

人流センサデータによる混雑予測モデルの構築と松江駅の特徴検出

古俣 勇樹[†] 中村 和幸[†]

[†] 明治大学

1 はじめに

新型コロナウイルス拡大に伴い、密を避けた行動様式が新たに求められている。もし仮に混雑予測の情報が簡単に手に入れば、混雑緩和につながると考えられる。また、ある観光地の混雑予測情報がわかることで、そこに存在する店舗の売り上げ予測や商品入荷の最適化をすることが可能になり、観光地や地域の活性化につながる。日立研究所による研究 [2] では、駅構内のモニタカメラを用いて人数を把握しており、画像データから数値データを取得し、人数推定を行なっている。本稿では、松江駅の人流センサで取得された数値データ [1] を用いた混雑予測の枠組みを提案し、そこから得られた松江駅の特徴を紹介する。

2 準備

2.1 本稿で用いるデータセット

本稿で用いるのは G 空間情報センター web ページにおいて公開されている松江駅人流センサデータである。松江駅構内にセンサが 19 個設置されており、センサがある出入り口から入ってきた人数や出ていった人数が 1 分ごとに取得されている。本稿では松江駅のコンコース内の人数予測に絞るため、センサ ID が 6 と 8 から 14 のものに限定する。データのカラムとしては、センサ ID、日付、時刻、入ってきた人数、出ていった人数、入ってきた人数の累計、出ていった人数の累計である。

2.2 データ加工

2 つの混雑指標データを作成する。1 つ目が stock データである。コンコース内に入っていく人数を足していき、出ていく人数を引いていくことで、コンコース内の人数指標を示すデータを作成する。その際、人数がマイナスの時刻が発生する場合がある。この問題点に対して、センサでの人数カウントの取りこぼしが発生しているという仮説を立て、区分線形モデルによる人数修正と

最適化を行う。区分線形モデルは、

$$f(x; a, c) = \begin{cases} a(x - c) + c & x > c \\ x & -c \leq x \leq c \\ a(x + c) - c & x < -c \end{cases} \quad (1)$$

で与えた。ただし、 a, c はパラメータ、 x は修正前の人数を表している。この人数について、1 日の間のコンコース内の出入り人数が同数であると考え、目的関数

$$L(a, c) = \left\{ \sum_{t=1}^T f(x(t); a, c) - 0 \right\}^2 \quad (2)$$

を最小にする。ただし、 t は 5 : 00 を $t = 1$, 24 : 29 を $t = T$ とした 1 分毎の時刻 $x(t)$ は時刻 t における出入りした人数の和で、流入を正、流出を負としたものである。

また、2 つ目に flow データを作成した。センサの前を通った人数が多ければ多いほどコンコース内は混んでいるという仮説のもと、センサの前を通った人数を 15 分間カウントすることで flow の値を取得した。

2.3 モデル作成と回帰分析

目的変数を stock, flow のそれぞれに対し、回帰モデル 2 種とテンソル分解モデル 1 種を適用した。まず、以下の 2 つの回帰分析モデルを示す。なお、回帰モデルの分析においては通常とは異なるゴールデンウィークと松江水郷祭という花火大会がある週は除外し、休日のみ用いた。

$$y = \beta_0 + \sum_{i=1}^2 \sum_{j=1}^{78} \beta_{i,j} x_{i,j} + \sum_{k=1}^{49} \beta_k x_k, \quad (3)$$

$$y = \beta_0 \prod_{i=1}^2 \prod_{j=1}^{78} \beta_{i,j}^{x_{i,j}} \prod_{k=1}^{49} \beta_k^{x_k}. \quad (4)$$

ただし、 i は曜日を表し、 $i = 1$ は土曜日、 $i = 2$ は日曜日である。また、 j は 5 : 00 から何番目の 15 分間であるかを表し、 $j = 1$ は 5 : 00 を表し、 $j = 78$ は 24 : 15 である。 k は週を表し、17, 18 週目はゴールデンウィーク、31 週目は松江水郷祭がある週であるため用いていない。そのため $k = 49$ が 52 週目である。 $x_{i,j}, x_k$ はダミー変数であり、例えば、 $x_{1,1}$ は土曜日の 5 : 00 であれば 1、それ以外は 0 である。また、 y は対象としている flow または stock の値である。このモデルを作成することで土曜日の動きと日曜日の動きを区別できる。(3) の

Construction of Congestion Prediction Models with Human Flow Sensor Data and Feature Detection of Matue Station

[†]Yuki Komata and [†]Kazuyuki Nakamura

[†]Meiji University

式を用いたモデルを 1 とし, (4) の式を用いたモデルを 2 とする.

