

World Wide Web アクセスにおける探索構造
～インターネット探索行動圏概念構築にむけた研究～

筑波大学審査学位論文（博士）

2003

下方 拓

筑波大学大学院

ビジネス科学研究科 企業科学専攻

論文概要
博士（経営学）

World Wide Web アクセスにおける探索構造
～インターネット探索行動圏概念構築にむけた研究～

下方 拓
筑波大学大学院
ビジネス科学研究科 企業科学専攻

第1章 序論

自律分散型のネットワークの構造をもつインターネットによって世界中から障壁がなく容易なアクセスが可能になり、あらゆる場面で情報技術（IT）が活躍することにより経済活動に効率性がもたらされるということに大きな期待が集まった。

しかし、世界的な規模でいわゆる IT バブルがはじけることで過剰な期待は影をひそめるようになった。インターネット上ではさまざまなサービスをもたらすアプリケーションの開発・試行が先行して進んだため、インターネット自体の性質やその上での人々の行動の性質を捉える試みはようやく始まったばかりである。

本研究は、Web の探索構造として「コンテンツ探索構造」というあたらしい構造を提案し、「インターネットは時間と空間を克服したのか」という命題を検証することを目的にしている。そして、その探索構造をもとにしたインターネット上での探索行動圏を展望しようというものである。

第2章 Web解析と探索行動のモデル化研究の動向と課題

探索行動のモデル化と Web 解析の研究の動向から命題に対する課題を探った。

Web 上の探索行動モデル化研究として、消費者情報探索行動研究は、インターネットの電子商取引が時間と空間の壁をこえて効率的な完全競争市場になったのかという疑問に対して、探索コストやブランド力を導入したモデル化によりインターネットにおい

ても製品に異質性が生まれることで別の意味での壁が存在していることを指摘している。情報採餌行動モデルは、数理生物学のアナロジーから情報を探索・消費していく行動をモデル化し、限界効用原理にもとづく情報獲得率の長時間平均が最大になる条件を示している。空間相互作用モデルでは通常の商圈モデルで2つのロケーション間の空間的な距離を定義したモデル化がおこなわれているが、インターネットの世界では距離に相当する変数を定義できずに具体的な成果は残していない。

Web 構造分析においては、[Cooley 99]がアクセスログ分析において参照時間長を用いたコンテンツランザクション処理を提案している。このコンテンツランザクションを用いることで探索コストを定量的に計測・分析できる。Web をグラフ構造ととらえるアプローチには、その特徴となるべき乗法則や Web サイズが議論されているが、URL 間リンクの強連結集合をもちいて Web のマクロ構造を描き出した[Broder et al. 00]の研究は興味深い。静的なリンク情報とアクセスログをどのように連携させられるが今後の課題となる。

第 3 章 ユーザのインターネットアクセス手段の選択行動

「探索コスト」は、ある商品情報を探索するためにつかわれる交通手段やサービスへの金銭支出とそれに要した時間に対する機会費用である[成生 94]。アクセスサービスの対価としての探索コストを検討するために、インターネットのアクセス手段の選択行動を分析することで、「探索コスト」に対してどのような要素が支配的で価値があるのかを検討した。

具体的には、ISP やアクセス回線の選択をロジットモデルやネスティッド・ロジットモデルを構築することで分析し、使用時間と単位時間コストで構成される総費用が大きな影響をもつことを確認した。さらに選択肢が数多くある状況や将来の曖昧な選択肢を評価する場合には、事業者のブランドを頼りに選択し、他の条件が同じであれば月額費用に対して数千円程度の価値をもっていることが明らかになった。

第 4 章 探索コストを用いた研究仮説の設定

第 4 章では、命題を検証する次の 2 つの仮説を設定した。

- 仮説 1 : 「インターネットでのコンテンツ探索コストは、Web サイト間のユーザ交流トラフィックに反比例する」
- 仮説 2 : 「Web 探索グラフにおいて、強連結されたグラフ構造に属する Web サイトには、他のクラスターより多くのアクセスが生じている」

これらを検証することで、機会費用としての探索コストの意味をもつコンテンツ探索距離がインターネットでのコンテンツ探索行動における距離の壁となり、Web ページリンクの空間構造が空間の壁となっていることを検証できる。

そのような構造を形成する原因を、Web グラフの物理的な特徴から検討するために、次の仮説を提示した。

- 仮説 3 : 「Web 探索グラフの直径は比較的小さく、少ない探索コストでコンテンツにたどり着けるスモールワールドを形成している」

さらに、ユーザの合理的な探索行動が Web グラフの構造を規定していると考え、より多くのコンテンツにアクセスすることが合理的だとして空間的な探索ヒューリスティックの合理性を検証するために、次の仮説を提示した。

- 仮説 4 : 「Web コンテンツ探索行動において、ユーザは探索の幅と深さを組み合わせたコンテンツ探索ヒューリスティックの中から、探索行動中により多くのコンテンツにアクセスできる合理的なヒューリスティックを選択する」

そして、探索活動全体の時間の中で獲得した情報量の比率を最大化する探索時間戦略配分の合理性を考察するためのシミュレーションをおこなうために、次の仮説を提示した。

- 仮説 5 : 「Web コンテンツ探索活動全体を通して、コンテンツの探索時間と消費時間を効用関数とした単位時間当たりの獲得情報量を最大化する探索コスト戦略が存在する」

第 5 章 探索距離概念とWebサイト探索グラフ構造の検証

第 4 章の仮説を検証するための基礎的な概念として、参照時間長を基準にしたコンテンツトランザクションにおける探索距離概念を定義した。この概念には、探索行動がもつユーザの明示的な「目的」「認知」「選択」という態度は扱われず、意味の重要性を失わずにアクセスログを探索シーケンスに変換し、定量的な距離概念として定義する

ことを重視した。[Cooley 99]が示したコンテンツランザクションにもとづいて、ユーザの URL へのアクセス行動を探索過程に変換し、コンテンツを探索するための探索コストを計算した。

データとして、123 人のユーザの行動をモニターし、3 ヶ月間にアクセスした Web の URL をトレースした十数万レコードのアクセスログデータを使用した。そして、ログデータからコンテンツランザクションを同定し、コンテンツの探索時間、探索ステップ数、消費時間を計算した。

これらをつかってコンテンツランザクションによる探索コストを算出することで、Web サイトのトラフィックを魅力度とし、Web サイト間の探索コストを距離項においた Web サイト間の空間相互作用を推定するグラビティモデルを構築した。探索コストがサイト間トラフィックにおいて距離として機能することを確認し、仮説 1 を確認した。

さらに、Web グラフ構造では、算出された強連結集合 (SCC) をもとにしたクラスター間でトラフィックに違いがあることが検証し、仮説 2 を確認した。

このようにして、Web サイト間リンクの探索グラフ構造においては、インターネットは空間と時間を克服しているのではなく、コンテンツを探索するための探索コストを距離とし、Web ページリンクを空間構造とした壁が存在していることを確認した。

しかし、一方そのサイズについては、仮説 3 による探索コストを用いたモデルにおいても、従来の研究同様、Web のクラスターの直径 (URL 間の最短パスの平均値) が 10 以下の比較的小さな構造でコンテンツが連結していることを確認した。同時に、SCC のクラスターにおけるマクロ構造においてどのクラスターでも平均探索ステップ数が 4 程度であることを考慮にいれると、「マクロな」構造のなかには、比較的小さなコスト負担でコンテンツにアクセス可能な「小さな」構造があると考えられる。

第 6 章 ユーザ探索パターンの表現と情報採餌行動モデル

第 6 章では、探索構造を検証する仮説 4 を確かめるために、ドメインジャンプ表現による探索パターンとユーザ探索行動タイプを規定したが、探索ヒューリスティックにおいてアクセスしたコンテンツの数の多さに合理性をおいた仮説 4 を検証することはできなかった。しかし、支配的な探索ヒューリスティックである「深さ優先探索」の時間配分からは、コンテンツの数ではなく、比較的短い探索時間で長くコンテンツを消費する

探索時間の組み合わせが合理的であることが伺える。

さらに、探索時間戦略の合理性を検討する仮説5を検証するために、情報採餌行動モデルによるシミュレーションをおこない、情報獲得率の長時間平均が増加しなくなるような探索参照時間が存在していることが明らかになった。ユーザが情報獲得率の長時間平均を最大化するような合理的な行動を仮定した場合に、ある探索時間を超えて探索することは合理的でない。仮説4と5を通して、Web探索構造からインターネットの時空間を考えると、距離と時間の制約に多くを縛られず、なおかつ長時間の情報獲得率を最大にするような効率的な探索が可能であるが、ある探索参照時間をこえて、さらに消費時間も最適値を超えて探索する場合には、距離と空間は現実世界と同様に非効率性をもたらしているといえる。

第7章 結論と今後の課題

空間と時間を克服したといわれるインターネットでの探索行動は、サイト側からみれば探索コストを距離とした経路に依存しており、道筋的であるといえる。しかし、ユーザ側からは探索パターンの違いはあってもアクセスしているコンテンツ数に違いはなく、かならずしも経路に依存しているとはいえない。このような一見矛盾した現象は、Webアクセスの探索構造が「マクロな」構造と「小さな」構造ともいうべき2つの側面をもつことで観察されるといえる。そして、この2つの構造的な側面においてここまでは探索してもいいとする探索参照時間が存在することにより、ある時間内での探索を効率的におこなう行動をとろうとする結果と考えることができる。

新しく提案したコンテンツランザクションによる探索コスト構造を適用し、インターネットの構造の一部を明らかにできたことは今後のサイトデザインやトラフィックを指向したマーケティング活動に大きな示唆をもたらすと考える。提案した探索コストにはユーザの探索行動自体を明示的に行っていないという課題はあるが、インターネットの中で探索コストを距離とした圏構造を構想するために寄与すると考える。

目次

部

第 1 章	序論	2
1.1	問題意識.....	3
1.2	インターネットの性質を捉える試みと問題点.....	5
1.3	研究の目的と概要.....	6
1.4	論文の構成.....	8
第 2 章	WEB 解析と探索行動のモデル化研究の動向と課題	10
2.1	インターネットの性質を探るアプローチと研究サーベイ分野.....	10
2.2	インターネットマーケティング・エコノミクス研究.....	12
2.2.1	分野の概観.....	12
2.2.2	消費者情報探索モデルと探索コスト.....	15
2.3	WEB アクセスログ解析・WEB マイニング研究.....	19
2.4	WEB 生態学 (WEBOLOGY) 研究.....	22
2.4.1	Web のグラフ構造.....	22
2.4.2	情報取得行動と情報採餌行動.....	25
2.5	関連研究分野のまとめと課題.....	27
2.5.1	インターネットマーケティング・エコノミクス研究における課題.....	27
2.5.2	Web アクセスログ解析・Web マイニング研究における課題.....	29
2.5.3	Webology 研究における課題.....	30
2.5.4	この章のまとめ.....	31

部

第 3 章	ユーザのインターネットアクセス手段の選択行動	34
3.1	本章の目的とアプローチ.....	34
3.2	関連研究.....	36
3.3	分析モデルのフレームワークとデータ.....	40
3.4	ISP の選択による多項ロジットモデル.....	43

3.4.1	多項ロジットモデルの定式化.....	43
3.4.2	選択肢とその属性.....	44
3.4.3	ユーザデータの概要.....	46
3.4.4	推定結果.....	49
3.5	アクセス回線とISPの選択によるネスティッド・ロジットモデル.....	52
3.5.1	ネスティッド・ロジットモデルの定式化.....	52
3.5.2	選択肢と属性.....	54
3.5.3	推定結果.....	55
3.6	考察とまとめ.....	58

部

第4章	探索コストを用いた研究仮説の設定.....	62
4.1	研究課題と仮説の設定.....	62
4.1.1	命題とアプローチ方法.....	62
4.1.2	探索コストを導入した距離と空間の考え方.....	65
4.1.3	命題を検証する仮説の設定.....	69
4.2	WEB探索構造の形成要因についてのアプローチ.....	71
4.2.1	Web探索構造形成要因を探る2つの方向性.....	71
4.2.2	スモールワールド仮説によるWeb探索構造形成要因.....	71
4.2.3	ユーザの探索行動によるWeb探索構造形成要因.....	72
4.3	本章のまとめ.....	76
第5章	探索距離概念とWEBサイト探索グラフ構造の検証.....	77
5.1	本章の目的とアプローチ～探索コストを距離としたWEBアクセス分析.....	77
5.1.1	目的とアプローチ.....	77
5.1.2	探索行動に対する立場.....	79
5.2	コンテンツランザクションによる探索コストの算出.....	79
5.2.1	探索行動モデルにおけるコンテンツランザクション.....	79
5.2.2	コンテンツランザクションの同定.....	82
5.2.3	コンテンツランザクションにおける探索距離の意味.....	87
5.3	データ概要.....	89

5.3.1	ユーザプロフィール	89
5.3.2	トランザクションデータ	90
5.4	グラビティモデルによる探索距離概念の検証	97
5.4.1	探索コストを距離としたグラビティモデルの構築	97
5.4.2	推定結果	99
5.5	WEB サイト探索グラフ構造による空間構造の検証	102
5.5.1	Web サイト探索グラフ構造分析に対するアプローチ	102
5.5.2	強連結集合 (SCC) によるサイト間交流のマクロ構造	103
5.6	WEB 探索グラフの特徴による構造形成要因	110
5.6.1	仮説とアプローチの確認	110
5.6.2	入次数の分布の検証	111
5.6.3	Web グラフの直径：スモールワールド仮説の検証	114
5.7	本章のまとめと考察	115
第 6 章	ユーザ探索パターンの表現と情報探餌行動モデル	120
6.1	本章の目的と仮説	120
6.2	ドメインジャンプ表現によるユーザ空間探索ヒューリスティックの評価	122
6.2.1	探索ヒューリスティックに関する仮説と検討課題	122
6.2.2	ドメインジャンプ表現による探索ヒューリスティックパターン	124
6.2.3	探索行動タイプ：ユーザ探索ヒューリスティックのパターンの分類	125
6.2.4	分析結果	130
6.2.5	探索ヒューリスティックにおける合理性の考察	131
6.2.6	この節のまとめ	134
6.3	情報探餌行動モデルによる探索行動要因の検証	134
6.3.1	本節の課題と検証仮説	134
6.3.2	情報探餌行動モデル	135
6.3.3	探索参照時間による数値シミュレーションと仮説 5 の検証	139
6.3.4	モデルとシミュレーションを通じた考察	143
6.4	まとめと考察	145
第 7 章	結論と今後の課題	147

7.1	結論.....	147
7.2	応用の方向性.....	150
7.3	今後の課題.....	150
7.4	補論：インターネット探索行動圏の展望.....	152
7.4.1	インターネット探索行動圏のイメージ.....	152
7.4.2	インターネット探索行動圏へのアプローチ.....	153
7.4.3	今後の研究課題.....	155
	謝 辞.....	158
	参考文献.....	160
	付録.....	I

部

部では、本研究の命題を提示し、全体の成果と研究の意義をまとめるとともに、関連分野の成果について述べ、本研究の課題を明らかにする。

第1章 序論

インターネットでは個人が閲覧するホームページから電子商取引まであらゆるジャンルでさまざまな試みが行われ、経済活動に対しても大きな影響をもつようになってきた。インターネットがこれほどもてはやされるようになったのも、インターネットがもつ可能性について人々が大きな期待を寄せているからである。その可能性とは、自律分散型のネットワーク構造によって世界中から障壁なくアクセスできる開放的でグローバルなアクセシビリティをもっているということである。それにより、コミュニケーションにおける時間と空間の壁を容易に乗り越えることができ、インターネットをとおしてあらゆる場面で情報技術（IT）が活躍することにより、経済活動に効率性がもたらされるということに大きな期待が集まったのである。

しかし、世界的な規模でいわゆる IT バブルがはじけることで過剰な期待は影をひそめるようになった。やはりインターネットに対する期待度は過度なものであったのか、インターネットにもその性質に因んだなんらかの限界があるのか。インターネットではアプリケーションの開発・試行が先行して進んだため、インターネット自体の性質やそのなかでの人々の行動の性質を捉えるという試みはようやく始まったばかりであるが、さまざまな分野で成果を結びつつあり、その沃野は大きい。

本研究は、インターネットでの人々の行動の性質を捉える試みとして、学際的なアプローチにより課題を浮き彫りにし、Webの探索構造として「コンテンツ探索構造」というあたらしい構造を提案・検証することで、「インターネットは時間と空間を克服したのか」という命題を検証することを目的にしている。そして、その探索構造をもとにしたインターネット上での探索行動圏を展望しようというものである。

本章では、研究課題に対する問題意識を明らかにし、本研究の目的と意義について述べる。

1.1 問題意識

インターネットに対する大きすぎる期待

インターネットには、自律分散型のネットワークという性質や HTML による Web のハイパーリンク構造という柔軟性をもった性質がある。そして、世界中が高速なバックボーンのもとで接続されることで、世界中からインターネットに容易に接続することが可能になり、多様なコンテンツを楽しめるようになったことから、「インターネットにより距離と時間が克服された」「個人、企業、国家といった各主体間の距離と時間を克服」[e-Japan 01]したといわれるようになった。また、商取引をインターネットベースでおこなうことで、経済活動にもインパクトをもつようになり、取引コストやマーケティングコストを激減し、商取引が効率化され、電子商取引市場がブレイクすると考えられるようになった。さらに、「情報量や距離に対する変動費が低い」[国領 97a]ので「伝達できる情報のリッチネスとリーチのトレードオフが解消される」[Evans 00]などという期待が高まっていった。

このような考え方には、インターネットを利用して必要な価格情報や製品情報を瞬時に収集することが可能になれば、現実の市場においても情報が完備された完全市場が実現するという大きな期待があったからだともいえる。

メディアによる空間的な距離の変容

社会学的な視点からは、メディアとしてのインターネットが社会に対してどのようなインパクトをもつのかということに感心が注がれてきた。電話やTVなどと同様にインターネットをメディアとテクノロジーの関係で捉え、「テクノロジーの進化は常に空間的な距離を短縮する量的な方向で行われてきたが、インターネットによりさらにその空間のもつ意味も変質してきている」[黒崎 99]と指摘されている。さらに、そこで行われるコミュニケーションの空間的な特徴について以下のような興味深い指摘がある。

ポール・ヴィヴィリオは、現代の情報メディア技術が可能にする関係のあり方を、先に述べた道筋的 = 移動経路的 (トランジェクティブ) に対して、「望遠レンズ的 = 遠隔对象的 (テレオブジェクティブ)」という言葉で表現している。電氣的・電子的なメディアテクノロジーが、それまで一体だった「情報の移動」と「情報を運ぶ媒体の移動」を切り離した結果、情報が伝達される道筋 = 経路の空間やそこでの経験を情報伝達の過程から取り去り、望遠レンズを通して遠方をみるように遠隔地の事物や出来事をリアル・タイムで「いま・ここ」に現前させることを可能にし

たことを指している。そこでは、道筋的 = 移動経路的な関係においては存在した距離感や遠近法的な空間の連続性が、コミュニケーションの空間から消失していく[若林 02]。

このようなインターネットによる都市環境からの経路と距離の喪失を危惧する言説*は、人間の拡張[マクルーハン 87]というスキームでメディアを捉えることで人間の身体性や空間把握の考え方と強く結びつき、インターネットが従来の社会的機能を変容し、人々のコミュニケーションスタイルを変える「IT 革命」ともいうべき社会的なインパクトをもたらすという考えにつながっている。

期待と現実とのギャップ

確かに、インターネットで Web サーフィンをやってみれば、ポータルサイトや検索エンジン・価格探索エージェント(ショップボット[Brynjolfsson 00a])を通じて、商品の製品情報や価格情報を収集することができるし、高機能化された Web ブラウザや Web サイトのさまざまなナビゲーション機能のより、自由にさまざまなコンテンツを手に入れることができる。そのような期待が高まり、日本での 2001 年の B2C での電子商取引市場規模は 1 兆 4840 億円と前年比 1.8 倍と爆発的に伸びた[総務省 02][商務省 00][インターネット白書 02]。

しかし、それが人々の満足につながっているかということ現状では必ずしもそうではない。「インターネットにアクセスする料金がなくて帯域スピードが遅い」、「欲しい情報が思うように検索できない」、「魅力的な商品をもってインターネットに電子店舗を出店しても思うように集客できない」などといった不満が多く聞かれる。

さらに、インターネットに対する期待の高さも、IT バブルが崩壊したこともあいまって一時の熱狂的な喧伝のされ方は少なくなってきた。日本の総務省の情報通信白書やアメリカの商務省のデジタルエコノミーに関するレポートでも、2000 年前後から「IT 革命を喧伝する方向」から「IT 技術をどのように有効に活かせるのか」という視点にかわってきている[総務省 02][商務省 00]。

果たして、インターネットは時間と空間が克服された効率的な場だったのか。その性質はどうなっているのか。その物理的な構造はどうなっているのか。その性質やそこで行われる人々の活動をどのように計測することができるのか。このような問いかけに対して、さまざまな分野から学際的な試みが行われてきた(第 2 章)。

* [Dodge 01]などのサイバースペースと現実世界をマッピングさせる試みはこの危惧に対する逆のアプローチである。

1.2 インターネットの性質を捉える試みと問題点

Web アクセスログ解析・Web マイニング研究

Web アクセスログやリンクを解析することで、Web に対する構造的な知見を得ようというアプローチが行われている。ログ解析のためにデータ整理の手法[Cooley 99]が整備されて、そのデータをもとに関係性をさぐる Web マイニングが実施される。分析の方法論としては、関係性ルール（アソシエーションルール）[Agrawal 95]や HITS アルゴリズム[Kleinberg 98]などのほかに、マルコフモデル[Pirolli 99b]や N-gram[風間 99][佐藤 99]を用いたものまで多様な試みが行われている。

さらに、Web をグラフ的にとらえてビジュアルな形にマッピングしたりする Web 可視化の技術[柴山 01][Hochheiser 99][Dodge 01][Spiliopolou 99]が提案され、インターネットという新しいメディアが生み出した空間を把握するために「Web 空間で喪失した経路と距離」[若林 02]を都市空間のアナロジーを用いて視覚化することによって回復する試みも行われている。しかし、Web 全体の性質を捉えようとすると、複数ユーザの大量の URL 情報を処理しなければならず、具体的な距離尺度の設定が難しい。

インターネットマーケティング・エコノミクス研究

電子商取引におけるマーケティングを検討するために、インターネット上の消費者の商品情報探索行動を捉えようとする議論が盛んに行われている。探索コストを導入した電子商取引市場（electronic marketplace）での一般均衡分析モデル[Bakos 97]が導入され、探索コストがもたらす価格水準、価格分散などを測定し、従来型の市場にくらべてインターネット上の電子商取引がもたらす「効率性」について多様な議論が展開されている[Smith 99][Brynjolfsson 00b][Lynch 00][Varian 99]。

さらに、サーチエンジンをつかった探索候補から商品を選択する消費者選択行動を多項ロジットモデルやネステッド・ロジットモデルをつかって分析した実証研究[Brynjolfsson 00a][Chatterjee 98]が行われるなど、数多くの成果が提出されている。しかし、具体的な探索コストを定量的に捉える方法がなく、効率性を示す傍証を議論していることが多い。

Web 生態学(Webology)研究

アクセスログやリンク情報を単純に解析するのではなく、Web 全体を生態系と捉え

て Web の物理的な特徴量やマクロ構造を描き出そうとする試みを「Web 生態学 (Webology)」と呼ぶことがある。この分野では、Web サーフィンを行うユーザの行動や Web サイトのリンク構造のもつ特徴的な規則性を検討している [Pirolli 99b][Huberman 98][Pitkow 98]。また、Web サイトでのユーザの行動を、数理生物学におけるパッチ内での捕食行動のアナロジーで捉えた情報採餌行動 (Information Foraging) [Pirolli 99a] のアプローチやサイト間のリンク構造の関係性をもとに Web のマクロ構造を描き出そうという試み [Broder et al. 00] もある。いずれも興味深い報告が多いが、分析が静的な Web リンク構造だけに留まっていることが多く、ユーザの行動を直接測定し、双方を結び付けたアプローチは少ない。

1.3 研究の目的と概要

目的と分析の対象

本研究の目的は、前節のような問題意識のもとで、インターネットは「時間と空間の壁を克服した効率的な場」であるのかということや、インターネットでのコミュニケーションが「道筋的 = 移動経路的 (トランジェクティブ) ではなくて、遠隔对象的 (テレオブジェクティブ)」なのか、という命題を検証しようというものである。

ここで注目されるのは、インターネット自体の構造というよりも、そこで行われる売買や探索活動といったコミュニケーション活動を繰り広げる人々の行動である。そのために必要とするデータは Web の静的なハイパーリンク構造ではなく、活動するユーザの URL へのアクセス行動である。そして、検証のためには定量的な測定が必要である。

しかし、ユーザのアクセス行動における「空間と時間の壁」を考えるにあたり、Web のアクセス空間とそこでの人々の活動を扱おうすると、その空間自体を数学的にも社会学的にも厳密に定義することが必要になる。また、ユーザの行動を説明するための周辺データは膨大で、かつ入手は難しい。そのため、ユーザの Web 空間での体験を把握するためのもっとも基礎的な方法として、URL へのアクセス行動をユーザが実際に体験した「経路」として取り扱うアプローチ [原 96] をとる (第 4 章)。

本研究の概要

本研究の命題に対して、探索コストを距離としたユーザの URL へのアクセス行動を

分析した。探索コストはある（商品）情報を探索するためにつかわれる交通手段やサービスへの金銭支出とそれに要した時間に対する機会費用である[成生 94]。まず、アクセスサービスコストの面では、探索コストとインターネットアクセス行動の関係について検討し、インターネットアクセスサービス選択行動をロジットモデルで表現し、アクセスサービス料金や利用時間、商品ブランド等に依存していることを確認した（第 3 章）。

また、機会費用としての探索コストに着目し、ユーザの URL へのアクセス行動を定量的に扱うために、コンテンツランザクションという概念を用いて「コンテンツ探索距離」を新しく定義し、「探索時間」や「探索ステップ数」を探索コストとして計算する手法を提案した（第 4 章）。

そして、百人以上の数ヶ月間にわたる実際のアクセスログデータをもとに、コンテンツ探索距離概念を使って以下の 2 つの仮説を検証した（第 5 章）。

- 仮説 1：「インターネットでのコンテンツ探索コストは、Web サイト間のユーザ交流トラフィックに反比例する」
- 仮説 2：「Web 探索グラフにおいて、強連結されたグラフ構造に属する Web サイトには、他のクラスターより多くのアクセスが生じている」

これらの仮説を検証することで、コンテンツ探索距離がインターネットでのコンテンツ探索行動における距離の壁となり、Web ページリンクの空間構造が空間の壁となっていることを確かめた。

次に、この探索構造を形成する要因を Web グラフの物理的な特徴から検討するために、以下の仮説を検証した（第 5 章）。

- 仮説 3：「Web 探索グラフの直径は比較的小さく、少ない探索コストでコンテンツにたどり着けるスモールワールドを形成している」

さらに、ユーザの合理的な探索行動が Web グラフの構造を規定しているとして、より多くのコンテンツにアクセスすることが合理的だとする場合における空間的な探索ヒューリスティックの合理性を検証するために、以下の仮説を検証したが、確認できなかった（第 6 章）。

- 仮説 4：「Web コンテンツ探索行動において、ユーザは探索の幅と深さを組み合わせたコンテンツ探索ヒューリスティックの中から、探索行動中により多くのコンテンツにアクセスできる合理的なヒューリスティックを選択する」

しかし、仮説 4 の検証で判明した事実をもとに、探索活動全体の時間の中で獲得した情報量の比率を最大化する探索時間戦略配分の合理性を考察するために、以下の仮説にもとづいたシミュレーションをおこなった。

- 仮説 5 : 「Web コンテンツ探索活動全体を通して、コンテンツの探索時間と消費時間を効用関数とした単位時間当たりの獲得情報量を最大化する探索コスト戦略が存在する」

以上の仮説検証やシミュレーションを通じて、本研究の命題である Web 探索における探索距離構造を明らかにし、それを形成する要因として Web 探索グラフの特徴的な構造とユーザ探索行動の合理性について議論を展開した。(第 6 章、第 7 章)

最後に、コンテンツ探索距離を距離概念としたインターネット探索行動圏構造を構想し、その研究課題を展望した(第 7 章)。

1.4 論文の構成

本論文は大きく 3 部に別れ、全体では 7 つの章から構成される(図 1-1)。

序論と第 2 章を全体の概要と関連分野を概観する 部、第 3 章を本論に入るまえの事前の分析をおこなう 部、第 4 章から第 7 章までを仮説検証を踏まえた本論とする 部とした。

序論では、本研究の目的と概要について述べた。

第 2 章では、インターネットの性質を捉える各分野の取り組みを概観し、これからの分析にむけた課題について述べる。

第 3 章では、サービスの対価としての探索コストとインターネットアクセス行動との関係を調べるため、インターネットにアクセスする回線と ISP の選択構造を分析する。

第 4 章では、命題を検証する研究仮説を提示し、探索コストとしてのコンテンツ探索距離概念とそれに基づいたデータの処理方法について述べる。

第 5 章は、実際のアクセスログデータを用いたグラビティモデルを構築することで探索距離概念の有効性を検証する。さらに、探索距離に基づいた Web グラフのマクロ構造を検証する。

第 6 章は、探索構造を形成する要因を探るために、ドメインジャンプ概念によりユーザのアクセス行動の探索戦略を類型化し、ユーザ探索パターンとトラフィックの間の関

係を探る。さらに、情報採餌行動モデルを用いたシミュレーションによりユーザの探索時間戦略を検証する。

第7章は、結論として、本研究の成果とモデルの限界について述べ、以上の成果をもとに、インターネット探索行動圏概念を構想し、今後の課題についてまとめる。

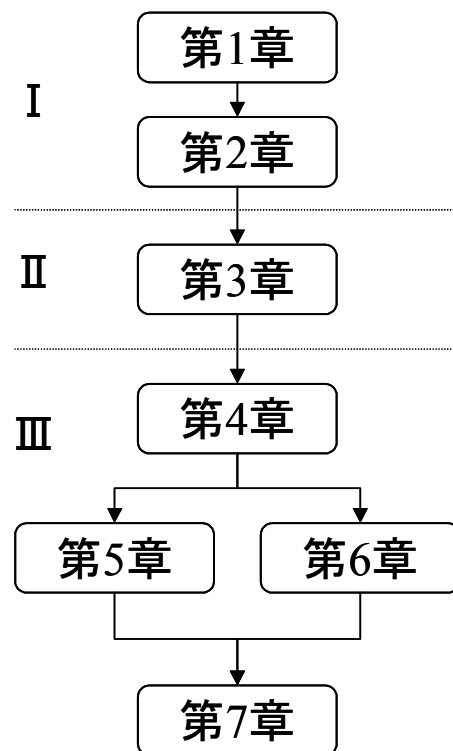


図 1-1 論文の構成

第2章 Web解析と探索行動のモデル化研究の動向と課題

本章では、関連分野を概観し、ユーザの Web アクセス行動を経路的に捉えて分析するための方法論と、得られたデータを探索行動として解釈するためのモデル化の方法論についての研究動向を述べ、本研究の目的となる Web アクセスにおける探索行動のモデル化のための課題について検討をおこなう。

2.1 インターネットの性質を探るアプローチと研究サーベイ分野

関連分野

インターネットの性質を捉える試みは多方面にわたっているが、アプローチには大別して2つの方向性がある。既存の概念（たとえば経済学）をインターネット上に適用することでインターネット自体をその枠組みの中で評価しようという方向と、Web 上のリンクやアクセスログといった構造的なデータから知識を獲得するための方法論を提示して実証研究を行うという方向である。おもに、前者は

- インターネットマーケティング・エコノミクス研究
という領域を形成しており、後者は、
- Web アクセスログ解析・Web マイニング研究
という領域での研究が盛んに行われている。

そして、その両者を融合したようなアプローチとして、

- Web 生態学（Webology）研究
といった分野が形成されつつある。

本研究との関連性

本研究では、『Web を代表とするインターネットは「空間と時間の壁」を克服した

のか』という命題に対して、ユーザの URL へのアクセス行動を経路的な情報として扱うことでアプローチする。このアプローチのもとで研究を進めるには、ユーザのアクセス行動に適用する行動概念モデルを想定し、それにもとづいたデータハンドリングにより必要な統計データを獲得し、Web アクセス行動を構造的に表現し、分析・検証をおこなうという検討手順が必要になる。

- Web でのアクセス行動に対する概念的な研究は、インターネットマーケティング・エコノミクス研究において消費者情報探索というトピックスでクローズアップされている。
- Web におけるデータハンドリングについては、Web アクセスログ解析・Web マイニング研究においてランザクシオン処理の問題として議論されている。
- Web アクセス行動の構造的な表現については、Web 生態学研究において Web グラフの構造というトピックスでさかんにとりあげられている。

そこで本章では、これらの分野の主要な研究を概観してその課題について述べ、本研究との関連性を明確にしていく。表 2-1には、これから述べる関連分野の主要研究と本研究の関連性をマトリックスで示した。

表 2-1 研究サーベイ分野の枠組みと主要研究

			本研究との関連			
			第3章	第4章	第5章	第6章
関連研究分野	インターネットマーケティング・エコノミクス	消費者情報探索	電話サービス選択ロジットモデル [Train 87] ショッピングボット選択行動モデル [Brynjoffson 00a]	消費者情報探索モデル [Bakos97] デジタル市場の効率性 [Smith 99] [Brynjoffson 00b]		
	Webアクセスログ解析・Webマイニング	クランザクシオン		コンテンツランザクシオンによる解析 [Cooley99]		
	Web生態学 (Webology)	Webグラフ			Webグラフとサイズ [Broder 00] Webグラフの直径 [Albert 99]	情報探餌行動モデル [Pirolli99a]

2.2 インターネットマーケティング・エコノミクス研究

2.2.1 分野の概観

マーケティングや経済学の方面では、従来の枠組みや概念道具のなかでインターネットのもつ性質を評価しようというアプローチが多い。そのため、研究者のバックグラウンドと分野の数だけ多くの研究例がある。経済的な分野としては、マクロ的な視点、ミクロ的な視点、情報財の特質論、消費者選択行動、都市産業立地、経営組織、料金制度など多岐にわたっている。

デジタル経済論

インターネットで行われる商取引をマーケティング・経済学の視点から分析し、ITやデジタルエコノミーが従来の経済と全く異なる新しいものではなく、伝統的なマーケティング・経済学の道具を利用することで分析可能だとする主張が展開されている[Shapiro 99][Varian 00]。そこでは「情報財としてのゼロに近い限界費用」、「バージョン化された生産」、「プライシング」、消費者効用に関する「ネットワーク外部性」の問題と「ロックイン」現象などの特徴が列挙され、デジタル経済の全体像を浮き彫りにすることが試みられている。特に、[Varian 00]は、amazon.comのアフィリエイトプログラムの重要性を指摘し、ロイヤリティプログラムが価格を探索するサーチボットなどのツールをほぼ無効にすることを議論している。

インターネット調査

マクロ的なアプローチでは、ITがGDPにおよぼす影響について、いわゆるニューエコノミー論を検証するような試みが多くおこなわれている[熊坂 01][商務省 00][Novak 97]が、この分析には常にメジャメントの問題や調査データの代表性の議論がつきまとう[Kannan 98][Hoffman 96b][Haltiwanger 00]。

Hoffman & Novakの一連の研究[Hoffman 96b]では、インターネット上で行うアンケート調査における代表性の問題とその対処方法として Post-stratification adjustment を利用したサンプルの補正の議論が行われ、さらに、ユーザの人種・言語や利用時間などによるセグメントはユーザのインターネットの閲覧行動に影響を及ぼすことが指摘されている[Hoffman 97]。

ネットワークサービスの消費者選択行動

インターネット上のサービス選択行動については、多項ロジットモデルやネスティッド・ロジットモデルをつかった消費者行動モデルや行動概念モデルを構築した実証研究が行われている。

[Train 87]は、住宅用ローカル電話の顧客のサービスオプションと通話ポートフォリオの相互選択を記述するモデルとしてネスティッド・ロジットモデルを適用した実証分析を行った。その効用関数を推定することでサービスオプション間の代替可能性 (substitutability) や価格弾力性を測定し、消費者が通話ポートフォリオ選択と割引サービスの組み合わせて選択している構造を明らかにした。この分析はインターネットが普及する以前の分析ではあるが、インターネットアクセス環境の選択が電話サービスの選択に関係することを考えると分析手法として有効なアプローチといえる。

[Brynjolfsson 00a]は、EvenBetter.com (現 DealTime.com) でショッピングボット (ワンクリックでネット店舗の価格調査と選択可能な店舗を表示するサービス) を利用した場合の商品・店舗選択の選択行動を分析した。彼らの研究の知見によると、推定された効用関数には、各サービス属性や店舗ブランド力などが強い正の要因として検出され、時間と頻度を乗じた機会費用と価格費用などが負の要因として検出されている。さらに、ブランドの有無を考慮したネスティッド・ロジットモデルを構築し、効用がより明確にブランドロイヤリティと価格に依存していることを確かめ、代替的な価格差別化・サービス差別化戦略の構築を検討している。そのほかにロジットモデルを適用した例として、バナー広告の選択構造の研究 [Chatterjee 98] がある。

また、ショッピングボットに関連して研究では、ショッピングボットが提示するサンプル数を戦略としたゲーム論的考察 [Kephart 99] が行われている。ショッピングボットの検索部分を担当するサーチエンジンについては、[Bradlow 00] が6つの異なるサーチエンジンのパフォーマンスを測定するモデルを提案し、探索エンジンを組み合わせて使うことの有効性を指摘している。

空間競争論と相互作用

産業集積立地の観点からは、[Hotelling 29] 以来、市場競争を空間競争の問題ととらえて様々な研究が取り組まれている。経済地理学の立場からは [Krugman 96] [Fujita 99] が空間立地論を展開し、チューネンの孤立国による産業構造からクリスタラーの中

心地理論などを俯瞰した上で、一見無秩序な状態からも経済的な構造が創出する過程を立地シミュレーションによりモデル化している。また、[Maurer 00]は Web サイト間の競争をサイトシェアのダイナミックな変化ととらえ、2 企業の間での競争を定式化から多企業間での競争に拡張しようとしている。

一方、地域科学の空間相互作用を扱う立場から、インターネット以前の分析ではあるが、[三友 95]がグラビティモデルを用いて電話トラフィックの県間交流をモデル化している。各県間のトラフィック*数を目的変数とし、加入数と距離を説明変数としてグラビティモデルを適用し、推定した規模指数と距離指数により通話交流状況による地域のセグメントをおこなっている。

このグラビティモデルをインターネットトラフィックに用いた例としては、[豊島 00]が Web サイトをもつ大学の地理上の位置とその間の Web サイトトラフィックをグラビティモデルで検証したが、実際の地理上の距離を距離要因に用いているため、推定精度に難がある。そもそも、インターネットでは地理上の空間距離とサイト間の距離要因には直接の関係がないので、地理上の距離を抵抗要因としての距離に用いることは難しいと思われる。

産業組織論

組織的な行動と産業構造については、デルコンピューターの直販システムの例などをとりあげ、インターネットによってエンド・ツー・エンドの取引が進むことにより仲介業者が存在できなくなるという「中抜き現象 (disintermediary)」が多く語られる [Evans 00]。しかし、[後藤 93]は一箇所にリクエストが集中することでコストが増大しかえって効率性が低下することを指摘し、仲介業者をエージェントとして捉えて、Brain (技術的な先導者)、Gate Keeper (技術の中継者)、End User (一般利用者)の三段階構造が効率性をもたらすと主張している。

また、売り手と買い手だけではなく、顧客間で商品の情報を共有したりするビヘイビアの発生について「顧客間インタラクション」[国領 97b]という概念も議論されている。この分野はコミュニティの生成や産業組織論に大きなインパクトをもっている。

* トラフィックは一般には交通量を示す言葉であるが、本論文では、電気通信分野でよく使われるように、コミュニケーションネットワークのなかを行き交うデータ量を意味する用語として使用している。

インターネットサービス料金制度論

インターネットアクセスのサービス料金制度の立場からは、インターネット自体が自律分散型でその費用負担が不透明であるため、インターネットを介したサービスを提供する場合の費用や価格のあり方についても議論が多い。インターネットアクセス料金の制度的な議論については、インターネットアクセスが定額制であるべきなのか、または従量制であるべきなのかという議論がおこなわれている [Bailey 97a][Bailey 97b][McKnight 97]。定額制は短期的には定額制は顧客の効用に影響するものの、長期的にはプロバイダーの設備投資を招き、効率的ではないという主張もある [Dewan 98]。

2.2.2 消費者情報探索モデルと探索コスト

消費者情報探索モデルの研究対象

インターネットエコノミクス・マーケティング分野で理論的に多様な成果が得られているのが、消費者情報探索とその探索コストを扱ったミクロ経済的なアプローチである。ここでは、消費者の商品情報探索行動 [Stigler 61] をとりあげ、電子商取引市場 (electronic marketplace) での探索コストを導入した一般均衡分析モデル [Bakos 97] による理論的考察や、実証的立場から従来型の市場にくらべてインターネット上の電子商取引がもたらす「効率性」について論じられている [Smith 99][Brynjolfsson 00b][Lynch 00][Varian 99]。

消費者情報探索モデルは、情報の経済学のさきがけとなった [Stigler 61] から始まり、伝統的な新古典経済学が見落としていた情報という枠組みを導入しながら、財の情報的な側面をクローズアップし、消費者が価格情報や品質情報を探索するコストと財の価格を中心に分析をすすめてきた。

[Stigler 61] は製品として均質な単一財を探索するシチュエーションをとりあげ、価格に関する情報の確率的な不確からしさにより、一物一価に価格が決まるのではなく均衡価格が分散するという事実を指摘した。そして、最適な価格探索行動として、限界収入と追加探索費用が均衡する点を最適な探索サンプル数とする「固定サンプルサイズルール」を提案した。

消費者情報探索モデルが対象とするのは、このような最適な情報探索をおこなうルール (情報探索の最適なサイズルールやそれを逐次的にサンプルしていった場合の停止ル

ール)や、情報探索による不確からさによって発生する価格分散 (price dispersion) の度合いによりその市場の効率性を評価することである。分析の対象は、労働者の職探しである労働者市場からはじまり、いわゆる「情報財」をあつかった電子商取引市場へと広がってきた。

財としての情報は、複製可能性 (ゼロに近い費用で複製可能であり、コピーしたあとでもオリジナルが原型を完全に維持する) や強い外部性 (ある情報の価値が、同じ情報を共有する人数の多寡によって影響を受ける相互依存性) をもっていて、それらによる商品の売上における「ネットワーク効果」や「ロックイン効果」の存在が指摘されている [Varian 00][Shapiro 99][永谷 02]。この特性を踏まえた上で、[Alba 97]は IHS (Interactive Home Shopping) に関する消費者行動論のリサーチトピックスを整理している。さらに、[Smith 99]は電子商取引市場で行われる消費者情報探索についても同じ枠組みを適用し、「効率的で障壁がないといわれているインターネットの電子商取引市場の効率性」を考えるために、次の4つの尺度で評価することを提案している。

- 価格水準 インターネットでは価格はより低くなっているのか
- 価格弾力性 インターネットでは価格の変化により敏感になっているのか?
- メニュー価格 インターネットではより頻繁に値札を変更するのか
- 価格分散 インターネットでは最低価格と最高価格の差がより小さいのか?

