

改札通過データを用いた 旅客流動のリアルタイム推定手法

杉山 陽一* 松原 広* 明星 秀一**
田村 一軌*** 尾崎 尚也#

An Approach for Real-time Estimation of Railway Traveler Flow

Yoichi SUGIYAMA Hiroshi MATSUBARA Shuichi MYOJO
Kazuki TAMURA Naoya OZAKI

If traveler flow information is comprehensible in real time, such information is applicable to railway traffic operations. In this report, forecast of the number of gate passage according to time zone was attempted by using past data from automatic ticket checking machines, then the number of passage by each origin station was estimated by using the number of gate passage. As attributes of daily passage data were apparent, two types of prediction approaches were applicable to stable data and irregular data. Consequently, applicable criteria of each approach have been understood. Moreover, accurate forecast of the number of OD passage was done by the developed forecasting model.

キーワード：旅客流動，自動改札機，OD データ，通過人数，パターンマッチング，自己回帰モデル

1. はじめに

鉄道の運行管理や旅客案内において、旅客流動（本報告では時間帯ごとの旅客の出発駅から目的駅までの流量を指す）を正確かつリアルタイムに把握することにより、旅客のニーズに合致した輸送力の提供や、状況に合ったより正確な旅客案内が期待できる。本報告は、大都市圏の旅客流動を推定するための手段として、日々の鉄道利用実績および現時点までの利用経過をもとに駅利用人数および2駅間のOD（Origin -Destination：発着駅の組合せ）利用人数をリアルタイムに予測可能な手法を考案し、実際の改札データを用いて特性を検証した。

2. 関連する研究事例

旅客流動予測の方法を検討するため、実績値が日々得られる分野での需要予測事例を調査した。

田中ら¹⁾は重回帰モデルを用いて改札通過旅客数を予測しており、イベント需要の予測も行っている。松木ら²⁾は、自己回帰移動平均モデルなど3つのモデルにより、新幹線の年間利用人数を予測している。また、小島ら³⁾

はマルスシステムの指定席発売記録をもとにした新幹線の臨時列車設定について考察している。鉄道以外では、重回帰モデルで予測した電力需要をニューラルネットワークにより補正した事例⁴⁾、決定木とニューラルネットワークの組合せで電力負荷を予測した事例⁵⁾などがある。また、短期的傾向や長期的傾向を反映した予測として、曜日・天気などの要因ごとに分解するモデルにより飲食店売上を予測する事例⁶⁾や、重回帰モデルにより電話問合せ件数を予測する事例⁷⁾などがある。

しかし、これらの手法の多くは総需要などある特定のデータのみを対象にした予測や推定に特化しており、多数のデータを対象にした予測や推定への適用が難しい。また、モデルのパラメータには、即時に確定しない属性（天気、降水量など）が含まれている事例も多い。旅客流動の予測では多数の駅の利用データを扱うことが想定されるため、よりシンプルな予測手法の適用が求められる。また、リアルタイムに予測を行うためには、即時に得られる属性によって予測できることが望ましい。

3. リアルタイムな旅客流動予測のアプローチ

旅客流動をリアルタイムに予測する方法を検討した。既往の研究には、明星⁸⁾によって旅客流動のシミュレーションを行った事例がある。これは、自動改札機から得られる既知のODデータをもとに各列車に旅客数を

* 輸送情報技術研究部（旅客システム）
** 総務部（総務）
*** 財団法人福岡アジア都市研究所
輸送情報技術研究部（設備システム）

特集：輸送情報技術

配分するものであるため、リアルタイムに時間帯別ODを予測することができれば、結果を文献8)のシミュレーションに適用することで、旅客流動を推定できることが分かる。しかし現状では、改札データをリアルタイムに取得することはできない(通過人数データのみオンラインで取得可能であり、ODデータは情報量が多く、駅で直接取得する必要がある)。そこで本報告では、改札口を通過する人数がリアルタイムに得られるという前提で、①リアルタイムに時間帯別通過人数を予測し、②その結果から時間帯別ODを推定する、という2段階のアプローチにより、旅客流動の推定を試みた。

