

ユーザの Web アクセス履歴のべき乗分布傾向に着目した属性推定手法の提案

市川 裕介^{†1} 小林 透^{†1}

我々は、レコメンデーションなどでの利用を目的としたサイコグラフィック属性として、ユーザの多様性受容度をユーザ個々の Web アクセス履歴のべき乗則を利用して推定する手法の提案を行った。実履歴を用いた検証の結果、ユーザ個々のサイトアクセス傾向がべき乗分布を示すこと、べき指数が観測期間によらずユーザ固有の特性値であること、べき指数がユーザの多様性受容度に相関があることから、提案手法が有効であることを示した。

Analysis of Web Access Logs to Estimate the User's Preferences

YUSUKE ICHIKAWA^{†1} and TORU KOBAYASHI^{†1}

This research aims for the establishment of the calculation method of user's psychological situation by a simple statistical work of the web access log. Each user's access tendency showed power law distribution as a result of analyzing the access log. There was a correlation between the user's behavior and the scaling exponent.

1. はじめに

次世代ネットワーク (NGN) により安心安全なネットワーク環境が提供され、クラウドコンピューティングにより安価にネットワーク上の記憶スペースや計算環境を利用できるようになると、これまで以上に多種多様大量のデータがネットワーク上に保存されるように

ることが予想される。近年、これらのデータをユーザ/サービス横断的に分析・活用することによる新たなネットワーク・サービスの創出を模索する動きが活発化している¹⁾。

一方で、情報検索や電子商取引 (EC) を中心として、すべてのユーザに画一的なサービスを提供する “one-size-fits-all” アプローチによるサービス提供から、ユーザ個々の属性に合わせてサービスを最適化させる “personalization” アプローチへと移行する動きが活発化している²⁾。以前より、マーケティング・サイエンス分野では、商品の販促活動において、売り手は個々の買い手に向けて別々のマーケティングプログラムを設計することが理想であるといわれており³⁾、個々の買い手をとらえる方法として消費者を各種属性によって細分化する方法が提唱されてきていたが、この考え方がネットワーク・サービスにも適用され始めてきている。

従来、ユーザを細分化するための属性として、性別や年齢といったデモグラフィック属性がさかんに利用されてきたが、ユーザの価値観の多様化により性別や年齢だけではユーザの実態をとらえることが難しくなっている。そこで、ユーザの活動や関心事、意見、あるいはパーソナリティといったサイコグラフィック属性を用いてユーザをより細分化したグループに分けてサービスを最適化する取り組みが増えている⁴⁾。

また、商品や情報を推薦するレコメンデーション・システムの研究においては、近年、従来の予測精度などによる評価のみでなく、「セレンディピティ」などの推薦内容の新規性、意外性がユーザ満足度において重要だとする考え方が出てきている⁵⁾。

Ziegler ら⁶⁾ は、精度を犠牲にしても、より広範囲からの推薦によりセレンディピティを上げる話題多様化を提案し、話題多様化により予測精度がそれほど低下せず、ユーザの満足度が上がると報告している。天笠ら⁷⁾ は、ユーザがシチュエーションに応じて志向型が目的志向/非目的志向に分かれることから、シチュエーションごとにニーズに合った明確な情報提供と、ある程度遊びを持たせた多様な情報の提供とを使い分けることにより、ユーザの状況にそった情報提供となる確率が高くなると報告している。

ここで、我々は話題多様化の効果は、ユーザの属性によってその満足度が上がる度合いが異なると考える。たとえば、遠藤ら⁸⁾ は、ユーザの興味特性を興味対象への興味レベルの違いから Mania, Active User, Majority の 3 つに分類し、Majority が興味対象が不明確であることから、他者の影響を受けやすく、興味を持つ対象が多様であり、逆に Mania は他の興味に対する関心が薄いことを示している。

したがって、サイコグラフィック属性の一種として、ユーザを多様な情報への興味を持ちやすい傾向か、特定の興味へ固執する傾向であるかを推定し、その度合い (ここでは「多様

^{†1} 日本電信電話株式会社 NTT サイバソリューション研究所
NTT Cyber Solutions Laboratories, NTT Corporation

性受容度」と呼ぶことにする)を指標値として示すことができれば、話題多様化を効果の高いユーザのみへ提供することが可能となり、レコメンデーションへの満足度をさらに向上させることができる可能性が高い。

そこで我々は、今後ネットワーク上に保存され、活用が見込まれる履歴データを対象として、情報検索や電子商取引における情報提供への活用を目的として、その履歴データから得られるユーザの情報閲覧行動のパターンから多様性受容度に関するサイコグラフィック属性を推定する手法の検討を行った。

