

コーホート分析による JP ドメイン名登録数の推移予測

Predicting trends in JP domain name registrations through cohort analysis

宇井 隆晴/Takaharu UI¹・池原 翔太/Shota IKEHARA²・森 健太郎/Kentaro MORI²

尾崎 剛/Takeshi OZAKI¹・広瀬 啓雄/Hiroo HIROSE¹

¹ 公立諏訪東京理科大学 ・ ² (株) 日本レジストリサービス

[Abstract]

Domain name management and DNS operation are crucial elements in the infrastructure of Internet operations. Long-term management planning, such as strategies to improve the maintenance system is thus vitally important. The prediction of trends in the number of domain name registrations is an essential issue in terms of business planning since the revenues of domain name registries come from the registration fees of new domain names and the renewal fees of existing ones. This study aims to address this issue by adopting cohort analysis with machine learning in the SARIMA model and to build a model to estimate the number of new domain name registrations. The SARIMA model adds changing factors in various seasons to the elements of autoregression (AR), integration (I), and moving average (MA) for the provision of optimal data analysis for data with seasonal fluctuations. The results reveal that the cohort analysis in the model helps with analyzing and estimating registration trends effectively based on the periodicity of the domain name registration and renewal process, as well as the features of the variance and convergence over time of the renewal rate in the initial status for registration groups.

[キーワード]

ドメイン名、登録数予測、機械学習、コーホート分析、SARIMA モデル

1. はじめに

インターネットの基盤を構成する重要要素として、その黎明期から現在に至るまで、ドメイン名と IP アドレス、そしてこれらに関連付ける DNS (ドメインネームシステム) が存在する[1][2][3]。インターネット上でのコミュニケーションは、これらの要素に大きく依存している。ドメイン名と IP アドレスはインターネットのアドレス割り当てを行う国際組織である ICANN (Internet Corporation for Assigned Names and Numbers) を頂点とする階層構造によって管理されており、インターネットの運営における重要な役割を担っている[4][5]。

有限のインターネット資源とも言われる IP アドレスの分配においては公平性が求められる[6]。このため、大陸レベルや国・地域レベルで IP アドレスを管理する組織である「IP アドレスレジストリ」は非営利機関が担っていることが多い。日本では JPNIC (一般社団法人日本ネットワークインフォメーションセンター) がその役割を担っている[7]。

一方で、ドメイン名を管理する組織である「ドメイン名レジストリ」については、ドメイン名の商業価値の高まりとともに、民間のサービスという形も取り入れながら競争環境の中で発展してきた。日本では 2001 年以降 JPRS (株式会社日本レジストリサービス) がこの役割を担っている。ドメイン名レジストリは、サービスの公平・中立な提供はもちろん、利便性の追求も求められるが、何よりも最も重要なのは、サービスを安定的かつ継続的に提供することである。

DNS は IP アドレスを用いて通信を行うインターネットに、ドメイン名という名前空間に関連付けるもので、これが機能しなくなるとドメイン名によるあらゆるアクセスができなくなる。ドメイン名レジストリが運用する TLD (トップレベルドメイン) の DNS が障害により停止すると、その TLD に属するドメイン名を用いてのアクセスができなくなり、インターネット社会に大きな影響を与える。Web やメールだけでなく、多くの社会・経済基盤がインターネット上で実現されている現代において、この影響は計り知れない。

また、継続性の観点からも、ドメイン名レジストリが経営的に破綻し、サービスが停止・終了するような事態が発生すれば、DNS の障害以上に大きな社会的影響を及ぼすことになる。このため、ドメイン名レジストリにとっては、インターネット社会の成長の中で DNS を安定的に提供するための将来予測と、システムへの長期的な設備投資計画を持つことが経営上の重要課題となる。

ドメイン名レジストリのサービスは、ドメイン名の登録料や登録更新料からの収益によって成り立っている。一般にドメイン名はドメイン名レジストリに登録する際に登録料を支払い、1 年ごとの登録更新の際に登録更新料を支払う、という収益モデルとなっている。経営としては、ドメイン名の登録数がどのように変化していくのかを長期的に見通すことが重要である。ドメイン名の登録数は、日々新規に登録される数と、それらがどれだけの期間にわたって登録継続されるかという、二つの要素から成り立っている。

これらの数は、ミクロな観点から見れば、ドメイン名レジストリがサービス対象としている地域の状況や、展開しているサービス、登録されるドメイン名の種類、登録継続期間など、多様な要素の積み重ねであると考えられる。一方で、マクロな観点からは、ドメイン名レジストリがサービス対象とする地域の社会発展やインターネット利用拡大の時間軸の中で現在の状況を反映したものであるとも考えられる。

登録数の予測手法については、各ドメイン名レジストリにおける取り組みがあると考えられるが、それらの手法が学術的な研究として公開されることは少ない。インターネットの発展が先行した国や地域の事例は、後進にとってはこれから起こる将来の事象の予測となるものでもあり、このような取り組みを研究成果として公開していくことには一定の価値があると考えられる。

2. 目的と意義

本研究の目的は、ドメイン名の登録数の長期的な予測モデルを構築することである。目的を達成するため、人口予測等で用いられるコーホート分析の手法を応用することを試みた。具体的には、筆者らが持つ JP ドメイン名のデータの中で、まず CO. JP ドメイン名のデータを用い、時間系列データの分析と予測に使用される SARIMA モデル (季節性自己回帰和分移動平均モデル, Seasonal AutoRegressive Integrated Moving Average Model) を用いて分析した。次に、この分析手法の有用性を検証するために、汎用 JP ドメイン名のデータを用いて同様の分析を行い、それらの結果について考察した。この手法は、JP ドメイン名特有の要素に依存していないため、他の TLD にも適用可能であると考えられる。

本研究の意義は、ドメイン名の登録数の予測を通じて、情報社会の発展におけるインフラの安定性と継続性を確保し、インターネットを利用する様々なサービスやビジネスの基盤を支えることにある。また、ドメイン名の登録動向を予測することで、インターネット社会の未来を見据えた計画立案や政策策定に寄与することが期待される。

3. JP ドメイン名とその種類

ドメイン名は TLD の種類によって大きく二つに分けられる。国や地域に割り当てられた ccTLD (Country Code TLD) と、それ以外の gTLD (Generic TLD) である。ccTLD は 2 文字の文字列で構成される TLD で、その文字列は ISO 3166-1 に原則として従っている。gTLD は 3 文字以上の文字列で構成される TLD で、.com や .net がよく知られているが、2024 年現在 1,000 種以上の gTLD が存在する [8]。

本研究で対象とする JP ドメイン名は、日本に割り当てられた ccTLD である。JP ドメイン名は、1986 年に日本のインターネット黎明期を先導した村井純に委任された後、1989 年に今の属性型 JP ドメイン名の形で運用が始まった。当初はインターネットの利用も一部の研究機関に留まり、登録されるドメイン名の数も限られていたため、研究者によるボランティア的な登録管理に委ねられていた。1990 年代に入ると、インターネット利用とドメイン名登録数の拡大により、大学や ISP などのネットワーク運用組織を会員とする組織 JNIC が 1991 年に設立され、1997 年には日本法の下での公益法人の一つの形である社団法人 JPNIC となった。その後、インターネットの商業化が急速に進展する中で、ドメイン名の登録管理をスケーラブルで機動的かつ安定的・継続的なサービスとして提供することを目的に、JP ドメイン名登録管理業務を JPNIC から、新たに設立した株式会社である JPRS へと移管することとなった。このような経緯で JP ドメイン名は 2002 年以降 JPRS によって運営されている。

