

休日の特徴を考慮したJPIXネットワークトラフィックの高精度週間予測法に関する研究

UEDA, Takayuki / 上田, 崇介

(出版者 / Publisher)

法政大学大学院理工学・工学研究科

(雑誌名 / Journal or Publication Title)

法政大学大学院紀要. 理工学・工学研究科編

(巻 / Volume)

56

(開始ページ / Start Page)

1

(終了ページ / End Page)

5

(発行年 / Year)

2015-03-24

(URL)

<https://doi.org/10.15002/00011229>

休日の特性を考慮した JPIX ネットワークトラフィックの 高精度週間予測法に関する研究

A STUDY ON WEEKLY JPIX TRAFFIC VOLUME PREDICTION CONSIDERING DIFFERENCES IN DAILY TRAFFIC CHARACTERISTICS

上田 崇介
Ueda Takayuki
指導教員 八名和夫

法政大学大学院理工学研究科応用情報工学専攻修士課程

Internet has been established as one of infrastructures indispensable for global social and economic activities. According to the investigation of Ministry of Internal Affairs and Communications, penetration rate of Internet is 84.9% per one household at the year-end 2014 in Japan. Estimated Internet traffic volume reaches at 2.89 Tbps (May 2014) which showing 10.7% increase in a year. The internet traffic prediction is getting more and more important for appropriate Internet traffic management. This paper proposes a method for weekly internet traffic on JPIX (JaPan Internet eXchange), one of the largest Internet backbone in Japan. Singular value decomposition is applied to the daily JPIX traffic data updating every 5 minutes. The data were taken from graphical data presented at JPIX's web page. Two orthogonal vectors are found to be useful for efficient traffic prediction. The prediction is made based on monthly trend of weight vectors. The trend has been examined in detail and the best prediction scheme has been proposed. Three classes, *i.e.* weekday, weekend and holiday season showed characteristic patterns useful for prediction. Adaptive weight update algorithm has been also proposed to improve the short term prediction accuracy. The method will be useful for network routing management in daily bases.

Keywords: network traffic, IX traffic, network traffic prediction, singular value decomposition, MATLAB

1. はじめに

近年、スマートフォンやSNSなどの普及が進み、インターネットがより身近なものになったことによって、ネットワークトラフィックは増加の一途をたどり、通信インフラとして中核的存在へなりつつあり、近年の著しい普及に伴い、利用人口、ユーザー一人当たりのトラフィック消費量は増加している。これにより、通信障害が引き起こされ、ネットワークの状態が不安定になることは、現在のネットワーク社会に非常に大きな影響を与える。

これに伴い、通信設備の増強などによって、トラフィック上限を引き上げるなどの予防策が必要とされている。しかし、設備の増強には非常にコストがかかる。そのため、ネットワークトラフィックの監視を行う事で、異常を素早く検知する優れたモニタリングプログラムを開発することは、設備の増強費用以外にも、人件費などのコストを抑えることにも繋げることができる。

本稿では、ネットワークにおける通信障害を防ぐ足がかりとして、任意のネットワーク上を流れるトラフィックを短期的に予測する研究を行った。本稿での、予測対象として日本のインターネットバックボーンの中核である、日本インターネットエクスチェンジ上を流れるトラフィックを予測対象とし、将来の1日分(24時間分)のトラフィックを予測した。具体的な予測法としては、一定期間収集したトラフィックデータに対して特異値分解を行うことにより直交成分を得て、その直交成分を元に短期的にトラフィックデータを予測する。また、予測当日新たに得られる5分間隔のデータにより逐次的にトラフィック直交成分の重みを更新し最適な予測を得るアルゴリズムを提案する。