特に価格分散に関しては、それを招く原因として「製品の異質性」をあげている。さらに、製品やサービスの単純な属性の違いだけでなく、次のような要因が価格分散を生み出しているとして研究課題を設定している。

- 「(サイトの) 利便性」
(時間の価値: 商品検索の容易性などがもつ価値)
- 「(製品の存在の) 認知」
(ポータルサイトなどで、その商品がどれだけ消費者に認識されやすいかということに関するロケーション Neural Real Estate の価値)
- 「小売店のブランドと信頼」
(例えば、書籍販売における Amazon.com ブランド)
- 「ロックイン」
(景品やマイレージなどの施策によるブランドロイヤリティ形成)
- 「価格差別化」

(製品品質による価格の違い、価格比較エージェントサイト[ショッピングポット]によるライバル商品の価格比較に対応して掲示した価格を変更するマッチングポリシーなど)

特に、商品のサイトロケーションやリンク・検索の容易性は探索コストに関係し、高い探索コストは価格分散を増大させ、市場としての非効率性につながることを指摘している。

電子商取引市場での情報探索コストモデル

[Bakos 97]によると、電子商取引市場の役割は、すべての売り手は同一の製品を提供していると仮定される伝統的な商品市場に対して、売り手の存在と価格の情報を提供することである。伝統的な商品市場では探索コストを導入することで売り手による独占的なプライシングになるため、適度な探索コストでさえ本質的に限界コスト以上のプライシングにつながる。そして、電子商取引市場は、買い手にはコストがかからず売り手の価格についてすべての情報があるような古典的で理想的なワルラス的なオークションに商品市場を変化させ、寡占状態でのベルトランのゼロ・プロフィット均衡にむけて急激な値下げ競争が始まってしまう。売り手はそれを恐れてさまざまな抵抗や差別化をおこなう。

[Bakos 97]は、[Hotelling 29]の空間差異化を情報探索モデルにもちこみ、電子商取引市場での買い手と売り手のマッチングと探索コストの関係を分析した。彼のモデルは、単位円周上の属性空間に分布する財をもつ売り手に対して、買い手が一定の探索コストをかけて価格発見と適切な財を同定する過程をモデル化したものである。このモデルでは、ブランド属性は多次元空間におけるロケーションの選択としてあらわされ、消費者選好は理想的な製品ミックスのロケーションを決定し、製品属性空間の距離尺度は消費者が完全に代替的である財を買うことができない場合の効用のロスや適応コストの代用として扱われる。[Bakos 97]は買い手の期待利得と探索コストが均衡する一般均衡モデルを構築し、[Stigler 61]が前提とした品質が均質な単一財だけでなく、空間的に差異化された財においても、電子商取引市場で買い手の探索コストが下がることが市場に効率性をもたらすことを示した。

探索コストが及ぼす価格競争への影響

[Bakos 97]のモデルが示すように、電子商取引市場では垣根がなくなって、ベルトラ

ン型の価格競争が発生し、生産者はまったく利益を得ることができないのだろうか？

[Lynch 00]は[Smith 99]が提示した財の製品差別化要因（利便性、製品認知、価格差別化など）をコントロールして、差別化された製品に対する価格感応実験をおこなった。

彼らは、ある製品に対して製品情報や価格情報を得やすいかどうかということを差別化の要因と考え、Price-Usability（価格情報をすぐに利用できる環境か）と Quality-Usability（製品情報をすぐに利用できる環境か）、Store-Comparability（店舗間の価格の比較が容易に行えるのか）という要因について Web ページの階層や検索結果の提示方法の違いをコントロールすることで、情報に対するユーザビリティが「高い」「低い」の 2 水準をもった 3 種類の探索コストを設定した。彼らは、2 つの Web サイト上に仮想のワインショップを 2 店舗設けて、ワインという製品差別化された市場においてユーザに仮想ショッピングをやらせて価格弾力性を計測し、価格感応性やマーケットシェアや探索時間に対する影響を分散分析で確認している。

その結果、1) 差別化された製品に対する品質情報が手に入りにくい状況（品質に関する探索コストが高い）では顧客は価格感応性がたかまり、品質情報の探索コストを下げることによって価格感応性が減少し、2) 2 店舗間の比較を容易にすることで、価格感応性は高まり、3) 3 種類の探索コストを下げ、商品の情報の透明性を高めることで、消費者の情報環境の透明性を指標とする消費者の厚生は増加したということが判明した。

また、探索コストがマーケットシェアにおよぼす影響を測定し、店舗間比較が難しくければ、2 店舗に共通のブランドの販売シェアは流通マーケットシェアに比例するが、店舗間比較を容易にした場合は、一つの店舗に運び込まれたワインのなかでのシェアにくらべて、店舗間での販売シェアは高くなく、情報の差異がなくなったことでブランドによる効果が減少し、ノンブランドが売れる可能性があることを指摘している。

彼らは以上のような実験結果から、製品情報や価格情報への探索コストをコントロールすることで、過激な価格競争の影響度を和らげることができるということを指摘している。

市場の効率性の評価と価格分散の測定

ベルトラン競争モデルは均質財において最終的に限界費用による価格形成をおこなうため、供給者側からの独占的なプライシングの影響がない市場としては究極的に効率的

な市場モデルであると考えられている。前述したように、消費者が製品に関する情報を探索するコストはこの独占的なプライシングの要素となりうるので、それが少ないに越したことはない。[Bakos 97]のモデルは、インターネットの電子商取引市場では、探索コストが低くなるのが望ましいことは示しているが、いままでの伝統的な市場とくらべてそれが低くなるということを保証はしていない。逆に[Smith 99]でも情報探索コストが高くなると価格分散が大きくなる可能性があることが指摘されている。

インターネット電子商取引市場の効率性の実証研究を行うには、情報探索コストを直接的に計測することは難しいので、前述した効率性を計測するための4つの次元にあるように、価格水準や価格分散の度合いを測定することでベルトラン競争モデルでの情報探索コストの市場への影響度を推測し、その市場が効率的かどうかの議論をしている。

[Brynjolfsson 00b]は、インターネット市場と従来市場の CD と本の小売店でのサンプルから、インターネットの方が9~10%程度価格は安く、ヘドニック回帰により修正した価格分散はインターネットの方が低いことを確認している。CD や本といったものでは物理的な特性は均質であるので、価格分散が低いことにより探索コストはインターネットの方が低いと考えられるとしている。しかし、ブランド力のある小売店のマーケットシェアが大きくなっているので、商品認知の問題も含めて価格分散の要因となる製品の異質性に対する重要性を指摘している。

2.3 Web アクセスログ解析・Web マイニング研究

インターネットでは、Web サイトのリンク構造データやアクセスログによりデータ収集が容易かつ大量に行えるようになった。そして、Web アクセスログやリンクを解析することで、Web に対する構造的な知見を得ようというアプローチが行われている。Web サイト間の結びつきやユーザの URL 取得のビヘイビアを計測することで、その結果を検索エンジンの検索効率や Proxy サーバーのレスポンスの向上、B2C のシステムでのサイト設計に用いるなど応用範囲は広い。

トランザクション処理

ログ解析のためにデータをどのような単位で処理すべきかというトランザクション処理の方法が検討されている。[Cooley 99]によれば、セッション処理するためのもっ

とも簡単な同定方法は、ある一定時間を限度としてそれを超える滞在時間の場合にセッションが終了したと見なす方法である。彼らによると、多くの商用ログ算出ソフトウェアやアクセス調査サービス提供で採用されている基準は30分であり、実証的に調査した例では約25.5分程度との報告もある[Catledge 95]。

さらに、トランザクションを同定するためのカットオフポイント設定の考え方として、長く参照していた Web ページには意味があるとする「参照時間長 Reference Length (RL)」[Cooley 99]、あるセッションの中で以前に参照した URL に戻って参照する事象が発生した時点までを一つのトランザクションとして扱う「最大前方参照 Maximum Forward Reference (MFR)」[Chen 96]、あるセッションのなかでもっとも長く連続して繰り返して参照された列をトランザクションとする「最長反復シーケンス Longest Repeated Sequence (LRS)」[Pitkow 99]、一定時間で区切る Time Window (TW) [Cooley 99]というようなアプローチがある。

[Cooley 99]は、これらのカットオフポイントを検証し、参照時間長 Reference Length (RL) をつかって一定時間以上参照した Web ページを「コンテンツ」と考える「コンテンツトランザクション」という方法を提案した。この方法では参照された Web ページは目的とする「コンテンツ」とそれに至るまでの補助的(Auxiliary)なページに分類され、コンテンツを単位としたトランザクションが同定される。[Cooley 99]は、Reference Length、Maximum Forward Reference、Time Window のトランザクション処理アプローチの頑健性について、仮想的に生成した Web サイトのデータにアソシエーションルールを適用することで意味のあるルールをいくつ生成できたかという基準でトランザクション処理のアプローチを比較した。その結果、Reference Length アプローチが Web のグラフ構造に影響されることなく、もっとも確実に意味のあるルールを生成したと報告している。

Web マイニング

同定されたトランザクションをもとに、データ間の関係性をさぐる Web マイニングが実施される。その方法論には、関係性ルール(アソシエーションルール)[Agrawal 95]や HITS アルゴリズム[Kleinberg 98]などのほかに、マルコフモデル[Pirolli 99b]や N-gram[風間 99][佐藤 99]まで多様な試みが行われている。

[Agrawal 95]はアソシエーションルールとして Apriori というアルゴリズムを提案し

ており、アクセスログを処理して得られたトランザクションデータから Sequential な共通パターンを抽出し、それをルール化していく。URL のトランザクションのなかにある $X.url$ と $Y.url$ をともに含む相関ルールの組を $\{X,Y\}$ としたとき、対象セッションのなかに含まれる全トランザクションの中でこの相関ルールを含むトランザクションの割合を支持度と呼ぶ。この支持度の閾値として、最小支持度 (minimum support) と最小確信度 (minimum confidence) を設定し、閾値を超えるルールを求めていく。またアソシエーションルールを適用するツールとして CBA[Liu 98] などがある。

[Kleinberg 98]は、Web ページを多くのページを参照している「ハブ」と多くのページから参照されている「オーソリティ」とにわけて、各リンクの重要性を再帰的に計算することで、あるキーワードについてのサイトの重要性を示すことのできる HITS アルゴリズムを提案している。

データマイニングの方法としては、このように要素間の関係性を検討するにあたって、その間の距離や離れ具合を定義しているものが多い。C4.5[Quinlan 93]などのツールでは、基準として情報エントロピー獲得量が最大になるような基準を設けて探索木を生成する。[風間 99][佐藤 99]は 2gram の URL の相関度や共起係数を利用することで、URL 間の距離を計算して探索・レコメンデーションの効率をあげるような工夫を提案している。[Pirolli 99b]は、Xerox.com のサイトにあつまった Web ユーザのクリックパスを Web ログから計測し、次にどのパスをたどるのかを予測しようという研究をおこなっている。ユーザの探索パスをセッションごとに記録し、セッション単位でユーザ間のパターンの類似度についてレーベンシュタイン距離をもちいて計測している。これらのどの距離尺度がふさわしいかは分析対象によって異なる。

さらに、Web をグラフ的にとらえてビジュアルな形にマッピングしたりする Web 可視化の技術[柴山 01][Hochheiser 99][Dodge 01][Spiliopolou 99]もこの分野のアプリケーションと考えることができる。Web 上のドメインに展開する一群の URL を「サイト」と呼んだりするように、Web 可視化の技術はインターネットという新しいメディアが生み出した空間を把握するために、「Web 空間で喪失した経路と距離」[若林 02]を都市空間のアナロジーを用いて視覚化することによって回復する試みでもある。

具体的な方法として、HPG (Hypertext Probabilistic Grammar) という各ノード間の推移を確率で表現する方法[Borges 99]や、Web サイトのアクセスログデータからサイトの構造とユーザのトランザクションをビジュアル化する WUM (Web Utilization

Miner) [Spiliopolou 99]というツールが開発されている。

2.4 Web 生態学(Webology) 研究

アクセスログやリンク情報を単純に解析するのではなく、ユーザの行動モデルを仮定し、その行動に基づいた Web の特徴量やマクロ構造を描き出そうという試みは Web 生態学 (Webology) とも呼ばれる。Webology という言葉は、Information Ecologies + the World Wide Web の造語である [Pirolli 99b]。Information Ecology については、[Pitkow 97b]で定義されており、Ecology に関する WebSter の定義のアナロジーをつかって、「人間の集団をある情報環境に適応させる関係性に関する研究」としている。さしずめ、Webology は Information Ecology 研究を Web について展開したものといえる。Webology はおもに、Xerox PARC の研究グループで使われている言葉であるが、興味の対象としてわかりやすいので本論文でも使うことにする。

2.4.1 Web のグラフ構造

この分野での代表的な研究として、Web サーフィンなどのユーザの行動や Web サイト自体のリンク構造の特徴的な規則性が検討されている [Pirolli 99b][Huberman 98][Pitkow 98][Albert 99]。

べき乗法則

なかでもさまざまな報告がされているのが、Web でよく観察される分布のべき乗法則 (The Power Law) である。[Albert 99]は nd.edu ドメインのドキュメントをすべてしらべ、あるドキュメント (URL) が k 個の入次数*または出次数をもつ確率は、パラメータ γ をもつ $P(k) \approx k^{-\gamma}$ であったことを報告している。同様の報告として、サイトのユーザ数とおなじアクセス数をもつサイトの数の分布 [Adamic 99]、サイトの入 (出) 次数とその入 (出) 次数をもつサイトの数 [Faloutsos 99] [Broder et al. 00]などがあり、都市人口分布と順位の間に関する Zip の法則の Web 版ともいうべきものである。

上記の分布は Web サイトの物理的なリンク構造に関するものであるが、ユーザがク

* 有向グラフにおいて、ある頂点を始点とする有向辺の本数を出次数、終点とする有向辺の本数を入次数という。

リックした探索過程をあつかったものに、[Huberman 98]がある。彼らは Web サイトのなかでユーザがクリックする数の長さ L が逆ガウス分布となっていることを報告している。これはユーザが Web サイトを閲覧するときの探索クリック数ともいえるもので、それがべき乗法則のように指数系の分布をとることは非常に興味深い。

[Adamic 99]はこのようなべき乗分布の連結が生み出される背景を考察しているが、[中井 97]は[Simon 55]が都市のサイズと順位とがべき乗分布にしたがうことを説明した方法のアナロジーをもとに、ホームページの規模分布をシミュレーションし、実際 ISP のホームページ規模分布と比較しながらホームページの成長過程をモデル化している。

Web の直径: スモールワールド仮説

[Albert 99]は N 個の頂点をもつ有向ランダムグラフを作成し、2つの頂点間の最短パス d を計算した結果、その平均 $\langle d \rangle = 0.35 + 2.06 \log(N)$ となることを報告している。Web は広大な探索空間であるが、 $\langle d \rangle$ が N の対数関数に依存しているので、 N が大きくなってもサイズはあまり変わらない。

[Albert 99]は $\langle d \rangle$ を Web グラフの直径と呼び、大抵の Web サイトでは平均 20 リンク程度でコンテンツにたどり着ける構造になっていると報告している。彼らはこれをスモールワールドネットワークとも呼び、人間社会においてお互いの知人をたどると案外簡単に共通の知人を見つけることができることとの類似性を指摘している。

Web のマクロ構造*

Web のマクロ構造に注目したものとして、[Broder et al. 00]が興味深い報告をおこなっている。具体的には、おおよそ 2 億ページの 15 億のハイパーリンクを調べ、ある入出次数をサイトがもつ確率がべき乗法則 (The Power Law) にしたがっていることを検証したあとに、サイトのコネクティビティをしらべ、有向グラフの強連結集合 Strongly Connected Components (SCC) を計算し、Web 空間のマクロ構造を描き出している。巨大な SCC とそこに連結するサイト、そして SCC から連結するサイト、孤立するサイトなどの巨視的な特徴が描き出されている (図 2-1)。

* ここでいうマクロ構造とは、Web サイト内のページの HTML リンク構造を最小単位だと考えたときに、インターネット全体を対象にして、Web サイトを越えて凝集されたリンク構造が大規模なクラスターを形成しているようなケースを意味している。本文中の「マクロ構造」という表現はこれを意味する。

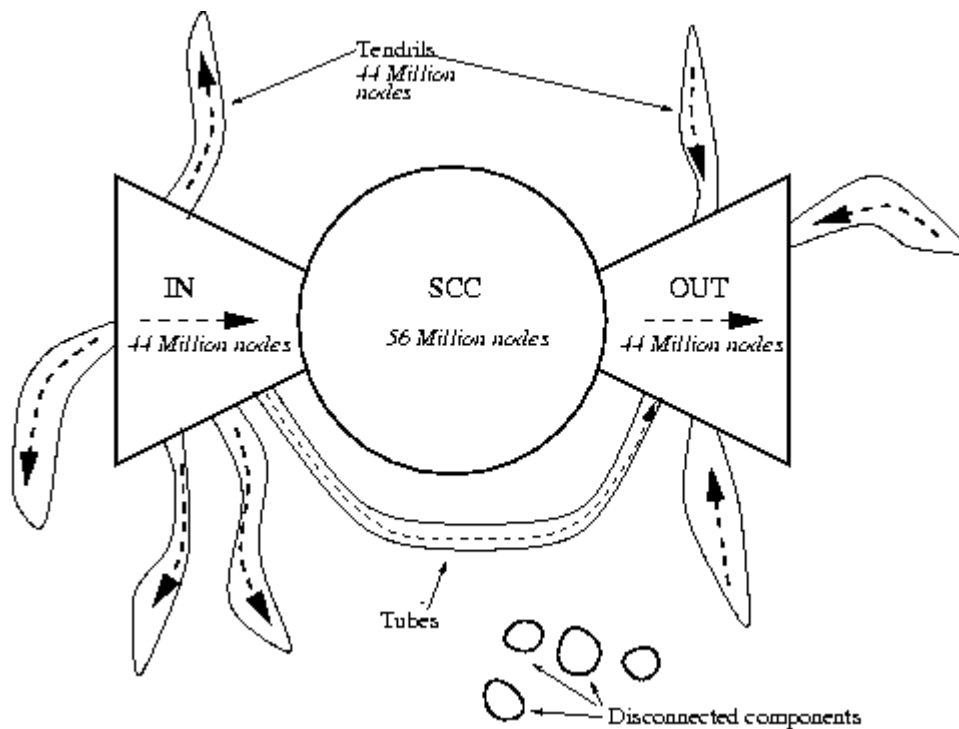


図 2-1 Web のコネクティビティの図示[Broder et al. 00]

特徴的なのは、中心となる SCC に対してほぼ同数の流れ込むノード (IN) と出て行くノード (OUT) が存在していることである。そして、同数のノードが SCC 以外の方に流出して (Tendrils)、SCC に流れこむノードと SCC から出て行くノードをつなぐ経路が見つかった (Tubes)。その各最小経路の平均距離は無向グラフでの計算では約 7 であった。

このように、無限に広がるように思われた Web 空間が、SCC を中心とした凝縮した大きな流れのなかで、比較的小さな世界スモールワールドにまとまっている様子が窺い知れる。この Web 空間で情報を探索すると、Web 空間にある情報については大きな探索努力なく「出会う」可能性が考えられる。その意味でこの Web 空間はいろんな情報が比較的アクセスしやすいようになっているとも思われる。しかし、この空間の中で欲しいと思う情報を探索する努力をどこまで人々が注入しているかどうかは、調べられたものが静的なハイパーリンクのグラフであるためにわからない。そのためには、Web 空間を探索しているアクセスログを大量に調べる必要がある。

2.4.2 情報取得行動と情報採餌行動

情報取得行動

Web などの CME (Computer Mediated Environment) 環境下である目的 (たとえばある省庁のサイトを見つけるなど) をもった行動を行うときに、その環境下の情報をどのように利用しているかという情報取得行動についても研究が行われている。[野島 00]は現実空間と Web 空間のアナロジーについて触れ、Web 空間で目的とするコンテンツにたどりつけないで「迷う」現象をとらえて、Web 空間を移動しているときに頭のなかで認識される「認知地図」を手がかりに、現実空間とのナビゲーションの違いについて研究を行っている。

また、[中村 01]は、学生の Web 検索行動とその学生のアクセス時間などの属性をアンケートと Web ログの両面から検討し、検索時のキーワードの使い方と個人の属性との関係を論じている。これらの中では、バーチャル環境と実空間の行動の共通点や相違点、行動過程に及ぼす影響因子などについて言及されている。そして、ポータルサイトの影響度や探索深度といった HTML ナビゲーション上の違いに着目して、それがユーザの習熟度の違いによって異なることが[中村 01]でも指摘されている。

さらに、[Hoffman 96a]はインターネット上でのハイパーメディア環境での消費者行動を捉える方法論として、「フロー」という概念を用いることを提案し、Web 空間のナビゲーション行動をメディアリッチネス[Daft 86]や本人のスキル・環境を要素とした共分散構造分析により検証している。

さらに、CMC の代表的なものとして、電子コミュニティ内で個人がメンバーと行う一対多のコミュニケーションネットワークモデルが分析されている。主に電子メールのメーリングリストや電子会議室、電子掲示板といったアプリケーションを対象にして、個人間でとりかわされた発言をもとにコミュニケーションを分析することで、組織内部での発言者の役割の発見や知識共有の方法が議論されている[小林 96][池田編 97]。手法としては社会ネットワーク分析[安田 01]として発言のやりとりと頻度をグラフ構造[高橋 99]にして表現することもしばしばあり、「ソシオマトリクス」とも呼ばれる隣接行列などでコミュニケーションを表現し、そのネットワークの特徴量や指標を算出していくことが多い。

情報採餌行動

情報を捜し求めて発見し、それを処理して次に移るという Web なかでおこなっている情報取得行動は、一見では数理生物学でいうところのパッチでの捕食行動[巖佐 98]と考えることができる。[Pirolli 95]はこれを情報採餌行動 (Information Foraging) として扱い、大規模データベースやオフィスのデスク上の情報探索過程などをモデル化している。たとえば、Web 空間のアナロジーでは、Web 空間に多数存在する Web サイトをパッチと捉えて、人はそのパッチ (サイト) 間を移動して情報を探索し、サイト内でその情報を消費すると考える。

このようなモデル化を行った場合に、パッチの中で獲得する情報の長時間平均を最大化するように生物が行動すると考えると、各パッチでの獲得価値の限界獲得率は獲得情報の長時間平均に等しくなるという Charnov の定理 [Priolli 99a] が確かめられている。これは探索計画のなかで探索努力を最適配分するという問題における限界効用原理 [多田 73] とおなじ意味合いをもつ。

この情報採餌行動モデルを Web 空間のアナロジーに適用した場合には、コンテンツの数やリンク構造というサイトの環境下での制約条件のもとで、効用関数としての情報獲得関数、ユーザの探索時間に対する機会費用、ユーザの処理時間に対する機会費用を包含したコンテンツ探索行動をモデル化することが可能と思われる。

情報痕跡 (Information scent)

情報採餌行動をおこなうときに、ターゲットとなる情報までにたどり着く過程を Information scent (情報痕跡) と呼ばれている [Card 01][Pirolli 95]。Information scent は、WWW リンクや参考文献の引用などといった隣接した流れのなかで、情報源の価値やコストやアクセスパスをどのように受容しているかということを表している。[Card 01]は、14 人の被験者にあるサイトの探索という目的タスクをあたえ、視線検出とクリックスルーを同時に測定する装置と連動させて、この Information scent を測定している。そして、1) Web 探索空間はパッチ構造のような特徴をもっている、2) 探索パターンはヒューリスティックに行われている、3) Web ヒューリスティック探索では一つの問題空間だけではなく、複数を組み合わせておこない、URL 探索 (URL 自体を探索、検索する)、リンク探索 (リンク自体をたどっていく)、キーワード探索 (キーワードを次々と入力して探す)、ビジュアル探索 (1つのページ上で、目で見

探す)ということを行っている。4)バックトラッキングするときはランドマークまで戻る、といった特徴を抽出している。

このような方法での行動パターンの表現は詳細なデータを取得することが可能でわかりやすい反面、被験者に対して必ず目的を設定する必要があるので、一般的なログデータを用いて目的行動かどうかを判別することは難しい。しかし、情報取得行動がサイト・ドメインを中心としたパッチ構造をもっているという知見は重要な指摘である。

なお、ここで扱っている対象は情報取得行動で触れたトピックスと近いが、対象へのアプローチが微妙に異なる。前述のモデルではメディア自体での人々の活動やそこで頭の中に描かれている認知マップに興味があるのに対して、このモデルでは残した足跡自体の関連性にフォーカスされている。

2.5 関連研究分野のまとめと課題

本章では、『Web を代表とするインターネットは「空間と時間の壁」を克服したのか』という本研究の命題についてユーザの URL へのアクセス行動を経路的な情報として扱うことでアプローチするために、インターネットマーケティング・エコノミクス研究と、Web の構造を分析する研究の両面から関連する分野を概観した。それらを踏まえて、具体的な分析に入る前に、本研究と照らし合わせて各分野の課題を整理する。

2.5.1 インターネットマーケティング・エコノミクス研究における課題

消費者情報探索モデルと探索コスト～具体的な探索コスト計測

消費者情報探索行動は、おもに情報の経済学の分野での成果を踏まえて、実際の伝統的な市場とインターネット市場での効率性についての議論にフォーカスされている。彼らの研究の視座は、インターネットでの電子商取引が時間と空間の壁をこえたインターネットのコミュニケーション行動によって効率的な完全競争市場になっていくのかという点にある。分析では、[Bakos 97]の探索コストモデルをもとに、[Lynch 00] や [Brynjolfsson 00b]が、均衡価格の分散やインターネット販売と実際の価格の違いなどを検証している。価格分散がネット販売の方が小さいことや価格がネットの方が低くなっていることなどといった具体的な結果から、インターネットでの電子商取引は効率性

をもたらしているが、探索コストの存在やブランドなどによる製品の異質性により、必ずしも理想的な完全競争をもたらすようなことにはならないとしている。

この分野の研究は、本研究の命題に対する経済学分野のアプローチとして注目すべき結果を提出している。探索コストやブランド力などにより消費者の電子市場での情報探索行動のモデル化が可能であること、空間や時間の壁を克服したといわれるなかでも、やはり製品に異質性が生まれることで別の意味での壁が存在していることを指摘している。ただ、彼らの実証分析では「情報探索コストが高いと価格分散が大きくなる」という消費者情報探索理論の知見から、価格分散を測定することで効率性を議論しており、直接探索コストを計測した議論はしていない。[Lynch 00]にしても、探索コストの大・小を探索しやすいのかそうでないのかという基準に置き換えて議論している。

さらに、探索コストの扱い方に注目してみると、[Bakos 97]のモデルでは探索コスト一定という仮定のもとで議論されている。議論のなかではフィットコストを持ち出すことで幅をもたせているが、情報探索にかかるコストが常に一定であるという議論は不自然である。[Kephart 99]はさらに、探索コストが可変で非線形な場合を探索エージェントのコストモデルのなかで議論しているが、一つの帰結を得るというより、複雑性がますますという状況を説明したにすぎない。

彼らの議論のなかでは、探索コストを一つのブラックボックスとして扱い、間接的に電子商取引の情報探索行動がフリクシオンレスであるのかどうかを調べているので、本研究の命題を検討するうえでは、具体的な探索コストの扱い方とアクセスの関係を明示していくことが課題になっていく。それが可能になれば、探索コストをキーにして、[後藤 93]が提起したような電子商取引市場での市場構造について実証データをもとに説明することができるようになる。

空間相互作用モデル～距離概念の構築

空間相互作用モデルは、いままでの過去の商圈分析やトラフィックモデルなどでは十分な成果はあるが、インターネットの世界では[豊島 00]の例がある程度で具体的な成果は残していない。それはモデルの構造として、距離に相当する変数を定義できていないからである。通常の商圈に相当するモデルでは2つのロケーション間の空間的な距離を定義することができる。普通は、それが直線距離であったり、交通手段を利用した場合の距離であったりする。

この問題は本研究の重要なトピックスになる。時間と距離を克服しているとする、ここでいう空間的な距離は定義できない。しかし、消費者情報探索モデルでみたように、インターネットをつかった電子商取引のなかでも探索コストという概念は探索行動のなかで重要な役割を果たしている。買い手の探索コストが増大しすぎると、市場自体が成立しなくなってしまうように、探索コストは消費者情報探索においてその行動を左右する重要な変数となる。前述したように、消費者情報探索モデルでは探索コストを一定として留保効用との大きさのバランスを議論しているだけなので、探索コストを測る方法論が必要になる。そのためには、Web のなかの距離をどのように考えるかが重要な議論になる。

2.5.2 Web アクセスログ解析・Web マイニング研究における課題

Web アクセスログ解析～コンテンツトランザクションの探索過程への応用

この項では、データクリーニング・ユーザの同定・セッションの同定・トランザクションの同定というステップのなかで、Web データのアクセスログを解析するために必要な過程と処理の方法論を概観した。本研究の命題のためには、ユーザが特定されたデータが必要になる。ある特定の Web サイトだけではなく、複数の Web のサイトを探索してまわる行動が Web での探索行動ではよくみられることを考えると、ユーザをパネルデータとして持つておく必要があり、その行動を追跡する必要がある。幸い、ユーザの了解をえることができればそのようなソフトウェアをブラウザと連動させることで、データの収集は可能になる。

むしろ、分析においてトランザクションをどのように同定するかが重要である。Web での行動をどのように定義し、なにを調べたいのかによって採用する方法論はかわる。特に、Maximum Forward Reference なのか、Long Repeated Sequence なのか、Reference Length なのかという選択は重要である。データマイニングのための方法論としては、[Cooley 99]が示したコンテンツトランザクションにおいて、Reference Length を用いたトランザクション処理をするのが有効なようである。[Cooley 99]では、アソシエーションルールを用いるデータマイニングのためのトランザクション処理として、補助的 (Auxiliary) な URL を削除し、トランザクション自体と選択肢の情報を縮退することで、ルール生成・発見の効率をあげることが目指されていた。しかし、見

方を変えてみると、コンテンツの URL と補助的 Auxiliary な URL というトランザクションは、あるコンテンツを探索している過程を表している。探索目標のコンテンツにたどり着くまでにどれくらいの時間をかけ、どれくらいの URL を探索したのかという情報を量的に把握することが可能になる。このコンテンツトランザクションを用いることで前述した課題の探索コストを定量的に計測・分析できる可能性がある。

Web マイニング～パターン抽出の方法論

Web データから何らかの知見を得る方法と、それを分別するための距離やパターン化の手法を概観した。パターン間の距離概念は、ユークリッド距離からレーベンシュタイン距離などの符号距離や類似性尺度までさまざまな定義が可能である。本研究で課題となっている Web アクセス行動を経路的に扱うという点ではこれらの手段は意味があるかもしれないが、複数のサイトで複数のユーザの経路情報からある意味のあるパターンを読み取るのは難しい。そのためには、新しい計測手法が必要である。

2.5.3 Webology 研究における課題

Web のグラフ構造～アクセスログの活用

Web をグラフ構造ととらえるアプローチには、べき乗法則や Web サイズの測定といったグラフの理論をもちいた Web 構造を記述する方法論が提出されている。ここで用いられている距離は URL をグラフの頂点とした頂点間の最短パスの平均値である。グラフの構造は基本的に頂点と辺で構成された経路であるので、ある意味単純である。その接続の状態を強連結集合への接続状態で表し、Web のマクロ構造を描き出した [Broder et al. 00]の研究は興味深い。彼らの分析で用いられているデータはあくまで、静的な Web のリンク情報であるので、実際の人があるリンク情報のとおり探索するかどうかはわかっていないが、探索する上での選択肢としては、探索エージェントを利用しない場合には、このリンク情報以外にはない。

しかし、実際のユーザがおこなう探索行動はほとんどの場合になんらかの探索エージェントを利用している。そのため、静的なグラフ構造で議論された測地的 (geodesic) な距離ではなく、探索エンジンなどをつかってでてきた候補から瞬間的に目的地に移動している可能性が高い。この静的なリンク情報では、本研究の命題をさぐることにはで

きないが、実際のアクセスログデータから計算されるグラフ構造には探索エンジンなどをつかった過程が織り込まれた接続状態になっている。本研究のためには静的なグラフ構造にアクセストラフィックを加味する必要がある。

情報採餌行動モデル～探索行動モデルへの応用

情報採餌行動モデルは、数理生物学のアナロジーから情報を探索・消費していく行動をモデル化しており、興味深い。なにより、彼らのモデルのなかでは探索している過程が時間にそって明示されており、前述の消費者情報探索における探索コストの課題を解決する方向性が示されている。ただ、彼らのモデルでは情報を獲得する獲得関数の設定がはっきりしていない。あくまで効用関数として効用逓減型のものを想定した議論がおこなわれており、それにもとづき、限界効用原理にもとづく長時間平均の獲得率最大となる行動を示している。人間の合理性に限界があり、限定合理性のもとでの最適行動が議論されるべきかもしれないが、ネットでの効用をある関数だけで記述するのは十分な結果を得られるとは限らない。しかし、うまく状況をコントロールすることで、ネットでの行動を説明するためのツールとしては機能することが考えられる。

2.5.4 この章のまとめ

本研究の命題と関連分野の課題を整理することで、具体的な分析にむけて以下のような検討が必要になる。

- インターネットマーケティング・エコノミクス分野では、消費者情報探索モデルの考え方において、製品の異質性やサイトの利便性や製品ブランドに対する信頼性などの要素を「探索コスト」が代替して表現することで、市場の非効率性を示す価格分散と関連付けられた説明が試みられている。これを一歩進めることで、「探索コスト」によりインターネットに距離や時間の壁が存在することを説明できる可能性が考えられる。そのためには、具体的な「探索行動」の定義と「探索コスト」の定量的な表現方法が課題になる。
- Web アクセスログ解析・Web マイニング分野では、コンテンツトランザクションにおけるトランザクション処理が試みられているが、この考え方をユーザのURL へのアクセス行動に応用することで、コンテンツ探索行動として経路的に処理する方法論が構築できる可能性がある。具体的なトランザクション処理の方

法論の構築が課題になる。

- Web 生態学の分野では、Web グラフ構造において、静的なリンク情報を用いた Web グラフのマクロ構造やサイズの分析に一定の成果が見られるが、本研究のためにはユーザの URL へのアクセス行動を反映した分析を行う必要がある。そして、情報採餌行動モデルにおいては Web をパッチに見立てたモデル化が行われているが、本研究のためには探索コストを取り込んだ探索行動に応用する必要がある。

次章では、以上の課題の具体的な解析のまえに、まずインターネット自体にアクセスしているときのアクセス手段の選択行動を分析することで、アクセス行動を支配している要因から「探索コスト」を考察するうえで必要なインプリケーションを探ることにする。

部

部では、アクセスサービスの対価としての探索コストを検討するため、インターネットアクセスサービス選択構造を分析することで、コンテンツ探索行動の鍵を握る探索コストに影響する要因を調べる。

第3章 ユーザのインターネットアクセス手段の選択行動

本章では、インターネットアクセスサービスの対価としての「探索コスト」に着目し、ユーザがインターネットを探索している時にどのような要素に影響を受けているのかを調べることで、4章における仮説導出に寄与するようにする。具体的には、ユーザがインターネットにアクセスしている時に利用しているISPとアクセス回線の選択構造をモデル化し、その効用関数のパラメータを推定する。まず、ロジットモデル・ネスティッドロジットモデルをつかった分析手法について述べ、アンケート調査データの概要を検討する。そののち、具体的な分析モデルとして、ISP選択のロジットモデルとアクセス回線とISPを組み合わせたネスティッド・ロジットモデルを推定することで選択構造をあきらかにする。さらに、パラメータの推定結果から探索コストとアクセス手段の選択構造の関係を考察する。

3.1 本章の目的とアプローチ

目的

本章の目的は、インターネットアクセスにおけるユーザのアクセス手段（ISPとアクセス回線）の選択行動の分析から、インターネットアクセスにおけるユーザ効用が何に依存しているかを確かめることで、インターネット上での情報の探索コストに影響する要素を検討することである。前章ではユーザのURLへのアクセス行動をモデル化していく方法論を概観したが、中でも消費者情報探索モデルでは「探索コスト」に着目した研究が行われていた。この探索コストは、ある（商品）情報を探索するためにつかわれる交通手段やサービスへの金銭支出とそれに要した時間に対する機会費用である[成生94]と考えることができる。インターネットにアクセスしているときには常にコンテン

ツをもとめて探索しているとする、アクセスサービスの選択はサービスの対価としての探索コストと関連があると考えられる。目的とするコンテンツを探索しているときの実際の探索コストを直接調べることは難しいが、インターネットアクセスコストに関する消費者の態度から探索コストに対する態度を推測することが可能である。

このように探索コストに対する影響を推測することで、第4章において本研究の命題を検証するための仮説導出の基礎となる探索コスト概念形成に寄与することを目指す。

分析アプローチ

分析アプローチとしては、[Train 87]や[Brynjolfsson 00a]でも採られていたロジットモデルによる離散的行動分析アプローチを用いた。消費者行動分析モデルについてはロジットモデル以外にも多数ある[片平 87]が、1)想定される選択肢の中の属性間で効用が補償的な構造になっているおり、属性間の比較が比較的容易なこと、2)[Brynjolfsson 00a]での例との比較が可能なこと、3)インターネットアクセスを交通手段と見立てると交通手段選択等での分析としてよく利用されている事例が多いこと等の理由からロジットモデルを利用した分析モデルを構築することにした。

ロジットモデルの効用関数を調整することにより、インターネットで探索を行っているときにかかる探索コストのなかで、実際にアクセスサービスの対価として発生する費用等の属性に対するユーザの反応を測定することができる。

さらに、選択手段はアクセス回線とISPによって提供されることから、IIA(文脈からの選択確率の独立)特性を緩和する意味でも両者の選択構造をネストにいったネステッド・ロジットモデルを構築し、アクセス回線サービスとISPの選択構造を検証する。これにより、アクセス回線の選択も考慮にいった探索コストの影響を推測することができる。

また、このような効用関数の推定はインターネットでの探索コストの推測に影響をもつほか、特に、帯域、料金、時間、物理的なサービス提供地域やアクセスポイントのカバーといった接続サービスの属性を設計するうえで、ISPや回線事業者が今後のサービス戦略オプションの検討する際にも役立つ。

本章ではまず、3.2節でインターネットアクセス手段選択に関連した研究を概観し、3.3節でデータをハンドリングするアプローチについて述べる。3.4節で将来選択したいサービスについての多項ロジットモデルを推定する。3.5節ではアクセス回線とISPを

組み合わせたネスティッド・ロジットモデルを推定して両者の選択構造を検証する。3.6節では効用関数の特徴や選択構造と探索コストの関係についていままでの議論をまとめ、考察を加える。

3.2 関連研究

ネスティッド・ロジットモデルをネットワークサービスに適用した事例として、[Train 87]と[Brynjolfsson 00a]を取り上げる。両研究については前章のインターネットマーケティング・エコノミクス研究でも取り上げたが、ここでは定式化を含めて詳細にレビューする。

ローカル通話サービスでのネスティッド・ロジットモデルの適用事例

[Train 87]はローカル通話サービスにおいてネスティッドロジットモデルを適用した分析をおこなった。ローカル電話での住宅用 (residential) サービスには2つのカテゴリーがあり、一つはある特定地域や時間帯での通話に対して通話数の上限のない定額料金を設定した flat-rate service で、もう一つは通話数の上限があるがそれまでは割引された通話料金が適用され、上限を越えた場合は追加料金がかかる measured service である。さらに、flat-rate service はその通話地域の数や範囲によりさまざまなサービスにわかれ、measured service も通話数の上限のレベルにより異なったサービスが設定されている。これらのサービスは顧客の通話ポートフォリオ (いつ、どこの地域へ、どれくらいの長さ通話するかということ) と関連している。

彼らは、住宅用ローカル電話の顧客のサービスオプションと通話ポートフォリオの相互選択を記述するモデルとして、ネスティッド・ロジットモデルを使用し、その効用関数を推定することでそのサービスオプション間の代替可能性 (substitutability) や価格弾力性を測定している。

ある顧客が t 時に z 地域に通話した通話数を N_{tz} 、そのときの平均通話時間を D_{tz} とすれば、その顧客の通話ポートフォリオはベクトル $(N_{11}, \dots, N_{TZ}, D_{11}, \dots, D_{TZ})$ と記述される。選択可能なポートフォリオ A のなかのあるポートフォリオを $i, j \in A$ とする。また、利用可能なサービスオプションを $s = 1, \dots, S$ とする。その組み合わせは (s, i) であらわされる。ロジットモデルでは、通話ポートフォリオ i やサービスオプション s を組み

合わせた選択確率 $P(s, i)$ は効用関数 Y_{is} を選択肢集合全体の効用のなか評価した割合で記述されるが、ネスティッド・ロジットモデルでは、通話ポートフォリオ i を選択する確率とその場合にサービスオプション s を選択する条件付き確率の積として、

$$P(s, i) = P(s | i)P(i) = \frac{e^{Y_{is}} \left(\sum_s e^{Y_{is}} \right)^{\lambda-1}}{\sum_j \left(\sum_s e^{Y_{js}} \right)^{\lambda}}$$

と記述される。このとき、

$$P(s | i) = \frac{e^{Y_{is}}}{\sum_s e^{Y_{is}}}$$

$$P(i) = \frac{\left(\sum_s e^{Y_{is}} \right)^{\lambda}}{\sum_j \left(\sum_s e^{Y_{js}} \right)^{\lambda}}$$

である。ここで、 Y_{is} はサービスオプション s とポートフォリオ i の両方の選択に影響される項 w_{is} とポートフォリオのみに影響される項 v_i にわけて記述される属性で、

$$Y_{is} = w_{is} + v_i / \lambda$$

と書くことが可能で、これにより、

$$P(s | i) = \frac{e^{w_{is}}}{\sum_{s=1}^S e^{w_{is}}}$$

となる。また、 $I_i = \ln \left(\sum_{s=1}^S e^{w_{is}} \right)$ とすれば、

$$P(i) = \frac{e^{v_i + \lambda I_i}}{\sum_{j \in A} e^{v_j + \lambda I_j}}$$

とかける。このとき、 I_i はポートフォリオ i の包括的な価格 (inclusive price) である。

このときのパラメータ λ は代替サービス間の代替性を測定したものと考えられ、ポートフォリオ間の代替性より同一のポートフォリオのなかでのサービスの代替性の方が大

きければ、 $0 < \rho < 1$ となる。逆にポートフォリオ間の代替性の方がたかければ、 $\rho > 1$ となる。 w_{is} はポートフォリオ*i*でサービスオプション *s* を利用した場合のコスト C_{is} に依存する定数として記述されている。

また、 v_i は少々複雑で、簡単のために*i*を省いて記述して、

$$V = \theta N \log \phi D - \alpha ND = \theta N \log D - \gamma N - \alpha ND$$

where, $\gamma = -\theta \log \phi$,

と定義されている。 V は通話から得られる効用を定式化したもので、 N は通話回数、 D は平均通話時間でデータとして与えられる。つまり、その平均通話時間 D に対して、 θ のレートで情報が意味のあるものとして変換され、それが通話回数 N に比例すると考えている。それによって、パラメータ θ はこの通話から得られる情報の利益を計測しているものとして考えられている。それに対して、その時間を別のことにあてれば得られる機会費用を γN として、それをさきほどの利益からマイナスしている。

以上のような構造の効用関数を想定し、サンプルからそれぞれのパラメータを推定している。推定にあたっては、サンプルにおける対数尤度関数を設定し、最尤法によって推定している。

実際の推定では、ログサム変数を通した包括的な価格(the coefficient of inclusive value)は1を越えた。これにより、顧客は価格変化に対してサービスオプションを別のものに変えるより、通話パターン(通話ポートフォリオ)を調整することで対処していることがわかるとしている。

さらに別途求めたサービス選択の価格弾力性より、通話料金値上げに対しては、通話料金自体をさげる方向で対処していることもわかるとしている。月間の通話料金に対してサービスのオプションは高い価格弾力性をもつので、通話料金の値上げは、多くの顧客がそれにあわせて flat rate service などに移行しやすく、ローカル通話収入に関してはマイナスの効果があることを指摘している。

なお、この研究には[Taylor 94]による詳細なレビューがある。

インターネットショップボットの消費者選択行動分析

[Brynjolfsson 00a]は本を扱う EvenBetter.com (現 DealTime.com)における消費者の選択行動を分析した。このサイトでは、ある商品を選択すると、同じ商品を取り扱うネット上の小売店のデータを提示する。提示されるデータは、小売店名、トータル価格、

商品価格、配送コスト、販売税（州）、発送時間、仕入れ時間、配送時間（発送時間+仕入れ時間）、発送方法などである。ここに表示されたデータを吟味してユーザは小売店を選択して発注する。

彼らは、1999年の8月25日から11月1日までの69日間の合衆国ユーザの7478人でのべ1,513,439件に及ぶ本の要求のデータをつかって、ユーザのクリックストリームの追跡もおこなった。各小売店候補に対する選択行動として、クリックスルー（クリックしたかどうか）、ラストクリックスルー（最後にクリックしたかどうか。選択したか）を計測し、ブランドロイヤリティを知るために、ユーザをクッキーで認識して、前回の買い物のおきに同じ店舗をクリックスルーしたか、ラストクリックスルーしたかを測定した。そして、ショッピングボットを利用したユーザの小売店選択行動を規定する効用関数をロジットモデルやネスティッド・ロジットモデルをつかって推定した。

構築したロジットモデルは、セッション i でのショッピングボットがだすオファ- t が選択される確率は

$$P_i(x_i, \beta) = \frac{e^{\mu\beta'x_{it}}}{\sum_{\tau=1}^{n_i} e^{\mu\beta'x_{i\tau}}}$$

で表現され、 x は各商品の属性ベクトルで、 β はそれに対するウェイト、 n_i はそのセッション時の選択肢の数である。各商品の属性変数としては、総価格（アイテムの価格、発送費用、税のダミー変数）、平均配送時間、配送ダミー、ブランド小売店ダミー（Amazon, Barnes and Noble, Borders）、前回のラストクリック、前回のクリックなどである。

推定の結果、総費用を始めとする価格・配送時間にはマイナスの効用を、店舗ブランドや前回選択した店舗の経験値については高いプラスの効用が測定された。書店のブランドネームに対する反応は、ブランドがない店に比べて他の条件が同じなら

$$\Delta p = \frac{-\beta_{BRAND}}{\beta_{PRICE}} = 1.13 \text{ ドル有利であり、さらに、過去の購買をした店 (Prior Last$$

Click) についても平均 2.49 ドル以上のアドバンテージがあるなど、ブランドロイヤリティは高かったと報告されている。

また、彼らはブランドロイヤリティを基準にした効用を測定するために、ブランドの有無をもとにしたネスティッド・ロジットモデルを構築した（図 3-1）。

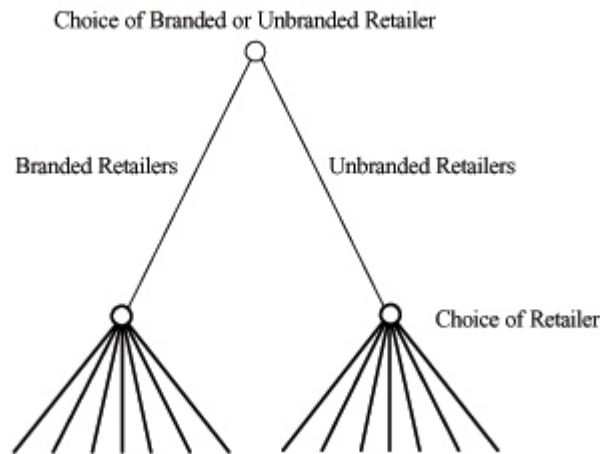


図 3-1 ネスティッド・ロジットモデルの構造[Brynjolfsson 00a]

その推定結果は、提示候補が最も低い価格帯に入っているかどうかによって強烈に依存している、さらにブランド力の有無にも依然として強い選好が現れているものであった。

価格に対する反応としては、トータルコストには敏感であったが、なかでも配送コストに敏感であった。配送時間のような契約不可能な要素については、ブランド力をその店に対する信用の代用として用いていることが考えられるとも報告されている。

消費者行動選択において、ロジットモデル等を用いてその選択効用関数のパラメータを推定することで、各要素に対して今後どのような戦略的な商品属性戦略、価格差別化戦略をとることが考察できるようになっている。

3.3 分析モデルのフレームワークとデータ

モデル構築のターゲット

この章の主要課題は「探索コスト」の環境を構成しているインターネットアクセスサービスの選択構造を調べることで、探索コストモデルを構築するための含意を得ることである。

具体的には、ISP の提供するインターネットアクセスサービスと回線事業者が提供するアクセス回線サービスの選択をモデル化したロジットモデルを構築し、その効用関数を測定することでインターネットにアクセスしているときの効用が何に依存しているの

かを検討する。

この ISP のサービスに対する効用関数はユーザの経験値や嗜好によって変化することも想定されるので、利用頻度の違いをユーザ属性に対応させて、ヘビーユーザとノーマルユーザにわけたモデル構築も試みる。

変数の考え方

ロジットモデルによる効用関数の推定においてはサービス属性の選択が重要である。探索コストを「ある（商品）情報を探索するためにつかわれる交通手段やサービスへの金銭支出とそれに要した時間等に対する機会費用である[成生 94]」と考えれば、インターネットアクセスサービスの対価としてのユーザ利用料金と機会費用としての時間の両方を組み入れることができればいい。

しかし、[Train 87]では、効用関数には機会費用（平均通話時×利用回数）が組み込まれているが、直接ユーザの利用料金に対する選好は組み入れられていない。これは通常のユーザ料金が通話ポートフォリオによって規定されているため、料金と通話ポートフォリオの両方を盛り込むと効用関数の変数間の独立性が保てなくなるからだと考えられる。機会費用と実際のユーザ利用料金が通話ポートフォリオと割引料金オプションを通して影響するので、これらを同時に組み入れたモデルを構築するのは妥当ではない。

また、[Train 87]のネスティッド・ロジットモデルでは一つの事業者において割引オプションより通話ポートフォリオの方が選択構造上は上位の包含関係にあると推定された。インターネットアクセスにおいては回線サービスと ISP サービスを提供する事業者が異なり、サービスの選択過程についても調査当時は別々に行なわれていたので、彼らの選択構造を直接使うのではなく、サービスを提供する ISP と回線事業者の事業構造を利用し、それぞれの総費用の中で割引オプションと料金ポートフォリオを同時に組み入れて検討する。料金制度については従量性と定額性があるが、それについては制度扱ったダミー変数として扱うことで対処する。

