

# モバイル端末位置履歴を用いた 都市動態予測のための核ノルム最小双線形ポアソン回帰

Bilinear Poisson regression with nuclear norm minimization  
for urban dynamics prediction using mobility logs

下坂 正倫<sup>\*1</sup>      和田 英之<sup>\*2</sup>      坪内 孝太<sup>\*3</sup>      築地 毅<sup>\*2</sup>  
Masamichi Shimosaka      Hideyuki Wada      Kota Tsubouchi      Takeshi Tsukiji

<sup>\*1</sup>東京工業大学      <sup>\*2</sup>東京大学      <sup>\*3</sup>Yahoo! JAPAN 研究所  
Tokyo Institute of Technology      The University of Tokyo      Yahoo! JAPAN Research

位置履歴の膨大な蓄積を利用した群集の活動分析の基盤となる活動人口の時系列挙動（都市動態）の予測手法を論ずる。予測手法として核ノルム正則化を取り入れて最適化する双線形ポアソン回帰モデルを提案する。従来の行列分解による最適化と比べ、大域的最適解が得られる利点がある。大規模な携帯端末位置履歴データに対する提案手法の有効性を示す。

## 1. 序論

近年、モバイル端末の普及に伴い、膨大な量のモバイル端末位置履歴データが蓄積されてきている。それらの位置履歴データを用いた都市の分析を行う研究 [Fan 14][Nishi 14] が盛んである。本研究では、モバイル端末位置履歴を用いた都市規模での活動人口の予測を行う枠組みを扱う。都市規模での活動人口の予測を行うことで、混雑状況の事前予測を行うことや、ユーザー属性ごとに分割したデータの分析により客層に合わせた屋外広告への応用が期待できる。

都市の活動人口の時系列挙動（都市動態）は天気や曜日などの様々な要因から影響を受けている。本研究では、それらの要因を取り込んだポアソン回帰モデルによる人数予測を行う。本研究では時間帯による要因を時間的要因とし、都市動態に影響を与えたと考えられる時間帯以外の要因を外的要因とする。我々のこれまでの取り組み [Shimosaka 15] では、外的要因の変化による都市動態のピーク変化を表現するために、双線形ポアソン回帰による定式化を行っている。従来広く用いられてきた線形ポアソン回帰と比べ大幅な性能向上を達成する一方で、パラメータ行列の低ランク化として行列分解に基づく最適化により局所解に陥る問題がある。

本研究では、低ランク化を明示的に行う代わりに、核ノルムを正則化項に含めた最小化問題を考えることでより精度の高い都市動態予測を行う。実際のモバイル端末位置履歴を用いた評価実験を行うことで、核ノルム最小化による最適化の優位性を示す。

## 2. 核ノルム最小化による双線形ポアソン回帰

### 2.1 都市動態予測の問題設定

分析をする対象地域の1日の活動人口の時系列挙動のモデル化を行う。1日を任意の時間帯に分割し、ある時間帯でのモバイル端末位置履歴の総量をその時間帯での対象地域の活動人口と定義する。非負整数である各時間帯の活動人口がポアソン分布に従うという仮定に基づき、都市動態のモデル化を行う。24時間を  $S$  分割した時間帯でのそれぞれの活動人口を  $\mathbf{h} = [h_1, h_2, \dots, h_S]$  として表すと、 $s$  番目の時間帯の活

動人口  $h_s$  に対して、ポアソン分布によるモデル化は  $p(h_s) = \text{Pois}(h_s | \lambda_s) = \lambda_s^{h_s} \exp(-\lambda_s) \Gamma(h_s + 1)^{-1}$  と表される。各時間帯の活動人口が従うポアソン分布のパラメータ  $\lambda_s$  を求める。時間帯や曜日、天気などに応じて変化するポアソン分布のパラメータ  $\lambda$  を求めることが本研究の中心的課題となる。

### 2.2 既存モデル：低ランク双線形モデル

時間的要因だけでなく外的要因も取り込むことを考える場合、時間的要因と外的要因の掛け合わせを考慮したほうが、精度の高い予測が行えると考えられる。そこで、それを達成する枠組みとして双線形の形でポアソン回帰の問題を考えると、 $\lambda$  の算出は次のように表せられる。ここで、時間的要因を  $t \in \mathbb{R}^S$ 、外的要因を  $d \in \mathbb{R}^M$ 、パラメータ行列を  $\mathbf{W} \in \mathbb{R}^{M \times S}$  とする。

$$\ln \lambda(t, d) = d^T \mathbf{W} t \quad (1)$$

時間的要因  $t$  は、該当する時間帯  $s$  が平均になるガウス分布の広がりを持つ特徴量である。つまり、時間帯  $s$  での時間的要因  $t$  の  $k$  ( $k = 1, 2, \dots, S$ ) 番目の要素  $t_k$  は、ガウス分布の分散を  $\sigma$  とすると、 $t_k = \mathcal{N}(k | s, \sigma)$  として表すことができる。外的要因  $d$  は、曜日、天気、祝日か否かなど  $M$  次元の特徴量からなり、該当する  $m$  番目の要素のみが1となる関数である。例えば、祝日か否かの要因を表す2次元の特徴量関数  $d$  は、祝日である時には  $d = (1, 0)^T$ 、祝日ではない時には  $d = (0, 1)^T$  で表せる。

