Osaka University

Extracting Information on Traffic Changes from Social Media Data for Predictive Traffic Engineering

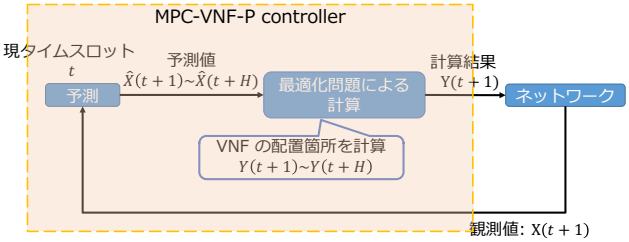
予測型トラヒックエンジニアリングのための ソーシャルメディアデータから トラヒック変動に関する情報の抽出

大阪大学 大学院情報科学研究科
情報ネットワーク学専攻 村田研究室
河島 滉太

Osaka University

モデル予測制御にもとづく 予測型トラヒックエンジニアリング [1]

- 需要変動の予測を用いることにより環境変動に追随可能な手法
  - 予測にもとづき、需要変動に先駆けて VNF の配置変更を開始
  - 観測結果をもとにした予測の補正



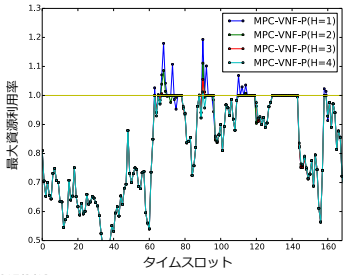
[1] K. Kawashima, T. Otoshi, Y. Ohsita, and M. Murata, "Dynamic placement of virtual network functions based on model predictive control," in *Proceedings of IEEE/IFIP NOMS 2016 Workshop: International Workshop on Analytics for Network and Service Management (AnNet 2016)*, pp. 1037–1042, Apr. 2016.

Osaka University

予測型トラヒックエンジニアリングの効果

- トラヒック変動に先駆けた制御を実現
  - 将来のトラヒック変動を予測
  - VNF の配置箇所を前もって変更

予測精度が高いことが前提



- ネットワーク環境
  - トポロジ：Internet2
  - トラヒック：トレースデータ[2] (2011/11/12 ~ 2011/11/18)
  - 1 タイムスロットあたりの移動可能な VNF 数：1
- 評価指標
  - 最大資源利用率
- 予測誤差を含まない環境下を想定
- 予測対象区間 (H) を変化させ、予測による効果を確認

[2] "Internet2 data." available from http://internet2.edu/observatory/ archive/data-collections.html.

Osaka University

既存のトラヒック予測の問題点

- 様々なトラヒック予測手法が存在
  - ARIMA[3]、SARIMA[4]
- トラヒックの突発的な変動の発生を予測することは困難
  - 現実世界でイベントが発生し、特定地域で人が集中したことによる通信トラヒックの急増
  - 直前のトラヒック変動に兆候が含まれないため

トラヒック変動の予兆が含まれている情報が必要

↓

予兆の抽出元としてツイート情報に注目

[3] K. Papagiannaki, N. Taft, Z.-L. Zhang, and C. Diot, "Long-term forecasting of Internet backbone traffic: Observations and initial models," in Proceedings of IEEE INFOCOM 2003, vol. 2, pp. 1178–1188, Mar. 2003.
[4] Y. Shu, M. Yu, J. Liu, and O. W. W. Yang, "Wireless traffic modeling and prediction using seasonal ARIMA models," in Proc. 2003 IEEE ICC, pp. 1675–1679.

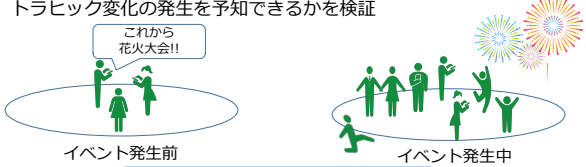
Osaka University

研究目的とアプローチ

- 研究目的
  - 対象地域から流入するトラヒック変動の発生の予知に有用な情報を抽出
- アプローチ

現実世界のイベントに関連するツイート数の増加に注目

  - ある時間帯に投稿されたツイートからイベント関連単語を抽出
  - イベント関連単語を含むツイート数の増加傾向からイベントに起因するトラヒック変化の発生を予知できるかを検証



人の増加による通信量増加を予知

Osaka University

イベント関連単語の抽出

- 対象地域内で、ある時間帯で投稿されたツイートから抽出
  - 当該時間帯で投稿されたツイートに含まれた単語からなる集合を生成
  - 集合に含まれる単語 w の重要度を算出
    - 文章に出現する単語の重要度を計算する手法 [5] をもとに、各時間帯の単語の重要度を計算する指標を考案

$$M_{w,T,D} = \frac{F_{T,D}(w)}{\sum_{s \in N_{T,D,z}} F_{T,D}(s)} \cdot \log \frac{|B_{T,D}|}{|B_{w,T,D}|}$$