[Brynjolfsson 00a]の例では、効用関数は総費用とブランド属性に対する強く依存していたので、それらの要素は本研究でも必要だと思われる。それ以外の属性として、本の配送時間など付加的なサービス属性が影響していることを考えると、アクセス回線や ISP サービスのサービス品質に関わる変数を取り込む必要もあると考えられる。

ネスティッド・ロジットモデルの構造

[Train 87]は割引サービスオプションと料金ポートフォリオの選択の包含関係を検証するために、両者を組み合わせたネスティッド・ロジットモデルを構築したが、前項で述べたように、本研究では割引サービスオプションと料金ポートフォリオは同時に扱う。

[Brynjolfsson 00a]では、ショッピングボットが提示した料金とサービススペックを変数属性としてもつ店舗を選択肢集合として扱いながら、ブランドの影響力の有無がもたらす選択構造要因を推定するためにネスティッド・ロジットモデルを構築した。本研究ではブランドの影響力に関心があるのではないので、このような方法はとらない。

インターネットアクセスサービスを選択する場合には、そのまえにユーザの環境には電話サービスが導入されていることが多い。そのため、アクセス回線は事前に選択されていることが多く、その環境を意識しながら、ISPのサービスを選択している可能性を考えると、アクセス回線とISPサービスのネスティッドな選択構造を想定するのが自然である。

また、ロジットモデルには IIA (文脈からの選択確率の独立) 特性があるので、推定されたモデルを新サービスの予測などには利用するのは現実的には難しい。

そこで、回線事業者のアクセス回線サービスとISPのインターネットアクセスサービスを組み合わせたネスティッド・ロジットモデルを構築することで、実際の選択構造と IIA 特性を緩和したモデルの推定をおこなう。それにより、新サービスについてを開発する際の示唆が得られるほか、アクセス回線とインターネットアクセスサービスの両方に関係し、インターネット上の探索コストに影響する要素を推定することができる。

選択データの構造

分析に使用するデータは、アンケートを利用して各個人のISPとアクセス回線の選択肢に対する選好データを収集する。また選択肢については、世の中のすべてのISPとそのサービスの組み合わせを提示するわけにはいかないため、妥当な変数選択ののちに選択肢プロファイルを作成して提示した。詳細は次節の分析のなかで述べる。

実際の分析では、図 3-2に示されるように、個人 n が行った選択肢の属性データに対してそれが選択されたかどうかを示す 0-1 の変数があり、個人 n の選択結果としては選択された選択肢のみ 1 で他の選択肢は 0 というデータを作成して、最尤法として P_{in} の同時確率である尤度関数が最大になるようにパラメータを推定する。

個人が複数の選択肢からある選択肢（サービス）を選択するかしないかを数学モデルで表現

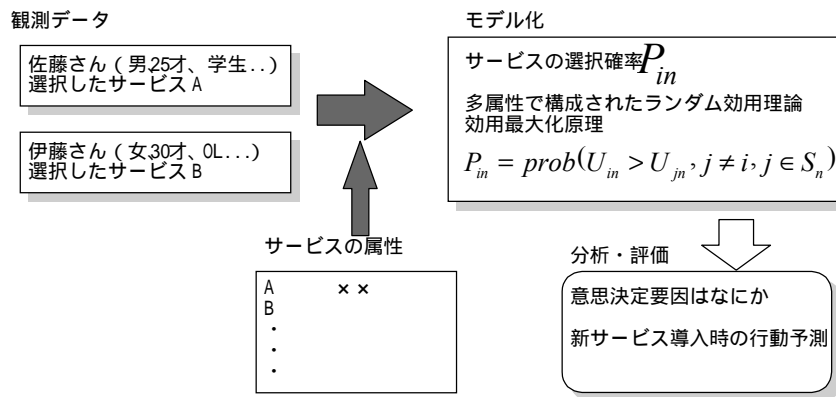


図3-2 離散的行動分析のアプローチ

3.4 ISP の選択による多項ロジットモデル

3.4.1 多項ロジットモデルの定式化

ISP のインターネットアクセスサービスについてユーザが行った選択行動についてロジットモデルを用いて記述する。

まず、ここで推定しようとしている多項ロジットモデルは、個人 n が ISP の選択肢集合 A_n の中から $ISP_i (i \in A_n)$ を選ぶ確率を P_{in} とすると、

$$P_{in} = \frac{e^{\lambda V_{in}}}{\sum_j e^{\lambda V_{jn}}}$$

となる。ここで λ は効用の確率項のばらつきをあらわすパラメータ。また、個人 n の効用関数 V_{in} は線形を仮定したもので

$$V_{in} = \sum_{k=1}^k \theta_k X_{ink}$$

とする。このとき、 β_k は効用関数の k 番目の変数 X_k のパラメータをあらわす。実際の推定では β_k と β_{k+1} だけを分離することができないので、通常 $\beta_k = 1$ として推定する。

3.4.2 選択肢とその属性

ユーザに対してこれからどのようなサービスを使用したいかという選択肢をあたえる。そのためには選択肢をどのように作成するかが重要な工程となる。水準としては業界でよく取り上げられるものを用いながら、直交表を用いて水準を減らしても変数間の独立性が保たれるようなプロフィールを作成する。

ここでは表 3-2 に示した 16 の選択肢から 1 つの ISP を選択するロジットモデルを考える。前述のように、ロジットモデルは効用関数と選択肢によって規定される。効用関数の変数選択については、前節で議論した探索コストと変数選択の関連性を考慮しながら、専門紙などで評価されている項目を念頭において選択する。それらを選択肢の属性として考えて、ISP の種類、ブランド、料金制度、単位料金、アクセスポイントの所在地、1 アクセスポイントあたりの加入制限の有無、ISDN 対応時の通信速度、バックボーンの通信速度、サポート体制などの項目を設定した。前半はおもに料金に関するもの、後半はサービス品質に関するものである。ただし、選択肢にはその ISP サービスを選択した場合の総費用は記載されない。これはユーザのインターネットアクセス時間などの利用頻度によって異なるからである。

選択肢は、それぞれの変数の属性に水準を設け、それを組み合わせて作成する。具体的な水準は、まず ISP の種類に関するものとして、東京・大阪を含む全国規模をサービスエリアする「ISP 全国」、一つの地域に集中してサービスを提供している「ISP 地方」、さらに「CATV」、「OCN」の 4 水準を想定する。そして、それぞれにダミー変数を割り当て、該当する場合は 1、そうでない場合は 0 とする。また、ブランドについては「大手」と「中小」の 2 水準を考え、「大手」の場合に 1 とする。料金制度は「定額」と「従量」の 2 水準で「定額」の場合に 1 とする。料金は「定額」と「従量」について標準的な 3 水準を割り当てる。なお、「定額」は月額料金、「従量」は 3 分あたりの単位料金を表示してある。アクセスポイントの所在地について、単一料金区域 (MA) 内に「ある」か、「ない」という 2 水準を設け、「ある」場合に 1 とする。加入制限についても同様で、ある場合に 1 とする。ISDN とバックボーンの数値について

ではそれぞれ2水準用意し割り当てる。サポート体制についてはサポートの時間により「24時間」と「昼間のみ」の2水準を想定し、「24時間」に1を割り当てる。

この属性の水準をすべて考慮した選択肢を作成すると膨大な数になることと、ロジットモデルの効用関数の変数は独立性を担保するために、コンジョイント分析のプロファイルの作成[片平 87][上田 94]を参考に表 3-1の水準を実験計画に用いる直交表に割り当て、表 3-2のような16 選択肢*を作成する。

表 3-1 選択肢の水準表

	水準			
	1	2	3	4
形態	ISP (全国型)	ISP (地域型)	CATV	OCN
制度選択	定額	従量		
料金選択	小	中	大	
定額制 (月額)	¥1,500	¥2,000	¥3,000	
従量制 (3分)	¥10	¥20	¥30	
アクセス回線あたり加入制限	なし	あり		
アクセスポイントまでの距離	隣接MA内	その他		
ISDN	同期 (64)	MP (128)		
バックボーンの数	256k	1500k		
サポート	日中	24時間		
ブランド	なし	あり		

* このように選択肢が多数存在する調査では、その推定の安定性からもサンプル数は多い方が望ましい。今回の調査ではサンプル数は全体的に少なく十分とはいえないが、モデルとしての安定性や説明力は示す適合度指標（後述の注参照）によると、推定されたモデルは調査データを十分に説明するものであると考えられる。

表 3-2 選択肢と属性

選択肢No	ISP 全国 ダミ -	ISP 地域 ダミ -	CATV ダミ -	OCN ダミ -	ブラン ド ダミ -	料金 制度 ダミ -	料金	MA 内ダ ミー	加入 制限 ダミ -	ISDN 速度	back- bone 速度	サポ ート ダミ -	選択 結果
	あり 1	あり 1	あり 1	あり 1	あり 1	定額 1	円	MA内 1	あり 1	kbps	kbps	24時 間を 1	(人 数)
1:ISP全国1	1	0	0	0	1	0	10	0	0	128	1500	1	11
2:ISP全国2	1	0	0	0	0	1	3000	0	1	64	256	1	11
3:ISP全国3	1	0	0	0	1	1	5000	1	0	128	256	0	3
4:ISP全国4	1	0	0	0	0	0	50	1	1	64	1500	0	0
5:ISP地方1	0	1	0	0	1	1	1500	1	1	64	256	0	12
6:ISP地方2	0	1	0	0	0	0	30	1	0	128	1500	0	2
7:ISP地方3	0	1	0	0	1	0	50	0	1	64	1500	1	2
8:ISP地方4	0	1	0	0	0	1	5000	0	0	128	256	1	1
9:CATV1	0	0	1	0	0	0	10	1	0	0	256	1	6
10:CATV2	0	0	1	0	1	1	3000	1	1	0	1500	1	32
11:CATV3	0	0	1	0	0	1	5000	1	0	0	1500	0	1
12:CATV4	0	0	1	0	1	0	50	1	1	0	256	0	0
13:OCN1	0	0	0	1	1	1	1500	0	1	128	1500	0	25
14:OCN2	0	0	0	1	1	0	30	0	0	64	256	0	4
15:OCN3	0	0	0	1	1	0	50	1	1	128	256	1	1
16:OCN4	0	0	0	1	1	1	5000	1	0	64	1500	1	8

3.4.3 ユーザデータの概要

分析するユーザデータはアンケート調査*によって得た。1996年の10月に約2000人を対象とした全国規模のメーリングリストに送付し、総回答数は136人（有効回答数119）を得た。回答率は低いものの、プロフィールは男性中心（約8割）、20・30代中心（8割以上）技術系中心という他の調査[CSJ 1996][ECN 96][東大・野村 96]と似たよく見られるものであった（表 3-4、図 3-3）。回答者の居住地†については、8割近くが関東、近畿、中部にある（表 3-3）が、東京、大阪などの大都市とそれ以外にわけた場合、都市部はおよそ3割であった‡。

インターネットへの接続状況は、平均で一回あたり45分、一週間に6回、月17時

* 調査表については付録参照のこと。

† このようなアンケート調査の場合、前節で検討したように、常に代表性の問題が付きまとうが、地方的なサンプルについては、人口分布との比率や回答者のプロフィールに大きな違いは見られないことから、特に修正することなくデータを使用している。

‡ このような分布の場合、都市と地方における選択行動の違いが考えられるが、都市と地方にわけた場合の選択パターンを²検定した結果からは有意な選択差は検出できなかった。そのため、本分析には居住地域の違いを考慮した変数は採用されていない。

間程度というものであった。ユーザ層別による効用の違いを測定するため、接続時間分布（図 3-4）の四分位点から、月 700 分以上をヘビーユーザ（n=32）、700 分未満をノーマルユーザ（n=87）というユーザ層*を便宜的に設定した。

表 3-3 回答者の居住地分布

地方	人口		本調査		csj第4回
	人口	人口比率	サンプル数	サンプル比率	比率
北海道	5,684,842	4.55%	4	2.94%	2.3%
東北	9,865,006	7.90%	1	0.74%	2.7%
関東	39,159,557	31.35%	78	57.35%	52.9%
北陸	5,614,151	4.49%	6	4.41%	
中部・甲信越	17,515,894	14.02%	14	10.29%	12.6%
近畿	20,383,100	16.32%	14	10.29%	16.7%
中国	7,763,515	6.22%	9	6.62%	5.8%
四国	4,220,707	3.38%	3	2.21%	
九州・沖縄	14,707,601	11.77%	7	5.15%	4.7%
その他					1.5%
全国	124,914,373	100.00%	136	100.00%	99.2%

表 3-4 回答者の職種と性別

職種	男		女		総計	
	人数	比率	人数	比率	人数	比率
技術開発	54	45.8%	4	22.2%	58	42.6%
企画総務経理	25	21.2%	6	33.3%	31	22.8%
営業	18	15.3%	6	33.3%	24	17.6%
学生	3	2.5%	0	0.0%	3	2.2%
自営業	1	0.8%	0	0.0%	1	0.7%
その他	17	14.4%	2	11.1%	19	14.0%
総計	118		18		136	

* モデルの推定時にはここから更に矛盾のあるデータ（アクセス回線に ISDN を選びながら CATV のインターネットサービスを選択するなど）を削除しているため、推定したモデルにより対象数は異なる。

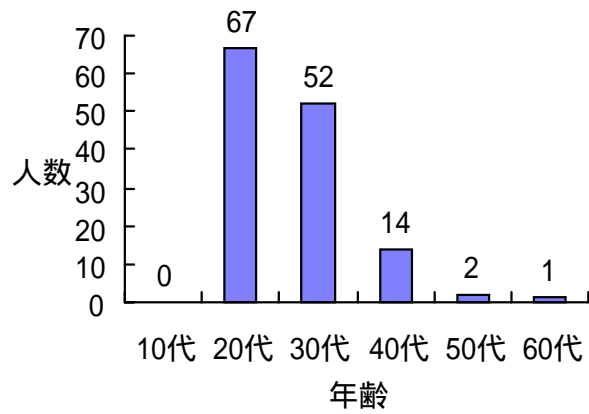


図 3-3 回答者の年齢構成

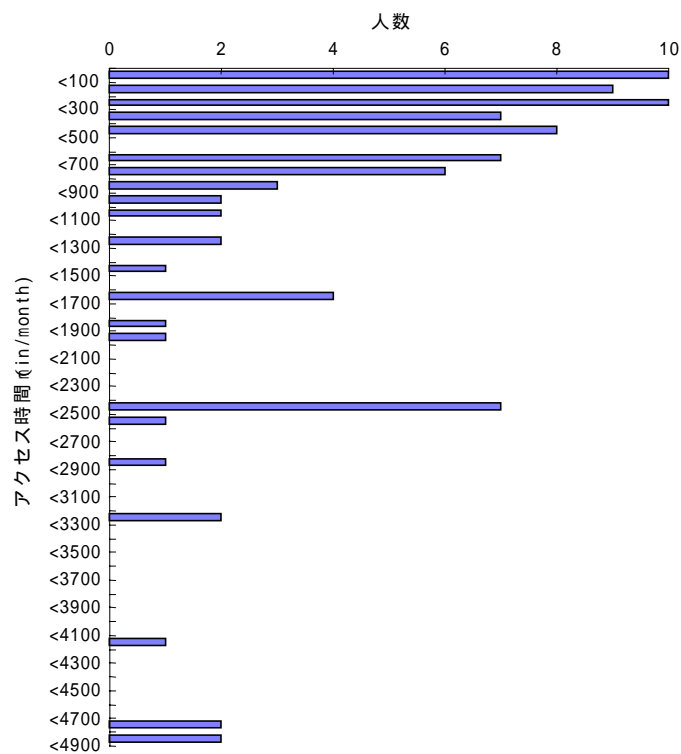


図 3-4 アクセス時間（月間）の分布

3.4.4 推定結果

推定にあたっては、さきほどのプロファイルの選択肢属性にはないがユーザの利用時間をもとに月間の総費用*を計算し、これについてのパラメータを追加している。探索コストとしてサービスの使用コストはこの総費用に含まれる。また、モデルのフレームワークで議論したように、機会費用としての利用時間は直接変数に使用せずに、利用時間を通して総費用のなかに反映されていると考える。

変数選択としては Stepwise 法、Backward 法、Forward 法などを用いて p 値が有意になる変数を選択している。なお、推定のプログラムとしては SAS の PHREG プロシージャを利用した。さらに、推定ではユーザの利用度に応じてモデルを別に推定した。モデルに違いがあるかについては尤度比検定をおこない、変数のパラメータ間の違いがあるかどうかについては t 検定で検定した。この検定は[藤原 93]の方法による。

モデルの適合度の指標 $-2(L(0)-L(\hat{\theta}))$ は、モデルのパラメータすべてが 0 であるかないかの χ^2 検定量で、 R^2 は McFadden の決定係数で尤度比を示すもので、 $R^2=1-L(\hat{\theta})/L(0)$ で計算され、0.2 ~ 0.4 程度あればモデルとしての説明力はあると考えられる[土木学会 95]。

ISP の選好についての効用関数のパラメータを、全体・ノーマル・ヘビーの 3 つのユーザ層別で推定した結果を表 3-5 に示す。 R^2 は全体とノーマルユーザのモデルで低く、ヘビーユーザでは説明力をもっている。総じてモデルの説明力は低いが、ここで得られたモデルのパラメータに対して、平均値法†により集計した結果は図 3-5 のとおりで、分析に使用したデータの中での ISP サービスのシェアがある程度再現されている。このことから、推定したモデルが調査データを表現したものとなっていると考えられるので、以下この推定結果を用いた議論をおこなう。

まず、探索コストと大きな関連がある料金制度オプションや総費用については、モデルの説明力が必ずしも高くないので詳細な検討は難しいが、どのユーザ層別モデルでも総費用に対して一様に負の効用が測定されている。そして、料金制度の定額制に対する指向は高く、特にノーマルユーザで高い。

* 定額料金+従量料金×利用時間

† 推定されたモデルに対して、利用時間など平均的なユーザプロファイルを代入し、選択肢のシェアを検討する方法。非集計モデルであるロジットモデルは個人の行動を表現するモデルではあるが、実務的には個人データを集計して議論することが多いため、このように簡便的にユーザデータを代入して選択肢間のシェアを議論することが多い[土木学会 95]。

また、サービス品質については、ヘビーユーザでは一加入あたりの加入制限やバックボーンスピードなどインターネットアクセスの速度に対する効用が有意になってきている。また、ノーマルユーザではサポートの効用がヘビーユーザに比べて高いのも注目される。ノーマルユーザには比較的初心者も多く含まれることが考えられるから、サポートの重要性も高いことは想像できる。

ブランド力については、正の要因として検出されている。他の要因が等しい場合のブ

ランド力[Brynjolfsson 00a] $\Delta p = \frac{-\beta_{BRAND}}{\beta_{PRICE}}$ は $0.658/0.000229 = \text{約 } 2873$ 円となり、一

ヶ月の料金に対してブランドが存在することでノンブランドに対して3千円近い競争力をもっていることが推定されている。しかし、モデルの説明力が低いことを考えるとこれは少々誇大に推定されているとも思えるが、[Brynjolfsson 00a]でも指摘しているように、選択肢のサービスをユーザが実際に使ったことがなく、サービス品質が曖昧で不明確であるので、サービス品質に対する信頼の代替の指標としてブランド力を参考にした選択を行っている可能性が考えられる。この傾向はノーマルユーザに顕著で、サービス品質について多くの情報をもつことが想像されるヘビーユーザはブランドに固執することなく、料金による効用に反応している。

ユーザ層別のモデル間の差については、尤度比検定やt検定の結果、5%の危険率で、ユーザ層別でのモデル全体の違いがみられるが、料金に対するパラメータの違いは検出できない(表 3-6、表 3-7)。

推定結果は以下のように要約できる。

- 探索コストに関連する料金制度や費用については、総費用に対してマイナスの選好が推定された。総費用は機会費用としての利用時間をもとに計算されたものである。また、ノーマルユーザでは地域限定のサービスに否定的で、定額料金制度に選好がある。
- サービス品質については、ヘビーユーザに顕著にあらわれ、アクセスポイントあたりの加入制限やバックボーンスピードなどに対する選好が測定された。
- ブランドについては、サービス品質に対する情報の少ないノーマルユーザはサービスに対する信頼の代替の指標としてブランド力を利用している。サービス品質に関する情報をもつと考えられるヘビーユーザは利用していない。

表 3-5 SP データによる多項ロジットモデルの推定結果

ユーザ層別 変数名	全体 (n=119)		Normal User(n=87)		Heavy User(n=32)	
	parameter	p値	parameter	p値	parameter	p値
全国ISPダミー						
地方ISPダミー			-0.738	0.0282		
catvダミー						
ocnダミー						
ブランドダミー	0.658	0.0175	0.792	0.0045		
制度定額ダミー	0.685	0.0065	1.03	0.0001		
総費用	-0.000229	0.0001	-0.000357	0.0001	-0.000124	0.0061
MA内ダミー						
加入制限ダミー	0.532	0.018			1.49	0.0014
ISDNサポートスピード						
バックボーンスピード	0.000401	0.0259			0.00102	0.0029
サポートダミー	0.623	0.0024	0.573	0.0182		
適合度指数						
$-2(L(0)-L(\hat{\theta}))$	120.478	0.0001	77.276	0.0001	50.993	0.0001
	0.182		0.160		0.287	

表 3-6 SP データでのモデルの違いの検出 (尤度比検定)

model	-2logL	DF	$-2\log(L(a+b)-L(a)-L(b))$	DF of Chisq	p
All	539.085	6	7.478	2	0.0238
Normal:<700	405.154	5			
heavy:>=700	126.453	3			

表 3-7 SP データでのモデルの違いの検出 (t検定)

variable	model	parameter	se	t	n	t-value	p
total fee	normal:<700	-0.00036	7.26E-05	-4.92	87	1.89	0.0680
	heavy:>=700	-0.00012	4.52E-05	-2.74	32		

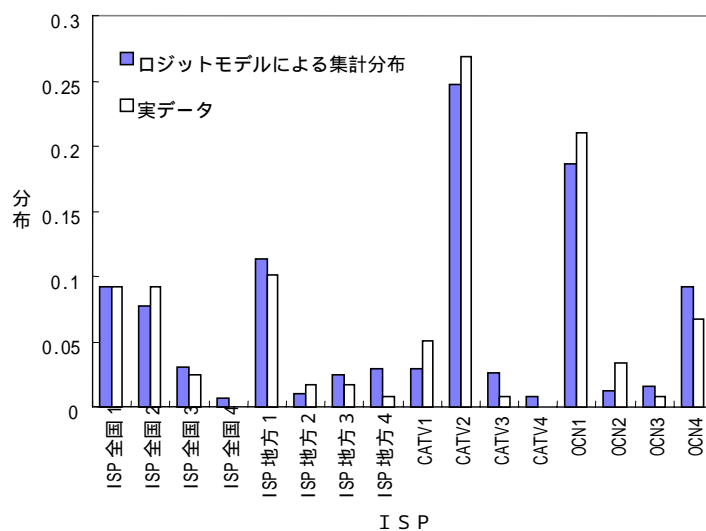


図 3-5 平均値法による多項ロジットモデルの集計結果

3.5 アクセス回線と ISP の選択によるネスティッド・ロジットモデル

3.5.1 ネスティッド・ロジットモデルの定式化

分析のフレームワークでおこなった議論を踏まえて、図 3-6のようなアクセス回線と ISP の選択を扱う選択構造を想定する。

ここで、個人 n がアクセス回線 m と ISP r の組み合わせ rm を選択する確率 $P_n(rm)$ は、アクセス回線 m の選択確率 $P_n(m)$ と m を選択したという条件のもとでの ISP r の選択確率 $P_n(r|m)$ の積であらわせられる。

$$P_n(rm) = P_n(r | m)P_n(m)$$

$$P_n(r | m) = \frac{e^{\lambda_1 V_{(r|m)_n}}}{\sum_{r=1}^{R_{mn}} e^{\lambda_1 V_{(r|m)_n}}}$$

$$P_n(m) = \frac{e^{\lambda_2 (V_{mn} + V_{mn}^*)}}{\sum_{m=1}^{M_n} e^{\lambda_2 (V_{mn} + V_{mn}^*)}}$$

このとき、

$$V_{mn}^* = \frac{1}{\lambda_1} \ln \sum_{r=1}^{R_{mn}} e^{(\lambda_1 V_{(r|m)_n})}$$

m : アクセス回線選択肢 $m=1,2,\dots,M_n$ r : ISP の選択肢 $r=1,2,\dots,R_{mn}$.

アクセス回線 m を選択した条件のもとでの ISP r の選択をレベル 1 の選択、アクセス回線 m の選択をレベル 2 の選択と呼ぶことにする。レベル 1 の ISP の選択は、個人 n の ISP r に対する効用の確定項 $V_{(r|m)_n}$ をもとに他の選択肢との間で評価されてきまる。一方、レベル 2 のアクセス回線の選択は、アクセス回線 m に対する効用の確定項 V_{mn} とレベル 1 の選択における効用がログサム変数を通して反映した V_{mn}^* との和により他のアクセス回線との間で評価されてきまるような構造になっている。このレベル 1、2 での λ_1 はスケールパラメータである。

それぞれの効用関数はパラメータに対して線形を仮定し、

$$V_{(r|m)_n} = \sum_{k=1}^{k_1} \beta_k X_{(r|m)_nk}$$

$$V_{mn} = \sum_{k=1}^{k_2} \theta_k X_{mnk}$$

とあらわす。ここで $X_{(r|m)_nk}$ はレベル 1 での効用関数の k 番目 ($k=1,\dots,k_1$) の変数であり β_k はその未知パラメータをあらわす。同様に、 X_{mnk} はレベル 2 での効用関数の k 番目 ($k=1,\dots,k_2$) の変数で θ_k はその未知パラメータである。

このモデルの推定では、まずレベル 1 の推定を行い、それからログサム変数を計算し、レベル 2 の推定をおこなう段階推定を行う。それぞれの段階でのモデルに対する適合度の指標は基本的にはロジットモデルと同じものを用いることができる。

さらに、レベル 1 のスケールパラメータ λ_1 は β_k と分離して推定することができない

ので、通常 $\lambda_1=1$ として推定する。レベル2の λ_2 はログサム変数のパラメータとして推定できる。ただし、 λ_2 については注意が必要である。モデルの性質からこの λ_2 は

$$0 \leq \lambda_2 \leq 1$$

となっていることが必要である[土木学会 95]。この時にはじめてこのネストの構造において効用最大化行動をとっているとみなすことができる*。

そこで、図 3-6のようなアクセス回線と ISP を組み合わせた選択構造を想定したネスティッド・ロジットモデルで定式化を行い、ユーザ層別に効用関数を推定する。

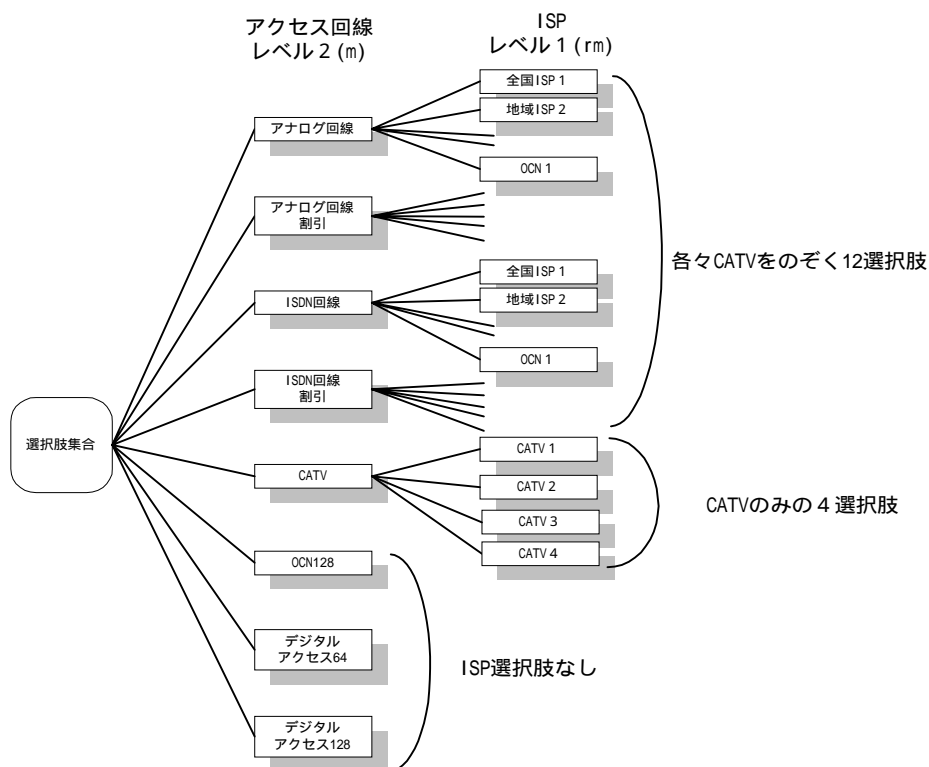


図 3-6 ネスティッド・ロジットモデルの選択ツリー

3.5.2 選択肢と属性

レベル1のISP 選択肢は前述のISP 選択の多項ロジットモデルと基本的におなじであるが、アクセス回線との物理的關係により異なる。図 3-6に示すとおり、たとえばア

* 逆に λ_2 が1を超えてしまう場合には、この選択オプション構造を重視していないことになる。

アクセス回線に CATV 回線を用いた場合、ISP は CATV の ISP 選択肢の範囲内で選択し、アクセス回線が CATV 回線以外では逆に ISP に CATV を選択できないような構造になっている。

また、レベル 2 のアクセス回線では表 3-8のような 8 つの選択肢*を用意し、アナログ回線、ISDN 回線、デジタルアクセス（廉価版専用線サービス）、OCN、CATV というアクセス手段にそれぞれダミー変数として、利用する場合に 1 を割り当て、そうでない場合は 0 を割り当てる。同様に、割引があるかどうか、定額サービスかどうかでもダミー変数を割り当てる。

表 3-8 アクセス回線の選択肢

No	アクセス回線区分	速度 (kbps)	初期費用†	月額料金	選択数
1	アナログ回線+定額サービス	28.8	0	3550	6
2	アナログ回線	28.8	0	1750	15
3	ISDN回線+定額サービス	64	2000	5230	28
4	ISDN	64	2000	2830	34
5	デジタルアクセス64	64	90000	28000	3
6	デジタルアクセス128	128	90000	38000	0
7	OCN128	128	90000	20000	7
8	CATV	10000	30000	0‡	40

3.5.3 推定結果

アクセス回線と ISP の選択を考慮したネスティッド・ロジットモデルの推定結果は表 3-9のとおりである。まず、ログサム変数のパラメータはいずれの層別でも 0 と 1 の間にあり、アクセス回線を関連付けた ISP の選択選択構造の合理性を確認した。ユーザはサービス選択においてアクセス回線と ISP を関連付けて選んでいるといえる。

レベル 1 の ISP の選択に関しては、さきほどのモデルとほぼ同じような傾向が読み取れるものの、推定されて有意なパラメータが異なる。これは推定時のデータ数の違いによるもので、ネスティッド・ロジットモデルの推定時には、アクセス回線と ISP の組み合わせで矛盾のある選択をおこなっているユーザデータをはずしてあるために異な

* アクセス回線の属性は調査時点 1996 年当時のもの

† アナログ回線、ISDN 回線の初期費用にはアナログ電話は加入されているとして電話加入料金は入っていない。

‡ CATV の月額料金が 0 になっているのは、ISP サービスとアクセスサービスを分離できないため。ここでは通常のロジットモデルとネスティッドロジットモデルの整合性を保つために基本的に ISP サービス側の課金にしている。

る効用が推定されたと考えられる。これは推定におけるデータ数が少ないため、かならずしもモデルが安定していないことを示している。しかし、モデルの適合度である²は向上し、アクセス回線の構造を導入することでモデルの適合度も改善されている。

ISP のブランド力の価格との代替性をしめす $\Delta p = \frac{-\beta_{BRAND}}{\beta_{PRICE}}$ は全体のモデルでは 7 5 1

3 円と高く、ノーマルユーザでは 2 5 8 2 円と前回とほぼ同様の数値を示している。これはモデルの構造がノーマルユーザでは比較前回の ISP の選択モデルと近いためだと考えられる。

また、レベル 2 のアクセス回線の選択では、各回線のダミー変数は有意にならず、アナログのみマイナスの効用が確認された。ユーザの選好としては、非アナログの選好が測定された。

特に、レベル 2 のアクセス回線でも探索コストと関連が考えられる経常費用については、ISP 同様にマイナスの選好が測定された。しかし、回線速度などの品質情報や特定サービスブランドについての選好は意味のあるモデルからは検出できなかった。これは調査時点ではまだブロードバンドが普及しておらず、一般のユーザにとってはアクセス回線の品質はアナログ回線でモデムをつかう 28.8 kbps か ISDN の 64kbps 程度の違いしかなかったためだと考えられる。そのため、非アナログという選好はこれを反映していると考えられる。

モデルの差異の検出方法は前出と同様である。レベル 1 では、尤度比検定では層別されたモデル間の違いは検出されなかった（表 3-10）。ただし、共通のパラメータである定額制に関する指向性においては棄却率 5% で有意な差がある（表 3-11）。さらに、レベル 2 では、尤度比検定ができず、パラメータの t 検定も有意な差は検出できなかった（表 3-12、表 3-13）。

表 3-9 ネスティッド・ロジットモデルの推定結果

	ユーザ層別 変数名	全体(n=88)		Normal User(n=68)		Heavy User(n=20)	
		parameter	p値	parameter	p値	parameter	p値
LEVEL1	全国ISPダミー			0.676118	0.03		
	地方ISPダミー						
	catvダミー						
	ocnダミー	-0.641	0.0437				
	ブランドダミー	1.36	0.0001	0.909	0.0049		
	制度定額ダミー	0.989	0.001	0.772	0.0089	2.20	0.0032
	総費用	-0.000181	0.0019	-0.000352	0.0001		
	MA内ダミー	-1.12	0.001				
	加入制限ダミー					1.73	0.0056
	ISDNサポートスピード						
	バックボーンスピード						
	サポートダミー						
	適合度指標						
	$-2(L(0)-L(\hat{\beta}))$	79.266	0.0001	54.902	0.0001	25.54	0.0001
		0.210		0.182		0.330	
LEVEL2	アナログダミー	-1.24	0.0001	-1.22	0.0002	-1.40	0.028
	割引ダミー						
	INSダミー						
	デジタルダミー						
	OCNダミー						
	CATVダミー						
	回線速度						
	初期費用						
	経常費用	-0.000214	0.0001	-0.000182	0.0046		
	定額ダミー						
	ログサム変数()	0.378	0.0406	0.518	0.0158	0.492	0.0037
	適合度指標						
		$-2(L(0)-L(\hat{\theta}))$	100.52	0.0001	74.392	0.0001	16.247
		0.274		0.263		0.195	

表 3-10 NLモデルレベル1での層別での尤度比検定

model	-2logL	DF	-2log(L(a+b)-L(a)-L(b))	DF of Chisq	p
All	298.752	5	1.175	1	0.278
normal:<700	245.693	4			
heavy:>=700	51.884	2			

表 3-11 NL モデルレベル 1 での層別の t 検定

variable	model	parameter	se	t	n	t-value	p
制度ダミー	normal:<700	0.772	0.295	2.61	68	2.11	0.0437
	heavy:>=700	2.20	0.745	2.95	20		

表 3-12 NL モデルレベル 2 での層別での尤度比検定

model	-2logL	DF	-2log(L(a+b)-L(a)-L(b))	DF of Chisq	p
All	265.462	3	-9.88	2	#NUM!
normal:<700	208.412	3			
heavy:>=700	66.93	2			

表 3-13 NL モデルレベル 2 での層別の t 検定

variable	model	parameter	se	t	n	t-value	p
アナログダミー	normal:<700	-1.22	0.322	-3.78	68	0.262	0.795
	heavy:>=700	-1.40	0.636	-2.20	20		
ログサム変数	normal:<700	0.518	0.215	2.41	68	0.0637	0.950
	heavy:>=700	0.492	0.169	2.90	20		

3.6 考察とまとめ

本章では、インターネットアクセスサービスの選択構造を検証することで、アクセスサービスの対価とそれを利用する機会費用で構成される探索コストに対するユーザの態度を推定し、次章以降の探索コスト概念形成に寄与することが目標であった。

探索コストを含む料金・費用変数のほかにサービス品質、ブランドに関する変数をもった効用関数を定義したうえで、ISP 選択のロジットモデルを構築したのちに、アクセス回線と ISP 選択を組み合わせたネスティッド・ロジットモデルを構築した。これにより、ユーザはアクセス回線を規定し、その上で ISP のサービスを選択するというインターネットアクセスにおける基本構造を確認し、モデルとしても説明力を向上させることができた。

探索コストに直接関係する料金制度オプションや総費用については、ISP、アクセス回線ともにマイナスの選好を測定した。機会費用としての利用時間は料金制度オプションを通して総費用の中に反映されていると考えられる。ユーザのレベルにもよるが、料金制度の定額オプションやサービス品質面での速度指標に対する高い選好からは、ユーザが機会費用としての利用時間を重視していることがわかる。

また、ブランドは、[Brynjolfsson 00a]が指摘したとおり、曖昧なサービス品質に対する信頼の代替指標として機能していると考えられ、ノーマルユーザとヘビーユーザのブランドに対する態度の違いはユーザがもっている品質についての情報量の違いを反映しているとも考えられる。サービス品質に関する属性の中で、機会費用としての利用時間（探索コスト）に関係する回線速度の属性を通じて、ブランドに対する態度が探索コストに影響する可能性もある。

さらに、本研究で対象とする URL へのアクセス行動においては、どの URL を選択するかということについて信頼の代替指標としてのブランドが大いに影響することが考えられる。ブランドが存在することで目的とするコンテンツに早く到達し、それにより探索コストが小さくて済むような状況が考えられるからである。しかし、URL のブランド情報をどのように収集し、データに反映するかは非常に難しい。Web サイトがブランドを頼りにリンクを張り巡らし、それをもとに検索エンジンが Web ページの順位を作成する PageRank や HITS[Kleinberg 98]などのアルゴリズムをつかっている場合には、ユーザは意識することなくブランド情報を利用していることが考えられる。いずれにしても、ブランド情報を分離しながら URL へのアクセス行動を検討するためにはさらに進んだ研究が必要である。

インターネットでユーザが行動しているときには、それにもなまって発生する探索コストを常に意識することになるので、選択行動において探索コストは重要な役割を果たし、探索に用いているブランド情報も含めてブラックボックスとして探索行動自体を表現しているものといえる。

以上の知見を踏まえて、次章以降では探索コストをつかって URL の探索行動を捉える仮説や方法論を検討していく。

部

部では、探索コストとしてコンテンツ探索距離概念を提示し、仮説検証を通じて Web コンテンツ探索構造を検証していく。

第4章 探索コストを用いた研究仮説の設定

本章では、本研究の『Web を代表とするインターネットは「空間と時間の壁」を克服したのか』という命題に対して、ユーザの URL へのアクセス行動を経路的な探索行動として扱うことでアプローチし、探索行動の概念を導入することで探索コスト（距離）を使ってこの命題を検証する2つの仮説を提示する。

さらに、Web 探索構造を形成する要因を考察するために、構造的な特徴とユーザの行動パターンに関する3つの仮説を提示する。

4.1 研究課題と仮説の設定

4.1.1 命題とアプローチ方法

アクセス行動による命題へのアプローチ

本研究の命題は『Web を代表とするインターネットは「空間と時間の壁」を克服したのか』ということを検証することである。この命題は抽象的なので、検証するためには具体的な対象と検証すべき量を定義した仮説が必要である。

まず、インターネットにおける「空間と時間の壁」をどのように表現するかが問題になる。関連する分野のアプローチを見てみると、第2章の消費者情報探索行動モデルでは、ショッピングボットでの価格分散の測定から電子商取引の効率性を議論し、価格探索における価格情報、店舗ブランドや過去の購買履歴の依存性について指摘することで、一見フラットで平等なアクセスが可能なインターネットにも現実世界と同様の「壁」が存在していることを示すというアプローチがとられている[Brynjolfsson 00b]。また、Web グラフの解析では、HTML リンクの接続構造を分析して、強連結集合 SCC を中心としたマクロなクラスター構造という特徴を指摘している[Broder et al. 00]。しかし、

前者は、電子商取引市場での価格探索行動を取り上げてはいるものの、具体的な「壁」となる探索コストを計測することはしておらず、その壁を定量的には扱うことはしていない。また、後者の研究は、静的な Web ページのリンク構造の特徴は示していても、そこを利用するユーザの行動は取り扱っていない。さらに、サーチエンジンなどを利用して行動するユーザが多いことを考えると、そのグラフ的な特徴が空間的な「壁」とは言い切れない。

そこで、本研究では、「ユーザがおこなった URL へのアクセス行動を経路的に取り扱う」というアプローチをとる。そうすることで、この「壁」をユーザの行動をもとに定量的に取り扱うことに道が開ける。インターネットや Web 自体は HTML をベースにした URL へのアクセスを積み上げた空間と考えられるので、このようなアプローチをとることでインターネットのもつ特徴の一般性を失うことはない。

このように URL へのアクセス行動ととらえることで、Web サイトに生じるユーザのアクセストラフィックを扱うことができる。また、第 3 章のインプリケーションから、アクセスサービスの対価や機会費用として、アクセスしている「時間」が探索コストとして選択効用に大きな影響をもつことがわかっている。そこで、「『時間』がアクセスすることの壁となる」ということは、アクセス時間の違いによって生じた Web サイト間の測定可能な「距離」尺度が存在することであると考える。具体的に、Web サイト間の空間相互作用[大友 97]を考えると、「距離」が増加すると Web サイト間トラフィックは「距離」に反比例して減少することになる。

そして、「距離」が決まればそれによって形作られる空間の構造を考えることができる。その構造の違いによって、Web サイトのアクセス数が変わることを検証できれば、「空間の壁」が存在するといえる。しかし、この「距離」と「空間」を定義するためにはさまざまな課題がある。

「距離」と「空間」を定義するための課題

第 2 章で概観したように、Web における「距離」と「空間」の考え方は多様である。たとえば、空間上に 2 点を定義できれば次のような距離を計測できる。

- ベクトルで定義された 2 点間の各成分の差の p 乗の和から算出するユークリッド距離を代表とした L_p 距離やその共分散を考慮したマハラノビス距離
- 2 つのベクトルを符号として見なし、2 つの符号の違いを訂正するオペレー

ション（置換、脱落、挿入）のコストの最小のものを距離と考えるレーベンシュタイン距離や成分が0と1の二値だけのときのハミング距離

- 二次元上の正方形の格子を辿る経路の辺の数で定義する市街地距離（4近傍距離）や斜めの動きを許すチェス盤距離（8近傍距離）
- 標準パターンのベクトルに対して、あるベクトルとの内積と長さをつかって

定義した類似度をもちいるコサイン類似度 $S(\mathbf{x}, \mathbf{y}) = \frac{(\mathbf{x}, \mathbf{y})}{\|\mathbf{x}\| \cdot \|\mathbf{y}\|} = \cos \theta$

- 類似したパターンを集めたクラスターを想定したときのクラスター間の距離を示すための、最短距離、最長距離、平均距離
- 2つのパターンの相互情報量とエントロピーから定義されるカルバック・ライブラー距離（ダイバージェンス）、c4.5での情報利得基準
- 木構造をもつグラフの2点間の経路の最短経路数を距離とする考え方

このような考え方を背景に、いくつかの距離尺度を利用した研究がある。たとえば、レーベンシュタイン距離をつかってユーザが探索した URL 列のパターン間の類似度 [Pirolli 99b] を計算したものや、n-gram や共起係数を利用して各 URL が連続しやすいかどうかを表した URL 相関度 [風間 99]、Web サイトの静的なリンクをグラフ構造に置き換え、そのグラフの直径を各 URL 間の最小経路探索アルゴリズムを用いて計算したグラフ構造間の経路リンク数 [Broder et al. 00] [Faloutsos 99] などがある。

しかし、URL に対するアクセス行動を扱うにあたって、単純に以上のような概念を持ち込むことは課題が多い。一つは、URL 空間が膨大な選択肢空間をもっていることである。一つの Web サイト内に限定すれば、有限な数の URL を扱えるが、数ヶ月の期間にユーザがブラウズした全 URL を測定することを考えると、その数は膨大になり、計算量が爆発する可能性がある。ましてや、ハミング距離のようなパターン間の構造距離の定義やコサイン測度のような類似性尺度も膨大な URL 空間にはあまり有効とはいえず、この研究で計測したい Web サイト間の「距離」を扱うには不向きである。

[Broder et al. 00]らの研究のように、Web グラフのパスを距離と考えることもできるが、やはり URL が膨大であることにはかわりがない。さらに、ユーザが Web を移動するときには検索エンジンなどショッピングボットなどのエージェントを経由してアクセスする 경우가多く、静的な HTML リンクのないところにリンクが形成されてしまうことを考えると、静的なリンク情報だけでは判断できない。しかし、検索エンジンやショ

ップボットがダイナミックに割り当てるリンク情報はユーザの要求によって発生するので、事前にその経路を知ることはできない。

そこで、膨大な URL 空間を扱うには、Web グラフを縮退し、ユーザの行動を反映した距離と空間の概念が必要になる。そのためには、ユーザの行動の特徴を記述する方法を選択する必要がある。

4.1.2 探索コストを導入した距離と空間の考え方

本研究における探索行動の考え方

本研究では、前節の距離・空間の課題を解決するために、Web 上のユーザの URL へのアクセス行動を「目的とする URL を探索する行動」と考え、これを「探索行動」と呼ぶことにする。このとき、この「目的」をどのように考えるのかを検討する必要がある。

Web 上でユーザが活動している例として、明確な目的をもって、さまざまな Web サイトを渡り歩いたり、検索エンジンを利用したりしながらその情報を探し求めようなケースがある。例えば、所在が不明で入手が難しいオールドビンテージワインを安価に求めようとして、ワインを扱っているサイトを検索し、そのワインサイトのリンクをしらみつぶしに探索したり、ショッピングボットを利用して関連するネットショップの一覧を出しながら比較したりして、目的とするワインを捜し求めるような行動である。この場合には「所在が不明で入手が難しいオールドビンテージワインを安価に求める」という明確な行動目的がある。第 2 章の情報取得行動[中村 01][Card 01][野島 00][Hoffman 96a]の実験では、被験者に行動目的を与えることでその行動を観察している。

