

トラヒックの時間変動を考慮した 動的トラヒックエンジニアリングのためのトラヒック予測

大歳 達也[†] 大下 裕一[†] 村田 正幸[†]

高橋 洋介^{††} 石橋 圭介^{††} 塩本 公平^{††}

[†] 大阪大学 大学院情報科学研究科

〒 565-0871 大阪府吹田市山田丘 1-5

^{††} 日本電信電話株式会社 NTT ネットワーク基盤技術研究所

〒 180-8585 東京都武蔵野市緑町 3-9-11

E-mail: [†]{t-otoshi, y-ohsita, murata}@ist.osaka-u.ac.jp, ^{††}{takahashi.yousuke, ishibashi.keisuke, shiomoto.kohei}@lab.ntt.co.jp

あらまし 時間変動の大きなトラヒックを収容するには、トラヒックの時間変動やネットワークの状態の変化に対して動的に経路を変更するトラヒックエンジニアリングと呼ばれる手法が有効である。しかし、各時刻のトラヒック量のみを考慮したトラヒックエンジニアリングでは、頻繁に大規模な経路変更が発生し、トラヒックを安定して収容することができない可能性がある。そのため、動的な経路変更を行う際にも、将来にわたるトラヒック変動を予測し、それを踏まえた制御が必要となる。本稿では、予測を用いた動的トラヒックエンジニアリングの入力として必要なトラヒック予測の要件を議論し、その要件に合致したトラヒック予測手法について検討を行う。予測手法の検討において、本稿では、実際のトラヒックには、予測可能な変動と、予測困難な変動が混在している点に注目し、予測の前処理として予測可能な変動のみを取り出す手法と、予測モデルの組み合わせについて評価を行う。そして、Internet2 で観測されたデータを用いた評価により、主成分分析を用いた SARIMA モデルによる予測が過大評価を小さく保ちつつ、過小評価を1%に抑えるということを明らかにする。

キーワード トラヒック予測, トラヒックエンジニアリング, データマイニング, 主成分分析, 季節自己回帰和分移動平均モデル

Traffic Prediction for Dynamic Traffic Engineering Considering Traffic Variation

Tatsuya OTOSHI[†], Yuichi OHSITA[†], Masayuki MURATA[†],

Yousuke TAKAHASHI^{††}, Keisuke ISHIBASHI^{††}, and Kohei SHIOMOTO^{††}

[†] Grate School of Information Science and Technology, Osaka University

1-5 Yamadaoka, Suita, Osaka, 565-0871, Japan

^{††} NTT Network Technology Laboratories, NTT Corporation

9-11, Midori-Cho 3-Chome Musashino-Shi, Tokyo 180-8585 Japan

E-mail: [†]{t-otoshi, y-ohsita, murata}@ist.osaka-u.ac.jp, ^{††}{takahashi.yousuke, ishibashi.keisuke, shiomoto.kohei}@lab.ntt.co.jp

Abstract Traffic engineering with traffic prediction is one approach to accommodate time-varying traffic adaptively. In this approach, the routes are calculated so as to avoid congestion based on the predicted traffic. The accuracy of the traffic prediction however has large impacts on this approach. In this paper, we investigate the traffic prediction methods suitable to traffic engineering, focusing on their underestimation that may cause the congestion. We evaluate the traffic prediction methods by using the actual traffic traces, and clarify that the SARIMA with the principal component analysis achieves the small underestimation without a large overestimation.

Key words Traffic Prediction, Traffic Engineering, Data Mining, Principle Component Analysis, Seasonal Autoregressive Integrated Moving Average Model

1. はじめに

近年、ストリーミング配信やクラウドサービス等のインターネットを介したサービスが普及するにつれて、ネットワークを流れるトラフィックの時間変動は大きくなっている。バックボーンネットワークは、大きなトラフィック変動が生じた場合にも輻輳を生じることなく全トラフィックを收容する必要がある。

輻輳を発生させずに時間変動の大きなトラフィックを收容する方法として、トラフィックエンジニアリングと呼ばれる手法に関する研究が進められている [1-3]。トラフィックエンジニアリングでは、定期的なトラフィックの観測を行い、観測されたトラフィックを收容できるように経路を動的に変更する。しかしながら、観測されたトラフィック量のみを考慮した経路制御では、経路変更後に生じたトラフィック変動への対応は次の経路制御時となり、輻輳の発生を防ぐことはできない。トラフィックの観測・制御の間隔を短くすることにより、トラフィック変動への対応速度を向上させることは可能であるものの、頻繁な観測・制御はネットワークの不安定化、観測・制御負荷の増加といった問題を招く。観測値から外れたトラフィックも考慮することで、観測値に対する最適化を行いつつ、トラフィック変動の影響を抑える方法も提案されている [3] もの、この方法では実際には発生しないトラフィックも考慮してしまうため、余分な資源も必要とってしまう。