JP ドメイン名は、1989 年からサービスが続いている属性型 JP ドメイン名 (属性型 JP ドメイン名にはその後いくつかの属性が追加されている) に、2001 年より新たに登録受付が始まった汎用 JP ドメイン名のサービスが加わった。また、2012 年には都道府県型 JP ドメイン名を追加している。JP ドメイン名は、2024 年 6 月 1 日現在

表-1 JP ドメイン名の登録数 (2024年6月1日現在)

種類	登録数
汎用 JP ドメイン名	1, 197, 693
(内: ASCII)	(1, 115, 248)
(内: 日本語)	(82, 445)
属性型 JP ドメイン名	555, 677
都道府県型 JP ドメイン名	10, 320
(内: ASCII)	(9, 012)
(内: 日本語)	(1, 308)
合計	1, 763, 690

表-2 属性型 JP ドメイン名の登録数内訳

種類	登録対象	登録数
△△△.co.jp	企業	481, 649
△△△.or.jp	企業以外の法人組織	40, 734
△△△.ne.jp	ネットワークサービス	12, 737
△△△.gr.jp	任意団体	5, 284
△△△.ed.jp	小中高校など初等中等教育機関	6, 396
△△△.ac.jp	大学など高等教育機関	3, 844
△△△.lg.jp	地方公共団体	1, 915
△△△.go.jp	政府機関	820
△△△.ad.jp	JPNIC 会員	253
地域型	地方公共団体、個人など	2, 045

で約 176 万件が登録されている[9]。表-1 は JP ドメイン名の登録数を示したもので、表-2 は属性型 JP ドメイン名の内訳を示したものである。JP ドメイン名全体の登録数のうち汎用 JP ドメイン名が約 7 割となっている。また、属性型 JP ドメイン名の中では co.jp が 9 割以上となっている。

属性型 JP ドメイン名は、TLD の直下に 2 文字の属性ラベルがあり、その下の第 3 階層に文字列を登録する形式 (example.co.jp) となっている。「△△△.co.jp」なら企業、「△△△.ac.jp」なら大学、「△△△.go.jp」なら政府機関というように、属性の種別ごとに登録資格が定められている。組織・団体による登録が前提で、個人による登録はできない。また、1 登録者あたり 1 ドメイン名のみ登録可能というドメイン名登録数の制限がある。

汎用 JP ドメイン名は、TLD の直下の第 2 階層に文字列を登録する形式 (example.jp) で、登録者は国内にサービスを受けるための連絡先住所を持つことが要件となっているが、登録者が組織か個人かは問われない。さらに、1 登録者あたりのドメイン名登録数に制限がない。

ドメイン名に用いることができる文字は初期においては ASCII 文字 (英数字とハイフン) のみであったが、その後国際化ドメイン名が標準化され[10]、JP ドメイン名においても汎用 JP ドメイン名と都道府県型 JP ドメイン名において日本語の文字 (漢字、ひらがな、カタカナ等) が利用可能となっている[11]。

登録者がドメイン名を登録するためには、ドメイン名レジストリに対して、希望するドメイン名の文字列と登録者自身に関する情報 (名義、連絡先などの情報) を登録する必要がある。また、登録にあたっては登録料が必要である。加えて、これらの情報登録と料金の支払いは、レジストラと呼ばれるドメイン名登録手続きを取り次ぐ事業者 (ホスティング事業者や ISP 事業者であることが多い) を介して行う必要がある。

4. C0. JP ドメイン名の特徴

本研究においては、まず C0. JP ドメイン名を登録数予測の対象とすることとした。C0. JP ドメイン名は、先に述べた通り属性型 JP ドメイン名の種別の一つであり、属性型 JP ドメイン名の中で最も登録数が多い。そして C0. JP ドメイン名の登録資格は日本国内で登記された企業等であることであり、ドメイン名の登録者の属性と、それらによって登録されたドメイン名の用途が同じような属性であることが期待できる。これらより、本研究の目的であるドメイン名登録数の予測モデルを構築する最初の対象として C0. JP ドメイン名が適切であると考えた。

ドメイン名の登録にあたっては属性ごとの登録資格を有する組織であることが必要だが、属性型 JP ドメイン名には仮登録という制度があり、組織設立の 6 か月前からドメイン名を登録することができることも特徴である (仮登録から 6 か月以内に登録資格を有する組織を設立して本登録の手続きを行うことが必要)。

また、C0. JP ドメイン名を含むすべての JP ドメイン名は、登録から 1 年後の月末までに廃止の手続きが取られなければ登録が 1 年間更新される。ある年のある月に登録されたドメイン名が 1 年後にどれだけ更新されたか、また更新されたドメイン名がさらに 1 年後にどれだけ更新されたか、という割合を更新率という

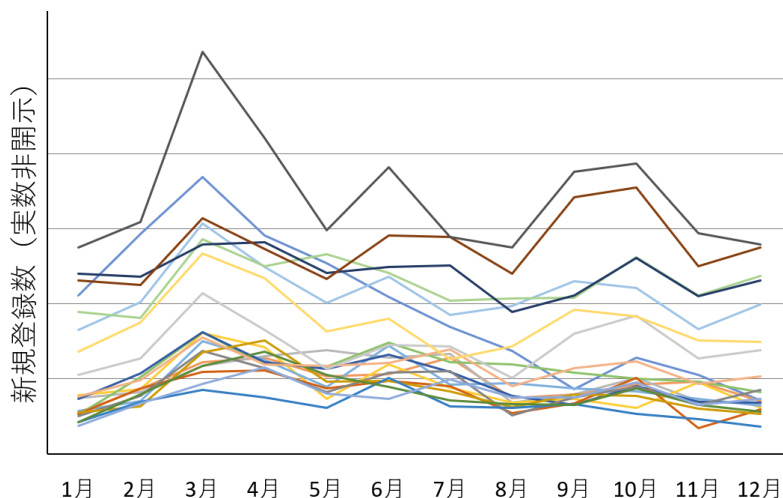


図-1 .co.jp ドメイン名の新規登録数年次推移

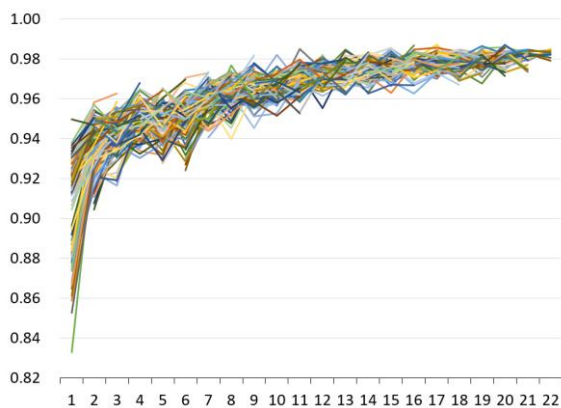


図-2 .co.jp ドメイン名の更新回数ごとの更新率推移 (年次は非開示)