我々のこれまでの取り組み [Shimosaka 15] では、双線形回帰により増加したパラメータの過剰適合を回避するため、パラメータ行列の低ランク化を明示的に行い、学習の安定化を行った。しかし、低ランク化を明示的に行う場合、パラメータ行列の全変数に関して非凸な最適化問題となるため、大域解が保証されなくなり、予測精度にばらつきが生じる問題がある。

### 2.3 核ノルム最小化による双線形モデルの学習

低ランク化を明示的に行う代わりに、低ランク制約の凸緩和である核ノルム制約を考えることで、予測精度のばらつきを抑え、予測性能の向上が期待できる。これは、核ノルムは非零の特異値の和で表され、スパースな特異値を持つように促されるためである。負の対数尤度関数に正則化項として核ノルムを含めた目的関数  $F = -\sum_{n=1}^N \sum_{s=1}^S \ln \text{Pois}(h_{n,s} | \lambda_{n,s}) + \|\mathbf{W}\|_{\text{tr}}$  が最小となる最適化を行う。ここでは、 $\|\cdot\|_{\text{tr}}$  はトレースノルムを表す。

核ノルム正則化の最適化については劣微分を考えねばならず、解決策として [Toh 10] のように近接勾配による計算が考

連絡先: 下坂正倫, simosaka@miubiq.cs.titech.ac.jp

連絡先: 和田英之, wada@miubiq.cs.titech.ac.jp

連絡先: 坪内孝太, ktsubouc@yahoo-corp.jp

連絡先: 築地毅, tsukiji@miubiq.cs.titech.ac.jp

えられる．近接勾配による計算は特異値分解を内部で行うため，計算コストが大きくなることが想定される．

そこで本研究では計算コストの軽減を念頭に，近接勾配による最適化ではなく，Frank-Wolfe 法に基づく最適化 [Jaggi 10] を採用する．具体的には初期値  $W_0 \in \mathbb{R}^{M \times S}$  を零行列で与え，核ノルム制約の大きさ  $r$ ， $k$  回目 ( $k \in \mathbb{N}$ ) に得られる  $W_k$  における負の対数尤度関数の勾配行列の最大特異値  $s_k$  に対する特異値ベクトル  $u_k \in \mathbb{R}^M$ ， $v_k \in \mathbb{R}^S$  を用いて，次の式で更新を一定回数繰り返す．

$$W_{k+1} = \left(1 - \frac{2}{k+2}\right)W_k + \frac{2}{k+2}r u_k v_k^\top \quad (2)$$

最大特異値と対応する特異値ベクトルの算出は，FRSVT[Oh 15] などの簡易な計算法により，特異値分解そのものと比べると計算時間を削減できる．

### 3. 性能評価実験

2.3 節で述べた核ノルム制約を満たしながら最適化する都市動態予測の性能を評価するために，実際の位置履歴データを用いた性能評価実験を行った．

#### 3.1 実験概要

##### 3.1.1 実験データ

大規模なモバイル端末から取得される位置履歴データを用いて，都市内部での活動人口データに対する動態予測の性能評価実験を行う．本実験では，ヤフーが提供する防災アプリから得られた位置情報を入力データとして用いた．完全な匿名化処理が施され，時間と位置情報のみのデータで，ユーザに関する情報は一切含まれていないデータである．300POI などの選定は Shimosaka ら [Shimosaka 15] に従う．

##### 3.1.2 実験設定

本実験では，曜日，祝日か否か，天気の3つの要因を外的要因として取り込む．訓練データ日数として，30日，90日，180日の3パターンを用いて学習させ，テストデータ日数をそれぞれ180日でテストを行う．訓練データ数の3パターンに対して各地域 (Point of Interests, 以後 POI) で5交差検定を行い，300POI での平均値を比較する．

##### 3.1.3 比較手法

本実験では，比較手法として，2.2 節で述べた既存手法である低ランク双線形モデル [Shimosaka 15] を用いる．

##### 3.1.4 評価指標

本実験では評価指標として，テストデータと予測値との平均絶対誤差 (Mean Absolute Error, MAE)，平均相対誤差 (Mean Absolute Error Rate, MAER)，平均ピーク時刻誤差 (Mean Absolute Peak Error, MAPE) の3つを用いる [Shimosaka 15] ．

