

改札通過データを用いた旅客流動のリアルタイム推定アプローチ

○杉山 陽一 (鉄道総研)

松原 広 (鉄道総研)

田村 一軌 (福岡アジア都市研究所)

明星 秀一 (鉄道総研)

尾崎 尚也 (鉄道総研)

An Approach for Real-time Estimation of Railway Traveler Flow

Yoichi SUGIYAMA, Railway Technical Research Institute 2-8-38, Hikari-cho, Kokubunji-shi, Tokyo

Hiroshi MATSUBARA, Railway Technical Research Institute

Kazuki TAMURA, The Fukuoka Asian Urban Research Center

Shuichi MYOJO, Railway Technical Research Institute

Naoya OZAKI, Railway Technical Research Institute

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.

Keywords: traveler flow, OD-data, automatic ticket checking machine, passage data, pattern matching, autoregressive model

1. はじめに

大都市圏の鉄道において、旅客流動（本報告では時間帯・区間ごとの旅客数とする）を正確かつリアルタイムに把握することにより、旅客ニーズに合致した輸送力の提供や、状況に合った旅客案内が期待できる。本稿は、大都市圏の旅客流動を推定する手段として、過去の鉄道利用実績および当日の利用経過から駅利用人数や2駅間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について紹介する。

2. 関連する研究事例

過去に蓄積した実績値から需要予測を行う事例には、重回帰モデルによる改札通過旅客数予測¹⁾、自己回帰移動平均モデルによる新幹線の年間利用人数予測²⁾、決定木とニューラルネットワークで電力負荷を予測する事例³⁾などがある。

しかし、これらの手法の多くは総需要など特定のデータを対象にした予測や推定であり、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即時に確定しない属性（降水量など）が含まれる事例が多い。旅客流動の予測では多数の駅のデータを扱うことが想定されるため、よりシンプルで、かつ即時に得られる属性から予測できる手法の適用が求められる。

3. リアルタイム旅客流動予測のアプローチ

既往研究に、自動改札機から得られるODデータに基づき、各列車に利用人数を配分する事例⁴⁾がある。リアルタイムに時間帯別OD利用人数を予測できれば、結果を文献4)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。し

かし、現状では改札データをリアルタイムに取得できないため、本報告では自動改札機の改札通過人数を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る予測を試みた（Fig.1）。



Fig.1 Three Steps for Estimation of Traveler Flow

4. 改札データの分析

4.1 取得した改札データ

実際の通過人数を分析するため、下記8駅の自動改札機から直接改札データを取得した。

- ターミナル駅：P駅、Q駅、R駅
- 乗換駅：S駅、T駅
- 中間駅：U駅、V駅、W駅

2007年通年のデータを予測のための入力データとし、2008年6月～9月のデータを検証用とした。改札データには、通過人数データとODデータがあるが、前者は時間帯（No.1～35に分割されている）・券種ごとの入出場人数であり、後者はさらに発駅ごとに分類された出場人数である。また、券種を通勤定期・通学定期・定期外にまとめた。

4.2 年間の傾向分析

通過人数の年間の周期性を調べるため、2007年のデータについて離散フーリエ変換を行ったが、曜日以外に起因する周期性は見られなかった。さらに、時間帯別通過人数から曜日相互のユークリッド距離（各時間帯の人数差の2乗和）を求めた結果、日曜、平日、土曜どうしの類似性は高く、相互間の類似性は低いことが分かった。

次に、各曜日について券種ごとの傾向について調べた。1日合計出場人数の年間推移の例をFig.2に示す。

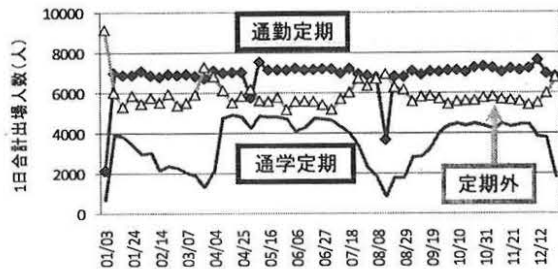


Fig.2 Annual Change of One-Day Passengers
(Station "S", Wednesdays)

通勤定期券については同じ曜日であれば年間安定した通過人数が観測されるが、平日以外では人数が大幅に減少する。通学定期券は学期と通学定期券の期間がずれるため、学期末・学期初の通過人数がなだらかに増減している。定期外は年間安定しているが、休学期の平日は通過人数がやや多い。