また、別の例では、特に明確な目的をもたずに、心の趣くままに Web サイトののリンクたどっていくようなケースがある。例えば、なんとなくニュースサイトを見ていて、関心のあるニュースに心を動かされてどんどんリンク項目をクリックしていくようなケースである。このような行動はよく Web サーフィンとも呼ばれるが、明確な目的はなくてもユーザのなんらかの嗜好に沿った行動であり、Web サーフィンをすること自体が目的であったりもする。[Hoffman 96a]は困難な登山に挑戦し続けたりスポーツに打ち込んだりする行為をもとに、至高体験が連続して発生している状態を「フロー」という言葉で定義し、行動自体が目的となっている Web サーフィンでの行動を規定して共

分散構造分析により要素間の関係を調べている。

この 2 つの例は最終的な目的の内容が明確であるかどうかの違いはあるものの、なんらかの目的 URL を求めて探索していることにはかわりはない。もちろん、探索ルートや探索方法は目的によって変化するし、第 2 章の研究例でも明らかのようにユーザのスキルや環境に左右される[中村 01][Card 01][Hoffman 96a]。また、3 章で議論したように、探索する目的物がブランド情報をもっているときには、検索エンジンでの提示順位が高くなることもあり、検索を行うコストも小さくなる可能性も高い。しかし、Web で活動をしているユーザの目的は多種多様であり、目的の内容を一般的に定義することは難しい。本研究の命題のようにインターネットにおける距離や空間の壁を捉えようとする場合には、探索の目的を定義することが課題ではなく、Web 上の距離や空間構造をユーザの行動を反映した形で定義することが課題である。ユーザの行動を「目的とする URL を探索する行動」と定義しても、その目的が何であるかを問うことは必ずしも必要ではない。

また、消費者行動論の枠組みでは、ユーザが商品（情報）を探索する過程についてのメカニズムを想定することがある。例えば、ユーザの異質性、目的とする商品に対する関与を考慮しながら、喚起されたニーズにもとづいた考慮集合をあげ、それを評価し、選択するようなメカニズムが考えられる。

しかし、本研究ではあえてそのメカニズムを規定しない。前述したように、本研究の「探索行動」は明確な目的の内容を規定しない行動であるので、選択や評価のメカニズムに一定の枠組みを当てはめるのは難しい。本研究の目的はそのメカニズムを議論することではなく、メカニズムはブラックボックスのままで、探索の結果たどり着いた目的とする URL までの「距離」やそれがつながった「空間の構造」にそのメカニズムやユーザの特徴が反映されていると考える立場をとる。

探索行動のもとでの探索コストと距離の考え方

探索行動のもとで発生した距離や空間の構造に対して、本研究では、[Bakos 97]の消費者情報探索モデルでも利用されていた「探索コスト」という考え方を利用する。彼らのモデルでは、ある財についての価格情報や製品情報を手にいれるために一定の「探索コスト」を費やして探索すると仮定して、探索コストと期待効用の均衡分析を行った。

[Bakos 97]のモデルでは、同一属性空間上の円周上に分布した売り手に対して、買い

手が最も望ましい価格 S と最も望ましい距離 D をもっているとして、買い手の期待利得は $g(S, D)$ は、

$$g(S, D) = 2 \int_{x=0}^{D+S/D} \int_{p=0}^{S+Dt-xt} (S + Dt - xt - p) f(p) dp dx$$

であらわされる。このとき、 x は買い手が探索するロケーションであり、 t は探索した財が望ましいものではないときに探索したロケーションー単位あたりに発生する調整コストである。 $f(p)$ は買い手が探索する前の価格 p の確率密度関数である。これがある価格 $p=p^*$ ですべての売り手の価格が均衡しているとすれば、期待利益 $g(S, D)$ は

$$\begin{aligned} g(S, D) &= 2 \int_{x=0}^{D+S/D} (S + Dt - p^* - xt) dx \\ &= \frac{1}{t} (S + Dt - p^*)^2 \end{aligned}$$

である。このとき、期待利益が探索コスト c (一定) と同じ点までは探索すると考えられるので、 $g(S, D)=c$ とすれば、

$$\begin{aligned} c &= \frac{1}{t} (S + Dt - p^*)^2 \\ S + Dt &= p^* + \sqrt{ct} \end{aligned}$$

となり、探索コストはユーザの望ましい価格や探索距離と調整コストを反映していると考えることができる。

[Bakos 97]の消費者情報探索モデルでは商品の価格に対する態度が重要な要素となるので、そのモデルをそのまま使うわけにはいかないが、探索コストに関する基本的な立場を利用することができる。一般に「探索コスト」は、情報を探索するためにつかわれる交通手段やサービスへの金銭支出とそれに要した時間に対する機会費用である[成生 94]と考えることができる。しかし、[Bakos 97]では「探索コスト」とはいうものの、具体的な行動にもとづく「コスト」ではなく、行動した結果の対価となる量であり、抽象的な概念である。

そして、[Bakos 97]のモデルにおける探索コストがユーザの望ましい価格や探索距離と調整コストを反映しているように、第3章で議論したインターネットにアクセスしているときの探索コストは、インターネットアクセスサービスの対価を通して機会費用である利用時間に対する選好を反映している。このように、Web 上の「探索コスト」が、目的とした URL を探索する「探索行動」により移動した空間上の距離や時間を反映し

た量であると考えすることは自然である。

また、[Bakos 97]の研究では「探索コスト」と市場の効率性の関係について議論され、[Lynch 00]でも、大・小の二水準の「探索コスト」とシミュレーションされた市場価格の関係が取り扱われているが、「探索」に関する具体的な行動は規定されていない。彼らの主張は、激的な価格競争によりベルトラン型のゼロ・プロフィットな市場になってしまいかねないインターネットの電子商取引市場において、探索コストをうまくコントロールして製品差別化をおこなうことで、競争の激しさを緩和できるというものであり、探索コストを測定することには関心は払われていない。インターネットエコノミクスの情報探索行動モデルのアプローチでは、購買行動の枠組みでのみ分析し、定量的な探索コストを扱ってこなかったが、本研究では、ユーザの URL へのアクセス行動全般を対象に探索コストを定量的に扱う必要がある。

本研究では、前節で定義した「目的となる URL を探索する行動」として探索行動を考えることで、メカニズムをブラックボックスにしながらも目的となる URL にいたるユーザの嗜好やサービスコスト、機会費用などまでのさまざまな要素を反映した「探索コスト」を考えることができるとする立場をとる。その考え方を導入することで、「探索コスト」を探索行動における距離（探索距離）と見なし、本研究の命題にそって、一見、空間と時間の壁を克服しているように見えるインターネットで、「探索コスト」が新たな壁として機能しているかどうかを検証するアプローチをとる。

このようにして「探索行動」と「探索コスト」を定義できれば、目的とする URL の探索に費やした時間や HTML リンクのホップ数という「探索コスト」を「距離」として計算することが可能になる。また、目的の URL とそれまでの「距離」という表現は、膨大な URL 空間を縮退するので、少ない計算時間で効率よく Web を分析することが可能になる。

さらに、実際の探索行動を扱うので、探索行動で規定された Web グラフでは、静的 Web ページの HTML リンクに加えて、検索エンジンの検索結果やショッピングボットによる動的なリンクを探索した結果を織り込んだトラフィックを同時に扱うことになり、[Broder et al. 00]らの分析を一步進めることになる。

探索構造における Web 空間の考え方

本研究では、前節で定義したユーザの探索行動に対して、[Broder et al. 00]らが発見

した Web グラフにおける強連結集合 SCC を中心にしたクラスター構造を適用するアプローチをとる。そうすることで、Web 空間を単純な HTML リンクのグラフ構造で捉えるのではなく、ユーザのコンテンツ探索行動を反映した目的 URL と探索距離をもった Web 探索グラフ構造として考えることができる。膨大な URL 空間は目的 URL と探索距離で記述できるので Web グラフとしてサイズを縮退することができる。そして、ユーザのアクセスログを利用するので、探索構造とトラフィックを同時に扱うこともできるようになる。

SCC ではどの 2 つのサイトを取り出してもそこに有向グラフが存在するような密接なリンク構造になっているので、そこにユーザのトラフィックが集中することが予想できる。探索構造を利用した Web グラフの空間をインターネットの空間の構造と代表させることで、SCC のトラフィックの多さを他のクラスターとの間で検証できれば、インターネットにおける空間の「壁」の一例を確認できたといえる。

4.1.3 命題を検証する仮説の設定

前節で述べた探索概念を導入することで、冒頭の命題は次のような仮説に解釈・設定することができる。

- 仮説 1 : 「インターネットでのコンテンツ探索コストは、Web サイト間のユーザ交流トラフィックに反比例する」
- 仮説 2 : 「Web 探索グラフにおいて、強連結されたグラフ構造に属する Web サイトには、他のクラスターより多くのアクセスが生じている」

Web アクセスが空間と時間の壁を克服し、第 1 章で引用した表現では「道筋的 = 移動経路的 (トランジェクティブ) ではなくて、遠隔对象的 (テレオブジェクティブ)」[若林 02]であれば、Web サイトへのアクセス数は距離やアクセスする経路に依存しないような Web の空間構造になっていると考えられる。

仮説 1 は、コンテンツトランザクション[Cooley 99]における探索距離がインターネットのコンテンツアクセスに対する障壁として機能していることをあらわしている。インターネットが空間と時間を克服していれば、インターネット上のコンテンツに対するアクセスはその魅力度だけで決定され、ネット上で探索する際にかかる探索コストには依存しない。インターネットエコノミクス分野で扱われている価格を中心にした研究で

はなく、コンテンツに対する探索トランザクションを考えることで、探索コストを空間的かつ経路的に扱い、探索コストを空間的かつ探索経路的に依存した「距離」とすることで定量的に扱うことができる。空間的な2つのロケーション間のトラフィックをその魅力度と距離でモデル化した空間相互作用モデル[大友 97]において、この探索コストは距離として機能するはずである。

具体的には、Web アクセスログのなかからユーザの探索行動の目的とする URL を同定するために、コンテンツトランザクション[Cooley 99]という考え方を利用する。探索行動における目的とする URL を「コンテンツ」と呼び、あらゆるユーザの Web 行動を「コンテンツ」である URL へのアクセス行動と考える。これにより、インターネットエコノミクスで取り扱われている価格を伴った探索コスト以上に柔軟かつ空間的な意味をもった探索コストを考える。目的コンテンツの URL の同定方法や計算方法の詳細は次章で述べるが、ある基準時間を超えて参照される URL は意味のあるものと判断して「コンテンツ」と同定する立場をとる。さらに、ユーザの URL へのアクセス行動データから計算された探索コストをもとに、コンテンツが存在するドメイン・サイト単位にトラフィックを集計し、コンテンツ探索時間やコンテンツ探索ステップ数を探索コストとして考えることで、コンテンツをもつドメイン・サイト間のトラフィックに対する探索コストを距離項とする空間相互作用モデル(グラビティモデル)を構築できる。このグラビティモデルを推定することにより仮説を検証する(第5章)。

仮説2は、コンテンツが他のコンテンツとどのような探索経路に接続をしているかを Web 探索グラフ構造と呼び、[Broder et al. 00]らの SCC(強連結集合)を利用した Web グラフのクラスター構造を導入した Web グラフのマクロ構造ともいべき空間構造がインターネットアクセスにおける壁となっていることを表現している。具体的には、探索グラフ構造とトラフィックを同時に扱うことで、SCC に対するコンテンツへのアクセス数が SCC 以外の他のクラスターと違うことを検証する。この仮説のもとで、インターネットは従来いわれているような、コンテンツが魅力的であれば瞬時にそれを探し出し、瞬時に対象にアクセスすることができるというフリクションレスな場ではなく、コンテンツのアクセス数はコンテンツを探索する道筋に影響されるものであるということを確認する(第5章)。

次章以降では、このように解釈された仮説1と2を検証することで、本研究の命題「インターネットは時間と空間を克服したのか」を確かめていく。

4.2 Web探索構造の形成要因についてのアプローチ

4.2.1 Web 探索構造形成要因を探る2つの方向性

命題である「インターネットは時間と空間の壁を克服したのか」という問いに対して、抽象的だが定量的に計測可能な探索コストを導入することで検証可能な仮説を設定したが、探索行動の定義で述べたように、探索行動自体をブラックボックスとして扱って議論しているので、このような Web 探索構造が生じる原因については言及していない。しかし、第2章で述べた他の研究成果から Web 探索構造に影響すると考えられる要因について2つの方向性が考えられる。

一つは Web 探索構造のグラフ的な特徴的な性質から要因を推測する方向性である。この考え方は探索行動自体の性質はブラックボックスのまま、グラフ的な特徴量を測定し、それに起因する要因について検討する立場である。

もう一つの方向性は、ユーザの探索行動の特徴が要因となって Web 探索構造に影響しているとする考え方である。いままでは、ユーザの探索行動自体はブラックボックスとして扱わずにその結果に着目した「距離」と「空間構造」だけを扱ってきたが、要因を検討するために合理的なユーザ行動を想定して検討する立場である。ただ、扱い方によって、直接探索行動を扱う方法と探索行動の結果としての探索パターンを扱う方法の2つのアプローチが考えられる。

4.2.2 スモールワールド仮説による Web 探索構造形成要因

Web 探索構造のグラフ的な特徴的な性質から形成要因を推測する具体的な方向性として、[Albert 99]で確かめられた Web のスモールワールド仮説が考えられる。それによると、静的な HTML リンク情報で構築された Web グラフの直径は比較的小さく、どんなコンテンツにも平均すると 10 から 20 ホップという比較的小さなリンクでたどり着ける構造となっているという。この構造は Web 探索グラフ構造にも存在している可能性があり、スモールワールド仮説は探索行動や測定される探索コストに影響している可能性がある。

実際には Web グラフから直径等の特徴量を抽出することでスモールワールド仮説が

Web 探索グラフ構造に当てはまるかどうかを確かめることができる。そこで、次の仮説を確かめる。

- 仮説3：「Web 探索グラフの直径は比較的小さく、少ない探索コストでコンテンツにたどり着けるスモールワールドを形成している」

コンテンツトランザクションを導入していることで、Web 探索グラフをコンテンツだけのグラフ構造に縮退した構造を考えることができる。そのコンテンツだけで構成されたグラフの直径を計算し、探索目的の URL であるコンテンツのグラフがスモールワールドを形成しているのかどうかを検証する（第5章）。

4.2.3 ユーザの探索行動による Web 探索構造形成要因

ユーザ探索行動に起因する2つのアプローチ

ユーザの探索行動自体によって Web 探索構造が規定されているとも考えられる。前節までの命題の仮説検証のために、Web グラフの探索構造は Web サイトを中心にその探索構造を眺めてきた。探索の結果生じた Web サイトやコンテンツ間の距離と構造だけに着目し、ユーザの探索行動自体はブラックボックスとして扱い、そのメカニズムには言及しない立場で検討をおこなってきた。しかし、ユーザの探索行動自体にこのグラフ構造の形成要因を求めるためには、ユーザの探索行動自体を取り扱う必要がある。その探索行動自体の特徴が合理的であれば、ユーザには合理的な探索戦略が存在して、それにそった行動をユーザが取ることで Web グラフの構造に影響を及ぼしている可能性がある。

つまり、「ユーザの合理的な探索行動が Web 探索グラフの構造に影響している」ことを検証することが課題である。しかし、このままでは検証すべき仮説としては抽象的なので、それを具体化するために、つぎの2つの立場が考えられる。1つはいままでの仮説と同様に、ユーザの探索行動のメカニズムにはあえて言及しないで、結果として検出されたパターンからユーザの行動の合理性を推測する立場である。もう一つは、ユーザの探索行動に対してある合理的な行動モデルを仮定して、今回の探索行動がそのモデルで説明することができるかどうかを試みる立場である。

探索ヒューリスティックパターンの合理性に関する仮説

Web 探索においてどのような探索パターンの表現を用いるべきであろうか。明確な探索目的が存在するユーザは検索エンジンやキーワードを頼りに、Web サイトのリンク構造に依存しないでダイレクトにコンテンツに到達できるような探索を試みるケースが考えられる。また、探索目的を明確にできないで、関心のある項目をリンクにしたがって Web サーフィンするケースや検索エンジンやショッピングボットをうまく使いこなせないような探索スキルが低い場合は、Web サイトのハイパーリンクをひたすらクリックしていくという Web サイトの構造に依存した探索をおこなうことが考えられる。

前者のような探索をおこなうケースでは Web サイトを複数渡り歩くような探索パターンになり、後者のような探索をおこなう場合には、一つの Web サイトの中味を深く掘り下げていくような探索になることが想像される。

探索パターンからユーザの行動を推測する立場からは、本研究で定義した「探索行動」からの情報として、ターゲットとする URL とそこまでの探索コスト、その URL で費やした時間を利用できる。以上のことを考慮してパターンを検出するアプローチとして、空間的な探索ヒューリスティックを用いる方法が考えられる。木構造を形成する問題空間において解を探索する場合に、探索の深さを優先するのか、幅を優先するのかという2つの探索戦略*を想定することがある[Russell 95]。問題空間の状態によって、どちらの戦略をとるかは解を探索する時間やそれを計算機のアルゴリズムで処理する場合の計算時間に大きく影響する。Web グラフの探索空間では、「探索の深さ」を「同一 Web サイト（ドメイン）内での探索」、「探索の幅」を「異なる Web サイトでの探索」と考えることで、この探索ヒューリスティックを表現することが可能になる。この考え方によって以下のような仮説を検討する。

- 仮説4：「Web コンテンツ探索行動において、ユーザは探索の幅と深さを組み合わせたコンテンツ探索ヒューリスティックの中から、探索行動中により多くのコンテンツにアクセスできる合理的なヒューリスティックを選択する」

深さ優先か幅優先かを組み合わせたユーザの探索ヒューリスティックによって、そのユーザのコンテンツアクセス数に違いが認められれば、コンテンツアクセス数の多い探

* ここでいう探索戦略は、探索ヒューリスティックを表現する場合の「探索方針」やパターンを示す言葉であり、情報探索メカニズムを前提にせずに使用される[サイモン 96]。この分野の語法として、誤解が生じない場合は「探索戦略」という言葉を使用するが、ユーザの意思や行動の合理性などを議論しなければならない場合と混同する恐れがある場合は「探索パターン」と記述することにする。

探索ヒューリスティックを採用する方が単位時間内で多くのコンテンツにアクセスできるので効率的であると考え。そして、合理的な探索ヒューリスティックの採用ユーザ数が増えると考え。

本来の情報探索行動を考える上では、探索する対象の具体的な性質や個人の探索スキルレベルの違いが探索の道筋に影響することは十分考えられる[中村 01][Card 01][Hoffman 96a]。本研究では、あえてユーザのスキルの違いに言及することなく、単位時間内に目的コンテンツに数多くアクセス可能な状態が効率的で合理的であると考えユーザを想定する立場をとっている。合理的なユーザを想定することで、ユーザのスキルの違いや目的の性質に関わらず、合理的なユーザは目的コンテンツの探索においては単位時間内のコンテンツ数の多い探索ヒューリスティックを採用すると考えられる。もちろん、人間の限定合理性[サイモン 96]のもとではこの「合理的」と想定するユーザが利用可能な情報は本人のスキルにも依存するので、すべてのシチュエーションにおいて、この考え方が通用するとは限らない。しかし、Web サイトを訪問するユーザは多種多様で、そのすべてのユーザに対応した Web サイト構造を設計することは非常に難しい。ここでいう合理性を想定することで、このような効率性と合理性をもつユーザには満足を与える Web サイト設計を行うことが可能で、Web サイト設計者がユーザの行動をこのように考えることで、Web サイトの探索構造が決定されてくる可能性も十分にあると考えられる。

以上のように、Web 探索構造の要因をユーザの探索ヒューリスティックパターンに求めることで、合理的な探索行動による構造が形成されているのかを判断することが可能になるのではないかという立場にたつてこの仮説 4 を検証していく。なお、具体的な探索ヒューリスティックの決定方法は後章で述べる（第 6 章）。

情報採餌行動モデルによる合理的な探索コスト仮説

もう一つのユーザの探索行動に対するアプローチは、いままでブラックボックスとしていた探索行動に具体的なモデルを導入することで、いまの構造を説明できるかどうかを試みる立場である。

分析するためには、探索コストと効用の関係を定義したモデルが必要になるが、本研究では第 2 章で参照した「情報採餌行動モデル」を用いた分析を試みる。情報採餌行動モデルでは、ユーザが獲得情報量の長時間平均を最大化させるような合理的な行動を

とるとした場合に、限界効用定理により情報採餌行動をとる各パッチ内での捕食速度は長時間平均に一致する。このアナロジーから餌のパッチを Web サイトと考え、Web サイトで獲得した情報量を仮定し、捕食速度を探索コストと滞在時間で表現することで、本研究で測定された探索コストの分布をもとにしたシミュレーションを行うことができる。それによって、探索戦略としての探索コストの合理性を評価することができる。このとき以下の仮説に従ってシミュレーションをおこなう。

- 仮説 5 : 「Web コンテンツ探索活動全体を通して、コンテンツの探索時間と消費時間を効用関数とした単位時間当たりの獲得情報量を最大化する探索コスト戦略が存在する」

合理的な探索戦略の存在自体は情報採餌行動モデルでの一つの帰結である。この仮説を Web コンテンツ探索に適用することで、Web コンテンツ探索でも同様に成立するかどうか、成立するとするとどのような探索戦略なのかということを実際の探索コストデータの分布を用いたシミュレーションにより確かめる。本来の情報採餌行動モデルにおける効用関数はパッチに滞在し餌を処理（消費）していた時間だけであるが、探索行動における合理性を消費時間とともに検討するために、探索時間にまで拡張した効用関数を想定した表現になっている。おなじユーザの合理的な行動に基準をおいていても、さきほどの仮説 4 とはその測度が異なる。

情報採餌行動モデルを探索行動を規定するモデルとして採用することで、合理的な探索コスト戦略が規定できれば、それをユーザが採用することで探索グラフ構造の特徴や形成要因を説明することが可能になる（第 6 章）。これらの分析・検証により、探索の結果として表現されたユーザの戦略と構造形成要因の関係を知らることができれば、合理的なユーザに対して実際の Web サイト運営やサイトデザインを効率よくおこなう指針をえることができる。

これらの仮説 4・5 の背景には、単位探索時間かまたは長時間の探索時間平均における情報の獲得量を最大化したいと考える「合理的な」ユーザを設定している。ユーザの行動がすべてこの合理的な行動によって説明できるとは限らない。ユーザの非合理的な探索目的やスキルの未熟さにより合理的な行動がおこなわれないケースも十分考えられる。本研究では、この仮説 4・5 を通してこのユーザの合理的な行動において説明を試み、十分説明できない部分はこの考え方の限界と考え、非合理的な事由の可能性を考察する。

4.3 本章のまとめ

本章では、『Web を代表とするインターネットは「空間と時間の壁」を克服したのか』という命題に対して、ユーザの URL へのアクセス行動を経路的な情報として扱うアプローチの考え方について述べた。あえてコンテンツ探索行動のメカニズムには言及せず、ユーザの URL へのアクセス行動を「目的とする URL を探索する行動」と定義することで、探索行動を反映した探索コストを探索距離に適用して、検証すべき2つの仮説、仮説1：「インターネットでのコンテンツ探索コストは、Web サイト間のユーザ交流トラフィックに反比例する」、仮説2：「Web 探索グラフにおいて、強連結されたグラフ構造に属する Web サイトには、他のクラスターより多くのアクセスが生じている」を提示した。

また、その探索構造を決定している要因をさぐるために、Web グラフの構造に対して仮説3：「Web 探索グラフの直径は比較的小さく、少ない探索コストでコンテンツにたどり着けるスモールワールドを形成している」を提示した。

さらに、ユーザの探索行動自体が探索行動を決定していることを検証するために、仮説4：「Web コンテンツ探索行動において、ユーザは探索の幅と深さを組み合わせたコンテンツ探索ヒューリスティックの中から、探索行動中により多くのコンテンツにアクセスできる合理的なヒューリスティックを選択する」で探索パターンからユーザの行動の合理性を推測し、仮説5：「Web コンテンツ探索活動全体を通して、コンテンツの探索時間と消費時間を効用関数とした単位時間当たりの獲得情報量を最大化する探索コスト戦略が存在する」から情報探餌行動モデルの Web コンテンツ探索グラフ構造に対する合理性を検証する立場を提示した。

次章以降は、これらの仮説を検証するためのデータの処理方法を提示しながら、具体的な分析を行うことにする。

第5章 探索距離概念とWebサイト探索グラフ構造の検証

本章では、探索コストを探索距離と考えるアプローチを用いてユーザの Web サイトへのアクセス行動を分析し、第 4 章で提示した仮説 1・2・3 の検証を試みる。まず、探索コストの概念をもったコンテンツの探索行動を仮定し、Web アクセスログからコンテンツ探索時間とコンテンツ探索ステップ数という 2 つの量をつかってコンテンツ探索距離概念を新しく定義する。

そして、その概念をつかったコンテンツトランザクションに関する基礎的な統計データを提示し、Web サイト間のトラフィックとコンテンツトランザクションを用いた探索距離の関係についてグラビティモデルを構築する。

さらに、コンテンツトランザクションのマクロ的なグラフ構造を描き出し、Web サイト間のグラフ構造とトラフィック・探索距離の関係を分析する。

これらの分析結果をもとに、仮説 1・2・3 を検討し、本研究の命題をとなるインターネットにおける空間と時間の壁について考察を加える。

5.1 本章の目的とアプローチ～探索コストを距離としたWebアクセス分析

5.1.1 目的とアプローチ

第 2 章で概観したように、従来の Web アクセスログ解析や Web マイニング分析では、ユーザのアクセスログから抽出したアクセスパターン間の類似性を求めることでユーザの Web アクセス行動の特徴を考察しており、それには様々な距離尺度が用いられている[Pirulli 99b][風間 99]。さらに、Web 生態学の分野では、HTML のリンク情報をもとに Web グラフ構造を描き出すことが行われている[Broder et al. 00]。しかし、

従来のアプローチではパターン抽出のためのデータ解析手法に関する議論や HTML の静的なリンク情報だけを用いていたりするなど、ユーザのアクセス行動を反映した分析は十分に行われてはいない。

本章では、情報の経済学分野の消費者情報探索モデルで議論されている「探索コスト」[Bakos 97]の概念を距離尺度に応用し、ユーザの Web アクセス行動を探索行動と捉えた分析を2つのアプローチに沿っておこなう。

第一のアプローチでは、Web サイト間のユーザトラフィック交流分析において探索コストを距離と定義したグラビティモデルを構築し、探索コストが空間相互作用において距離尺度として機能していることを、探索コストが Web サイト間の交流トラフィックに反比例することで確かめるようにする。

第二のアプローチでは、探索コストを距離と定義した Web グラフのマクロ構造 [Broder et al. 00]を描き出し、探索行動を反映した Web 探索グラフ構造のもつ特徴を検討する。具体的には、ユーザのアクセスログから Web グラフの強連結集合 (SCC) を算出し、SCC と連結する距離によって、Web グラフ空間を分類し、各クラスター間におけるアクセス数や探索における特徴量を比較する。

本章の目的は、これら2つのアプローチを用いて第4章で提示した仮説を検討し、本研究の命題『Web を代表とするインターネットは「空間と時間の壁」を克服したのか』を考察することである。第一のアプローチにより、探索コストが Web アクセス行動における距離として機能していることを確かめることで、仮説1：「インターネットでのコンテンツ探索コストは、Web サイト間のユーザ交流トラフィックに反比例する」ことを確認し、一見、時空の制約を超えたように見えるインターネットにおいて、探索コストが Web サイトにアクセスする距離としての「壁」となっていると考えることができるかどうかを考察する。

そして、第二のアプローチにより、SCC によって生成されたクラスター間で Web サイト単位の平均アクセス数を比較することで、仮説2：「Web 探索グラフにおいて、強連結されたグラフ構造に属する Web サイトには、他のクラスターより多くのアクセスが生じている」ことを確かめ、探索コストが空間的な探索経路によってアクセス数に違いを生む空間的な「壁」となっていることを考察する。

そして、このような構造を生んでいる要因を検討するために、2番目のアプローチにおいて Web グラフの直径 [Broder et al. 00] [Albert 99] など Web グラフ全体の特徴量を

算出することで、仮説 3 : 「Web 探索グラフの直径は比較的小さく、少ない探索コストでコンテンツにたどり着けるスモールワールドを形成している」を確かめる。これにより、Web 探索空間に生じた「壁」を形成するグラフ的な構造要因を考察する。

5.1.2 探索行動に対する立場

第 4 章で議論したように、本章では、ユーザの URL へのアクセス行動を「目的とする URL を探索する行動」と考える。本章での分析では、目的とする URL を探索した結果生じたコスト（距離）を利用するので、実際の探索行動を性格づける探索目的の性質やユーザの探索スキルを反映した探索・選択情報処理を行うメカニズムをあえて考慮しない立場をとる。「探索コスト」には探索行動のさまざまな要素がブラックボックス的に反映されているとし、それを使って探索した範囲「探索距離」を定量的に表現することで、目的とするコンテンツとその間の探索距離を用いた Web 探索グラフ構造を構築していく。

次節では、具体的な分析に入る前に、目的 URL コンテンツの同定方法や探索コスト・距離の計算方法とそれによって計算されたデータの概要を検討していく。

5.2 コンテンツトランザクションによる探索コストの算出

5.2.1 探索行動モデルにおけるコンテンツトランザクション

探索コスト・探索時間・探索距離の関係

本節では、前節のアプローチの方法論を前提に探索コストの具体的な定義について述べる。

まず、探索コストを「情報を探索するためにつかわれる交通手段やサービスへの金銭支出とそれに要した時間に対する機会費用である」[成生 94]と考える。そして、探索コストをつかって探索した範囲の長さを「探索距離」と呼ぶことにする。定義に従うなら、探索コストは、機会費用としての利用時間とアクセスサービスの対価によって決ま

る。

サービスの対価としてのインターネットのアクセスコストについては、第 3 章で議論したように、料金制度や機会費用を通して大きく利用時間に依存していたといえる。ブロードバンドアクセスが普及するにつれ、時間に依存しない定額の常時接続サービスが一般的になりつつあるが、設備とサービスを設計する上で平均的な接続時間やピークトラフィックを考慮した設計になっていることを考えると、定額制であっても接続時間に対する態度がアクセスサービスに影響を及ぼしていると考えられる。探索行動をおこなっていた時間を「探索時間」とすると、探索時間が探索コストに対して重要な役割をもっていると考えることができる。

本章の分析では、アクセスサービスの支出やその他の情報を取得するための金銭支出については取り扱わないため、探索コストを機会費用としての探索時間として測ることができる。

また、逆の見方をすれば、探索時間は時間を単位にした「時間距離」であり、探索した長さを表しているので「探索距離」として考えることができる。本研究では具体的な探索サービス対価を考えないので、事実上は「探索コスト」=「探索時間」=「探索距離」のように扱える。

コンテンツトランザクションによる「探索距離」

アクセスログのなかで「目的とする URL」を同定できれば、探索行動のトランザクションが決定され、目的とする URL までの探索時間や探索にした経路のホップ数、目的とした URL への滞在時間などが計算できる。

[Cooley 99]を参考にして、ログに記録された URL を目的としたもの（コンテンツ Contents）とそれにいたるために通過した補助的なもの（Auxiliary）と呼ぶ。コンテンツの同定においては、基準を超えて長く参照した URL には意味があると考え参照時間長（Reference Length）方式を採用し、基準値となる滞在時間を超える URL をコンテンツ（Contents）、それ以下のものを補助的なもの（Auxiliary）とする。このとき、補助的に通過したものを考慮するトランザクションを Auxiliary-Contents Transaction、最終目的の URL だけに着目した関係を Contents Only Transaction と呼ぶ。

本研究では、目的とする URL であるコンテンツとそれに至るまでの経路情報（URL、

探索時間、探索ステップ数、滞在時間)をもったトランザクションを総括して「コンテンツトランザクション」と呼ぶことにする。コンテンツトランザクションの同定方法については次節で議論する。

以下では本研究で使用する探索距離の概念として、Auxiliary-Contents Transaction から探索経路の長さや探索時間を測定し、そのコンテンツを探すために費やす探索コストと距離を考える方法を提示する。これにより、Contents Only Transaction をサイト間で集計することで、サイト間の探索コストまたは距離を算出することもできる。

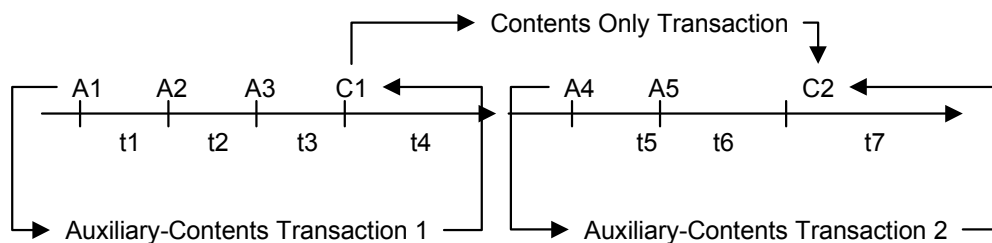


図 5-1 コンテンツトランザクションの考え方

図 5-1はアクセスログを時間軸上に展開したものである。時間軸上に配置されたURL (A1 ~ A5、C1 ~ C2) で、Aは Auxiliary をCはコンテンツを表している。このとき、目的とするコンテンツを探索する Auxiliary-Contents Transaction での各 URL の経過時間の合計を「コンテンツ探索時間」と考える。図 5-1の Auxiliary-Contents Transaction1 における C₁を探索する探索(時間)コストはそれまでの経過時間を合計して以下のように計算する。

$$SearchTimeCost_1^{Transaction1} = \sum_{i=1}^3 t_i$$

さらに、コンテンツトランザクションにおける URL 間の探索時間は、あるコンテンツトランザクションが終了してから次のコンテンツトランザクションが終了するまでの時間である。例をあげると、図 5-1の C₁と C₂の間の探索コストは以下のように計算する。

$$SearchTimeCost_{12}^{ContentsTransaction1to2} = \sum_{i=5}^6 t_i$$

このときの探索距離については、探索コスト = 探索時間 = 探索距離と考えることもできるが、空間的なポジションを明確に区別した表現としては、「探索ステップ数」という表現を用いる。これは、URL から URL への経路の移動を 1 ステップと数えて、目標とするコンテンツまでにたどった経路 (URL から URL) を数えあげた数値である。例えば、図 5-1 Auxiliary-Contents Transaction1 の C₁ までの探索ステップ数は

$$SearchStep_1^{Transaction1} = 4$$

図 5-1 の C₁ と C₂ の間のコンテンツトランザクションでの探索ステップ数は

$$SearchStep_{12}^{ContentsTransaction1to2} = 3$$

となる。

以上のことをまとめると、セッション m のコンテンツトランザクション n におけるコンテンツ URL_i からコンテンツ URL_j へのコンテンツトランザクションの探索コスト C_{ij}^{mn} は、探索時間でも探索ステップ数でも定義することが可能で、

$$SearchTime_C_{ij}^{mn} = \sum_k t_k$$

$$SearchStep_C_{ij}^{mn} = LastNumber_k + 1$$

このとき、 t_k はコンテンツトランザクション n における k 番目の Auxiliary-URL の滞在時間である。ただし、コンテンツトランザクション n においてコンテンツが連続して Auxiliary-URL が含まれない場合は $t_k = 0, LastNumber_k = 0$ である。数上げられたコンテンツ URL_i からコンテンツ URL_j へのコンテンツトランザクションの数を L とすると

$$SearchTime_C_{ij} = \frac{1}{L} \left(\sum_m \sum_n SearchTime_C_{ij}^{mn} \right)$$

$$SearchStep_C_{ij} = \frac{1}{L} \left(\sum_m \sum_n SearchStep_C_{ij}^{mn} \right)$$

と記述することができる。

5.2.2 コンテンツトランザクションの同定

コンテンツトランザクションという仮定のもとでの探索時間と探索ステップ数は定義したが、その前提となるコンテンツトランザクションを事前に同定する必要がある。第

2章の Web アクセスログデータハンドリングで述べた手順で、Web アクセスログからコンテンツランザクションを同定するためのステップを考えると、まず、「セッションの同定」し、「コンテンツの閾値の設定」してから、「セッション内のコンテンツ探索シーケンスを同定」し、「コンテンツ探索時間、コンテンツ探索ステップ数の計算」するという手順をとる。

セッションの同定

まず、セッションを同定する。セッションは一連のランザクションのまとまりであるが、明確な目的をもっているとはかぎらない。たとえば、PC のまえにすわって HP をクリックしはじめて、その作業が一段落して一休みをとるまでの連続した作業と考えるとわかりやすい。このようなセッションの考え方を導入するのは、Web のアクセスログのなかでは、あるサイトに留まってそのサイトの内容を読んでいるのか、閲覧を止めてしまって席を立ってしまったかを判断できないためである。そのために、ある数値をもってセッションの区切りを設ける。第 2 章でも述べたように実務上は 30 分以上のコンテンツ（サイト）の閲覧時間が記録された場合にセッションが終了したと判断することが多い。本研究でも同じ基準を採用することにした*。

参照時間長の基準値の設定とコンテンツの同定

次に、コンテンツを同定する。第 2 章で検討したように、ランザクションの同定には、Maximum Forward Reference[Chen 96]や Longest Repeated Sequence[Pitkow 99]、Time Window[Cooley 99]などの方式があるが、[Cooley 99]で指摘されているようにデータマイニング上でも比較的頑健な評価であった参照時間長（Reference Length）を使う方式を採用する。つまり、ある程度の長さを参照していた URL には意味があると考え、それを本研究では「コンテンツ」と呼ぶ。そのためには、コンテンツランザクションの定義で述べたように、URL をコンテンツと補助的な Auxiliary にわけて考えるために、コンテンツと考える滞在時間（閲覧時間）の判断基準が必要である。

[Cooley 99]では、URL の参照時間長（Reference Length） t が指数分布をすることをとらえて、基準となる参照時間長 RL_t を

* ひとつの URL へのアクセス時間は 30 分（1800 秒）を超えることはなくなる。セッションについては、URL での視聴時間が 30 分を超えると別セッションとして扱うようにした。

$$RL_t = \frac{-\ln(1-\gamma)}{\lambda}$$

とすることを提案している。ここで、 γ は t 全体の分布のなかの補助的な参照 auxiliary reference の割合で、 t は観測された平均参照時間長 (observed mean Reference Length) の逆数である。これは期待値 $1/\lambda$ のパラメータをもつ t の確率密度関数を $f(t) = \lambda e^{-\lambda t}$ としたときの分布関数を $F(t) = 1 - e^{-\lambda t} = \gamma$ として、 t について解いたものである。ここでいう参照時間長 t は、1つのレコードとしてアクセスログに記録された URL から次の URL が記録されるまでの時間で、アクセスログ上の n 番目の URL に記録された時間と $n+1$ 番目に記録された時間の差分である。また、[Cooley 99]によると、 γ は 70% から 90% 程度でも生成された意味のあるルールの数には変化がなく参照時間長方式での頑健性が確認されていたので、仮に 70% とする。

実データの詳細な分析は次節に譲るが、本研究で利用するデータの URL 参照時間長 t の平均値は 65 秒であった (図 5-2)。よって、 $1/\lambda = 65$ であるので、 $\lambda = 0.02$ で近似すれば、70% ($\gamma = 0.7$) を auxiliary が占めるとして、コンテンツトランザクションの閾値 RL を計算すると、およそ 60 秒程度になる。

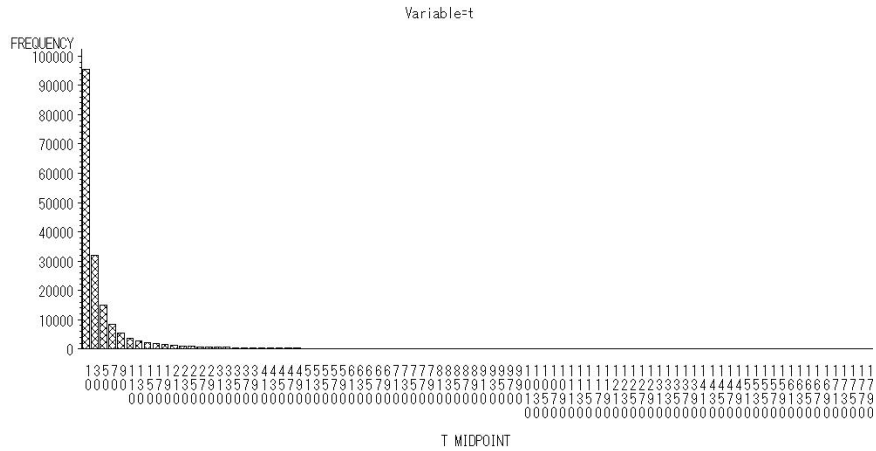
$$\begin{aligned} RL_t &= \frac{-\ln(1-\gamma)}{\lambda} \\ &= -\ln(1-0.7)/0.02 \\ &\cong 60.20 \end{aligned}$$

一方、実際の分布をもとに指数分布しているとして、 $f(t) = \lambda e^{-\lambda t}$ のパラメータ λ を直接推定すると、 $\lambda = 0.04$ 程度に推定される*。 $\lambda = 0.04$ として、先ほどと他の条件は同じにした場合は、RL は 30 秒に計算される。

図 5-2にあるように t の分布の 75% 点は 50 秒程度であることを考えると、分布推定パラメータを用いた閾値では実際のデータに適用すると Auxiliary の比率が 60% に近くになってしまい、[Cooley 99]の頑健性の検証結果ともずれてしまうことから 60 秒をコンテンツトランザクションの閾値と考えることにする。

* 推定にあたっては、統計プログラムの SAS の非線形最小二乗法をつかうプロシージャを利用した。

Transaction Time Distribution



Variable=T

Moments				Quantiles(Def=5)			
N	180799	Sum Wgts	180799	100% Max	1799	99%	937
Mean	64.84055	Sum	11723107	75% Q3	49	95%	277
Std Dev	163.7943	Variance	26828.58	50% Med	17	90%	138
Skewness	5.69611	Kurtosis	39.4881	25% Q1	6	10%	2
USS	5.6107E9	CSS	4.8506E9	0% Min	1	5%	1
CV	252.6109	Std Mean	0.385213			1%	1
T:Mean=0	168.3239	Pr> T	0.0001	Range	1798		
Num ^= 0	180799	Num > 0	180799	Q3-Q1	43		
M(Sign)	90399.5	Pr>= M	0.0001	Mode	1		
Sgn Rank	8.1721E9	Pr>= S	0.0001				

図 5-2 参照時間長 t の分布

アクセスログからのコンテンツ探索時間、コンテンツ探索ステップ数の算出

アクセスログは URL と時間などが記録されているだけなので、図 5-3 のような形に加工する。参照時間長の基準値を越えるものには「c_ch」という項目に 1 のフラグを立てる。図 5-3 のデータでは、OBS はオブザベーション番号、「Domain2」から「Domain」に対するトランザクションになっており、ID は個人番号、SID は個人ごとのセッション番号となっている。

OBS	ID	SID	DOMAIN2	DOMAIN	TC	TS	TSCOUNT	C_CH
141	2	10	nih.gov	nih.gov	179	1	2	1
142	2	10	nih.gov	nih.gov	.	.	1	0
143	2	10	nih.gov	nih.gov	.	.	2	0
144	2	10	nih.gov	nih.gov	.	.	3	0
145	2	10	nih.gov	128.231.164.19	.	.	4	0
146	2	10	128.231.164.19	128.231.164.19	.	.	5	0
147	2	10	128.231.164.19	128.231.164.19	.	.	6	0
148	2	10	128.231.164.19	128.231.164.19	540	120	7	1
149	2	10	128.231.164.19	128.231.164.19	.	.	1	0
150	2	10	128.231.164.19	128.231.164.19	.	.	2	0
151	2	10	128.231.164.19	ccr.co.jp	1800	58	3	1
152	2	11	ccr.co.jp	tsukuba.ac.jp	1800	0	1	1
153	2	12	tsukuba.ac.jp	tokutoku.com	.	.	1	0
154	2	12	tokutoku.com	machiko.or.jp	1800	0	2	1
155	2	13	machiko.or.jp	machiko.or.jp	.	.	1	0
156	2	13	machiko.or.jp	machiko.or.jp	.	.	2	0
157	2	13	machiko.or.jp	machiko.or.jp	.	.	3	0
158	2	13	machiko.or.jp	machiko.or.jp	.	.	4	0
159	2	13	machiko.or.jp	plan-net.co.jp	.	.	5	0
160	2	13	plan-net.co.jp	machiko.or.jp	62	122	6	1

図 5-3 コンテンツトランザクションの同定ステップ*

コンテンツが決定したので、コンテンツ探索シーケンスを同定することができる。ここでデータハンドリングのために、いままで述べてきた探索時間や距離を変数として呼びなおしておく。あるコンテンツの閲覧が終了し、次のコンテンツを探索するという場合、次のコンテンツが見つかるまでの時間を探索時間（変数 *ts*、コンテンツにたどり着くまでの Auxiliary Transaction の参照時間の合計）、そのコンテンツにたどり着くまでのトランザクションの数を探索ステップ数（変数 *tscount*）とする。あるコンテンツを参照していた時間を参照時間または消費時間（変数 *tc*）とする。セッションとコンテンツトランザクションの関係は、図 5-4の通りである。これが個人ごとにユニークに決定される。

図 5-3は SAS を使って、データ処理をした例であるが、コンテンツの閾値を 60 秒に設定し、プログラム処理により自動的に、セッションの同定、コンテンツの同定、探索時間、探索ステップ数、コンテンツの消費時間を計算するようになっている。例えば、図例の一番下にあるオブザベーション番号 160 の URL はコンテンツとみなされ、そこまでの探索時間は 122 秒、探索ステップ数は 6、そのコンテンツの消費時間は 62 秒であることを表している。

* このデータにあるようなセッション区切り前後の参照は、セッションの基準を 30 分 = 1800 秒に設定しているため、1800 秒のコンテンツに数えられてしまうことがあるが、実際のデータ処理からは削除して計算している。

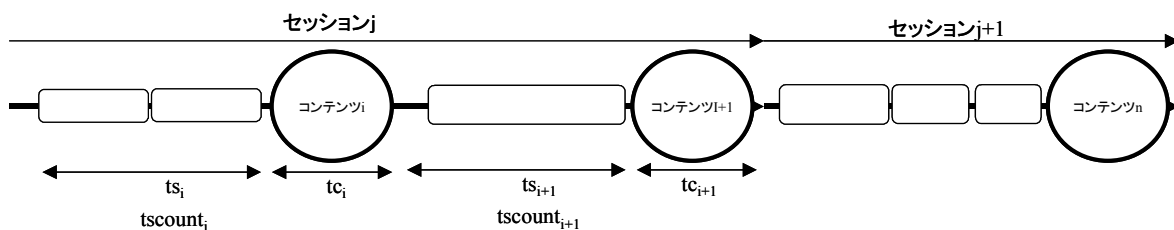


図 5-4 コンテンツトランザクション考え方と変数の与え方

5.2.3 コンテンツトランザクションにおける探索距離の意味

探索参照長をもちいたコンテンツ同定による意味

アクセスログ上で、基準となる参照時間長以上の長さの参照時間をもつ URL をコンテンツとした。通常「探索行動」という言葉で想起されるのは、「探索する目的」があり、その目標に向かって選択肢集合を評価し、目標に一步でも近づこうとする行為である。しかし、本研究の考え方では、実際にユーザがこの URL を目的に探索していたかどうかということの問題にしていない。手順として、アクセスログデータを、参照時間長に重要度においた「コンテンツ」と「探索距離（探索時間）」という概念をもった探索シーケンスに変換していることになる。