時間変動の大きなトラフィックを安定的に收容する方法として、観測したトラフィックを基に将来のトラフィックを予測し、予測されたトラフィックを考慮に入れたトラフィックエンジニアリングが考えられる。トラフィック予測を考慮に入れたトラフィックエンジニアリングを行うことにより、予測を外れた変動が発生しない限りは頻繁な経路変更は必要なく、安定してトラフィックを收容することが可能となる。トラフィック予測を用いたトラフィックエンジニアリングでは、トラフィックエンジニアリングに適した精度のトラフィック予測を行うことが重要となる。

ネットワークのトラフィック予測に関しては、多数の研究がなされている [4-9]。文献 [4] では、コンマ数秒から数分間隔の短期的な予測についてニューラルネットワークとファジィ論理式を組み合わせた α -SNF モデル、ARMA モデル、ARIMA モデルを用いて学習期間やデータの粒度等の予測のパラメータと予測の精度の関係を調べている。文献 [5] では、ウェーブレット変換による多重解像度分析を用いて、長期的な変動を取り出し、ARIMA モデルを用いた予測を行うことで、トラフィック変動の上端と下端を含めた 6ヶ月先の予測が可能であることを示している。しかしながら、従来のトラフィック予測の研究では、予測誤差を小さくすることのみを目的としており、トラフィックエンジニアリングに用いるという観点での評価はされていない。トラフィックエンジニアリングでは、トラフィック量が過小評価された場合、輻輳が発生する可能性があるため、過小評価は避ける必要がある。一方、過大評価は多めの資源を要求するものの、ネットワークの通信品質に与える影響は少なく、過小評価ほどの影響はない。本稿では、トラフィックエンジニアリングに用いるという観点において、上記の点を考慮した上で、トラフィックエンジニアリングに適したトラフィック予測手法を検討する。

以降の本稿の構成は次の通りである。2章では、トラフィック

エンジニアリングにおける予測と制御の連携について述べる。3章では、予測を行う際の予測前の処理と、予測モデルについて述べ、4章では、各予測方法の評価方法を述べる。5章では、評価の結果を示し、6章で、まとめと今後の課題について述べる。

2. 予測を用いたトラフィックエンジニアリング

予測を用いたトラフィックエンジニアリングでは、定期的なトラフィックの観測・予測を行い、予測される輻輳を回避するように経路制御を行う。

このアプローチにおいてトラフィックエンジニアリングの制御間隔より短い間隔で発生するトラフィック変動の予測は重要である。制御間隔よりも短い間隔で発生する変動にトラフィックエンジニアリングで対応することはできない。短い間隔で発生する変動を正確に予測することにより、その変動により輻輳が生じないような経路を設定することが可能となる。

一方、トラフィックエンジニアリングにおいて、数時間後から1日後といった、より長い時間粒度のトラフィック予測を用いることも有用である。経路の変更時には、経路が収束するまでの間、パケットのロス率が高まるため、多くのトラフィックに影響を与えるような大規模な経路変更は避ける必要がある。トラフィックエンジニアリングを行う際に、長い時間粒度のトラフィック予測を用いることにより、将来発生しうるトラフィック変動を踏まえた経路を設定することができ、大規模な経路変更が頻繁に発生することを防ぐことができる。

トラフィック予測を用いたトラフィックエンジニアリングでは、上記のトラフィック予測を用い、予測されたトラフィックが流れた場合に輻輳が発生しないように経路を設計する。実際のトラフィック量が予測されたトラフィック量よりも大きい場合は、予想外の輻輳が発生してしまうこととなるため、過小評価の影響は大きい。それに対して、実際のトラフィック量が予測されたトラフィック量よりも小さい場合は、通信品質への影響はない。そこで、本稿では、過小評価を可能な限り小さくすることに焦点を当て、予測手法の検討を行う。

また、トラフィック予測を用いたトラフィックエンジニアリングでは、各時刻に入手した情報をもとにトラフィック予測をした上で、適切な経路を計算する。そのため、トラフィック予測もトラフィックエンジニアリングを行う間隔に合わせて周期的に行う必要があり、計算時間のかかるトラフィック予測手法を用いることはできない。そこで、本稿では、予測の精度のみではなく、予測にかかる計算時間についても検討を行う。