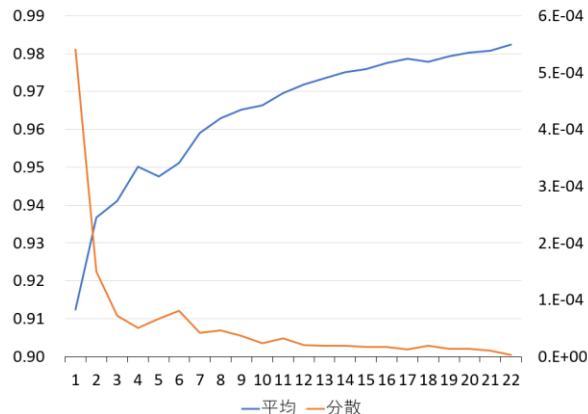


図-3 .co.jp ドメイン名の更新回数ごとの更新率の平均と分散

.co.jp ドメイン名の新規登録数は図-1 に示すように、1 年ごとで見ると絶対値は違うものの月ごとの登録数の大小は同じ変動傾向を示す (12 か月を周期とする季節性変動がある) ことがわかっている。そして、.co.jp ドメイン名の更新率は図-2 および図-3 に示すように、1 回目の更新では年次によって大きな差があるが、2 回目以降は徐々に収束していくことがわかっている。仮登録から本登録となったものは1 回目の更新率は新規登録 (最初から本登録のもの) と大きな差があるが、2 回目以降の更新率はその差がなくなることがわかっている。これらの統計上の特徴に適したモデルを構築する必要がある。

5. コーホート分析の設計

5.1 SARIMA モデルの採用

統計学やデータ分析の手法の一つであるコーホート分析 [12] [13] は、特定の期間内に共通の特徴や経験を持つ集団 (コーホート) を追跡し、その行動パターンやパフォーマンスを時間経過とともに分析する手法である。コーホートごとの時間経過による比較、将来の予測を行うために用いられる。本研究では、.co.jp ドメイン名における 1 か月ごとの新規登録数と仮登録数をコーホートとして定義し、これらの時間経過による変化を分析、予測する。

コーホートの時系列分析には先に述べた .co.jp ドメイン名の新規登録の年間月別変動と更新率推移の特徴から、SARIMA (Seasonal AutoRegressive Integrated Moving Average) モデル [14] [15] を用いるのが適していると考え

た。SARIMA モデルは、時系列データの分析と予測に使用される統計モデルであり、COVID-19 の感染を予測した研究にも用いられている[16][17]。このモデルは、自己回帰 (AR)、和分 (I)、移動平均 (MA) の三つの要素に季節性の要素を組み合わせることで、時系列データの複雑なパターンを捉えることができる。具体的には、季節性要素 (例えば 12 か月周期のデータ) を考慮することで、季節ごとの変動をモデルに組み込む。先に図-1 で示したように CO. JP ドメイン名の新規登録には 12 か月周期の季節性変動が見られるため、これを捉えることができる SARIMA モデルを用いることでドメイン名登録数をより正確に予測することが可能になる。AR (AutoRegressive) モデルと MA (Moving Average) モデルは、それぞれ以下のように表される。

$$y_t = c + \sum_{i=1}^p \phi_i y_{t-i} + \varepsilon_t, \varepsilon_t \sim N(0, \sigma_\varepsilon^2), \tag{1}$$

$$y_t = c + \sum_{j=1}^q \theta_j \varepsilon_{t-j} + \varepsilon_t, \varepsilon_t \sim N(0, \sigma_\varepsilon^2), \tag{2}$$

ここで c は定数、 ϕ, θ は回帰係数、 ε_t は t 時点での誤差項である[18], [19]。AR モデルと MA モデルを組み合わせた ARMA モデルは定常な時系列のモデルであり、非定常な時系列は定常な時系列に変換した後に ARMA モデルに当てはめる必要がある。 d 階差分をとった系列に対して $ARMA(p, q)$ を考えるモデルが ARIMA (AutoRegressive Integrated Moving Average) モデルであり、 $ARIMA(p, d, q)$ は以下のように表される。

$$y_t - y_{t-d} = c + \varepsilon_t + \sum_{i=1}^p \phi_i y_{t-i} + \sum_{i=1}^q \theta_i \varepsilon_{t-i} \dots, \tag{3}$$

SARIMA モデルは、時系列方向の説明に $ARIMA(q, d, q)$ モデルを使うだけでなく、周期方向の説明にも $ARIMA(P, D, Q)$ モデルを用いる。これに周期 s を加えて、 $SARIMA(p, d, q)(P, D, Q)[s]$ と表記する。

ここで、 p は非季節性 AR 成分の次数、 d は非季節性差分の階数、 q は非季節性 MA 成分の次数、 P は季節性 AR 成分の次数、 D は季節性差分の階数、 Q は季節性 MA 成分の次数、 s は季節周期を意味する。

5. 2 CO. JP ドメイン名の新規登録数・仮登録数

2015 年 1 月 ($n=1$) から 2023 年 9 月 ($n=105$) までの 1 か月ごとの新規登録数と仮登録数のデータを元データとして、SARIMA モデルを用いて 2033 年 9 月 ($n=225$) まで予測する。ここで用いる値を表-3 に示す。

表-3 新規登録数の予測に用いる値

n	2015 年 1 月=1 とした経過月数
$sinki[n]$	n か月目の新規登録数
$kari[n]$	n か月目の仮登録数

5. 3 CO. JP ドメイン名の 1 回目の更新率

前述の通り、CO. JP ドメイン名の更新率は、1 回目の更新では登録月別に分散が大きく、新規登録と仮登録での差も大きいので、2 回目以降の更新とは分けて予測することが必要である。

2015 年 1 月 ($n=1$) から 2021 年 12 月 ($n=84$) までの 1 か月ごとの新規登録数と仮登録数の更新率データを元データとして、SARIMA モデルを用いて 2033 年 9 月 ($n=225$) までを予測する。ここで用いる値を表-4 に示す。

表-4 1 回目の更新率の予測に用いる値

$k_sinki[n]$	$sinki[n]$ の 1 回目の更新率
$k_kari[n]$	$kari[n]$ の 1 回目の更新率

5. 4 C0. JP ドメイン名の 2 回目以降の更新率

2 回目以降の更新率は新規登録と仮登録の差がなくなるとともに分散も小さくなるため、直近 3 年間（2020 年 1 月から 2022 年 12 月）の更新回数別更新率の実績値の平均値を用いる。ここで用いる値を表-5 に示す。

表-5 2 回目以降の更新率の予測に用いる値

year[z]	z 回目の更新率 (z : 2 以上)
---------	---------------------

5. 5 C0. JP ドメイン名全体の登録数推移

ここまでの値を用いて、n か月目に登録された C0. JP ドメイン名が m か月後にどれだけ登録され続けているかを求めることができ、これを積算することで C0. JP ドメイン名全体の登録数推移を求めることができる。ここで用いる値を表-6 に示す。

表-6 登録数推移の予測に用いる値

$sinki_rem[n][m]$	$sinki[n]$ の m か月後の継続登録数
$kari_rem[n][m]$	$kari[n]$ の m か月後の継続登録数
total2014	2014 年 12 月時点での C0. JP ドメイン名の登録数
total[n]	n か月目の C0. JP ドメイン名の登録数

