用いたモデルが最も偽陰性が低く，輻輳予測に適しているという結果が得られた．

4.2 今後の課題

輻輳が起きてないことを予測し続けることで正解率を高めてしまう結果を踏まえて，再現率が高いモデルの正解率をある一定の精度高めることでより最適化されたモデルを作成できると考えている．そのため，良性ラベルの精度が 50%以上であることを条件に，予測精度を最適化することを目標としてモデルの実験を行いたい．加えて，学習するトラフィックデータのその他の特徴量を追加して比較実験を行うことで，輻輳検知と関連性が高い特徴量を発見したい．

参考文献

[1] 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.

Copyright ©2022 Information Processing Society of Japan.
All Rights Reserved.