3. トラフィック予測方法

一定の時間間隔ごとに、各フローのトラフィック量を集計することで、フローのトラフィック変動が表現される。 i 番目の時間間隔におけるフロー j のトラフィック量を第 i, j 成分 x_{ij} とした行列 X により、フローのトラフィック量が表現される。この時、 X はトラフィックマトリクスと呼ばれ、 X の第 j 行 X_j は、フロー j の時間変化を表す。トラフィック予測を用いたトラフィックエンジニアリングは、このトラフィックマトリクスで表される全フローの時間変動の適切なリンクへの收容を試みる。そのため、トラフィックエンジニアリングでは、トラフィックマトリクスの予測を行うことが必要となる。

3.1 予測前の処理

実際のトラヒック変動では、周期的な変動や、長く安定した変動のような予測の容易なもの、ノイズやスパイクのような予測の困難な変動が混在している。予測の困難な変動を除き、予測の容易な変動を取り出すことで、精度よく予測できると考えられる。

本稿では、予測の容易な変動を取り出す手法として、以下の3つの予測前処理を検討する。

- 全フローの特徴を少数のフローで表す主成分分析
- フロー毎にノイズ等の短期変動を除いたトレンド成分
- フロー毎にトラヒックの上端の変化を追従する包絡線

3.1.1 主成分分析

主成分分析は、多数の変数を含むデータから主要な成分を取り出す手法である。主成分分析をトラヒックマトリクスに対して行うことで、フロー全体を少数のフローで近似できる [10]。主成分分析を前処理として行った上でトラヒック予測を行うことで、主要なトラヒック変動にのみ注目した予測が可能となり、ノイズの影響を除去し、予測精度を向上させることができると考えられる。また、大規模なネットワークでは、全てのフローを個別に予測すると膨大な計算時間を必要とするが、主成分分析により抜き出された少数のフローのみを予測することで、予測にかかる計算時間を短縮することも可能となる。

主成分分析を用いるためには、時間間隔の数 t と、フローの数 p に対し、 $t > p$ である必要がある。これは、十分長い期間のデータを用いることで容易に達成可能であるため、以降では、これが成り立つものとして考える。

主成分分析では、フローの線形和により、 p 個の新たなフローを生成する。生成されたフローは固有フローと呼ばれ、この時、元のフローの係数は次式に従う。

$$\mathbf{v}_k = \arg \max_{\|\mathbf{v}\|=1} \left\| \left(X - \sum_{i=1}^{k-1} X \mathbf{v}_i \mathbf{v}_i^T \right) \mathbf{v} \right\|, k = 1, \dots, p \quad (1)$$

ここで、 X はトラヒックマトリクスであり、 \mathbf{v}_k は第 k 番目の固有フローに対する係数ベクトルで第 k 番目の主成分と呼ばれる。また、 $\|\cdot\|$ はユークリッドノルムを意味する。この時、係数に従って元のフローの時系列の線形和を取り、正規化した $\mathbf{u}_k = \frac{X \mathbf{v}_k}{\|X \mathbf{v}_k\|}$ が第 k 番目の固有フローである。

式 (1) により、既に捉えた変動を除いた上で、最も大きく変動を捉える係数が順に求まる。よって、固有フロー \mathbf{u}_k は k 番目に X の中の大きな変動を捉えたフローとなる。

本稿では、文献 [10] と同様に、主成分分析を行う前に、各フロー j のトラヒック時系列 X_j の平均 μ_j を用いて、 $X_j - \mu_j$ と中心化する。これは、主成分分析により、単純に大きなトラヒック量を持っているものが上位に取り出されることを防ぐためである。

式 (1) より、既に捉えられた変動を含まないように、主成分 \mathbf{v}_k が求められるため、 \mathbf{v}_k は互いに直交する。よって、固有フロー \mathbf{u}_k は、元のフローを直交座標変換したものであり、逆変換が存在し、固有フローから元のフローを再構成することができる。このとき、小さな変動である下位の固有フローを無視した、上位 r 個の固有フローのみによる X の近似が次式により求まる。

$$X \approx \sum_{k=1}^r \|X \mathbf{v}_k\| \mathbf{u}_k \mathbf{v}_k^T \quad (2)$$

予測を行う際には、上位 r 個の固有フローのみを予測することで、全フローの主要な変動の予測を行うことができる。

トラヒックエンジニアリングに用いるためには、トラヒック量の過小評価は避ける必要がある。そのため、式 (2) で、除外した変動も考慮に入れ、過小評価を避けるような予測値を得る必要がある。本稿では、式 (2) で、除外した変動については、変動の予測を行わず、変動の大きさのみを見積もることを考える。これは、各フロー i について、フロー i の近似 X'_i による近似誤差の標準偏差 $\sigma_i = \sqrt{V[X_i - X'_i]}$ を用いて行う。ここで、 $V[\cdot]$ は分散を意味する。上位 r 個の固有フローの予測と再構成による X'_i の予測値を X''_i とすると、 $X''_i + \alpha \sigma_i$ が式 (2) で除外された変動の大きさも含めたフロー i の予測値となる。ここで α は、トラヒックの変動をどれだけ安全側に見積もるかを表すパラメータである。

