

Weblogにおける長期間履歴データの可視化法

Visualization for Long-Term Logs of Weblog Sites

森谷 友昭[†] 田中 明通[‡]
Tomoaki Moriya Akimichi Tanaka

高橋 時市郎[†] 加藤 泰久[‡]
Tokichihiro Takahashi Yasuhisa Kato

1. はじめに

近年、Weblog サイトの増加と共に Weblog に関する様々な研究が行われている。岡野らは Weblog 作者による記事の追加、またそれに対するサイト来訪者によるコメント、トラックバックなどをユーザ行動と定義し、約 1 週間収集した各ユーザ行動の回数から、定量的に Weblog サイトを役割、特徴別に分類する手法を示した[1]。

次なる段階として Weblog サイトの役割、特徴の時系列変化を検証するため、長期間に渡り各ユーザ行動の回数履歴を収集した。しかし得られたデータは時系列かつ多次元のデータであり通常のグラフを用いた可視化では、多数のグラフが必要である。そのため可視性が悪く、特徴を発見することが困難である。そこで、長期間に渡り収集した各ユーザ行動の回数履歴データを一度に俯瞰することが可能な可視化法を提案する。また可視化結果から得られた知見についても報告する。

2. 収集した履歴データ

ユーザ行動の回数履歴データの収集を行った期間は 2005 年 10 月 4 日から 2006 年 2 月 21 日までの約 5 ヶ月間で、各日、goo ブログ[2]内のアクセスランキング上位 100 位にランクインした Weblog サイトを対象に収集を行った。収集の対象としたユーザ行動を表 1 に示す。ユーザ行動に含まれている「ランキング順位」は他ユーザ行動の値が反映された値と考えられるので、収集、可視化対象とした。

表 1 収集対象としたユーザ行動の種別

①	ランキング順位
②	アクセス数
③	エントリ数
④	コメント数
⑤	トラックバック数

3. グラフによる可視化例とその限界

図 1 は、ある Weblog サイト 1 個分の、全回数履歴データを時系列に、折れ線グラフで可視化した例の一部である。図 1 に示したように各ユーザ行動別にグラフが必要となる。また図 1 では、可視化の対象とした 1 つの Weblog サイト内での“絶対的”な回数変動のみしか知ることが出来ない。他の Weblog サイトと比較するために、“相対的”なユーザ行動の回数の変動を可視化するには更なるグラフが必要となる。履歴データが長期間であるため、全期間の可視化を行った際、グラフが巨大化する傾向にある。

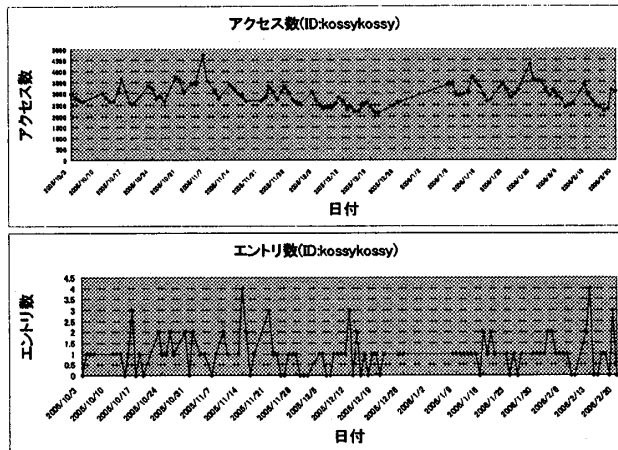


図 1 折れ線グラフによる可視化例
(アクセス数、エントリ数)

4. 提案手法とその可視化ツール

3. で述べたグラフの弱点を改善するため、次の条件を満たす可視化手法とその可視化ツールを検討し、開発した。

- 1 枚の画像に全てのユーザ行動の回数履歴データを表示できる
- ユーザ行動の回数の絶対的、相対的な変化を一度に可視化可能である
- 可視化範囲を長期間から短期間まで自由に調整可能である

条件(A)を実現するためには、各ユーザ行動回数を 1 次元で表示する必要がある。そこで擬似カラー表示を利用し、ある日付における行動回数を表す。色の決定式を式(1)に示す。

$$p = \frac{a_t - a_{\min}}{a_{\max} - a_{\min}} \quad \text{color} = (p, 1, 0) \quad \dots \text{式(1)}$$

$$\text{color}_{\text{rank}} = (1 - p, 1, 0)$$

a_t は日付 t におけるユーザ行動 a の回数、 a_{\min} 、 a_{\max} はそれぞれ履歴データ収集期間内におけるユーザ行動 a の最小値、最大値、 color の各要素は、0 から 1 の範囲の RGB 値をそれぞれ表す。 $\text{color}_{\text{rank}}$ は表 1 ① ランキング順位の擬似カラー化に用いられる。図 1 のアクセス数を、同じ期間、式(1)によって可視化した例を図 2 に示す。

また式(1)を用い、全てのユーザ行動回数を可視化した例を図 3 に示す。表示期間は図 1 と等しい。図 3 における①～⑤はそれぞれ表 1 に対応する。式(1)は図 1 の折れ線グラフと同じく絶対的なユーザ行動回数の変化を表している。

次に条件(B)を実現するために式(1)を拡張する。

$$q = \frac{a_t - a_{\min}}{a_{\max} - a_{\min}} \quad \text{color} = (p, 1, q) \quad \dots \text{式(2)}$$

$$\text{color}_{\text{rank}} = (1 - p, 1, 1 - q)$$

[†] 東京電機大学 大学院工学研究科 情報メディア学専攻

[‡] 日本電信電話 (株) サイバーソリューション研究所

a_{MIN} 、 a_{MAX} は履歴データを取得した全Weblogサイトにおけるユーザ行動 a の最小値、最大値を表す。式(2)を用い、全てのユーザ行動回数を可視化した例を図4に示す。