前章の研究仮説の設定でも触れたように、このような探索シーケンスへの変換は、具体的な目的をもった探索行動を浮き彫りにするのではなく、大量の URL データをもつ Web グラフの意味を保ったまま縮退して、基準値以上の参照時間長という重要度のある「意味をもつ」URL に出会うまでの時間と距離を定量的に測定することを目指している。具体的な探索の意味を探ることはできないが、探索距離や探索コストを定量的に扱うことで、次節以降の仮説を検証するために必要な情報を抽出することができる。

距離データとしての性格

空間上の2点の1点を n 次のベクトルで表現するような尺度 (measure) を決めると、その2点間がどれだけ離れているのか、どれだけ近いのか、似ているのかどうかを表現する「距離」を定義することができる。そして、距離を定義した場合には以下の性

質が満たされることが求められることが多い[長尾 00]。任意のベクトル x, y, z について距離関数 d が定義されると、

2. 正定値性：0以上の距離をとり、自分自身との距離に限っては0、
 $d(x, y) \geq 0, d = 0 \text{ if } x = y$
3. 対称性：自分と相手の距離は相手と自分との距離に等しい、 $d(x, y) = d(y, x)$
4. 三角不等式：途中を経過すると必ず距離は長くなる。 $d(x, y) + d(y, z) \geq d(x, z)$

しかし、感覚的に2点が離れていることを示すことができればそれを距離とするところもあるので、数学的な厳密さを求めなければ、この2と3は必ずしも成立していなくても慣習的には距離と呼ぶことがある[長尾 00]。

本研究で算出されたコンテンツトランザクションの距離概念では、1) 正定値性、2) 対称性、3) 三角不等式のうち、1は定義上明らかに満たすが、2と3は必ず満たすとは限らない。

2の対称性については、 i から j への探索トランザクションと j から i へのものが同じURLであることは、コンテンツ探索ステップが1のときには比較的起こりやすいが、それ以外のケースでは必ずそうなるとは限らない。

また、3については $i \rightarrow j \rightarrow k$ というリンクが成立している場合に成立するが、ハイパーリンクの性格や検索エンジンによるダイナミックなリンクが発生することも考えると、 $i \rightarrow k$ が $i \rightarrow j$ より短くなることも十分考えられる。

距離概念としては、市街地距離などのグラフ上のリンク数を計算する方法と近い。さらに、Webグラフ上の距離を扱ったもので[Broder et al. 00]はWebのURLリンクの静的グラフ構造からグラフの直径や連結性を探っているが、本研究はおなじグラフ構造ではあっても、静的なリンクではなく、ユーザアクセスログデータから算出する。そのため、静的なリンク構造では検索エンジンやショッピングボット[Brynjolfsson 00a]を利用したトランザクションは含まれないが、本研究のコンテンツトランザクションを用いた探索構造では、検索エンジンのトランザクションの特定はおこなわないものの、そのダイナミックなリンク・選択の結果を取り込んだデータになっている。

そのため、次節のコンテンツトランザクションの統計概要に示すように、ワンクリックでコンテンツに到達しているケースでは、コンテンツ探索ステップ数は1程度のものが多くなる傾向がある。

5.3 データ概要

5.3.1 ユーザプロフィール

データとして、(株)NTT データ開発本部システム科学研究所が 1997 年に実施した「WWW利用状況調査」を借りて分析した。このデータはユーザが事前に登録されているため、通常の Web ログ分析で行うユーザを特定する過程[Cooley 99]を必要としない。ユーザはID管理され、年齢層、性別、Web でのショッピング経験などのプロフィールが登録されている。総数 607 ユーザが登録されているが、今回はそのうち約 200 のユーザプロフィールを抜き出し、実際は 123 ユーザの 3 ヶ月間のログデータ 18 万レコードを利用した。なお、ユーザが ID 管理されているので、アクセスログからユーザを同定するプロセスは必要ない。

このデータには、複数のユーザが数ヶ月にわたって移動した URL が記録されており、特定の Web サイトに偏ることなく、Web での URL のアクセス行動を捉えたものである。厳密な意味でのユーザの無作為抽出をおこなっているわけではないので偏りは避けられないが、これだけの規模と多様性をもっていることで、本研究の仮説検証のためのコンテンツトランザクションの探索距離を計算するには十分なデータであると考えられる。

- ユーザ数：123 人
- 男女構成：男性 56 人 女性 67 人
- 年齢構成比

年齢層	人数	パーセント
回答なし	4人	3.3%
15才以上 - 19才以下	1人	0.8%
20才以上 - 24才以下	12人	9.8%
25才以上 - 29才以下	26人	21.1%
30才以上 - 34才以下	31人	25.2%
35才以上 - 39才以下	19人	15.4%
40才以上 - 44才以下	22人	17.9%
45才以上 - 49才以下	4人	3.3%
50才以上	4人	3.3%

図 5-5 調査ユーザプロフィール

5.3.2 トランザクションデータ

ここでは、データの処理に関する基本的な方針や Web アクセスログ解析で使用される用語の概念について述べて、全体のデータを概観する。

データの取得とクリーニング

ユーザはあるサイトにおいてユーザ登録をおこない、アクセスした URL を記録送信するネットローバーというソフトウェアをインストールしている。ユーザのアクセスデータはネットローバーを通じて、ブラウザに表示される URL が変更されるたびに、ユーザ ID、日時、URL 名などがサーバーに記録されている。そのため、ブラウザのバックボタンでブラウズしたものも含めてすべて記録される。

また、アクセスデータに記録される URL には、HTML ファイル以外に CGI の呼び出し、gif や jpeg などの画像ファイルなどが記録されているが、一枚のページにかなりの枚数が記録されていることもあり、探索過程を検討するにあたっては不要なトレースが自動的に記録されるので削除し、' h t m ' を含むページビューに限定した。このような URL から URL へのアクセスをトランザクションと呼ぶことにする。

トランザクションデータの処理については[Cooley 99]に従い、HTML ファイル単位でのページビューを基本にした。

ドメイン

インターネットサイトや Web サイトという表現もあいまいな表現であり、ドメイン名と混同してつかわれることが多い。ISP に簡易な形でホスティングしているサイトの場合、ISP のドメインを使うこともしばしばある。サイトという表現の感覚としては、同一ドメイン内における同じ看板・ブランドを背負ったページ群といったところである。今回の分析では、分析の効率性から URL の情報はデータのクリーニングにのみ利用し、詳細は分析の対象としていない。そのためインターネットサイトという表現では、サーバー名を含めたドメイン名をひとつの Web サイトとして認識している*。

トランザクションにおける URL へのアクセス時間については、先ほどの記録方法によりその URL にアクセスした（クリックした）時間から次の URL が記録されるまで

* このような場合、大手プロバイダーのドメインに個人 HP などを持っているケースやサーバーホスティングして独自ドメインを取得しないケースなどではいっしょにされてしまうが、本研究ではそれは無視して一つのドメインとする。

の時間となる。また、サイトへの滞在時間ということ考えると、同一サイト名が連続する場合はおなじサイトにいたと考えられるので、配下の異なる URL へのアクセス時間を合計する必要がある。計算アルゴリズムの詳細は付録にゆずるが、同一サイト名がブレイクするまで URL アクセス時間の合計をドメイン滞在時間としている。

トランザクションデータの概観

いままで示した手順で処理されたデータを集計し、表 5-1 に表記した。3 ヶ月間の 123 人のログデータということもあり、約 18 万レコード、月間の平均接続時間は 8 時間 50 分、平均セッション時間は 20 分程度である。あわせて、Nielsen//NetRatings 調査データ[岡橋 00]を記したが、3 年以上の前のデータとはいえ、平均セッション時間や訪問サイト数、接続時間をみても今回のデータセットは比較的汎用的な性格をもっていると考えることができる。

表 5-1 トランザクションデータの比較

項目*	本研究データ	Nielsen//NetRatings	
		日本	米国
データ総対象レコード数 (html) :	180,674		
データ総セッション数 :	9,378		
月間平均セッション数	25	17	18
平均セッション時間 :	00:20:50	00:28:54	00:31:26
月間平均ページビュー	490	780	719
月間平均ページビュー/セッション	19	47	41
月間平均接続時間	08:49:29	08:11:18	09:14:18
コンテンツトランザクション数 :	47,135		
セッション平均コンテンツトランザクション数 :	5.0		
ドメイン数 :	4,309		
月間訪問サイト数	11.7	9	10

コンテンツトランザクションにおけるコンテンツ探索行動データの分布

前述のコンテンツトランザクションの考え方にに基づき、トランザクションデータの分布から参照時間長の基準を 60 秒と設定し、コンテンツに対する探索時間 t_s 、探索ステップ数 t_{scount} 、消費時間 t_c を算出し、そのデータ概要 (表 5-2) と分布 (図 5-6、図 5-7、図 5-8) を示した。

* 「月間平均セッション数」は月当たりの一人平均のオンライン・セッション数。「月間平均サイト訪問数」は月当たり一人平均の訪問サイト数。「月間平均ページビュー数」は月当たりの一人平均の閲覧ウェブ・ページ数。「月間平均接続時間」は月あたりオンラインで過ごした一人平均の時間。

コンテンツまでの探索時間 t_s については、データのなかに探索時間が 0 というデータが全体の 4 分の 1 程度存在している。これはコンテンツが連続している場合に生じる現象で、ブックマークなどに目的のコンテンツを登録しているケースやあるコンテンツが集中している場合が考えられる。

また、コンテンツまでの探索ステップ数 t_{count} についても探索時間同様、探索ステップ 1 というケースが 4 分の 1 程度存在し、考えられる理由は探索時間と同じである。

コンテンツの消費時間 t_c の分布は最小時間を 60 秒としたため、60 秒より必ず大きくなり、セッションの上限を 30 分としているので、コンテンツの最大時間も 1800 秒以下になっている。単純に計測するとセッション区切りのコンテンツはすべて 1800 秒となってしまうので、セッション区切りのデータはコンテンツとしないようにグラフ上は補正してある。

探索時間、探索ステップ数、消費時間の各数値は上記のようにデータの分布は指数的に分布しているのが特徴的である。ちなみに指数分布だと仮定して、パラメータを推定すると、

$$f_{t_c}(t_c) = \beta e^{-\beta(t_c-60)} = 0.03e^{-0.03(t_c-60)}$$

$$f_{t_s}(t_s) = \gamma e^{-\gamma t_s} = 0.27e^{-0.27 t_s}$$

t_s が $\gamma = 0.27$ 、 t_{count} が $\beta = 0.68$ 、 t_c が $f(t_c) = \lambda e^{-\lambda(t_c-60)}$ とした場合、 $\lambda = 0.03$ という推定値*を得ている。

表 5-2 探索データの概要

	平均	最小	最大
探索時間	49.6秒	0秒	4873秒
探索ステップ数	3.83	1	206
コンテンツ消費時間	552秒	61秒	1800秒

* 前出の SAS 非線形最小二乗法のプロシーダを使用した。

Search Time Distribution

Variable=ts

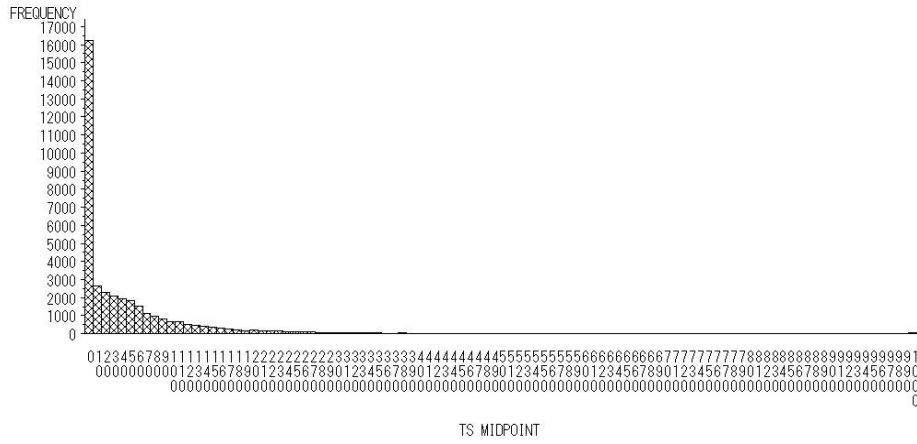


図 5-6 探索時間 ts の分布

Search Step Number Distribution

Variable=tscount

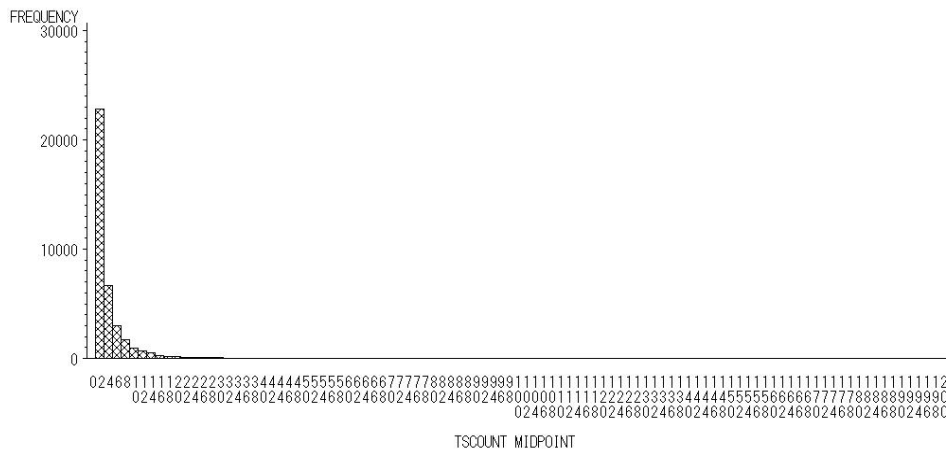


図 5-7 探索ステップ数 tscount の分布

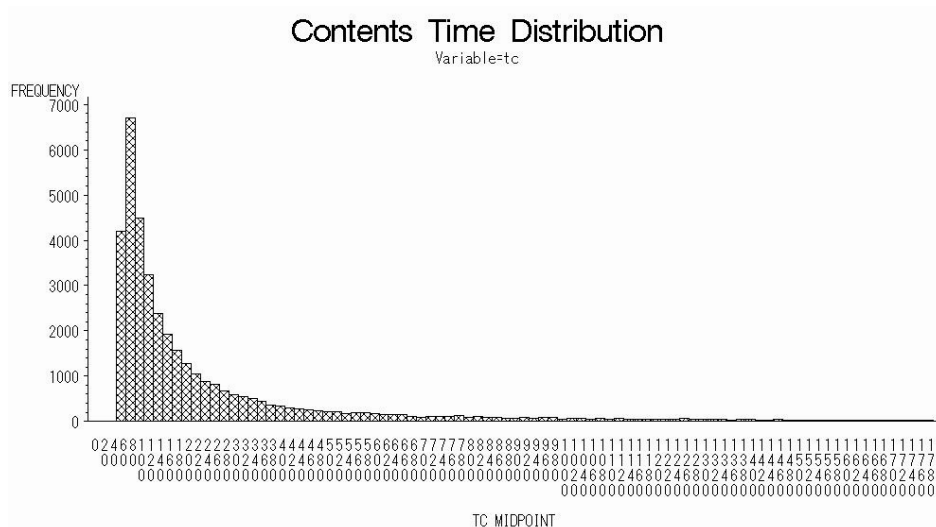


図 5-8 コンテンツ消費時間 tc の分布

探索時間、探索ステップ数、消費時間の関連性

コンテンツに対する探索時間、探索ステップ数、消費時間について、各コンテンツトランザクションのユーザ単位の平均値をプロットした（図 5-9、図 5-10）。探索時間と探索ステップ数については比例関係にあることが読み取れる。長時間探索する場合は探索ステップ数が増加するのは自然な考えである。探索コストとして探索時間を考える場合に、探索の結果生じた探索距離である探索ステップ数もその探索においてどれだけの距離を探索に使ったのかということでは、探索コストと考えることもできるし、このような比例関係を考える場合は、どちらか一方で代用することも問題はないと考えられる。

また、コンテンツの消費時間と探索ステップ数については、ユーザ単位の平均値をプロットしても図 5-10に見られるように相関はみられない。図 5-11はコンテンツトランザクション単位で比較したコンテンツ消費時間と探索時間の平均値の関係をプロットしたものであるが、こちらも同様に関係を読み取ることはできない。

現時点ではコンテンツの消費時間を探索コストに明確な関係性を見出すことはできず、シミュレーションや統計的な調査を行う場合には、消費時間と探索時間は異なる変数として扱う必要がある。

Relation between Search Time and Search Step by User

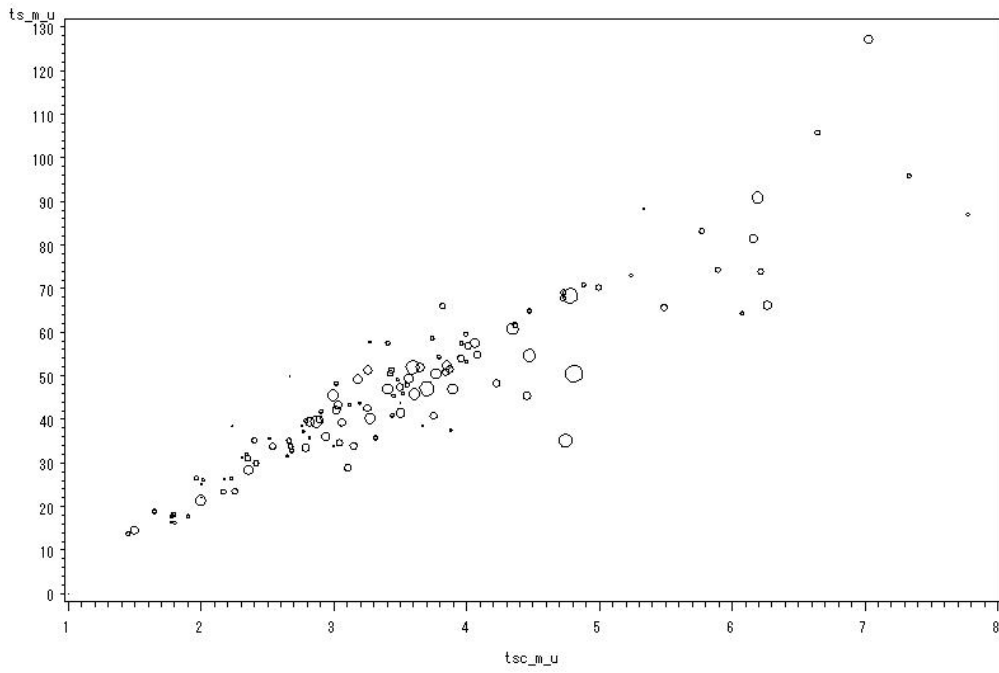


図 5-9 コンテントランザクションにおける探索時間 ts (縦軸) と探索ステップ長 tsc (横軸) の関係

Relation between Contents Time and Search Step by User

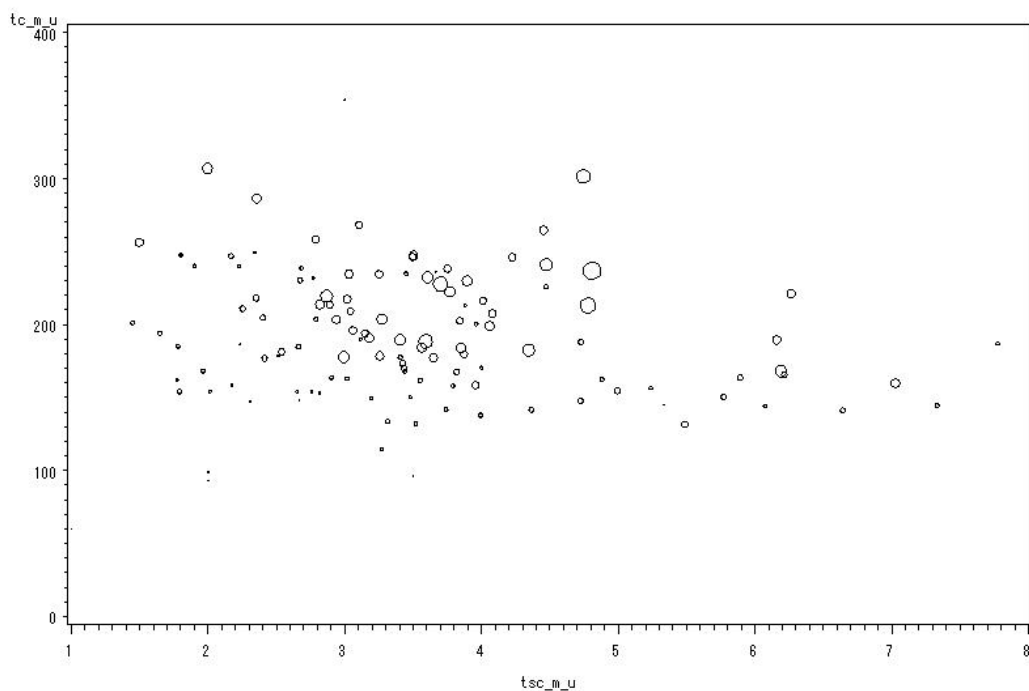


図 5-10 ユーザー単位で集計したコンテンツ消費時間 tc (縦軸) と探索ステップ長 tsc (横軸) の関係

Relation about Searchtime ContentsTime SearchStep by ContentsTransaction
Bubble=ts_s

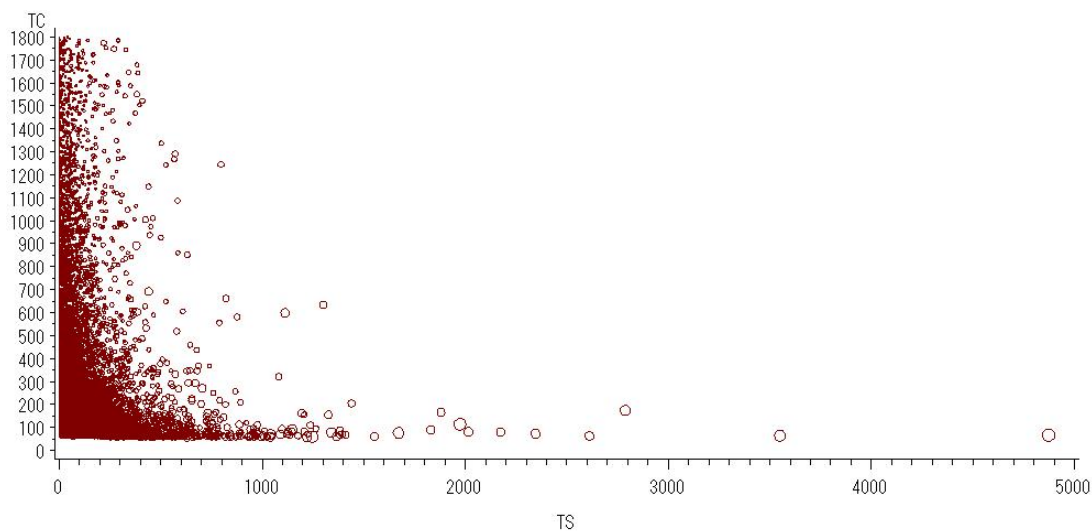


図 5-11 コンテンツトランザクションにおける消費時間 tc (縦軸) と探索時間 ts (横軸) の関係

5.4 グラビティモデルによる探索距離概念の検証

5.4.1 探索コストを距離としたグラビティモデルの構築

グラビティモデルによる Web サイト間のユーザトラフィック交流分析

本節では、ユーザの Web 探索構造分析における第一のアプローチとして、前節で定義した「探索コスト」の概念を距離尺度に応用して Web サイト間のユーザトラフィック交流を分析するためのグラビティモデルを構築する。探索コストが空間相互作用において距離尺度として機能していることについてグラビティモデルを推定することで確かめ、仮説 1：「インターネットでのコンテンツ探索コストは、Web サイト間のユーザ交流トラフィックに反比例する」を検討する。

Web サイトへアクセスとユーザの探索行動を扱って探索コストとトラフィックの関係を分析するためには、空間相互作用モデルや確率選択モデルでの表現が可能である。しかし、第 3 章で用いたロジットモデルのような確率選択モデルのアプローチでは、すべてのユーザに共通の選択肢として大量の選択肢を用意することになる。ロジットモデルはデータごとに選択肢集合が変わってもかまわない[土木学会 95]が、経路や URL リンクごとに大量の選択肢データを用意するのは計算上非現実的である。

そこで、空間相互作用モデルとして代表的なグラビティモデルを採用する。空間相互作用モデルは、都市工学の分野で開発され、商圈立地とマーケティング戦略の分析に用いられることが多く、その一般形は

$$t_{ij} = k u_i^{\beta_1} v_j^{\beta_2} d_{ij}^{\beta_3}$$

t_{ij} は発地区 i から着地区 j への流動量、 u_i は発地区 i の規模（放出性）、 v_j は着地区 j の規模（吸引力）、 d_{ij} は ij 間の距離、 k 、 β_1 、 β_2 、 β_3 はパラメータである。

これを利用したものとして重力（グラビティ）モデルがある。引力が距離の二乗に反比例することのアナロジーによる表現であるが、現実の分析対象が、魅力度とそれを隔てるものに分けられた複数エージェントの交流の特徴を表現するには格好のモデルである[大友 97]。

Web サイト間のユーザトラフィック交流についてグラビティモデルを適用するためには次のように考える。インターネットの世界では都市や店舗に相当するのが、ドメイ

ン、サイト、URL*と表現されるインターネットのサイトである。魅力度や規模についてはそのサイトに集まるアクセス数が人気をあらわしていると考え、サイトのトラフィック†を用いる。そして、距離に相当する抵抗要因をどのように考えるかが大きな課題になる。

そこで、その距離要因にコンテンツの探索コストを用いることで、サイト間のトラフィックが探索コストに反比例することを確認できれば、グラビティモデルにおいて、探索コストが距離要因として機能していることを検証できることになる。

探索距離グラビティモデル

インターネットでの探索距離概念として、コンテンツランザクションにおける探索距離を距離変数において、サイト間トラフィック交流をあらわすグラビティモデルを推定する。

このモデルでは、サイト間ランザクションのトラフィックはサイトの魅力度に比例し、サイトランザクションの距離に反比例すると考える。ランザクションは次のOD (Origin-Destination) 表で考えるとわかりやすい。

表 5-3 グラビティモデルにおける OD 表

		Destination					計
		Site-1	Site-2	Site-3	..	Site-N	
Origin	Site-1	T_{11}	T_{12}	T_{13}		T_{1n}	
	Site-2	T_{21}	T_{22}	T_{23}		T_{2n}	
	Site-3	T_{31}	T_{32}	T_{33}		T_{3n}	
	...						
	Site-N	T_{n1}	T_{n2}	T_{n3}		T_{nn}	
計							

モデルとして、放出側 i の出側トラフィックの合計 ($\sum_j T_{ij}$) を放出側の変数 O_i とし、

* 探索コストの定義でも述べたが、ドメイン、サイト、URL という表現はこのモデルではほぼ同一の意味である。それはこの研究用データがある特定のサイトを対象にしたのではなく、ユーザが探索したサイトを対象にしているため、個別 URL をふくめて膨大な選択リソース量から計算するとモデルによっては計算量が多くなったり、場合によっては少なすぎたりするのを防ぐためでもある。

† トラフィックは一般には交通量を示す言葉であるが、本論文では、電気通信分野でよく使われるように、コミュニケーションネットワークのなかを行き交うデータ量を意味する用語として使用している。交流データなので、ひとつのサイトからみると外に出ていく出側のトラフィック、中に入ってくる入側のトラフィックがある。

吸引側 j の入側トラフィックの合計 ($\sum_i T_{ij}$) を吸引側の変数 D_j とし、 ij 間のトラフィックを T_{ij} 、そのときの ij 間の探索コストを距離 C_{ij} とする。また、トランザクションの性質として、同一サイト内でのコンテンツが多数存在するケースが考えられ、それが大きく影響することが考えられる。そこで、サイト内部のトランザクションの場合にダミー変数 IN_{ij} は 1、それ以外では 2 をとることにした。これは、ダミー変数に対して対数をとることを想定して、対数をとったうえで同ドメインの場合には 0 となり影響せず、他のドメインへのトランザクションの場合に距離抵抗が起こるように調整した。

推定されるモデルは次のようなものになる。

$$T_{ij} = \frac{O_i^\alpha \cdot D_j^\beta}{IN_{ij}^\gamma \cdot C_{ij}^\theta}$$

実際にはこのまま推定するのではなく、両辺の対数をとって、

$$\log(T_{ij}) = \alpha \log(O_i) + \beta \log(D_j) - \gamma \log(IN_{ij}) - \theta \log(C_{ij}) + intercept$$

を重回帰分析してパラメータ α 、 β 、 γ 、 θ を推定する*。

推定にあたっては探索距離として探索ステップ数である `tscout` を用いた。それぞれのパラメータは正の値をとるので、重回帰では α と β は見かけ上 > 0 、 γ と θ として推定される。これらのパラメータがマイナスで推定されればこのモデルの構造が正しいことを示すことができ、探索コストがコンテンツ間のトラフィックに影響するということを検証できる。

5.4.2 推定結果

コンテンツトランザクションの探索コストを用いたグラビティモデル

実際の推定では、ある URL へのトランザクションが一回だけというものが多く存在するため、それがノイズとなって十分なモデル構築ができない。そこで、コンテンツトランザクションのトラフィックが 10 以上または 20 以上のサイト間トランザクションをに限り推定した。(表 5-4)

* 誤解を防ぐため、あえて別の表記をしてあるが、 $T_{ij} = L_{ij}$ であり、探索コストは実際にかかった探索時間や探索ステップ数の合計を ij 間のトラフィックで割って平均したものである。探索コストのなかのトラフィックを分離して説明変数にいれると意味のない推定モデルになってしまうが、探索時間や探索ステップ数の影響を組み込んだ探索コストで推定するという事は大きな問題にはならない。

いずれのモデルでも探索距離係数の符号はマイナスで推定され、このグラビティモデルで探索距離がモデル上の抵抗要因としての距離概念に合致することが確かめられた。また、ドメイン内外の交流をあらわすダミー変数も距離要因に推定され、2つのサイトに同じ規模の魅力があってもドメインがかわるとそれが距離要因になることをあらわしている。トラフィックが少ないサイトを含むケース（モデル1、モデル2）ではこのダミー変数の影響が大きい、ある程度トラフィックが多いサイトについては、このドメインダミー変数がなくても、探索コストを距離要因においたモデルで十分高い説明力をもつことがわかる（モデル3、モデル4）。これはトラフィックが少ないときに内部トラフィックが多くなるケースで、外部トラフィックに対する十分な探索距離データを得られていないことが考えられる。

表 5-4 サイト間交流のグラビティモデルの推定値* () 内は P 値

モデル パラメータ	モデル1 ダミー変数あり トラフィック10 以上	モデル2 ダミー変数なし トラフィック10 以上	モデル3 ダミー変数あり サイト間トラフ ィック20以上	モデル4 ダミー変数なし サイト間トラフ ィック20以上
(放出側の魅力度係数)	0.327 (0.0001)	0.114 (0.0420)	0.338 (0.0001)	0.280 (0.0002)
(吸引側の魅力度係数)	0.397 (0.0001)	0.282 (0.0001)	0.499 (0.0001)	0.442 (0.0001)
(同ドメイン内ダミー変数)	-2.671 (0.0001)		-1.829 (0.0001)	
(探索距離係数)	-0.545 (0.0001)	-0.414 (0.0001)	-0.719 (0.0001)	-0.627 (0.0001)
切片	0.055 (0.7175)	1.445 (0.0001)	-0.18 (0.369)	0.293 (0.258)
自由度修正済み決定係数	0.653	0.270	0.793	0.637
データ数	348	348	160	160

ワンステップのサイト間の推移時間を用いたグラビティモデルの推定

コンテンツ探索行動を仮定しない場合はどうなるのだろうか？同様のグラビティモデルについてコンテンツトランザクションを用いなくて URL 間の推移時間を距離要因においたモデルを推定することができる。単純に時間がインターネットアクセスの壁となっているのなら、URL 間の推移時間が距離として機能するとも考えることは無理な話ではない。

* 表中の推定値は偏回帰係数である。

コンテンツランザクションではないので、探索する過程を反映するのではなく、単に、ある URL から次の URL を移動するワンステップの経過平均時間 t_{ij} をつかう。この t_{ij} はもちろん先ほどの推定したモデルの探索コストとして代替的な探索時間コスト $SearchTime_C_{ij}$ とは異なり、パラメータは τ とする。推定されるモデルは、

$$\log(T_{ij}) = \alpha \log(O_i) + \beta \log(D_j) - \gamma \log(IN_{ij}) - \tau \log(t_{ij}) + intercept$$

である。また、コンテンツランザクションによるデータ圧縮がないので、データ数は数倍ある。

推定結果は表 5-5のとおりであるが、ダミー変数がある場合は説明力があり、一見経過時間が距離要因のように推定されているが、ダミー変数がない場合はほとんど説明力が低く、経過時間の符号もマイナスにはならない上に、パラメータ τ も統計的な確かさをもって検出されないので安定したモデルであるとはいえない。

表 5-5 コンテンツランザクションを用いないサイト間交流のグラビティモデル () 内は P 値

モデル パラメータ	モデル1 ダミー変数あり トラフィック10 以上	モデル2 ダミー変数なし トラフィック10 以上	モデル3 ダミー変数あり サイト間トラフ ィック20以上	モデル4 ダミー変数なし サイト間トラフ ィック20以上
(放出側の魅力度係数)	0.367 (0.0001)	0.096 (0.0001)	0.416 (0.0001)	0.184 (0.0001)
(吸引側の魅力度係数)	0.364 (0.0001)	0.141 (0.0001)	0.420 (0.0001)	0.287 (0.0001)
(同ドメイン内ダミー変数)	-3.756 (0.0001)		-3.618 (0.0001)	
(推移時間係数)	-0.029 (0.0037)	0.080 (0.0009)	-0.074 (0.0001)	0.007 (0.8228)
切片	0.846 (0.0001)	1.959 (0.0001)	0.752 (0.0001)	1.793 (0.0001)
自由度修正済み決定係数	0.714	0.141	0.816	0.389
データ数	1581	1581	885	885

推定結果と仮説 1 の考察

第 4 章で提示した仮説 1 : 「インターネットでのコンテンツ探索コストは、Web サイト間のユーザ交流トラフィックに反比例する」は、コンテンツランザクションの探索コストを用いたグラビティモデルの推定結果において、探索コストを距離項としたパラメータがマイナスで推定され、モデルとしても説明力が高いことから支持された。コンテンツを探するという探索シーケンスでは時間や探索距離といった探索コストを導入することで、時間を含めた探索距離がインターネットアクセスにおける「壁」となってい

ると考えることができる。

一方、コンテンツ探索行動を仮定しないワンクリックだけの URL 間の移動については移動時間が抵抗要因にはなっていないことも確認した。伝統的な現実空間の世界では、空間交流においては移動時間や空間的な距離がその交流の抵抗要因になるとを考えるのは自然なことだが、インターネットの空間交流においては、ワンステップの瞬間の移動では時間の壁が影響しない。

この2つ事実の対比はどのような意味をもつのであろうか。本研究の命題が示すように、インターネットは時間や空間を克服したものとしてもてはやされた。その理由は、どのような URL にもその所在がわかっているならば瞬時に移動できるからである。ワンステップの推移時間はこの URL の所在を知っている場合や現在のページの URL リンクを一回だけクリックする移動のケースを示している。その場合の推移時間は「克服」されたものとして、距離としては機能しない。しかし、目的とするコンテンツ URL を探索するというシチュエーションを想定した仮説 1 においては探索にかかる探索コスト・距離は距離項として機能し、この仮説は支持される。ワンクリックによる URL の瞬時の移動では時間や空間の障壁がなく、目的とするコンテンツを探索するという場合には探索コストが時間や空間の障壁となると考えることができる。第 1 章で述べたようにインターネットに対して過大な期待が集まったのは、この両者を混同してしまってインターネットが全く障壁のない空間だと考えてしまったことによるといえる。

5.5 Web サイト探索グラフ構造による空間構造の検証

5.5.1 Web サイト探索グラフ構造分析に対するアプローチ

分析アプローチと仮説2の関係

本節では、Web 探索構造の分析における第二のアプローチとして、探索コストを距離と定義した Web 探索グラフのマクロ構造[Broder et al. 00]を描き出し、探索行動を反映した Web 探索グラフ構造のもつ特徴を検討することで、仮説 2 : 「Web 探索グラフにおいて、強連結されたグラフ構造に属する Web サイトには、他のクラスターより多くのアクセスが生じている」ことを確かめる。

この分析においては、Web 上の連結グラフの状態をしらべた[Broder et al. 00]の研究が参考になる。彼らは、Web サイトのリンク構造についてロボットをつかってデータを集め、各 URL 間のリンク構造を描き出した。その際の距離には URL 間のリンク数をつかっている。彼らの主な知見は、1) グラフの URL リンクの入次数・出次数の分布がべき乗法則によっていることを確認したことで、2) グラフ構造としての直径を計算し、3) リンクの強連結集合 (SCC) を中心にした Web 全体のマクロ構造を描き出したことである。

本研究であつかう探索構造と[Broder et al. 00]の研究の大きな違いは、同じ Web のグラフ構造をあつかっても、本研究はアクセスログを用いているため、構造と同時にトラフィックを扱えることでもある。そのために、静的なリンク情報だけでなく、検索エンジンやショッピングボットのユーザの要求に応じて生成されるダイナミックなリンク構造を反映した分析になっている。探索行動や探索距離という文脈ではこの違いは決定的である。

また、SCC がクラスターよりトラフィックが多くなることに対して明確なモデルは設定されていないが、密接に連結されたクラスターである SCC では魅力あるコンテンツが連結されやすいとすれば、お互いのトラフィックの相互作用により平均トラフィック数が他のクラスターより多くなる可能性は高い。

本節では、まず彼らの発見した SCC の構造を反映したクラスター構造を Web 探索グラフ構造のなかで確認し、その構造クラスター間でサイト単位のアクセス数 (トラフィック) に違いがあるかを検証していく。それにより、インターネットにおける空間構造の壁が存在しているか、どうかを検証していく。

5.5.2 強連結集合(SCC)によるサイト間交流のマクロ構造

コンテンツトランザクションで構成された Web グラフ構造を探索距離によってサイトを分類し、クラスターとしての構造からトラフィックを比較するために、[Broder et al. 00]が示した強連結集合を核としたマクロ構造を調べる。

サイト 1 からサイト 2 へのトランザクションが計測された場合、それをノード 1 から 2 への有向グラフ $\{1,2\}$ と考えることができる。例えば、 $\{2,1\}$ 、 $\{1,2\}$ 、 $\{3,2\}$ 、 $\{2,5\}$ 、 $\{6,2\}$ 、 $\{2,3\}$ 、 $\{3,6\}$ 、 $\{4,4\}$ 、 $\{1,4\}$ 、 $\{4,5\}$ 、 $\{7,4\}$ 、 $\{4,7\}$ 、 $\{8,6\}$ 、 $\{9,7\}$ 、 $\{10,11\}$ というトラ

ンザクション情報を有向で表現したものが、図 5-12である。

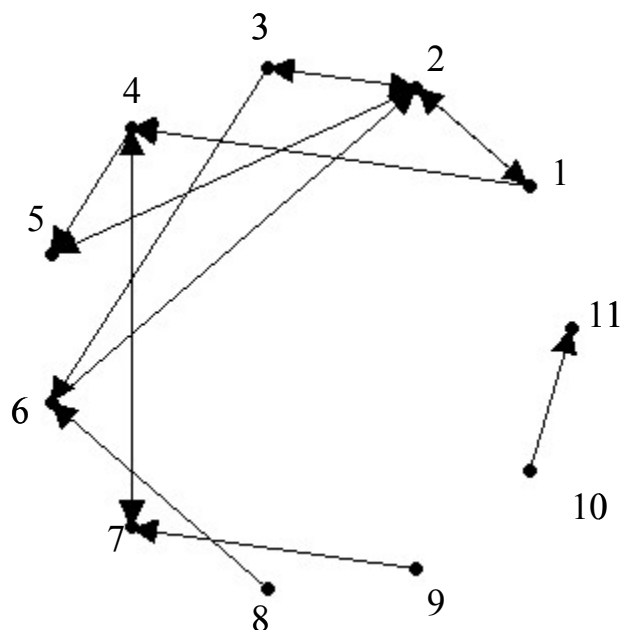


図 5-12 トランザクションのグラフ表現例：有向グラフ

ここで、サイトの連結の様子を考えるために、強連結性という性質を扱う。強連結性は、有向グラフにおいて連結されているどの2つの頂点を選んでもその間に有向グラフが存在するというものである。その頂点の集合を Strongly Connected Components (以下 SCC) という。

たとえば、先ほどの図 5-12の例では SCC は $\{6,3,2,1\}$ と $\{7,4\}$ である。 $\{6,3,2,1\}$ を核にして、各頂点がそこからどれくらいの距離にあるかを示すことができる。このとき、 $\{5,4\}$ は距離1にあり、他の $\{7,8,9,10,11\}$ はこの SCC から連結されていないことになる。

本研究で使用したトランザクションデータを用いてこの SCC のマクロなクラスター構造を描く。コンテンツトランザクションを用いない単純なトランザクションをもとに SCC を計算すると、ひとつのドメイン内のトランザクションが多いため、ほとんどのサイトは強連結されてしまう。

そこで、扱うトランザクションをコンテンツトランザクションにすることで情報を圧縮して、強連結性を検討する。計算には、コンテンツトランザクションのサイトトランザクションを SAS のデータセットから取り出して、Mathematica にインポートし、

Mathematica の Discrete Math ライブラリにある強連結集合を計算するプログラムを使用した。コンテンツトランザクションを使用しているため、サイト数は圧縮されて全体で 3152 である。その結果を図 5-13 に図示した。

SCC を計算した結果、全体の 80% 強を占める巨大な SCC が算出された。そして、全体の 3% 程度のものが SCC から孤立していることがわかった。[Broder et al. 00] でも全体の 5 割を占める大きな SCC が算出されている (図 2-1) が、それと同様に SCC が中心となる構造を確認できた。

以上の準備を通じて、冒頭の仮説検証のために、この SCC からの距離による空間的なクラスターに属していることで、トラフィックなどに違いがあるかどうかを検証していく。

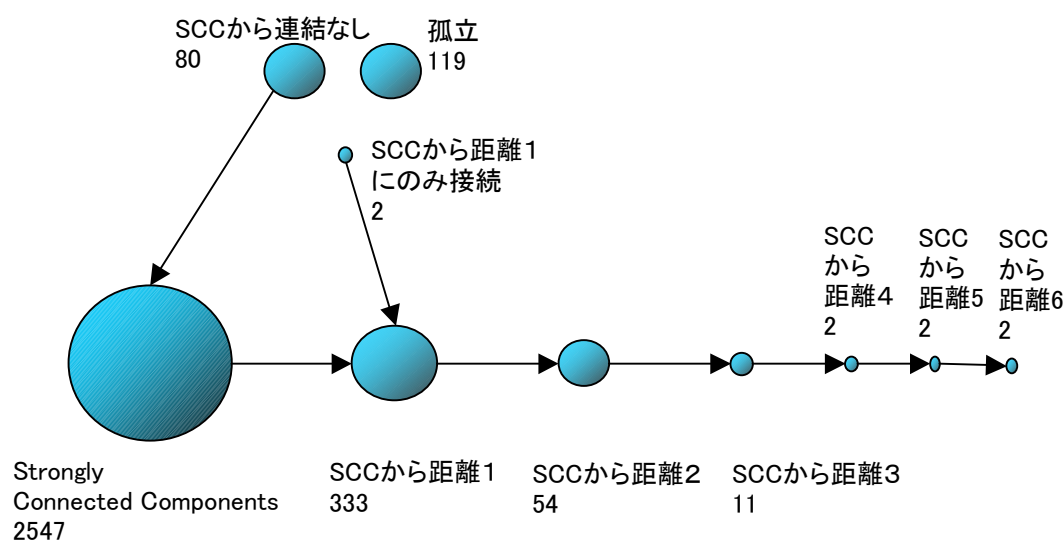


図 5-13 SCC を中心にしたサイト連結のマクロ構造

SCC からの距離によるクラスター間の探索行動指標の比較と仮説2の検証

Web の空間的ロケーションによるアクセス数の違いを検証するために、SCC によるクラスターを利用する。SCC からの連結のステップ距離により、SCC からの連結なし、SCC、SCC からの距離 1、SCC からの距離 2、SCC からの距離 3、SCC からの距離 4、SCC からの距離 5、SCC からの距離 6 というように分類した。この分類をもとに、そのクラスター内での Web サイト単位の平均トラフィック、平均サイト滞在時間、平

均コンテンツ消費時間、平均探索ステップ数を計算し、それぞれを縦軸に、SCC からの距離を横軸に、そのクラスター内のデータ数をバブル状にとり、グラフにプロットした(図 5-14、図 5-15、図 5-16、図 5-17)。Web サイトの SCC からの距離を横軸にしたグラフにプロットしたので、連結しないものは便宜上-1 のポジションにおいてある。クラスター内の Web サイト単位の平均トラフィックとは、表 5-3の Web サイトを単位とした OD 表において、ある Web サイト*i*の出側トラフィック $\sum_j T_{ij}$ と入側トラフィック $\sum_j T_{ji}$ の合計値 T_i を Web サイト*i*のトラフィックとし、それぞれのクラスター内で平均した数値であり、それぞれのクラスターに属している個別の Web サイトへのアクセスの頻度をあらわしたものである。

各グラフをみてもわかるように SCC は他のクラスターと比べて、トラフィックが多く、コンテンツ消費時間が長い(図 5-14、図 5-15、図 5-16)。しかし、コンテンツ探索ステップ長は他のクラスターと変わらない(図 5-17)。これは Tukey の多重比較を SAS の GLM プロシージャで実施して 5 %の危険率で確かめられている(表 5-7)。

SCC は強固に連結されており、そのノード間は有向グラフで結ばれていることで交流が盛んになり、各 Web サイトへのアクセス数も多くなり、さらにコンテンツの魅力度としての時間も長くなるのではないかと推測することができる。

以上のことから、仮説 2 : 「Web 探索グラフにおいて、強連結されたグラフ構造に属する Web サイトには、他のクラスターより多くのアクセスが生じている」と考えることができる。