4. 改札データの分析

時間帯別通過人数予測および通過人数からのOD推定のためのデータとして、乗車人員報告、車両ばね圧、券売実績なども考えられるが、①データの精度や利用実態の再現性が高いこと、②大都市圏が予測対象であること、③文献10)で使用したこと、などの理由から、改札データを使用することが最も現実的と考えられる。

4.1 改札データ取得

4.1.1 取得駅

様々な駅の特徴を把握するため、ある鉄道事業者の下記8駅から改札データを取得することとした。

- 大規模駅：P 駅， Q 駅， R 駅
- 乗換駅：S 駅（他社乗換）， T 駅（自社乗換）
- 中間駅：U 駅（並行他社線あり）， V 駅
- イベント駅：W 駅（コンサート会場最寄り）

複数の乗継がある大規模駅3駅は、主要な目的地でもある。それ以外の5駅については、乗換の有無、並行他社線の有無、イベントの発生などを選定基準とした。

4.1.2 取得期間

通過人数予測手法の適用および予測結果の検証を行うため、2007年通年および2008年6月中旬から9月下旬まで（出勤・登校日、盆休み、夏季休暇など特徴的な日が含まれる）の期間に改札データを各駅の改札機から直接取得した。2007年のデータは予測手法の適用に、2008年のデータは予測結果の検証に用いる。

4.2 データの様相

改札データには、通過人数データとODデータがある。通過人数データは時間帯(No.1-35に分割されている)・券種(利用目的や媒体によって細分されている)ごとの入場人数および出場人数が記録され、ODデータはさらに発駅ごとに分類された出場人数が記録されている。特記しない限り、改札口ごとのデータを1駅分に集約した。また、券種については特徴が捉えやすい通勤定期・通学定期・定期外の3種類に集約した。

4.3 通過人数の傾向

4.3.1 年間の通過人数変動

1日合計通過人数の年間推移の周期性を調べるため、検証用データについて離散フーリエ変換(DFT: Discrete Fourier Transform)を行った。その結果、曜日に関与する波形は見られたが、他に目立つ波形は見られなかったため、通過人数を左右する主要な要素は曜日と考えられる。

4.3.2 曜日による通過人数の違い

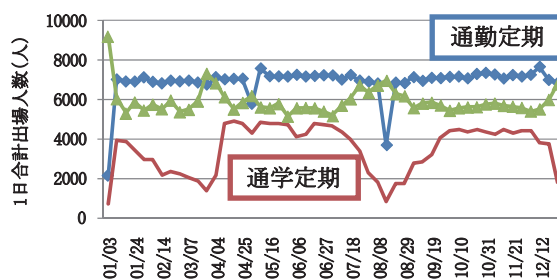
曜日による通過人数の違いを検証するため、時間帯別通過人数から得られる曜日相互のユークリッド距離(人数差を2乗した値の総和)を求めた。結果を表1に示す。日曜どうし、平日相互、土曜どうしの類似性は高いが、相互の類似性は低いとわかる。

表1 曜日間のユークリッド距離(2007年6月S駅入場)

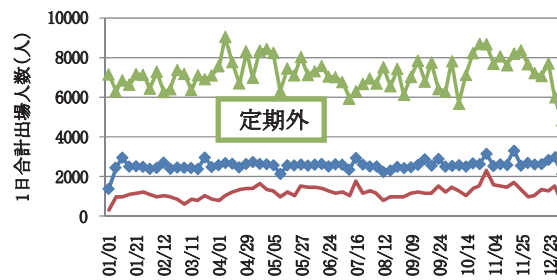
	16(土)	17(日)	18(月)	19(火)	20(水)	21(木)	22(金)	23(土)	24(日)
16(土)	-	710	1497	1476	1493	1582	1624	250	838
17(日)	710	-	1768	1751	1746	1858	1892	783	433
18(月)	1497	1768	-	322	379	489	468	1508	1736
19(火)	1476	1751	322	-	282	411	456	1502	1751
20(水)	1493	1746	379	282	-	386	359	1508	1745
21(木)	1582	1858	489	411	386	-	523	1591	1874
22(金)	1624	1892	468	456	359	523	-	1643	1878
23(土)	250	783	1508	1502	1508	1591	1643	-	920
24(日)	838	433	1736	1751	1745	1874	1878	920	-

