

潜在意味解析を用いた情報ウェブサイトの Markov 連鎖によるモデル化と評価

狩谷典之*

北島宗雄†

高木英明‡

張勇兵‡

*筑波大学大学院システム情報工学研究科

nkariya@sk.tsukuba.ac.jp

†産業技術総合研究所人間福祉医工学研究部門

kitajima@ni.aist.go.jp

‡筑波大学社会工学系

{takagi, ybzhang}@shako.sk.tsukuba.ac.jp

コンピュータとインターネットの発展と普及により、我々は容易にして膨大な量の情報にアクセスし、必要な情報を得ることが可能となった。しかしながら、求めている情報がそこにあるはずとわかっていても、その情報へのアクセスルートが、予想していたものと違っていたりするために、目的の情報を見つけだすことができなったり、見つけだすまでに必要以上に時間を要してしまうことも少なくない。そこで本研究では実際に存在する情報データベースウェブサイト为例にとり、ウェブユーザビリティ評価法の一つである、ウェブ認知ウォークスルーを利用するとともに、Markov 連鎖でモデル化し、目的の情報にたどりつくまでに要するクリック数を調べることによって、ウェブサイトのユーザビリティを評価する。また簡単なリンク構造の変更によって、既存のユーザビリティの問題が改善され、ユーザが求める情報を探し易くなることも示す。

1. はじめに

昨今のコンピュータとインターネットの発展と普及により、我々は必要な情報が欲しいとき、時間と場所を選ばずに、容易に膨大な量の情報にアクセスし、目的の情報を探ることができる手段を得ることができるようになった。しかし、それらの情報は日々増えていく一方であり、またそれらの情報は整理されて、蓄積されているわけではない。実際、インターネットによって情報収集を行おうとしているユーザは、サーチエンジンを利用したり、関連するウェブサイトからリンクをたどるなどして目的の情報を探し出すのが一般的である。しかしながら、ユーザがそのサイトに目的の情報があるとわかっていながら、探し出すことができない場合が 50% を越しているという調査結果 [1] や、サイトサーチエンジンを利用してサーチを行った場合の情報探索成功率が 30% になり、利用しないで行った場合の成功率が 53% になったという調査結果 [2] もある。

このことは目的の情報までのアクセスルートや情報のありかが、ユーザの予想していたものと違うなどの、ウェブサイトの構造やインターフェースのデザインがうまくいってないといった、ユーザビリティに問題があることに原因があると考えられる。このような問題により、利用するユーザは情報を取得できなかったり、探すことができたとしても多大な探索時間を要してしまう可能性がでてくる。このことは情報を提供することを目的としているウェブサイト側にとっても、深刻な問題であるといえる。そのような問題が起きないようにするには、ユーザが間違えることなく、少ない手間で目的の情報を獲得できるようにウェブサイトデザインすることが重要である。そこで本論文では、実際に存在する情報ウェブサイト为例にとり、ウェブ認知ウォークスルーや Markov 連鎖などの手法を用いて、理論的にウェブサイトのユーザビリティを簡単に評価できることを示す。また、簡単なウェブデザインの変更で、ユーザビリティが改善されることも示す。

2. ウェブ認知ウォークスルー (CWW)

本章では、ウェブサイトユーザビリティ評価法のひとつである「ウェブ認知ウォークスルー (CWW: Cognitive Walkthrough for the Web)」と、その基礎となるモデルについて紹介するとともに、それについての簡単な説明を行う。詳しい内容は [5] を参照していただきたい。

ユーザの、WWW ブラウザによる階層化されたリンク構造をもつウェブサイトの探索の主な方法は、ウェブサイト上のハイパーリンクを順次選択していくという操作である。このような階層化されたメニュー構造は、オフィス用アプリケーションや ATM (現金自動預払機)、携帯電話など日常で利用されるものに幅広く適応されている代表的なインタフェースデザインである。そしてユーザがこのような機器を利用するとは、操作の系列を覚え、記憶から呼び出して実行していくのではなく、表示される内容を見て、理解し、その時点で適切な項目を選択することによって操作されることが多く、実際、メニューベースの機器の操作の場合、繰り返し行

われるタスクであってもユーザはメニューの項目を覚えていないという実験結果がある [3]。このような場合、ユーザは項目選択には「ラベル追従ストラテジ」を用いることが知られている [4]。ここで、選択の主な基準は、タスクの表現と項目との意味的適合度である。したがって、項目選択を行うためのインタフェースのデザインが、タスクの内容と適合していない場合には、タスクをうまく遂行できないことが予測される。

2.1. ウェブ探索過程の認知モデル (CoLiDeS)

CWW は、ユーザがウェブサイトを探査する過程を、認知モデルに基づいてシミュレートすることによってユーザビリティを評価し、問題発見を行う。ここで認知モデルに基づくシミュレーションは、理解に基づく意図的探索過程の認知モデルである CoLiDeS モデル (Comprehension-base Linked model of Deliberate Search) に基づいて行われる。CoLiDeS モデルは、ユーザがインタフェースを探りながら機器操作を行う過程の認知モデルである LICAI モデル (Linked model of Comprehension-based Action planning and Instruction taking) を拡張したモデルである。これらのモデルの核となっている心的プロセスは、領域の選択、操作対象の選択、操作の選択という 3 つの選択のプロセスである。それぞれの選択は 2 段階で行われる。まず、インターフェースディスプレイに表示されている情報を操作の目的に照らして理解する。次に、その理解に基づいて目的との適合性を評価し、目的に最も近いものを選択する。前者は理解過程、後者は手段-目標分析 (means-ends analysis) とよばれる問題解決過程である。CoLiDeS および LICAI モデルでは、理解過程を人間がどのように文章を理解するかということに関する認知理論、Construction-Integration 理論 [6] によってモデル化している。