2. ネットワークトラフィックの短期予測

(1) 特異値分解

特異値分解とは、観測される複数の時系列信号を無相関

で直交な信号に分解することである。観測される m 本の時系列信号の1本あたりの点数を n 点として表す。 m 行 n 列の観測データの行列を X とすると、特異値分解では以下のように表す、

$$X = U \sum V^T$$

ここで、 U は $m \times m$ の正規直行行列、 V は、 $n \times n$ の正規直行行列、 Σ は $\mu_1, \mu_2, \dots, \mu_r$ を対角要素にもつ $m \times n$ の対角行列である $\mu_i(i=1,2,\dots,r)$ を行列 X の特異値といい、 $\mu_1 \geq \mu_2 \geq \dots \geq \mu_r$ である。 U と V は互いに直交しており、以下の式が成り立つ。ただし、 I は単位行列である。

$$UU^T = I$$

$$VV^T = I$$

(2) 各曜日の特徴抽出

特異値分解の式 U には、特異値分解されたJP IXトラヒックの各日の特徴が抽出されており、各行には各日における各成分の大きさが数値として示されている。また、各曜日の特徴が以下のように並んでいる。

$$U = \begin{bmatrix} \text{月曜日の特徴} \\ \text{火曜日の特徴} \\ \vdots \\ \text{月曜日の特徴} \\ \vdots \end{bmatrix}$$

本研究では、1 週間ごとにデータを分け、第1 週目から第4 週目までの28 日間を1 カ月分とした。28 日分のJP IX トラヒックを特異値分解して得られた行列 U から、同じ曜日を取り出して平均を取り、それを $\sum V^T$ と掛け合わせることで、1 カ月における各曜日の平均的なトラヒックである、予測トラヒックを作成することが可能である。

(3) ランダムサーチ

実トラヒックは様々な要因によって変動するため、誤差が小さくなるように予測トラヒックを更新する。

本研究では、予測データを作成する際に特異値分解で得られた、特徴成分 u の第一成分と第二成分を用いており、JP IX トラフィックは5分に1点得られるため、新たな1点を得るごとにランダムサーチを用いて u_1, u_2 の最適な大きさを求め、更新した u_1, u_2 を平面に見立てて、平面上の全方向を探索し最適解を求める。

以下にランダムサーチのアルゴリズムを示す。

1. 新たなデータが得られた時で、現在の u_1, u_2 の組み合

わせを基準点としてランダムサーチを開始する。

2. 現在の基準点の近傍を探索する。現在の基準点と、基準点から 12 方向に同じ距離だけずらした点を全て調べる(合計 13 点)。
3. 最良の点を新たな基準点とし、手順2を繰り返す。基準点が最良の点となれば、最良の u_1, u_2 が得られたという事なので終了する。

(4) グリッドサーチ

より正確な予測トラヒックを作成するため、ランダムサーチを行った後にグリッドサーチを行う。基本的な更新手順は同じであるが、 u_1, u_2 の決定方法が異なる。

また、グリッドサーチで得られた u_1, u_2 の値よりも、ランダムサーチで得られた値の方が優れていた場合、グリッドサーチで得られた値は棄却する。

以下にグリッドサーチの手順を示す。

1. ランダムサーチで得られた u_1, u_2 を基準点とし、 u_1, u_2 それぞれを設定した値の範囲内で変化させたものを 400 個作成する。
2. 400 個の組合せと基準点を比較し、その中から最も誤差が小さくなったものを最良の u_1, u_2 とする。