3.1.2 トレンド成分

トラヒックの変動には、ノイズのような短期的な変動、一日周期の変動、長期的な増減傾向等、様々な時間スケールの変動が含まれる。注目する時間スケールの変動を取り出すことで、より精度の良い予測が可能になると期待される。ここでは、文献 [11] のトレンドモデルを用いて、データから、ノイズによる短期的な変動を取り除き、トレンド成分として長期的な変動を取り出す。トレンド成分は、以下の式により定式化される。

$$x_k = t_k + \epsilon_k \quad (3)$$

$$\Delta t_k = \Delta t_{k-1} + w_k \quad (4)$$

ここで、 x_k は元のデータ、 t_k はトレンド成分であり、 $\Delta t_k = t_k - t_{k-1}$ である。 $\epsilon_k \stackrel{\text{i.i.d.}}{\sim} N(0, \sigma^2)$ と $w_k \stackrel{\text{i.i.d.}}{\sim} N(0, \tau^2)$ は観測誤差と、トレンドの変化の誤差で、互いに独立である。

式 (3) は、観測されるデータがトレンド成分と、誤差の和として表されることを示す。式 (4) は、トレンド成分の差分が誤差を伴って変化することを示している。

実際にトレンド成分を計算するには、まず、 σ, τ を最尤法により推定する。 σ, τ が決定すると、式 (3), (4) から求まる遷移確率から各時刻でのトレンド成分の確率分布が決定するため、観測値の条件付期待値 $E[t_i | x_1, x_2, \dots, x_i]$ として、トレンド成分 $t_i (i = 1, 2, \dots, t)$ を求める。そして、主成分分析の場合と同様に、 $t_k + \alpha \sigma$ として求めた値を各時刻のトラヒック変動におけるトレンド成分値とすることにより、取り除いた変動を考慮する。

3.1.3 包絡線

トラヒックの上端の変化を取り出し、予測することで、短期的なトラヒック変動も収容するために必要な帯域を予測可能であると考えられる。ここでは、上端の変化を一定間隔に区切った区間ごとのピークの変化を追うことで取り出すこととする。ピークを求める間隔を s とすると、観測値 x_1, \dots, x_t は、 $l = \frac{t}{s}$ 個の区間に分割される。分割された各区間に属する時刻の集合 I_k と、各 I_k でのピークの時刻 p_k は次式で表される。

$$I_k = \{(k-1)s + 1, \dots, ks\}, k = 1, \dots, l \quad (5)$$

$$p_k = \arg \max_{i \in I_k} x_i \quad (6)$$

この時、各区間におけるピークの値は x_{p_k} となる。予測に最新のデータを反映させるために、 x_t も含めた各ピークの値 $x_{p_1}, \dots, x_{p_l}, x_{p_{l+1}} = x_t$ を結ぶ線を包絡線とする。ただし、 x_{p_l} と x_t が重複する場合は、 x_t を除いて考える。各ピーク間の $p_k - p_{k-1} - 1$ 個の点は、線形補完によって補完することとし、包絡線の各点は次式により計算される。

$$x_i = x_{p_{k-1}} + \frac{x_{p_k} - x_{p_{k-1}}}{p_k - p_{k-1}} (i - p_{k-1}) \quad (7)$$

$$i = p_{k-1}, \dots, p_k, k = 2, \dots, l + 1$$

3.2 予測

各前処理によって、トラヒックの特徴を取り出した後、予測を行う。予測は予測モデルを元に行う。本稿では、予測モデルとして、ARIMA モデルと Seasonal ARIMA (SARIMA) モデルを用いる。予測モデルに基づく予測では、入力されたデータを元に、モデルの各パラメータを学習した後、得られたモデルに従って、将来の値を計算する。以下、ARIMA モデルと SARIMA モデルの概要を説明する。

3.2.1 ARIMA モデル

各時刻のデータを、過去の数ステップのデータと、誤差の線形和から表すモデルとして、ARMA モデルがあり、次の式で表される。

$$x_n = \sum_{i=1}^p a_i x_{n-i} + \sum_{i=0}^q b_i \epsilon_{n-i} + c \quad (8)$$

$$b_0 = 1$$

ここで、 p, q は何ステップ過去までのデータ・誤差を用いるかを表し、 a_i, b_i は i ステップ前のデータ・誤差に対する依存係数、 c は定数項である。