Fig.2などから、曜日と券種によって1日の通過人数が大きく異なることが分かったが、年末年始や祝日など必ずしも暦日通りの曜日による分類が適当でない点も見受けられた。

4.3 券種ごとの日付クラスタリング分析

例外的な日と標準的な日を分類するため、2007年の各日の時間帯別通過人数をもとに365日分のデータについてクラスタリング分析を行った (Fig.3)。

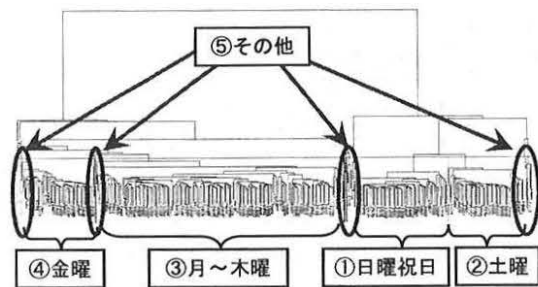


Fig.3 Clustering Result of 365-day Data
(Station "P", Commuters)

通勤定期の場合、土日と平日に分類でき、さらにFig.3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、系統樹の下方で各データ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。定期外では、通勤定期と同様のグループに分かれ、一部の連休が他の日曜祝日と別のグループに分類される。また、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、さらに各曜日に分かれるが、休みの日数が多く通常の平日の日数が少ない。

券種ごとのクラスタリング分析により、通常の日と通過人

数が特殊な日を分離でき、券種の特徴が詳しく分かった。

- 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)。
- 曜日のみの分類では例外的な日が含まれる。

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。2つの予測手法および通過人数データを適用した場合の予測結果について述べる。

5.1 過去の類似データからの統計的な予測手法

4.3節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予測値を算出するシンプルな方法として自己回帰 (AR: Auto-Regression analysis) モデルを用いる。

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、ある時点 t の値 x_t が式(1)のように過去の値の線形和で定義されるモデルである⁶⁾。

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

誤差 ε_t を最小にする $a_1, \dots, a_n$ と最適次数 p の求め方は文献⁶⁾で述べられている。予測イメージをFig.4に示す。

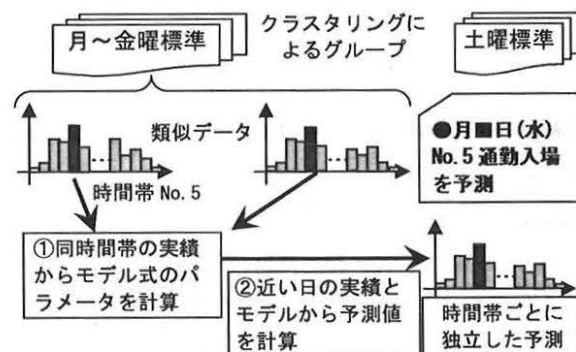


Fig.4 Image of Estimation by Auto-Regression Model

参照する類似データは予め券種ごとのクラスタリング分析により、曜日を基本としたグループに分類されている。Fig.4で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「月～金曜標準」が参照するグループとなる。この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およびそれに対応する係数を求める。モデル式に当日に近い日の実績値を代入して得られる値を予測値とする。

5.2 利用経過からリアルタイムに予測可能な手法

4.3節で行ったクラスタリング分析で例外的であると判明し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方法として、検索対象の波形と類似度の高い波形をシンプルに特定するパターンマッチング (PM: Pattern Matching) を用いた。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。

例えば、Fig.5で予測当日の時間帯 No.1～5 の通過人数が既知であるとするとき、既知の実績と過去の実績 (No.1～5の部分のみ) を比較し、実績と類似度が高い実績 (終日: No.1～35) を抽出する。こうして抽出した時間帯別通過人数実績の平均のうち No.5 以降 (No.6～35) の値を予測人数とする。

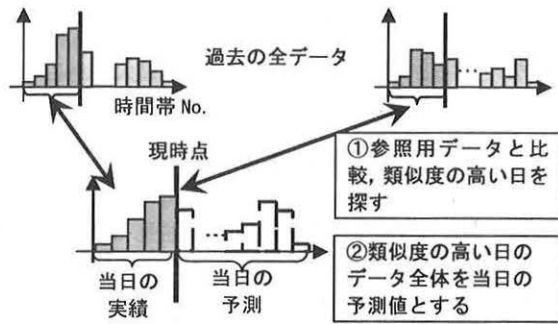


Fig.5 Image of Estimation by Pattern Matching

5.3 通過人数予測結果および手法の評価

5.1節と5.2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。

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・参照用データは2007年通年、検証用データは2008年6月中旬～9月下旬のものとした。自己回帰モデルについては、全てのケースで最適次数が1であった。ここでは、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