(5) トラヒックの予測結果

2章(1)(2)(3)で述べた方法を用いて、2013 年 8 月 6 日(火)のJP IX トラヒックの短期予測を行う。

結果として得られた予測データと実測データを図 1 に示す。

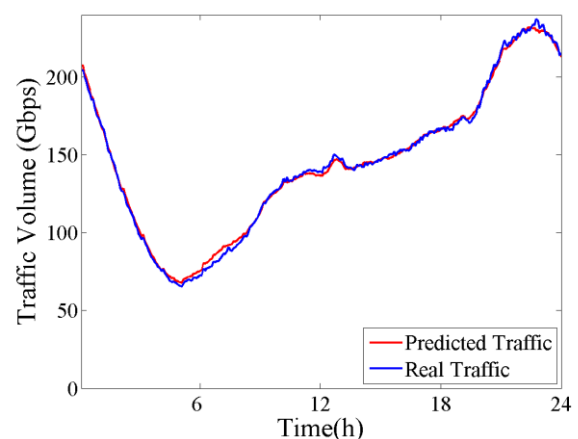


図 1. 2013 年 6 月 24 日の予測データと実測データ

時系列毎の相対誤差の変移を図 2 に示す。

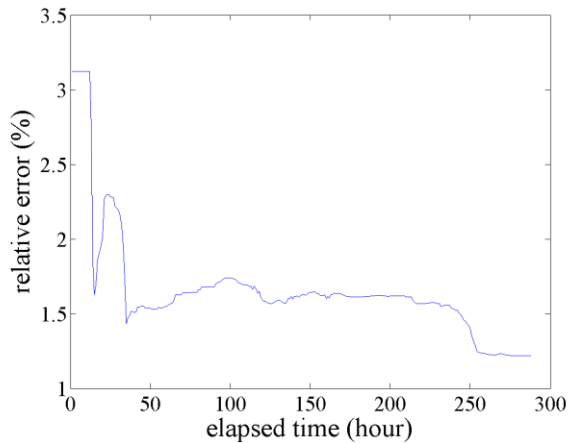


図 2. 2013 年 6 月 24 日の相対誤差の推移

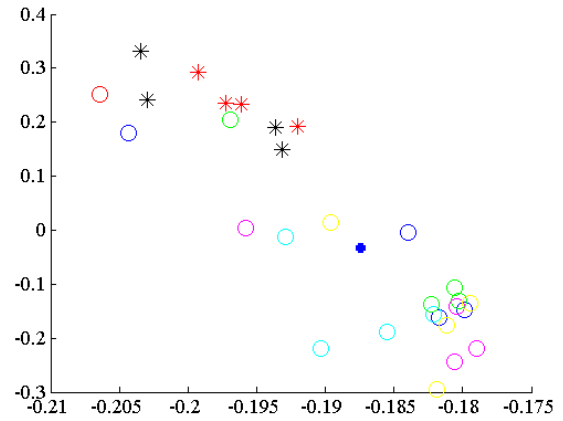


図 4. u の第一成分と第二成分の分布図

(6) 祝日,振替休日のトラヒックの予測

本来平日である曜日での祝日,振り替え休日には, 平日の通りに会社での仕事や大学など授業を行う人のトラヒックと, 休日のようにインターネットを使用する人のトラヒックの二つの特徴が見られるのではないかと考えた

実際に 2013 年 4 月 8 日から 2013 年 5 月 5 日の 28 日分のトラヒックを特異値分解し, 振替休日である 2013 年 5 月 6 日のトラヒックの予測を行う.

実測データと予測結果を図 3 に, u の第一成分と第二成分の分布を図 4 に, 図 5 に相対誤差の推移を示す. また, 各曜日は

月曜日: ブルー,
火曜日: マゼンタ,
水曜日: シアン,
木曜日: イエロー,
金曜日: グリーン,
土曜日: レッド,
日曜日: ブラック,

のマーカで表示されている. また, 青い点は 5 月 6 日の予測トラヒックの u の第一成分と第二成分の始点, 赤い丸は 5 月 6 日の u の第一成分と第二成分を示す

