

少量データ環境における人間の知識を取り入れた 交通需要予測手法の開発

Forecasting Transportation Demand in Scenarios with Prior Knowledge and Limited Data

井波玄龍^{1*} 稗方和夫^{1,2}
Genryu Inami¹ Kazuo Hiekata^{1,2}

¹ 東京大学工学部システム創成学科

¹ Systems Innovation, Faculty of Engineering, The University of Tokyo

² 東京大学大学院新領域創成科学研究科

² Graduate School of Frontier Sciences, The University of Tokyo

Abstract: Demand forecasting based on past data sets is useful for efficient operation of transportation. However, it is difficult to prepare an enough amount of data right after changes in the environment such as the spread of COVID-19 and a change of a fare. In this study, we propose a new demand forecasting method that incorporates qualitative effects of changes in the environment and estimates parameters of the model using past data. In the case study, it is shown that the proposed method can predict more accurately when an amount of data is limited.

1 はじめに

需要予測では、過去のデータに基づき、トレンドや周期性、ノイズなどを説明する統計的モデルを構築し、そのモデルを外挿することによって予測値を算出する。しかし、モデルに含まれない外部の要因により需要が大きく変動する場合など、需要データを説明するのに必要なデータ量が不足している場合、十分な精度が得られないことがある [1]。そのような少量データ環境において、外部要因の影響の大きさに関する人間の知識をモデルに取り入れることで予測精度の改善を目指す研究は小売業等で数多く存在する。一方で、交通需要予測において知識を取り入れたヒューリスティックなモデルを開発している例は多くない。そこで本研究は、人間の知識を取り入れた交通需要予測を行うための手法を開発することを目的とする。

2 提案手法

図 1 に提案手法の概要を示す。以下の段階に分けて需要の予測を行う。まず需要データを利用頻度に基づいて二分割する。高頻度のユーザーによる需要を人間の知識を取り入れたヒューリスティックな需要モデリ

ング手法により予測する。また低頻度のユーザーによる需要を統計モデリングにより予測する。

1. 対象とする需要の時系列分析を行う。
2. 需要全体を、統計的に予測できる部分と人間による判断が必要な部分に分割し、それぞれにおいて予測を行う。
3. 2つの予測値を足し合わせることで最終的な全体の予測値とする。

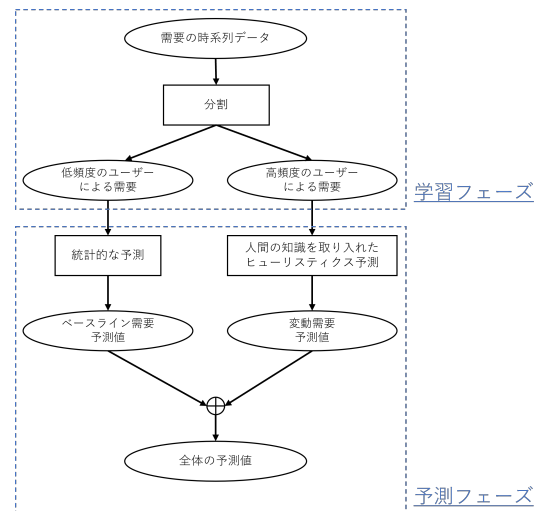


図 1: Overview of the proposed method

*連絡先：東京大学大学院新領域創成科学研究科
〒 277-8563 千葉県柏市柏の葉 5-1-5 環境棟 272 号室
E-mail: inami-genryu@g.ecc.u-tokyo.ac.jp