4.3.3 曜日・券種による通過人数の年間変動の違い

1日合計出場人数の曜日ごとの年間推移を図1に示す。通勤定期券ではどの曜日でも年間安定した通過人数が観測されている。通学定期券では学期に対応した推移であ



(a) S駅2007年平日(水曜)



(b) S駅2007年休日(日曜・祝日)

図1 1日出場人数の曜日別年間推移の例

るが、学期末～長期休暇～学期初の通過人数はなだらかに増減している。これは、学期と通学定期券の期間のずれによるものと推察できる。定期外は年間安定しているが、通勤定期券ほどではなく、休日は平日に比べると不安定である。また、休学期の平日は通過人数がやや多い。

図1より、曜日で各券種の1日合計通過人数が大きく異なることが分かる。一方、必ずしも暦日通り曜日によって単純に分類することが適当でない点も見受けられる。例えば、通勤定期であっても年末年始、大型連休などは平日でも通過人数が落ち込んでいる。また、本来の曜日が平日である祝日には通勤定期の通過人数が日曜日よりもやや多い傾向もある。

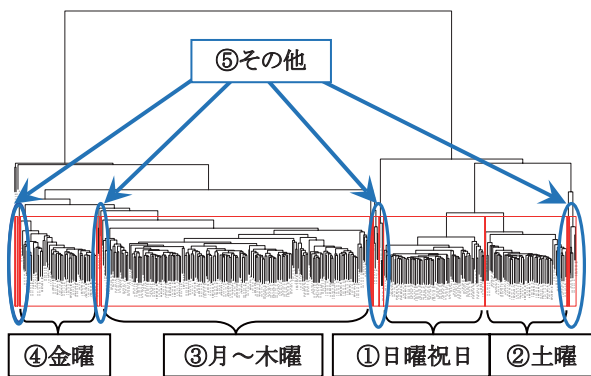
4.3.4 券種ごとの日付クラスタリング分析

1年間を曜日別に分類しても例外的な日が含まれることから、1日の時間帯別通過人数をもとに365日分のクラスタリング分析を行った。結果の系統樹を図2に示す。

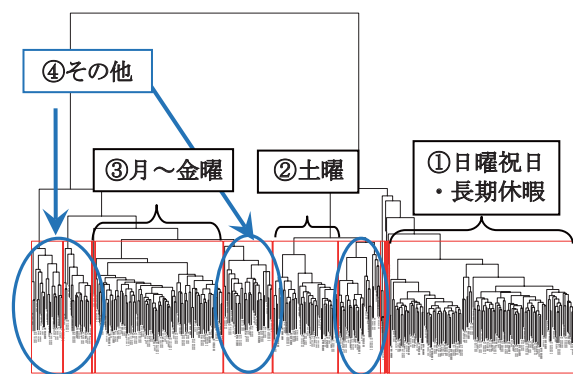
通勤定期の場合、土日と平日に分かれ、さらに図2(a)の①～⑤の5種類に分けることができる。⑤は、大型連休や大規模イベント日、大幅にダイヤが乱れた日などが該当する。また、系統樹の下方で各日が分岐していることから、グループ内各日の類似度が高いことがうかがえる。

定期外も通勤定期と同様のグループに分かれるが、連休のうち一部が他の日曜祝日と別のグループを形成する。グループ内の類似度は通勤定期ほど高くない。

通学定期の場合、通学期と休校日に大きく分かれ、さ



(a) P駅通勤入場



(b) P駅通学入場

図2 年間のクラスタリング分析例

らに図2(b)の①～④に分かれるが、通勤定期や定期外とは異なり、休みの日数が多く通常の平日の日数が少ない。また、④は学期末や試験期間など、利用人数や帰宅時間が通常と異なる複数のグループから成り立っている。

