論文・事例研究

HTPモデルを用いたWebサイト訪問者の興味把握

本吉　夏樹，朝日　弓未，山口　俊和

1. はじめに

2000年から2007年の調査によると，電子商取引（企業間取引，企業と一般消費者間取引）の市場規模は拡大傾向にある[6]．電子商取引とは，インターネット等で契約や決済を行う取引のことである．主だったものにインターネット上で製品を販売するオンラインショップや，ビジネス用のソフトウェアをインターネットを通じて顧客にレンタルするASP（Application Service Provider）が挙げられる．市場規模拡大の理由には，インターネット利用者数増加による普及率上昇，インターネット通信速度の高速化がある[8]．

インターネットの爆発的普及に伴い，企業にとってWebサイト（以降，サイト）は顧客との関係構築をもたらす重要チャネルとなった[9]．電子商取引市場では，サイトの目的として「販売促進」，「集客」，「サポート」等がある[4]．これらの目的は，自社サイトの訪問者をサイトの目的ページに誘導することに集約される．目的ページは，製品の申込後に表示される「送信完了」ページや，顧客サポートのための「問合せ」ページ等，サイトの目的によって様々である．「サイト訪問者が，サイト設計者の設定した目的ページに到達すること（コンバージョン）」は，製品のブログやコミュニティサイトを立ち上げるだけでは増やすことはできない[9]．インターネットには距離的な制約が無く，あらゆる人がサイトへアクセスできる．また，欲しい情報の入手も容易であり，インターネット利用者が欲する製品や情報は実社会以上に多様である[9]．サイト訪問者（以降，訪問者）をコンバージョンさせるには，サイト改良の手段に終始するのではなく，訪問者が何に興味を持っているか（興味対象）を真摯に把握することが求められる[9]．

サイトでは，サーバ等から訪問者全員が残したサイト閲覧行動の軌跡であるアクセスログを取得できる．アクセスログからは「誰が，どのページからどのページへ移動したか」という顧客ベースのデータを取得できるため，サイトにおける訪問者の興味対象を把握することができる[9]．

サイトに関する研究は主に3種類に分類される[3][12]．

(1) サイトを構成するデータ（テキスト，デザイン，サウンド等）を対象とした知見の獲得（Web Content Mining）
(2) サイトに設置されたリンクの構造を基にページ間の関連性を発見する等の技術（Web Structure Mining）
(3) アクセスログやクリック履歴に基づくサイト訪問者の行動分析（Web Usage Mining）

本研究は，(3)のWeb Usage Miningに位置づけられる．Web Usage Miningには，訪問者のページ移動過程をマルコフ連鎖と捉えマルコフモデルを適用した研究や，訪問者のページ移動パターンを抽出した研究が多い．狩谷ら[5]と山本ら[13]は，マルコフモデルによってサイト内の各ページから目的ページまでの平均クリック数や平均到達時間を求めている．マルコフモデルにより，目的ページに訪問者を誘導しやすい関連の強いページを把握できる．ただ，目的ページとの関連の強さに関わらず，訪問者が強い興味を抱くページが存在する可能性がある．マルコフモデルでは，目的ページを閲覧する直前にアクセスされたページほど目的ページとの関連が強く，目的ページ閲覧前の早い段階でアクセスされたページほど目的ページとの関連が弱くなりやすい．そのため，サイトを訪れた早い段階で訪問者が最も興味を持っているページを閲覧す

もとよし　なつき
東京理科大学 大学院工学研究科
あさひ　ゆみ，やまぐち　としかず
東京理科大学 工学部経営工学科
〒102-0073 千代田区九段北1-14-6
受付08.4.30　採択09.3.24

る場合には，訪問者の興味を正確に把握できない可能性がある．

訪問者の真の興味対象をつかむためには，目的ページとの直接的な関係に加え，サイトを訪れてからコンバージョンに到るプロセスの把握が重要である[4]．このため，訪問者の興味対象把握には，訪問者のサイト内における全体的な移動の流れ（移動ルート）を把握する必要がある．マルコフモデルでは，全ページ間の移動確率（遷移率）を計算し，ページを「点」，遷移率を「矢印」として，訪問者のページ移動過程を状態遷移図で把握することもできる．しかし，規模の大きい（ページ数の多い）サイトでは図が煩雑となるため，ページをいくつかのグループに分類して訪問者の移動を視覚化する．そのため，グループ化される前のどのページに対して訪問者が興味を抱いているかという詳細な興味対象の把握が困難になる．