2.1 交通需要データの分割

本手法では、対象とする交通機関を利用する頻度によってユーザーを二分割する。頻繁に交通機関を利用するユーザーの行動が交通機関の運行上の変更や、感染症の蔓延などに影響を受ければ、全体の需要が変動する可能性は高い。一方、交通機関を利用する頻度が低いユーザーの発生させる需要は突発的であり、個別の予測が困難である上に全体の需要値への影響は小さい。高頻度のグループを H 、低頻度のグループを L とし、期間 p 、回数 C を用いて次のように定義することができる。

$$H = \{i \mid \text{ある期間 } p \text{ における利用回数が } C + 1 \text{ 以上}\}$$

$$L = \{i \mid \text{ある期間 } p \text{ における利用回数が } C \text{ 以下}\}$$

また時刻 t における個人 i の需要を $d_i(t)$ とすれば、全体の需要は式 1 のように分割できる。

$$D(t) = D_H(t) + D_L(t) = \sum_{i \in H} d_i(t) + D_L(t) \quad (1)$$

1 式の第一項は人間の知識を取り入れた提案モデルによりモデリングし、第二項は状態空間モデルによりモデリングする。

2.2 利用頻度が低いユーザーによる需要の統計的モデリング

$D_L(t)$ を状態空間モデルによりモデリングする。一般に交通機関の需要を示す時系列データは曜日ごとの周期的な変動を示すと考えられる。

よって、以下の式によってモデリングする。

$$x_t = x_{t-1} + s_{t-1} + \beta \quad (\text{状態モデル}) \quad (2)$$

$$D_L(t) = x_t + w_n \quad (\text{観測モデル}) \quad (3)$$

ただし β は一定のトレンドを表す定数であり、 w_n は平均 0、分散 σ^2 の正規分布に従うとする。また s_t は曜日成分を表す。

2.3 利用頻度が高いユーザーによる需要のヒューリスティックモデリング

H に属する個人 i ごとに $d_i(t)$ を提案モデルによりモデリングする。本手法では、過去の行動履歴や外生変数などに対する個人の行動の応答が線形であると仮定し、図 2 における個人の行動決定関数は、ロジスティック

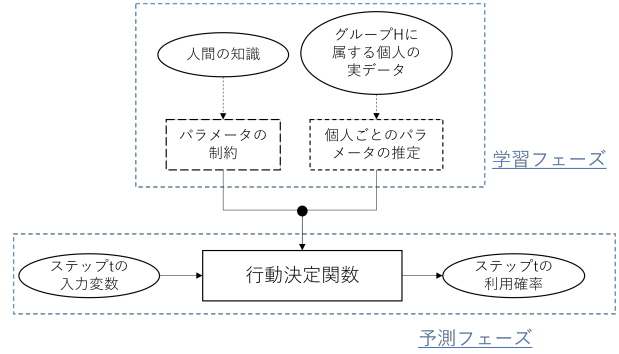


図 2: Predicting demand by frequent users

ク回帰モデルを用いて構築する。ロジスティック回帰モデルは以下の式で記述できる。

$$p_i = \frac{\exp(X_i)}{1 + \exp(X_i)} \quad (4)$$

$$X_i = \beta_{i0} + \beta_{i1}x_{i1} + \beta_{i2}x_{i2} + \beta_{i3}x_{i3} + \dots \quad (5)$$

ただし、 β_{ik} は k 番目のパラメータ、 x_{ik} は k 番目の入力、 p は出力である。パラメータの推定は制約付き非線形最適化問題として定式化でき、実数値遺伝的アルゴリズムを用いて計算する。

2.4 人間の知識のモデル化

需要に影響を及ぼしうる外的要因は判断要因と呼ばれ [2]、図 3 のように分類できる。

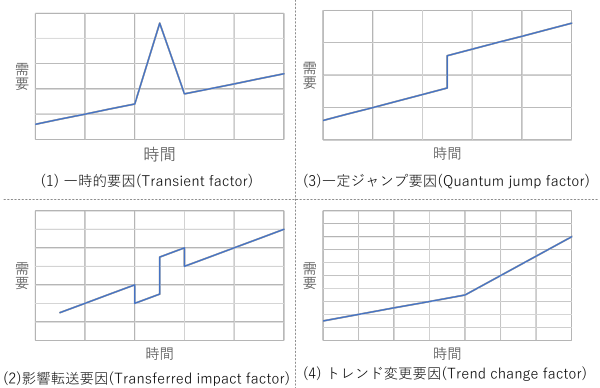


図 3: The categorization of judgmental factors by Marmier et al.[3]