CoLiDeS モデルでは、ウェブページ上のオブジェクトであるテキストやグラフィックスを理解し、評価するプロセスが重要である。ここでは、ウェブページ上のオブジェクトの心的表象を作り出し、関連する既有知識を用いて精緻化し、目的との適合性を評価する。そして最も目的とゴールとの適合性の高いオブジェクトが操作対象として選択される。CoLiDeS モデルは、上述の 3 つの選択プロセスを、注意段階、操作選択段階の 2 段階に分けて、ユーザのウェブナビゲーション行動をモデル化している。まず、注意段階では、ユーザはウェブページを部分領域に分割し、その各々に対して適当な記述を与える。それには、例えば、見出しやページレイアウトを表現する言葉が利用される。次に、ユーザは、現在の目的との適合度が最も大きい部分領域を選択する (領域選択プロセス)。操作選択段階では、訪問者は選択された部分領域内のすべてのウィジェットの心的記述を生成し、現在の目的との適合度が最も大きい操作対象を選択する (操作対象選択プロセス)。そして、そのウィジェット (ハイパーリンクなど) に対する操作 (通常はクリック) を選択する (操作選択プロセス)。

2.2. 潜在意味解析 (LSA) による意味的類似度の評価

CoLiDeS モデルでは、初めて遭遇したウェブページに表示されているウィジェットを選択するとき、各々に対して与えられる記述と目的との意味レベルでの適合度と文字レベルでの適合度を総合的に評価し、最も適合度の高いものを選択する。これは、モデルの基礎となっている Construction-Integration 理論 [6] に基づいている。ここで、総合評価は、インタフェース上のオブジェクトや活性化された知識をノードとするネットワークに活性を伝播させることによって行われる [7]。CWW では、この手続きを簡略化して、意味的類似度のみを用いて適合度の評価を行う。

意味的類似度の定量化には潜在意味解析 (LSA: Latent Semantic Analysis) [8] として知られている手法を用いる。LSA は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、各語および各文脈はそれぞれに対して定義される約 300 次元の意味空間内のベクトルとして表現される。複数の語で形成される合成語の意味は、個々の語のベクトルを合成することによって表現される。2 つの合成語間の意味的類似度は、対応する 2 つのベクトルのなす角の余弦として定義される。たとえば、*human computer interaction* と *software engineering* の意味的類似度は 0.64 である。これは、これらの言葉が同時に現れる文脈が多い、つまり類似した言葉であることを反映している。一方、*parenting* という *human computer interaction* と同時に現れる文脈がないと思われる言葉については、類似度は 0 となっている。このように、LSA を用いることにより意味的類似度を客観的に定量化できる。なお、<http://lsa.colorado.edu> には米国学生のいくつかの学年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提供されている。

LSA を利用することにより、ユーザの目的の記述をより現実的に行うことができる。従来、ユーザインタフェースの中から、目的を達成するために必要なオブジェクトを選択する過程のモデル化の研究では、たとえば「心臓病に関する情報を探す」のような簡潔かつ特定の意味を持つ目的記述が用いられていた。しかし、ユーザがウェブなどで探索を行うときに、必ずしもこのように明確に記述された目的を心に抱いているわけではない。LSA を用いることにより、探索したい直接的な目標ばかりでなく、一般的な関心や動機、また背景を目的の記述に含めることができる。

3. CWW の適用と分析対象ウェブサイト

本章では、以上で説明した CWW を、実際に存在する情報ウェブサイトに応用し、ユーザビリティ問題の発見を行った結果について述べる。本論文では分析対象のウェブサイトとして、Microsoft が提供しているオンライン百科事典サイト Encarta Encyclopedia (<http://encarta.msn.com>、以下 Encarta) を設定とした。

3.1. ウェブサイト概要

ここでは、今回分析の対象にしたウェブサイトである、Encarta の構成について説明する。なお、Encarta 自体は日々更新され、細かい内容が頻繁に変わることがあるが、ここでは分析をした時点である 2003 年初めの情報を用いて説明する。Encarta には、総数 41952 個のアーティクルが掲載されているが、これらのアーティクルは 9 個のカテゴリとそれに属する 94 個のトピックを用いて階層的に分類されている。目的のアーティクルを表示させるには最初にカテゴリを選択し（領域選択プロセス）、次にトピックを選択する（操作対象選択プロセス）。すると、そのトピック下に掲載されているアーティクルのタイトル一覧がアルファベット順に表示されるので、その中から目的のアーティクルを探し出し、クリックする。なお、全アーティクルのうち約 9 割は唯一のアクセスルートをもつが、残りの約 1 割は複数のアクセスルートをもっている。トピックを選択することで現れるアーティクルの数は 46406 個あるので、4454 個のアーティクルが複数のアクセスルートを持つことがわかる。表 1 に各カテゴリ名とそれらに属するトピックの数、アーティクル数を示す。CWW では、ナビゲーションに用いられるリンクのみが評価されるので、その点で Encarta は評価しやすいというメリットがある。

カテゴリ名	トピック数	アーティクル数
Art, Language, & Literature	13	5309
Geography	13	8978
History	9	6087
Life Science	14	5153
Performing Arts	6	4845
Physical Science & Technology	16	4930
Religion & Philosophy	7	2900
Social Science	12	6562
Sports, Hobbies, & Pets	4	1640