券種ごとのクラスタリング分析により、通常の日と時間帯別通過人数が特殊な日を分離できることが判った。

5. 時間帯別改札通過人数予測

3章で述べた旅客流動予測の第1段階のアプローチとして、時間帯別通過人数予測を試みた。予測手法および通過人数データを適用した結果について以下に述べる。

5.1 通過人数予測手法

4章では、以下の通過人数データの性質が判明した。

- 年内の通過人数は曜日以外に周期的変動はない。
- 同曜日のデータの類似度は高い（特に通勤定期）。
- 曜日ごとの分類では例外的な日が含まれる。

これを踏まえ、以下2種類の予測手法を考案した。

- ① 過去の類似データからの統計的な予測手法
- ② 当日利用経過からリアルタイムに予測可能な手法

5.1.1 過去の類似データからの統計的な予測手法

4.3.4項で行ったクラスタリング分析を踏まえ、過去の標準的なデータを参照することで通過人数の予測を行う。過去の類似データから未来の予測値を算出する方法に自己回帰分析（AR: Auto-Regression analysis）があり、最もシンプルな方法であるため、ここでは自己回帰モデルを用いた予測手法とした。

自己回帰モデルとは、古い順に x_1, \dots, x_n で与えられるデータ系列があるとき、式(1)のように過去の値の線形和である時点 t の値 x_t が定義されるモデルである⁹⁾。

$$x_t = \sum_{i=1}^p a_i x_{t-i} + \varepsilon_t \quad (1 \leq p \leq n, 1 \leq t \leq n) \quad (1)$$

x_t の誤差 ε_t を最小にする係数 a_1, \dots, a_p は Yule-Walker 方程式から算出でき、赤池情報量規準値をもとに最適次数 p が求められる¹⁰⁾。予測イメージを図3に示す。

参照する類似データは予め券種ごとに各日のクラスタ

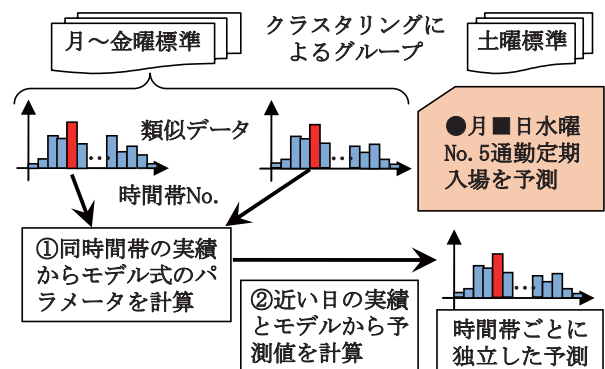


図3 自己回帰モデルによる予測イメージ

特集：輸送情報技術

リング分析を行い、曜日を基本としたグループに分類されている。例えば、図3のように水曜日の通勤定期入場人数を予測する場合、通勤定期入場のうち「月～金曜標準」のグループが参照すべきグループとなる。この参照グループ中の時間帯No.5の実績から、自己回帰モデルの最適次数と対応する係数を求める。モデル式に当日に近い過去の実績値を代入して得られる値を予測値とする。

5.1.2 利用経過からリアルタイムに予測可能な手法

4.3, 4項で行ったクラスタリング分析で例外的であるとされた日の予測にも対応するため、現時点までの利用経過をもとに、以後の通過人数の予測を行う。パターンマッチング (PM: Pattern Matching) はユークリッド距離などの指標をもとに2つの波形の類似度を算出することで、検索対象の波形と類似度の高い波形を特定するシンプルな方法であるため、ここでは予測当日の途中までの実績を検索対象とし、類似した日の時間帯別通過人数を当日の予測とするよう適用した。イメージを図4に示す。

予測当日の時間帯別通過人数は営業開始から「現時点」まで実績として既知であるものとする。例えば、図4のようにNo.1～5の通過人数が既知であるとする。既知の実績と、過去の実績 (No.1～5のみ) を比較し、実績と類似度が上位の数日の実績 (No.1～35の終日) を抽出する。抽出された時間帯別通過人数実績の平均をとったもののうちNo.6～35の値を当日の予測人数とする。