本稿では、まず既存研究でのアプローチとその課題について 2 章で述べ、その課題を解決する手法について 3 章で仮説と提案を述べる。4 章、5 章でその仮説検証方法と結果、6 章で考察を述べ、最後に 7 章でまとめを記述する。

2. 既存研究

サイコグラフィック属性の推定には、主にアンケート回答を利用した分類を行う手法、ユーザの行動観察により行動の振舞いの違いから分類を行う手法がある。

アンケートによる手法では、マーケティング・サイエンスの分野において多くの取り組みがあるほか、検索結果のパーソナライズを行う目的で、検索クエリなどとともに検索目的を入力してもらい、情報行動の分類を行う取り組みなどがある^{9),10)}。アンケートによる手法は、ユーザにアンケート回答の負担がかかるという問題があり、適用ユーザ数を増やすのが難しい。また、回答の正確性にばらつきがでる傾向にあり、信頼性が低いという問題がある。

一方、ユーザ観察によりユーザの行動特性を分類する手法は、実際のユーザ行動を第三者が観察する方法であるため、ユーザへの入力負担や、信頼性の問題は解決できる。たとえば、高久ら¹¹⁾は、Web 利用時の視線の動きから情報行動のパターン分類を行っている。ユーザ観察の手法の場合、1 日単位などの小規模な分析のみが限界であり、長期的な変化の有無などユーザ固有の行動特性が否かを調べることができない問題がある。

このため、White ら¹²⁾は、ユーザの PC 端末へインストールしたソフトにより Web アクセス履歴データを収集したログからユーザの行動特性を分析することで、長期間のデータからユーザ分類する手法を提案している。White らは、Web アクセス履歴データをグラフパターンとして扱い、パターン分類によりユーザが毎回同じ検索行動を行う“navigators”タイプと、毎回異なる行動を行う“explorer”タイプに属性付けし、属性ごとに検索結果のパーソナライズを行うことを提案している。

Ichikawa ら¹³⁾は、Web アクセス履歴から、アクセス先のページタイプごとのアクセス

頻度をベクトルデータとして扱い、SVM (Support Vector Machine) などの機械学習の手法により、ユーザが新しもの好きな「革新的ユーザ」タイプか、その対極にある性質を持つ「保守的ユーザ」タイプか推定する手法を提案している。

これらの、Web アクセス履歴をパターン分類や、機械学習により属性を推定する手法は、推定処理のための計算コストが大きく、また、サイトやページヘラベル付けを行わなければならない、ラベル付けの手間(人的コスト)がかかるという問題がある。

以上のように、ネットワーク・サービスへサイコグラフィック属性を利用する目的のためには、既存手法は、ユーザへの負担、適用範囲の限界、計算コストやメンテナンス・コストの問題がある。そこで、我々はサイト/サービス横断的に収集した Web アクセス履歴を単純に統計解析して得られた特徴量に着目する。Web アクセス履歴の単純な統計解析結果からサイコグラフィック属性を推定することができれば、ユーザへの負担なくネットワーク・サービス横断的な属性推定が提供可能となる。

3. Web アクセス履歴からの属性推定手法の提案

我々は、Web アクセス履歴から多様性受容度を表す統計的な特徴量として、ユーザ個々のアクセス数分布のべき乗則に着目し、そのべき指数を用いる手法を提案する。

ユーザの行動の多様性を表す統計的な指標値としては、平均や標準偏差を用いた方法が考えられる。しかし、一般的にユーザが情報や物を選択する選好行動の結果の分布はべき乗則になることが知られている¹⁴⁾。

べき乗則とはある観測量がパラメータのべき乗に比例することを示し、ネットワーク分析、言語解析など、多くの分野でべき乗則が成り立つという解析結果が報告されている。特に、近年の研究により、人が情報や物を選択する行動(選好行動)の結果得られる履歴データの統計をとると、多くの場合べき乗則を示すことが報告されている。たとえば、井庭ら¹⁵⁾は、書籍販売の POS データを解析した結果、販売数の分布がべき乗分布になることを示し、コンテンツ商材市場の市場モデルの構築に活用している。よく知られる「ロングテール」の法則もべき乗則をベースとする経験則である。我々の以前の研究においても、ディレクトリサイトにおけるウェブサイトの選択や、EC サイトにおける商品選択の結果がべき乗則になることを示した¹⁶⁾。

べき乗則を示す場合、同一観測対象でも観測ユーザ数が増減するだけで平均・分散が大きく変化し、観測数が十分に多い場合、平均は 1 に近く、分散は ∞ に近くなり、平均や標準偏差といった正規分布を前提とした指標が役立たない。その一方で、観測数によらずそのべ