しかし、ARMA モデルは、データの定常性を仮定しており、平均値が変動する等、非定常なデータに対しては、適用できない。そこで、データを定常化させる処理が必要となる。データの平均が線形な変化をする場合、データの差分 $\Delta x_n = x_n - x_{n-1}$ を取ることで、平均の変化が除去される。同様に、 d 次の多項式に従う平均の変化傾向の場合は、 d 回繰り返して差分を取ることで除去できる。差分を d 回繰り返す処理を Δ^d で表すこととする。データの差分 $\Delta^d x_n$ を取り平均値の変動を除き、データを定常化してから ARMA モデルを適用するモデルは、ARIMA モデルと呼ばれる。

データに対するモデルの当てはめは、まず d の決定から行う。データの定常性の検定を行い、非定常と判定されると、データの差分を取り、再び定常性の検定を行う。これを繰り返し、データが定常となった時点の d を用いる。定常性の検定は、単位根過程と呼ばれる非定常な過程 $x_t = x_{t-1} + \epsilon$ に従うか否かを統計的検定により検定する。ここでは、 ϵ の分散 σ^2 が 0 となる定常過程を帰無仮説とした KPSS 検定 [12] を用いて、検定を行う。

p, q の決定は、 p, q を与えた時に、最尤法で得られるモデルの良さが最も高くなる p, q の探索によって行う。モデルの良さの基準には、次式で表される赤池情報量基準 (AIC) を用いる。

$$AIC = -2 \log L + 2k \quad (9)$$

ここで、 L は最尤法により、最大化された尤度で、最大尤度と

呼ばれる。 k は、パラメータの数であり、ここでは、 $k = p + q$ である。パラメータ数を増やすことにより、よりデータに適合し、 L が大きくなるが、ノイズのような偶発的な変動も追従する可能性が高くなる。AIC を用いた場合、パラメータ数 k によりペナルティを科すことで、過剰適合を避けたモデルを良いモデルとして選択する。 p, q の探索は、文献 [13] の手法を用い、 p, q を 1 ずつ変化させたモデルの中で、AIC が改善されるモデルが得られなくなった時点で終了する。

モデルの各パラメータの推定は以下のようにになる。

(1) d の値を決定

- (a) データの定常性検定を行う
- (b) 定常なら d を決定し、終了する
- (c) データの差分をとり、(a) に戻る

(2) p, q とモデルの決定

- (a) p, q の値を指定する
- (b) 最尤法によるデータへ当てはめを行い、誤差の分散、係数を決定する
- (c) 得られたモデルの AIC を評価する
- (d) AIC が改善されなければ終了する
- (e) (a) に戻る

3.2.2 Seasonal ARIMA モデル

SARIMA モデルは、データの周期性を考慮した ARIMA モデルの拡張である。周期毎の平均の変化等による非定常性を取り除くために、1 周期前のデータとの差分 $\Delta_s x_n = x_n - x_{n-s}$ を取る。ここで、 s はデータの周期である。ARIMA と同様に、周期的な差分を繰り返し行うことで、データの定常化を行い、周期的な差分を D 回繰り返す操作を Δ_s^D で表す。周期的なデータの定常化を行った後は、ARIMA と同様に直前のデータとの差分による定常化を行う。よって、データの定常化は、 D 回の周期的な差分と、 d 回の差分を取った $\Delta^d \Delta_s^D x_n$ によって行われる。トラヒック変動では 1 日毎の周期的な変化が見られるため、 s は 1 日とする。

定常化後のデータに対しては、周期性を考慮した ARMA モデルの拡張を適用する。この拡張されたモデルは SARMA モデルと呼ばれる。SARMA モデルでは、各時刻のデータが局所的な過去のデータ・誤差に加え、過去の周期におけるデータ・誤差の線形和として表され、次式に従う。

$$x_n = \sum_{i=1}^p a_i x_{n-i} + \sum_{i=0}^q b_i \epsilon_{n-i} + c + \sum_{j=1}^P A_j \sum_{i=1}^p a_i x_{n-s_j-i} + \sum_{j=1}^Q B_j \sum_{i=0}^q b_i \epsilon_{n-s_j-i} \quad (10)$$

$$b_0 = 1$$

ここで A_j, B_j は j 周期前のデータ・誤差に対する依存係数である。各パラメータの推定は、ARIMA モデルと同様の方法で行う。ただし、 D を決定する際の非定常性の検定には、周期的な変動上での非定常性を検定する。ここでは、 $x_k = \sum_{j=1}^{s/2} (\gamma_{k,2j-1} \cos \frac{k}{s} 2\pi j + \gamma_{k,2j} \sin \frac{k}{s} 2\pi j) + \epsilon_k$ と、三角関数の線形和でデータを表現した時の、係数 $\gamma_{k,i}$ の変化により周期ごとの変化を見る。この時の $\gamma_{k,i}$ が単位根過程 $\gamma_{k,i} = \gamma_{k-1,i} + \epsilon_k$ に従うか否かを、 ϵ_k の分散が 0 となる定常過程を帰無仮説とした、Canova-Hansen 検定 [14] により検定