ドメイン名は 1 年ごとの更新のタイミングで更新率に応じて減少するものとする、 $sinki_rem[n][m]$ は以下の式で求められる（ $kari_rem[n][m]$ も同様）。

$$\begin{aligned}
 &1 \text{ 年目 } (0 \leq m \leq 11) \\
 &\quad sinki_rem[n][m] = sinki[n] \\
 &2 \text{ 年目 } (12 \leq m \leq 23) \\
 &\quad sinki_rem[n][m] = sinki[n] \times k_sinki[n] \\
 &3 \text{ 年目 } (24 \leq m \leq 35) \\
 &\quad sinki_rem[n][m] = sinki[n] \times k_sinki[n] \times year[2] \\
 &x \text{ 年目 } ((x-1) \times 12 \leq m \leq x \times 12 - 1, x \geq 3) \\
 &\quad sinki_rem[n][m] = sinki[n] \times k_sinki[n] \times \prod_{i=2}^{x-1} year[i]
 \end{aligned}$$

これらを用いて、total[n] は以下の式となる。
 （total2014 の更新率には一律 year[13] を適用する）

$$\begin{aligned}
 total[n] = & total2014 \times year[13]^{\lfloor \frac{n-1}{12} \rfloor} + \\
 & \sum_{i=1}^n sinki_rem[i][n-i] + \sum_{i=1}^n kari_rem[i][n-i], \quad (4)
 \end{aligned}$$

6. C0. JP ドメイン名を対象とした予測実験

6. 1 C0. JP ドメイン名の新規登録数・仮登録数の予測

SARIMA モデルで予測するにあたり、7 つのパラメーター (p, d, q, P, D, Q, s) を決定する必要がある。これには考えられる中で、最も AIC（赤池情報量基準）[20] が小さい組み合わせを採用する。(p, d, q) の取り得る範囲を 0 ~ 3、(P, D, Q) を 0 ~ 1、s を [6, 12, 36] として、これらのすべての組み合わせで最少となる AIC を求めた結果、新規登録・仮登録それぞれに以下のモデルを採用した。

表-7 一括予測と反復予測との比較実験結果

	a : 一括予測	b : 反復予測
訓練データ	2015/01-2022/09	2015/01-2021/09
予測期間	2022/10-2023/09	2021/10-2023/09
検証期間	2022/10-2023/09	2022/10-2023/09
MAE	141.333	126.583

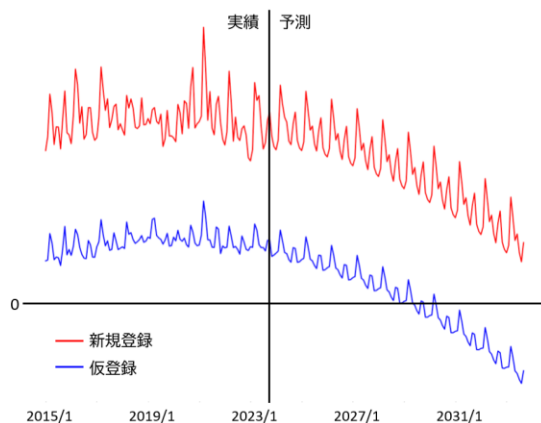


図-4 CO. JP ドメイン名の新規登録・仮登録の予測結果 (縦軸実数非開示)

新規登録 : SARIMA(2, 2, 2)(1, 1, 1)[12]

仮登録 : SARIMA(2, 2, 2)(1, 1, 1)[12]

また、予測にあたり、以下二つの手法の比較を試みた。

- a. 実績値での学習をもとに期間全体を一括予測する
- b. 実績値での学習をもとに次の1年を予測し、次はその1年分の予測結果を含めて再度学習し、さらに次の1年を予測し、これを反復し期間全体を予測する

表-7がその結果である。aの一括予測とbの反復予測でMAE(平均絶対誤差)を比較した結果、bの反復予測の方が小さいため、本件により適した手法であるとして採用した。

以上を踏まえ、SARIMAモデルで長期予測した結果が図-4である。周期性をよく捉えているが、新規登録・仮登録ともに大きく右肩下がりの傾向を示している。この現象は、学習に用いた実績値が新型コロナ禍における急速なインターネット利用拡大とその後の反動による影響を受けたことなどが原因と考えられる。特に仮登録の予測数値がマイナスの値にまで下がっているという結果は、モデルの予測限界を示しており、現実にはあり得ない。したがって、この結果は短期的なトレンドが長期予測に過度に影響を与えたものであると考えられる。

予測にあたって、短期的なトレンドと長期的なトレンドのバランスをどう捉えるか、というのは機械学習による予測に加えて、ある振れ幅の中で人間が意思を持って判断する領域であるとも考えられる。この振れ幅を実現する調整係数を λ ($0 \leq \lambda \leq 1$) として以下のように導入する。

$$\hat{x}_{t_i} = x_{t_i} + \frac{x_{t_i} - x_{t_0}}{t_i - t_0} \lambda x_{t_i} \tag{5}$$

- t_i i時点の日付
- λ 調整係数
- x_{t_i} t_i 時点の登録数
- x_{t_0} 予測開始時点 (t_0) の登録数
- \hat{x}_{t_i} t_i 時点の調整後の登録数

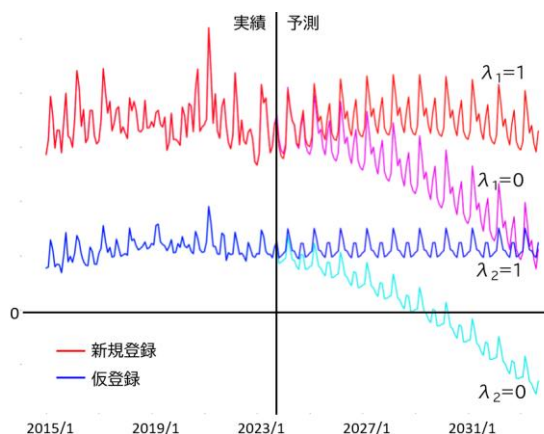


図-5 調整係数λによる振幅 (縦軸実数非開示)

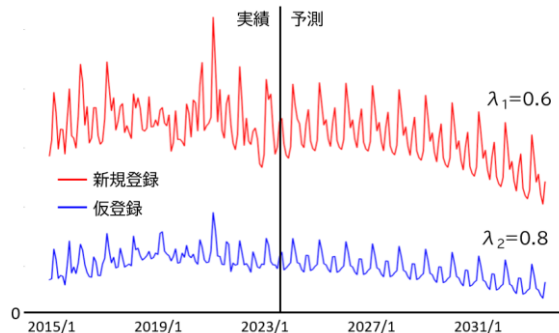


図-6 λにより調整した
新規登録・仮登録の予測結果
(縦軸実数非開示)

新規登録の調整係数を λ_1 、仮登録の調整係数を λ_2 とし、それぞれが0の時と1の時の二つのパターンをグラフにしたものが図-5である。この範囲の中で、過去の実績値の変動をどう考え、今後取り得る施策や社会・経済状況などを考慮して λ を決定することが必要であるが、 λ の決定方法については本稿では論じない。今回の実験では $\lambda_1 = 0.6$ 、 $\lambda_2 = 0.8$ とし、これを用いて予測した結果が図-6である。