べき指数などの分布形状が維持される「スケールフリー」などの特徴を持つことが知られている¹⁷⁾。このため、行動履歴の活用においては、従来の正規分布を前提とした平均や分散による閾値設定などではなく、べき乗則の特性を生かしたアプローチをとることが重要となってくる。

ここで、べき乗則の発生原理に着目する。べき乗則の発生原理をモデル化する取り組みの中で示されているものに、Zipf の「労力最小の原理」や、小野寺が示した「成功累積モデル」などがある¹⁷⁾。これらの共通する特徴として、べき乗則の「少数への集中、多数への分散」が、労力を最小化しようとする人間の努力や、成功が成功を生む状態など、人間行為の結果よるものとしている点にある。

このモデルを Web アクセスにあてはめると、ユーザは有用と感じたサイトを集中的に利用する傾向にあり、そのべき指数が小さいほどその傾向が強まるという仮説が立てられる。つまり、Web 利用の目的がはっきりしているユーザほどその目的に合致したサイト（有用と感じたサイト）を集中的に利用する傾向が強くなるため、べき指数が小さくなり、逆に多様な目的で利用するユーザはべき指数が大きくなる傾向にあり、べき指数によってユーザの行動傾向が集中的傾向にあるか、多樣的傾向にある多様性受容度の高いユーザであるか分類できる可能性があるといえる。

べき指数をユーザの多様性受容度を表す特徴量として適用可能であるためには、すべてのユーザのサイトへのアクセス数分布がべき乗則を示し、近似関数からべき指数を求められること、および、そのべき指数が観測期間、時期によらずユーザ固有の値として変化しない必要がある。

さらに、ユーザのアクセス先の Web サイトの利用傾向が多様性のあるユーザと、そうでない固定的なユーザとでべき指数に差異が現れていること。いい換えると、ユーザの選好行動特性とべき指数の間に相関が見られることが必要となる。

以上をまとめると、提案手法が成立する要件は以下の 3 つがあげられる。

- 要件 1) ほとんどすべてのユーザのアクセス分布がべき乗則を示すこと
- 要件 2) ユーザ個々のべき指数値は観測時期によらず、時系列変化の幅はユーザ間の差異に比べて小さく、ユーザの固有値として利用できること
- 要件 3) アクセス先 Web サイトの利用傾向など、ユーザの選好行動特性とべき指数に相関が見られること

次章より、これらの要件について成立するか、検証方法とその結果について述べる。

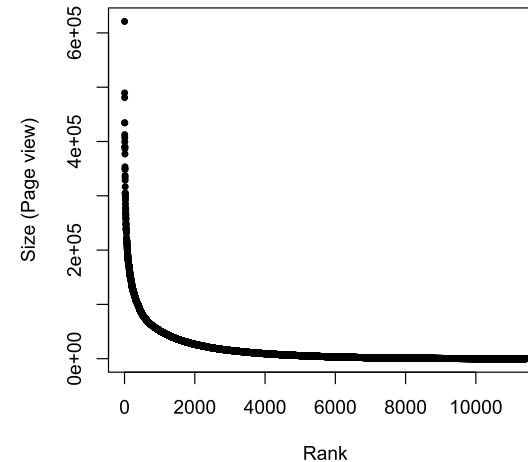


図 1 ユーザのアクセス数分布

Fig. 1 Rank-size distribution of page access.

4. 実ログを用いた仮説の検証

4.1 分析データ概要

検証には、ビデオリサーチインタラクティブ社調べの Web 視聴データ^{*1}を使用した。本データはユーザの端末 (PC) にインストールしたソフトにより、Web アクセス先の URL、アクセス日時、閲覧時間を記録したものである。データ収集の対象ユーザは、男女比、年代構成は日本全体のインターネット利用者分布と同等分布となるよう、日本全国から集められたパネルユーザで構成されている。

使用したデータの観測対象ユーザ数は月間のアクティブユーザ数が約 7,000 人となるよう調整された、のべ 11,424 人となっており、また、観測期間は 2007 年 1 月 1 日～2007 年 12 月 31 日までの 1 年間分となっている。本検証では、この中から各月の閲覧数 (ページビュー数) が 100 PV 以上あるユーザ 2,677 人に観測対象を絞って検証を行った。

図 1 は、観測対象の全ユーザについて、アクセス数 (PV) の多い順に、横軸をランク、縦軸をアクセス数でプロットしたグラフである。図 1 から、本データについても、前章で