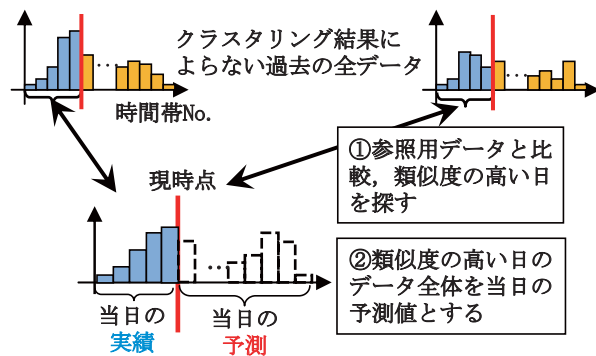


図4 パターンマッチングによる予測イメージ

5.2 通過人数予測結果および手法の評価

5.1節で挙げた2種類の予測手法の特性を比較し、手法の選択について考察する。

5.2.1 予測手法の適用条件

データを収集した8駅について、パターンマッチングと自己回帰モデルによる予測を行った。学習・参照データは2007年の1年間、検証用データは2008年6月中旬から9月下旬のデータとした。パターンマッチングの類似データ採用数 (近傍数) を1, 3, 5の3種類、予測当日の実績が明らかな時点を7, 8, 10, 15時の4種類とし、それぞれ予測した。自己回帰モデルにおいては、参照用データの各券種について線形和の最適次数を求めたところ、

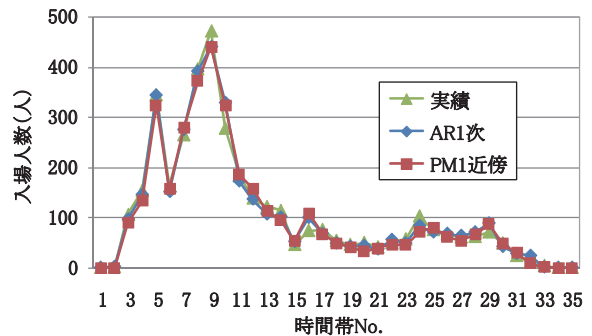
全てのケースで最適次数が1であったが、それ以外の次数での傾向を見るため、1～3次の3種類に対して予測した。

5.2.2 両手法の比較

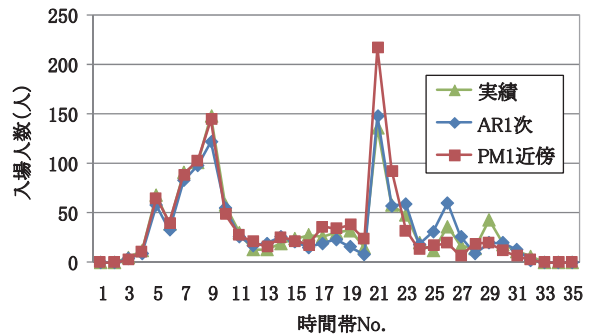
ここでは、データ取得駅のうち一般的な中間駅であるV駅を例にとり、パターンマッチングと自己回帰モデル双方による予測の傾向について述べる。

(1) 1日の傾向分析

両手法による予測結果の例を図5に示す。ただし、パターンマッチングの予測では、実績は8時まで既知とし、近傍数は1とする。自己回帰モデルの次数は1とする。



(a) 通勤入場人数



(b) 通学入場人数

図5 予測結果の例 (2008年6月26日 (木))

午前中の早い時間帯にピークを迎える通勤定期入場の場合、両手法とも実績と近い値を予測できたが、ピークが昼の定期外や夕方以降の通勤出場の場合は、パターンマッチングによる予測結果が実績と大きく外れている。一方、自己回帰モデルについては、予測する日と類似度の高いデータを参照していることから、誤差はデータ中に元から含まれる日々の変動分と考えられる。

通学定期は、通勤定期や定期外に比べて利用人数が少ないため、両手法とも予測誤差の割合が通勤定期や定期外より大きい。また、ピークの立ち上がりが遅い出場に関しては特にパターンマッチングの予測誤差が大きい。

