

機械学習の汎化性能獲得のための 特徴量選択に基づく RTP フロー分類手法の検討

RTP traffic identification based on tailored feature selection toward generalization in machine learning

持丸 雄匡¹ 妙中 雄三² 塚本 和也¹
Yuta Mochimaru Yuzo Taenaka Kazuya Tsukamoto

九州工業大学¹ 奈良先端科学技術大学院大学²
Kyushu Institute of Technology Nara Institute of Science and Technology

1 研究背景・目的

アプリケーションの多様化に伴い QoE(Quality of Experience) を活用したネットワーク制御が期待されている。また、SDN(Software Defined Network) を用いたアプリケーション単位の柔軟な制御を実現できている [1]。この QoE 計算法はアプリケーション毎に標準化されているため、QoE に基づくネットワーク制御にはアプリケーション識別が欠かせない。しかし、近年のペイロード暗号化によって直接的なアプリケーション識別が困難となったため、機械学習を用いたアプリケーション分類が検討されている。[2] では、IP アドレス等の特定のクラウドサーバの情報自体を特徴量としているが、多様なデバイスが接続する IoT 環境では、IP アドレスが頻繁に変化する事が予想されるため、汎化性能は期待できない。そこで本研究では、RTP(Real-time Transport Protocol) の動画フローを環境非依存で検出するための特徴量の検討を行う。具体的には、OpenFlow(OF) を用いて異なる環境で訓練データとテストデータを生成し、環境が異なる場合でも高い検出率を実現できることを示す。

2 提案手法

本研究では、RTP フロー開始直後 15 パケットの時間的挙動に着目し、OF の Packet In メッセージの到着間隔を元にパケット到着間隔を取得した。しかし、通常は到着間隔そのものは環境毎に異なるネットワーク遅延、ジッタ、パケットロスにより変化するため、本研究で目標とする汎化性能獲得は期待できない。そこで、提案手法では到着間隔の統計量に着目し、環境の影響を受けやすい最大値、中央値、最小値を除いた、平均値、分散、標準偏差を特徴量とした。また、アプリケーション独自の特徴といえるポート番号、プロトコル番号、パケットサイズも特徴量として使用する。このようにパラメータを適切に選択することで、訓練・テスト時間を削減しつつ、高い汎化性能を持つモデルを作成する。

3 評価実験

本実験でのデータセット作成から RTP フロー検出までの流れを図 1 に示す。本検証では提案手法の汎化性能を検証するため、表 1 に示すネットワーク環境を tc コマンドを用いて構築することで、訓練時とテスト時で異なる環境のデータセットを取得した。また、特定のクラウドサーバを学習する比較手法として L3 ヘッダ、L4 ヘッダに含まれる情報に加え、パケット到着間隔値を直接的

に特徴量として用いるモデルを用いた。

本実験では訓練時と異なる環境での汎化性能を評価するため、訓練データにテストデータを含む場合と含まない場合について検証を行った。得られた検出率 (F1 値) を表 2 に示す。結果として、比較手法よりも、環境に依存しない特徴量を選択、生成することが、環境非依存な汎化性能獲得に重要となることがわかった。また、これによって、訓練・テスト時間を比較手法に比べて約半減 (比較: 訓練 34.2 秒, テスト 0.52 秒, 提案: 訓練 14.7 秒, テスト 0.17 秒) 出来ることが分かった。

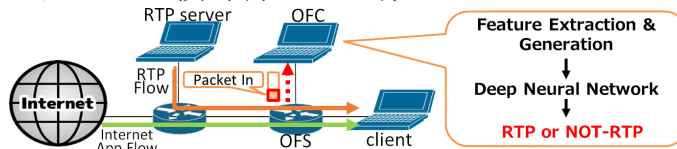


図 1 実験の構成

表 1 訓練・テスト時の環境パラメータ

	遅延	ジッタ	パケットロス率
訓練時	0 ms	0 ms	0 %
テスト時	350 ms	160ms	1 %

表 2 比較手法と提案手法の検出率 (F1 値)

	比較手法	提案手法
訓練データにテストデータを含む	100%	100%
訓練データにテストデータを含まない	70.0%	100%

4 まとめ

本研究では、訓練時とテスト時で異なる環境で取得したデータセットを用いて評価した結果、適切な特徴量選択・生成の重要性を示した。また、これにより、訓練・テスト時間を削減できることを示した。今後は、より高い汎化性能を持たせるための、訓練・テストで用いるデータセットやフローの割合を変化させて調査を行う。

参考文献

- [1] S. Shimokawa, et al., "SDN Based in-Network Two-Stage Video QoE Estimation With Measurement Error Correction for Edge Network," in IEEE Access, vol. 9, pp. 39733-39745, 2021.
- [2] A. Nakao, et al., "Toward In-Network Deep Machine Learning for Identifying Mobile Applications and Enabling Application Specific Network Slicing" IEICE TRANS. Comm., VOL.E101-B, NO.7, pp.1536-1543.

本研究は JSPS 科研費 JP21H03430 及び国立研究開発法人情報通信研究機構の委託研究の助成を受けたものです。