*1 <http://www.videoi.co.jp/>

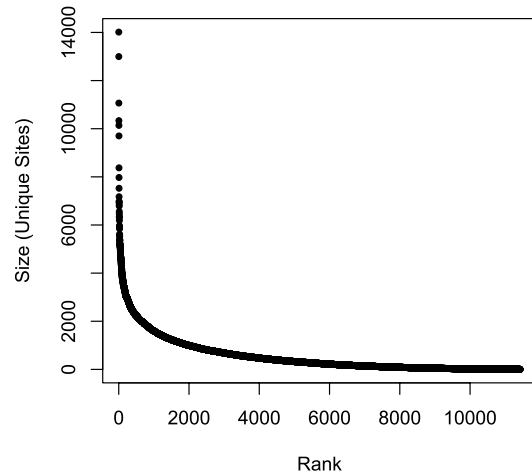


図2 ユーザのアクセスサイト数分布
Fig.2 Rank-size distribution of access sites.

述べたように、ユーザのアクセス数分布がべき乗則を示していることが分かる。同様にアクセスサイト数（ユニークサイト数）のランキングでプロットした場合についても、べき乗則を示していた（図2）。

4.2 検証方法

4.2.1 べき指数の算出手順

要件1, 2を検証するためのべき指数の算出手順は以下のとおり。

- (1) ユーザ個々についてサイトごとのアクセス数の分布（ランク-サイズ分布）を集計
- (2) ランク-サイズ分布の x, y 両軸について値を対数変換
- (3) 最小二乗法により $y = ax + b$ の近似関数を求める（ここで a がべき指数）

なお、近似関数の算出には Perl のモジュール Statistics::OLS を用いた。また、分布のプロットすべてを対象に近似関数を求めると、ランク低位の分布の影響が大きいため、対数軸上 0.1 等間隔になるようデータを間引いてから近似式を求めている。

4.2.2 サイト利用傾向の分析方法

サイトの利用傾向として、多様な目的で Web を利用するユーザは、まず検索・ポータルサイトを窓口として利用し、特定の興味領域に絞った情報収集に特化したユーザは掲示板・ブログサイトを定点観測する傾向にあるという仮説を立て、ユーザのよく利用するサイト

（アクセス数上位のサイト）が検索・ポータルサイトか、掲示板・ブログサイトかによってユーザを分類し、べき指数の比較を行うことで要件3の検証を行う。

ここで、本検証に用いた履歴データにはアクセス先の URL のみの記録となっているため、単純にアクセス数上位を抽出するだけでは、上位に検索・ポータルサイトが含まれるか、掲示板・ブログサイトが含まれるかを判別し、ユーザを分類することはできない。したがって、アクセス先の URL に対してサイト種別によるラベル付けを行う仕組みを構築する必要がある。

本検証では、URL に対するラベル付けは、goo カテゴリ検索^{*1}において、「インターネット > 検索サービス」、「インターネット > ブログ」および「インターネット > 掲示板」のそれぞれのカテゴリに登録されているサイト・リストを利用し、ドメインでサイト種別が判別できるサイト、ホスト名でサイト種別を判別する必要のあるサイトを手作業で振り分け、ラベル付けのルールベースを構築して行った。

5. 検証結果

5.1 ユーザのサイトアクセスの傾向

ユーザのサイトへのアクセス数ランキングを横軸に、アクセス数を縦軸にプロットしたグラフを、べき指数の大きいユーザ、小さいユーザ、それぞれについて、図3、図4にその例を示す（グラフは両対数軸で作成）。グラフからユーザごとのアクセス傾向についても、べき乗則を示していることが分かる。

対象ユーザすべてのべき乗近似式を求めたところ、べき指数の範囲は $-0.88 \sim -3.32$ （平均 -1.24 、標準偏差 0.17 ）、決定係数 R^2 は $0.88 \sim 1.00$ （平均 0.97 ）の間に収まっており、年齢、性別に関係なくすべてのユーザがべき乗分布を示していることが分かった。図5にべき指数の分布を示す。なお、図中の赤線は正規分布を示す。

以上の結果から、ユーザの Web アクセス履歴が要件1を満たすことが分かる。

5.2 時系列変化の傾向

ユーザごとに月ごとのべき指数を算出し、各月のべき指数値の年間（12カ月分）でのばらつきを示す標準偏差を算出したところ、その平均は 0.11 となった。図6に縦軸を頻度（ユーザ数）、横軸を標準偏差の区間にとった分布図を示す。

