

# Markov 連鎖を用いたウェブナビゲーション過程の評価

山本哲生\* 北島宗雄† 高木英明‡ 張勇兵‡

\*筑波大学第三学群社会工学類  
yamamo70@shako.sk.tsukuba.ac.jp

†工業技術院 生命工学工業技術研究所  
kitajima@nibh.go.jp

‡筑波大学社会工学系  
takagi@shako.sk.tsukuba.ac.jp  
ybzhang@shako.sk.tsukuba.ac.jp

## 1. はじめに

2000年1月現在のウェブサイトの総数は1000万にも及んでいると推定されている<sup>[1]</sup>。ウェブサイトを開設する目的も、エンターテインメント、ショッピング、教育、情報提供、組織紹介、コミュニティなど、多岐にわたっている。さらにインターネットの急速な発達に伴い、インターネット上で入手可能な情報は日々増加しており、もはやそのすべてをフォローすることも困難な状況にまで達している。ことに最近の通信インフラ整備の拡充により、インターネットアクセス環境も飛躍的に向上しており、今後の情報の入手の方法は、以前の雑誌・本・CDから、インターネットに次第にその中心を移していくと思われ、益々、インターネット上の情報を絞りこむ技術が必要となると考えられる。現在、我々がインターネット上の情報を検索する際の手助けとなるものとして、Yahoo、Goo、Infoseekなどのサーチエンジンが知られている。しかしながら、実際は、そのウェブサイトに探している情報があるとわかっていながら、その情報を探し当てることのできないことがしばしばあり、ウェブの利用が必ずしも効率的に行われていないというのが現状である。実際、Spoolら<sup>[2]</sup>はいくつかのウェブサイトを対象として、被験者にそれぞれのサイトで提供されている情報を探索させたが、その成功率は12%から43%と非常に低かった。

どのようにしたらより使いやすいウェブサイトをデザインできるかということに関していくつかのガイドラインが提案されている。しかしながら、他のガイドラインがそうであるように、ウェブサイトやウェブページのデザインのためのガイドラインもその全体が必ずしも首尾一貫しているわけではない。例えば、ハイパーリンクの内容をわかりやすくするためには、どこにリンクしているのかがわかるように、リンクのラベルはあまり単純で簡潔すぎではない。しかし、その一方で、ユーザがウェブページを開いたときにはページの内容を「読む」のではなく、「走査」することが知られている。走査を効率的に行うにはラベルは簡潔なほうがいい。

情報(コンテンツ)を提供するためのインフラ(インターネット)が整備される中で、そのインターフェース(ウェブサイト)のデザインは、情報環境の有益さを高める上で非常に重要な役割を果たす。しかしながら、現在のウェブサイトのインターフェースデザインがその役割を十分に果たしているとはいえない。

マルチメディアを「データの提供」に陥らせずに「情報提供」にしていくために、次に述べる情報デザインの概念から迫ってみたい。生データは時として情報になり得るが、必ずしも情報ではないため、情報とデータは混同してはいけない。受け手に意味があって初めて情報と呼べるものであることを理解しておく必要がある。生データを伝える能力と意味のあるメッセージを生成する能力を混同してはいけない。データと知識の間にある情報不安症に受け手を引き込んでしまうからである。大量の情報を目前にして、知りたいことにたどり着けずに苛立っている状態が、まさに情報不安である。我々は知りたいことが何かを特定でき、その内容が満たされたとき「理解」が起こる。極めて個人的なレベルにおいてしか、データが情報として活かされることはないのである。自分の好奇心に従って、興味あるものをつなぎ合わせ、興味対象の発見する順に従ってばらばらのデータを組み替えていく作業を通じて知識を得ていくのである。つまり、情報とは「解釈」ということができる。データが情報に変わるまでの過程を、ここでは仮に「アクセス」と呼ぶことにする。問題の多くには、何を解決したいかはひとつしかないが、どう解決するかという解決方法は沢山ある。そのひとつしかない問題を発見し、どう解決していくか、その過程(つまりアクセス)を選択して、わかりやすく表現することが情報デザインである。理解していなかったことを理解するところにコミュニケーションは生まれるわけであるから、

情報デザインは、何も知らないユーザを常に想定して、ユーザの理解を促すようなアクセスを選択していくことが必要である。この問題に対するひとつのアプローチのし方は、ウェブを利用しようとするユーザが、どのようにして個々のページと接するのかを認知のレベルで理解し、そこでの知見をウェブサイトのデザインに反映させるといったものである。

本論文では、ユーザがウェブサイト上でタスクを遂行する過程の認知モデルである CoLiDeS モデル (2.1. 参照) を考慮に入れ、ウェブページに表示されるリンクを順次選択しながら、与えられたターゲット情報が提供されているページにたどり着くというゴールを達成する過程を Markov 連鎖を用いてモデル化・評価する。

## 2. ウェブナビゲーション過程のモデル化

まずウェブのリンク構造を定義し、各リンクの類似度 (2.2. 参照) を付与する。ここでウェブページが表示されたときに個々のリンクが選択される確率は、意味空間において定義されるターゲットとリンクとの類似度を適当に変換して求めることができる。また各ウェブページには、直前のページに戻るためのリンク「BACK」が存在する。「BACK」を選択する確率は、同じページの他のリンクの選択確率に依存する。