AgrawalとSrikantは，サイト内におけるシーケンスパターンを発見するアルゴリズムを提案している[1][2]．シーケンスパターンとは，どのような順序でページを閲覧したかという移動パターンである．AgrawalとSrikantは，Aprioriという相関規則発見アルゴリズムを用い，アクセスログ内に指定回数以上出現したシーケンスパターンを発見している．ただ，発見されるパターンの長さ（連続したページ列）は短く，パターン同士の関係も分からない．そのため，訪問者行動の全体的な特徴をつかめないという問題点が存在した．長いパターンを抽出し訪問者の全体的な傾向を把握するため，小柳ら[7]はMatrix Clusteringという手法を提案している．これは，Aprioriで発見されるパターン同士を適切に関係づけたスーパシーケンスを生成し，訪問者行動の全体的な特徴が把握可能な手法である．ただ，複雑な計算を避けるためスーパシーケンスは重複したページやループを含まないものとされ，実際の訪問者行動と乖離する可能性がある．

以上の先行研究を踏まえ，訪問者の興味対象把握のためには，サイト内における訪問者の興味ページを抽出しつつ移動ルートを把握することが求められる．また，移動ルートはサイト内での訪問者の全体的な移動の流れをありのまま記述する必要がある．以降，興味ページとは訪問者が強い関心を示すページと定義する．

本研究では，アクセスログを用いて訪問者の興味ページによる移動ルートを抽出できる分析手法の提案を目的とする．

2. データの概要

本研究では，アクセスログ解析ツール「シビラ」を販売するK社サイトのアクセスログを用いる．「シビラ」は企業向け製品であり，サイトのPV（Page View）は，企業の業務時間帯に当たる平日の午前9時から午後8時が全体の約70%を占め，その多くが企業からの業務目的と考えられる．PVとは各ページが閲覧された回数である．また，サイトの総ページ数は100である．

〈使用データ〉提供：平成18年度データ解析コンペティション

- データ期間：平成18年1月1日～6月30日
- データ件数：56,353件
- データ内容：時間のデータ：年，月，日，曜日，時，分，秒
  ：顧客のデータ：ユーザID，セッションID
  ：サイトのデータ：リクエストURL，リファラURL

ユーザIDは，Cookieを用いた個々のブラウザ（Internet Explorer等）を識別するIDである．Cookieとは，ブラウザを通して訪問者のコンピュータにデータを書き込み，訪問者を識別する仕組みである．セッションIDは，ブラウザの起動ごとにふられるIDであり，同じ訪問者（同じユーザID）でも，ブラウザを起動するごとにセッションIDは異なる．セッションIDが等しいページ移動の集合は訪問者の一連の移動とされる．リクエストURLは，移動先ページを表し，リファラURLは，移動元ページを表す．例えば，ある訪問者が「トップページ」から「価格表」へ移動した場合，リクエストURLは「価格表」，リファラURLは「トップページ」となる．

3. 分析

3.1 分析準備

3.1.1 分析対象

アクセスログでは，次のような訪問者の行動を把握できる．

(1) 他サイトからサイトに入ってくる訪問者
(2) サイト内からサイト内に移動する訪問者
(3) サイト内から他サイトに移動する（離脱する）訪問者

(1)の訪問者からは「どの検索サイトや他サイトから

訪れたのか」,「どのような検索キーワードでサイトを見つけたのか」という情報が分かる．(1)の訪問者を分析することで，サイトに多くの訪問者を集める手がかりを見つけることができる．(2)の訪問者からは，訪問者のサイト内における移動経路の把握や，各ページにおける滞在時間といった情報を得ることができる．(3)の訪問者からは，離脱した先のページは把握できないが，サイト内のどのページで閲覧を終えたかを把握できる．(2)と(3)の訪問者行動を分析することで，訪問者が興味を持っているページ，何度も閲覧されるページ等を把握できる．

アクセスログでは(1)～(3)の訪問者を分析できるが，どれほど多くの訪問者をサイトに集めても，訪問者がほとんどページを閲覧せずに離脱するのでは意味がない．そこで本研究では，(2)と(3)の訪問者を中心に分析した．

本研究では(2)と(3)の訪問者を分析する前に，訪問者がサイト内で閲覧を開始するページ（他サイトからの入口ページ）を特定した．まず，サイトを訪れて1ページのみ閲覧して離脱した訪問者（1ページ離脱者）のデータを全データから除き，データ件数を41,800件とした．1ページ離脱者のデータからはサイト内における移動ルートを把握できないため，本研究では対象外とした．次に，リファラURLがシビラ以外のサイトで，リクエストURLがシビラサイトのデータに着目し，他サイトからのPVが最も多いページを入口ページとした．他サイトからのPVはトップページが約62%を占め，次にPVが多いキャンペーンページ（約16%）と比べても圧倒的に多かった．このため，ほとんどの訪問者はサイトのトップページを入口として訪れ，サイト内の移動を始めるといえる．