する。

4. 評価方法

4.1 評価に用いるデータ

評価には、米国の研究教育ネットワーク Internet2 のバックボーンネットワークのトラフィックデータ [15] を用いて行う。このトラフィックデータでは、NetFlow を用いてトラフィックの計測が行われており、計 9 個の Point of Presence(PoP) のルータのインターフェースで 100 パケットに 1 パケットの割合でサンプリングが行われる。サンプリングされたパケットは 5 分毎にまとめて出力される。

本稿では、2011/11/28–2011/12/25 までの 4 週間分のデータを用い、BGP のルーティング情報を用いて、PoP 間のフローに集約したものを扱う。

4.2 評価の手順

評価の手順は以下の通りである。

- (1) 予測開始時点から過去 2 週間のデータを取り出す
- (2) 取り出したデータに対し、それぞれの前処理を行う
- (3) 前処理を行った後のデータを用いて予測を行う
- (4) 予測したフローをリンク上にマッピングする
- (5) 予測誤差の評価を行う
- (6) 予測開始時点をずらしながら、1~5 を繰り返す

1 日の周期性を考慮する SARIMA モデルでは、ARIMA モデルより長い時間スケールの予測に向いている。一方、1 時間以内といった比較的短いスケールのトラフィック変動を予測する際には、周期性を考慮することの効果は薄く、ARIMA モデルで十分な予測が可能であると考えられる。そこで、本稿では、短期の予測に ARIMA モデル、長期の予測に SARIMA モデルを用いるものとし、ARIMA モデルでは、5 分毎の観測データを用い、SARIMA モデルでは、3 時間粒度にまとめた観測データを用いる。

予測結果は予測開始時点の変動に大きく影響を受けるため、本稿では、予測時点をずらした予測を ARIMA モデルでは 72 回、SARIMA モデルでは 56 回行った。以降の評価では、予測誤差は、得られた結果の平均値を用いて計算する。

また、トラフィックエンジニアリングへの適用を考えると、各フローの変動の予測精度よりも、特定の経路に收容した際に輻輳が発生するかを予測できるかが重要となる。そこで、本稿では、予測誤差の評価はフロー単位ではなく、経路に従って各フローをリンクに收容した時に流れるリンク上のトラフィック量の予測誤差を評価する。ここで、経路は評価開始時点の最大リンク利用率が最小となる経路を用いる。

トラフィックエンジニアリングにトラフィック予測を用いた場合、トラフィックの過小評価は予想外の輻輳を生じるなどの影響が大きい。そのため、過小評価を可能な限り少なくすることが必要となる。そこで、本稿では、予測値を用いるのではなく、予測の信頼区間を用い、トラフィックの変動の上方を予測したものを予測値として用いることで、輻輳の発生を避けることを考える。この時、予測誤差の信頼区間を広くとることで、過小評価を減少させ、輻輳を避けることができるが、過大評価は増加し、資源の利用効率は低下する。本評価では、各手法において、各リンクを流れるトラフィック予測値の過小評価の最大値をそのリ

ンクのトラフィック流量の平均値の 0.01 倍以下に抑えるように、各信頼区間を設定した。これは、フローの平均値に対して、各時刻で最大 1% までの過小評価を許容することを意味する。また、その際に、主成分分析やトレンド成分を抽出する前処理における α の値も、予測の信頼区間と同じ水準となるように定めるものとした。

本評価では、発生する過大評価の相対値を評価に用いる。過大評価の相対値は、次式で定められる。

$$e_{ij} = \begin{cases} \frac{\hat{y}_{ij} - y_{ij}}{\mu(\mathbf{y}_j)} & (\hat{y}_{ij} > y_{ij}) \\ 0 & (\hat{y}_{ij} \leq y_{ij}) \end{cases} \quad (11)$$

ただし、 \hat{y}_{ij} はリンク j 上の i 期先のトラフィック量の予測値、 y_{ij} は実際のトラフィック量、 $\mu(\mathbf{y}_j)$ はリンク j 上の実際のトラフィック量の平均値である。