ウェブページには、探索開始時に表示されているホームページ、リンクとして BACK リンクのみを持つターミナルページ、それ以外の中間ページがあるとす。ホームページには、ラベル付きリンクのほかにリンク「中止」があり、それが選ばれた時には探索は終了する。さらに、あるリンクを選択してターミナルページが表示され、それがターゲットと異なるということは、ゴールが達成されていないことを意味するので、その場合にはリンク「BACK」を選択して直前のページに戻ると仮定する。次のページの図 1 は簡単なウェブナビゲーションのモデルである。

### 2.1. CoLiDeS モデル

ここで CoLiDeS モデル<sup>[7]</sup>について簡単に説明する。CoLiDeS とは、Comprehension-based Linked model of Deliberate Search の頭文字をとったもので、オフィスアプリケーションを利用するユーザの認知モデルを、ウェブを利用する場面に適用できるように拡張したものである。ベースにしている認知アーキテクチャーは文章理解の理論から生まれてきた Construction-Integration Architecture<sup>[3]</sup> である。これらのモデルでは、ユーザはウィンドウやメニューなどのインタフェース要素に関する知識、及びタスクに関する知識は持っているとして仮定し、ユーザが今までに経験したことの無い局面、例えば、普段利用しているアプリケーションで利用したことのない機能を使う、あるいは、Windows で行っていたタスクを Macintosh で行う、といったような局面で、すでに持っている知識を利用して、タスクを遂行するのに必要な操作を選択する操作選択プロセスのモデル化を行う。ここで生成される操作系列は、以前に経験し記憶に蓄積されているものを想起して生成されるのではなく、ディスプレイに表示されている情報をそのときのタスクの文脈に即して理解し、インタフェース要素に関する知識を利用して生成されるものであり、柔軟で適応的な操作選択過程がモデル化されている。

さて、ウェブページが表示されたとき、ユーザはハイパーリンクの選択やナビゲーションボタンの選択を行って、タスクを遂行しようとする。ウェブの利用とオフィスアプリケーションの利用の決定的な相違点は、1回のマウスクリックで生じるディスプレイ上での変化の大きさにある。実際、オフィスアプリケーションでは、インタフェース要素であるメニュー、ツールバー、ラジオボタン (Windows や Mac OS などの GUI 環境で複数の選択項目の中から 1つだけを選択するための仕組み)、チェックボックスなどをクリックしたときに生じる変化は局所的であることが多い。したがって、次の操作の選択をディスプレイのどこに着目して行えばよいかということに迷うことは少ない。一方、ウェブページのハイパーリンクをクリックするとページの大部分が書き換えられることが多い。そのため、次に操作を行う場所を選択すること (アテンションプロセス) が重要になる。CoLiDeS モデルでは、オフィスアプリケーションのモデルにアテンションプロセスを付加し、ウェブユーザのインタラクション過程のモデル化を行っている。

### 2.2. 類似度の定量化法

CoLiDeS モデルでは、どのオブジェクトを選択するか判断はゴールとそのオブジェクトの類似度に基づいてなされる。ここで、類似度は、LSA (Latent Semantic Analysis : 潜在的意味解析) として知られている手法を用いて計算する<sup>[4]</sup>。LSA は単語とそれが現れる文脈 (ドキュメント) の関係を統計的に評価する手法であり、各単語および文脈は約 300 次元の意味空間内のベクトルとして表現されている。単語が複数ある場合には、合成語の意味はベクトルを合成することによって表現される。合成語間の類似度は、この意味空間を利用して定義できる。米国大学生の語彙レベルに基づいて構成した意味空間をもとに、インタラクティブに、語や合成語間の類似度を得ることができるウェブページが <http://lsa.colorado.edu> に提供されている。