空間的なクラスターに依存しない性質

仮説 2 を確認したが、一方、図 5-17では探索ステップ数は SCC の距離にかかわらずほぼ一定の値に収束している。つまり、どのようなクラスターに属していても、目的とするコンテンツ URL を探索しようとした場合の探索コストがほぼ一定であることを表している。これは、表 5-7においても同様に Tukey の多重検定により、SCC と他のクラスター間では差がないことが 5 %の危険率で検証されている*。

* 違いが検出されるクラスターもあるが、そのクラスターの構成数が小さすぎるのでグラフや表中には記したが、議論の対象外だと考える。

SCC にはサイトトラフィックが高くなることから魅力的なコンテンツが集まりやすい構造になっていると思われるが、それは単純に探索コストだけの違いではないことが考えられる。

このことは、次節において、Web のグラフ構造的な特徴から考えられる構造要因を検証する過程で再度議論したい。

表 5-6 SCC からの距離によるクラスターの属性

SCCからの距離によるクラスター	サイト数	平均トラフィック (アクセス数)	平均ドメイン滞在時間 (秒)	平均コンテンツ消費時間 (秒)	平均探索ステップ数
孤立	201	20.04	401.71	161.81	4.13
SCC	2547	110.12	273.71	213.86	4.38
SCCから 1	333	13.92	176.98	105.93	4.82
SCCから 2	54	12.52	134.39	109.44	4.81
SCCから 3	11	15.09	152.18	115.08	3.04
SCCから 4	2	11.00	320.00	226.00	31.00
SCCから 5	2	49.00	78.93	74.75	18.00
SCCから 6	2	2.00	256.75	161.75	1.25

SCCからの距離と平均トラフィック

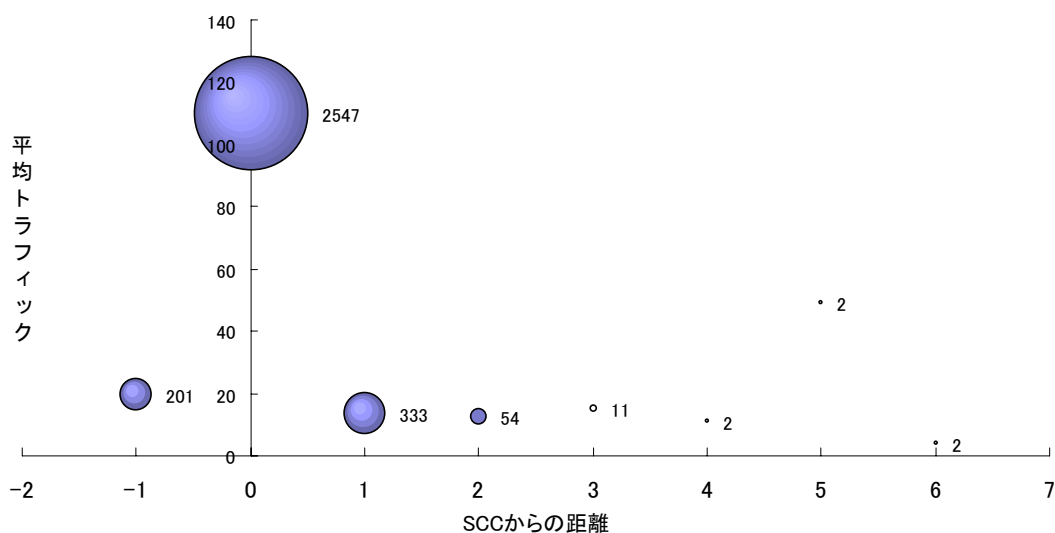


図 5-14 SCC からの距離分類比較 (平均トラフィック)

SCCからの距離と平均ドメイン滞在時間

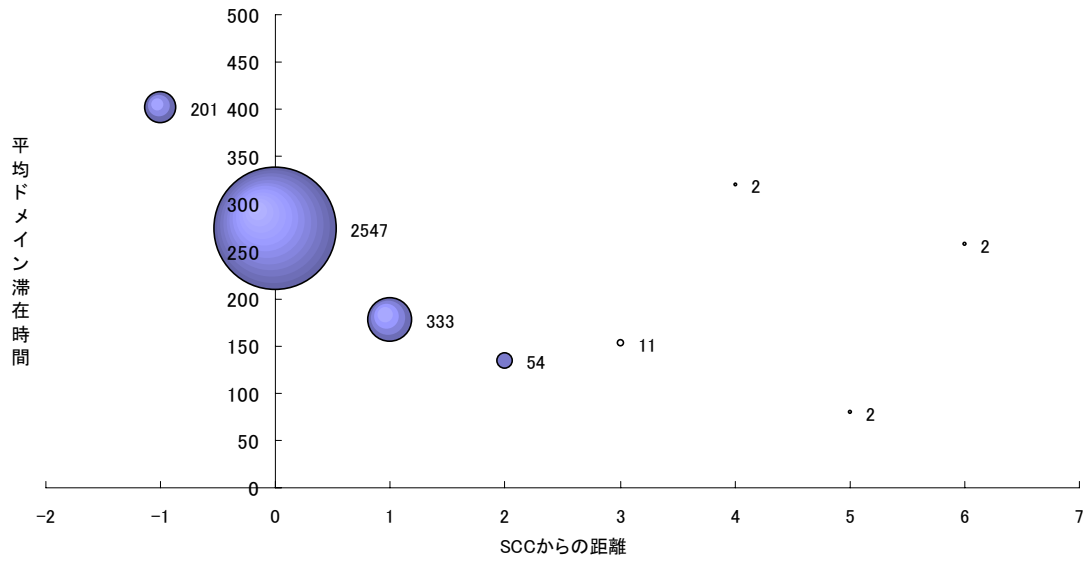


図 5-15 SCC からの距離分類比較 (平均ドメイン滞在時間)

SCCからの距離と平均コンテンツ消費時間

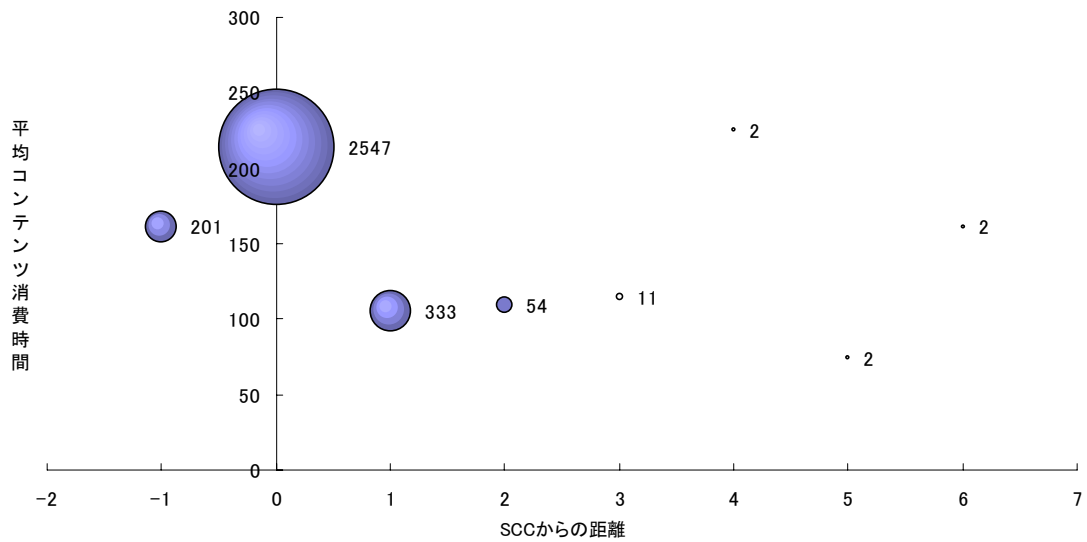


図 5-16 SCC からの距離分類比較 (平均コンテンツ消費時間)

SCCからの距離と平均探索ステップ数

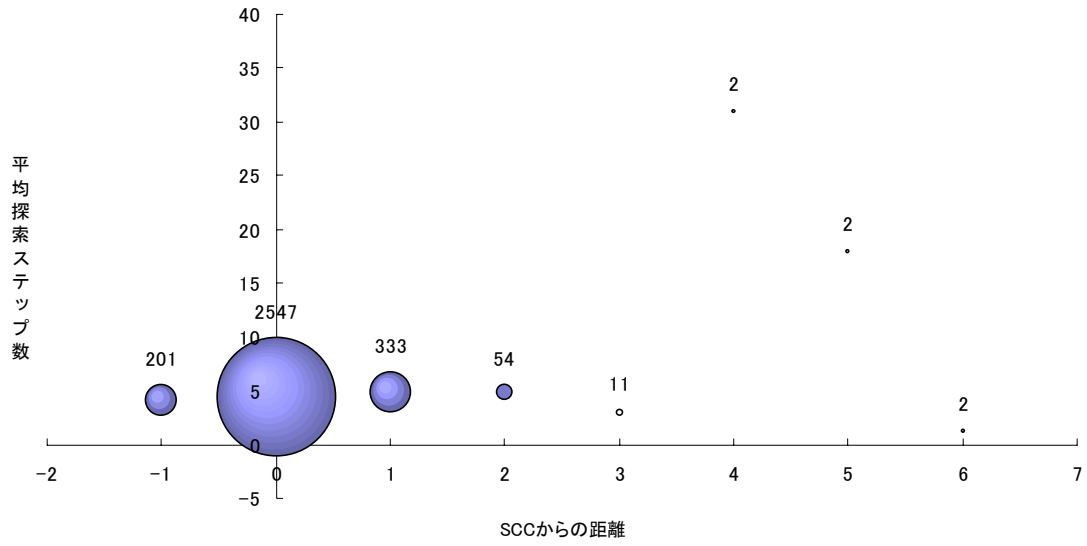


図 5-17 SCC からの距離分類比較 (平均探索ステップ数)

表 5-7 サイト間データの Tukey 多重検定結果

変数名	比較対象		平均の差	5% 有意差
平均ドメイン滞在時間	SCC-	SCCから連結なし	-127.994	***
	SCC-	SCCから距離 1	96.73	***
	SCC-	SCCから距離 2	139.32	***
	SCC-	SCCから距離 3	121.529	
	SCC-	SCCから距離 4	-46.289	
	SCC-	SCCから距離 5	194.783	
	SCC-	SCCから距離 6	16.961	
平均コンテンツ消費時間	SCC-	SCCから連結なし	52.05	***
	SCC-	SCCから距離 1	107.926	***
	SCC-	SCCから距離 2	104.413	***
	SCC-	SCCから距離 3	98.779	
	SCC-	SCCから距離 4	-12.145	
	SCC-	SCCから距離 5	139.105	
	SCC-	SCCから距離 6	52.105	
平均探索ステップ数	SCC-	SCCから連結なし	0.2535	
	SCC-	SCCから距離 1	-0.4414	
	SCC-	SCCから距離 2	-0.4285	
	SCC-	SCCから距離 3	1.3347	
	SCC-	SCCから距離 4	-26.6198	***
	SCC-	SCCから距離 5	-13.6198	***
	SCC-	SCCから距離 6	3.1302	
トラフィック	SCC-	SCCから連結なし	90.086	***
	SCC-	SCCから距離 1	96.205	***
	SCC-	SCCから距離 2	97.602	
	SCC-	SCCから距離 3	95.03	
	SCC-	SCCから距離 4	99.121	
	SCC-	SCCから距離 5	61.121	
	SCC-	SCCから距離 6	106.121	

5.6 Web探索グラフの特徴による構造形成要因

5.6.1 仮説とアプローチの確認

前節の仮説 2 の検証において、マクロなクラスター構造間ではコンテンツの平均探索ステップ数に違いがないことがわかった。[Albert 99]は Web のリンク構造において

も、各ページが比較的小さな距離で接続されていることを報告し、Web グラフ構造の大きな特徴とした。同様に、Web 探索構造においても Web 探索グラフ自体が比較的小さな直径の範囲にコンパクトな形にまとまっていることによって、探索ステップ数に違いがでなかったと推測できる。この Web グラフの特徴を形作る要因と推測されるスモールワールド仮説について、Web グラフの直径を計算することで、[Broder et al. 00]の結果と比較しながら検討を加え、仮説 3 : 「Web 探索グラフの直径は比較的小さく、少ない探索コストでコンテンツにたどり着けるスモールワールドを形成している」ことを検証する。

本節では、まず Web グラフの特徴として多方面で報告されている入次数のべき乗法則を確認し、コンテンツトランザクションの Web 探索グラフ構造でもスモールワールドが成立しているのかどうかを確認するとともに、このような構造を生む背景について考察を加える。

5.6.2 入次数の分布の検証

Web サイトのリンク構造のグラフ的な特徴として、ある URL をノードとした入次数 i を計算し、その次数をもつ確率は $1/i^x$ に比例するとする power law が確認されている [Broder et al. 00][Albert 99]。

[Albert 99]は nd.edu ドメインのドキュメントをすべてしらべ、あるドキュメント (URL) が k 個の入次数または出次数をもつ確率 $P_{in}(k)$, $P_{out}(k)$ は、

$$P(k) \approx k^{-\gamma}$$
$$\text{with } \gamma_{out} = 2.45 \text{ and } \gamma_{in} = 2.1$$

であったことを報告している。同様の報告として、サイトのユーザ数とおなじアクセス数をもつサイトの数の分布 [Adamic 99]、サイトの入 (出) 次数とその入 (出) 次数をもつサイトの数 [Faloutsos 99] [Broder et al. 00] などがある。特に [Broder et al. 00] は $\gamma_{out} = 2.72$ and $\gamma_{in} = 2.1$ と報告している。これらは、都市人口分布と順位的相关に関する Zip の法則の Web 版ともいえるべきものである。

そこで、各 URL について前節の OD 表のようなマトリックスをつくり、サイト (URL) 単位で入次数を計算した。入次数の観測される分布は図 5-18、図 5-19、図 5-20 のとおりである。

分布としては、指数分布的でもあり、べき乗分布で近似しても問題ない程度だと思われるが、これを指数分布で近似した場合は $P(i) = \lambda e^{-i}$ のパラメータ $\lambda = 0.4$ で、べき乗分布の確率関数を $P(i) = \alpha/i^\beta$ とした場合の $\beta = 1.4$ で、 $P(i) = 1/i^\beta$ として推定した場合は、 $\beta = 1.8$ となった。[Broder et al. 00]の研究では $\beta = 2.1$ 前後を推定している。今回の推定値はトラフィックから見てもほぼそれと同等の推定値となっており、経験的にべき乗分布を確かめたともいえる。

表 5-8 入次数のパラメータ比較 (べき乗分布)

パラメータ	本研究	[Broder et al. 00]
モデル $P(i) = \alpha/i^\beta$ の	1.4 ~ 1.8	2.1

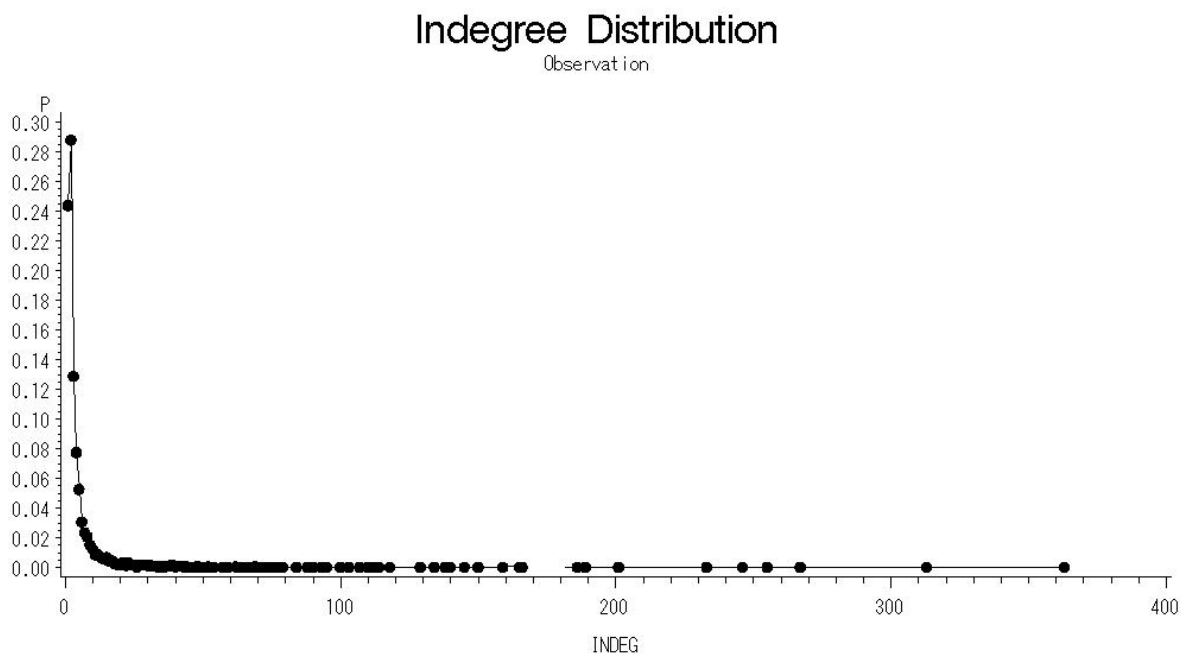


図 5-18 入次数の確率分布 (横軸：入次数)

Indegree Distribution

by Exponential Distribution

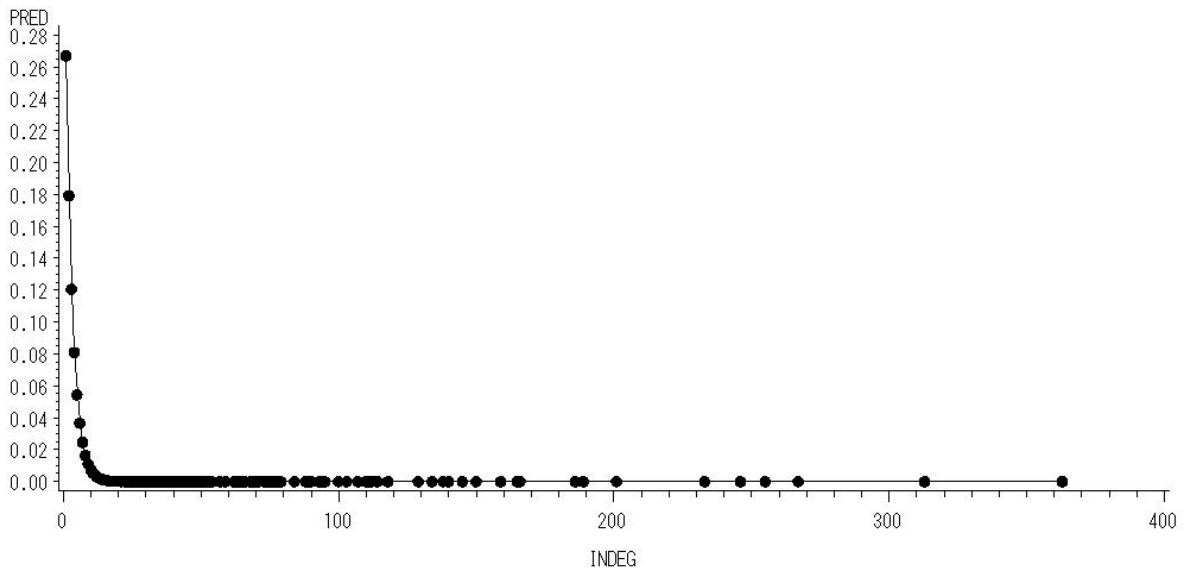


図 5-19 入次数の指数分布での確率分布推定予測値 ($\alpha=0.4$)

Indegree Distribution

by Power Law model $\log p = a - r * \log indeg$

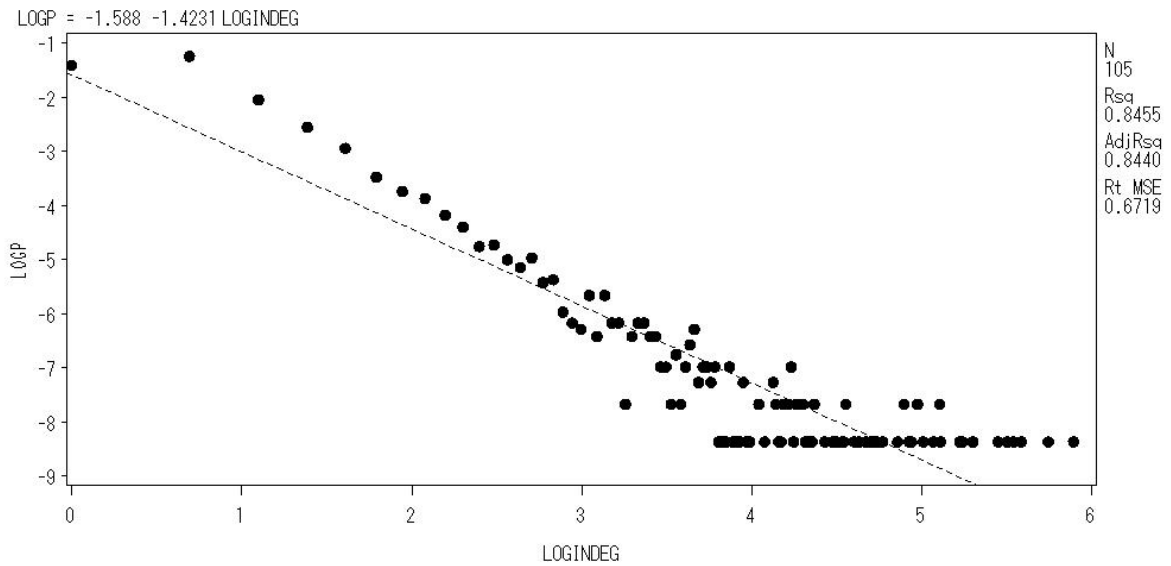


図 5-20 入次数のべき乗分布での対数表現での回帰直線とデータの分布 (予測線は $\alpha=1.4$, $\beta=1.5$)

5.6.3 Web グラフの直径:スモールワールド仮説の検証

[Albert 99][Broder et al. 00]は各 URL から他の URL までの最短パスの長さの平均値をその Web グラフの直径として計算している。[Albert 99]はN個の頂点をもつ有向ランダムグラフを作成し、2つの頂点間の最短パス d を計算した結果、その平均 $\langle d \rangle$ は、

$$\langle d \rangle = 0.35 + 2.06 \log(N)$$

となることを報告している。実際に、 $N = 8 \times 10^8$ では $\langle d \rangle = 18.59$ を得ている。このように、Web は広大な探索空間であるが、 $\langle d \rangle$ が N の対数関数に依存しているため、 N が大きくなってもサイズはあまり変わらない。

[Broder et al. 00]も同様に計測したが、 $\langle d \rangle$ は無向グラフで 6.83、有向グラフで入のリンクで 16.12、出のリンクで 16.18 となっていた。

このような構造は探索空間の形状やその空間の制限要素として様々な示唆をもたらす。

本研究ではコンテンツトランザクションを適用した探索空間を調べている。ここではグラフ構造として探索コンテンツのリンク構造を扱う。計算の仕方は同様にリンク間のパスを計算した。計算にあたっては、数式計算ソフトウェア *Mathematica* の Dijkstra のアルゴリズムを使った。ただし、計算機の性能や時間の制約上、Web 探索構造全体に集合からランダムに 100 個の URL の対を抜き出し、その間の最短経路を計算して平均値を算出した。

その結果、100 個のサンプルの分布は図 5-21 のようになり、有向グラフでの直径は $\langle d \rangle = 3.97$ であった。従来の研究に比べると小さいが、もともとあったトランザクションをコンテンツ間のトランザクションに変換しているため、探索ステップ数の平均値が 3.83 (表 5-2) あることを考えてそれを乗じると、15 から 16 程度になる。コンテンツトランザクションのサイト総数は 3152 個なので、[Albert 99]の法則にいれて計算すると、 $\langle d \rangle = 7.56$ となるので、彼の法則が正しいとしても若干小さめにはなっている (表 5-9)。

他の研究で得られた数値同様、コンテンツ探索シーケンスを用いた Web グラフの構造においても、最短パスで言えば十数個の短いパス (距離) で連結された小さな空間であるといえる。このように仮説 3 にあげたスモールワールド仮説も Web グラフの特徴量から確認できる。

このようなスモールワールドが形成されていることにより、探索を行った場合にはどのような空間構造の位置にあらうと、少ない探索コストですぐコンテンツにたどり着いてしまう。このような事実から、前節の探索グラフのマクロ構造において平均探索ステップ数に差がでなかったのは、このスモールワールドの構造になっていることに起因すると推測できる。

表 5-9 Web の直径データ比較

	本研究		[Broder et al. 00]	[Albert 99]
	平均	× 平均探索ステップ数3.83		
Webサイズの直径<d>	3.97	15.20	16.18	18.59

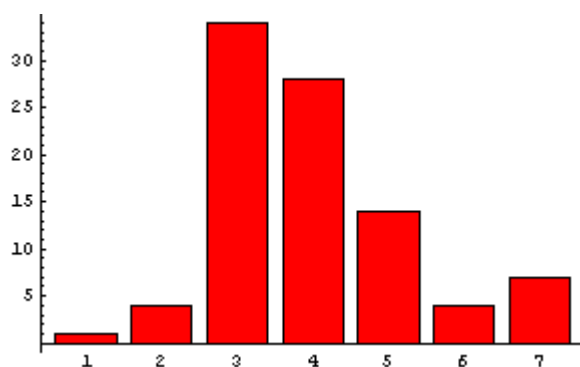


図 5-21 Web グラフの最短パスの分布（縦軸=個数、横軸=最短パスの長さ）<d>=3.97

5.7 本章のまとめと考察

コンテンツランザクションデータの算出

具体的なコンテンツランザクションにおける探索距離を定義するために、行動モデルとして[Bakos 97]らが分析した消費者情報探索モデルの探索コストの概念をもったコンテンツの探索行動を仮定し、[Cooley 99]が提示した参照時間長を用いる方法を使って Web アクセスログからコンテンツを探索するランザクションを同定し、「コンテンツ探索時間」と「コンテンツ探索ステップ数」という2つの量を新しく定義した。

この探索距離は実際のユーザが利用した記録であるアクセスログをつかうことで、Web の静的な URL リンクだけでなく、検索エンジンなどのダイナミックなリンクを反映した距離構造を提示しているが、数学的な厳密さという距離と違って対象性はもっていない。

このコンテンツランザクションの探索距離概念は、具体的な目的・対象をもった探索行動ではないが、重要性の基準をもったコンテンツとそれに出会うまでの時間をつかって、重要度を失わず Web グラフを縮退可能にする抽象的な距離概念である。コンテンツ探索時間はコンテンツを探索するための機会費用をあたえるための探索コストとして考えることができる一方、時間距離ということでは探索距離であるともいえる。探索ステップ数も同様で、実データを用いた比較でも探索時間と探索ステップ数は代替的で、コストであり距離でもあるということが出来る性格をもっている。

コンテンツランザクションの考え方を導入することで、コンテンツに留まっているコンテンツ消費時間を定義できるが、そのデータの上からは探索時間や探索ステップ数との関係を見出すことはできない。

仮説1の検証結果

本章では、探索コストを導入した Web 探索構造分析を通じて、第 4 章で提出した 3 つの仮説について検証することで、「Web をはじめとするインターネットは距離と空間を克服したのか」という命題について考察した。

仮説 1 : 「インターネットでのコンテンツ探索コストは、Web サイト間のユーザ交流トラフィックに反比例する」については、第 4 章で定義したコンテンツランザクションを導入し、探索コストとして探索ステップ数を距離要因に用いた Web サイト間のトラフィックのグラビティモデルを構築した。結果として、探索ステップ数はグラビティモデルでの距離要因になっていることが確認され、第 1 の仮説が成立することを確認した。

これによって、ユーザの URL へのアクセス行動について、インターネットは一見距離と空間を克服しているように見えるが、いままでの空間とは異なる Web 空間での探索行動上の距離がアクセスを阻む距離要因となることを示すことができた。

しかし、探索距離を導入しない URL 間のワンステップの推移時間は空間相互作用モデルの距離要因とならない。このことは探索行動分析におけるコンテンツランザクシ

ヨンの有効性を示すとともに、逆の見方では、URL の所在がわかっているワンステップの移動については時間の制約を離れた行動になっていることを示している。そして、URL の所在がわかっているケースと URL を探索するケースをいっしょに議論してしまっただことで、インターネットにはいつでも時間や距離の制約がなくなったと取り違え、その期待からインターネットへの過度の期待が集まったが、URL の所在を探索する時間や距離の壁にぶつかってしまうことで、Web サイトに思うようなトラフィックを集めることができず、事業として失敗を重ねてしまう原因となってしまったと考えられる。

仮説2の検証結果

仮説 2 : 「Web 探索グラフにおいて、強連結されたグラフ構造に属する Web サイトには、他のクラスターより多くのアクセスが生じている」について、Web グラフの静的な構造のみを扱った[Broder et al. 00]の議論にくわえ、本研究では探索行動モデルを導入することでグラフ構造とサイトアクセスを同時に議論することで、はじめてこの仮説を検証することが可能になった。

探索リンクの強連結集合 SCC を計算し、SCC からの距離による Web リンク空間のクラスターを生成することで、SCC と他のクラスターの間でアクセス数（トラフィック）の違いを確認し、仮説 2 を確かめた。このことにより、空間上のロケーションとして SCC に属しているかどうかコンテンツの平均アクセス数に影響することを確認したことになる。

空間的なリンクや他の Web サイトと提携的な関係を構築する上では、どのロケーションに置くが非常に重要なことになる。Web アクセスにおけるポータルサイトの重要性はいままでも指摘されてきたが、特定のブランド力だけでなく、空間的なコネクティビティが構造的に影響するということがこの仮説を通して確認できた。

仮説3の検証結果: Web グラフの小さな構造

Web グラフの特徴からは、Web リンクの入次数におけるべき乗法則を確認し、Web グラフの直径を計算することで、Web サイトのリンクは比較的小さな探索空間構造になっていることがわかった。これにより、仮説 3 : 「Web 探索グラフの直径は比較的小さく、少ない探索コストでコンテンツにたどり着けるスモールワールドを形成している」ことを確認した。

また、仮説 2 の検証過程で Web グラフの SCC を中心にしたクラスターのマクロな

構造を調べていく中で、クラスター間で平均の探索ステップ数に差がないことがわかった。これはどのクラスターでも探索する長さについては違いがないことを示している。Web リンクグラフの直径が比較的小さいスモールワールドであることが、このような探索構造に影響していると考えられる。この Web グラフにおいて、あるコンテンツを探そうとする場合に、4 ステップ前後の比較的短い探索距離（平均探索ステップ数=3.83）のなかでそれが発見できるのは、それぞれの探索コンテンツが 4 ステップ前後（Web の直径=3.97）でつながっている構造となっているからであると考えられる*。

つまり、Web 探索グラフの「小さな」構造が影響して、クラスター間の平均探索ステップ数に違いがなくなっているといえる。

しかし、今回の検証ではなぜ「小さな」構造が生まれるかはあきらかにされていない。このような特徴が複雑なグラフ構造に発生することについては様々な説明[Davidson 02]が試みられているが、まだ明確な説明はおこなわれていない。べき乗法則と凝集したクラスター構造の相互作用と考えることもできるが、具体的な生成過程を明らかにするのは今後の課題である。

マクロな構造と小さな構造

Web やインターネットが距離と空間を克服しているのかという命題に対して、Web サイトを中心とした視点から距離と空間の構造に焦点をあてた議論を展開した。

それぞれの仮説を検証するなかで、コンテンツ探索行動というスケールを導入することで、Web サイト間の距離とアクセス数の関係や Web サイト間のグラフ的な構造におけるロケーションとアクセス数の関係には、従来の概念でいう距離や空間に近い「壁」があることがわかった。これはいうなれば、Web 探索構造のなかの「マクロな」構造とでも呼べるような関係である。この点においては、インターネットも従来の現実世界も変わりはないといえる。

しかし、一方で、サイト間のワンステップの推移時間はサイトアクセス数には距離項として影響せず、平均のコンテンツ探索ステップ数に違いがなく Web グラフの直径が比較的小さな値に集約しているということからは、コンテンツが Web グラフのどの位置にあるのかに関係なく、コンテンツ探索行動が比較的「小さな」構造のなかに閉じて

* 数値は近いが、平均探索ステップ数と Web の直径（最短パスの平均値）は同じものを測定していない。平均探索ステップ数は、コンテンツトランザクションでの探索ステップ数の平均値であり、Web の直径はランダムに選んだ二点間のノード（コンテンツ）間の最短経路の平均値である。

いるといえる。

Web アクセスの探索構造を考えるには、この「マクロな」構造と「小さな」構造という 2 面性のある性質を深くとらえていく必要がある。ここで、あえて「小さな」構造を「ミクロな」構造と呼ばなかったのは、マクロな構造に対してミクロな構造と呼ぶべきものは主体となる探索者の行動を中心に分析すべきであるという考えによる。

この小さな構造は、ミクロな構造としてのユーザの探索行動に共通する性質が影響して形成されると推測することができる。たとえば、ユーザがもっとも好ましいとする探索参照時間が存在したとして、平均的にはこの探索参照時間をめざしてユーザが探索しているとも考えられる。

成果の応用と課題

第 4 章で定義したコンテンツランザクションの探索距離の概念が距離として、インターネットアクセスにおいて成立することが確かめられたことで、様々な応用が考えられる。

一つは、新規に Web サイトをオープンするときに、他の Web サイトとのリンク構造と探索距離をサンプル調査することで、仮説 1 で検証したグラビティモデルによりリンク先 Web サイトの間に生じるトラフィックを予測できる。それにより、最適な Web サイトの提携戦略を策定することができるようになる。もちろん、対象となる Web サイトのランザクションを知ることができれば、計算された SCC に属している Web サイトにリンクを張ることはアクセストラフィックを向上させる条件である。さらに、SCC に参加するためには片方向ではなく、相互リンクが必要である。

また、平均探索ステップ数と Web グラフのスモールワールドの知見からは、適度な Web サイト構造デザインを考えることができる。Web サイトのなかでの主要コンテンツは平均 4 ステップ以内でデザインされておくべきで、それを超える階層構造は望ましくない。

今後の課題としては、上述のような特徴を生むミクロ構造ともいうべきユーザの探索行動の特徴を抽出することである。次章では、そのミクロな構造を形成するユーザの行動にフォーカスして議論する。

第6章 ユーザ探索パターンの表現と情報採餌行動モデル

本章では、ユーザの探索行動にフォーカスし、Web 探索グラフ構造を生む要因をさぐる。合理的な探索をおこなうユーザを想定し、空間的な探索ヒューリスティックからその行動を推測するアプローチと探索行動モデルを規定する情報採餌行動モデルの2つのアプローチをとる。

前者の表現として、「深さ優先探索（ドメインに留まる）」なのか、「幅優先探索（ドメイン間を移動する）」なのかを記述するドメインジャンプ表現を新たに提案し、ユーザの探索パターンを記述するとともに、探索行動タイプの違いにより仮説4のアクセス構造の違いを検証する。

さらに、[Pirolli 95]による情報採餌行動モデルを導入し、ユーザの探索行動を定式化するとともに、探索戦略である探索参照時間を変化させたシミュレーションを実施することで、探索構造を形作る探索時間の要因について考察を加える。

6.1 本章の目的と仮説

目的とアプローチ

第5章においては、本研究の命題である「インターネットは時間と空間の壁を克服したのか」という問いに対して、抽象的だが定量的に計測可能な探索コストを導入することで、空間の壁を形成している「コンテンツ探索コスト（距離）構造」を検証した。その構造には、SCC のようなマクロ構造が平均トラフィックに影響する空間構造の「壁」として存在するとともに、比較的小さな直径をもつ Web 探索グラフ構造と同じ程度の平均コンテンツ探索ステップ数がマクロ構造に関係なく存在するという特徴が見出された。

第5章では Web サイトを中心にその探索構造を眺めてきたが、本章ではユーザを中心に捉えなおすことで、第5章で検証した特徴を形成する要因を探る。

これまで本研究では、インターネットの探索距離構造に関心があったので、あえて探索メカニズムを規定しない立場に立って仮説検証をおこなってきた。しかし、情報探索行動には探索する目的や探索するユーザのスキルにより探索経路や時間に影響があることが指摘されている[中村 01][Card 01][野島 00][Hoffman 96a]。構造を形成する要因の検証のためには、探索コストには様々な要素が反映されているとはいえ、ユーザの行動を規定する要因を想定する必要がある。

そこで、ユーザの行動原理として、ある期間内での情報取得を効率的にかつ最大化しようという合理的なユーザを想定する。つまり、ユーザの探索行動に合理的な探索戦略が存在していれば、それにそった行動をユーザが取ることで Web グラフの構造に影響を及ぼしていると考ええる。

本章では、第4章で議論したようにユーザの行動から Web 探索グラフ構造を説明する要因を探るために2つのアプローチをとる。

1つは、ユーザのアクセスログデータから探索ヒューリスティックパターンを検出し、そのパターンの合理性を検討することで、

- 仮説4：「Web コンテンツ探索行動において、ユーザは探索の幅と深さを組み合わせたコンテンツ探索ヒューリスティックの中から、探索行動中により多くのコンテンツにアクセスできる合理的なヒューリスティックを選択する」

を検証する。ユーザの探索ヒューリスティックに、アクセスできるコンテンツ数が多いことで効用が高まるという合理性を仮定し、ユーザの選択したヒューリスティックによる Web 探索構造への影響を検討する。

2つめは、ユーザの探索行動として長時間の平均情報獲得量を最大化するという合理的な仮定をもつ情報探餌行動モデルを想定し、実際のユーザの探索行動データ分布を使って、シミュレーションをおこなうことで、

- 仮説5：「Web コンテンツ探索活動全体を通して、コンテンツの探索時間と消費時間を効用関数とした単位時間当たりの獲得情報量を最大化する探索コスト戦略が存在する」

ことを検討し、ユーザの探索コスト戦略の合理性に考察を加える。

探索行動の合理性に関する立場

本章においては、第 5 章で用いた探索行動の考え方を引き続き踏襲している。URL へのアクセス行動を「目的とする URL を探索する行動」と捉えて、その結果発生した「探索コスト」にはブラックボックス的に探索に関わる様々な要素が反映されていると考える。

本章における探索行動の合理性については、探索行動から得られる獲得情報量が大きいものを求めて探索するという合理性を仮定している。この獲得情報量が何に起因するのかによって、効用関数の考え方が仮説 4 と仮説 5 では少々異なる。

仮説 4 は、探索行動を通じてアクセスしたコンテンツの数が多い探索ヒューリスティックを合理的と考える立場で、コンテンツの数が獲得情報量を決定していると考えられる。一方、仮説 5 の詳細は後述するが、獲得情報量は探索時間と消費時間を組み合わせた効用関数を想定して考察を進めている。そのいずれの考え方が支持されるかを踏まえて、ユーザの探索行動における合理的な行動が及ぼす影響について考察を加えていく。

6.2 ドメインジャンプ表現によるユーザ空間探索ヒューリスティックの評価

6.2.1 探索ヒューリスティックに関する仮説と検討課題

検討の進め方

本節では、合理的なユーザの探索行動が Web 探索グラフ構造に影響しているかどうかを確かめるために、仮説 4 : 「Web コンテンツ探索行動において、ユーザは探索の幅と深さを組み合わせたコンテンツ探索ヒューリスティックの中から、探索行動中により多くのコンテンツにアクセスできる合理的なヒューリスティックを選択する」を検証する。

探索空間ヒューリスティックを表現するために、第 4 章で提案したコンテンツトランザクションの概念を利用し、コンテンツを見つけるまでおこなわれる「探索の深さ」と「探索の幅」の 2 つの空間探索ヒューリスティックを組み合わせたパターンを検討する。

そのために、URL のシーケンスに探索戦略としての意味をもたせるドメインジャンプ表現を提案し、ユーザの探索行動パターンの特徴を抽出する方法を試みる。

そして、ユーザを探索行動タイプに分類し、前述した合理性の仮定に従いユーザごとのコンテンツのアクセス数を集計し、探索ヒューリスティックによるアクセス数の違いを検定することで、探索ヒューリスティックの合理性を検討する。そして、その合理的なヒューリスティックを採用するユーザの数を比較することで仮説 4 を検証していく。

探索ヒューリスティックを表現し、仮説 4 を検証していくためには、以下のような課題がある。

- ユーザの探索行動要因の検討
- 探索ヒューリスティックの表現方法の検討
- 探索ヒューリスティックの評価尺度の検討

ここでは、ユーザの探索行動要因について考える。本研究ではユーザの行動要因をあえて決定しない立場で研究をすすめてきた。行動要因にはこの項の冒頭で述べたような探索行動に影響及ぼす「探索目的」「ユーザの探索スキル」などの多くの要因が存在する。それを一意に決定するのは膨大な検討が必要になるし、それは本研究の目的ではない。

そこで、本研究では、情報獲得量が多い行動を選択する「合理的な」ユーザを想定し、それぞれの探索ヒューリスティックの合理性を検証するという立場をとる。そう考えると、「深さ優先」と「幅優先」という探索パターンはユーザの具体的な目的にかかわらず、Web の空間のなかで木構造を深堀する方が獲得情報量が高く、ユーザの効用が高くなるのか、広く浅く探索する方が効用が高くなるのかを評価することになる。

そして、その探索ヒューリスティックが合理的だとみなされれば、それを採用するユーザの数は多くなることが予想され、結果としてどの探索ヒューリスティックを選択するのかという探索戦略が Web グラフの探索構造に大きな影響を及ぼすことが考えられる。

このような考えに従うと、今後の Web の発達の動向次第では、Web 探索グラフが深さ方向に発達したり、あるいはコンテンツデザイナーがユーザ効用を想定して幅方向に配置するコンテンツを充実したりする可能性がある。

果たしてそのような影響を計測することができるかどうか、次に具体的にアクセスログ情報からこの探索ヒューリスティックを表現するための手法について検討する。

6.2.2 ドメインジャンプ表現による探索ヒューリスティックパターン

ドメインジャンプ表現

アクセスログのなかでは、ユーザのアクセス行動は、たとえば、www.nifty.com www.machiko.or.jp www.yahoo.co.jp www.yahoo.co.jp www.nikkei.co.jp といったような URL のトランザクションの列として表現されている。この表現は実際の足跡を表現しているので Web サイトの定性的性質を観測するには適しているが、コンテンツトランザクション間の比較やユーザ間の定量的な特性比較をおこなうには明らかに情報が多すぎる。また、URL をサイトのドメインに絞った場合でも、URL 数が非常に多いため共通構造を見つけるためには適さない。

そこで、コンテンツトランザクションにおいて、目的とするコンテンツにいたるまでのトランザクションで同一ドメインにいる場合は0、次のステップで別のドメインに異動した場合（これを「ドメインジャンプ」とよぶ）は1と表現することで、それぞれのコンテンツトランザクションを0 1の数値の列で表現できる。たとえば、あるユーザのコンテンツトランザクションを0 1の組み合わせパターンとして表現し、パターンごとに頻度を集計することで図 6-1のようなパターン表現が可能になる。このようにして、膨大な URL 情報に左右されることなく探索パターンを表現できる。

第5章でのコンテンツトランザクション表現との大きな違いは、コンテンツだけに注目しているのではなく、コンテンツに至る過程に注目しているということである。

LIST2	Frequency	Percent
1	221	41.2
1,0	84	15.6
1,0,0	39	7.3
1,0,0,0	29	5.4
1,1	25	4.7
1,0,0,0,0	12	2.2
1,0,0,0,0,0,0,0	10	1.9
1,0,1	9	1.7
1,0,0,0,0,0,0	7	1.3
1,0,0,0,0,0,0,0,0,0	7	1.3
1,0,0,0,0,0	6	1.1
1,1,1,1	5	0.9
1,0,0,0,0,0,0,0,0	4	0.7
1,0,0,1	4	0.7
1,1,0,0	4	0.7
1,0,0,0,0,1	3	0.6
1,0,0,0,1	3	0.6
1,0,1,1	3	0.6
1,1,1	3	0.6
1,0,0,0,0,0,0,1,0	2	0.4
1,0,1,0	2	0.4
1,0,1,1,0	2	0.4
1,1,0,0,0	2	0.4
1,1,1,0,1	2	0.4
1,0,0,0,0,0,0,0,0,0,0,0	1	0.2
1,0,0,0,0,0,0,0,0,0,0,0,0	1	0.2
1,0,0,0,0,0,0,0,0,0,0,0,0,0	1	0.2

図 6-1 ドメインジャンプパターンの例

6.2.3 探索行動タイプ: ユーザ探索ヒューリスティックのパターンの分類

幅優先探索と深さ優先探索

ユーザのアクセスログには、ユーザが「コンテンツ」探索にもちいた探索ヒューリスティックの結果が表現されていると考えることができる。そこで、膨大な URL 情報の中から抽出するために、前項のドメインジャンプ表現を利用する。

先ほどの 01 で表現したアクセスのパターンコードを探索ステップ数とドメインジャンプ回数に分解することで、ひとつのドメインにとどまって深く探索するのか、あるいはドメインをジャンプしながら探索するのかを検討することができる。つまり、一つのドメイン内を深堀していく「深さ優先」探索か、ドメインを渡り歩き、浅くバリエーションを探す「幅優先」探索を表現することが可能になる。

コンテンツ探索トランザクションのドメインジャンプ表現のままでは、この2つの空間探索ヒューリスティックパターンを表現できないので、ドメインジャンプ表現による 01 のユーザ探索パターンを（探索ステップ数、ドメイン移動回数）に変換する。たとえば 1001 なら探索ステップ数は 4、ドメインジャンプ数は 2 となり、パターンコードを（4,2）と表現することができる。パターンコード別に頻度を計算し、それを個人の

アクセスパターンベクトルとして表現できる。

$$\begin{aligned} & \textit{PatternVector_IDn} \\ & = (\textit{Frequency}(1,1), \textit{Frequency}(2,1), \textit{Frequency}(2,2), \dots, \textit{Frequency}(m, m)) \end{aligned}$$

これらの内積を計算することで、個人アクセスパターンの類似度を計算できる。しかし、このままでは探索過程の特徴を空間的には表現できない。

ユーザの探索行動タイプの分類

ここでの課題はどのようにして、このドメインジャンプ表現によるユーザの探索行動パターンを「深さ優先」「幅優先」というタイプに分類できるかということである。

そこで、探索ステップ数を縦軸にとドメインジャンプ回数を横軸にとった 2 次元のグラフで表現してみると、ID 番号 2 番と 25 番の探索をプロットすると図 6-2、図 6-3 のようになる。ID2 番はドメインジャンプ回数と探索ステップ数が比例するような探索をしている傾向がグラフから読み取れる。また ID25 番はドメインジャンプ回数より探索ステップ数の方が多いような探索をしているような傾向がうかがえる。

このような両者の傾向をヒントに、このグラフをもとに次のようなユーザの探索行動タイプの表現を考えることにする。

- ドメインホッパー：幅優先探索を採用し、ドメインを次々に飛び歩きながら探索するケース。探索ステップ数とドメインジャンプ回数が等しくなることが多いので、グラフ上では傾き 1 に近い直線上に分布する。
- ドメインスティッカー：深さ優先探索を採用し、ひとつのドメインを掘り下げて探索するケースで。縦軸方向に広がって分布する。
- バラエティシーカー：両方を組み合わせて探索するケース。明確な探索パターンを見出せない行動タイプ。