これに対し、各月の全ユーザのべき指数の標準偏差、すなわちユーザ間のべき指数のばら

*1 <http://dir.goo.ne.jp/>

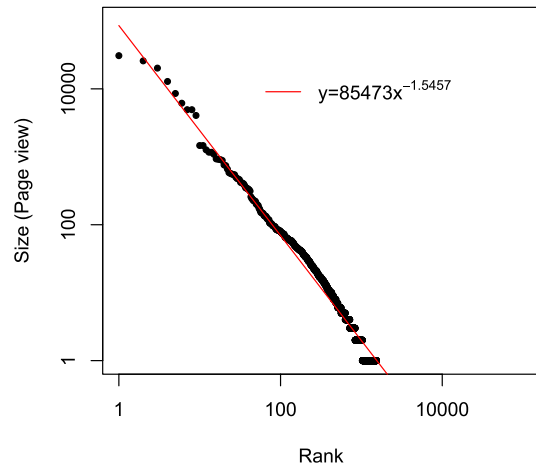


図 3 ユーザのサイトへのアクセス数分布 (べき指数の小さいユーザの例)
Fig. 3 Examples of the distribution of site access (steep slope).

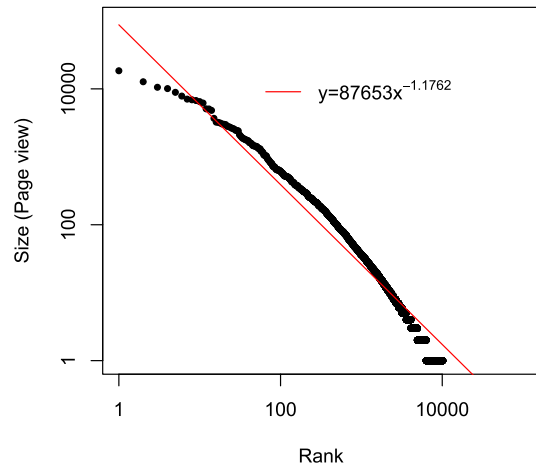


図 4 ユーザのサイトへのアクセス数分布 (べき指数の大きいユーザの例)
Fig. 4 Examples of the distribution of site access (gentle slope).

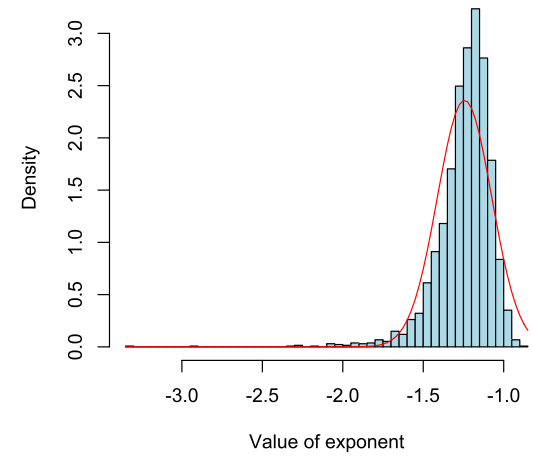


図 5 ユーザのサイトアクセス分布のべき指数の分布
Fig. 5 Exponential distribution of site accesses.

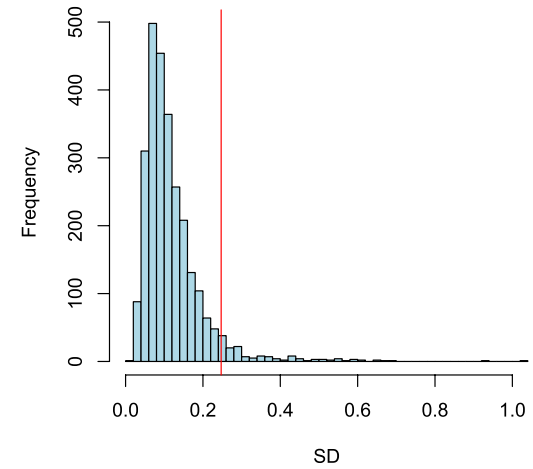


図 6 べき指数のユーザごとの変化 (標準偏差) の分布
Fig. 6 Distribution of standard deviation.

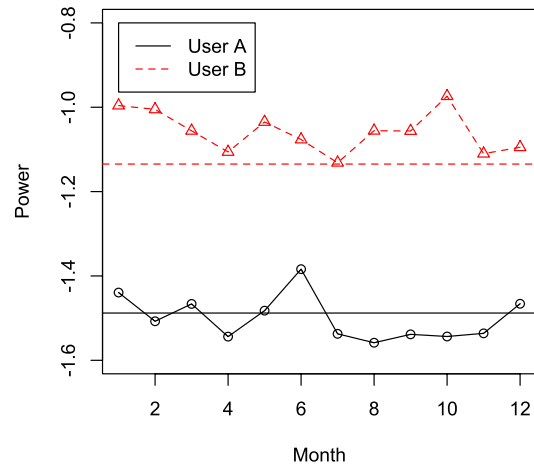


図 7 ユーザごとのべき指数の時系列変化
Fig. 7 Time series variation of exponential.