表 1: Encarta のカテゴリとそれに属するトピック数とアーティクル数。

3.2. CWW によるユーザビリティ問題の定義と評価 (Usability Problem)

本節では、CWW によるユーザビリティの問題の評価法を定義し、実際に Encarta について評価する。Encarta に限らず、このような情報ウェブサイトについて重要な点は、ユーザができるだけ簡単に、間違えることなく求めている情報にたどりつけるかという点である。ユーザビリティ問題は、後述の Markov 連鎖によるモデル化とそれによる評価への橋渡しを考えて、以下のように比較的簡略に定義する。CWW による詳細なユーザビリティの分析については [9] を参照していただきたい。

見出しの問題性：

1. Weak Scent Problem(Category) (以下, C-WS) : 目的のアーティクルと各カテゴリとの意味的類似度の値がある閾値 (δ) 以下になるとき、このアーティクルは、「見出しを選択する際の手がかりの十分性に問題がある」と判断する。
2. Goal-Specific Competing Heading Problem (以下, GSCH) : 見出しの手がかりの十分性についての問題はないが、意味的類似度の値が最も大きくなるカテゴリが、目的のアーティクルの属するカテゴリでない場合、「正しい見出しの選択に関して問題がある」と判断する。

リンクの問題性について：

1. Weak Scent Problem(Topic) (以下, T-WS) : 目的のアーティクルの属するカテゴリ下の各トピックについて、正解であるトピックとの意味的類似度の最大値と、不正解であるトピックとの意味的類似度の最大値のいずれもがある閾値 (δ') 以下になるとき、このアーティクルは、「リンクを選択する際の手がかりの十分性に問題がある」と判断する。
2. Goal-Specific Competing Link Problem (以下, GSCL) : リンクの手がかりの十分性についての問題はないが、正解であるトピックとの意味的類似度の値と、不正解であるトピックとの意味的類似度

の値の比が、ある閾値 (γ) 以下になるトピックがあるとき、「正しいリンクの選択に関して問題がある」と判断する。

以上のように定義したユーザビリティの問題を、Encarta の各ア－ティクルについて調べた結果を表 2 に記す。なお、今回の Encarta の分析では、ユーザが求める目的のア－ティクルについては、そのア－ティクルの最初のパラグラフを用いる。これは大規模な情報ウェブサイトにおいて、目的のア－ティクルのデータをそのまま全て用いることは時間的、量的に評価上効率的ではないことと、Encarta のア－ティクルは、最初のパラグラフで、そのア－ティクルに関する要旨が記述されていることが多いからである。そして今回の分析では、LSA により目的のア－ティクルと、各カテゴリやトピックとの意味的類似度を計算するが、できるだけ実際のユーザの判断に近いものを求める必要がある。Encarta では、カテゴリを選択するときには、それらに属するトピックも見ながら判断し、選択することができる。そこで今回の分析では、実際のユーザの判断に近いものとして、簡易的に、見出しとして、カテゴリとそれに属するトピック全てを含んだものを適用し、リンクとしては、トピックをそのまま用いる。また分析に必要な各言葉の LSA による意味的類似度は、米国大学生のよる語彙レベルを表現する意味空間を用いている。なお前述の定義で現れたパラメータは、今回は $\delta = 0.1$ 、 $\delta' = 0.1$ 、 $\gamma = 0.8$ とした。

	ア－ティクル数	計		ア－ティクル数	計
問題なし	6648	6648	C-WS のみ	1454	6050
C-WS のみ	1454		C-WC かつ T-WS	1908	
GSCH のみ	3008	C-WS かつ GSCL	2688		
T-WS のみ	191	4462	T-WS のみ	191	3730
GSCL のみ	17863		C-WC かつ T-WS	1908	
		18054	GSCH かつ T-WS	1631	
C-WS かつ GSCL	2688	12788	GSCH のみ	3008	11200
GSCH かつ T-WS	1631		GSCH かつ T-WS	1631	
C-WC かつ T-WS	1908		GSCH かつ GSCL	6561	
GSCH かつ GSCL	6561			11200	
		41952	GSCL のみ	17863	27112
			C-WS かつ GSCL	2688	
			GSCH かつ GSCL	6561	

表 2: Encarta における Usability Problem の結果 (左: 一覧, 右: 各問題別)。

表 2 の結果は、Encarta の見出しとリンクを評価した、各ア－ティクルのユーザビリティの結果である。この表から Encarta は、ユーザビリティに問題がないア－ティクルは全体の 15%程度しかなく、残りは何らかの問題があると確認される。正しい見出しやリンクを発見するための手がかりが不十分になるア－ティクル (C-WS, T-WS) は全体の 10~15%と少ないが、正しい見出しの選択に問題があるア－ティクル (GSCH) は全体の 30%程度もあり、正しいリンクの選択に問題があるア－ティクル (GSCL) は全体の 65%程度もあることがわかる。このことは Encarta は、見出しやリンクを選択する際の手がかりの十分性にはそれほど問題がないが、正しいアクセスルートを探し出すことが困難なデザインになっているといえる。ユーザが見出しやリンクの言葉を認識したときに類似した言葉を想起する行動をシミュレートしたものと比較しても、とりわけ大きな違いは見られず、既存の見出しやリンクに使われる語をそのまま使用して分析した場合でも、実際のユーザが、それらの言葉を認識し、類似した言葉を想起して操作する行動に近い結果が生じていることがわかる。見出しを発見するための手がかりの十分性に問題があるア－ティクルが減少していることがわかるが、これは前述の通り見出しについては、カテゴリとそれに属するトピックの語を全て含んでいるので、その分、見出しの認識によって想起されるであろう、類似語が非常に多くなり、ユーザは見出しについて、手がかりの選択に幅を増えたためと考えられる。