このようなユーザの探索行動タイプをグラフ上で表現すれば、横軸にドメインジャンプ回数を、縦軸に探索ステップ数をとることで、図 6-4 のように各行動タイプを表現できる。

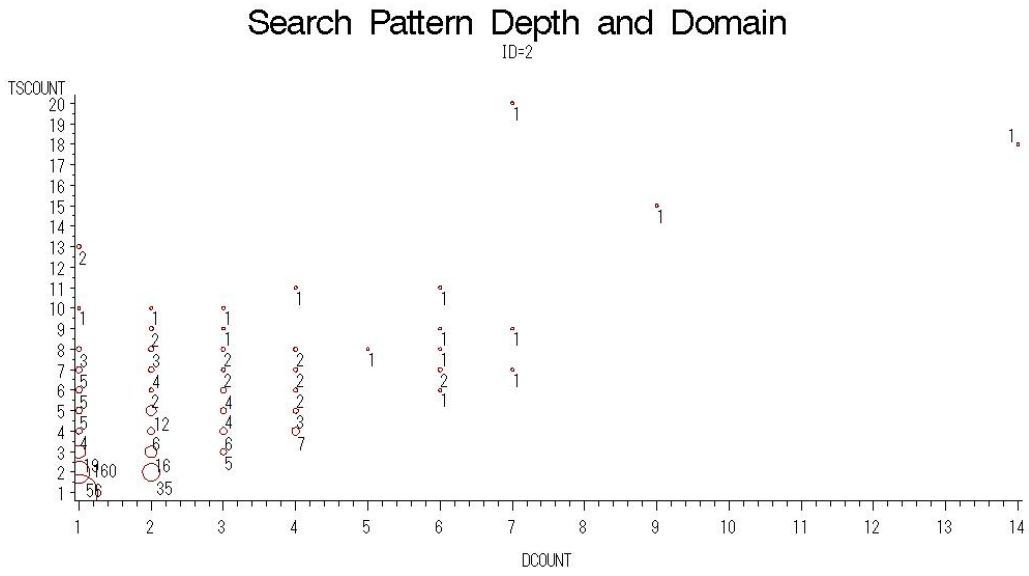
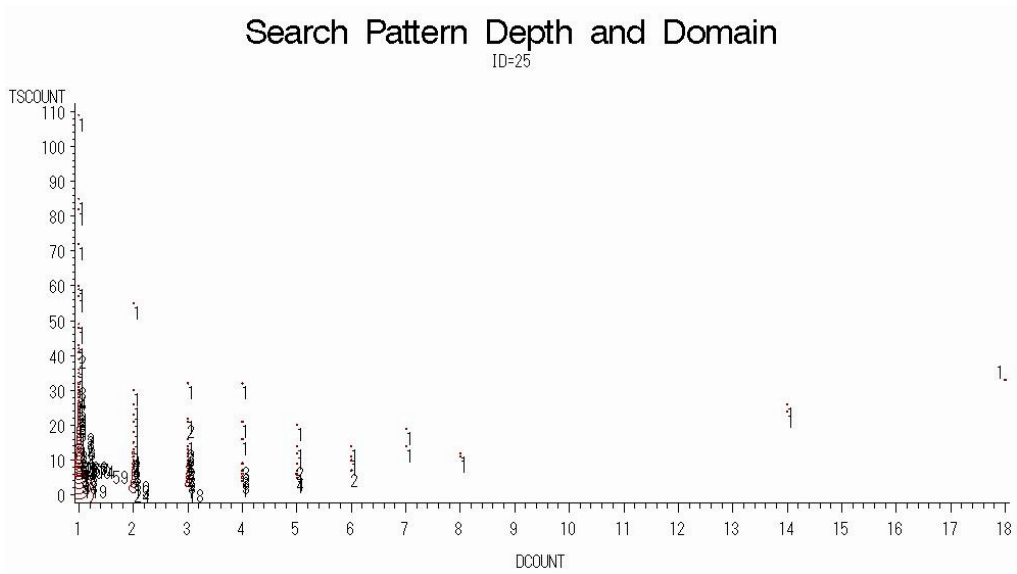


図 6-2 ID2 番の探索パターン (横軸=ドメインジャンプ回数、縦軸=探索ステップ数)



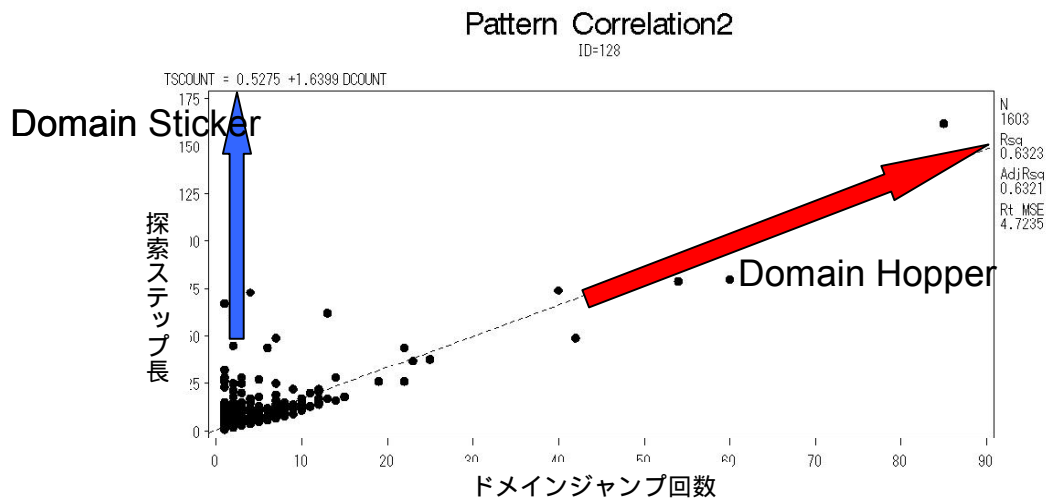


図 6-4 探索ステップ長とドメインジャンプによるパターン

探索パターン分類基準の設定

ユーザの探索行動タイプを上記のような分類に当てはめるためには、ある基準をきめて数値的に定義する必要がある。そこで、回帰分析の回帰係数と決定係数を利用してそれぞれの閾値を定めることを提案する。等分散性のないデータに回帰分析を用いては正確なパラメータを推定することはできないが、それを逆にとる。

図 6-4に示されるように、ドメインホッパーでは回帰係数が1に近く決定係数のあてはまりもよく、また、ドメインスティッカーでは決定係数のあてはまりが非常に悪くなることが予想できる。回帰係数と決定係数を組み合わせて、ユーザ探索行動パターンの類型化を試みる。

それぞれの区分となる閾値を算出するためには、ユーザごとに回帰係数と決定係数を計算したデータを作成し、SAS Fastclus プロシージャをつかって図 6-5のようなおよそ8つのクラスターをつかった。このクラスターをもとに、回帰係数が1に比較的近いグループ(1, 2, 8, 6のクラスター)、決定係数で別れるクラスターが明確なポイント(7と1, 2のクラスターの境目)ということに基づに、図 6-5からポイントを探し、回帰係数が2.5以上のものをバラエティシーカー、回帰係数が2.5以下で自由度修正済み決定係数が0.3以上のものをドメインホッパー、前後を境界に、回帰係数が2.5以下で自由度修正済み決定係数が0.3未満のものをドメインスティッカーとした(図 6-6)。

このように閾値の数値を設定することで、クラスター分析で創出されるグループに意味付けできると同時に、その意味が恣意的ではなく、深さと幅の探索パターンの類型を前提にした表現と分類をあらわすことができる。

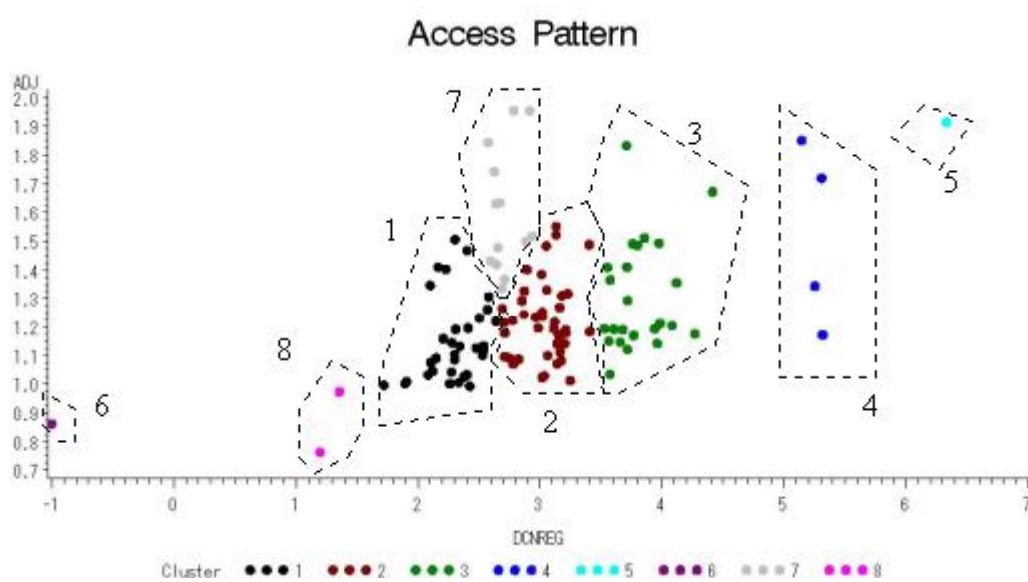


図 6-5 sas fastclus プロシージャで 8 つのクラスターを定義した場合のクラスター（縦軸 = 決定係数、横軸 = 回帰係数、数値は表示上で計算値に+1 されている）

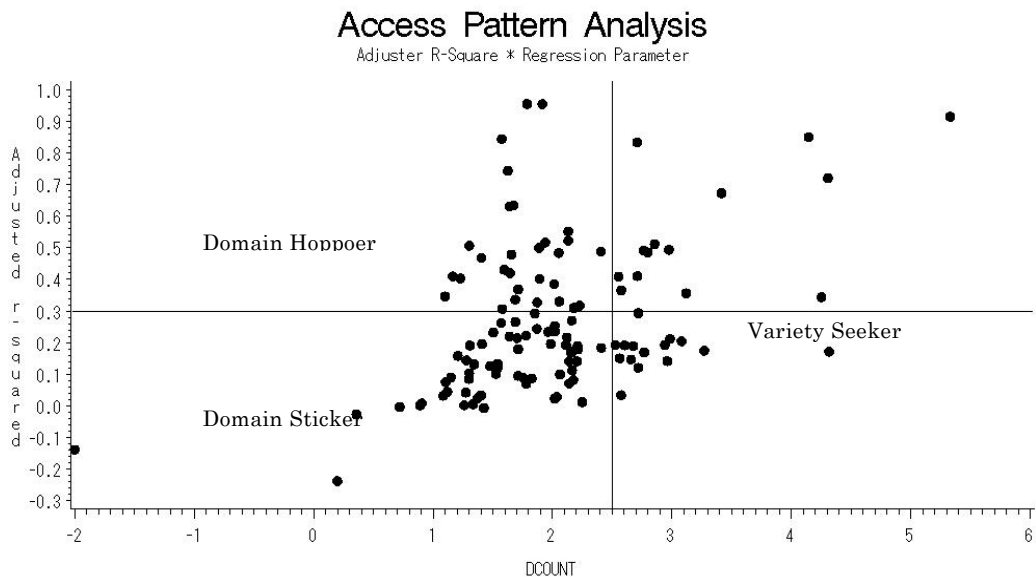


図 6-6 回帰係数と決定係数によるパターン分け（横軸 DCOUNT=傾き、縦軸=決定係数）

6.2.4 分析結果

利用データ

分析に使用したデータは第 5 章の Web 探索グラフ構造の検証にもちいたアクセスログとおなじものである。今回は Web サイト単位で集計したものを、今回の分析ではユーザ個人単位に集計しなおし、コンテンツトランザクション毎に整理する。目的とするコンテンツまでの URL の変化をドメインジャンプ表現に置き換え、前節の探索行動タイプの基準に当てはめて、ユーザの探索ヒューリスティックパターンによるアクセス数の違いを検証した。

探索行動タイプによる仮説4の検証(探索時間、消費時間、コンテンツアクセス数)

仮説 4 : 「Web コンテンツ探索行動において、ユーザは探索の幅と深さを組み合わせたコンテンツ探索ヒューリスティックの中から、探索行動中により多くのコンテンツにアクセスできる合理的なヒューリスティックを選択する」ことを検証するために、ユーザの探索行動タイプを、ドメインホッパー(幅優先)、ドメインスティッカー(深さ優先)、バラエティシーカー(混合)の3つに類型化して、それぞれのコンテンツ探索における探索時間、探索ステップ数、コンテンツ消費時間、一人あたりのアクセスしたコンテンツ数の平均値に違いがあるかどうかを確かめた(表 6-1)。

まず、探索ヒューリスティックの合理性を探索行動中にアクセスしたコンテンツの数が多い方が効用が高いと仮定したが、Tukey による多重検定ではそれぞれの行動タイプによって、一人のユーザがアクセスしたコンテンツ数の差を統計的に検出することができなかった。それぞれの探索ヒューリスティックのシェアはスティッカーがもっとも多いことを考えると、この時点で仮説 4 は成立しない。

さらに、一つの探索シーケンスを構成するコンテンツ探索時間と探索ステップ数、消費時間については、探索時間はホッパーの方がスティッカーやバラエティシーカーより長く、コンテンツ消費時間はホッパーが最も短く、ついでバラエティシーカー、スティッカーという順に長い。この違いは Tukey の多重検定の結果 0.05% の危険率で確かめられている。

表 6-1 探索行動タイプによる探索データ* ()内の符号については脚注参照

コンテンツ探索データ	Domain Hopper	Domain Sticker	Variety Seeker
探索時間	59.35秒(+)	45.39秒(++)	58.39秒(+)
探索ステップ数	4.31(+)	3.56(++)	4.38(+)
消費時間	225.18秒(+)	257.54秒(++)	246.49秒(+++)
一人当たりがアクセスしたコンテンツ数	341.52個/人	361.65個/人	191.07個/人
人数	29人	62人	29人

6.2.5 探索ヒューリスティックにおける合理性の考察

アクセスしたコンテンツ数が多い探索ヒューリスティックの方が効用が高いとする合理性を前提にした仮説 4 は成立しなかったが、ユーザが最も多く採用したヒューリスティックはスティッカーであった。

スティッカーの探索ステップ数や探索時間は他のヒューリスティックより短く、単純にコンテンツに遭遇するということだけから考えると、単位時間当たりではスティッカーがもっとも効率がよい。しかし、ユーザー一人当たりがアクセスしたコンテンツ数の違いに明確な違いが見出せなかったのは、スティッカーのコンテンツ消費時間が他のヒューリスティックより長いいため、結果として全体の探索行動の中ではコンテンツへのアクセス数に明確な違いが見出せなかったと推測することができる。

スティッカーとホッパーの探索時間配分を比較すると、ホッパーは各サイトをめぐりながらスティッカーより長く探索し、一つ一つのサイトで過ごす時間は短くなっている。逆に、スティッカーはじっくりと一つのサイトを掘り下げて、短い探索ステップ数でコンテンツに遭遇するが、消費時間は長くじっくりと参照するので、最終的に両者のコンテンツ数の違いは明確でなくなってしまった。このように、両者の探索パターンの違いは、空間的な探索構造だけではなく、探索時間配分にも違いを見出すことができる(図 6-7)。ユーザのアクセス数の分布からも、スティッカーには短い探索時間でコンテンツに多く遭遇するユーザも多いが、長く消費時間を使うユーザも多く存在する傾向が読み取れる(図 6-8)。

ドメインスティッカーの人数は他のパターンの2倍程度存在し、この3つのパターン

*表中の符号+,++,+++はTukeyの多重検定により異符号間の差が5%の危険率で有意である。また符号*と**は二群の平均値の検定で5%の危険率で有意であった。

の中では支配的な行動タイプであることを考えると、この探索ヒューリスティックにはなんらかの合理性があることが推測される。しかし、探索行動中にアクセスしたコンテンツ数はそれを測定する指標にはなりえず、ドメインスティッカーの深さ優先探索ヒューリスティックが示すように、比較的短いコンテンツ探索時間を持ち、比較的長くコンテンツを消費する探索方法の合理性を示している可能性が考えられる。

課題として、探索時間や探索ステップ数などの探索コストとコンテンツの消費時間を組み合わせた効用関数を想定する必要がある。

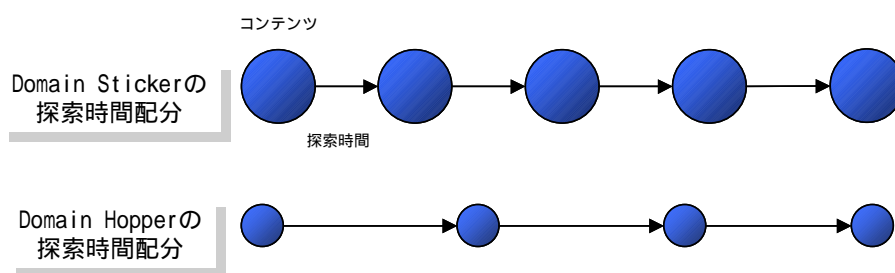


図 6-7 探索戦略による時間配分パターン（円のサイズ：コンテンツ消費時間、：探索時間の長さ）

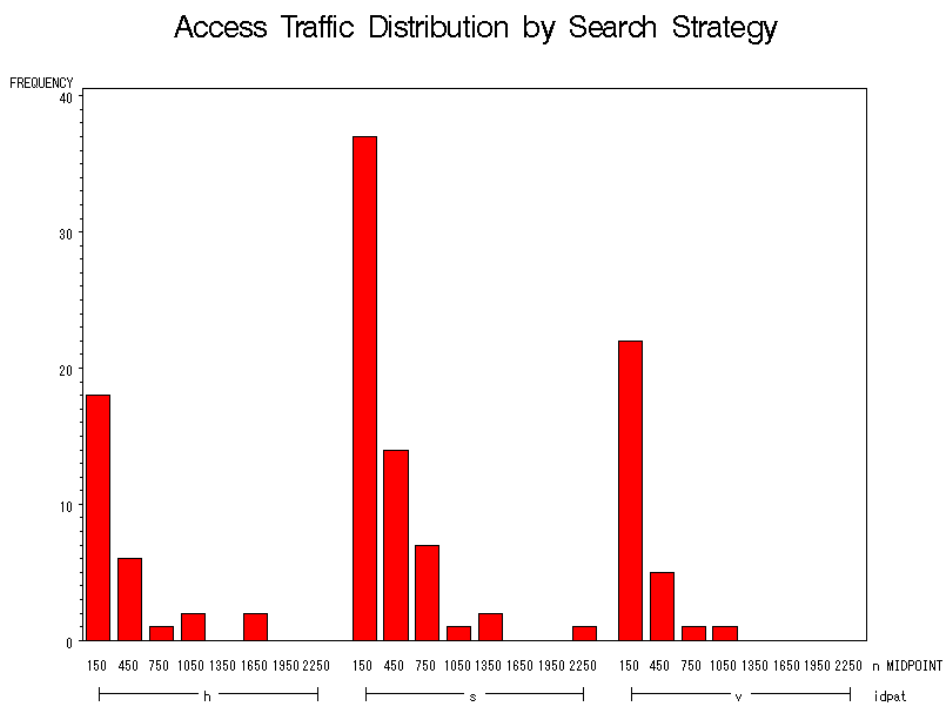


図 6-8 探索戦略によるアクセス数の違い（h=ホッパー、s=スティッカー、v=バラエティシーカー）

(補足)探索行動タイプによる Web サイト分類

探索ヒューリスティックを反映した行動タイプの応用例として、これをもちいて Web サイトの性格を分類できる。都市のアナロジーでは、ハイセンスな人が集まる街やおばあちゃんの前宿などという表現がこの考え方に相当する。ここでは、試みとして Web サイトごとにどのユーザ探索行動タイプのシェアが多いかでそのサイトを特徴づけてみた。

タイプとして、もっともシェアの多いユーザ探索行動タイプを Web サイトのキャラクターと考え、ホッパー中心のサイト、スティッカー中心のサイトと呼ぶようにする。その結果、ホッパーの多いサイトはスティッカーの多いサイトに比べて平均滞在時間が短いことを確認した(図 6-9)。さらに他のサイトに比べると、明確に戦略をもっているホッパーのサイト、スティッカーのサイトの方が探索ステップ数は短く、トラフィックも多いことも確認できる。

表 6-2 ユーザ探索行動タイプによる Web サイト分類* ()内の符号については脚注参照

	Hopper Site	Sticker Site	Other Site
平均トラフィック	248(+)	277(+)	140(++)
平均滞在時間	263(+)	350(+)	395(++)
平均探索ステップ数	4.10(+)	4.83(+)	5.92(++)
サイト数	238	644	153

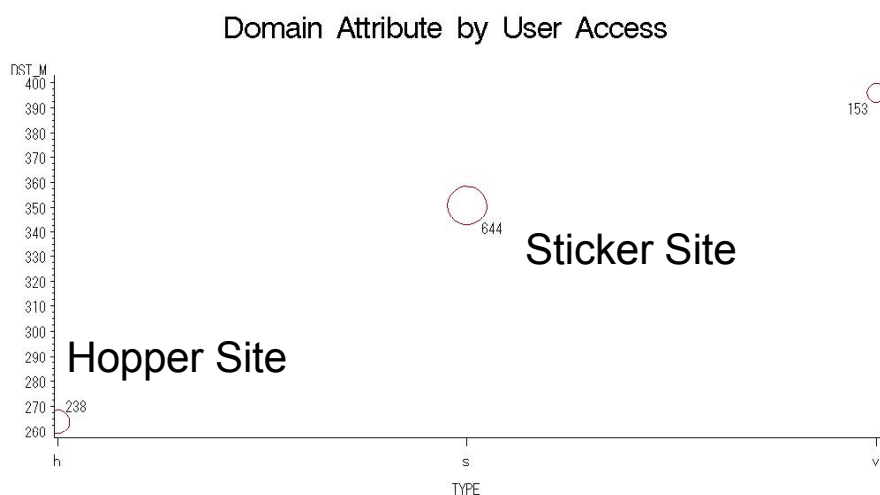


図 6-9 サイトのアクセスパターンによる滞在時間

*表中の符号+,++,+++は Tukey の多重検定により異符号間の差が 5%の危険率で有意である。サイト総数は 1035 でサイトシェアによってサイトのタイプを決めるのでトラフィックが 5 以上のサイトを対象とした。

6.2.6 この節のまとめ

本章の一つの目的となっている仮説 4 を検証するために、ドメインジャンプ表現を用いて探索ヒューリスティックを類型化する方法を提案し、幅優先探索と深さ優先の 2 つの探索パターンを基準にした 3 つのユーザ探索行動タイプを決めた。そのタイプ間でコンテンツへのアクセス数が異なるかを検証した結果、探索ヒューリスティックを基準にした行動タイプの間では違いは検出できず、仮説 4 は成立しなかった。

しかし、短い探索時間で効率よくコンテンツに遭遇し、コンテンツを長く消費する「深さ優先探索」をおこなうドメインスティッカーの採用数が支配的であることから、この探索の合理性を示している可能性が考えられる。今後の分析では探索時間と消費時間を組み合わせた合理性を検討する必要がある。

6.3 情報採餌行動モデルによる探索行動要因の検証

6.3.1 本節の課題と検証仮説

本節の課題

前節では、探索ヒューリスティックにおいてコンテンツのアクセス数だけではなく、ユーザの探索時間と消費時間を考慮する必要性を議論した。本節では、ユーザの探索行動モデルの消費時間に合理性を仮定したモデルを導入し、それを探索時間に拡張し、実際のデータ分布を適用したシミュレーションをおこなうことで、探索時間戦略の合理性について考察をおこなう。その妥当性が確認されれば、合理的な探索時間戦略により、Web 探索グラフ構造が規定されることが説明できる。

適用するモデルとして[Pirolli 95]の情報採餌行動モデルを採用し、それを拡張して探索時間の範囲を示す探索参照時間を変化させることで探索時間戦略の合理性を検討し、仮説 5 : 「Web コンテンツ探索活動全体を通して、コンテンツの探索時間と消費時間を効用関数とした単位時間当たりの獲得情報量を最大化する探索コスト戦略が存在する」ことを考察する。仮説 5 は他の仮説と異なり、ユーザの探索行動と戦略の関係性を実際のデータをもって検証するのではなく、前節に続く探索戦略の合理性を考察する

ためのシミュレーションをおこなうことが目的である。

情報採餌行動モデルとの関連性

情報採餌行動モデルでは、行動生物学の採餌行動のアナロジーから餌のパッチを Web サイトと考え、Web サイトで獲得した情報量を仮定し、捕食速度を Web サイトの探索時間・コストと処理（消費）時間で表現する。ユーザが獲得情報量の長時間平均を最大化させるような合理的な行動をとるとした場合に、限界効用定理により情報採餌行動をとる各パッチ内での捕食速度は長時間平均に一致するという特徴をもつ。このような合理的な消費（処理）時間戦略の存在自体は情報採餌行動モデルでの一つの帰結である。

情報採餌行動モデルの探索戦略はパッチ内処理時間に関するモデルであるので、本研究ではこのモデルを Web コンテンツ探索に適用することで、探索時間についても合理的な探索戦略があるのかということをも第 5・6 章で利用したコンテンツトランザクションデータから実際の探索コストの分布を用いて検討する。

Web 探索構造において、合理的な探索時間戦略を確認できれば、その行動によって Web 探索グラフ構造に影響があると考えられる。これを確認する手順としては、実際のデータから関係する分布を推定し、その推定値をモデルに挿入し、探索時間を連続的に変化させて合理性を考える。

また、実際の探索を考えると探索に合理性があるとは限らない。このモデルを適用して合理的に現在のデータの探索構造を説明できない場合には、その合理性の限界について考察を行う必要がある。

6.3.2 情報採餌行動モデル

モデルの枠組み

探索時間を探索戦略として評価するために、ユーザのアクセス行動を規定するモデルとして、数理生物学で使われる採餌行動モデル[巖佐 98]のアナロジーをつかった情報採餌行動モデルを用いる。情報採餌行動モデルはパッチ間の探索と処理のサイクルを繰り返し、そのサイクルにおける情報獲得量の長時間平均の性格を分析する[Pirolli 95]。

数理生物学の採餌モデルでは、餌が n 種あり、 i 種の餌には単位時間あたり、 i の

確率で出会うと考える。もし、 i 種の餌を食べると g_i カロリー得られるが、餌を処理し、次の餌を探索しはじめるまでに処理時間 h_i がかかる。捕食者は i 種の餌に出会ったときに、確率 p_i で食べる。捕食者は探して判断して処理してまた探してというサイクルを繰り返す。捕食速度の長時間平均（平均価値獲得率） R は、1サイクルあたりに得られるカロリーの平均を1サイクルにかかる平均時間で割った値に等しくなると考えると、長時間平均 R は次のように計算する。

$$R = r(p_1, p_2, \dots, p_n) = \frac{\sum_i \lambda_i p_i g_i / \sum_i \lambda_i}{1 / \sum_i \lambda_i + \sum_i \lambda_i p_i h_i / \sum_i \lambda_i}$$

これに対して、餌たる情報が集まっている「パッチ」というものを形成していることを考える（図 6-10）。たとえば、それは Web 空間に多数存在する Web サイトや、職場に複数配置された作業机でおこなうタスク（仕事）のようなものと考えることができる。このようなパッチモデルを考えると、そのパッチが n 種あると考え、あるパッチ i 内で得られる情報獲得量は消費した時間 t_i の関数 $G_i(t_i)$ であると考えたとさきほどの採餌モデルに照らしてみても、

$$R(t_1, t_2, \dots, t_n) = \frac{\sum_i \lambda_i G_i(t_i) / \sum_i \lambda_i}{1 / \sum_i \lambda_i + \sum_i \lambda_i t_i / \sum_i \lambda_i}$$

というように書くことができる。このときの (t_1, t_2, \dots, t_n) は、各パッチの中での餌を消費する時間の組で、これが一種の探索戦略と考えることができる。Web 空間をこのようなパッチと考えることは[Card 01]でも確かめられている。

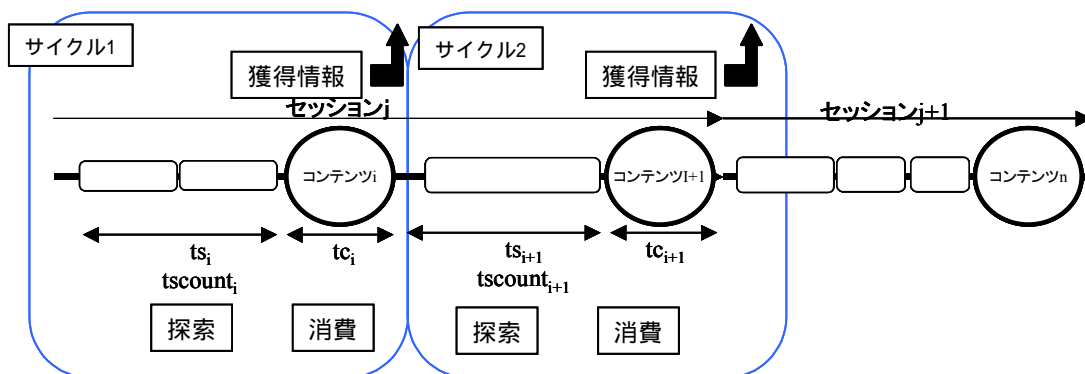


図 6-10 情報採餌行動モデルの枠組み

Web サイト探索への適用

[Pirolli 95]での展開を本研究にあわせて記述すると、Web 空間のアナロジーでは、人はそのサイト間を移動して情報を探索し、サイト内でその情報を消費すると考える。総探索時間を Tt_s 、餌を処理している総消費時間を Tt_c として、各々のサイト単位の平均をサイト間の平均探索時間を t_s 、サイト内での餌の消費時間を t_c とし、サイトに出会う平均的な割合を λ とすると、長時間での獲得情報量平均 R は

$$R = \frac{G}{Tt_s + Tt_c}$$

となる。このとき、定義より λ, G, T_c は以下のようにかけるので、

$$\begin{aligned}\lambda &= 1/t_s \\ G &= \lambda T t_s g \\ Tt_c &= \lambda T t_s t_c\end{aligned}$$

最終的に長時間平均 R は次のように書き換えられる。

$$\begin{aligned}R &= \frac{\lambda T t_s g}{Tt_s + \lambda T t_s t_c} \\ &= \frac{\lambda g}{1 + \lambda t_c}\end{aligned}$$

それぞれのサイトの集合 $\rho = \{1, 2, \dots, P\}$ での消費時間を $(t_{c1}, t_{c2}, \dots, t_{cP})$ とした場合、 R を最大化させる最適な Web サイト消費時間の配分を考える。このとき、上記の R は以下のように書ける。

$$R = \frac{\lambda_i g_i(t_{ci}) + k_i}{c_i + \lambda_i t_{ci}} \quad \text{数式- 6-1}$$

このとき、 $k_i = \sum_{j \in \rho - \{i\}} g_j(t_{cj})$ 、 $c_i = 1 + \sum_{j \in \rho - \{i\}} \lambda_j t_{cj}$ である。 R をこの t_{ci} について偏微分する

と、

$$\frac{\partial R}{\partial t_{ci}} = \frac{\lambda_i g'_i(t_{ci})[c_i + \lambda_i t_{ci}] - \lambda_i [\lambda_i g_i(t_{ci}) + k_i]}{(c_i + \lambda_i t_{ci})^2}$$

g が消費時間について収穫逓減するものと考えて、 R の最大化の条件は、 $\frac{\partial R}{\partial t_{ci}} = 0$

となるので、

$$\lambda_i g'_i(t_{ci})[c_i + \lambda_i t_{ci}] - \lambda_i [\lambda_i g_i(t_{ci}) + k_i] = 0$$

$$g'_i(t_{ci}) = \frac{\lambda_i g_i(t_{ci}) + k_i}{c_i + \lambda_i t_{ci}} = R \quad \text{数式 6-2}$$

となり、長時間平均 R を最大化するような探索では、それぞれの Web サイトでの消費時間での限界獲得率は長時間平均 R に等しいということになる。これは Charnov の限界値定理と呼ばれている。

ユーザは長時間平均 R を最大化するような合理的な行動を選択していると仮定すると、結果として各 Web サイトで消費する時間あたりの情報獲得率に等しいような消費時間配分を戦略として採用していることになる。つまり、各 Web サイトにおいては全体として最適になるようなコンテンツ消費時間が存在することを示している。

探索コストを考慮した情報獲得関数の拡張

情報探餌行動モデルの枠組みを用いて、いままで検討してきたコンテンツトランザクションにおける探索モデルに適用することを考える。

いままでの議論から、探索距離や探索時間が探索コストとなって URL 間のコンテンツのアクセス数に影響することや、SCC のグラフ構造のクラスターに依存しない一定の平均探索時間が確認されている。このことから、前述の情報獲得関数は消費時間 t_{ci} だけでなく、探索時間 t_{si} にも影響をうけていると考えることは自然である。

そこで、2変数関数 $G_i = g_i(t_{si}, t_{ci})$ として、消費時間からは $\frac{\partial R}{\partial t_{ci}} = 0$ の条件を考え

ると、前項と同様に、

$$\frac{\partial g_i(t_{si}, t_{ci})}{\partial t_{ci}} = \frac{\lambda_i g_i(t_{si}, t_{ci}) + k_i}{c_i + \lambda_i t_{ci}} = R$$

が導かれる。

探索時間については、 $\lambda_i = \frac{1}{t_{si}}$ として、

$$\begin{aligned}
R &= \frac{\lambda_i g_i(t_{si}, t_{ci}) + k_i}{c_i + \lambda_i t_{ci}} \\
&= \frac{g_i(t_{si}, t_{ci})/t_{si} + k_i}{c_i + t_{ci}/t_{si}} \\
&= \frac{g_i(t_{si}, t_{ci}) + k_i t_{si}}{c_i t_{si} + t_{ci}}
\end{aligned}$$

極値の条件を $\frac{\partial R}{\partial t_{si}} = 0$ としたとき、

$$\frac{\partial R}{\partial t_{si}} = \frac{\left(\frac{\partial g_i(t_{si}, t_{ci})}{\partial t_{ci}} + k_i \right) (c_i t_{si} + t_{ci}) - (g_i(t_{si}, t_{ci}) + k_i t_{si}) c_i}{(c_i t_{si} + t_{ci})^2} = 0$$

このとき、

$$\begin{aligned}
\frac{\partial g_i(t_{si}, t_{ci})}{\partial t_{si}} &= \frac{c_i [g_i(t_{si}, t_{ci}) + k_i t_{si}]}{c_i t_{si} + t_{ci}} - k_i \\
&= R \cdot c_i - k_i && \text{数式 6-3} \\
&= R \left(1 + \sum_{j \in \rho - \{i\}} t_{cj} / t_{sj} \right) - \sum_{j \in \rho - \{i\}} g_j(t_{sj}, t_{cj})
\end{aligned}$$

のような条件になる。

Web サイト内での消費時間の最適戦略の条件は前項のとおりであるが、探索時間については、探索時間による限界効用は、長時間平均 R から探索時間に対する消費時間の比と情報獲得量の差の積み重ねの分だけシフトする点において、 R は極値をとる。

このシフトの意味については、情報獲得関数の形状や消費時間との組み合わせを考慮しなければならず、このままでは最適な探索時間を議論するのは難しい。

そこで、いままで得られた知見から探索時間戦略の違いによる R の変化をシミュレーションする。

6.3.3 探索参照時間による数値シミュレーションと仮説5の検証

情報探餌行動モデルのコンテンツランザクションへの適用

情報探餌行動モデルをコンテンツランザクション探索行動に適用し、探索時間戦略をシミュレーションすることで、情報獲得率の長期平均 R の振る舞いを観察する。

まず、コンテンツランザクシヨンの枠組みに情報採餌行動モデルを適用する。採餌行動モデルの R の分母分子の意味を考えると、 n 種の餌のどれかに会える確率が単位時間 t_i なので、分母の第 1 項は会えるまでの平均待ち時間で、会ったものが i 種である確率は i/n なので、分子は平均情報獲得量、分母第 2 項は平均処理時間である。

本研究におけるコンテンツランザクシヨンの枠組みでは、各個人の個別の離散的な Web サイトの選択ではなく、各個人が集まった全体のマクロな個人の挙動を記述する意味で、それぞれ探索時間や消費時間に対して連続な形を考えて、長時間平均 R を記述すると、

$$R = \frac{\iint U(ts, tc) f_{ts}(ts) f_{tc}(tc) dt ds}{\int ts f_{ts}(ts) dt ds + \int tc f_{tc}(tc) dt ds} \quad \text{数式 6-4}$$

となる。

ここで、 u は探索時間とコンテンツ時間による効用関数、 f_{ts} は探索時間 ts の密度関数、 f_{tc} はコンテンツ消費時間 tc の密度関数と考える。

シミュレーションにおいて、今回のアクセスデータを分布として反映させると、第 5 章のデータ概要で述べたように、コンテンツ消費時間とコンテンツ探索時間の分布関数は、図 5-6、図 5-8 のデータから指数分布を前提にパラメータを近似することで、

$$f_{tc}(tc) = \beta e^{-\beta(tc-60)} = 0.03e^{-0.03(tc-60)}$$

$$f_{ts}(ts) = \gamma e^{-\gamma ts} = 0.27e^{-0.27ts}$$

をえることができる。

効用関数の考え方

シミュレーションのために探索行動における効用関数を定義する必要がある。効用関数はさまざまな形が考えられるが、前節の仮説 4 の検証結果を踏まえて、コンテンツ消費時間とコンテンツ探索時間に依存する $U(tc, ts)$ と考える。集計データで、コンテンツ消費時間とコンテンツ探索時間の関係性を確認すると、図 5-10、図 5-11 のように一見しては大きな関係性を見出せない。データとしては、コンテンツ探索時間のほかにコンテンツ探索ステップ長もあるが、これは図 5-9 のようにコンテンツランザクシヨンの単

位ではほぼ比例関係にあるので、ここではコンテンツ探索時間に代表させることができると考える。

以上のことから、効用関数を簡単のために、コンテンツ消費時間、コンテンツ探索時間を含む以下の関数を定義する。

$$G = U(ts, tc) = u tc^{\varpi} - \alpha ts^{\tau}$$

ここでは、 $u > 0, \alpha > 0$ で、コンテンツ消費時間が長いものはよりそこから得ている価値は高く、コンテンツ探索時間が長くなると価値が減少していくように考えることにする。この考え方は、前節のユーザ探索ヒューリスティック戦略において、比較的探索時間が短く消費時間が長い「深さ優先探索」が支配的であったということにも矛盾しない。

シミュレーション

検証すべき探索時間戦略はこの数式のなかで、コンテンツを探索する時間の最大値を示すコンテンツ探索参照時間 T_s で表現する。あわせて、コンテンツを消費時間の最大値をコンテンツ消費参照時間 T_c とする。このとき情報獲得率の長時間平均 R は、

$$R = \frac{\int_0^{T_c} \int_0^{T_s} U(tc, ts) f_c(tc) f_s(ts) dt ds}{\int_0^{T_s} f_s(ts) dt ds + \int_0^{T_c} f_c(tc) dt ds}$$

$$= \frac{\int_0^{T_c} \int_0^{T_s} (u tc^{\varpi} - \alpha ts^{\tau}) \beta e^{-\beta(tc-60)} \gamma e^{-\gamma ts} dt ds}{\int_0^{T_s} \gamma e^{-\gamma ts} dt ds + \int_0^{T_c} \beta e^{-\beta(tc-60)} dt ds}$$

のように書くことができる。本章の定義では、このときに、コンテンツの消費時間については、あらかじめ計算時点で範囲 $[60 < t_c < 1800]$ がきまっているので、このパラメータ T_c を操作できない。探索時間については、ここまでは探索してもいいとおもえる探索参照時間 T_s を変化させることで探索時間と R の関係を考察することができる。そこで、分布関数の推定値を挿入して、 T_s について R を考えると、ケース 1 ($u = \alpha = \varpi = \tau = 1$) のときは、

$$\begin{aligned}
R &= \frac{\int_0^{T_c} \int_0^{T_s} (u tc - \alpha ts) \beta e^{-\beta(tc-60)} \gamma e^{-\gamma ts} dt ds}{\int_0^{T_s} \gamma e^{-\gamma ts} dt ds + \int_0^{T_c} \beta e^{-\beta(tc-60)} dt} \\
&= \frac{\int_0^{1800} \int_0^{T_s} (tc - ts) 0.03 e^{-0.03(tc-60)} 0.27 e^{-0.27ts} dt ds}{\int_0^{T_s} 0.27 e^{-0.27ts} dt ds + \int_0^{1800} 0.03 e^{-0.03(tc-60)} dt} \\
&\cong 0.92 + \frac{0.02T_s - 0.89}{-0.01T_s + e^{0.27T_s} - 0.04}
\end{aligned}$$

となる。

同様にケース 2 ($u = \alpha = \tau = 1, \varpi = 0.9$)、ケース 3 ($u = 1, \alpha = 2, \varpi = 0.85, \tau = 1.8$) の 3 つのケースにおいて、探索参照時間 T_s を変化させて図 6-11 にプロットした。グラフからも明らかなように、いずれのケースでもある参照時間 T_s にかから R の増加率は逓減し、頭打ちになり一定になってくる。ケース 3 は R の増加率は逓減からマイナスに転じ、あるポイントで一定になる。

このことから、仮説 5 の情報獲得量の長時間平均 R を最大にする探索参照時間が存在することがシミュレーションから明らかになったといえる。別の見方をすればこの探索参照時間を超えて探索したとしても、これ以上、長時間平均 R は改善されないことになる。

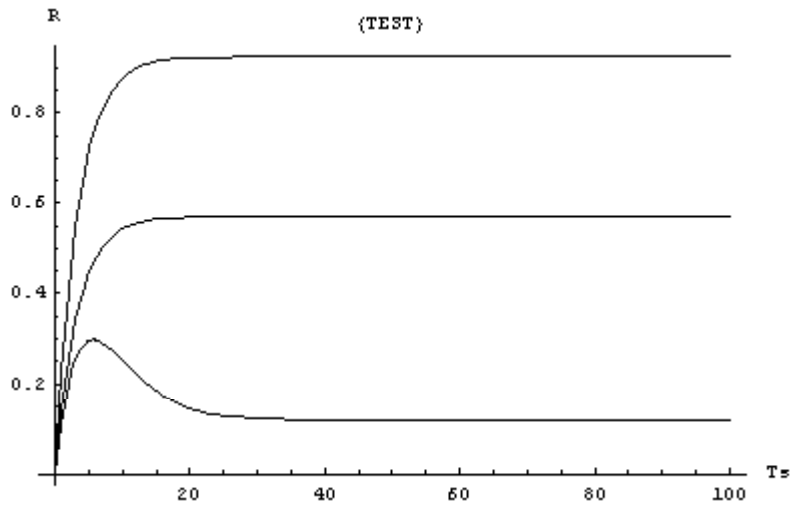


図 6-11 探索参照時間による長時間平均情報獲得率（上からケース1、ケース2、ケース3）

6.3.4 モデルとシミュレーションを通じた考察

本節では、情報採餌行動モデルをコンテンツランザクションにおける Web 探索行動モデルに適用し、探索時間戦略として、探索参照時間 T_s を変化させたシミュレーションを行うことで、探索時間と消費時間を考慮した情報獲得率 R の振る舞いを観察した。モデルの効用関数にコンテンツ探索時間を含めて拡張し、シミュレーションした結果、ある探索参照時間を超えて探索しても R が大きくなることが期待できないケースが明らかにされてきた。

仮説 5：「Web コンテンツ探索活動全体を通して、コンテンツの探索時間と消費時間を効用関数とした単位時間当たりの獲得情報量を最大化する探索コスト戦略が存在する」においては、ユーザが探索活動をおこなう場合に、コンテンツの探索時間と消費時間の両方を考慮にいたった効用関数の全探索活動を通じた情報獲得率を最大化するという合理性が前提にされてきた。この合理性にもとづいてユーザが探索活動をおこなっているとすると、シミュレーションで明らかになったように情報獲得率 R を最大化する探索時間（コスト）があることがわかる。探索活動をおこなっているユーザはその探索コストを知っているのだろうか。

ユーザが何度か Web サーフィンを繰り返すうちにこれ以上情報を探索しても情報獲得率のあまり上がらなくなり、探索を諦める探索時間を経験的に定めていることも考えられる。しかし、その探索時間が全ユーザに共通であるとは限らない。ユーザはその探

索時間を直接知っているのではなく、Web サイトや情報の配置といった環境がそれを決めている可能性もある。たとえば、コンテンツの配置や Web リンク的环境によってそれ以上探索することを諦めざるをえないポイントも存在することもある。この探索参照時間には、実際のユーザが共通にもっている時間というよりも、全ユーザや Web を一つの環境と考えた場合に、合理的な探索活動においてはこの探索参照時間が全体のふるまいを規定している可能性が考えられる。

第 5 章において、Web 探索グラフ構造ではどのクラスターにおいても平均探索ステップ数が 3 から 5 程度の小さな探索範囲になっており、Web 探索グラフ構造がスモールワールドを形成していた。また、第 6 章の前節においては、比較的短い探索時間をもちある程度長く消費するドメインホッパーという探索ヒューリスティックが支配的であった。このことは、シミュレーションで明らかになった探索参照時間という探索活動を通じた情報獲得率を最大化する探索コストが存在することを Web 探索グラフ構造の側面とユーザ探索ヒューリスティックの側面から表現されたものとも推測できる。全体の因果関係は明らかになっていないが、これは Web 探索構造の特徴を形成している一因と考えられる。

今回のシミュレーションでは探索参照時間が 20 から 30 秒前後で R は飽和していく。ケース 3 ではそれが低下し 40 秒過ぎにまた一定に近づいていく。探索参照時間を超えて探索すると分母にあたる平均探索時間が大きくなりすぎ、長時間平均が減少していく。さらに、探索参照時間を長くしても大幅に R が減少しないのは、そのような探索時間の発生確率が非常に低くなっていて、全体の平均に影響しないような指数的な分布になっているからである。今回のシミュレーションでは消費時間の範囲を一定にしているので、探索時間の大小がこれを左右する。

第 4 章で計測した探索時間の平均は 40 秒台であり、ここで行ったシミュレーションより長く探索している。シミュレーションのパラメータによって若干の変化はあるものの、理論的な R の飽和ポイントを越えて探索している。これにはいくつかの理由が考えられる。