図2 式(1)によるユーザ行動回数履歴可視化
(図1、アクセス数と同じデータ)

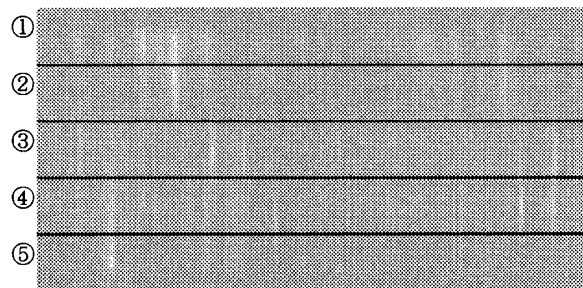


図3 式(1)によるユーザ行動回数の可視化例

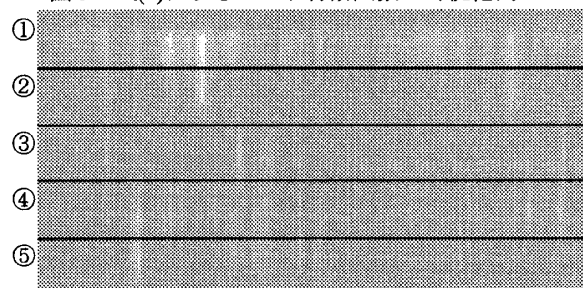


図4 式(2)によるユーザ行動回数の可視化例

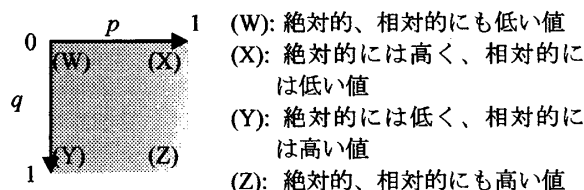


図5 p 、 q による color の変化

図4において現れた色の意味を図5にまとめる。

図4①ランキング順位を見ると、可視化対象サイト内では大きく変動を繰り返して、絶対値の変動は大きい。しかしその平均は水色となっている(図5(Y)~(Z)の間で変化)。すなわち相対値は高い値を保っていることから、他Weblog サイトと比較すると安定して高ランクに位置していることが分かる。

条件(C)を実現するために、今回構築した可視化アプリケーションには表示期間の拡大、縮小及びスライド機能を実装した。これにより長期間のデータであっても一定の画面領域で全データを俯瞰することが可能となる。

5. 結果及び考察

実際の各 Weblog サイトから得られたユーザ行動の履歴データを今回の提案手法で可視化した。可視化結果から特徴的なものを取り上げ、考察する。

図6の例では、①ランキング順位がほぼ図5(W)~(X)の間で変動していることから高ランクには至らないが安定し

たランクを維持していることが分かる。③エントリ数、④コメント数が同時に高くなっているのは、定期的に訪れている閲覧者が新たなエントリに対してコメントを寄せているためと考えられる。このことから定期的に訪れる閲覧者が存在し、そのため安定したランクを維持していると推測できる。

一方、図7の結果では、時間経過によって①ランキング順位が白色から緑色へ変化していく。これは高ランクに位置していたサイトが時間経過によってランクが低下していることを示している。図中の(I)では③エントリ数の増加に反応して②アクセス数が上昇しているが、(J)では起きていない。つまり、この Weblog サイトを定期的に訪れる閲覧者が漸減していることが推測できる。

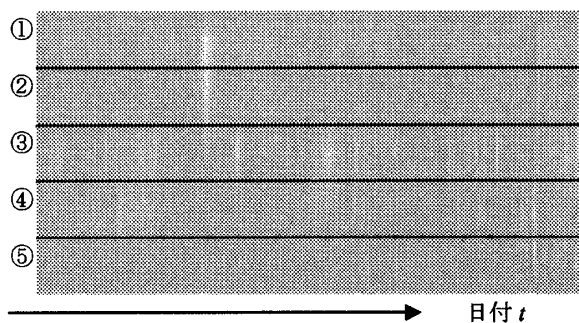


図6 可視化結果-1

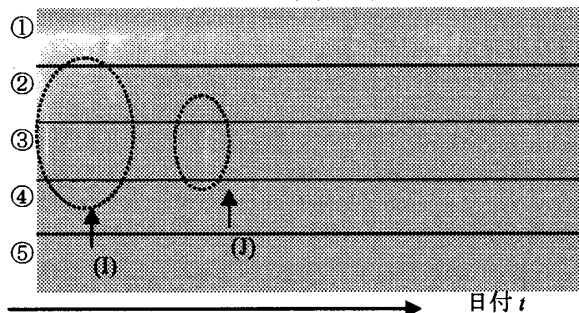


図7 可視化結果-2

本報告の可視化手法では、以上述べたようにユーザ行動間の相関的な変動を容易に発見することができ、Weblog サイトの時間的変位を捉えることができる。

6. まとめ

本稿では、擬似カラー表示を利用して、長期間に渡り収集した各ユーザ行動の回数履歴データを一度に俯瞰可能な可視化手法を提案した。また本可視化手法を利用し、Weblog サイトの時間的変位を発見し、その特長についての考察を述べた。今後は、複数の Weblog サイトを可視化、比較するための可視化手法を検討するとともに、類似した Weblog サイトの発見手法や、Weblog サイトの特徴づけを進める。

文献

- [1] 岡野他 「ユーザの行動に基づく Weblog の特徴別分類手法の検討」 電子情報通信学会総合大会 D-13-8 (2005)
- [2] <http://blog.goo.ne.jp/>