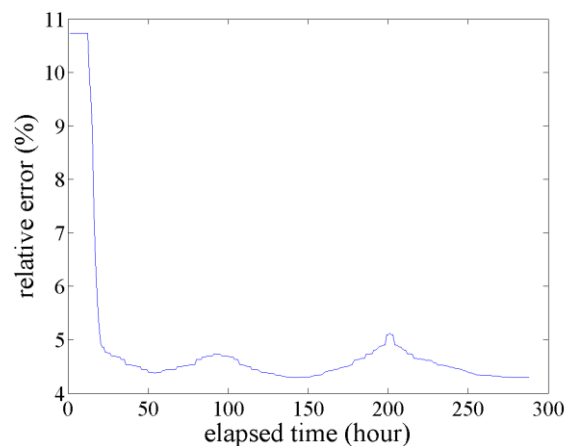


図 5. 2013 年 5 月 6 日の相対誤差の推移

図 4 より平日の特徴成分を用いて u の第一成分と第二成分の初期値を決めているので, 5 月 6 日の予測では u の第一成分と第二成分は平日の特徴を表した結果となっている. それに対し, 平日と休日の二つの特徴を持っているため中間に分布されているという仮定通りではなく休日のトラヒックの特徴成分を表していることがわかる.

同様の手法を用い検証したところ, 他の祝日, 振替休日においても休日の特徴成分を表していることがわかった.

また, ゴールデンウィーク内の休日と祝日の間の平日においては休みをとり長期の連休にする人と休みを取らず働いている人がいるため, u の第一成分と第二成分が平日と休日の特徴の中間に位置していることも新たに分かった.

よって, 新たに祝日, 振替休日は休日の特徴成分を用い, 休日と祝日の間の平日では, 図 6 で示す重み a_1, a_2 を用いて予測を行うことで予測の精度があげられるのではないかと考える

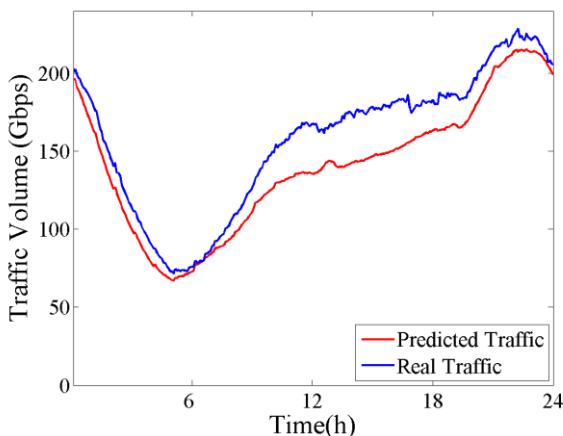


図 3. 2013 年 5 月 6 日の予測データと実測データ

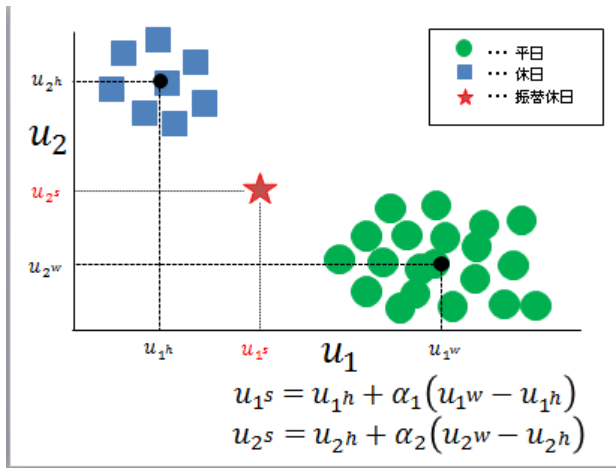


図 6. 予測に用いた 28 日間のデータの u_1, u_2 の値の分布

2013 年 4 月 8 日から 2013 年 5 月 5 日の 28 日分のトラヒックを特異値分解し、休日の成分を用いて振替休日である 2013 年 5 月 6 日のトラヒックの予測を行い、実測データと予測結果を図 7 に、図 8 に相対誤差の推移を示す。