入口ページの把握後，他サイトからシビラサイトにアクセスしたデータを除き，分析対象とするデータ件数を32,471件とした．

3.1.2 遷移率の算出方法

遷移率とは，リファラURLを固定した上で，あるリクエストURLへ移動する訪問者の割合である．ページiからページjへの遷移率t_{ij}は，(1)式で表される．訪問者がナビゲーション等から同一ページにアクセスする可能性もあるため(1)式は$i=j$の場合にも成り立つ．本研究では$n=100$である．

$$t_{ij}=(x_{ij}/X_i)*100 \quad (1)$$

x_{ij}：ページiからページjへの移動件数

X_i：ページiを移動元とする全移動件数

$i,j=1,2,\cdots,n$

例えば「トップページ」から「シビラの特徴」へ移動しているデータが2つ，「トップページ」から「価格表」へ移動しているデータが3つあったとすると，前者の遷移率は(2/5)＊100，後者の遷移率は(3/5)＊100となる．

3.2 分析方法

本研究では，訪問者の興味ページはその他のページよりも閲覧される可能性が高いと仮定し分析方法の提案を行った．閲覧されやすいページとは，他ページから高い遷移率で訪問者を集めるページである．そのため，興味ページによる移動ルートは，各ページから最も高い遷移率（MTP: Maximum Transition Probability）を用いてページをつなぐことで表現した．ただし，ページ間にMTPによるループができる可能性がある．MTPループが出現すると，それ以上ページをつなげないため，移動ルートの抽出が終了する．これは訪問者の全体的な移動傾向を把握するという本研究のコンセプトに反する．そこで，MTPのみを用いるのではなく，必要であれば高い遷移率から順に移動ルートを作成する方法を提案した．以降，ページLからN番目に高い遷移率を$HTP_{L,N}$（High Transition Probability of order N）と表す．また，HTP_1はMTPと同義である．HTP_Nを上位から用いることで，ループができても移動ルートの作成は制限されない．そして，入口ページを移動ルートの始点とし，目的ページを終点とした．本研究では目的ページを，サイト設計者が訪問者を特に誘導させたいページと定義する．目的ページはサイトの目的次第で複数存在することもあるが，本研究では分析者が目的ページを一つ選び移動ルートを作成する方法を提案した．この方法により，訪問者がサイトを訪れてから目的ページに到るまでの大まかな移動ルートと，その過程に現れる興味ページの把握を行った．

以降では，本研究で提案した方法によって抽出する移動ルートをHTPモデルと呼び，移動ルートを視覚的に表現するためにページを「点」，ページ間移動を「矢印」とした．

3.2.1 同一遷移率存在時の対処方法

あるページiから次のページに矢印をつなぐ際，iからの同一な遷移率が複数存在する場合がある．例えば，「トップページ」から「シビラの特徴」と「価格表」双方への遷移率が等しい場合が挙げられる．その際，ページiから矢印をつなぐページの選択基準を設

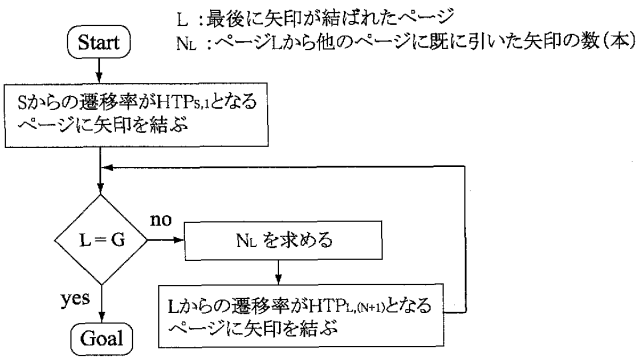

図1　HTP モデル作成フローチャート

ける必要があった．本研究の分析コンセプトは「訪問者の興味ページによる移動ルート」を表現することであるが，興味の度合いが同じなら，より多くの訪問者が移動を行っているページを抽出する方が，訪問者行動の把握において有益と考えた．そこで，もしページ i からの遷移率が同じページ j，k がある場合，j と k の PV を比較し，その多い方に矢印をつなぐ．

3.2.2　HTP モデル作成 step

以下の step で訪問者の移動ルート（HTP モデル）を作成する．

Step 1　全ページにページ番号を付け，t_{ij} を全組合せ求める．また，各ページにおいて t_{ij} の高いものから降順に順位付けする．

Step 2　他サイトからの PV から入口ページ，サイトの目的から目的ページを決める．また，S：入口ページ（始点）G：目的ページ（終点）とする．

Step 3　図1に示したフローチャートを実行する．

3.3　分析結果・考察

3.3.1　HTP モデル

分析対象データに3.2.2のモデル作成 step を適用し，作成されたモデルを図2に示した．図2のモデルは，「トップページ(1)」を始点，申込の「送信完了(76)」を終点とした．