(2) 長期的な傾向の分析

35の時間帯それぞれ実績値と予測値の誤差が生じるが、1日ごとの誤差の傾向を調べるため、複数時間帯の誤差率 e_a を式(2)で求める。

$$e_a = \sqrt{\frac{\sum_{i=a}^k (v'_i - v_i)^2}{\sum_{i=a}^k v_i}} \quad (2)$$

ここで、 k は時間帯の数(本研究のデータでは35)、 v_i は時間帯*i*における実績値、 v'_i は時間帯*i*における予測値を示す。これは、時間帯*a*から時間帯35までの予測値と実績値のユークリッド距離を、同じ時間帯分の総通過人数(実績)で正規化した値であり、曜日に関係なく統一的な基準で比較できる。

両手法の予測精度の比較のために、2008年6月中旬～9月下旬について予測を行った。誤差率を図6に、6月中旬～7月中旬の誤差率の傾向を表2に示す。

パターンマッチングの場合、近傍数が増えるに従い、誤差率が低くなり日ごとの散らばりも収束する傾向がある。最も類似したデータのみを採用するよりは、数個の類似データの平均の方が、再現性が高い場合があると考えられる。ただし、近傍数を多くしすぎると、類似していないデータまで採用し、却って精度が落ちる可能性もある。自己回帰モデルについても、次数が上がれば誤差率が低くなる傾向があるが、冗長なモデルとなる。

両手法を比較すると、曜日をキーにしている自己回帰モデルの予測では、お盆にかなり予測精度が落ちるため、特殊な時期はパターンマッチングの適用が望ましい。また、通学定期についても、夏休み周辺の平日は自己回帰モデルの精度が落ちる。通学定期は例外的な日が多いため、「夏休みは休日扱い」などの特例を設けることも必要と考えられる。一方、通勤定期と定期外について、

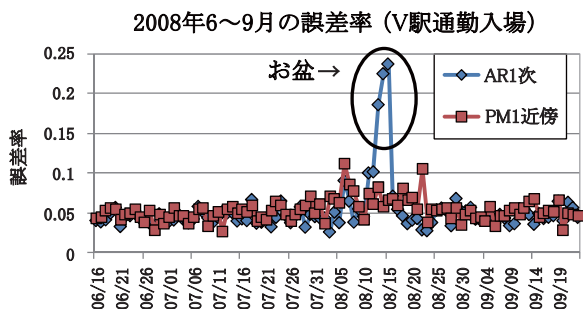


図6 誤差率の推移 (PMは8時までの実績を参照)

表2 誤差率の平均・標準偏差 (6月中旬～7月中旬)

誤差率の平均	AR1次	AR2次	AR3次	PM1近傍	PM3近傍	PM5近傍
通勤入場	0.0447	0.0417	0.0410	0.0468	0.0406	0.0395
通勤出場	0.0383	0.0336	0.0313	0.0419	0.0365	0.0349
通学入場	0.1258	0.1236	0.1182	0.1567	0.1084	0.1054
通学出場	0.1053	0.0956	0.0963	0.1367	0.1234	0.1230
定期外入場	0.0489	0.0415	0.0392	0.0562	0.0418	0.0403
定期外出場	0.0511	0.0446	0.0421	0.0525	0.0441	0.0420

誤差率の標準偏差	AR1次	AR2次	AR3次	PM1近傍	PM3近傍	PM5近傍
通勤入場	0.0063	0.0070	0.0063	0.0081	0.0082	0.0079
通勤出場	0.0131	0.0125	0.0120	0.0144	0.0128	0.0113
通学入場	0.0559	0.0544	0.0525	0.0499	0.0331	0.0368
通学出場	0.0373	0.0319	0.0332	0.0416	0.0386	0.0374
定期外入場	0.0103	0.0076	0.0071	0.0253	0.0103	0.0109
定期外出場	0.0092	0.0086	0.0086	0.0110	0.0117	0.0117