当該時間帯における w の単語出現頻度

過去の時間帯のうち、w が出現した時間帯の割合の逆数

- T：時間帯
- D：日付
- $F_{T,D}(w)$：w の出現回数
- $N_{T,D,z}$：日付 D の時間帯 T のツイートに含まれた名詞のうち、上位 z の出現回数の名詞の集合
- $B_{T,D}$:日付 D の時間帯 Tと比較するタイムスロットの集合
- $B_{w,T,D}$：$B_{T,D}$ に含まれるタイムスロットのうち、w が出現したタイムスロットの集合

  - 上記の式によって抽出される単語
    - 当該時間帯で出現頻度が高い単語
    - 他の時間帯には出現していない単語

[5] G. Salton and C. Buckley, "Term-weighting approaches in automatic text retrieval," *Information Processing and Management*, vol. 24, no. 5, pp. 513–523, 1988.

Osaka University

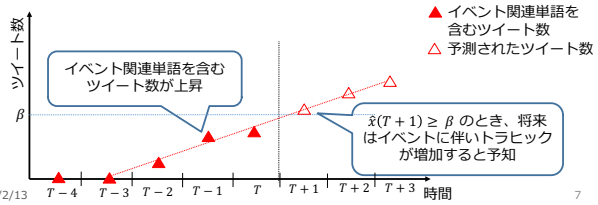
イベント関連単語にもとづく予知

・イベント関連単語のいずれかを含むツイート数を予測
  ・イベント関連単語を含むツイート数の過去の時系列データをもとに、将来の予測値を取得

$$\hat{x}(T+1) = F(x(T-s+1), \cdots, x(T))$$

- $\hat{x}$：予測ツイート数　・x：観測ツイート数
- F：予測モデル　・s：モデルの長さ

・予測されたツイート数を用いて、イベントに起因するトラヒック変化の発生を予知



Osaka University

評価方法

・予知すべきトラヒック変動の定義
  ・過去の同時間帯のトラヒック変動から外れた変動
    ・当該地域内からの位置情報付きツイート数をもとに以下の条件を満たす

本発表では $\epsilon = 70$ の結果を紹介

$$\text{時間帯 } T \text{ のツイート数} > \text{同時間帯の平均ツイート数} + \epsilon$$

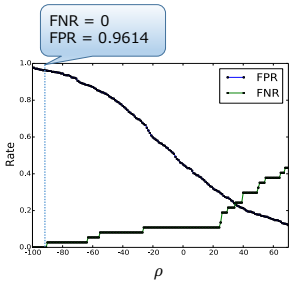
    ・特定地域からの流入トラヒック量とツイート数は高い相関があるため [6]
・データセット
  ・対象地域
    ・渋谷駅周辺
    ・難波駅周辺
  ・取得期間：2016/10/4 ~ 2016/12/23
    • ただし、2016/11/07 14:00 ~ 2016/11/07 17:59, 2016/11/09 04:00 ~ 2016/11/09 14:59, 2016/11/22 06:00 ~ 2016/11/22 12:59, 2016/12/01 20:00 ~ 2016/12/02 11:59 は除く
・比較手法
  ・各時間帯の総ツイート数を用いた予知
    ・総ツイート数の時系列予測の結果と同時間帯の平均ツイート数の差が閾値 ρ を超えた際に発生を予知
・指標
  ・偽陰性率 (FNR; False Negative Rate), 偽陽性率 (FPR; False Positive Rate)

[6] B. Yang, W. Guo, B. Chen, G. Yang, and J. Zhang, "Estimating mobile traffic demand using Twitter," *IEEE Wireless Communications Letters*, vol. 5, pp. 380-383, Aug. 2016.

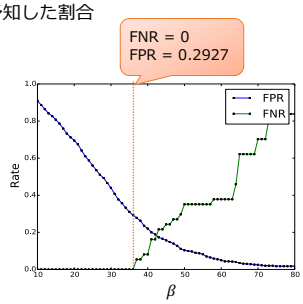
Osaka University

評価結果

・イベント関連単語に含まれる情報を用いることで予知精度が向上
  ・FNR：予知すべきトラヒック変動を予知できなかった割合
  ・FPR：トラヒック増加を誤って予知した割合



総ツイート数の時系列予測にもとづく予知



抽出された単語をふくむツイートの増加にもとづく予知

Osaka University

まとめと今後の課題

・まとめ
  ・対象地域から流入するトラヒック変動の予兆をツイートから抽出
    ・ある時間帯で投稿されたツイートからイベント関連単語を抽出
    ・イベント関連単語を含むツイート数の予測をもとに、トラヒック変動の発生を予知
  ・評価結果
    ・トラヒック変動の予兆をソーシャルメディアデータから抽出可能
    ・スパムツイートの影響による誤検知が発生
・今後の課題
  ・スパムの影響を除外する手法の改善
  ・抽出情報をとりいれたトラヒック予測手法の提案