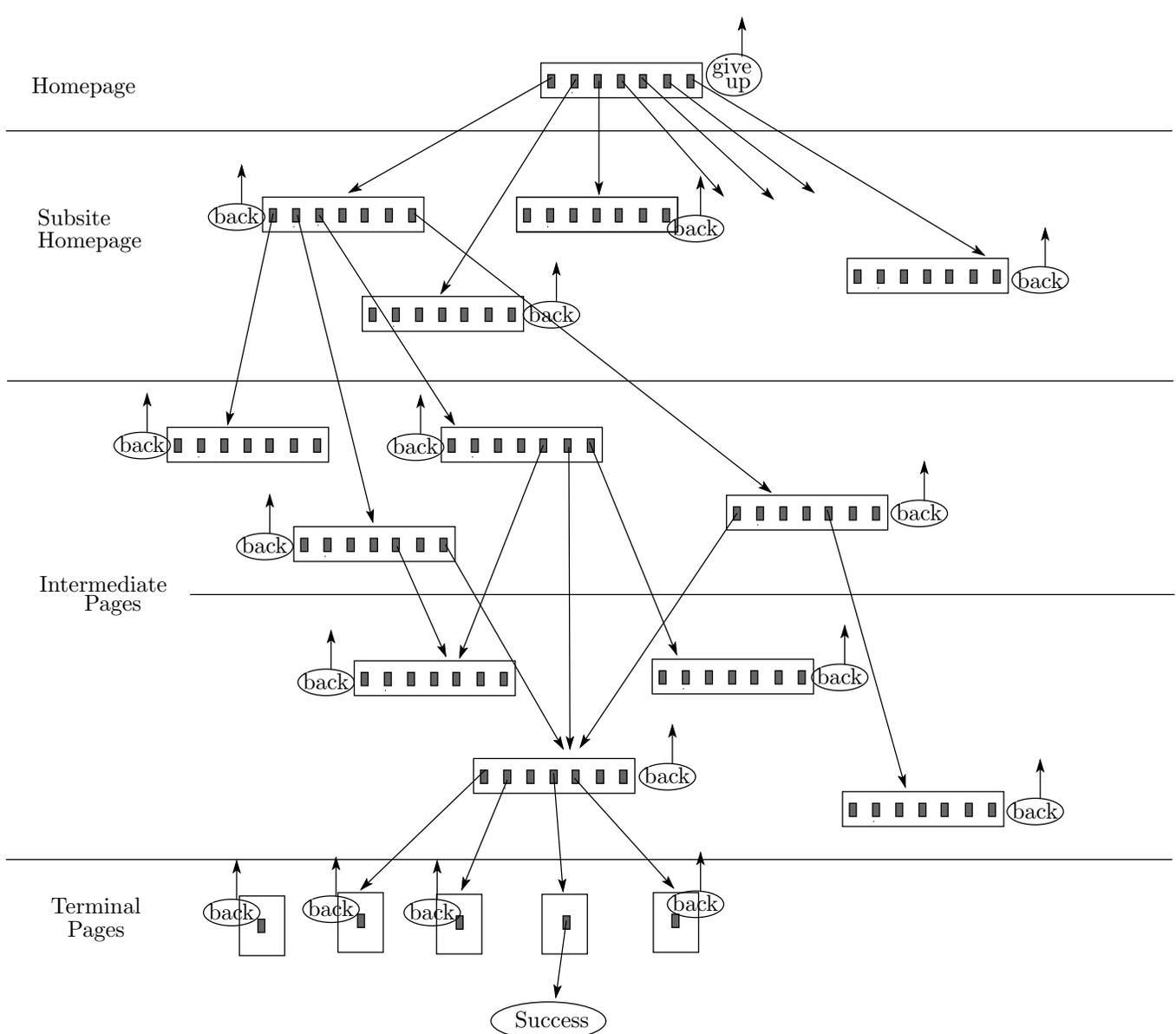


図 1: ウェブナビゲーションのモデル

### 2.3. Markov 連鎖への当てはめと変形

ウェブのナビゲーション過程を考えると、ウェブページのハイパーリンクをクリックするとページの大部分が書き換えられることが多いことは既に述べた。しかし、ある1つのページに着目すると、そのページへたどり着くには直前のページを分かっていたら十分である。また、各ページを1つの状態として、クリックを時刻パラメタとして考えると、ウェブのナビゲーション過程は離散的時刻パラメタと離散の状態空間  $S$  を持つ Markov 連鎖といえる。すなわち、 $X_n$  を  $n$  回目のクリックで現れるページとすると、以下の式が成り立つ [5,6]。

$$P\{X_{n+1} = j \mid X_0 = i_0, X_1 = i_1, X_2 = i_2, \dots, X_n = i_n\} = P\{X_{n+1} = j \mid X_n = i_n\} \quad (1)$$

また最初のステップ後の状態  $X_1$  で条件付けることにより、任意の  $i, j \in S$  について、次の方程式が成り立つ。

$$f_{ij}^{(1)} = P_{ij} \quad ; \quad f_{ij}^{(n)} = \sum_{k \neq j} P_{ik} f_{kj}^{(n-1)} \quad n = 2, 3, \dots \quad (2)$$

ここで、 $P_{ij}$  と  $f_{ij}^{(n)}$  の定義を述べる。状態空間  $S = \{0, 1, 2, \dots\}$  をもつ Markov 連鎖  $\{X_n; n = 0, 1, 2, \dots\}$  において、状態  $i$  から  $j$  への推移確率を

$$P_{ij} := P\{X_{n+1} = j \mid X_n = i\} \quad i, j \in S$$

と定義する。また、状態空間  $S$  をもつ Markov 連鎖  $\{X_n; n = 0, 1, 2, \dots\}$  が初期状態  $X_0 = i \in S$  から始まる  
とき、状態が初めて  $j \in S$  になるまでのステップ数  $T_{ij}$  を、状態  $i$  から  $j$  への初到達時間という。その確率分布

$$f_{ij}^{(n)} := P\{T_{ij} = n\} = P\{X_1 \neq j, X_2 \neq j, \dots, X_{n-1} \neq j, X_n = j \mid X_0 = i\}$$

$$n = 1, 2, \dots \quad ; \quad i, j \in S$$

を、状態  $i$  から  $j$  への  $n$  ステップ初到達確率と定義する。ここで、もちろん

$$f_{ij}^{(1)} = P_{ij} \quad i, j \in S$$

である。(2) の方程式の意味は、「初期状態  $X_0 = i$  から始まる Markov 連鎖が  $n \geq 2$  ステップ目において初めて状態  $j$  になるということは、最初のステップで  $j$  でない状態  $k$  になり、そこから数えて  $n-1$  ステップ目において初めて状態  $j$  になる」ということである。

(2) の両辺に  $\omega^n$  を掛け、 $n = 1, 2, \dots$  について加えると

$$\sum_{n=1}^{\infty} f_{ij}^{(n)} \omega^n = P_{ij} \omega + \sum_{n=2}^{\infty} \omega^n \sum_{k \neq j} P_{ik} f_{kj}^{(n-1)} = P_{ij} \omega + \omega \sum_{k \neq j} P_{ik} \sum_{n=2}^{\infty} f_{kj}^{(n-1)} \omega^{n-1} \quad (3)$$