お盆などを除いた通常の時期は自己回帰モデルの方が誤差率の平均・標準偏差とも小さい傾向にあるため、誤差を補正する観点からも自己回帰モデルが望ましい。

表2は、8時までの実績をキーとしてパターンマッチングによる予測を行ったものであるが、それ以前の時間帯までしか実績が分からない場合は、特にピークの立ち上がりが遅い券種について、自己回帰モデルが優位である。

(3) 傾向分析を踏まえた手法の評価

(1) (2) から、予測手法について以下のことが分かった。

- ①時間帯：ピーク前の時間帯には自己回帰モデルによる予測が適当である。
- ②時期：お盆などの特殊な時期はパターンマッチングによる予測が適当である。
- ③券種：通学定期は安定した時期が短く、自己回帰モデルでは精度が落ちる。
- ④補正：自己回帰モデルの方が年間のぶれが小さく、安定した券種・時期では、前時間帯のずれを反映した補正に向いている。

これらの性質を利用して、券種・時期・時間帯に最適な予測手法を選択することが望まれる。例えば夕方にピークを迎える通勤出場の場合、ピークまでは自己回帰モデル、ピーク後はパターンマッチングによる予測が適していると考えられる。

6. 時間帯別通過人数からのOD推定

自動改札機からはODデータも得ることができるが、現状ではオンラインで取得することは不可能である。また、各駅のODデータがリアルタイムで取得できたとしても、時間帯別通過人数を発駅別に分割すると通過人数が少なくなり、5.1節で挙げた通過人数予測手法を直接適用することは困難である。そこで、3章で述べた旅客流動予測の第2段階のアプローチとして、時間帯別通過人数をもとにしたOD推定手法を開発した。

6.1 重力モデルを応用した駅間OD予測モデル

発駅の時間帯*i*の入場人数 o_i 、着駅の時間帯*j*の出场人数 d_j 、発駅から着駅への時間帯*j*に出场したOD x_j はデータから分かる。一方、*i*に発駅に入場し*j*に着駅に出场する人数 y_{ij} はデータからは直接分からない。

ここで、 o_i および d_j が多いほど y_{ij} は多くなり、*i, j*の間隔が大きいと y_{ij} は少なくなると予想されるため、式(3)のモデル式によりODの予測を行うこととした。

$$y_{ij} = \alpha o_i d_j \exp(-\beta c_{ij}) \quad (3)$$

ここで、 α および β は発駅と着駅の組合せに依存して決まる値であり、 c_{ij} は時間帯*i, j*の間隔や発着駅間の標準所要時分に依存するコスト関数である。 $x_j = \sum y_{ij}$ である

特集：輸送情報技術

ため、 x_j は式 (4) で表すことができる。

$$x_j = \alpha d_j \sum_i o_i \exp(-\beta c_{ij}) \quad (4)$$

6.2 パラメータ推定方法

前節で示したモデル式のうち、 x_j , d_j , o_i は自動改札機から取得可能なデータであり、 c_{ij} は時間帯 i, j により決まるコストである。それらの実績データから、式 (4) に最もよく当てはまるパラメータ α および β を非線形最小二乗法により求めることができる。得られたパラメータ α, β 、さらに駅の入出場人数の予測値である d_j, o_i を式 (4) の右辺に与えれば、左辺の時間帯別 OD x_j を間接的に予測することが可能となる。推定イメージを図7に示す。

なお、この α, β は発駅と着駅の組合せに依存して決まる値である。様々な発着駅の組合せについて α, β を推定できれば、旅客数が少ない発駅と着駅の組合せでも、入出場人数から時間帯別ODを予測することが可能となる。

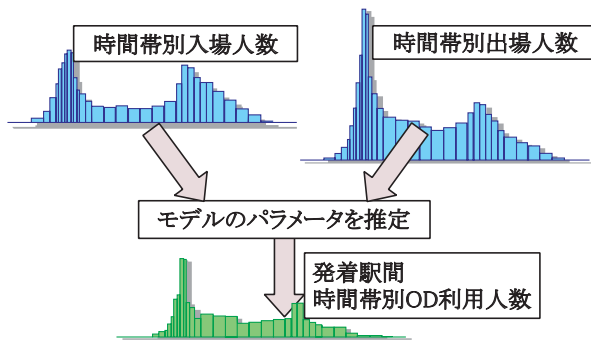


図7 入出場人数からのOD推定イメージ

6.3 通過人数をもとにしたOD推定の検証

1年間の参照用データ (T駅の時間帯別入場人数、P駅中央口の時間帯別出場人数、およびT駅からP駅中央口への時間帯別OD利用人数) を使って6.1節式(3)で表されるモデルのパラメータ α, β を非線形最小二乗法により推定したところ、 α, β ともに年間を通して安定した値が得られた。ただし、 β については、平日と休日で値に差があることも分かった。