1つ目は一時的要因と呼ばれ、祭りの開催など一時的に需要を変化させるものである。2つ目は影響転送要因と呼ばれ、このような需要の変化の形は値下げキャンペーンの実施が事前に発表されたときなどに観測される。3つ目は一定ジャンプ要因と呼ばれ、新路線の追加など需要が一度に変化するとき観測される。最後はトレンド変更要因と呼ばれ、運賃の値下げなど需要が一定割合増加する場合に観測される。

判断要因	追加する入力変数の例	パラメータの制約の例
(1)一時的要因	$x_{(1)} = \begin{cases} 0 & (t \neq T) \\ 1 & (t = T) \end{cases}$	$\beta_{(1)} > 0$
(2)影響転送要因	$x_{(2)} = \begin{cases} 0 & (t < T_0, t \leq T_2) \\ -1 & (T_0 \leq t < T_1) \\ 1 & (T_1 \leq t < T_2) \end{cases}$	$\beta_{(2)} > 0$
(3)一定ジャンプ要因	$x_{(3)} = \begin{cases} 0 & (t < T) \\ 1 & (t \geq T) \end{cases}$	$\beta_{(3)} > 0$
(4)トレンド変更要因	$x_{(4)} = \begin{cases} 0 & (t \leq T) \\ T-t & (t \geq T) \end{cases}$	$\beta_{(4)} > 0$

図 4: Input variables that represents judgmental factors

提案手法では判断要因をロジスティック回帰関数の入力変数として取り込む。係数パラメータの符号を制限し、人間の知識に適合するようにパラメータ推定を誘導する。具体的な変数の値、係数の符号の制限は図 4 に示した。

3 ケーススタディ

提案手法を実データに適用し、予測精度を検証した。データは成田市のオンデマンド交通の運行実績データを用いた。

データの分割においては簡単に p を一ヶ月、 C を一日とする。

成田市オンデマンド交通では 2020 年 4 月 1 日に月に 8 回まで利用回数制限が施行されている。加えて 2020 年 3 月末から COVID-19 の蔓延が始まっており、同時期に需要パターンが大きく変化している。利用回数制限、COVID-19 の流行の影響はそれぞれ図 5、図 6 に示したような構造を持つとして判断した。

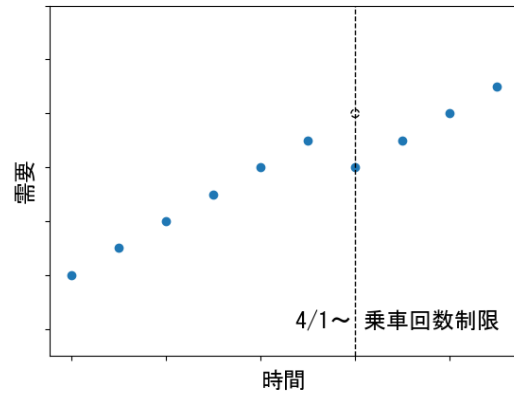


図 5: The estimated effect of the ride restriction

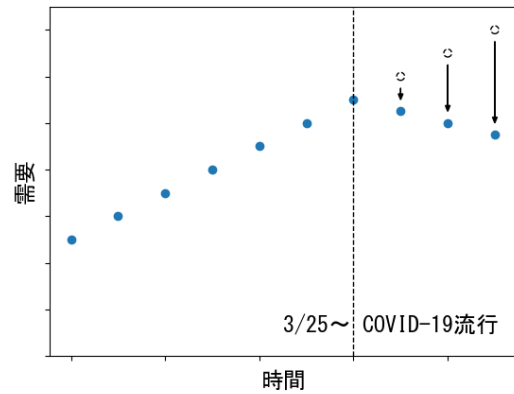


図 6: The estimated effect of the spread of covid-19

具体的には、表 1 に示したように入力変数およびその係数パラメータの制限を決定した。

表 1: The input variables of the prediction model

入力変数	取りうる値	係数パラメータの制限
COVID-19 感染者数	感染者数	$\beta < 0$
乗車回数制限	0 or 1	$\beta < 0$

以上に基づき、 H に属する個人 i ごとに $d_i(t)$ を提案モデルによりモデリングした。

2020 年 3 月 1 日から 14 日までのデータを用いてパラメータを推定した。フィッティングの結果を図 7 に示す。ただし 3 月 20 日は祝日のため運行が休止されていた。