6.2 00.JP ドメイン名の1回目の更新率の予測

1回目の更新率について、SARIMAモデルで(p, d, q)の取り得る範囲を0~3、(P, D, Q)を0~1、sを[6, 12, 36]として、これらのすべての組み合わせで最少となるAICを求めた結果、以下のモデルを採用し、新規登録数の予測と同様に反復予測した。図-7がその結果である。

新規登録 : SARIMA(2, 1, 1)(0, 1, 1)[12]
 仮登録 : SARIMA(1, 2, 2)(0, 1, 1)[12]

新規登録数の予測と同様に、短期的なトレンドに長期予測が大きく影響を受けていることがわかる。仮登録の更新率は1.0を超えており、これは実際にはあり得ない。1回目の更新率は分散が大きい、長期的なトレンドとして数値が大きく動くことは経験則から考えにくい。これについても、調整係数 μ ($0 \leq \mu \leq 1$) を導入することで予測に振幅を持たせることとし、新規登録の調整係数を μ_1 、仮登録の調整係数を μ_2 とした。今回の実験では $\mu_1 = 0.6$ 、 $\mu_2 = 0.8$ とし、これを用いて予測した結果が図-8である。

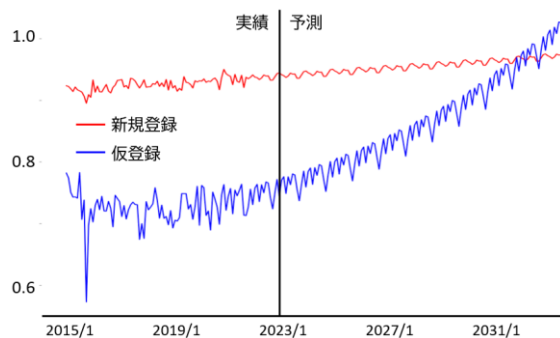


図-7 1回目の更新率の予測結果

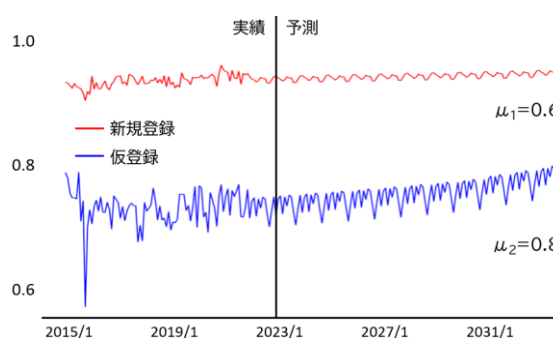


図-8 μ により調整した
1回目の更新率の予測結果

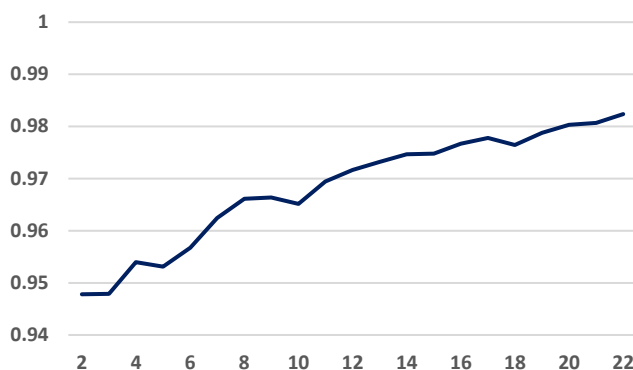


図-9 2回目以降の更新率の予測値

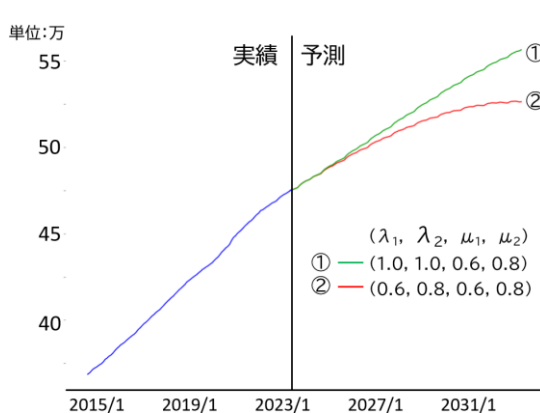


図-10 CO.JP ドメイン名全体の登録数推移の予測結果

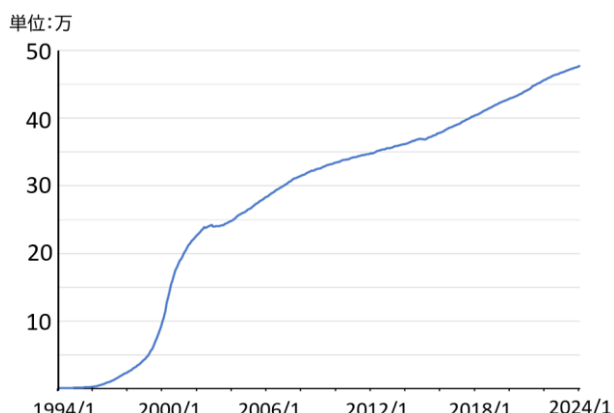


図-11 CO.JP ドメイン名の登録数推移 (実績)

6.3 CO.JP ドメイン名の2回目以降の更新率の予測

4.4節で論じたように、2回目以降の更新率は直近3年間（2020年1月から2022年12月）の更新回数別更新率の実績値の平均値を用いる。図-9は2回目以降の更新率の予測値である

6.4 CO.JP ドメイン名全体の登録数推移の予測と評価

これまでの個別の予測値をもとに、式(4)を用いてCO.JPドメイン名全体の登録数推移の予測を計算したものが図-10である。図-11は、CO.JPドメイン名の登録数の推移について、1994年1月から2024年1月までの30年分を示したものである。これからわかるように、登録数の推移は、指数関数的な部分、対数関数的な部分、直線的な部分と、一様ではない。これは、インターネットの発展や社会情勢、サービスの变化などの要因によってその登録数の傾向が変化するためであると考えられる。

図-10の予測結果は2015年からの実績データをもとにしているが、図-11で同じ期間を見ると2015年から2021年にかけては直線的に推移してきたものが2021年から2022年にかけては若干の傾向の変化（1年間程度の傾きの増加）を見て取ることができる。時期的には新型コロナ禍でインターネット活用が急速に進んだ時期と重なっており、これの影響を受けたものと考えられる。今回の実験結果はこの変化を捉えたものと考えことができ、この短期的な変化がその後どの程度継続して影響を及ぼすかを推し量るものが調整係数 λ, μ であると言える。図-10の予測の①は2021年からの変化は短期的に収束し長期的な傾向が維持されるとしたものの、②は一旦大きくなった傾きがその後衰える傾向が長期的に継続されるとしたもので、調整係数の範囲はこれまでの経験から設定している。

以上の結果より、ドメイン名の1か月ごとの新規登録数をコーホートとして定義し、SARIMA モデルを用いて新規登録数の月ごとの変動と更新率を予測、調整係数により長期的トレンドと短期的トレンドの影響を考慮した上でこれを積算することで全体の登録数の推移を予測する今回の手法は、CO. JP ドメイン名に対しては有効であると言える。

7. 汎用 JP ドメイン名を対象とした予測実験

前節までで構築した CO. JP ドメイン名の登録数の推移を予測する手法が他の種類のドメイン名においても有効であることを確認するため、ASCII 文字による汎用 JP ドメイン名（以下、単に「汎用 JP ドメイン名」とする）を対象として同様の手法で予測した。