ページに矢印が結ばれた順序は表1に示した．まず，始点の(1)から $HTP_{1,1}$ を用い(13)へ矢印をつなぐ．そして，新たに抽出された(13)から他ページへの矢印が出ていない（N＝0）ので $HTP_{13,1}$ を用い(14)へ矢印をつなぐ．以降，step を繰り返し終点(76)に矢印がつながれたらモデル作成が完了する．

HTP モデルはモデル内のすべての点を一筆書きで結んだものである．HTP モデルでは高順位の遷移率を使用するため，多数の訪問者のページ選択傾向を表現できる．

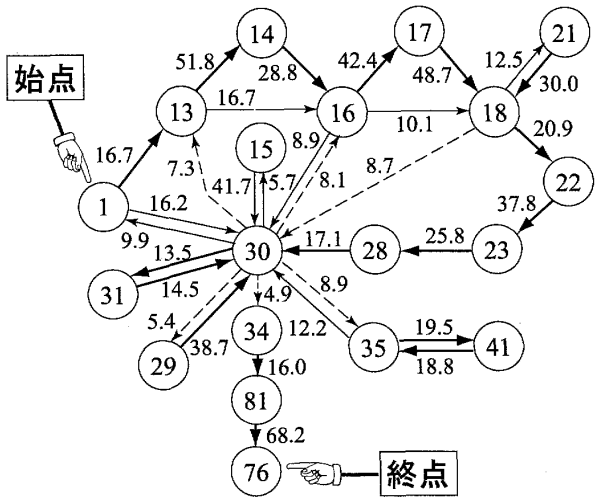

図2　HTP モデル

表1　ページ連結順序

順序	ページ番号	順序	ページ番号
1	1	18	30
2	13	19	16
3	14	20	18
4	16	21	21
5	17	22	18
6	18	23	30
7	22	24	13
8	23	25	16
9	28	26	30
10	30	27	15
11	31	28	30
12	30	29	29
13	1	30	30
14	30	31	34
15	35	32	81
16	41	33	76
17	35		

図2の矢印上の数値は遷移率（%）である．また，各点に対応するページ内容と矢印数を表2に示した．シビラサイトの左端には，リンク（他ページへ移動する場所）が縦に並べて設置されている．この設置順番を基にサイトの全ページに1～100の番号を付けた．表2の番号は，モデルに抽出された19個を示した．矢印数とは，点に出入りする矢印の数の合計である．矢印数が多いほど，多くの他ページから高確率で訪問者を獲得していることを表す．そのため矢印数の多い点は，訪問者の関心が高く選択されやすいページ，もしくは迷ってたどりつきやすいページと考えた．ただしシビラサイトには，全ページの左端にほとんどのペ

表2　ページ内容と矢印数

番号	ページ内容	矢印数(本)
1	トップページ	3
13	シビラとは	4
14	シビラの特徴	2
15	ページビュー数とは	2
16	サイト解析	6
17	ページ解析	2
18	経路解析	6
21	逆引き経路解析	2
22	その他の機能	2
23	広告測定効果	2
28	SEO効果測定	2
29	申込の流れ	2
30	価格表	16
31	シビラ導入事例	2
34	申込	2
35	解析レポート	4
41	コンサルティング	2
76	送信完了	1
81	入力内容の確認	2

ージへ移動できるサブメニューがあり，整理された設計になっている．そのため本研究では訪問者が迷うことは考えず，矢印数の多い点を訪問者の関心が高いページとした．

矢印の種類については図2の下部に示した．HTP_1のみを用いると，図2の「解析レポート(35)」と「コンサルティング(41)」のようにループができてしまう．そこでHTP_2以降の遷移率を使うことで，モデル内にループができても終点まで移動ルートを作成できる．モデル内の点はまず，始点から最も移動しやすい点が抽出される．次に，抽出された点から最も移動しやすい点が抽出され，連鎖的に点が増える．このため，始点から終点までの移動プロセスに入らない点は，PVが多くても抽出されない．例えば，シビラサイトの「会社概要」ページは，モデル内の「申込(34)」や「コンサルティング(41)」よりもPVが多い．「会社概要」は製品購買とは関係が弱く，就職活動等の別の目的でアクセスされることがモデルに抽出されなかった理由と考えた．