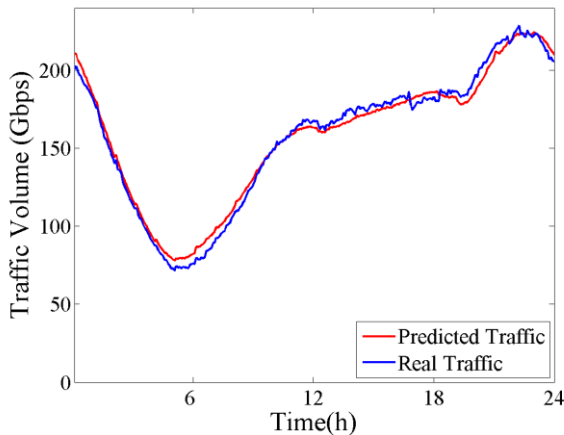


図 7. 休日の特徴を用いた 2013 年 5 月 6 日の予測データと実測データ

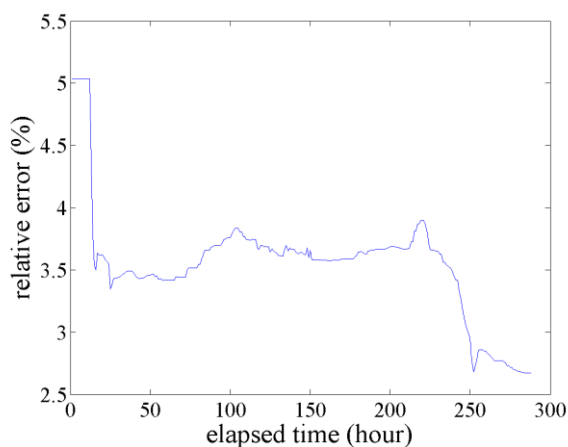


図 8. 休日の特徴を用いた 2013 年 5 月 6 日の相対誤差の推移

初期の相対誤差が 11% だった従来の予測を休日の特徴を用いることで、初期の相対誤差を 5% まで抑えることが出来た。また、最終的な相対誤差も 2.6% となり従来の結果より良い結果となり改善できたといえる。

次に 2013 年 4 月 1 日から 2013 年 4 月 28 日の 28 日分のトラヒックを特異値分解し、重み a_1, a_2 を用いて休日と祝日の間の平日の 2013 年 4 月 30 日の予測を行う。従来の方法での予測結果を図 9、相対誤差の推移を図 10、重みを用いた場合を図 11、相対誤差を図 12 に示す。