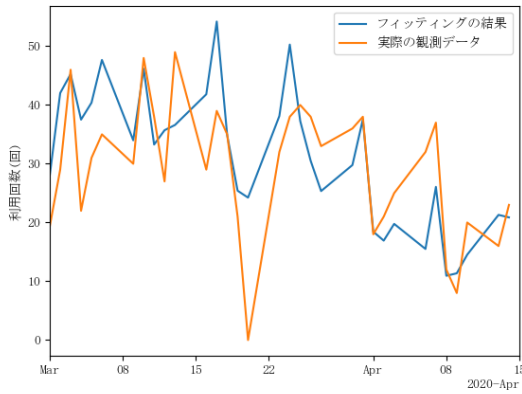


図 7: Fitting proposed model that predicts use

以上のモデルを用いて 4 月 15,16,17,20,21 日の需要の予測を行った。ヒューリスティックを交えた提案手法による予測値の MAE が 10.7753、状態空間モデルのみを用いたときの予測値の MAE が 11.5830 であり、与えたデータにおいて提案手法の精度が良いことを確認できた。

3.1 考察

提案手法におけるデータの分割の有効性を検証するために、 $D(t)$ および $D_L(t)$ の統計モデルへの当てはまりをそれぞれ算出した。すると、分割後の $D_L(t)$ データに対してフィッティングされた統計モデルのほうが対数尤度が大きく、当てはまりが良かった。また、推定されたトレンドの分散も分割後の $D_L(t)$ に対するモデルにおける値の方が小さくなっており、外的要因によるトレンドの変化が除外できていることが分かる。

また、データ量の大小が提案手法による予測精度に与える影響の検証を行った。図 8 に、利用するデータを制限したときの予測精度の変化を示す。「32 日分」とある点は 32 日分のデータを提案モデルの学習に使用し予測を行ったときの予測誤差を示している。その他の点も同様である。提案手法による予測精度はいずれの場合も純粋な統計モデルよりも高いことがわかる。

以上より、少量データ環境において本提案手法に基づいて人間の知識を取り入れたモデルを構築することにより、高い精度で需要予測を行うことが可能であることが示された。

4 結論

本研究では、人間の知識を取り入れた交通需要予測を行うための手法を提案し、開発した。ケーススタディ

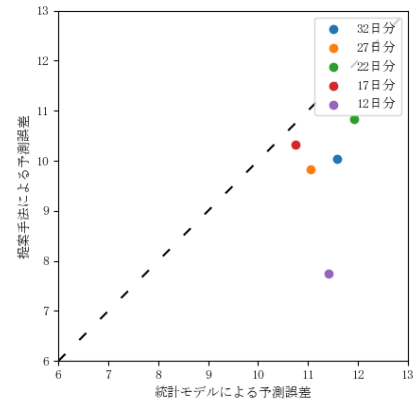


図 8: Forecast error based on reduced data

において、提案手法による需要予測の精度を検証し、ベースラインとした状態空間モデルよりも精度よく需要予測ができることを示した。

参考文献

- [1] Meysam Arvan, Behnam Fahimnia, Mohsen Reisi, and Enno Siemsen. Integrating human judgement into quantitative forecasting methods: A review. *Omega*, Vol. 86, pp. 237–252, 2019.
- [2] Jae Kyu Lee, Sang Bong Oh, and Jung Cheol Shin. Unik-fcst: Knowledge-assisted adjustment of statistical forecasts. *Expert Systems with Applications*, Vol. 1, No. 1, pp. 39–49, 1990.
- [3] François Marmier and Naoufel Cheikhrouhou. Structuring and integrating human knowledge in demand forecasting: a judgemental adjustment approach. *Production Planning and Control*, Vol. 21, No. 4, pp. 399–412, 2010.