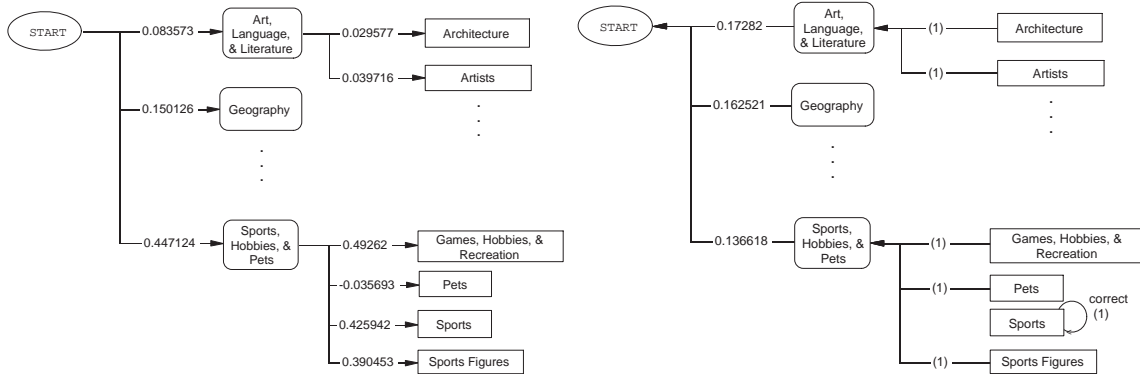


図 1: Markov 連鎖によるモデル化の例 1 (左: 順推移方向, 右: 逆推移方向)。

4. Markov 連鎖によるモデル化

本章では, Encarta が階層構造になっている点を利用し, このウェブサイトを Markov 連鎖でモデル化する方法を示す. また, そのモデル化により, ユーザが目的のアーティクルにたどりつくまでに要する平均クリック数を求めることによって, ウェブページのユーザビリティの解析と評価が容易に行えることを示す. このような階層的な構造のモデル化については, 階層メニュー構造をモデル化したり [10], e-コマースにおけるユーザ行動やシステム解析を行っているもの [11][12], 実際の家電などの操作に対してモデル化するなどして [13], インターフェースデザインを成功率や実行速度といったユーザビリティの観点から評価し, ユーザインタフェースの効率の良さや有用性を確認している研究結果もある. 本論文では, 簡単なモデル化によっても, ユーザビリティの評価の有用な評価ができることを示し, また簡単なリンクの張替えによって, ユーザビリティの問題が改善され, ユーザが目的の情報を探しやすくなることを示す.

4.1. サイトのモデル化

ここでは, 実際に Encarta のウェブサイトを Markov 連鎖でモデル化する方法を述べる. 前述のように, Encarta は, カテゴリ (見出し), トピック (リンク), アーティクルの 3 層構造になっており, ユーザが目的のアーティクルにたどりつくためには, 以下のような選択を順に行うと仮定することができる.

1. 見出しの選択: まずホームページ (Start 状態) から, 9 つのカテゴリの中から適当と思われるカテゴリ (見出し) を探して選択する.
2. リンクの選択: つぎに, そのカテゴリ下のトピック一覧が現れるので, 「見出しの選択」同様, そこから適当と思われるトピック (リンク) を見つけて選択する. あるいは, トピック (リンク) の選択時に, 目的のアーティクルへのルートとして適当なものが存在しないと思えば, 「見出しの選択」に戻り, カテゴリ (見出し) の選択から始める.
3. アーティクルの選択: トピック (リンク) を選択すると, そのトピックに属するアーティクルのタイトル一覧が得られるので, 目的のアーティクルのタイトルがあるか探す. 目的のアーティクルのタイトルがあれば, そのタイトルを選択することにより目的のアーティクルを発見したことになる. 目的のアーティクルが見つからなければ, 「リンクの選択」状態に戻る.

上述のユーザの選択行動は, Encarta の簡単な階層構造と, CWW により, 現実のユーザの行動に近い選択行動になっていると考えられる. しかし実際のユーザの行動には, 途中で探索をあきらめてしまったり, 途中で考え込むといった行動も含まれる. このような現実的なユーザの行動も考慮してユーザ行動の分析を行うために, 様々なモデルを構築し, 評価を試みているものもある [10] が, 今回のモデル化の目的は, ユーザビリティの問題を見出すことにあり, 忠実にユーザ行動をシミュレートすることではない. 実際の行動において何らかの問題が生じる場合, ユーザは様々な問題解決手段を講じるので, そのことをシミュレートするのは難しい. しかし, 問題のない場合には, 上述のような行動パターンに近い行動をとるものと考えられる.

図 1 に, Markov 連鎖によるモデル化のための, 実際の Encarta の一部を例として示す. 図 1 は, アーティクル *Baseball* を目的として探すときの Encarta のモデルである. 図 1 左図が, 各状態から目的のアーティクルに向かう順方向の推移についてモデル化したものである. 直線上の数値は, このアーティクルと各見出しやリンクとの LSA による類似度の値である. 見出しへの推移の類似度については, 前述で説明した通り, カテゴリ

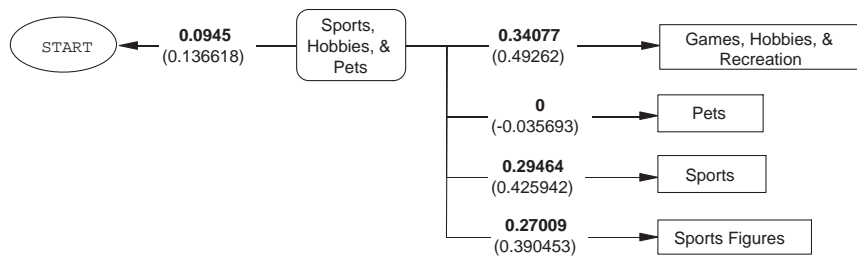


図 2: マルコフ連鎖 (推移確率例) .