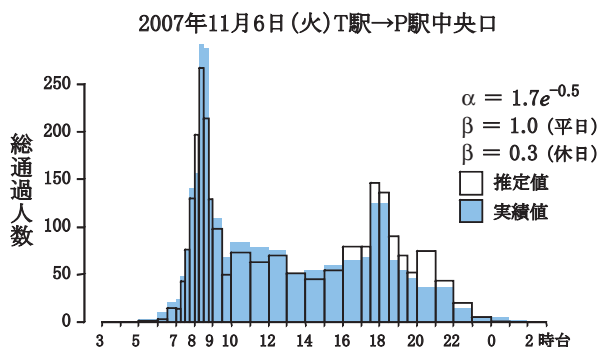


図8 OD予測モデルによる推定値と実績値

2007年11月6日(火曜)のT駅の入場人数と、P駅中央口の出場人数を用いて、モデルからT駅→P駅中央口の時間帯別OD利用人数を推定した。結果を図8に示す。

実績値と推定値の相関係数は0.93であった。朝のピーク後と夕方へのピーク後に若干乖離があるものの、高い精度でODを推定できていることが分かる。

7. おわりに

自動改札機から得られる過去の通過人数データ・ODデータをもとに、時間帯別通過人数や2駅間の時間帯別利用人数推定を行う手法を考案するとともに、各手法に実際のデータを用いることで手法の特性を検証した。今後は、旅客流動把握のための課題として、①予測精度の改善、②リアルタイムでのデータ取得、③駅数増加への対応、④ダイヤ乱れ対応、⑤予測の活用方法、について引き続き検討していきたい。

文献

- 1) 田中幹夫, 佐藤紀生, 佐久間靖, 眞船和敏, 後藤浩一: データマイニング手法による旅客流動データ等の分析と活用, 鉄道総研報告, Vol. 16, No. 11, pp.37-42, 2002
- 2) 松木恭太郎, 寺部慎太郎: 都市間旅客鉄道における時系列分析を応用した短期需要予測, 高知工科大学大学院修士論文, 2004
- 3) 小島泰昌, 川井慎吾, 佐藤敬範: 東海道新幹線の列車設定の今一 日 310 本以上走るようになった理由, JREA, Vol.49, No.8, pp.31825-31829, 2006
- 4) 田中英一, 長谷川淳, 伊藤正義: 重回帰分析と階層型ニューラルネットワークによる翌日電力需要予測, オペレーションズリサーチ, Vol.49, No.9, pp.499-503, 1996
- 5) 森啓之, 小瀬村紀行: データマイニング手法による短期電力負荷予測, 電気学会論文集, 121-B,2, pp.234-241, 2001
- 6) 山口類, 土屋映子, 樋口知之: 状態空間モデルを用いた飲食店売上の要因分解, オペレーションズリサーチ, Vol.49, No.5, pp.316-324, 2004
- 7) 伊藤稔: コールセンターにおけるインバウンド予測, UNISYS TECHNOLOGY REVIEW, Vol.87, No.5, pp.19-30, 2005
- 8) 明星秀一: 自動改札機データを活用した旅客流動推定手法, 鉄道総研報告, Vol. 20, No. 2, pp.351-354, 2006
- 9) 菅正信: ARMAモデルを用いた時系列解析, 数理設計研究所, 2004, http://www.madlabo.com/mad/book/REP040701_TimeSeries.pdf
- 10) 坂元慶行, 石黒真木夫, 北川源四郎: 情報量統計学, pp.142-149, 1983