つきを示す値を算出し、12 カ月分の平均をとったところ 0.24 となった（図 6 に縦線で示す）。したがって、ユーザ個々の月ごとのべき指数の変化が、ユーザ全体でのばらつきに比べて小さいことが分かる。なお、平均の差について t 検定を行ったところ、 $p < 0.001$ で有意差が認められた。

図 7 は、前節で例を示したべき指数の大きいユーザ、小さいユーザそれぞれについて、1 カ月ごとのべき指数の時系列変化を、横軸は観測時期（2007 年の何月）、縦軸をべき指数にプロットしたグラフである。水平に引かれている線は、1 年分の履歴に基づくべき指数値である。

図 7 から、観測期間によらず 1 年分の履歴から算出したべき指数と月ごとのべき指数値はほぼ同じ値を示し、べき指数が観測期間によらずユーザの固有の特性を表す値となっており、要件 2 を満たしていることが分かる。

5.3 ネット利用傾向との相関

上位 3 位サイトまでに検索・ポータルサイトが 1 つでも含まれるユーザ（866 ユーザ）と検索・ポータルサイトが含まれず、掲示板・ブログサイトが含まれるユーザ（217 ユーザ）のべき指数の比較を行った。その結果、検索・ポータルサイトが含まれるユーザのべき指数平均は -1.15 、含まれないユーザのべき指数平均は -1.34 と大きな差があった（ p 値 < 0.001

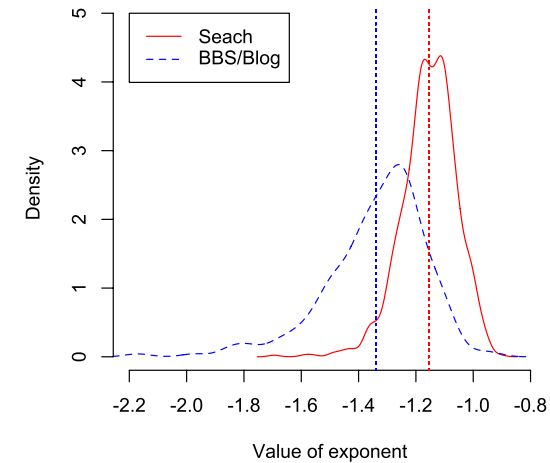


図 8 ユーザの利用サイト別のサイトアクセス分布のべき指数の比較
Fig. 8 Distribution of standard deviation.

で有意差あり）。

各ユーザ群ごとのべき指数とユーザ数の確率密度分布を図 8 に示す。図 8 から、検索・ポータルサイト利用中心のユーザと掲示板・ブログサイト利用中心のユーザでべき指数の分布が大きく異なり、要件 3 を満たしていることが分かる。

一方、アクセスサイト数でも同様に、検索・ポータルサイトが上位に含まれるユーザと、掲示板・ブログサイトが上位に含まれるユーザでの分布の比較を行った結果が図 9 である。サイト利用特性によらず、アクセスサイト数の分布がほぼ重なっており、アクセスサイト数の大小による順位付けは、ネット利用傾向との相関が見られないことが分かる。なお、サイト数の平均値について検定を行ったところ、 p 値 $= 0.41$ で有意差が認められなかった。

6. 考 察

6.1 べき指数を用いたサービス適用パターンの考察

ここで、実際にサイトの利用傾向と多様性受容の関係を考察するため、5.1 節で例を示した、べき指数の小さい図 3 のユーザ A と、べき指数の大きい図 4 のユーザ B のアクセス上位サイトへのアクセス URL をピックアップし、実際にその URL にアクセスして内容を確認することで、両ユーザのアクセス傾向を定性的に比較してみた。表 1 は、それぞれの

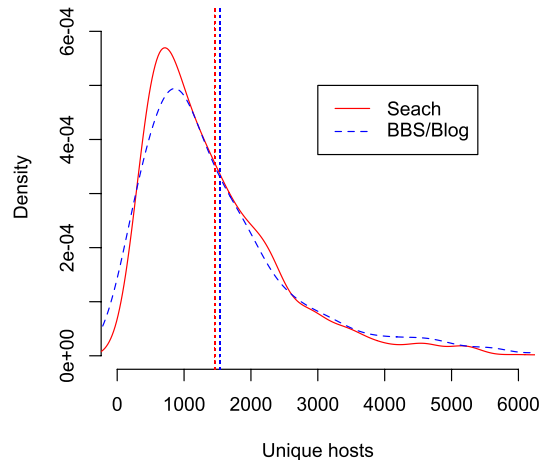


図 9 ユーザの利用サイト別のアクセスサイト数の比較
Fig. 9 Distribution of standarddeviation.

