

プログラマブル API を用いた宿泊者数予測分析

諸田健太郎[†] 小川滉介[†] 松井伸司[‡] 佐藤彰洋[‡]

横浜市立大学データサイエンス学部[†] 総務省統計改革実行推進室[‡]

1. はじめに

近年コロナウイルスにより観光業の経営は厳しい状況に置かれていて、少しでも早く復興することが求められている。そこで効果的な観光施策を行うために、現在の観光業の状況についてより早く正確に知る必要がある。

総務省と横浜市立大学は共同研究として、流動人口データ[4]を活用して宿泊者数の推計を行い、宿泊旅行統計調査[3]の公表を近似的に速報化する方法を研究している。宿泊旅行統計調査とは、国土交通省が全国の宿泊施設を対象に延べ宿泊者数などを毎月調査するものである。ここでは各都道府県の日本人のべ宿泊者数を統計値と定義する。統計値としては第二次速報値というものを使っていて、現在の月より2ヶ月~3ヶ月前のデータとなっている。それに対して流動人口データは1ヶ月前のデータを入手できるため、速報性を保つことができる。また、この研究では宿泊者数の説明変数として、指標値というものを定義する。指標値とは流動人口データの朝4時のデータから国勢調査の常住人口を引いた値で、宿泊者数を概算した値である。[1]

$$\text{指標値} = \sum_i (4\text{時流動人口}(i) - \text{常住人口}(i)) \quad (i: \text{メッシュコード})$$



図1 指標値の計算方法[1]

宿泊者数推計の流れとしては、まず分析に使う統計値と指標値のデータを揃える。その後回帰分析などの手法を使ってモデルを作成する。最後にモデルを適用させて予測を行い、結果の評価を行う。

この研究の問題としては、分析するのに時間がかかること、データの再現性が保てないこと、開発作業が共同でやるのが難しいことがあげられる。

本研究では、MESHSTATS[3]上でプログラマブル API の発想を通して分析自動化プログラム作成をおこなった。

Logger count analysis using programmable API
 Kentaro MOROTA[†], Hirotsuke OGAWA[†], Aki-Hiro Sato[†], Shinzi Matsui[‡]
[†]Yokohama City University, [‡]Ministry of Internal Affairs and Communications

2. MESHSTATS

MESHSTATS とは、日本産業規格地域メッシュコード(JIS X0410)の独自拡張である世界メッシュコードに基づく多言語対応 DIaaS(Data Infrastructure as a Service)である。[3]世界メッシュコードとは、緯度経度を基にして地球上の平面をタイルのようにして区切り、番地のようにして番号のようなものを割り振っているものである。MESHSTATS では世界メッシュコードに紐づいた、その区域の常住人口や飲食店の数、標高などのさまざまなデータを保管していて、扱うこともできる。データに基づく分析などができるアプリケーションの開発や閲覧もすることができて、例えば、該当するメッシュコードの周辺地域の人口分布などを色付きで理解したり、コロナウイルスの感染者数を自治体ごとに把握して人数を予測したりすることができる。

3. 提案手法

プログラム全体の流れとしては2つのプログラムを作成した。一つは自動区分線形回帰プログラムで、もう一つは分析結果表示プログラムである。特に自動区分線形回帰プログラムの部分はプログラマブル API の発想で開発をしており、2つのプログラムは共同で独立に開発して動かすことができる。

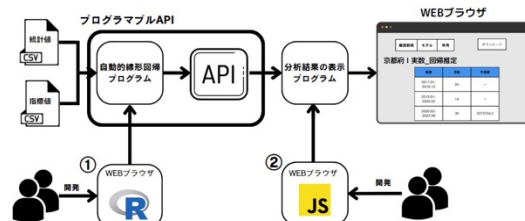


図2 プログラム全体の流れ

3.1 プログラマブル API

プログラマブル API とは、API で受け渡す情報をプログラム可能にすることにより、共同で独立した開発作業を行えるようにするものである。本研究では R プログラムで分析した結果を作成し、PHP を通してデータを受け渡せるように開発をおこなった。

3.2 自動区分線形回帰プログラム

宿泊者数の予測の流れとして、まず統計値と指標値を用いて、区分的線形回帰をおこなったモデルを作成する。そのモデルを適用して結果の反映と評価を行う。区分的線形回帰とは、時系列データのトレンドの切り替わりを自動で検知を行い、区間を分割した上で回帰分析を

行う方法である。切り替わる区間を検知するために尤度比検定を行う。

尤度比検定では分割する左右の区間での対数尤度の差の二倍を算出してから p 値を算出して尤度比検定を行う。 p 値はモデルパラメータ数の差である 2 を自由度とするカイ二乗分布に従う。分割区間を切り替えて、 p 値が有意水準を棄却するかつ、 $D(i)$ が最大となる箇所を見つけて分割を行うことを繰り返す。

$$y_t = a + bx_t$$

$$y_{L,t} = a_L + b_L x_{L,t}, \quad y_{R,t} = a_R + b_R x_{R,t}$$

$$D(i) = 2 \left(\sum_{t=1}^i \ln p(\varepsilon_i; \mu_L, \sigma_L^2) + \sum_{t=i+1}^T \ln p(\varepsilon_i; \mu_R, \sigma_R^2) - \sum_{t=1}^i \ln p(\varepsilon_i; \mu, \sigma^2) \right)$$

$$\approx -i \ln \sigma_L^2 - (T-i) \ln \sigma_R^2 + T \ln \sigma^2$$

$$p = \exp \left(-\frac{\max_i D(i)}{2} \right) \quad p(\varepsilon; \mu, \sigma^2) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp \left(-\frac{(\varepsilon - \mu)^2}{2\sigma^2} \right)$$