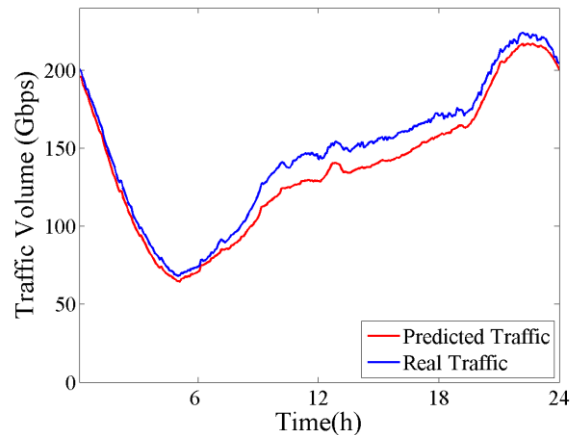


図 9. 2013 年 4 月 30 日の予測データと実測データ

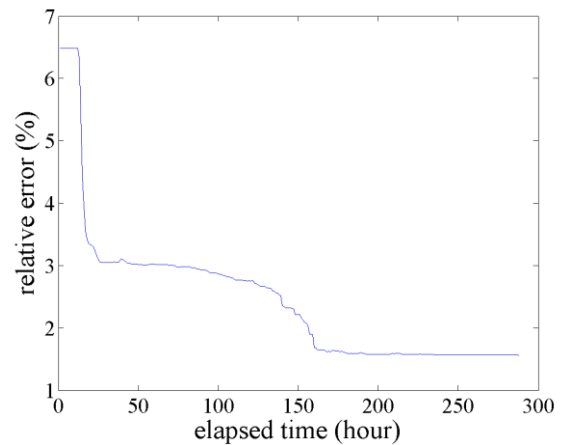


図 10. 2013 年 4 月 30 日の相対誤差の推移

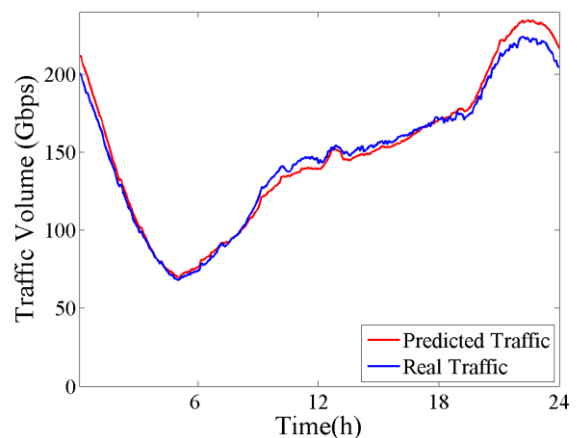


図 11. 重みを用いた 2013 年 4 月 30 日の予測データと実測データ

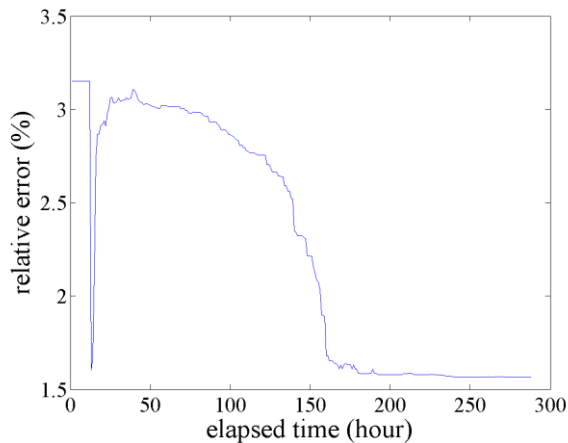


図 1 2. 重みを用いた 2013 年 4 月 30 日の相対誤差の推移

図 10,図 12 より,重みを用いることで予測した 2013 年 4 月 30 日の初期予測の相対誤差が下がり,従来の予測より初期予測精度が向上した.また,追従後の予測結果は従来の手法と同等の結果が得られ,その結果は非常に良い結果となっている.

(7) 結論

新たに提案した祝日,振替休日の予測においては,初期の予測,追従後の予測,相対誤差全てにおいて,従来の手法より良い結果が得られ,今後予測を行っていく上で有用な手法であるといえる.また,祝日と休日の間の日においては,初期予測の初期精度を向上することができ,追従後,相対誤差いずれも従来の手法と同等の結果を得られ,有用であるといえる.

(8) 参考文献

- [1] 我が国のインターネットにおけるトラヒックの集計・試算/総務省
- [2] 石橋弘隆, "JPIX トラヒックの短期予測"2010 年電子情報学科卒業論文
- [3] JPIX- マルチメディア/インターネット事典 ¥¥(<http://www.jiten.com/dicimi/docs/j/6589s.htm>)
- [4] 加藤学, "主成分分析"京都大学大学院工学研究科化学工学専攻プロセスシステム工学研究室
- [5] 鈴木悠太, "IX トラヒックの成分分解"2009 年電子情報学科卒業論文
- [6] 山下雅人, "ネットワークトラヒックの短期予測と公開システムの開発"2014 年情報電子工学修士論文
- [7] Shen Fu-ke, Zhang Weim Chang Pan, "An Engineering Approach to Prediction of Network Traffic Based on Time-Series Model" Artificial Intelligence, 2009. JCAI '09. International Joint Conference on, 25-26 April 2009, 432 - 435