モデルに点が抽出された順番は，始点から高い確率を追うことで把握できる．図2より，まず訪問者は「シビラの説明 (13, 14)」，「機能 (16, 17, 18, 21, 22, 23, 28)」，「価格表(30)」の順に流れている．これは，シビラのリンク配置順通りの移動である．その後，「機能」，「申込の流れ(29)」，「事例(31)」，「オプションサービス (解析レポート(35), コンサルティング(41)」を「価格表(30)」とともに閲覧し終点へ移動する．つまり，「価格表(30)」までは興味というよりもリンクに依存した移動であり，その後は「価格表(30)」を常に確認しながら終点へ向かっている．その結果，「価格表(30)」の矢印数は他ページに比べてダントツに多くなった（表2参照）．最も多くのページからアクセスされやすい「価格表(30)」は，訪問者が最も関心を示すページといえる．この結果からサイトへの実務的提案を行うとすれば，「価格表(30)」ページ内にサイト設計者が強調したいページへのリンクを配置することが挙げられる．このサイトの訪問者の多くは「価格表(30)」を訪れる可能性が高いため，リンク設置により多くの訪問者を強調したいページに誘導しやすくできる可能性がある．特に，訪問者の入口ページが多数存在するサイトの場合では，以上のように多数の訪問者が共通して閲覧するページを把握することで訪問者のページ誘導が行いやすくなるといえる．

次に，「申込(34)」へ結ばれた矢印に着目した．図2では，「価格表(30)」から「申込(34)」へ移動しており，「価格表(30)」によって訪問者が最終的な意思決定をしているといえる．実際に，訪問者が「申込(34)」へ移動する直前ページとして「価格表(30)」が最多であることはデータの集計から確認した．

「機能（16～28)」に着目すると，「サイト解析(16)」と「経路解析(18)」の矢印数が比較的多いことが分かる．このため，シビラが持つ機能の中では，アクセスされやすい「サイト解析(16)」と「経路解析(18)」が訪問者の主な関心であると考えた．

3.4　訪問回数別HTPモデル

3.4.1　訪問者の製品購買プロセスと訪問回数

図2のモデルでは，他サイトからのアクセスを除いた全データを使用した．これにより，サイト全訪問者という大きな視野の移動傾向を一つのパターンとして把握できた．ただ，訪問者の製品購買行動には段階があり，それによって訪問者のサイト閲覧行動は変化する[10][11]．そのため，訪問者の購買プロセスに関連した傾向を捉えることで，サイトの現状把握に生かせると考えた．購買プロセスの初期段階としては，様々なサイトから積極的に情報を集めようと試みる「情報収集」がある．そして訪問者の評価基準で製品を比較検討し，購買を決定する．製品に対する満足度が大きければ，購買後に再び購買を行うこともある．

本研究では，購買決定に近い段階と情報探索のような初期段階では，サイト閲覧行動が異なると考えた．そして購買決定段階の訪問者を分析すれば，訪問者が製品購買を決定した理由を把握できると考えた．購買

表3　訪問回数別 PV

訪問回数（回）	PV	（%）
1	22883	70.47
2	4754	14.64
3	1988	6.12
4	1199	3.69
5	626	1.93
6	388	1.19
7	255	0.79
8	152	0.47
9	88	0.27
10	77	0.24
11	21	0.06
12	12	0.04
13	4	0.01
14	6	0.02
15	2	0.01
16	2	0.01
17	1	0.00
18	3	0.01
19	5	0.02
20	5	0.02
総計	32471	100.00

決定とそれ以外の段階を見いだす指標として，本研究では訪問者がサイトを訪れた回数（訪問回数）と訪問者の全移動件数中コンバージョンした件数の割合（コンバージョン率）を用いた．2つの指標を用いたのは，何度も訪問する人はサイトへの強い関心を持っており，訪問回数によってコンバージョンしやすさが異なると考えたためである．本研究のデータにおける訪問回数別の PV を表3に示した．また訪問回数別のコンバージョン率を表4に示した．

表3より，初回訪問者の PV は，全 PV の約70%を占めることが分かる．つまり本研究のデータは，初回訪問者のページ移動が大半である．そのため，図2のモデルには，初回訪問者の行動が強く表れていたといえる．

表4では，訪問回数4～20回目を一区分として扱った．その理由は，訪問回数4～20回目ではデータ数が少なくコンバージョン件数が0件や1件と極端に少なかったためである．表4より，初回訪問時にはコンバージョン傾向が弱く，情報探索の段階に当たると考えた．そして，2回目・3回目の訪問で購買を決定する傾向が見られる．4回目以降のコンバージョン率が低いのは，製品の再購買がされにくい状況を表すと考えた．実際，全データ中コンバージョンした訪問者は58人であり，そのうち再購買を行ったのは1人（約1.7%）と少なかった．

表4から，訪問回数によって購買決定とそれ以外の段階を見いだすことができた．次に，訪問回数別の行動傾向の違いを把握するため，訪問回数別に HTP モデルを作成し，分析結果を表5に示した．

