

機械学習を用いた複数のネットワーク間の 時系列データの補間に関する研究

A Study on Interpolation of Time Series Data Between Multiple Networks Using Machine Learning

會澤一輝
Kazuki Aizawa

渡部康平
Kohei Watabe

中川健治
Kenji Nakagawa

長岡技術科学大学 大学院工学研究科
Graduate School of Engineering, Nagaoka University of Technology

1 目的

ネットワークの研究において、実験的にネットワークのパフォーマンスを測定する際に、試験用のトラフィック生成のためトラフィックジェネレータが使用される。しかし、既存のトラフィックジェネレータでは、複雑な特性を持つネットワークのトラフィックの再現が困難な場合がある。そこで、機械学習を用いて複数の異なる実ネットワーク上のトラフィックデータを元に、新たなデータセットを生成する手法を検討する。

2 IP2Vec

IP2Vec [1] とは自然言語処理で利用される Word2Vec を元に考案された、ニューラルネットワーク (NN) を用いてカテゴリ値である IP アドレスを連続的特徴空間 \mathbb{R}^m 上のベクトルに変換する手法である。ネットワークにおける 2 つのホスト間の開始から終了までの一連の通信に関する情報を集約したフローベースデータについて、データセットに含まれる IP アドレス、ポート番号、プロトコルをベクトル表現に変換する。これにより、似た振る舞いをするホストの IP アドレス同士を \mathbb{R}^m 上の近い位置に存在するベクトルとして表現できる。

3 提案法

ネットワークのフロー情報をデータセットとし、2 つの異なるデータセットに対して別々に IP2Vec による学習を行う。学習により 2 つのデータセットから得られた特徴表現ベクトルの集合について、ニューラルネットワークによりベクトル間の変換を学習する。学習には IP アドレスの情報を利用せず、2 つのデータセット間で共通して表れるポート番号とプロトコルを同一の属性値として扱い、変換後のベクトルの差が小さくなるように学習を行う。提案法のモデルを図 1 に示す。

4 結果と考察

異常検出システム評価用のフローベースのデータセット CIDDs-001 の 2 日分のデータを分割し、別々のデータセットとして利用した。データセットに提案法を適用した結果を図 2 に示す。提案法を用いて 1 日目のデータの埋め込みを変換した結果が 1 段目、2 日目のデータの IP2Vec による埋め込みが 2 から 4 段目である。1 日目のデータセットに提案法を適用して得られたベクトルと、2 日目のデータセットの埋め込みの比較について、コサイン類似度の高いベクトルを示す。結果から、ベクトルの変換により同じ用途で用いられるホスト間で類似度が

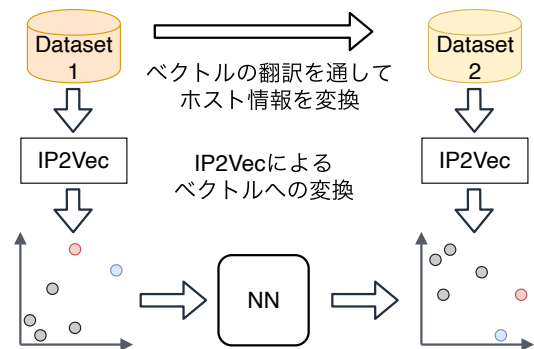


図 1 提案法のモデルの概要

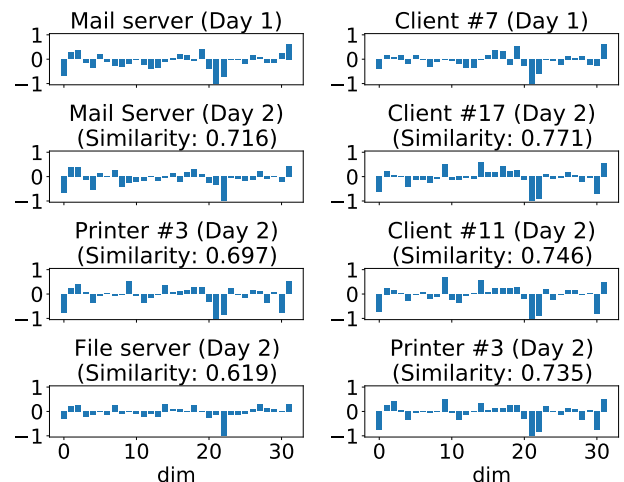


図 2 IP2Vec と提案法による埋め込みの変換

高くなることが確認された。

5 まとめ

Word2Vec による翻訳の手法と、IP2Vec によるカテゴリ値のベクトルへの変換を利用することで、ネットワーク上のホストの情報を他の時間のネットワークのベクトル空間上に変換できることを確認した。今後は、複数の異なるデータセットを用いて本手法の有用性を検討する。

6 謝辞

本研究の一部は科研費 20H04172 の助成を受けたものである。

参考文献

- [1] Ring, Markus *et al.* IP2Vec: learning similarities between ip addresses. In *ICDMW 2017*, 2017.