となるので、母関数の関係式

$$F_{ij}(\omega) = \omega \left[ P_{ij} + \sum_{k \neq j} P_{ik} F_{kj}(\omega) \right] \quad |\omega| < 1 \quad (4)$$

が得られる。

(4) で  $\omega \uparrow 1$  とすると

$$f_{ij} = P_{ij} + \sum_{k \neq j} P_{ik} f_{kj} \quad i, j \in S \quad (5)$$

となる。ここで状態  $i$  から  $j$  への初到達確率  $f_{ij}$  の定義は、初期状態  $X_0 = i$  から始めて、有限時間内に状態が  $j$  になる確率

$$f_{ij} := P\{T_{ij} < \infty\} = \sum_{n=1}^{\infty} f_{ij}^{(n)} \quad i, j \in S$$

である。(5) の方程式の意味は、「初期状態  $X_0 = i$  から始まる Markov 連鎖が状態  $j$  になるのは、最初のステップで  $j$  になるか、最初のステップで  $j$  でない状態  $k$  になり、そこから状態  $j$  になる」ということである。

また、(4) を  $\omega$  で微分してから  $\omega \uparrow 1$  とすると平均初到達時間 (平均ステップ数)

$$\mu_{ij} = \lim_{\omega \rightarrow 1} \frac{dF_{ij}(\omega)}{d\omega} = P_{ij} + \sum_{k \neq j} P_{ik} f_{kj} + \sum_{k \neq j} P_{ik} \mu_{kj} = f_{ij} + \sum_{k \neq j} P_{ik} \mu_{kj} \quad (6)$$

が得られる。ここで状態  $i$  から  $j$  への平均初到達時間  $\mu_{ij}$  は、状態  $i$  から  $j$  への初到達時間  $T_{ij}$  が確率 1 で有限であるとき、その期待値

$$\mu_{ij} := E[T_{ij}] = \sum_{n=1}^{\infty} n f_{ij}^{(n)} \quad i, j \in S$$

として定義する。

有限 Markov 連鎖の全ての状態  $S$  は、有限個の既約な正再帰的同値類  $C_1, C_2, \dots, C_R$  ( $R < \infty$ ) と、非再帰的な状態の集合 (一般に、いくつかの非再帰的同値類の和集合)  $T$  に一意的に分割される。

$$S = C_1 \cup C_2 \cup \dots \cup C_R \cup T$$

つまり Markov 連鎖の初期状態  $X_0 = i$  がどれかの正再帰的同値類に属せば、その後の状態は同じ同値類に留まる。初期状態  $X_0 = i$  が非再帰的な状態ならば、有限 Markov 連鎖は永久に非再帰的な状態であることはできず、

必ずいつかどれかの正再帰的同値類に属する状態になり、その後は同じ同値類に留まる。

ここで、このウェブナビゲーションの Markov 連鎖モデルでは、2つの吸収状態があると考えているので、2つの吸収状態は正再帰の状態であり、それ以外は非再帰の状態であるということが出来る。したがって (5) より、状態  $i$  は非再帰的であり、状態  $j$  は再帰的であるとするとき、初到達確率  $f_{ij}$  は、初期状態  $X_0 = i$  から始まる Markov 連鎖が  $j$  の属する再帰的同値類に吸収される確率を表す。再帰的状态  $j$  の属する同値類を  $C(j)$  とし、非再帰的状态の集合を  $T$  とすれば、初ステップ解析の方法により

$$f_{ij} = \sum_{k \in C(j)} P_{ik} f_{kj} + \sum_{k \in T} P_{ik} f_{kj} + \sum_{k \notin C(j), T} P_{ik} f_{kj} \quad (7)$$

が成り立つ。(7) で、 $k \in C(j)$  なら  $f_{kj}=1$  であり、 $k \notin C(j), T$  なら  $f_{kj}=0$  であるから、与えられた再帰的状态  $j$  に対して、連立方程式

$$f_{ij} = \sum_{k \in C(j)} P_{ik} + \sum_{k \in T} P_{ik} f_{kj} \quad i \in T \quad (8)$$

が得られる。ここで右辺の  $\sum_{k \in T} P_{ik} f_{kj}$  を左辺に移行すると

$$\sum_{k \in C(j)} P_{ik} = f_{ij} - \sum_{k \in T} P_{ik} f_{kj} \quad i \in T \quad (9)$$

となるから、各  $j$  に対して、これを解いて吸収確率  $f_{ij}$  を求めることができる。

さらに (6) の右辺の  $\sum_{k \neq j} P_{ik} \mu_{kj}$  を左辺に移行すると

$$f_{ij} = \mu_{ij} - \sum_{k \neq j} P_{ik} \mu_{kj} \quad (10)$$

となり、各  $j$  に対して、これを解いて平均初到達時間  $\mu_{ij}$  を求めることができる。

### 3. ナビゲーションの履歴を覚えなない Markov 連鎖モデル

この章では、Markov 連鎖を用い、具体的な数値をあてはめてウェブナビゲーション過程を評価する。ここで、ナビゲーションの履歴を覚えなないモデルについて説明する。もともと Markov 連鎖では、過去と現在のすべての履歴の条件の下での将来の状態確率は現在の状態のみに依存し、過去の履歴とは独立であるという Markov 性を持っている。しかし、実際のウェブナビゲーション過程では、一度訪れたページにターゲットが存在しなかった場合、もう一度そのページを訪れることは考えにくい。つまり過去の履歴に少なからず影響を受ける。モデルを簡単にするため、ここで述べる「ナビゲーションの履歴を覚えなない Markov 連鎖モデル」とは、過去にどこを訪れたかを考慮にいれないで、自分の探しているターゲットを見つけるまで探索を続けるモデルである。したがって、過去に一度訪れたページをもう一度訪れる可能性がある。