りとそれに属する全トピックのにより計算される . つまり , $C+$ をカテゴリ C を表す見出し , $sim(A, X)$ を A というアーティクルと , X という見出しまたはリンクとの類似度とすると , アーティクル *Baseball* と見出し *Sports, Hobbies, Pets* との類似度は次のようになる . なお前述の通り , 類似度を求める際のアーティクルについては , Encarta での最初のパラグラフを用いる . アーティクル *Baseball* の Encarta での最初のパラグラフは *Baseball, competitive game of skill played with a hard ball and bat between two teams of nine players each. Baseball is often called the national pastime of the United States, because of its strong tradition and great popularity. It is played throughout the world by people of all ages.* となっている .

$$\begin{aligned}
 0.220824 &= sim(Baseball, \{(Sports, Hobbies, Pets)+\}) \\
 &= sim(Baseball, \{\{Sports, Hobbies, Pets\}, \\
 &\quad \{Game, Hobbies, \&Recreatrion\}, \{Pets\}, \{Sports\}, \{SportsFigures\}\}) \quad (1)
 \end{aligned}$$

図 1 右図が , 各状態から , 1 つ前の状態に戻る逆方向の推移についてモデル化したものである . 各見出し (カテゴリ) から Start への直線上の数値は , 見出しから Start へ戻る類似度を $sim((\text{全ての見出し}) \setminus \{\text{対応する見出し}\})$ として定義し , LSA で計算させた値である . 例えば *Baseball* の例では , $Back(A, C+)$ をアーティクル A における , 見出し $C+$ (カテゴリ C) から Start へ戻る推移の類似度とすると , 見出し *Sports, Hobbies, Pets* から Start へ戻る推移の類似度は次のようになる .

$$\begin{aligned}
 0.136618 &= Back(Baseball, \{(Sports, Hobbies, Pets)+\}) \\
 &= sim(Baseball, \{\{(Art, Language, \&Literature)+\}, \{(Geography)+\}, \\
 &\quad \dots, \{(Sports, Hobbies, Pets)+\}\} \setminus \{(Sports, Hobbies, Pets)+\}) \\
 &= sim(Baseball, \{\{(Art, Language, \&Literature)+\}, \{(Geography)+\}, \dots, \} \\
 &= sim(Baseball, \{\{Art, Language, \&Literature\}, \{Architecture\}, \{Artists\}, \\
 &\quad \dots, \{Geography\}, \dots\}) \quad (2)
 \end{aligned}$$

また , 各トピックからカテゴリへの線上にある括弧内の数値は , 「リンクの選択」後の状態から「見出しの選択」後 (再び「リンクの選択」) へ戻る推移の , Markov 連鎖での「確率」である . ユーザは「リンクの選択」後 , 目的のアーティクルが見つからなければ , 再度「リンクの選択」(正しいリンクを探している状態) に必ず戻るため , 確率 1 でその状態に戻る推移をする . もし , 選択したリンクが , 目的のアーティクルへの正しいルートであれば , ユーザは目的を達成したことになるので , 「リンクの選択」に戻ることがなく , その選択されたリンクが Markov 連鎖の吸収状態になる . 図 1 右図のリンク *Sports* がその状態である .

このように類似度をもとにして , 各推移にウェイトを付与することによって , ある状態から , 各状態への推移確率を設定することができる . ただし , 正しくない見出し下でリンクを選択する時には , 各リンクの類似度が全てある閾値 (ここでは δ' とした) 以下になるときは , リンクについて手がかりが不十分になり , それらのリンクが選択されることはないとして , 「リンクの選択」をすることなく , Start に確率 1 で戻る . また , 類似度の値が負になるものについては , 正しいルートとは連想が結びつかないものと判断されるので , ユーザはそれらのルートは選択しないと考え , 推移確率を 0 とする . 類似度の値から , 推移確率にする例を図 2 に示す . これは図 1 で示した類似度をもとに , 見出し *Sports, Hobbies, Pets* を選択した状態からの推移確率を表したものである . 線上の値が推移確率の値であり , その値の下にある括弧の中の値が類似度の値である .

4.2. Markov 連鎖による定式化

前節のようにモデル化することにより , Encarta のような階層構造をもつウェブサイトにおけるユーザのリンク選択行動は , Markov 連鎖で表すことができる . つまり , 各ページを 1 つの状態 , クリックを時刻パラメ

タして考えると、ウェブページのリンク推移過程は、ある 1 つのページに着目すると、そのページへたどりつくためには、直前にいるページが分かれば十分であることを考えれば、離散の状態空間と離散的時刻パラメタをもつ有限 Markov 連鎖とみなすことができる。すなわち、 X_n を n 回目のクリックで現れるページの状態とすると、以下の式が成り立つ。

$$P\{X_{n+1} = j | X_0 = i_0, X_1 = i_1, X_2 = i_2, \dots, X_n = i_n\} = P\{X_{n+1} = j | X_n = i_n\} \quad (3)$$

ここでいうクリックとは、ウェブページのリンクをたどるのに必要な選択決定クリックであり、その他の操作を目的とした、選択決定に関係ないクリックは考慮にいれない。この Markov 連鎖は、目的のアーティクルのタイトルが現れる、正しい「リンクの選択」後のページが唯一の吸収状態となる。このようにモデル化することによって、Start 状態であるホームページから出発して、目的のアーティクル（のタイトル）発見という吸収状態に吸収されるまでの、ユーザの行動を解析することができる。