#### 3.2 実験結果

評価指標 MAE, MAER, MAPE での比較をそれぞれ図1, 図2, 図3に示す．縦軸が評価指標を示し，3つの指標とも数値が小さいほどモデルの評価がよいことを意味する．図には300POI での予測性能の平均と標準偏差を示している．赤色 (BilinearLR) が既存手法による結果を示し，緑色 (BilinearFW) が提案手法による結果を示す．また，横軸は訓練データ日数を示し，図の左から順に訓練データとして30日分，90日分，180日分のデータを使った結果を示す．

いずれの評価指標でも既存手法に比べて提案手法が優位であることを確認した．このことから提案手法により高精度な都市動態予測が行えることが示された．

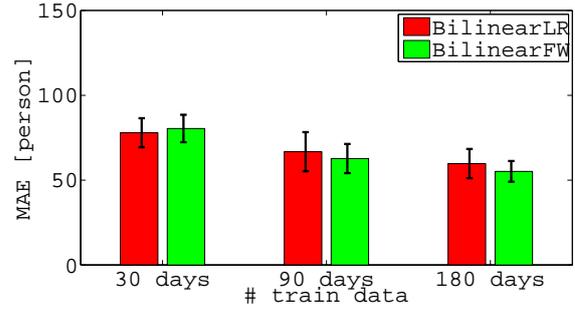


図1: 平均絶対誤差 MAE

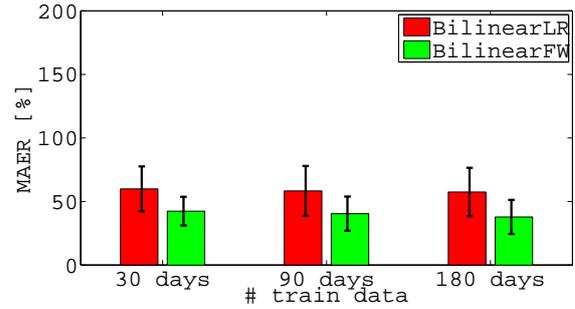


図2: 平均相対誤差 MAER

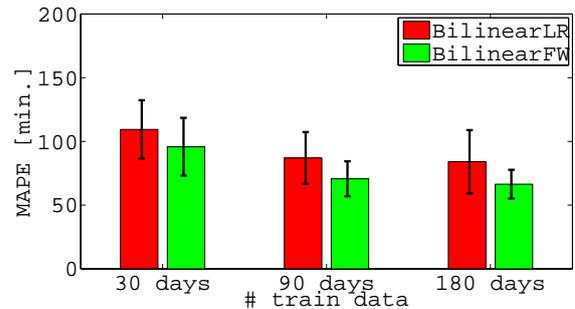


図3: ピーク時刻誤差 MAPE

## 4. 結論

本研究では，高精度な都市動態予測を行う枠組みの構築を目的として，核ノルム最小化によりパラメータの低ランク化を促す双線形ポアソン回帰を提案し，提案モデルによる性能の向上を実際の位置履歴データを用いた実験を行い検証した．3つの評価指標で性能の向上が確認された．これにより明示的な低ランク化を行う既存手法よりも提案手法を用いることでより精度の高い都市動態予測を行えることが示された．

## 参考文献

- [Fan 14] Fan, Z., Song, X., and Shibasaki, R.: CitySpectrum: A Non-negative Tensor Factorization Approach, in *Proc. of UbiComp* (2014)
- [Jaggi 10] Jaggi, M. and Sulovsky, M.: A Simple Algorithm for Nuclear Norm Regularized Problems, in *Proc. of ICML* (2010)
- [Nishi 14] Nishi, K., Tsubouchi, K., and Shimosaka, M.: Extracting land-use patterns using location data from smartphones, in *Proc. of Urb-IoT* (2014)
- [Oh 15] Oh, T., Matsushita, Y., Tai, Y., and Kweon, I.: Fast Randomized Singular Value Thresholding for Nuclear Norm Minimization, in *Proc. of CVPR* (2015)
- [Shimosaka 15] Shimosaka, M., Maeda, K., Tsukiji, T., and Tsubouchi, K.: Forecasting Urban dynamics with Mobility Logs by Bilinear Poisson Regression, in *Proc. of UbiComp* (2015)
- [Toh 10] Toh, K. and Yun, S.: An accelerated proximal gradient algorithm for nuclear norm regularized linear least squares problems, *Pacific Journal of Optimization* (2010)