

# Web Site 構成の検討方法

## －非対称クラスター分析法を用いて－

### Evaluating the Structure of Web Site by Asymmetric Cluster Analysis

01009690 立教大学 岡太 彬訓 OKADA Akinori

立教大学 大川 英恵 OKAWA Hanae

#### 1 はじめに

本発表で Web site 構成というのは、Web site の各ページ間に張られているリンクによるページ間の関連を意味している。本発表の目的は、Web site 構成が、実際のページ間の移動に基づくページ間の近さを反映しているかどうかを評価し、実際のページ間の移動に基づくページ間の関連を反映するように、Web site 構成を改善するための方法を提案することである。

#### 2 データ

2001 年 3 月から 4 月にかけての 21 日間に、ある Web site 内での各ページの滞在時間を IP アドレス毎に記録したデータをもとに、あるページから別のページへの移動頻度を算出した。移動頻度の算出にあたっては、あるページでの滞在が目的ではなく、他のページに行くことが目的である場合の単なる通過（例えば、ボタンを探すなど）を、そのページでの滞在と区別する必要がある。単なる通過であるページに滞在した場合、その前のページから当該ページへの移動ではなく、その前のページから当該ページの次のページの移動と考える。そのために、一定の値以下の滞在時間は、そのページへの滞在ではなく、通過と考えることにした。2 種類の滞在（通過のための滞在と真の滞在）があるならば、通過のための滞在は比較的短時間であり、真の滞在は比較的長時間であると考えられる。滞在時間のヒストグラムを作成し、2 種類の滞在時間の分布が分けられる時間を境にして、通過のための滞在と真の滞在と考えることにした。作成した滞在時間のヒストグラムにもとづき、1 つのページでの滞在時間が 13 秒以下の場合にはそのページを通過したとみなした。

分析対象となった Web site は、32 個のページをもつが、その内の 2 ページは、滞在時間が全て 13 秒以下であったので、真の滞在が 1 回もなかったと考え、分析から除いた。30 個のページ間の移動頻度からなる  $30 \times 30$  の表が求められた。この表の  $(j, k)$  要素は、ページ  $j$  からページ  $k$  への移動頻度である。

#### 3 非対称クラスター分析法

ページ  $j$  からページ  $k$  への移動頻度は、ページ  $k$  からページ  $j$  への移動頻度と等しいわけではない。したがって、ページ間の移動頻度からなる  $30 \times 30$  の表は、非対称である。移動頻度の表の  $(j, k)$  要素を、ページ  $j$  からページ  $k$  への近さと考えれば、この表はページ間の類似度行列と考えることができる。非対称類似度を分析するためには、データのもつ非対

称を考慮した分析方法が必要である (Arabie & Hubert, 1994).

非対称クラスター分析法を用いてページ間移動頻度行列を分析した。従来の非対称クラスター分析法 (Okada, 2000; Okada & Iwamoto, 1996) では、適切でない点がある。ページ  $j$  からページ  $k$  への移動頻度を  $s_{jk}$  とする。従来の方法は、クラスターを構成する際

$s_{pq} < s_{qp}$  ならば ページ  $p$  がページ  $q$  を吸収して新たなクラスターを構成する

$s_{qp} < s_{pq}$  ならば ページ  $q$  がページ  $p$  を吸収して新たなクラスターを構成する

ことにより、階層クラスターを構成する凝集的クラスター分析法である。Web site の構成を考える際には、これとは反対に、あるページから他のあるページへの移動頻度が、逆方向の移動頻度よりも大きい場合に、前者が後者を吸収して新たなクラスターを構成する方がリンクを張るページが中心になる、という意味で結果の理解がしやすい。本報告では、従来の非対称クラスター分析法でのクラスター間の吸収方向を逆転させるため、ページ間移動頻度行列を転置した行列を分析した。

#### 4 結果と今後の課題

非対称クラスター分析の結果は、玄関ページ (トップページ) が、他の多くのページを吸収し、大きなクラスターを構成した。しかし、玄関ページ以外のページも、他のページを吸収して、ある程度の大きさのクラスターを構成していた。また、これらのクラスターの中には、直接リンクの張られていないページに吸収される場合があり、新たにリンクを張るべきページを示唆していると考えられる。また、現在玄関ページからリンクが張られているが、玄関ページに吸収されないページは、玄関ページからのリンクを通じての移動が少なく、このようなリンクを除くことが考えられる。

今後解決すべき、いくつかの問題点の中の3点を挙げる。(a) 本発表で用いた IP アドレス毎に記録したデータにおいては、ブラウザでの「戻る」、「進む」、「次へ」などのボタンをクリックしたページ間の移動は捉えられていない。(b) IP アドレス毎にページ間の移動が記録されているため、複数の人が同一の IP アドレスのパソコン等を利用している場合に、それらの人達を区別していない。(c) あるページでの滞在時間が 13 秒以下の場合に、通過のための滞在としたが、これは通信速度の違いを考えていない。さらに、さまざまな条件 (Web site 利用時間や滞在頻度などの違い) のもとでページ間移動頻度行列を求め、これらを分析して条件間の差異を明らかにすることができれば、よりきめの細かい Web site の構成を考えることが可能になる。

#### 参考文献

- Arabie, P., & Hubert, L. (1994). Cluster analysis in marketing research. In: R. P. Bagozzi (Ed.), *Advanced methods in marketing research* (pp. 160-189). Oxford: Blackwell.
- Okada, A. (2000). An asymmetric cluster analysis study of car switching data. In W. Gaul, O. Opitz, & M. Schader, (Eds.), *Data analysis: Scientific modeling and practical application* (pp. 495-504). Berlin, Germany: Springer-Verlag.
- Okada, A., & Iwamoto, T. (1996). University enrollment flow among the Japanese prefectures: A comparison before and after the Joint First Stage Achievement Test by asymmetric cluster analysis. *Behaviormetrika*, 23, 169-185.