#### 3.1. 具体的なモデル

例として、ウィンドウ画面で表示されるツールバーの項目であるメニュー階層のウェブナビゲーション過程を考える。メニュー項目は、File、Edit、General、Assign、Transformation、Appearance、Tools の 7 項目あり、そのそれぞれにプルダウンメニューが存在する。また、さらにもう 1 階層下にサブプルダウンメニューが存在する。したがって、リンク構造はウィンドウの最初の画面をホームページ、メニュー項目とプルダウンメニューを中間のページ、サブプルダウンメニューをターミナルページとする 4 階層からなる。

表 1 に、メニューに現れる言葉とターゲットである “change the graph type to column” という言葉との類似度を示す。この表を作成するためには、ウェブサイト <http://lsa.colorado.edu> に行き、one-to-many という項目を選び、“change the graph type to column” というターゲットとメニュー項目を入力すると、ターゲットとメニュー項目間の類似度を得ることができる [8]。ここで、類似度が 0.15 以下の項目については、下限を設けてそこに訪れる可能性がないものとした。その場合には、1 つ上の階層に戻るリンク「BACK」の選択に加えることで考慮している。またホームページから出るパスにおいて、そのような状態がある場合は、「中止」へのリンクの選択に加える。ホームページからスタートして、ターゲットもしくは「中止」に到達するまで、有限時間でこのタスクは実行される。また、このタスクにおいて、ターゲットまでの正解のパスは、ホームページ → General → Change Type → Graph Type である。Graph Type を選択することにより、タスクのゴールにマッチしたダイアログボックスが表示されるので、成功したことがわかる。

| メニュー項目               | プルダウンメニュー              | サブプルダウンメニュー       | 類似度  |
|----------------------|------------------------|-------------------|------|
| Edit                 | Repeat                 |                   | 0.16 |
|                      | Copy                   |                   | 0.22 |
|                      | Find                   |                   | 0.24 |
|                      | Replace                |                   | 0.24 |
|                      |                        |                   | 0.21 |
| General              | <b>Change Type</b>     |                   | 0.38 |
|                      | Table                  |                   | 0.64 |
|                      | Document Info          | <b>Graph Type</b> | 0.56 |
|                      | Preferences            | Create            | 0.18 |
|                      |                        |                   | 0.19 |
|                      |                        | Application       | 0.43 |
|                      |                        | Text file         | 0.16 |
|                      |                        | Graph Format      | 0.23 |
|                      |                        | Save Preferences  | 0.17 |
|                      |                        |                   | 0.43 |
| Assign               | Headings               |                   | 0.19 |
|                      |                        |                   | 0.25 |
|                      |                        | Heading 1         | 0.46 |
|                      |                        | Heading 2         | 0.4  |
|                      |                        | Heading 3         | 0.44 |
|                      |                        | Heading 4         | 0.53 |
|                      |                        | Heading 5         | 0.49 |
|                      | Layout view            |                   | 0.52 |
|                      | New Data               |                   | 0.16 |
|                      | Bullet List            |                   | 0.32 |
|                      | Number List            |                   | 0.23 |
|                      | Apply Format           |                   | 0.37 |
|                      |                        |                   | 0.35 |
|                      | Block quote            | Standard Layout   | 0.23 |
|                      | Coordinates Definition |                   | 0.2  |
| Esoteric preferences | Table                  | 0.16              |      |
|                      | Math Function          | 0.18              |      |
|                      |                        | 0.23              |      |
|                      |                        | 0.16              |      |
| Transformation       | Rotate                 |                   | 0.16 |
|                      | Leaning function       |                   | 0.18 |
|                      | Change Font            |                   | 0.23 |
|                      |                        |                   | 0.53 |
|                      | Sort                   | References        | 0.22 |
|                      | Graph Values           | Graph Font        | 0.4  |
|                      |                        |                   | 0.18 |
|                      |                        | Into              | 0.33 |
|                      |                        | Bar Points        | 0.28 |
|                      |                        | Over              | 0.25 |
|                      |                        | 0.2               |      |
| Appearance           | Fit Layout             |                   | 0.22 |
|                      |                        |                   | 0.22 |
|                      |                        | Layout Level      | 0.19 |
|                      | Replenish              | Build             | 0.16 |
|                      | Graph Artwork          |                   | 0.29 |
|                      |                        |                   | 0.39 |
|                      | Rows                   | Reduction         | 0.19 |
|                      |                        | 0.26              |      |