両手法での予測結果をFig.6に示す。ただし、自己回帰モデルの次数は1とし、パターンマッチングでは8時までの実績を既知、抽出する類似データ数（近傍数）を1とする。

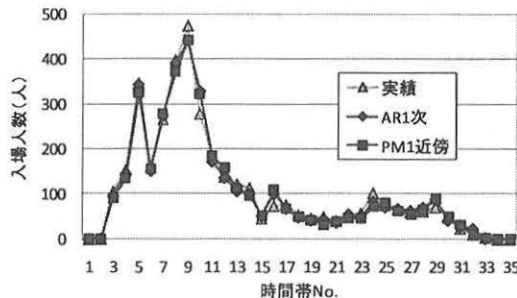


Fig.6 Result of Estimation of Passengers by Each Time Section (Commuters on 26th, June, 2008, Thursday)

早い時間帯にピークを迎える通勤定期入場では、両手法とも実績と近い値を予測できたが、予測している時点よりもピークが遅いケースでは、パターンマッチングによる予測結果が実績と外れやすい傾向にある。一方、自己回帰モデルでは、予測する日と類似度の高いデータを参照するため、誤差は日々のデータに含まれる変動分と考えられ、予測精度が良い。

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で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

曜日や通過人数の大小に関係なく誤差の傾向を比較できるようにするため、誤差率 e_a を式(2)で定義する。

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

ここで、 k は時間帯の数（本報告では35）、 v_i は時間帯 i の実績値、 $\hat{v}_i$ は時間帯 i の予測値である。 e_a は、時間帯 a か

らNo.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、通過人数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月下旬について誤差率を計算した（Fig.7）。

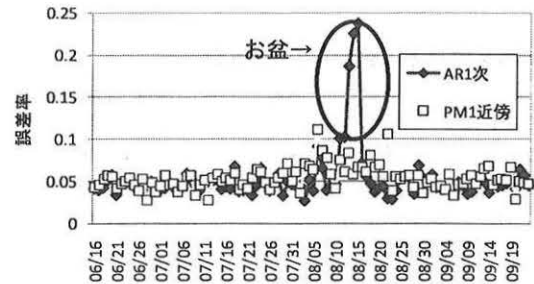


Fig.7 Change of Error Rate (PM Refers Data 3:00-8:00)

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ため、夏休みを休日扱いするなどの条件設定も必要と考えられる。一方、通勤定期と定期外について、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、基本的に自己回帰モデルの適用が望ましい。

6月中旬～7月中旬の誤差の傾向をTable 1に示す。

Table 1 Average and S.D. of Error Rate (from Middle of June to Middle of July)

誤差率 の平均	AR 次数			PM 近傍数		
	1	2	3	1	3	5
通勤入場	0.045	0.042	0.041	0.047	0.041	0.039
通勤出場	0.038	0.034	0.031	0.042	0.036	0.035
通学入場	0.126	0.124	0.118	0.157	0.108	0.105
通学出場	0.105	0.096	0.096	0.137	0.123	0.123
定期外入場	0.049	0.041	0.039	0.056	0.042	0.040
定期外出場	0.051	0.045	0.042	0.052	0.044	0.042

誤差率の 標準偏差	AR 次数			PM 近傍数		
	1	2	3	1	3	5
通勤入場	0.006	0.007	0.006	0.008	0.008	0.008
通勤出場	0.013	0.012	0.012	0.014	0.013	0.011
通学入場	0.056	0.054	0.053	0.050	0.033	0.037
通学出場	0.037	0.032	0.033	0.042	0.039	0.037
定期外入場	0.010	0.008	0.007	0.025	0.010	0.011
定期外出場	0.009	0.009	0.009	0.011	0.012	0.012

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、近傍数を多く設定しすぎると類似していないデータまで抽出され、逆に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、次数を上げすぎるとAICの値が悪化し、冗長なモデルとなる。

(3) 傾向分析による予測手法の考察

(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：ピーク前は自己回帰モデルが適当である。

②時期：お盆など特殊な時期に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：自己回帰モデルでは年間のぶれが小さく、安定した条件では前時間帯の誤差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。例えば夕方がピークの通勤出場では、ピーク前は自己回帰モデル、ピーク後はパターンマッチングを採用するなどの選択が考えられる。