4.3. 平均クリック数

非再帰の状態 i から始まる Markov 連鎖がどこかの再帰的同値類に吸収されるまでの平均ステップ数、つまり平均吸収ステップ数を $\mu^{(i)}$ とすると、これは、非再帰の状態の集合を \mathcal{T} とすれば、連立方程式

$$\mu^{(i)} = 1 + \sum_{k \in \mathcal{T}} P_{ik} \mu^{(k)} \quad (4)$$

を満たす。式 (4) を意味的に解釈すると、状態 i から他の状態に移るのに 1 ステップかかり (右辺第 1 項)、その推移した状態 k から吸収状態にステップするのに平均 $\mu^{(k)}$ ステップかかるのを、状態 i から推移可能なすべての非再帰の状態 k について足し合わせている (右辺第 2 項) ということである。

これを用いて、ユーザがホームページである Start 状態から目的のアーティクル発見という吸収状態に到達するまでに要する平均クリック数を求めることができる。

4.4. リンクの張替えによる効果

ここでは前述までの Encarta の Markov 連鎖のモデル化における解析と評価において、簡単なリンクの張替えによって、ユーザビリティに問題があるアーティクルが改善されるようなモデルを考える。これらのモデルを解析し、リンク張替えなしの既存の Encarta をモデル化したものとを比較することによって、ユーザが目的のアーティクルに到達するまでに要するクリック数の負担が、どれくらい軽減されるかを容易に評価することができる。3.2 でのユーザビリティ問題の結果から、Encarta は C-WS や T-WS に問題があるアーティクルはあまりないが、GSCH と GSCL に問題のあるものが多い。つまり、「正しい見出しやリンクの選択に関して問題がある」ものが多く、ユーザは正しくない見出しやリンクを選択してしまう可能性が多いということであり、そこで、その問題を改善させることが、Encarta 全体のユーザビリティを改善させるために重要なことであるといえる。そこでまず、ユーザが正しい見出しの選択までを行えたとき、正しいリンクの選択をしやすいうようにリンクを張替えることにより、ユーザビリティの改善を試みる。このリンク張替えによって、Encarta 全体のユーザビリティで最も問題が見られる GSCL が改善されることが見込める。

リンク張替え 1: C-ws, T-ws とともに該当しない場合、全ての正しい見出し下の正しくないリンクで、目的のアーティクルとの類似度の値が最大となるものを正しいリンクとして追加する。

次に、ユーザは目的のアーティクルを探す際、まずは見出しを選択し、その後、その見出し下に現れるリンクを選択するという操作を行くことを考え、ユーザの選択行動が最も効率が良くなるように、目的のアーティクルとの類似度の値が最大となる見出しと、その見出し中の、同様に類似度の値が最大となるリンクを正しいリンクとして追加することにより、ユーザビリティの改善を試みる。この張替えによって GSCH がなくなり、また T-ws と GSCL の改善が期待される。

リンク張替え 2: C-ws に該当しないとき、目的のアーティクルとの類似度の値が最大となる見出しの中の、同じように類似度の値が最大となるリンクを、正しいリンクとして追加する。ただし、追加する見出しとリンクについては、その見出し下の全てのリンクの類似度の値が、閾値 δ' 以下にならないことが必要である。

さらに、これら 2 通りのリンク張替えを同時に行うことにより、Encarta 全体のユーザビリティの大幅な改善を試みる。

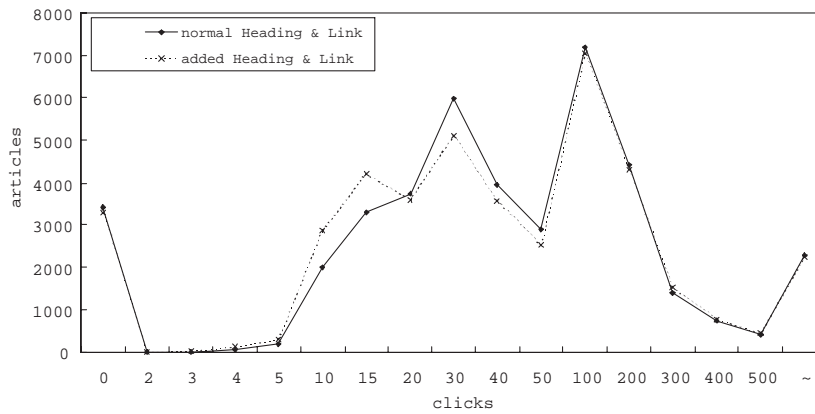


図 3: 既存の見出しとリンクのものと追加した見出しとリンクのもののクリック数の比較 .

リンク張替え 3: リンク張替え 1 と 2 を同時に実行する . 条件によりいずれか一方の張替えしかできないときは , 張替え可能なものだけ行う .

これらのリンク張替えによってユーザビリティの問題が改善され , ユーザが目的のアーティクルに到達しやすくなるのが期待される . そこで , これらを Markov 連鎖で解析し , 評価を行いリンク張替えの効果を確認するとともに , Markov 連鎖による解析と評価の容易さも確認する .