表 1: ターゲット (change the graph type to column) との類似度

### 3.2. 推移確率行列の作成

ウェブナビゲーション過程モデルの推移確率行列の前提となる定義を述べる。目標とするターゲット (change the graph type to column) が存在するページは1つに固定されている。また、ターミナルページに行き着いたときのリンク「BACK」の定義として、その場所がターゲットが存在するページではない場合、確率1で直前のページに戻ることを仮定する。ターミナルページ以外の状態からのリンク「BACK」の選択は、その状態から出る類似度 0.15 以下のパスが複数個あるとき、それぞれの状態のもつ類似度を加算する。そうでない場合は、類似度を一律に 0.15 と設定した。さらに「中止」という状態も考慮にいれなければならない。実際の人間のウェブナビゲーションにおいては、例えばリンクを4回選択してもターゲット情報が得られなかったら中止するなどといった様々なパターンが考えられるが、ここでは、最初のページに「中止」へのリンクが存在すると仮定し、全てのページを訪問してもターゲットに到達することができなかった場合、「中止」に至って探索を終了する。図2に履歴を覚えないウェブナビゲーションのモデルを表す。ここで、リンクに付けられた数は類似度を表す。

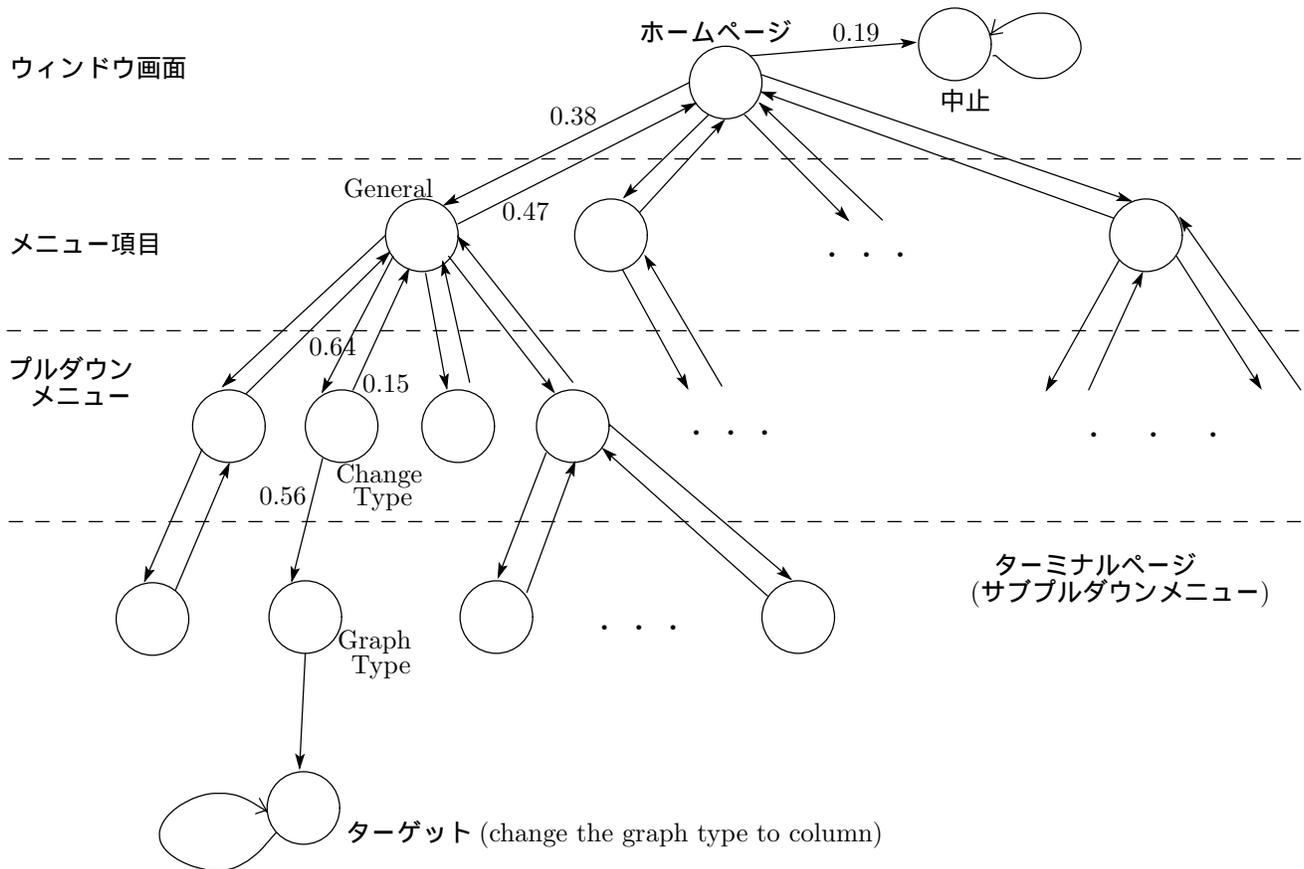


図 2: 履歴を覚えないウェブナビゲーションのモデル

次に推移確率行列の作成方法について述べる。離散的時刻 Markov 連鎖を考えると、行列はページ (状態) ごとに適当に番号を付け、ページ  $i$  からページ  $j$  に行く推移確率を  $(i, j)$  成分とする。そして先程計算した類似度を各行和が1になるように、推移確率に変換する。表1に「ホームページ」と「中止」の状態を追加すると、全状態数は55個となるので、ホームページからターミナルページへ順に状態の番号をつける。便宜上、吸収状態となっているターゲットを状態54とし、「中止」を状態55とすると、推移確率行列は(11)のように表される。

$$(P_{ij}) = \left( \begin{array}{ccc|cc} P_{11} & P_{12} & \dots & P_{1j-1} & P_{1j} \\ P_{21} & P_{22} & \dots & P_{2j-1} & P_{2j} \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ \hline 0 & 0 & \dots & 1 & 0 \\ 0 & 0 & \dots & 0 & 1 \end{array} \right) \quad (i, j = 1, 2, \dots, 55) \quad (11)$$