表5のモデル内ページ数は，モデルに抽出されたページ数の合計を示す．また，モデル内矢印数は，モデルに抽出された矢印の数の合計を示す．モデル内ページ数やモデル内矢印数が多い場合，多くのページや矢印を抽出しなければ終点にたどりつかなかったことを意味する．つまり，すぐに終点に向かう傾向が弱い，もしくは多くのページを網羅的に閲覧する訪問者である．逆にモデル内ページ数・矢印数が少ない場合，少ないページや少ない移動で終点にたどりつくコンバージョン傾向が強い訪問者といえる．

表4　訪問回数別コンバージョン率

訪問回数（回）	1	2	3	4～20
コンバージョン率（%）	0.10	0.42	0.40	0.08

表5　訪問回数別 HTP モデル

訪問回数（回）	1	2	3	4～20
モデル内ページ数	35	22	15	35
モデル内矢印数	132	43	26	87
モデル説明力 E_i	0.67	0.37	0.25	0.51
コンバージョン率（%）	0.10	0.42	0.40	0.08

モデル説明力 E_i は(2)式のように定義した．

$$E_i=(v_i/w_i) \tag{2}$$

v_i：訪問回数 i におけるモデル内の移動件数

w_i：訪問回数 i における全移動件数

i：訪問回数（回）　$i=1, 2, 3, 4\sim20$

E_i は，訪問回数 i のデータ中，モデルが説明する移動の割合である．E_i が1に近いほど，抽出された移動の全体に対する割合が大きいことを表す．E_i は抽出された移動傾向の代表性を表す．例えば，抽出ページ・矢印数が等しい訪問者でも，E_i が大きいグループの方がより代表的な移動傾向となっていることを示す．

表5のコンバージョン率以外の3指標は性質が類似しているが，一つが増えれば残りの指標も増えるという単純な関係ではない．モデル内ページ数が多い場合

でも，矢印数が少ないこともある．この場合，少ない矢印で多くのページを閲覧するため，訪問者は一様なルートを移動していると解釈できる．逆にページ数が少なく矢印数が多い場合は，訪問者は少数のページをあらゆる組合せのパターンで移動している（多様な移動傾向）と解釈できる．また，ページ数と矢印数にE_iの大小を加えた組合せでさらに複数の解釈が存在する．

以降では，訪問回数i回目のモデルを$Vi(Visit\ i)$モデルと表す．表5より，$V1$モデルではコンバージョン率が0.1%と低く，抽出ページ・矢印数・E_iが他モデルと比べて大きくなった．つまり，目的ページ以外が閲覧される割合が大きい．そのため，モデルは終点に向かわず多くのページ移動を抽出した．表5を見ると，$V1$と$V4$～20モデルは終点に向かい難い傾向が現れたといえる．ただ，$V1$と$V4$～20モデルのページ数は等しいが，矢印数とE_iには差がある．$V1$モデルは$V4$～20モデルよりも1ページ当りに接続される矢印数が多い．つまり，前者は後者より，同じページを複数回閲覧する傾向があるといえる．また，E_iの増加に関しては矢印数の増加に伴うものと考えた．以上より，$V1$モデルは他モデルと比べて訪問者が最も情報を求めている段階と考えた．コンバージョン率による解釈と同様に，$V1$モデルは購買決定以前の情報探索の段階に当たると考えた．

$V2$と$V3$モデルでは，コンバージョン率が約0.4%と全モデル中で高い訪問者群であり，$V1$と$V4$～20モデルより抽出されるページ移動が少なくなった．つまり，$V2$と$V3$モデルにはコンバージョンに強く向かう傾向が現れたといえる．ただ，コンバージョン率を見ると多少$V2$モデルの方が高い．それに対してモデル内のページ数・矢印数・E_iは$V3$モデルが最も小さい．直感的にはコンバージョン率が高いほどモデルの規模が小さくなるはずである．$V3$モデルは，モデル上に関していえば$V2$モデルより強いコンバージョン傾向が現れた．$V2$モデルでは，モデルに抽出されていないページ移動に強いコンバージョン傾向が存在すると解釈できる．

つまり，$V2$モデルでは$V1$モデルの情報収集傾向が残っており，それが全体的な傾向であるHTPモデルに反映されたと考えた．以上より，$V2$モデルは強いコンバージョン傾向を持つ訪問者群と，依然として情報を探索している訪問者群が混在すると考えた．$V3$モデル上のコンバージョン傾向が強くなった理由は，多くのページを探索する傾向が弱まり，コンバージョンへ向かう傾向が全体的に表れてきたためと考えた．

3.4.2 $V3$モデル

訪問回数別の分析から，$V3$モデルは情報収集から購買決定段階に移行した訪問者が多いと考えた．そのため，$V3$モデルを用いて購買決定段階に近い訪問者の興味対象の把握を行った．抽出された$V3$モデルのパス図を図3に示した．また，ページが抽出された順序を表6に，各点に対応するページ内容と矢印数を表7に示した．