5. 評価結果

5.1 予測誤差の時間変化

図 1 に最大の過小評価を 0.01 に保つように、信頼区間の幅を決定した時の、各時刻の過大評価を示す。図 1(a), 1(b) は、それぞれ、SARIMA モデルと ARIMA モデルを用いた場合の誤差を示す。横軸の i の最大値は 7 日と、12 時間であり、最小値は、予測の最小粒度である 3 時間と 5 分としている。また、グラフ中の過大評価は全フロー・全予測における平均値である。図中の Original は元のフローに対して予測前の処理を何も行わず、直接予測を行った場合を指し、Trend, Envelope はそれぞれトレンド成分、包絡線を取り出した後予測を行った場合、PCA は主成分分析により取り出された固有フローを予測し、その予測値から、元のフローを式 (2) に従って再構成した場合を示す。また、PCA($r = k$) はフローの再構成に上位 k 本の固有フローの予測値を用いることを意味し、Envelope の後の数値は、包絡線を取り出すピーク値の間隔を示す。

図 1(a) より、SARIMA モデルを用いた予測では、PCA($r = 5$) が過大評価を抑えていることが分かる。しかし、短い期間の予測では、元のフローの予測の方が過大評価を抑えることができてきている。これは、短い期間の予測では、主成分として抽出された固有フローに含まれない変動の方が、予測誤差よりも大きく影響するためと思われる。図 1(b) の、直後のデータの予測に関しては、包絡線が最も過大評価を抑えていることが分かる。これは、直後の予測に関しては包絡線が元のフローの予測よりも、変動の上端をうまく予測できているためと思われる。しかし、予測する期間を長くすると、包絡線の予測が困難となり、過大評価が急激に増加し、元のフローの方が過大評価を抑えている。

5.2 予測の計算時間

予測された変動を元にトラフィックエンジニアリングを行う場合、経路の最適化の処理の前に、トラフィック予測の処理が加わり新たな負荷となる。特に大規模なネットワークとなると、トラフィック予測に関わる計算時間も大きくなると考えられる。このため、予測の計算時間を抑えつつ、適切な予測を行える予測方法が望ましい。表 1, 2 に SARIMA モデルと ARIMA モデルの予測と、予測前の処理に要した合計時間を示す。ただし、表中の時間は 1 回の予測更新当たりの平均の時間とする。計

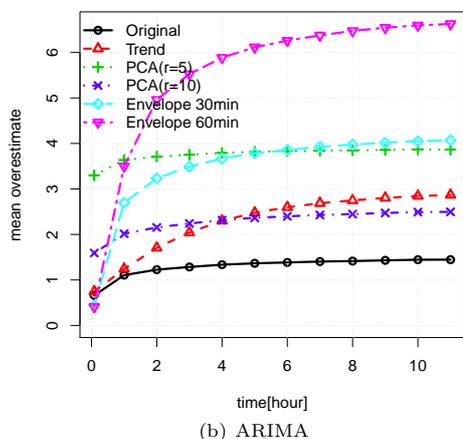
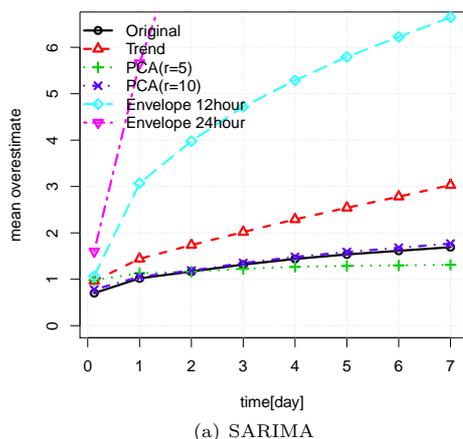


図 1 過小評価を 0.01 に抑えた時の i 期先の過大評価

表 1 予測に要した計算時間 [秒](SARIMA, 一点の時間粒度:3 時間)

包絡線 (12h)	包絡線 (24h)	トレンド成分	PCA($r=5$)	PCA($r=10$)	元のフロー
73.9	76.2	181	3.3	9.98	61.6

表 2 予測に要した平均の計算時間 [秒](ARIMA, 一点の時間粒度:5 分)

包絡線 (30m)	包絡線 (60m)	トレンド成分	PCA($r=5$)	PCA($r=10$)	元のフロー
61.5	59.6	101	2.09	4.16	60.0