この推移確率行列を4つに分け、左上の行列を  $Q$ 、右上の行列を  $R$  とすると、 $Q$  は  $53 \times 53$  の行列であり、 $R$  は  $53 \times 2$  の行列である。ここで  $53 \times 53$  の単位行列  $E$  を作成し、 $E$  から  $Q$  を引いたものを  $M$  とする。上記の推移確率行列から初到達確率  $f_{ij}$  を求めるが、このウェブナビゲーション過程の場合は、ホームページからターゲットへの到達確率  $f_{1,54}$  だけ計算できれば十分である。 $R$  の第1列を取り出し、線形連立方程式を作ると、(9) から

$$\begin{pmatrix} P_{1j} \\ P_{2j} \\ \vdots \\ P_{53j} \end{pmatrix} = M \begin{pmatrix} f_{1j} \\ f_{2j} \\ \vdots \\ f_{53j} \end{pmatrix} \quad (12)$$

となり、 $Ax = b$  の形になるので  $f_{1,54}$  が得られる。

$\mu_{ij}$  を求める場合は、 $j$  を固定して、同様に線形連立方程式を作ると、(10) から

$$\begin{pmatrix} f_{1j} \\ f_{2j} \\ \vdots \\ f_{53j} \end{pmatrix} = M \begin{pmatrix} \mu_{1j} \\ \mu_{2j} \\ \vdots \\ \mu_{53j} \end{pmatrix} \quad (13)$$

が得られ、ホームページからターゲットへ至るまでの平均クリック数  $\mu_{1,54}$  を得ることができる。

ここで、本論文の  $f_{1,54}^{(n)}$ 、 $f_{1,54}$ 、 $\mu_{1,54}$  の意味は

$$f_{1,54}^{(n)} = \frac{\text{ターゲットに } n \text{ 回のクリックで至るようなアクセスの数}}{\text{すべてのアクセスの数}} \quad (14)$$

$$f_{1,54} = \frac{\text{ターゲットに至るようなアクセスの数}}{\text{すべてのアクセスの数}} \quad (15)$$

$$\mu_{1,54} = \frac{\text{ターゲットに至るアクセスについて、そのクリックの合計数}}{\text{すべてのアクセスの数}} \quad (16)$$

であるから、(15, 16) より、ターゲットに至るアクセスだけについての平均クリック数は

$$\frac{\mu_{1,54}}{f_{1,54}} = \frac{\text{ターゲットに至るアクセスについて、そのクリックの合計数}}{\text{ターゲットに至るようなアクセスの数}} \quad (17)$$

となる。

### 3.3. 初到達確率 $f_{ij}$ と平均初到達時間 $\mu_{ij}$

上記の方程式を用いてホームページからターゲットへ、または「中止」への初到達確率  $f_{1,54}$ 、 $f_{1,55}$  と平均初到達時間  $\mu_{1,54}$ 、 $\mu_{1,55}$  を求めた。表2がその結果である。表2における「平均クリック数」の意味は、ターゲット、または「中止」に至るアクセスだけについての平均クリック数を表す。さらに類似度を適当に変えた場合、つまり推移確率を変えることで、 $f_{ij}$  と  $\mu_{ij}$  がどのように変化するかを表2で観察する。その際、表2の  $a$  は計算した類似度をそのまま使用した。このとき、正解となるパスの類似度は、ホームページ → General (0.38) → Change Type (0.64) → Graph Type (0.56) であり、リンク「BACK」の類似度は、Graph Type → Change Type

(0)→General (0.15) → ホームページ (0.47) である。 $b$  は  $a$  のなかでホームページからターゲットまでの正解のパスだけを類似度 0.8 に設定したとき、つまりホームページ →General (0.8) →Change Type (0.8) →Graph Type (0.8) である。 $c$  は  $a$  のなかでホームページからターゲットに到達するまでの状態のリンク「BACK」と中止となる類似度を 0.1 に設定したとき、つまり Graph Type→Change Type (0)→General (0.1) → ホームページ (0.1)→ 中止 (0.1) である。上と同様に、 $d$  は  $a$  のなかでホームページからターゲットに到達するまでのパスの類似度を 0.9、中止となる類似度を 0.1、そのほかの状態の類似度を全て 0 と設定したとき、 $e$  は  $a$  のなかでホームページからターゲットまでのパスだけを類似度 0.15 と低く設定したとき、 $f$  はこのモデルの全てのパスを同じ確率に設定したときを表している。以上の 6 通りのモデルについて計算を行った。