表 1 べき指数の大きいユーザ、および小さいユーザの利用サイト比較
Table 1 Compare top 5 sites.

順位	ユーザ A	ユーザ B
1 位	掲示板ブラウザサイト	ウィルスチェックソフト自動アクセス
2 位	ホビー関連掲示板サイト	検索サイト A
3 位	鎌倉市情報サイト	検索サイト B
4 位	掲示板ログアーカイブサイト	検索サイト C
5 位	恋愛関連掲示板サイト	オークションサイト

ユーザの利用上位サイトの比較表である。なお、両ユーザともに、両ユーザが示す、べき指数値周辺ユーザの典型的な行動を行っている。

べき指数の小さいユーザ A は日記・口コミ掲示板サイトが上位を占め（上位 6 サイトが口コミ掲示板と地域情報サイトであった）、すでに決まったジャンル（このユーザの場合は「ボーイズラブ」というジャンルにこだわっていた）についての情報収集を中心とした Web 利用がほとんどの行動を占めていた。対して、べき指数の大きいユーザ B の上位サイトは検索・ポータルサイトが上位を占め（上位 6 サイトのうち、ネットオークションサイトが 5 位に含まれる以外はすべて検索・ポータルサイトであった）、そのときどきで異なった目的での多様な Web 利用を行っている傾向にあった。

したがって、べき指数が小さいユーザは、特定の興味領域への執着が大きい多様性受容度が低い傾向があると考えられ、本ユーザに対しては領域を絞ったサービス提供が向いている可能性が高い。たとえば、過去の履歴から趣味嗜好を分析し、ユーザの好みのジャンルのみを絞り、セレンディビティを重視した推薦を行わないといった情報提供は、べき指数が小さいユーザに向いている。

一方、べき指数が大きいユーザは、目的達成のためにより最適なサイトがあればそちらを優先的に利用する冒険的なユーザであるといえる。したがって、あまり過去の履歴を考慮しない、多様な情報提供を行った方が満足度の高い情報提供が行える可能性が高い。

6.2 既存手法とのコスト面での比較

実際に White ら¹²⁾ の手法による属性推定を実装し、提案手法との計算速度の比較を行った。White らの手法は以下の手順で属性推定を行う。

- (1) セッション切り出し
- (2) 検索サイトか情報サイトに依ってラベル付け
- (3) レーベンシュタイン距離¹⁸⁾によりセッションごとの行動多様度を算出

比較を行ったプログラムはともに Perl で実装したものを使用。レーベンシュタイン距離の算出には Perl のモジュール Text::Levenshtein を用いた。

比較の結果、White らの手法では 1 人 1 年分のログから属性推定を行うのに、平均で 302.7 秒/人の計算時間がかかっているのに対し、提案手法では 0.2 秒/人と約 1/1000 の速度で計算ができた。したがって、計算コストの面で提案手法は既存のパターンマイニングによる手法に比べて優れていることが分かる。

また、White らの手法や、本稿の検証で行ったようなアクセス数上位のサイト種別に着目する手法では、サイト種別によるラベル付けが必須であり、ラベル付け対象サイトリストをつねに更新し続けるためのメンテナンスコストがかかる。また、レコメンドサービスをクライアントソフトとして実現しようとした場合、クライアントにあるサイトリストを更新する仕組みを構築するコストがかかってしまう。対して、提案手法ではサイト種別の判別は不要であり、メンテナンスコストがかからないという点においても優れている。

7. ま と め

本稿では、レコメンデーションなどでの利用を目的としたサイコグラフィック属性として、ユーザの多様性受容度をユーザ個々の Web アクセス履歴のべき乗則を利用して推定する手法の提案を行った。

検証の結果, (1) ユーザ個々のサイトアクセス傾向がべき乗分布を示すこと, (2) べき指数が観測時期によらずユーザ固有の特性値であること, (3) べき指数がユーザの多様性受容度に相関があることから, 提案手法が有効であることを示した. また, 提案手法は従来のパターンマイニングによる手法に比べて少ない計算コスト, メンテナンスコストで属性推定が可能な点において優れていることを示した.

ただし, 推定した属性に基づき多様性の調整を行ったレコメンデーションの効果の有無については, 未確認である. 今後はレコメンデーション・サービスなどへの適用により, 推定属性を用いたサービスへのユーザ満足度向上効果について検証を行っていく必要がある.