7.1 汎用 JP ドメイン名の特徴とコーホート分析の設計

汎用 JP ドメイン名が CO. JP ドメイン名などの属性型 JP ドメイン名とサービス仕様上大きく異なる点は、2 節で述べた通りドメイン名の登録者に属性型 JP ドメイン名のような登録資格が設けられていないことと、1 登録者あたりのドメイン名登録数に制限がないことである。これにより、CO. JP ドメイン名の場合には「企業が自社の名前のドメイン名を登録して Web の URL やメールアドレスとして利用する」というケースが多数を占めるのに対して、汎用 JP ドメイン名は多様な登録者が多様な用途のためにドメイン名を登録している。その結果として汎用 JP ドメイン名は JP ドメイン名の中で最も登録数が多くなっている。なお、CO. JP ドメイン名で設けられていた仮登録制度は汎用 JP ドメイン名には存在せず、通常の新規登録のみとなっている。

CO. JP ドメイン名では、月ごとの新規登録と仮登録をそれぞれコーホートとして捉えて分析したが、汎用 JP ドメイン名ではこのような違いと特徴を踏まえ、コーホート設計も特定の登録者属性に基づかず、ドメイン名の登録行動や文字列のパターンをもとに分類することとした。

今回の予測実験では、まず新規登録された汎用 JP ドメイン名を、そのドメイン名の文字列や過去の登録状況、登録者の登録行動といった特徴から表-8 に示す 1 から 5 の群に分類した。2 群と 3 群はドメイン名の再登録に着目している。一度登録されたドメイン名が何らかの理由によって廃止された場合、一定期間（汎用 JP ドメイン名は 1 か月、属性型 JP ドメイン名は 6 か月）後に開放され再び登録が可能となる。2 群は解放直後に登録されたもの、3 群は解放から 1 か月以上経過した後に登録されたものである。4 群はドメイン名の文字列に着目している。ドメイン名は一般に何らかの意味を持った文字列で登録されるが、一方でランダムに生成された意味のない文字列と考えられるドメイン名も登録されることがある。5 群は登録者のドメイン名登録の行動パターンに着目している。汎用 JP ドメイン名は 1 登録者あたりのドメイン名登録数に制限がないため、同一組織・同一人物が多くドメイン名を登録していることが珍しくないが、まとまった数を一度に登録するという行動を抽出している。

この 5 つの群をさらに新規登録数や更新率の傾向が似ている A、B の 2 つのグループに分類し、これをコーホートとして扱うこととした。

表-8 汎用 JP ドメイン名のコーホート分類

コーホート	特徴による群	説明
グループ A	1	以下 2~4 のいずれにも該当しないもの
	2	登録されていたドメイン名が廃止・開放されて 1 か月以内に再度登録されたもの
グループ B	3	登録されていたドメイン名が廃止・開放されて 1 か月以上経過してから再度登録されたもの
	4	ドメイン名の文字列が意味のないランダムなもの
	5	同一の登録者が同日に多数のドメイン名を登録したもの

7.2 汎用 JP ドメイン名の新規登録数および更新率の予測

A、B の 2 つのグループをコーホートとして、CO. JP ドメイン名の予測と同様に、汎用 JP ドメイン名の新規登録数について SARIMA モデルで予測したものが図-12 である。

新規登録数の実績部分 2019 年初あたりで大きなスパイクを観測しているが、1 登録者あたりのドメイン名登録数に制限のない汎用 JP ドメイン名ではこのようなスパイクが発生することがある。この現象は特定の登録者が短期間に大量のドメイン名を登録した際や、ドメイン名の登録を促進するキャンペーンのような施策が実施された際に観測される。このようなスパイクは、通常ドメイン名登録のパターンからは外れており、一時的なものに過ぎない。このため、時間軸的に前後の登録には影響を与えず、周期的に表れるものでもない。このような行動パターンも除外することなく一つのコーホートとして予測モデルに含めているが、事象としての特異性を SARIMA モデルがよく捉え、長期的な予測をしていることが読み取れる。

また、CO. JP ドメイン名の新規登録数予測と同様に、図-12 でも短期的なトレンドの影響を受けて両グループともに大きく減少していく予測となっているが、長期的トレンドの中でそこまで大きな影響を受けることはないとの考え方より、調整係数 λ を導入して調整したものが図-13 である。更新率についても CO. JP ドメイン名と同様に予測したものが図-14、調整係数 μ を導入したものが図-15 である。

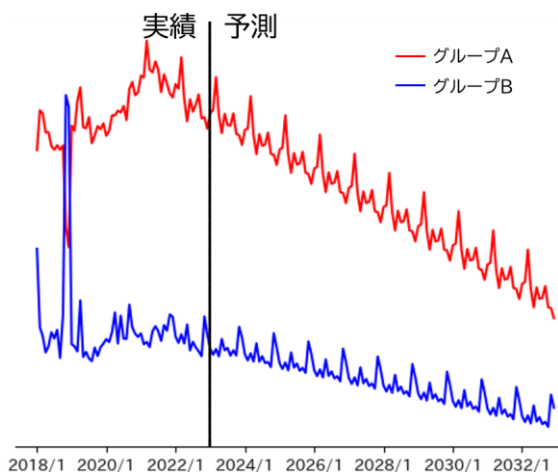


図-12 汎用 JP ドメイン名の新規登録の予測結果
(縦軸実数非開示)

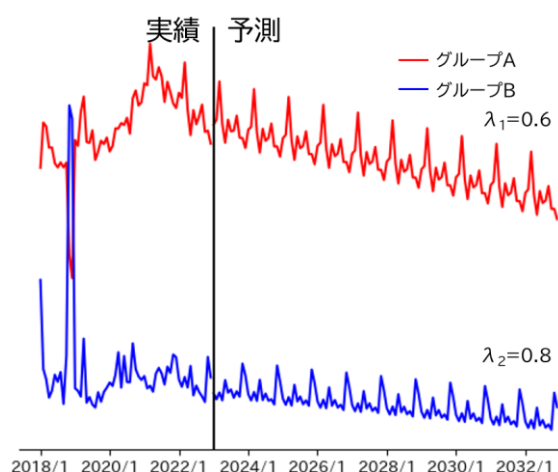


図-13 λ により調整した
汎用 JP ドメイン名の新規登録の予測結果
(縦軸実数非開示)