6. 時間帯別通過人数からの OD 推定

4.1 節で自動改札機から OD データも得ることができたと述べたが、現状ではオンラインで取得することは不可能である。仮にリアルタイムで取得できたとしても、OD の組合せ数が膨大であり、OD ごとの人数が少なくなる。利用人数が少ない OD に 5.1 節と 5.2 節で挙げた通過人数予測手法を適用すれば予測精度が悪くなるおそれがある。

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、オンラインで取得できる時間帯別通過人数を代入することで OD 利用人数を推定するモデルを開発した。

6.1 重力モデルを応用した駅間 OD 予測モデル

発駅から時間帯別入場人数が、着駅から時間帯別出場人数がオンラインで得られる（本研究では自動改札機から直接取得した）。また、着駅の自動改札機からは OD データとして発駅からの時間帯別出場人数が得られる。

発駅での時間帯 i の入場人数を o_i 、着駅での時間帯 j の出場人数を d_j とし、 i に入場し j に出場する利用人数を y_{ij} とすると、以下の関係が予想される。

- y_{ij} は o_i および d_j に比例する。
- i, j の間隔が大きいと y_{ij} は少なくなる。

これらの関係により、式(3)を仮定する。

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

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である。また、時間帯 j に出場した発駅からの OD 利用人数を x_j とすると、 $x_j = \sum_i y_{ij}$ であるため、発着駅から得られる値 o_i, d_j, x_j の関係は式(4)で表わされる。これを式(4)のようにモデル化して推定を行う。

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

6.2 パラメータ推定方法

一定期間のデータから入出場人数と OD 利用人数の組が得られるため、式(4)に最もよく当てはまるパラメータ α, β を求めることができる。得られたパラメータ α, β および発着駅の入出場人数の予測値である d_j, o_i を式(4)に代入すれば、時間帯別 OD 利用人数 x_j を間接的に予測することが可能となる。この推定イメージを Fig.8 に示す。

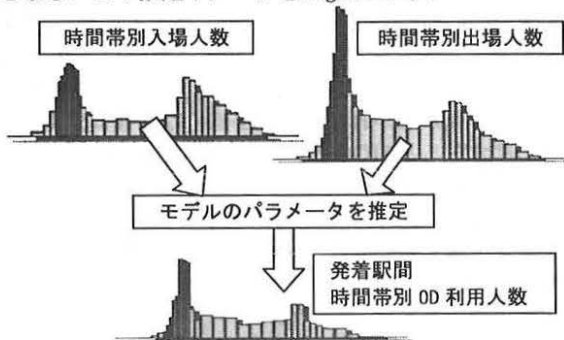


Fig.8 Image of OD-Estimation from Passage Data

なお、 α, β の最適値は発着駅の組合せに対して固有に決まる。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。

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 駅の出場人数、および T 駅から P 駅への時間帯別 OD 利用人数）を使って式(3)で表されるモデルのパラメータを推定したところ、 α, β ともに年間を通して安定した値が得られた。

2007 年 11 月 6 日の T 駅の入場人数と P 駅の出場人数を用いて、T 駅→P 駅の時間帯別 OD 利用人数を推定した (Fig.9)。

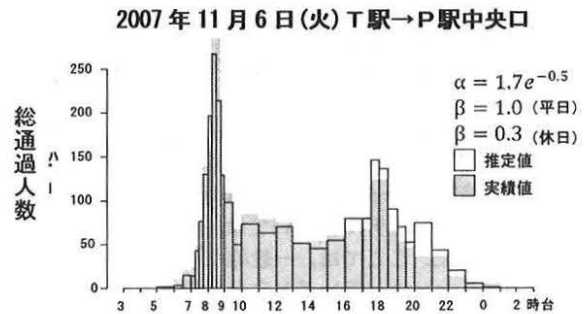


Fig.9 Result of OD-Estimation by Suggested Model

実績値と推定値の相関係数は 0.93 であった。夕方にピーク後に乖離が見られるが、高い精度で推定できていることが分かる。

7. 終わりに

自動改札機から得られる過去のデータをもとに、時間帯別通過人数予測や 2 駅間の時間帯別利用人数推定を行う手法を考案し、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、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、などの課題について引き続き検討したい。

謝辞

本研究にあたり、西日本旅客鉄道株式会社には、改札機データのご提供など、多大なご協力を頂きました。厚く御礼申し上げます。

参考文献

- 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) 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
- 4) 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
- 5) 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
- 6) 坂元慶行, 石黒真木夫, 北川源四郎: 情報量統計学, pp.142-149, 1983