また, 本稿ではべき指数が時系列での変化が少ないことを示したが, 時系列でユーザの利用特性が大きく変化した場合には, べき指数が大きくな変化を示す可能性も考えられる. したがって, 時系列でべき指数が大きく変化している箇所についてより詳細な検証を行い, 利用目的の変化イベントの抽出への利用といった応用の可能性についても検討していく予定である.

参 考 文 献

- 1) 市川裕介, 小林 透: 大量の行動履歴情報を扱うプラットフォーム技術, 情報処理, Vol.51, No.1, pp.18–21 (2010).
- 2) 土方嘉徳: 情報推薦・情報フィルタリングのためのユーザプロファイリング技術, 人工知能学会誌, Vol.19, No.3, pp.365–372 (2004).
- 3) Kotler, P. and Armstrong, G.: *Marketing: An Introduction, 4th Ed.*, Prentice Hall (1997).
- 4) 平久保伸人: 消費者行動論, ダイアモンド社 (2005).
- 5) 敏弘神嵐: 推薦システムのアルゴリズム (1), 人工知能学会誌, Vol.22, No.6, pp.826–837 (2007).
- 6) Ziegler, C.-N., McNee, S.M., Konstan, J.A. and Lausen, G.: Improving recommendation lists through topic diversification, *WWW '05: Proc. 14th International Conference on World Wide Web*, New York, NY, USA, ACM, pp.22–32 (2005).
- 7) 天笠邦一, 加藤文俊, 岡部大介: シチュエーション情報の収集とその活用可能性に関する一考察, 情報処理学会研究報告 2008-UBI-18, pp.1–8 (2008).
- 8) 遠藤雅和, 廣田啓一, 山室雅司: 情報流過程におけるユーザ種別を考慮した協調フィルタリング方式の検討, 情報処理学会研究報告 (GN), Vol.2004, No.31, pp.97–102 (2004).
- 9) Broder, A.: A taxonomy of web search, *SIGIR Forum*, Vol.36, No.2, pp.3–10 (2002).
- 10) Rose, D.E. and Levinson, D.: Understanding user goals in web search, *WWW '04: Proc. 13th International Conference on World Wide Web*, New York, NY, USA, ACM, pp.13–19 (2004).

- 11) 高久雅生, 江草由佳, 寺井 仁, 齋藤ひとみ, 三輪眞木子, 神門典子: サーチエンジン検索結果ページにおける視線情報の分析, 情報知識学会誌, Vol.19, No.2, pp.224–235 (2009).
- 12) White, R.W. and Drucker, S.M.: Investigating behavioral variability in web search, *WWW '07: Proc. 16th International Conference on World Wide Web*, New York, NY, USA, ACM, pp.21–30 (2007).
- 13) Ichikawa, Y., Nakamura, M., Hata, K. and Nakagawa, T.: Provision of Services According to Individual User Preferences over a Cross-section of Sites Implemented with “Personalized-service Platform”, *NTT Technical Review*, Vol.6, No.4, pp.– (2008).
- 14) 石田賢治: ネットワーク上の付和雷同的なふるまい, 日本学会事務センター『SCIENTIA』, No.11, pp.1–3 (2001).
- 15) 井庭 崇, 深見嘉明, 斉藤 優: 書籍販売市場における隠れた法則性, 情報処理学会論文誌 数理モデル化と応用, Vol.48, No.6, pp.128–136 (2007).
- 16) 市川裕介: 協調フィルタリングを用いたレコメンサービス導入事例と課題, 情報処理, Vol.48, No.9, pp.972–978 (2007).
- 17) 海野 敏: 計量情報学の分布法則, 図書館情報学における数学的方法, chapter 2, 日本図書館学会研究委員会 (編), pp.16–54, 日外アソシエーツ (1994).
- 18) Levenshtein, V.: Binary codes capable of correcting deletions, insertions, and reversals, *Soviet Physics-Doklady*, Vol.10, No.8, pp.707–710 (1966).

(平成 22 年 5 月 21 日受付)

(平成 22 年 11 月 5 日採録)



市川 裕介 (正会員)

NTT サイバーソリューション研究所主任研究員. 平成 8 年慶應義塾大学大学院理工学研究科修士課程修了. 同年日本電信電話 (株) 入社. 通信履歴活用サービスの研究開発に従事. 平成 6 年本会学術奨励賞, 平成 17 年本会山下記念研究賞受賞.



小林 透 (正会員)

NTT サイバーソリューション研究所主幹研究員。昭和 62 年東北大学大学院工学研究科修士課程修了。同年日本電信電話 (株) 入社。以来、ソフトウェア生産技術、ユビキタスコンピューティング、暗号応用技術に関する研究開発に従事。電子情報通信学会会員。