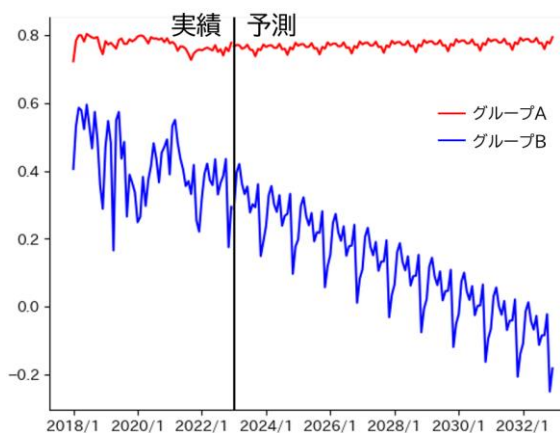


図-14 汎用 JP ドメイン名の1回目の更新率の予測結果

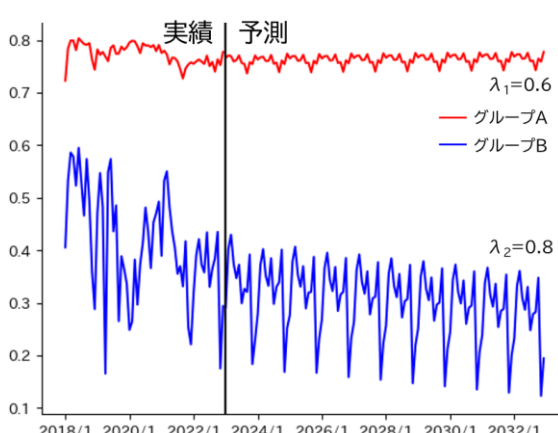


図-15 μ により調整した
汎用 JP ドメイン名の1回目の更新率の予測結果

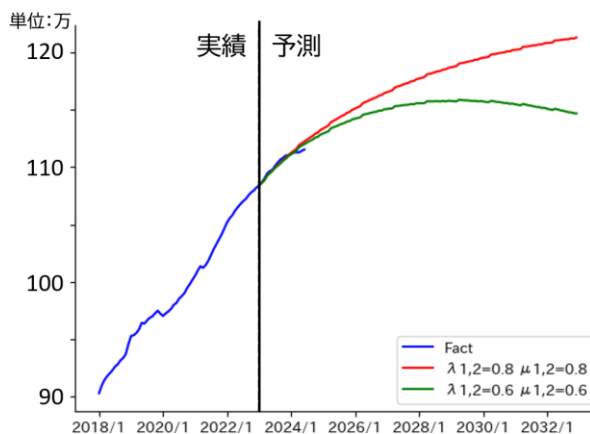


図-16 汎用 JP ドメイン名全体の登録数推移の予測結果

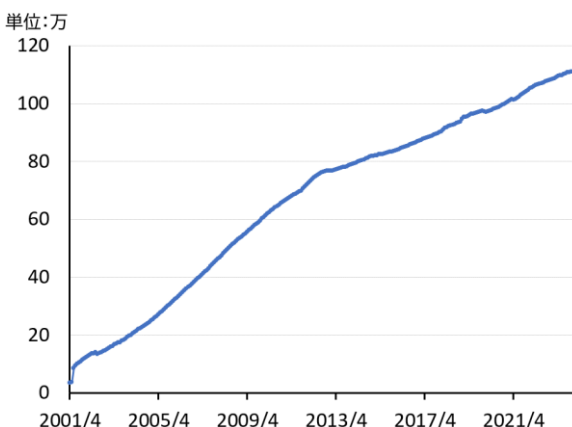


図-17 汎用 JP ドメイン名の登録数推移 (実績)

7. 3 汎用 JP ドメイン名全体の登録数推移の予測と評価

前節までの予測値をもとに、式(4)を用いて汎用 JP ドメイン名全体の登録数推移の予測を計算したものが図-16 である。図-17 は汎用 JP ドメイン名の登録数推移について、2001 年 4 月から 2024 年 1 月までの実績を示したものである。図-11 の C0. JP ドメイン名の登録数推移と比べると汎用 JP ドメイン名の方が実績部分について動きがあることがわかる。これは C0. JP ドメイン名がインターネットの発展や社会情勢など中長期的なマクロな状況を反映しているのに対して、汎用 JP ドメイン名はサービスの変化や突発的な個別事象など短期的・ミクロな状況を反映しやすいという違いが表れているものと考えられる。この違いは、ドメイン名の利用目的やターゲット層の違いにも起因していると考えられる。図-16 の予測は、そのような中長期的なトレンドに加えて今起こりつつある短期的なトレンドを反映した予測であると言える。これにより、ビジネスや政策決定に有用な情報を提供することが可能となる。

以上のように、C0. JP ドメイン名と同じ手法で汎用 JP ドメイン名全体の登録数の推移を予測し、有効な結果を得ることができた。

さらに、トレンドと季節変動を考慮した他の予測手法との比較として、Serrecchia が HOLT-Winter [21] [22] [23] 法を用いてイタリアのドメイン名の登録数推移を予測した手法 [24] を汎用 JP ドメイン名に適用した結果が図-18 である。ここでは、2001 年 5 月から 2022 年 12 月までの実績数値を学習用データとし、それ以降 2034 年 12 月までを季節周期を 12 か月として予測した。

予測モデルとして、本研究の手法と Serrecchia の手法は、中長期的なトレンドと短期的なトレンドを組み合わせ

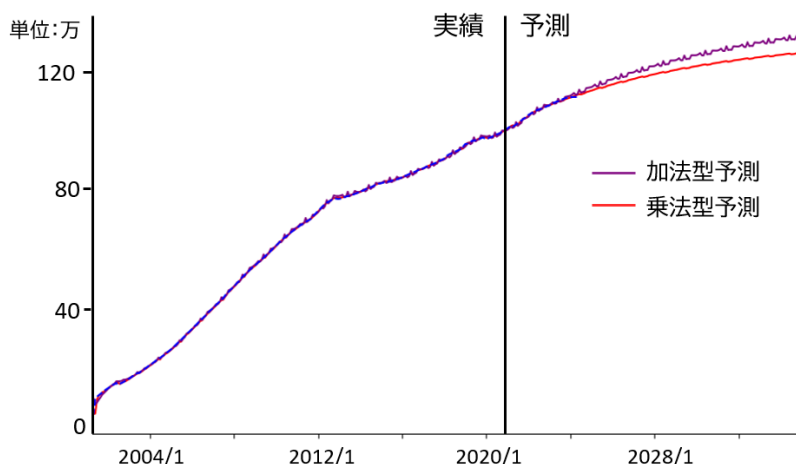


図-18 HOLT-Winter 法による汎用 JP ドメイン名の登録数推移予測

せる点では似ているが、本研究の手法では1か月ごとに登録されたドメイン名をコーホートとして、そこから新規登録数と更新率を予測し、それを積み上げることで全体の登録数推移を予測しているのに対して、Serrecchiaの手法では、ドメイン名登録数全体の推移から将来の推移を予測している点が異なっている。Serrecchiaはイタリアにおけるドメイン名の登録数は飽和の段階に達する傾向にあると述べているが、その先の減少傾向に至る状況についてまでは予測されていない。図-18の結果を見ても飽和の段階に至る傾向は見取れるが、その後の減少までは予測されていない。実際に将来登録数が飽和の段階の先に減少の傾向を見せるかどうかは時間の経過を待つしかないが、いくつかのTLDにおいてはそのような減少傾向に入っているものも現れてきている状況である[25]。本研究の優位性は登録数の推移を変化させるパラメータ（毎月の新規登録数とそれらの更新率）を予測し積み上げるところにあり、これにより全体推移からは観測できない動きまでを予測できていると言える。将来ドメイン名の登録数推移が減少傾向の局面に入る可能性も他のTLDの状況を考慮すれば予測の振れ幅として十分にあり得ることであるとと言える。