図3と表7より，まず訪問者は「トップページ(1)」から「価格表(30)」を選択しやすい．そして「機能(16)～(28)」ページを閲覧し，再び「価格表(30)」を閲覧してか

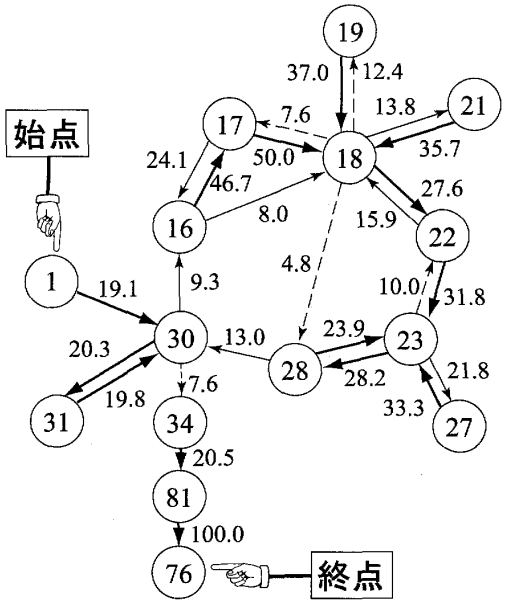

図3　$V3$モデル

表6　ページ連結順序

順序	ページ番号	順序	ページ番号
1	1	15	18
2	30	16	21
3	31	17	18
4	30	18	19
5	16	19	18
6	17	20	17
7	18	21	16
8	22	22	18
9	23	23	28
10	28	24	30
11	23	25	34
12	27	26	81
13	23	27	76
14	22		

表7　ページ内容と矢印数

番号	ページ内容	矢印数
1	トップページ	1
16	サイト解析	4
17	ページ解析	4
18	経路解析	10
19	経路解析の特徴	2
21	逆引き経路解析	2
22	その他の機能	4
23	広告測定効果	6
27	アドワーズ	2
28	SEO効果測定	4
30	価格表	6
31	導入事例1	2
34	申込	2
76	送信完了	1
81	入力内容の確認	2

ら「申込(34)」へ移動する．

全データを分析した図2では，最初はサイトのリンク配置順に依存した移動がされていた．しかし$V3$モデルでは，初めの移動からリンク配置順に従っていない．これは，訪問者が自分の興味あるページを選択しやすくなり，訪問者の興味が反映されやすくなった結果である．

また，基本的なページである「シビラの説明(13,14)」が抽出されていない．これは，既に情報収集の段階で閲覧し，もう見る必要がないページであるためと考えた．

モデルより，入口ページの「トップページ(1)」から初めに移動するのは「価格表(30)」であり，最終的に「申込(34)」に移動する直前にも「価格表(30)」を閲覧することが分かる．このことから，図2と同様に，「価格表(30)」が訪問者の意思決定に大きく影響を与えているといえる．図2で最も矢印数が多いのは「価格表(30)」であったが，表7から$V3$モデルでは「経路解析(18)」の矢印数が最も多い．経路解析はシビラの機能の内，訪問者の閲覧経路を把握する機能である．そのため，シビラを購入する訪問者は，シビラが提供する機能の中でも経路解析に強い興味を抱いていると考えた．

4．おわりに

本研究では，電子商取引市場においてコンバージョンを増やすため，訪問者の興味ページと移動ルートを把握できる分析手法を提案した．提案したHTPモデルでは，始点から終点を一筆書きで結ぶ過程において，訪問者の大まかな移動ルート，多くのページとの関連性が深いページを把握できた．また，訪問回数別にモデル作成を行うことで，購買決定段階の訪問者の行動傾向をつかむことができた．本研究では，特に訪問回数3回目の訪問者を「購買決定」の段階と考え，訪問者の興味対象の把握を行った．分析結果から，コンバージョンに近い訪問者の購買を促すには，「価格表(30)」と「経路解析(18)」を軸としたアピールを行う必要があるといえる．例えば，シビラの機能の中でも，経路解析の機能のみを安価で提供するといった提案が考えられる．

本研究が属するWeb Usage Miningにおいてマルコフモデル[5][13]やシーケンスパターン[1][2][7]に関する先行研究があった．HTPモデルにより，これらの研究では把握できなかった訪問者の全体的な移動パターンを抽出できた．マルコフモデルでは，確率が高いページを追えば中心的な移動ルートも把握できるが，それを一意に抽出可能にしたのがHTPモデルである．HTPモデルでは，入口ページから目的ページまでのパターンを一意に抽出し，矢印数から過程に現れる各ページの選択されやすさも把握できることが利点である．本研究では，矢印数が多いページを訪問者が関心を抱くページとして，訪問者の興味対象を把握した．HTPモデルは，訪問者の興味ページとコンバージョンまでのルートとの関わりを把握し，コンバージョンを増やすための重要コンテンツ把握に寄与できる．HTPモデルは，サイトのページ数が多い場合でも訪問者の興味ページと移動ルートを視覚化できる．また，閲覧されるページの順番よりも，全体的なパターンを表現し，直観的に把握しやすいモデルにした．そして，モデル内にループを含むことを許し，より訪問者行動に近い傾向を表現できた．