| 場合分け   |         | ターゲット    | 中止       |
|--|---------|----------|----------|
| $a$ : 類似度をそのまま使用   | 初到達確率   | 0.508766 | 0.491234 |
|  | 平均初到達時間 | 82.9498  | 76.8703  |
|  | 平均クリック数 | 163.041  | 156.484  |
| $b$ : ターゲットまでのパスの類似度を 0.8                                      | 初到達確率   | 0.71266  | 0.28734  |
|  | 平均初到達時間 | 68.4915  | 25.929   |
|  | 平均クリック数 | 96.1068  | 90.238   |
| $c$ : リンク「BACK」と「中止」の類似度を 0.1                                  | 初到達確率   | 0.762467 | 0.237533 |
|  | 平均初到達時間 | 111.767  | 32.7344  |
|  | 平均クリック数 | 146.586  | 137.81   |
| $d$ : ターゲットまでのパスの類似度を 0.9、<br>「中止」への類似度を 0.1、<br>その他の状態の類似度を 0 | 初到達確率   | 0.889024 | 0.110976 |
|  | 平均初到達時間 | 3.05738  | 0.137745 |
|  | 平均クリック数 | 3.43903  | 1.24121  |
| $e$ : ターゲットまでのパスの類似度を 0.15                                     | 初到達確率   | 0.097996 | 0.902004 |
|  | 平均初到達時間 | 29.0877  | 258.779  |
|  | 平均クリック数 | 296.825  | 286.893  |
| $f$ : 全てのパスを同じ確率   | 初到達確率   | 0.25     | 0.75     |
|  | 平均初到達時間 | 21       | 53       |
|  | 平均クリック数 | 84       | 70.6667  |

表 2: 初到達確率と平均初到達時間と平均クリック数

### 3.4. 数値結果の考察

付与する類似度によって結果は異なるものの、それぞれの初到達確率と平均初到達時間を求めることができた。予期されたように、ターゲットまでの類似度が高くなればなるほど、到達できる確率は高くなる。例えば、表 2 の  $d$  のように極端にターゲットとの類似度が高い場合、ほかの状態を無視して迷うことなく目的とするページにたどり着くことができる。さらに戻る確率や中止となる確率が低い場合も同様である。この場合は全ての状態を回るという前提条件のために到達確率が上がることが分かる。逆に、 $e$  の場合のようにターゲットまでの類似度が非常に低いと、目的のページに到達することが難しく探索は終了する。

初到達確率と平均初到達時間（平均ステップ数）との関係は、直感的には到達する確率が高くなるにつれて、平均ステップ数は減少していくように思われた。確かにそのような関係が読み取れなくもないが、はっきりとした結果として示すことができなかった。この理由として挙げられることは、このモデルでは全ての状態を探索し、なおかつ一度訪問したところにもう一度訪問する可能性があるからである。この問題を解決するために、一度訪問したところには二度と行かないようなモデルを定義しなければならない。

また、ここでの平均クリック数と平均初到達時間との関係を述べる。例えば、表 2 の  $c$  のように、ターゲットに到達する確率が非常に高いとき、「中止」となる平均初到達時間は、ターゲットまでの平均初到達時間よりもはるかに小さい値をとっている。ターゲットに到達する確率が高いのにもかかわらず、ターゲットまでの平均初到達時間が大きいのは、不自然である。これは、(16) より、「中止」に至るアクセスについて、そのクリック数の合計をすべてのアクセスの数で割っていることが原因である。したがって、(17) から、「中止」に至るアクセスについて、そのクリック数の合計を「中止」に至るようなアクセスの数で割ると、ターゲットへの平均初到達時間の値と近い値になっており、平均初到達時間の値は妥当だといえることができる。

#### 4. まとめと今後の課題

本研究では、メニュー階層（パソコンのウィンドウの画面で表示されるツールバーの項目）をモデルとして取り上げた。ここでは、ターゲットを1つ決定し、それについての類似度を計算した。そしてホームページからターゲット、または中止への初到達確率  $f_{ij}$  と平均初到達時間  $\mu_{ij}$  の値を Markov 連鎖を用いてウェブナビゲーション過程をモデル化することで計算することができた。しかし、実際の人間のウェブナビゲーション時における行動はまだ十分に記述されていない。さらに3章で述べたように、ウェブナビゲーション過程において一度訪問したところには二度と行かないようなモデルを定義しなければならない。このことについては現在検討中であり、議論を具体的な形で行い、シミュレーションなどを行って解析していこうと考えている。この研究がうまくいき、今後ユーザの視野に立った最適なウェブサイトの設計方法の確立に貢献していくことができれば幸いである。

#### 参考文献

- [1] Nielsen, J.: Designing Web Usability, Indianapolis: New Riders Publishing (2000). ヤコブ ニールセン : ウェブ・ユーザビリティ-顧客を逃がさないサイトづくりの秘訣、エムディエヌコーポレーション (2000).
- [2] Spool, J.M., Scanlon, T., Schroeder, W., Snyder, C., and DeAngelo, T.: Web Site Usability: A Designer's Guide, San Francisco: Morgan Kaufmann (1999). ジャレッド・M. スプール他 : Webサイトユーザビリティ入門~サイトの「使いやすさ」を考える~、トッパン (2000).
- [3] Kintsch, W.: Comprehension; A Paradigm for Cognition, Cambridge University Press (1998).
- [4] Landauer, T.K. and Dumais, S.T.: A solution to Plato's problem: The latent semantic analysis theory of acquisition, induction, and representation of knowledge, Psychological Review, 104, pp.211-240 (1997).
- [5] 尾崎俊治 : 確率モデル入門 (朝倉書店、1996).
- [6] 森村英典・高橋幸雄 : マルコフ解析 (日科技連出版社、1979).
- [7] 北島宗雄 : ウェブナビゲーションの認知モデル、ヒューマンインタフェースシンポジウム 2000.
- [8] Rodolfo Soto: Learning and performing by exploration: label quality measured by latent semantic analysis, Proceeding of the CHI 99 Conference on Human Factors in Computing Systems: The CHI is the Limit, pp.418-425, May 15-20, 1999, Pittsburgh, Pennsylvania, U. S. A.