8. 結論

以上より、ドメイン名の登録数を長期的に予測する手法として、ドメイン名の新規登録と更新プロセスにおけるサービスの周期性、さらには登録集団ごとの初期更新率の分散と経年にわたる収束のパターンなど、ドメイン名の登録行動に見られる特徴をコーホート分析の手法によって捉えるとともに、直近の実績値に基づく短期トレンドと長期的なトレンドの影響を調整係数 λ 、 μ を用いて考慮し分析する本研究の手法は、ドメイン名の登録数全体の変動傾向から予測する手法に比して有効であると判断できる。

9. おわりに

本研究では、コーホート分析の手法を、特に JP ドメイン名の登録数の長期予測に応用することに焦点を当てた。

JP ドメイン名の登録数においては、CO. JP ドメイン名と汎用 JP ドメイン名が占める割合が大きい。本研究では、まず CO. JP ドメイン名を実験対象とした。CO. JP ドメイン名の登録資格は日本国内で登記された企業等であり、ドメイン名の登録者の属性と、それらによって登録されたドメイン名の用途が同じような属性であることが期待できるため、予測モデルを構築する最初の対象として適していると考えた。CO. JP ドメイン名を対象として予測モデルの構築を行いその有効性を確かめた後、検証として汎用 JP ドメイン名への適用を試みた。CO. JP ドメイン名に比べて汎用 JP ドメイン名は登録要件がより広範であり、ドメイン名の用途も多岐にわたるため、ドメイン名登録者の種類や登録の行動などの特徴を捉えたコーホートを組み立てた上で予測を行い、本研究で構築した予測モデルが汎用 JP ドメイン名においても有効であることを確認した。

さらに、本研究の結果は、ドメイン名の登録数予測がインターネット社会の未来を見据えた計画立案や政策策定にどのように寄与できるかを示している。例えば、ドメイン名の登録数の減少が予測される場合、その原因を分析し、インターネットの利用促進策や新たなサービス提供を検討することが重要である。また、ドメイン名の登録動向はビジネスの動向や社会の変化を反映するため、経済状況や技術革新の影響を理解する上でも有用である。

総じて、本研究は、ドメイン名の登録数の長期予測にコーホート分析を応用することにより、ドメイン名レジストリの経営における戦略的な意思決定に貢献する可能性がある。今後は、分析手法のさらなる精緻化と、予測モデルの拡充に向けて研究を進めていくことが望まれる。さらに、社会状況の影響を受けたものとしてドメイン名の登録数があるのであれば、ドメイン名の登録数予測は社会状況の将来を予測するものであるとも考えられる。この観点も含め、他のTLDへの適用や、先進・後進の比較といった研究の拡大が考えられる。

[参考文献]

- [1] Mockapetris, P. (1983). RFC 883: Domain names - Implementation and specification. Internet Engineering Task Force. Retrieved from <https://www.ietf.org/rfc/rfc883.txt>
- [2] Mockapetris, P. (1987). RFC 1034: Domain names - Concepts and Facilities. Internet Engineering Task Force. Retrieved from <https://www.ietf.org/rfc/rfc1034.txt>
- [3] Mockapetris, P. (1987). RFC 1035: Domain names - Implementation and specification. Internet

Engineering Task Force. Retrieved from <https://www.ietf.org/rfc/rfc1035.txt>

- [4] Mueller, M. (2002). Ruling the Root: Internet Governance and the Taming of Cyberspace. Cambridge: MIT Press.
- [5] National Research Council. (2005). Signposts in Cyberspace: The Domain Name System and Internet Navigation. Washington, DC: The National Academies Press.
- [6] K. Hubbard, M. Kosters, D. Conrad, D. Karrenberg, J. Postel (1996). RFC 2050: Internet Registry IP Allocation Guidelines. Internet Engineering Task Force. Retrieved from <https://www.ietf.org/rfc/rfc2050.txt>
- [7] Japan Network Information Center. (2024). About JPNIC. Retrieved from <https://www.nic.ad.jp/>
- [8] IANA: Root Zone Database, <https://www.iana.org/domains/root/db>, The IANA functions an affiliate of ICANN, Accessed 28 June 2024
- [9] JP ドメイン名の登録数, <https://jprs.jp/about/stats/>, JPRS, Accessed 28 June 2024
- [10] Klensin, J. (2010). RFC 5890: Internationalized Domain Names for Applications (IDNA): Definitions and Document Framework. Internet Engineering Task Force. Retrieved from <https://www.ietf.org/rfc/rfc5890.txt>
- [11] JP ドメイン名の種類|使用できる文字, <https://jprs.jp/about/jp-dom/character/>, JPRS, Accessed 28 June 2024
- [12] Xiaofeng Wang, Michael W. Kattan. Cohort Studies: Design, Analysis, and Reporting. CHEST JOURNAL. VOLUME 158, ISSUE 1, SUPPLEMENT, S72-S78, (2020)
- [13] Samantha Teague, George J. Youssef, Jacqui A. Macdonald, Emma Sciberras, Adrian Shatte, Matthew Fuller-Tyszkiewicz, Chris Greenwood, Jennifer McIntosh, Craig A. Olsson, Delyse Hutchinson & the SEED Lifecourse Sciences Theme. Retention strategies in longitudinal cohort studies: a systematic review and meta-analysis. BMC Medical Research Methodology. Vol.18, Article number: 151 (2018)
- [14] Mehdi Khashei, Mehdi Bijari & Seyed Reza Hejazi. Combining seasonal ARIMA models with computational intelligence techniques for time series forecasting. Soft Computing. Volume 16, pages 1091-1105, (2012)
- [15] Korstanje, J. The SARIMA Model. In Advanced Forecasting with Python. Apress, Berkeley, CA. pp. 115-122 (2021).
- [16] Solomiia Fedushko, Taras Ustyianovych. E-Commerce Customers Behavior Research Using Cohort Analysis: A Case Study of COVID-19. Journal of Open Innovation; Technology, market, and Complexity. Volume 8, Issue 1, (2022),
- [17] 江谷典子: 時系列解析 SARIMA モデルを用いた日本の COVID-19 感染予測実証実験, DP レポート, DPR10019 (2019):1-9
- [18] 島田直希: 時系列解析-自己回帰型モデル・状態空間モデル・異常検知-, 共立出版株式会社, 2020
- [19] Marco Peixeiro、株式会社クイープ: Python による時系列予測, マイナビ出版, 2023
- [20] 赤池弘次ほか: 赤池情報量規準 AIC—モデリング・予測・知識発見—, 共立出版, 2007
- [21] C. Chatfield. The Holt-Winters Forecasting Procedure. Journal of the Royal Statistical Society Applied Statistics Series C. Volume27, Issue3, Pages 264-279. (1978)
- [22] Sarah Gelper, Roland Fried, Christophe Croux. Robust forecasting with exponential and Holt-Winters smoothing. Journal of FORCASTING. Volume29, Issue3, Pages 285-300. (2010)
- [23] Chantha Wongoutong. The Effect on Forecasting Accuracy of the Holt-Winters Method When Using the Incorrect Model on a Non-Stationary Time Series. Thailand Statistician, 19(3), 565-582. (2021)
- [24] Michela Serrecchia. Forecasting internet diffusion in Italy based on the “.it” domain names metrics. FORESIGHT, Vol.25, No.3, pp.400-419 (2023)
- [25] gTLD の登録数, <https://www.nic.ad.jp/ja/stat/dom/gtld.html>, JPNIC, Accessed 28 June 2024

(2024年8月28日受理)