(誤差は正規分布に従うと仮定する)

図3 区分的線形回帰における尤度比検定

実際の分析では、実数、前年同月比、対数前年同月比の3種類について分析を行う。実数については、指標値と統計値をそのまま用いて区分的線形回帰を行う。前年同月比分析に関しては指標値と統計値の前年同月比を用いて区分的線形回帰を行う。前年同月比の予測値を実数に割り戻した際に、マイナスになる可能性があるため、注意が必要である。対数前年同月比分析に関しては、指標値と統計値の前年同月比を対数にして 1000 を掛けたものを用いて区分的線形回帰を行う。予測値を実数に割り戻した際にマイナスになることはない。しかし、戻した後の値の誤差が大きくなってしまふことが懸念される。

回帰分析						
	from	to	相関係数	傾き	切片	
①	2017.1	2021.9	0.887	0.513	-528.342	
②	2019.10	2021.9	0.940	0.530	-596.933	
③	2020.4	2021.9	0.973	0.643	-856.662	

自動回帰分析結果						
	from	to	相関係数	傾き	切片	
④ $\alpha=0.01$	2017.1	2021.9	0.886	0.511	-523.620	
⑤ $\alpha=0.05$	2017.11	2021.9	0.907	0.515	-554.834	

採用値						
	from	to	相関係数	傾き	切片	
⑤ $\alpha=0.05$	2017.11	2021.9	0.907	0.515	-554.834	

推計結果						
中心	下限	上限	計算方法	統計値	乖離率	評価
1,132,515	1,019,263	1,245,766	⑤	1,149,620	-1.5%	○

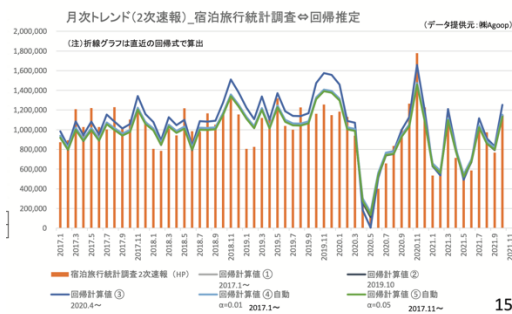


図4 京都府の宿泊者数分析の例[1]

区分的線形回帰の結果の例として、京都府の2017年1月~2021年9月のデータから2021年10月の統計値の予測を実数でおこなった結果は図4の通りである。表の④

⑤のモデルが有意水準を変えたもので、分割した区間の中で最新の区間のものを選んでいいる。有意水準 0.01 の場合は分割されていないが、有意水準 0.05 の場合は分割された。推計結果としては、分析モデル⑤が乖離率が一番小さく、推計値は 113.2 万人で統計値は 114.9 万人となり、統計値から見た乖離率は-1.5%という小さい値に収まった。

3.3 分析結果の表示プログラム

分析の表示設定としては、都道府県の指定と年月の設定をすることで分析結果を表示できるようにする。また、結果をダウンロードできるようなボタンも設定しておく。

まず API を用いて自動区分的線形回帰プログラムから分析結果を取得する。その後、結果をまとめてブラウザに表示する。ダウンロードボタンが押された際には csv ファイルで出力を行う。

4. まとめ

本研究では、宿泊者数の予測分析を自動化、MESHSTATS 上でプログラマブル API 機能を設計・実装、区分的線形回帰のアルゴリズムを API で実装した。今後の取り組みとしては、分析の完全自動化と評価基準の設定が挙げられる。分析の完全自動化についてはプログラマブル API を通して手作業で行われる部分を一つ一つ取り除いていきたい。また、評価基準についてはより納得のいくものを定めていきたい。

参考文献

- [1] 総務省統計改革実行推進室： 人流データを活用した宿泊旅行統計調査延べ宿泊者数の推計 https://www.soumu.go.jp/main_content/000791551.pdf (最終アクセス年月日: 2022年12月14日)
- [2] 一般社団法人世界メッシュ研所: MESHSTATS, <https://www.fttsus.org/worldgrids/ja/meshstats-ja>, (最終アクセス年月日: 2022年12月23日)
- [3] 国土交通省官公庁: 宿泊旅行統計調査 https://www.mlit.go.jp/kankocho/siryou/toukei/shuku_hakutoukei.html, (最終アクセス年月日: 2022年12月20日)
- [4] 株式会社 Agoop: 流動人口データ <https://www.agoop.co.jp/service/dynamic-population-data/> (最終アクセス年月日: 2022年12月20日)
- [5] Open Street Map Japan: Open Street Map <https://openstreetmap.jp> (最終アクセス年月日: 2022年12月20日)
- [6] Aki-Hiro Sato, A Comprehensive Analysis of Time Series Segmentation on Japanese Stock Price, Procedia Computer Science, Volume 24 (2013) pp. 307-314, <https://www.sciencedirect.com/science/article/pii/S1877050913011976> (最終アクセス年月日: 2022年12月20日)