5. 解析・評価

以上のことを踏まえ , Encarta を Markov 連鎖でモデル化し , ユーザが目的のアーティクルに到達するまでに要する平均クリック数を解析した結果を図 3 に示す . 図 3 は , Encarta のウェブサイトの前章で説明した方法によりモデル化し , ホームページ (Start 状態) から目的のアーティクルを発見するまでの平均クリック数を各アーティクルについて , 既存のウェブサイトの見出しとリンクのものと , カテゴリとトピックについて LSA により類似した語を追加した見出しとリンクのものを評価し , 比較したものである . LSA による類似した単語を探し出し追加することは , ユーザが見出しやリンクに使用されている言葉を認識した時 , 意味的に近い言葉を想起するというプロセスをシミュレートする狙いがある . 横軸に平均クリック数 , 縦軸にアーティクル数をとってある . 以降の図も , 特にことわりがなければ軸のとり方は同じである . なお前に説明した , 目的のアーティクルに到達することができない可能性があるものは , 計算不可なアーティクルとし , クリック数を便宜上 0 としている . これらのグラフから , 半数以上のアーティクルは , 目的のアーティクルに達成するまでに 100 クリック以内に到達しているのがわかる . しかし , 最短での到達クリック数が 2 クリックであり , 実際にウェブサイトでの操作を考えると , 数クリック ~ 数十クリックが適切と考えられる . このようにクリック数が多くなってしまふのは , Markov 連鎖によるモデルが , リンクの問題がある場合のユーザ行動を , 忠実にモデル化できていないためであると考えられる . しかしそれは , Encarta に問題があることを表しているので , ユーザビリティ発見の妨げになるものではない .

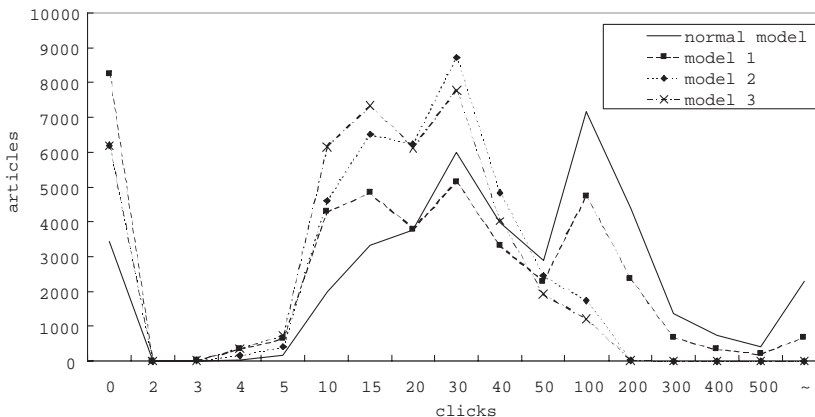


図 4: 既存の見出しとリンク張替えの各モデルによるクリック数の比較 .

	計算可	計算不可	計算不必要
リンク張替え 1	455	390	2577
リンク張替え 2	1268	123	2031
リンク張替え 3	1270	123	2029

表 3: リンク張替え前に計算不可だったアーティクルの張替え後の効果

図 4 は、Markov 連鎖による、既存の見出しとリンクのもと、前章で説明した各リンク張替えのモデルとの目的のアーティクルに到達するまでに要する平均クリック数の比較である。いずれの張替えによっても、数十～数百クリックを要していたアーティクルの数が減少し、二、三十クリックまでで目的のアーティクルに到達できるアーティクルが多くなっている。このことは、特にリンク張替え 2 と 3 で大きな改善が見られる。リンク張替えによってクリック数が 0 となるものが増えているが、これは、前章での説明の通り、リンクを張替えるための条件があるためであり、リンクを張替えられないもの（計算不必要なもの）が出てくるためである。また 3.2 のユーザビリティの結果から、34080 個のアーティクルがリンク張替え 1 の、そして 35902 個のアーティクルがリンク張替え 2 の、34080 個のアーティクルがリンク張替え 3 の計算対象候補である。ちなみに、リンク張替え前のもので、既存の見出しとリンクのものでは、3422 個のアーティクルが正解まで到達しない可能性のあるものとなっており、それらのアーティクルが、リンク張替え後に計算可となっているのか、もしくは正解までに到達しない可能性がなくなり計算不可のままとなっているのか、ユーザビリティの結果からリンクを張替えられなく計算不必要の状態になっているのかは、表 3 の通りであり、特にリンク張替え 2 もしくは 3 の効果によって、計算が可能になっているものが増えていることがわかる。

また、図 5 に、リンクを張替える前に計算不可だったアーティクルが、リンク張替え後に計算可になったものについて、それらがどれくらいのクリック数で正解までに到達できるかの結果を示す。