次にテンソル分解モデルを以下に示す. テンソル分解基底による回帰のための準備として, 土日, 週, 15 分毎の時刻を軸とする $\mathbb{Z}^{2 \times 49 \times 78}$ の 3 次元テンソルに対するテンソル分解を適用した. 分解アルゴリズムは PARAFAC を用いた. 次に曜日ごとの $\mathbb{Z}^{49 \times 78}$ の 2 次元テンソルに対するテンソル分解を行い, その時間成分である第 2 軸の基底を用いた回帰を行なった. 回帰モデルは次の通りである.

$$y = \beta_0 + \sum_{i=1}^2 \sum_{j=1}^{49} \sum_{k=1}^{78} \sum_{l=1}^3 \beta_{j,l} x_{i,j,k} z_k^{(i,l)} \quad (5)$$

$z_k^{(i,l)}$ は土曜日 ($i = 1$) または日曜日 ($i = 2$) の第 l 基底の第 2 軸第 k 成分を表す. 本稿ではテンソル分解によって得られた基底を 3 つ取り入れ, これをモデル 3 とする.

3 分析結果と考察

3.1 数値実験結果

表 1 各モデルの MAE の値

model	Stock			Flow		
	Linear	Lasso	Ridge	Linear	Lasso	Ridge
1	24.66	24.65	24.66	85.17	85.12	85.15
2	24.32	24.32	24.32	82.32	81.69	82.19
3	29.02	27.74	28.09	84.30	84.30	85.30

3 つのモデルに対して Leave-One-Cut CrossValidation(LOOCV) による予測誤差の平均絶対誤差を評価指標として平均絶対誤差を用いて評価した. 回帰係数の推定には通常の最小二乗法に加え, Lasso ならびに Ridge 回帰を適用した. 結果は表 1 であり, stock データ, flow データを用いたモデル共にモデル 2 が最も良いと言える.

また, stock データを用いたモデルと flow データを用いたモデルではどちらが優れているのかを確認するために, 標準化を行い, 重回帰分析を用いたモデル 1 のみ平均絶対誤差を算出した. Stock データを用いたものが 0.382, flow データを用いたものが 0.183 となり, flow データを用いたモデルが最も良いことが確認できた.

3.2 テンソル分解基底と回帰係数から観察される人出の特徴

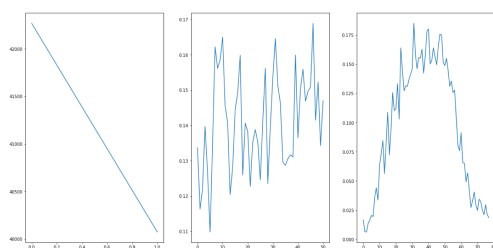


図 1 PARAFAC 分解による第 1 基底

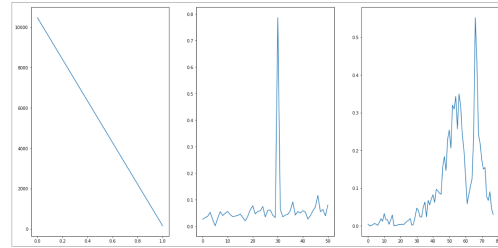


図 2 PARAFAC 分解による第 2 基底

表 2 モデル 1 の土日の各時刻の回帰係数 (上位 4 位)

土曜日の時刻	回帰係数	日曜日の時刻	回帰係数
12:45	515.3	15:00	521.5
15:00	505.1	14:45	492.8
16:45	475.6	16:45	460.9
18:00	464.1	12:45	456.0

図 1,2 は PARAFAC 分解によって得られた第 1,2 基底であり, 左から土日, 週, 時刻を表している. 両基底より土曜日の方が人数が多いことがわかり, 第 2 基底で松江水郷祭の特徴を検出することに成功した. また, 表 2 はモデル 1 の土日の各時刻の回帰係数のうち, 上位 4 位のものを記載したものである. この結果から, 00 分と 45 分台に人数が多くなることがわかる.

4 結論

本稿では, 松江駅の人流センサデータを用いて, 混雑分析・予測のための前処理とモデル構築の枠組みを作成した. 駅構内の人数に関する stock データ, 出入口の通行人数に関する flow データを作成した. stock データ作成時に, 駅構内の人数が負の値になる時刻が発生するケースがある場合は, 区分線形モデルと最適化によって解決した. これらのデータの対して回帰分析やテンソル分解を用いて混雑予測を行うと, flow データの予測の方が良かった. また 00 分台, 45 分台の人数が多いこと, 土曜日の方が人数が多いことが読み取れ, 8 月の 1 週目に行われる松江水郷祭の 1 日の動きを抽出することにも成功した.

謝辞

本研究では, G 空間情報センター 社会基盤情報流通推進協議会により公開されているデータを使用した.

参考文献

- [1] G 空間情報センター人流解析チーム. 「松江駅構内人流センサデータ」.
<https://www.geospatial.jp/ckan/organization/aigid-team>.
- [2] 日立製作所 (2017 年 4 月). 「駅構内モニタカメラを用いた混雑度可視化技術」.
<https://www.ipsj.or.jp/dp/contents/publication/30/DP30-T01.html>.