本研究では，購買決定段階の訪問者群を把握するために訪問回数とコンバージョン率を用いた．しかし，それ以外にも2指標を用いた実務的メリットがある．それは，シンプルな移動傾向を抽出できることである．本研究で全データを分析した際，比較的複雑で規模の大きいモデルとなった．もし，さらに規模の大きいモデルとなるサイト（例：コンバージョン率が低い）の場合，図が複雑で視覚的に把握しにくくなる．その際に購買決定段階に近い訪問者に着目することは，モデルの視覚化を行う上でもメリットが大きい．コンバージョン率の高い訪問者を用いても複雑なモデルとなる可能性もある．その場合は，目的ページを申込の送信完了ページより前の申込ページに設定することで，モデルが終点に収束しやすくなる．

HTPモデルの最大の利点は，モデルの使用方法が非常に容易であることである．サイトの全ページにおいて遷移率を計算しておけば，後は矢印をつなぐだけであり，手書きでもモデル作成ができる．そのためサイトが改良されても，特にモデルの使用方法を変えることなく素早い分析ができる．

HTPモデルの活用方法は，まず自社サイトでモデル作成を行い，訪問者の興味ページと移動ルートを把握する．次に，興味ページのコンテンツ強化やキャンペーン等のアピールを行い，再びモデル作成を行う．そしてモデル内の遷移率や移動ルートがどう変化するかを把握し，訪問者へのアピールによる効果の見極めを行える．モデル内の遷移率が増加すれば訪問者の移動パターンが一様化してきたことを表し，移動ルートが短くなればコンバージョンへ向かう傾向が強くなったことを意味する．

今後の課題としては，新たな属性別の分析を行うことが挙げられる．本研究の分析では，訪問回数で訪問者を分類し分析を行ったが，訪問回数以外の，訪問者の購買プロセスに影響を与える属性で分析を行うことで，新たな知見を得られると考える．

参考文献

[1] R. Agrawal and R. Srikant: "Fast algorithms for mining association rules," Proc. 20th VLDB Conf., pp. 487-499 (1994).

[2] R. Agrawal and R. Srikant: "Mining sequential patterns," Proc. Intl. Conf. Data Engineering, pp. 3-14 (1995).

[3] R. Kosala and H. Blockeel: "Web Mining Research: A Survey," ACM SIGKDD, Vol. 2, Issue. 1, pp. 1-15 (2000).

[4] 江尻俊章:「稼ぐホームページ損なホームページ」, 株式会社アスキー (2004).

[5] 狩谷典之, 北島宗雄, 高木英明, 張勇兵:「Markovモデルを用いたe-コマースサイトのwebデザイン評価」, 電子情報通信学会論文誌B, J85-B, 10, pp. 1809-1812 (2002).

[6] 経済産業省情報経済アウトルック (http://www.meti.go.jp/policy/it policy/statistics), 最終閲覧日 (2006/12/8)

[7] 小柳滋, 上原子正利, 久保田和人, 仲瀬明彦:「WWWアクセスシーケンスの新しいマイニング手法の提案」, 電子情報通信学会論文誌D, Vol. J87-D1, No. 2, pp. 232-243 (2004).

[8] 総務省情報通信統計データベース (http://www.johotsusintokei.soumu.go.jp/), 最終閲覧日 (2006/12/8).

[9] 武井由紀子, 遠藤直紀:「ユーザ中心ウェブサイト戦略」, ソフトバンククリエイティブ株式会社 (2006).

[10] 田中あゆみ:「Webマーケティングの入門教科書」, 毎日コミュニケーションズ (2005).

[11] フィリップ・コトラー, ケビン・ケラー, 恩蔵直人監修, 月谷真紀訳:「コトラー&ケラーのマーケティング・マネジメント第12版」, ピアソンエデュケーション (2008).

[12] 山西健司:「Webマイニングと情報論的学習理論」, 2002年情報学シンポジウム講演論文集, pp. 9-16 (2002).

[13] 山本哲生, 北島宗雄, 高木英明, 張勇兵:「Markov連鎖を用いたウェッブナビゲーション過程の評価」, 情報通信ネットワークの新しい性能評価法に関する総合的研究, pp. 189-198 (2002).