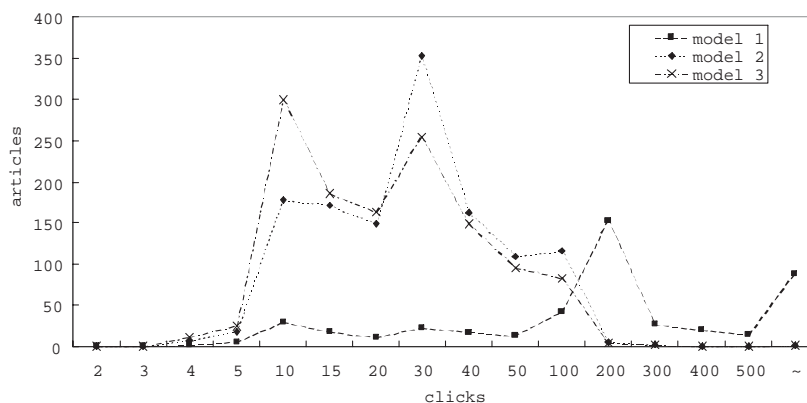


図 5: リンクを張替える前に計算不可だったアーティクルのリンク張替え後の効果。

これらの図より、リンク張替えによって、目的のアーティクルに到達するまでに要する平均クリック数が全体的に、大幅に改善されていることがわかる。どの張替えでも、100 ステップ以上かかるアーティクルが減少することが確認され、リンク張替え 2、リンク張替え 3 にいたっては、100 クリック以上かかるアーティクルがほぼなくなっている。さらにどの張替えによっても、20 クリック以下で目的のアーティクルに到達できるものが増えており、張替え 1 と 3 にいたっては 5 クリック以下で目的のアーティクルに到達できるものが大幅に増えていることがわかる。また、リンクを張替えることによって、目的のアーティクルに到達しない可能性があったものが、到達できるように改善されたアーティクルが多いことも確認される。どの張替えでも数十クリックまでで到達するアーティクルが増えているのがわかるが、リンク張替え 1 では特に、数クリックで目的に達するアーティクルが大幅に増え、リンク張替え 2 では特に、数百クリックかかるアーティクルが大幅に減らすことができ、ほとんどのアーティクルを数十ステップまでで目的のアーティクルに到達するように効果があることがわかる。リンク張替え 3 は、張替え 1 と 2 を同時に行う（条件によってはいずれか一方のみ）ので、どちらの効果も現れることがわかる。

6. まとめと今後の課題

今回の研究によって、認知工学のウェブ認知ウォークスルー (CWW) を用いて、実際に存在する大規模な情報をもつウェブサイトのユーザビリティを評価できることを示した。さらに、Markov 連鎖でモデル化することによって、ユーザが目的に到達するまでに要する平均クリック数を簡単に解析できることを示し、またユーザビリティの問題を減らすようにリンクを張替えることにより、目的に至るまでに要するクリック数が大幅に減少するという事も確認できた。しかし、今回の Markov 連鎖によるモデルは簡易的にモデル化し評価できることを示しただけであり、実際のユーザの行動を厳密に表したものではない。しかしながら、実際のユーザが、目的になかなか到達できないときに途中で探索を諦めたり、同じページでしばらくの間考えて操作を行ったりすることがあるが、こういった場合にはユーザビリティの問題があり、Markov 連鎖によるモデル化の評価では、推移過程は、見出しとリンクを何度も往来するというパターンになる。正しいリンクへの推移確率は小さいので、目的に至るまでに要する平均クリック数は多くなってしまう。このように Markov 連鎖では、ユーザビリティに問題のある場合の、ユーザ行動の振る舞いは忠実にモデル化できないが、ユーザビリティ問題を発見するという目的のためには、十分な性能をもったモデル化となっている。また見出しやリンクを理想のものにし、各アートをクラスタリングすることによって、理想的なカテゴリ化をすることも考えられる。今後はこのような分析手法に基づいたウェブサイトデザイン法を確立することにより、既存のウェブサイトや、さらには作成途中のウェブサイトなどを理想的な見出しやリンク構造に改善できる手段を構築していきたい。

参考文献

- [1] Spool, J. M., Scanlon, T., Schroeder, W., Snyder, C. and DeAngelo, T.: *Web Site Usability: A Designer's Guide*, Morgan Kaufmann, San Francisco, 1999. 篠原稔和 (監訳): *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) .
- [2] Ojakaar, E. and Spool, J. M.: Getting Them to What They Want, *UIE Reports: Best Practices Series*, 2001.
- [3] Mayes, T. J., Draper, S. W., McGregor, A. M. and Oatley, K.: Information flow in a user interface: The effect of experience and context on the recall of MacWrite screens, *Proc. HCI'88 Conference on People and Computers IV*, Cambridge, pp.275-289, Cambridge University Press, 1988.
- [4] Wharton, C., Rieman, J., Lewis, C. and Polso, P.: The cognitive walkthrough method: A practitioners's guide, *Usability inspection methods*, Nielsen, J. and Mack, R. L. (Eds.), New York, pp.105-140, John Wiley and Sons, 1994.
- [5] 北島宗雄: ウェブ認知ウォークスルーによるウェブサイトユーザビリティの評価, *ファジィ学会誌*, Vol.14, pp.446-460, 2002.
- [6] Kintsch, W.: *Comprehension: A Paradigm for Cognition*, Cambridge University Press, Cambridge, 1998.
- [7] Kitajima M. and Polson, P. G.: A comprehension-based model of correct performance and errors in skilled display-based, human-computer interaction, *International Journal of Human-Computer Studies*, 43, pp.65-99, 1995.
- [8] 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, 211-240, 1997.
- [9] 北島宗雄, 狩谷典之: 大規模情報発信サイトのウェブ認知ウォークスルーによるユーザビリティ評価, *電気学会研究会資料, システム・制御研究会 SC-03-8 ~ 14*, 9-14, 2003.
- [10] 北島宗雄, 高木英明, 山本哲生, 張勇兵: 潜在意味解析 (LSA) を利用した Markov 連鎖モデルによる階層メニュー探索過程の評価, *情報処理学会論文誌*, Vol.43, No.12, pp.3722-3732, 2002-12.
- [11] Daniel A. Menascé, Virgilio A. F. Almeida, Rodrigo Fonseca, and Marco A. Mendes: Business-oriented resource management policies for e-commerce servers, *Performance Evaluation*, Vol.42, pp.223-239, 2000.
- [12] Daniel A. Menascé and Virgilio A. F. Almeida: *Scaling for E-Business: Technologies, Models, Performance, and Capacity Planning*, Prentice-Hall, 2000.
- [13] Harold Thimbleby, Paul Cairns, and Matt Jones: Usability analysis with Markov models, *ACM Transactions on Computer-Human Interactions*, Vol.8, No.2, pp.99-132, June 2001.
- [14] Stuart K. Card, Thomas P. Moran, and Allen Newell: *The Psychology of Human-Computer Interaction*, Lawrence Erlbaum Associates, 1983.