算は、Intel Core i7-2600S の CPU を搭載したコンピュータ上で、統計処理ソフト R を用いて行った。

表 1, 2 より、主成分分析を用いた場合の計算時間が最も短くなっている。これは、他の予測方法では、72 個の全ての PoP 間のフローについて予測を行うのに対し、主成分分析を用いた場合、予測を行う対象が、上位 5, 10 個の固有フローで済み、主成分分析自体や元のフローの再構成にも時間がかからないためである。一方、トレンド成分の計算時間が最も長くなっているのは、トレンド成分を求める計算時間自体が長いためである。

6. まとめ

本稿では、トラフィック予測とトラフィックエンジニアリングを組み合わせるために、トラフィックエンジニアリングに必要なトラフィック変動を予測する方法について評価した。主成分分析を用いて、上位少数個の固有フローを取り出し予測することで、

計算時間の短縮と同時に、トラフィックの主要な変動が予測でき、長期的に過小評価と過大評価が共に小さい予測が行えることが分かった。しかし、直近の予測では包絡線、短期の予測では元のフローを直接予測の方が優っている。予測する期間の違いによって、精度の良い予測が異なっていることから、複数のスケールで異なる予測方法を用い、組み合わせることで、予測の精度が高められると期待される。

今後の課題としては、各タイムスケールに合わせた予測方法の定め方の検討と、予測されたトラフィック変動を元にトラフィックエンジニアリングを行った場合の評価が挙げられる。

文 献

- [1] N. Wang, K. H. Ho, G. Pavlou, and M. Howarth, "An overview of routing optimization for Internet traffic engineering," *IEEE Communications Survey & Tutorials*, vol. 10, no. 1, pp. 36–56, first quarter 2008.
- [2] D. O. Awduche and U. (mci Worldcom), "MPLS and traffic engineering in IP networks," *IEEE Communications Magazine*, vol. 37, pp. 42–47, Dec. 1999.
- [3] H. Wang, H. Xie, L. Qiu, Y. R. Yang, Y. Zhang, and A. Greenberg, "COPE: traffic engineering in dynamic networks," in *Proceedings of SIGCOMM*, vol. 36, no. 4. New York, NY, USA: ACM, Aug. 2006, pp. 99–110.
- [4] M. F. Zhani, H. Elbiaze, and F. Kamoun, "Analysis and prediction of real network traffic," *Journal of Networks*, vol. 4, no. 9, pp. 855–865, Nov. 2009.
- [5] K. Papagiannaki, N. Taft, Z.-L. Zhang, and C. Diot, "Long-term forecasting of Internet backbone traffic: Observations and initial models," in *Proceedings of INFOCOM*, vol. 2, Mar. 2003, pp. 1178–1188.
- [6] B. Zhou, D. He, Z. Sun, and W. H. Ng, "Network traffic modeling and prediction with ARIMA/GARCH," in *Proceedings of HET-NETs*, Sep. 2006, pp. 1–10.
- [7] H. E. Hag and S. Sharif, "An adjusted ARIMA model for Internet traffic," in *Proceedings of AFRICON*, Sep. 2007, pp. 1–6.
- [8] L. Xiang, "A new hybrid network traffic prediction method," in *Proceedings of GLOBECOM*, Dec. 2010, pp. 1–5.
- [9] M. L. F. Miguel, M. C. Penna, J. C. Nievola, and M. E. Pellenz, "New models for long-term Internet traffic forecasting using artificial neural networks and flow based information," in *Proceedings of NOMS*. IEEE, Apr. 2012, pp. 1082–1088.
- [10] L. Anukool, P. Konstantina, C. Mark, D. Christophe, E. D. Kolaczyk, and T. Nina, "Structural analysis of network traffic flows," in *Proceedings of SIGMETRICS/Performance*, vol. 32, no. 1. New York, NY, USA: ACM, Jun. 2004, pp. 61–72.
- [11] G. Kitagawa and W. Gersch, "A smoothness priors-state space modeling of time series with trend and seasonality," *Journal of the American Statistical Association*, vol. 79, no. 386, pp. 378–389, Jun. 1984.
- [12] D. Kwiatkowski, P. C. Phillips, and P. Schmidt, "Testing the null hypothesis of stationarity against the alternative of a unit root: How sure are we that economic time series have a unit root?" *Journal of Econometrics*, vol. 54, no. 1–3, pp. 159–178, Oct–Dec 1992.
- [13] R. J. Hyndman and Y. Khandakar, "Automatic time series forecasting: The forecast package for R," *Journal of Statistical Software*, vol. 27, no. 3, pp. 1–22, Jul. 2008.
- [14] F. Canova and B. E. Hansen, "Are seasonal patterns constant over time? A test for seasonal stability," *Journal of Business & Economic Statistics*, vol. 13, no. 3, Jul. 1995.
- [15] "Internet2 data," available from <http://internet2.edu/observatory/archive/data-collections.html>.