一つは、ユーザは常に合理的だとは限らないということである。実際の探索では、どれだけ労力をさいても必ず見つけなければいけないような探索課題があることがある。そのような場合にはこのモデルの定義する合理的なポイントは適用されない。また、このモデルではユーザによる効用関数の違いを考慮にいれていないが、第 3 章のユーザ

層別モデルを待つまでもなく、その違いが影響する可能性は大きい。

このような非合理性や効用に関する個人差など、このモデルで説明できない探索行動によって実際の探索は理論的なポイントより長く探索することが考えられる。これらを明確に規定するためには、今回の分析ではあえてブラックボックスとしてきた探索メカニズムを規定しなければならない。

6.4 まとめと考察

本章の課題である仮説4：「Web コンテンツ探索行動において、ユーザは探索の幅と深さを組み合わせたコンテンツ探索ヒューリスティックの中から、探索行動中により多くのコンテンツにアクセスできる合理的なヒューリスティックを選択する」の検証では、コンテンツトランザクションの探索行動タイプ間で平均アクセス数の違いを検出するというアプローチがとられた。結果として、仮説は成立しなかった。しかし、探索パターンの採用数と平均アクセス数の高さから、比較的短い探索時間で長く消費する「深さ優先探索」をとるドメインスティッカーが支配的であることにより、アクセスしたコンテンツの数ではなく、探索時間と消費時間を考慮に入れる必要性を議論した。

さらに、第4章の課題を考慮して、合理性を考察するために、仮説5：「Web コンテンツ探索活動全体を通して、コンテンツの探索時間と消費時間を効用関数とした単位時間当たりの獲得情報量を最大化する探索コスト戦略が存在する」ことを想定し、情報採餌行動モデルをコンテンツトランザクションに適用したシミュレーションをおこなうことで、情報獲得率の長時間平均が増加しなくなるような探索参照時間が存在していることが明らかになった。

さらに、最適な探索参照時間が存在するという事実と、前章で考察した Web グラフのマクロな構造と小さな構造には関連性が考えられる。両者の分析は、同じデータを Web サイトの側からとユーザ側からの二つの側面から眺めた結果ともいえる。Web グラフの直径が比較的小さな数字であり、SCC からの距離分類による Web サイトクラスター単位で集計した平均探索ステップ数に違いがないということは、Web サイトの探索において、比較的均等に小さな努力でコンテンツにたどり着くことができるが可能であるということである。それと同時に、そのような小さな構造が探索行動のなかには形成されているとする。そのような小さな構造が形成される要因として、最適な探索参照時間

内に探索を終えようとするユーザの合理的な行動から発生していると考えることができる。

Web 探索構造からインターネットの時空間を考えると、インターネットにはコンテンツ探索の小さな構造により、距離と時間の制約に多くを縛られず、なおかつ長時間の情報獲得率を最大にするような効率的な探索が可能である。しかし、ある探索参照時間をこえて、さらに消費時間も最適値を超えて探索する場合には、小さな探索構造にはおさまらず、距離と空間は現実世界と同様に非効率性をもたらすといえる。

第7章 結論と今後の課題

7.1 結論

命題と研究課題

本研究は、第1章において『Webを代表とするインターネットは「空間と時間の壁」を克服したのか』という命題に対して、ユーザのURLへのアクセス行動を経路的な情報として扱うというアプローチをとることを述べた。

第2章において、アプローチ方法を具体化するために、ユーザの行動を規定する行動モデルとWebアクセスを分析する方法について概観し、課題を検討した。消費者情報探索行動モデルでは、探索コストをつかった分析が成果を収め、Webグラフではリンク構造を反映したマクロな空間構造が描き出されている。

インターネットアクセス手段におけるユーザの選択行動

本研究のキーワードである探索コストはアクセスサービスの対価という性格とさまざまな機会費用を含んだものと考えられる。第3章では、アクセスサービスの対価としての探索コストがどのような要素に影響されるのかということについて、消費者のインターネットアクセス手段の選択構造を調べることで検討した。いずれのモデルにおいても、ユーザによる違いはあるが、アクセス時間を反映した総費用がコストとして重要な役割を果たしている。さらに、品質情報が不足しているとおもわれるノーマルユーザで、ブランドを信頼性の代替指標として使用している様子が認められた。

探索コストを考えるうえで、時間と魅力度、信頼性の代替指標としてのブランドという概念は重要である。コンテンツを探索するとき、時間は直接探索コストに機会費用として反映し、ブランドへの依存は、魅力度への高い選好がトラフィックとして集中することを示している。

研究仮説の提出とコンテンツ探索距離概念

第4章では、これまでの知見を踏まえて、命題を検証する次の2つの仮説を設定した。

- 仮説1：「インターネットでのコンテンツ探索コストは、Web サイト間のユーザ交流トラフィックに反比例する」
- 仮説2：「Web 探索グラフにおいて、強連結されたグラフ構造に属する Web サイトには、他のクラスターより多くのアクセスが生じている」

これを検証することで、機会費用としての探索コストの意味をもつコンテンツ探索距離がインターネットでのコンテンツ探索行動における距離の壁となり、Web ページリンクの空間構造が空間の壁となっていることを検証できる。

そのような構造を形成する原因を、Web グラフの物理的な特徴から検討するために、次の仮説を提示した。

- 仮説3：「Web 探索グラフの直径は比較的小さく、少ない探索コストでコンテンツにたどり着けるスモールワールドを形成している」

さらに、ユーザの探索行動が Web グラフの構造を規定しているとして、深さ優先と幅優先という2つの探索ヒューリスティックの合理性と、探索時間戦略配分の合理性を検証するために、次の2つの仮説を提示した。

- 仮説4：「Web コンテンツ探索行動において、ユーザは探索の幅と深さを組み合わせたコンテンツ探索ヒューリスティックの中から、探索行動中により多くのコンテンツにアクセスできる合理的なヒューリスティックを選択する」
- 仮説5：「Web コンテンツ探索活動全体を通して、コンテンツの探索時間と消費時間を効用関数とした単位時間当たりの獲得情報量を最大化する探索コスト戦略が存在する」

そして、これらを検証するための基礎的な概念として、参照時間長を基準にしたコンテンツトランザクションにおける探索距離概念を定義した。この概念には、いわゆる探索行動がもつユーザの明示的な「目的」「認知」「選択」という態度は扱われず、意味の重要性を失わずにアクセスログを探索シーケンスに変換し、定量的な距離概念として定義することを重視した。

Web サイトと探索構造

第4章で提案したコンテンツトランザクションによる探索コストを算出し、第5章で

は仮説 1・2 の検証をとおして、Web サイトに関する探索グラフ構造の特徴を明らかにした。Web サイトのトラフィックを魅力度、Web サイト間の探索コストを距離項において、Web サイト間の空間相互作用を推定するグラビティモデルを構築し、サイト間トラフィックには探索コストが距離として機能することが確認され、仮説 1 を確認した。

さらに、Web グラフ構造では、算出された強連結集合 (SCC) をもとにしたクラスター間でトラフィックに違いがあることが検証され、仮説 2 を確認した。

つまり、Web サイト間リンクのグラフ構造においては、インターネットは空間と時間を克服しているのではなく、コンテンツを探索するための探索コストが距離となり、Web ページリンクを空間構造とした壁が存在していることを確認した。

しかし、一方そのサイズについては、仮説 3 を通して、Web のクラスターの直径を最短パスの平均値で測ると 10 以下の比較的小さな構造でコンテンツが連結していることがわかった。これは既存の研究でも指摘されているが、探索コストを用いたモデルでも同様に確認できた。同時に、SCC のクラスターにおけるマクロ構造においても、クラスターに共通して平均探索ステップ数が 4 程度であることを考慮にいれると、「マクロな」構造のなかには、比較的小さなコスト負担でアクセス可能な「小さな」構造があると考えられる。「小さな」構造では少ない探索コストでサイト間を移動し、マクロ構造に左右されない平等な、ポール・ヴィヴィリオのいうところの遠隔对象的な行動が実現されている。

ユーザと探索行動

第 6 章では、探索構造を検証する仮説 4 を確かめるために、ドメインジャンプ表現による探索戦略とユーザ探索行動タイプを規定したが、探索ヒューリスティックにおいてアクセスしたコンテンツの数の多さに合理性をおいた仮説 4 を検証することはできなかった。しかし、支配的な探索ヒューリスティックである「深さ優先探索」の時間配分からは、コンテンツの数ではなく、比較的短い探索時間で長くコンテンツを消費する探索時間の組み合わせが合理的であることが考えられる。

さらに、探索時間戦略の合理性を考察するために仮説 5 を踏まえて、情報採餌行動モデルによるシミュレーションをおこない、情報獲得率の長時間平均が増加しなくなるような探索参照時間が存在していることが明らかになった。ユーザが情報獲得率の長時間

平均を最大化するような合理的な行動を仮定した場合に、ある探索時間を超えて探索することは合理的でない。仮説 4 と 5 を通して、Web 探索構造からインターネットの時空間を考えると、距離と時間の制約に多くを縛られず、なおかつ長時間の情報獲得率を最大にするような効率的な探索が可能であるが、ある探索参照時間をこえて、さらに消費時間も最適値を超えて探索する場合には、距離と空間は現実世界と同様に非効率性をもたらしているといえる。

7.2 応用の方向性

Web アクセスにおける探索構造を知ることにより、SCC やトラフィック、入出次数などの Web 探索グラフの性質がわかった。Web サイトをデザインする際には、SCC に対して相互リンクを張り、主要なコンテンツは平均探索ステップ数以内の 4 ステップ内に配置することが必要である、といった知見を活用できる。

Web アクセスのグラフ構造をもとに探索距離とサイト間トラフィックのグラビティモデルを構築することで、サイトでアフィリエイトプログラムを設定した場合のトラフィックを予測することができる。トランザクションの予備調査から探索ステップ長や探索時間を計測することで、アフィリエイトプログラム設定時の単価や販売キックバックの率や金額の戦略設定や最適化をおこなうことが可能になる。

また、Web サイトに集まるユーザのアクセス行動を、ホッパー、スティッカーなどに分類することで Web サイトの性格を規定することが可能で、それから今後のサイト運営の指針が得られる。たとえば、ユーザトラフィックを増やし、滞在時間を増やしていくために、どのようなユーザを増やすべきなのかというマーケティングターゲットの設定が可能になる。

7.3 今後の課題

本研究においては、探索行動における具体的な探索メカニズムや探索目的に対する態度、環境の影響などをあえて考慮することをせず、それらを包括的に表現する探索コストに依存して展開した。そのため、探索課題や個人の探索スキルといった探索行動自体の基本的な解析は行っていない。それを行うことで、探索行動タイプや探索時間戦略を

もっと具体的に評価することができるようになると考えられる。探索コストにおいても非常に抽象的に扱ったので、探索行動自体を明確にすることでもっとリアリティの高い議論ができるようになる。

また、データについては、アクセスログデータだけをつかっているのもので、実際の Web リンク構造はもっと複雑な構造を形成しているともいえるし、実際の Web の物理的な構造自体がこのような Web アクセス環境を形成していると考えられる。分析方法としては、Web リンク情報とアクセスデータの両方をつかい、Web グラフの隣接行列にウェイトを取り込んだような表現を開発することで、より深い分析が可能になる。そのときはリンクがあるという情報とトラフィックが1であるという情報が区別できるように行列を操作すればいい。

さらに、その際には Web サイト自体の属性を考慮にいれるようにすべきである。今回の分析では、ポータルサイトやショッピングボットといった Web サイト自体がもつ機能については全く言及されていない。そういった機能を使った結果として生じたアクセスログを使うことでその状態を包含したと解釈している。そのため、Web サイト個別の機能に対する説明というより、もっと抽象的な距離構造を提供した。Web サイトの属性を考慮にいれることでもっと精度の良く、探索行動の具体的な意味をとらえた推定を行える。

また、第3章で示したサービスの対価としての探索コストにおいて、インターネットアクセス手段の選択効用にみられるようなブランドを効用として取り入れる工夫が今後の分析では必要になる。第5章の探索距離と Web サイト間のトラフィックに対するグラビティモデルにおいても、Web サイト自体のブランドを考慮することが必要である。Web サイトに集中するトラフィックに比例する魅力度としてブランド表現を取り込むことができれば、探索コストを距離項として現実世界の都市間アクセス構造に近い状態を表現できる。

第6章の情報採餌行動モデルについてもブランドを取り入れることが望ましいが、そのほかに、ユーザごとの効用関数の異質性も取り込む必要がある。そのような拡張をおこなった分析をすることで、本研究では十分説明できなかった探索時間に対する演繹的な解析が可能になる。

7.4 補論: インターネット探索行動圏の展望

本節では、いままでの成果を踏まえて、インターネット探索行動圏というコンセプトを構想する。その問題意識について述べ、現実の商圈とのアナロジーから発想されたインターネット圏の定義について述べる。そして、それを検証するため課題と展望について述べる。

7.4.1 インターネット探索行動圏のイメージ

これまでの研究により、一見フラットで距離と時間を克服し、グローバルな均一空間のようなインターネットでも、そのアクセスには特徴的な構造があり、インターネットの空間上に探索構造をもった「圏」が形成されいると考えることができる。

地理上の世界では、たとえば、ストラスブールを中心にいくつかの小都市があつまって形成するフランスのアルザス地方などといった呼び方をする「地域圏」というものが存在する。また、地理的な結びつきの強い「地方」という言い方に対して、文化交流や経済交流を基盤にした結びつきを強調した「文化圏」や「経済圏」や、商業を中心に「商圈」という言い方がある。「圏」とはこれらを包含する広い概念として、ある地域がなんらかの関連性により結びついて広がり形成しているものといえる[大友 97]。

このように、現実世界の圏は、「人口」、「地理的な距離」、「交通」、「経済・文化」といったいろいろな要素が交じり合い、影響しあって形成されると考えられる。

いままでの研究の成果から、特に Web サイト間の関係には以下のようなことがいえる。

- コンテントランザクションを利用した探索距離は、グラビティモデルにおいて現実世界での距離概念とおなじように機能する。
- Web アクセスの探索グラフ構造は、強連結集合 (SCC) を中心にしたクラスター構造を形成する。
- Web アクセスの探索グラフ構造がもつ直径は比較的小さく、比較的簡単にコンテンツに到達しているケースがある。

インターネットでもこれらの成果と現実世界の「圏」のアナロジーをつかって、「インターネット探索行動圏」なるものが形成されていると考えることにする。そこでは、物理的なドメイン間のリンク構造やサイトのアクセス帯域、ドメインの店舗的な特徴な

どにより、インターネットアクセスに構造的な特徴が生じたり、また、その特徴により、ドメインが結び付けられたり、ユーザのアクセス行動に一定の特徴がうまれたり、といったことが起こっていると考え、図 7-1に示すようなイメージを想定する。そのような構造的な関係性を総称して、「インターネット探索行動圏」と呼ぶこととする。

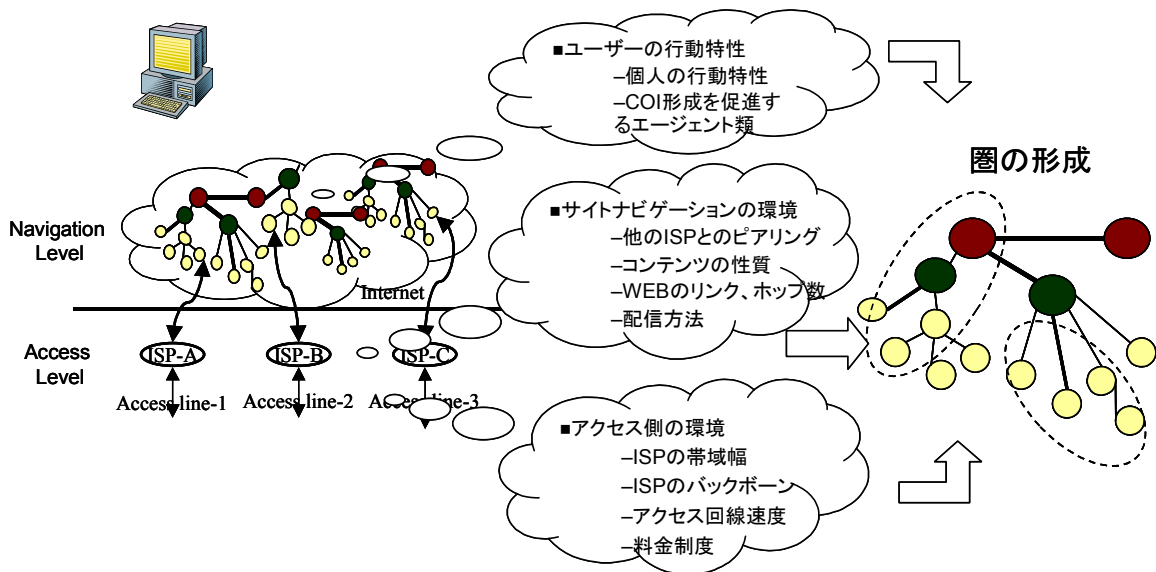


図 7-1 インターネット探索行動圏のイメージ

7.4.2 インターネット探索行動圏へのアプローチ

2つの方向性からの定義

商圈の定義として「顧客や生活者が店舗ないし地域にあつまる来店範囲、店舗ないし地域が顧客を吸引する力が及ぶ範囲の商業空間」[佐藤 98]といわれることがある。この考え方には2つのアプローチが含まれている。ひとつは、前半部の「顧客」の行動特性に着目したアプローチであり、2つ目は「店舗」に着目したアプローチである。従来、この考え方をモデル化するには、「顧客」の選択行動を選択肢の効用比で記述する「確率的選択モデル」や、「店舗」や「地域」間の「顧客」の空間上の交流状況をモデル化した「空間相互作用モデル」が用いられてきた。

そこで、インターネット上で圏を定義するために、前節での商圈定義に対する二つのアプローチのアナロジーを適用して、

- インターネット上で顧客や生活者がアクセスする Web サイトの属性空間
- ある Web サイトがインターネット上で吸引する顧客の属性空間

という2つのアプローチを考えることにする。ここでいう Web サイトの属性空間とは、Web サイトの運営目的や所属といった社会的なものからアクセス数などのネット視聴率に相当するようなアクセス状況、そのサイトに設置されるコンテンツの種類やサーバー数、セキュリティ、バックボーン回線帯域といったテクニカルな要素まで幅広いものが想定される。また、顧客の属性空間とは、集まる顧客の年齢・性別・職業などの社会的な属性からネット上での商品購入履歴・嗜好などいわゆるマーケティングデータといわれるような属性まで想定され、こちらも幅広い属性が考えられる。

アクセスログデータをもとに分析をすることを考えると、顧客ユーザを同定しながら、ユーザの探索行動をトレースするためには、ユーザ個人単位でデータを収集・分析する必要がある。前述の2つのアプローチはそれをユーザ属性で集計するのか、Web サイトの属性で集計するかの違いである。

また、インターネット上での購買行動だけに着目するのではなく、広く探索・情報取得行動（ナビゲーション行動）全般を対象にする。

その意味で、インターネット探索行動圏の概念は、「インターネット上のユーザ個別の探索（情報取得）行動を分析し、ユーザ属性またはサイト属性で集計した距離と構造をもつ情報空間」であるということができる。これを検討するためには、本論でも課題になっていた情報取得行動を記述する具体的なモデルが必要であることは論をまたない。

通常の商圈概念との違い

インターネット探索行動圏には、通常の商圈概念とはアナロジーを持ちながらも、いくつかの重要な違いがある。表 7-1にその概要を整理したが、たとえば、方向や位置に関する情報の扱い方や認知過程が両者では大きく異なる[野島 00]ことは非常に重要である。特に、通常の商圈では、位置や方向が地図上で明確に定義し、移動距離を計測できるのに対して、WWW 上では IP アドレスという準固定されたアドレスはあっても、WWW ではすべて HTML 上のノードとしての URL とそれをつなぐリンクで構成されているため、位置や方向が相対化されてしまうので、絶対的な位置を把握することは難しい。

インターネット探索行動圏の距離概念として、探索コストが想定されるが、本研究で

距離概念としての機能を確認した「コンテンツ探索距離」を導入するだけでなく、さらに、探索サービスへの対価としての探索コストやブランド特性も検討されるべきである。

表 7-1 商圏モデルとインターネット探索行動圏モデルの違い

	従来の商圏モデル	インターネット探索行動圏モデル
コンテンツ特性	<ul style="list-style-type: none"> • 基本的に有形小売品、対面サービスを扱う • 商品の複製にコストがかかる 	<ul style="list-style-type: none"> • 有形小売品の他、無形デジタルコンテンツの放送・配信まで情報流通全般 • 商品の複製コストは低くなる
アクセス方法と利用情報	<ul style="list-style-type: none"> • 交通（徒歩、電車、車など）手段。 • それを補助する広告ちらし、販促活動の影響がある • 複数箇所、複数商品の情報を入手するにはコストが非常にかかる 	<ul style="list-style-type: none"> • リンク、バナー広告、サーチエンジン、ショップボット • レコメンドサービスなどさまざまなエージェントが存在する • 複数箇所、複数商品の情報を入手するコストは非常に低く、扱う情報の種類も多い
距離の定義	<ul style="list-style-type: none"> • 実際の空間配置に近い形の距離 	<ul style="list-style-type: none"> • コンテンツ探索距離+サービス対価
顧客のデモグラフィックデータ	<ul style="list-style-type: none"> • メッシュ法などを利用して、空間的・量的に定義可能 	<ul style="list-style-type: none"> • 量と位置が不明
調査方法	<ul style="list-style-type: none"> • アンケートなど対象地域に対するランダムサンプリングが可能で、それをを用いた統計モデルでの展開が可能 	<ul style="list-style-type: none"> • 統計的代表性に問題がある • ユーザ行動測定・トラッキングシステムの構築が可能

7.4.3 今後の研究課題

現在はインターネット探索行動圏というイメージを述べたにすぎない。これを実現していくためには、前節の研究課題でも述べたような、アクセスログデータと Web リンクデータの統合的な分析、Web サイトの個別の機能とブランド力の表現などの方法論が開発されていなければならない。機会費用としてのコンテンツ探索距離だけでなく、サービスコストとしての探索コストを統合していかなければならない。

そうしていくことで、空間立地理論や複雑系経済学の流れを融合した空間経済学と連携してさまざまな展開を考えることができる。特に、サイトの立地問題を扱うことは今後の情報経済を占う面で非常に重要である。

また、実際の生活基盤である「都市」のもつ情報と接続することで、バーチャルな空間に展開しているインターネット探索行動圏と実際の都市情報を複合した「情報都市」[若林 02]の構想を実現できるものになる可能性もある。

このように構築されたインターネット探索行動圏では、消費者情報探索行動モデル [Smith 99] が扱う価格分散のしくみをロケーションやブランドロイヤリティを探索行動に組み込んで説明できるようになると考えられる。

謝 辞

本論文の作成過程はインターネットの探索構造を手探りで「探索」するようなものだった。多くの方々のご指導とご支援なくしては到底ここまではたどり着くことはできなかった。この場を借りて心より感謝申し上げたい。

そもそも最初から探索構造に関心があったわけではない。当初は、インターネットの世界と現実世界はどのような社会的な繋がりをもっているのかということを対象に研究をはじめた。しかし、インターフェースを議論するためにはそれぞれの社会自体の構造を定義する必要があるということに思い至り、まずはインターネットの世界をどうやって測るのかということを考えはじめた。インターネットはさまざまなハードウェア、ソフトウェア、ネットワーク、そしてそれを扱う「人」が複雑に絡み合った集合体である。その世界の捉え方は見る立場によって異なる。

専門的な学問領域をもたないことを学際的と称していた私には確固たる立場と呼べるものがなく、研究を進めるためには自分の問題意識に関係するものを片っ端から調べることで問題解決に関係する分野と分析の道具を絞り込んでいくしかなかった。それこそ、インターネット～とつくものはなんでも目を通すことになってしまった。これは羅針盤をもたないでインターネット研究の広大な海原に船出してしまったようなものである。

修士課程の経営システム科学専攻時代からご指導をいただき、いまは帝京大学でご活躍の橋田先生には、この広大な研究の沃野のなかで方法論を暗中模索しているときに、ネットワークに関する鋭い考察と幅広い視野から本当に貴重なアドバイスをいただいた。橋田先生はご自分の専門の枠に拘泥することなく、問題意識に応じて数理学から計算機科学・社会学・経済学・哲学の分野に至るまで様々な道具を軽々と使いこなされていた。学際的であることを習うのではなく、自ら実践される先生の姿勢に感服するとともに、自分の視野の狭さを思い知る機会を得たことは、いかなる感謝のことばをもっても尽くせるものではない。

橋田先生退官後に引き続いてご指導いただいた寺野先生には、既存の枠組みにはまり

きらないで路頭に迷いかけたこの研究をまとめ上げ、研究成果を意味付けるために多くのご示唆をいただいた。心より感謝申し上げたい。寺野先生の複雑さに関する知見にもとづいたアドバイスなくしては、複雑なネットワーク研究に通じる本研究の意味すらわからなかったかもしれない。

また、修士課程から長年お世話になったG S S Mの諸先生方に感謝申し上げたい。とりわけ副指導教官である久野先生と西尾先生には親身なご指導をいただいたことに感謝申し上げたい。多少なりともインターネットを理解し、そこで展開するマーケティングについて議論できるのも両先生のご指導に負うところが大きい。

本論文を査読していただいた先生方には、論文を書くという作業の意味を実践で教えていただいた。的確な指摘をなんとかこなしていくことで論文としてまとめ上げることができた。稚拙な論考と文章に我慢強くお付き合いいただいたことに感謝申し上げたい。橋田ゼミ、修士・博士入学同期の方々とのなにげない会話や議論も大いに刺激的であり、G S S Mという創発の場の魅力は何ごとにも代えがたいものであった。皆様に心より感謝申し上げたい。

(株)N T Tデータ開発本部システム科学研究所の勝川高志副所長、恒松直幸部長、廣岡康雄氏には貴重なデータを提供していただいた。このデータなくしては研究に手をつけることすらできなかった。なかなか成果を出せず大変ご迷惑をおかけしたが、温かく見守っていただいたことに心より感謝申し上げたい。

大学時代や社会人となってからも私の活動を理解してくれた友人や職場の方々にも感謝申し上げたい。社会人と学生の二足のわらじをはきながらも、専門依存型ではない学際的な「入力問題主導型」の問題解決が有効であることを実践を通して体得できたのは皆様のおかげである。

また、研究内容がわからずとも、研究の進捗や私の健康にいろいろと心を配ってくれた両親の暖かい眼差しにも感謝したい。

そして、この論文をまとめ上げる時期にいっしょに人生を歩み始めることを決断し、論文作成を何度も諦めかけた本当に大変な時期に私を支えてくれた妻恭子に感謝したい。

2003年

下方 拓

参考文献

- 1 [Adamic 99] Adamic, L. A., and Huberman, B. A.: The Nature of Markets in the World Wide Web, *Quarterly Journal of Electronic Commerce*, Vol.1, pp. 5-12, 2000.
- 2 [Agrawal 95] Agrawal, R., and Srikant, R.: Mining Sequential Patterns, *Proc. of the 11th International Conference on Data Engineering*, pp.3-14, 1995.
- 3 [Alba 97] Alba, J., Lynch, J., Weitz, B., Janiszewski, C., Lutz, R., Sawyer, A., and Wood, S.: Interactive Home Shopping: Consumer, Retailer, and Manufacturer Incentive to Participate in Electronic Marketplaces, *Journal of Marketing*, Vol.61, pp.38-53, 1997.
- 4 [Albert 99] Albert, R., Jeong, H., and Barabasi, A. L.: Diameter of the World-Wide Web, *Nature*, Vol.401, pp.130-131, 1999.
- 5 [Axelrod 97] Axelrod, R.: *The Complexity of Cooperation: Agent-Based Models of Competition and Cooperation*, Princeton University Press, 1997.
- 6 [Bailey 97a] Bailey, J., and Bakos, Y.: An exploratory study of the emerging role of electronic intermediaries, *International Journal of Electronic Commerce*, Vol.1, No.3, pp.7-22, 1997.
- 7 [Bailey 97b] Bailey, J. P.: The economics of Internet interconnection agreements, *Internet Economics*, pp.155-168, The MIT Press, 1997.
- 8 [Bakos 97] Bakos, J. Y.: Reducing Buyer Search Costs: Implications for Electronic Marketplaces, *Management Science*, Vol.43, No.12, pp.1676-1692, 1997.
- 9 [Borges 99] Borges, J., and Levene, M.: Data mining of user navigation patterns, *Web Usage Analysis and User Profiling*, Vol. 1836, pp.92-111, Springer-Verlag, 1999.
- 10 [Bradlow 00] Bradlow, E. T., and Schmittlein, D. C.: The Little Engines That Could: Modeling the Performance of World Wide Search Engines, *Marketing Science*, Vol.19, No.1, pp.43-62, 2000.
- 11 [Bray 96] Bray, Tim.: Measuring the Web, *Proc. of 5th International World Wide Web Conference*, pp.994-1005, 1996.
http://www5conf.inria.fr/fich_html/papers/P9/Overview.html

- 12 [Broder et al. 00] Broder, A. Kumar, R. Maghoul, F. Raghavan, P. Rajagopalan, S. Stata, R. Tomkins, A. Wiener, J.: Graph structure in the web, *Proc. of 9th International World Wide Web Conference*, pp. 309--320, 2000.
<http://www9.org/w9cdrom/160/160.html>
- 13 [Brynjolfsson 00a] Brynjolfsson, E., and Smith, M. D.: The Great Equalizer? Consumer Choice Behavior at Internet Shopbots, MIT working paper, 2000.
<http://ebusiness.mit.edu/erik/>.
- 14 [Brynjolfsson 00b] Brynjolfsson, E., and Smith, M. D.: Frictionless Commerce? A comparison of Internet and Conventional Retailers, *Management Science*, Vol.46, No.4, pp.563-585, 2000.
- 15 [Card 01] Card, S. K., Pirolli, P., Van Der Wege, M., Morrison, J. B., Reeder, R. W., Schraedley, P. K., and Boshart, J. :Information Scent as a Driver of Web Behavior Graphs: Results of a Protocol Analysis Method for Web Usability, Palo Alto Research Center User Interface Research, 2001.
<http://www2.parc.com/istl/projects/uir/pubs/items/UIR-2000-13-Card-CHI2001-WWWProtocols.pdf>.
- 16 [Catledge 95] Catledge, L. D., and Pitkow, J. E.: Characterizing Browsing Strategies in the World-Wide Web, *Computer Networks and ISDN Systems*, Vol.27, No.6, pp.1065--1073 1995.
- 17 [Cawley 97] Cawley, R. A.: Interconnection, pricing, and settlements: Some healthy jostling in the growth of the Internet, *Coordinating the Internet*, pp.346-376, The MIT Press, 1997.
- 18 [Chatterjee 98] Chatterjee, P., Hoffman, D. L., and Novak, T. P.: Modeling the Clickstream: Implications for Web-Based Advertising Efforts, elab working paper, 1998.
<http://elab.vanderbilt.edu/research/papers/html/manuscripts/clickstream/clickstream.html>
- 19 [Chen 96] Chen, M.-S., Park J. S., and Yu P.S.:Data Mining for Path Traversal Patterns in a Web Environment, *Proc. of the 16th International Conference on Distributed Computing Systems*, pp.385-392, 1996.
- 20 [Cooley 99] Cooley, R. Mobasher, B., and Srivastava, J.: Data Preparation for Mining World Wide Web Browsing Patterns, *Journal of Knowledge and Information Systems*, Vol. 1, No. 1, pp.5-32, 1999.
<http://maya.cs.depaul.edu/~mobasher/pubs-subject.html>.
- 21 [Crawford 97] Crawford, D. W.: Internet services: A market for bandwidth or communication? , *Internet Economics*, pp.379-400, The MIT Press, 1997.
- 22 [Daft 86] Daft, R.L., and Lengel, R.H.: Organizational Information Requirements, Media Richness and Structural Design, *Management Science*, Vol.32, No.5, pp.554-571, 1986.

- 23 [Davidsen 02] Davidsen, J., Ebel, H., and Bornholdt, S.: Emergence of a Small World from Local Interactions: Modeling Acquaintance Networks, *Physical Review Letters*, Vol.88, No.12, pp.128701-1-128701-4, 2002.
- 24 [Dewan 98] Dewan, R. Freimer, M. L., and Seidmann, A.: Internet Service Providers, Proprietary Content, and the Battle for User's Dollars, *Communications of the ACM*, Vol.41, pp.43-48, 1998.
- 25 [Dodge 01] Dodge, M., and Kitchin, R.: *mapping cyberspace*, Routledge, 2001.
- 26 [Economides 96] Economides, N.: The Economics of networks, *International Journal of Industrial Organization*, Vol.14, No.6, pp.673-699, 1996.
- 27 [Evans 00] Evans, P., and Wurster, T. S.: Blown to Bits (邦題：ネット資本主義の企業戦略), ダイヤモンド社, 2000.
- 28 [Faloutsos 99] Faloutsos, M., Faloutsos, P., and Faloutsos, C.: On Power-Law Relationships of the Internet Topology, *Proc. of ACM SIGCOMM*, pp.251-262, 1999.
- 29 [Fujita 99] 藤田昌久, Krugman, P., and Venables, A. J.: *The Spatial Economy: Cities, Regions and Intenational Trade*, MIT Press, 1999. (邦題：空間経済学-都市・地域・国際貿易の新しい分析, 東洋経済新報社, 2000.)
- 30 [GVU 97] Graphic, Visualization, & Usability Center's (GVU) 8th WWW User Survey.
http://www.gvu.gatech.edu/user_surveys/survey-1997-10/
- 31 [Gaylord 98] Gaylord, R.J., and D'Andria, L.J.: *Simulating Society: A Mathematica Toolkit for Modeling Socioeconomic Behavior*, TELOS Springer-Verlag New York, 1998.
- 32 [Gillet 95] Gillet, S.E.: Connecting Homes to the Internet: An Engineering Cost Model of Cable vs. ISDN, *MIT Laboratory for Computer Science Technical Report 654*, 1995.
<http://tns-www.lcs.mit.edu/publications/mitlestr654.html>
- 33 [Girardin 95] Girardin, L.: *Cyberspace Geography Visualization*, 1995.
<http://www.girardin.org/luc//cgv/report/>
- 34 [Gomory et al 99] Gomory, S., Hoch, R., Lee, J., Podlaseck, M., and Schonberg, E.: E-Commerce Intelligence: Measuring, Analyzing, and Reporting on Merchandising Effectiveness of Online Stores, IBM T.J Watson Research Center, *IBM Institute of Advanced Commerce Technical Papers*, 1999.
<http://researchweb.watson.ibm.com/iac/papers/eabs3.pdf>
- 35 [HERMES 95] HERMES : Consumer Survey of WWW Users, 1995.
<http://www-personal.umich.edu/~sgupta/hermes/>
- 36 [Haltiwanger 00] Haltiwanger, J., and Jarmin, R.S.: Measuring the Digital Economy, *Understanding Digital Economy*, pp.13-33, The MIT Press, 2000.
- 37 [Herzog 97] Herzog, S., Shenker, S., and Estrin, D.: Sharing Multicast Costs, *Internet Economics*, pp.169-212, The MIT Press, 1997.

- 38 [Hochheiser 99] Hochneiser, H., and Shneiderman, B.: Understanding Patterns of User Visit to Web Sites: Interactive Starfield Visualizations of WWW Log Data, *Technical Reports 99-3 Institute for Systems Research*, University of Maryland, 1999.
- 39 [Hoffman 96a] Hoffma, D. L., and Novak, T. P.: Marketing in Hypermedia Computer-Mediated Environments: Conceptual Foundations, *Journal of Marketing*, Vol.60, pp.50-66, 1996.
- 40 [Hoffman 96b] Hoffman, D. L., Kalsbeek, W. D., and Novak, T. P.: Internet and Web Use in the United States: Baselines for Commercial Development, eLab working paper, 1996.
<http://elab.vanderbilt.edu/research/papers/html/manuscripts/baseline/internet.demos.july9.1996.html>
- 41 [Hoffman 97] Hoffman, D. L., and Novak, T. P.: Diversity on the Internet: The Relationship to Access and Usage, eLab working paper, 1997.
<http://elab.vanderbilt.edu/research/papers/html/manuscripts/aspen/diversity.on.the.internet.oct24.1997.html>
- 42 [Hotelling 29] Hotelling, H.: Stability in Competition, *The Economic Journal*, Vol.39, pp.41-57, 1929.
- 43 [Huberman 98] Huberman, B., Pirolli, P., and Pitkow, J.: Strong Regularities in World Wide Web Surfing, *Science*, Vol.280, pp.95-97, 1998.
- 44 [Kannan 98] Kannan, P.K, Chang, Ai-Mei, and Whinston, A. B.: Marketing Information on the I-way, Data junkyard or information gold mine? *Communication of the ACM*, Vol.41, pp.35-49, 1998.
- 45 [Kephart 99] Kephart, J. O., and Greenwald, A. R.: Shopbot Economics, *Proc. of the 3rd conference on Autonomous Agents*, pp.378-379, 1999.
<http://www.ibm.com/iac/>
- 46 [Kleinberg 98] Kleinberg, L.: Authoritative sources in a hyper-linked environment, *Proc. of the 9th ACM SIAM Symposium on Discrete Algorithms*, pp. 668-677, 1998.
- 47 [Krugman 96] Krugman, P. : *The Self-Organizing Economy* (邦題 : 自己組織化の経済学), 東洋経済新報社, 1996.
- 48 [Liu 98] Liu, B. Hsu, and W. Ma, Y.: Integrating Classification and Association Rule Mining, *Proc. of ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD-98)*, pp. 80-86, 1998.
- 49 [Lohse 98] Lohse, G. L., and Spiller, P.: Quantifying the Effect of User Interface Design Features on Cyberstore Traffic and Sales, *Proc. of the SIGCHI conference on Human factors in computing systems*, pp.211-218, 1998.
- 50 [Lynch 00] Lynch, J. G., and Airely, D.: Wine Online: Search Costs Affect Competition on Price, Quality, and Distribution, *Marketing Science*, Vol.19, No.1, pp.83-103, 2000.

- 51 [Maurer 00] Maurer, S. M., Huberman, B. A.: Competitive Dynamics of Web Sites, working paper, Xerox Palo Alto Research Center, 2000.
- 52 [McKnight 97] McKnight, L.W., and Bailey, J.P. (Eds): *Internet Economics*, The MIT Press, 1997
- 53 [Morgan 96] Morgan, R. F.: An INTERNET Marketing Framework for the World Wide Web, *Marketing Management*, Vol.12, pp.757-775, 1996.
- 54 [Netratings] Nielsen//NetRatings : <http://www.netratings.co.jp>
- 55 [Novak 97] Novak, T., and Hoffman, D.: New Metrics for New Media: Toward the Development of Web Measurement Standards, *World Wide Web Journal*, Vol.2, No.1, pp. 213-246, 1997.
- 56 [Novak 00] Novak, T. P., Hoffman, D. L., and Yung, Yui-Fai. : Measuring the Customer Experience in Online Environments: A Structural Modeling Approach, *Marketing Science*, Vol.19, No.1, pp.22-42, 2000.
- 57 [Palmer 98] Palmer, J. W., and Griffith, D. A.: An Emerging Model of Web Site Design for Marketing, *Communication of ACM*, Vol.41, No.3, pp.45-51, 1998.
- 58 [Peppers 97] Peppers, D., and Rogers, M.: *Enterprise one to one "Tools for Competing in the Interactive Age"* (邦題 : one to one 企業戦略) , ダイヤモンド社, 1997.
- 59 [Pirolli 95] Pirolli, P., and Card, S.K.: Information foraging in information access environments, *Proc. of the ACM Conference on Human factors in Computing Systems(CHI'95)*, pp.51-58,1995.
- 60 [Pirolli 98] Pirolli, P.: Exploring Browser Design Trade-offs Using a Dynamical Model of Optimal Information Foraging, *Proc. of the ACM Conference on Human factors in Computing Systems(CHI'98)*, pp.33-40, 1998.
- 61 [Pirolli 99b] Pirolli, P., and Pitkow, J.: Distributions of surfers' paths through the World Wide Web: Empirical characterization, *World Wide Web*, Vol.2, pp.29-45, 1999.
- 62 [Pitkow 97a] Pitkow, J.: In Search of Reliable Usage Data on the WWW, *Proc. of the 6th International WWW Conference*, GVU Technical ReportGIT-GVU-97-13, 1997.
- 63 [Pitkow 97b] Pitkow, J.: Characterizing World Wide Web Ecologies, *Ph.D. Thesis, GVU, Georgia Institute of Technology*, 1997.
- 64 [Pitkow 98] Pitkow, J.: Summary of WWW characterizations, *Proc. of the 7th International WWW Conference*, pp.551-558, 1998.
<http://www7.scu.edu.au/programme/fullpapers/1877/com1877.htm>
- 65 [Pitkow 99] Pitkow, J., and Pirolli, P.: Mining longest repeating subsequences to predict world wide web surfing, *Proc. of the 2nd USENIX Symposium on Internet Technologies and Systems*, pp.139-150, 1999.

- 66 [Priolli 99a] Pirolli, P., and Card, S.K.: Information Foraging, *UIR Technical Report*, 1999.
- 67 [Quinlan 93] Quinlan, J. R. : *C4.5 : Programs for Machine Learning*, Morgan Kaufmann, 1993. (邦題 : AI によるデータ解析, トッパン, 1995.)
- 68 [Resnik 97] Resnik, P., and Varian H.R.: Recommender Systems, *Communication of the ACM*, Vol.40, No.3, pp.56-64, 1997.
- 69 [Russell 95] Russell, S. J., and Norvig, P.: *Artificial Intelligence A Modern Approach*, Prentice-Hall, 1995.
- 70 [Shapiro 99] Shapiro, C., and Varian, H. R.: *Information Rules*, Harvard Business School Press, 1999. (邦題: ネットワーク経済の法則, IDG, 1999.)
- 71 [Simon 55] Simon, H.: On a Class of Skew Distribution Functions, *Biometrika*, pp.425-440, 1955.
- 72 [Smith 99] Smith, M. D., Bailey, j., and Brynjolfsson, E.: Understanding Digital Markets: Review and Assessment, *Understanding Digital Economy*, pp.99-136, The MIT Press, 2000.
- 73 [Spiliopolou 99] Spiliopoulou, M., and Faulstich, L. C.: WUM: A Web Utilization Miner, *Proc. of EDBT Workshop WebDB98*, Springer Verlag, 1999.
- 74 [Stigler 61] Stigler, George. J.: The Economics of Information, *The Journal of Political Economy*, Vol.69, No.3, pp.213-225, 1961.
- 75 [Taylor 94] Taylor, L.D.: *Telecommunications Demand in Theory and Practice*, Kluwer Academic Publishers, pp.111-128, 1994.
- 76 [Train 87] Train, K.E., D.L. MacFadden and M. Ben-Akiva. : The demand for local telephone service : a fully discrete model of residential calling patterns and service choices, *Rand Journal of Economics*, Vol.18, No.1, pp.109-123, 1987.
- 77 [Varian 99] Varian, Hal.R. : Economics and Search, *SIGIR Forum*, Vol.33, No. 3, 1999.
<http://www.acm.org/sigir/forum/F99/Varian.pdf>
- 78 [Varian 00] Varian, Hal.R. : Market Structure in the Network Age, *Understanding Digital Economy*, pp.137-150, The MIT Press, 2000.
- 79 [Walrand 99] Walrand, J., and Varaiya, P.: *High-Performance Communication Networks-Second Edition Chapter10-Network Economics*, Morgan Kaufmann, 1999.
- 80 [Wigand 95] Wigand, T. R., and Benjamin, I. R.: Electronic Commerce: Effects on Electronic Markets, *Journal of Computer-Mediated Communication*, Vol.1, No.3, 1995.
<http://jcmc.huji.ac.il/vol1/issue3/wigand.html>
- 81 [CSJ 1996] サイバースペースジャパン (株) : 第 4 回 WWW 利用者調査, 1996. <http://www.csj.co.jp/www4/index.html>.
- 82 [ECN 96] 情報通信総合研究所 : 第 5 回 ECN 「インターネット利用に関するアンケート」, 1996. <http://www.commerce.or.jp>

- 83 [e-Japan 01] 高度情報通信ネットワーク社会推進戦略本部：e - J a p a n 戦略, 2001.
http://www.kantei.go.jp/jp/singi/it2/kettei/010122honbun.html
- 84 [相澤 98] 相澤彰子：H T T P ログファイルからの情報抽出によるインターネットドメインの分析, 電子情報通信学会論文誌 D-I, Vol.J81-D-I, No.11, pp.1201-1210, 1998.
- 85 [浅野 96] 浅野熙彦：入門多変量解析の実際, 講談社サイエンティフィック, 1996.
- 86 [池田編 97] 池田謙一編：ネットワーキング・コミュニティ, 東京大学出版会, 1997.
- 87 [インターネット白書 02] 日本インターネット協会：インターネット白書 2002, インプレス, 2002.
- 88 [井上 93] 井上明也・山本尚生：通信サービスにおけるユーザの選択行動分析法, 電子情報通信学会誌, Vol76, No.5, pp.510-517, 1993.
- 89 [井上 99] 井上哲浩：インターネット広告の特性と可能性, Japan Marketing Journal, Vol.71, pp.38-48, 1999.
- 90 [巖佐 98] 巖佐 備：数理生物学入門-生物社会のダイナミクスを探る, 共立出版, 1998.
- 91 [石野 95] 石野洋子：模擬育種法と機械学習を適用したマーケティング情報分析手法, 筑波大学経営・政策科学研究科経営システム科学専攻修士論文, 1995.
- 92 [上田 94] 上田徹：コンジョイント分析法に基づくサービス選好構造分析法の検討, 電子情報通信学会論文誌 B-I, Vol.J77-B-I, No.9, pp.542-549, 1994.
- 93 [大友 97] 大友篤：地域分析入門, 東洋経済新報社, 1997.
- 94 [岡橋 00] 岡橋孜, 矢川直弘, 及川直彦, 萩原雅之, 日高靖, 細井勉, 小林和夫：インターネット・マーケティングベーシックス, 日経 BP 社, 2000.
- 95 [風間 99] 風間一洋, 佐藤進也, 清水奨, 神林隆：WWW ユーザ操作履歴による HTML 文書の相関関係の解析, 情報処理学会論文誌, Vol.40, No.5, pp.2450-2459, 1999.
- 96 [片平 87] 片平秀貴：マーケティングサイエンス, 東京大学出版会, 1987.
- 97 [岸 97] 岸志津江：インターネット広告のオーディエンス測定と効果, 日経広告研究所報, 171号, pp.12-16, 1997.
- 98 [熊坂 01] 熊坂有三, 峰滝和典：IT エコノミー情報技術革新はアメリカ経済をどう変えたか, 日本評論社, 2001.
- 99 [黒崎 99] 黒崎政男：メディアの受容と変容, 情報の空間学～メディアの受容と変容～, pp.13-40, NTT 出版, 1999.
- 100 [下方 97] 下方 拓：インターネットアクセスにおけるユーザーの選択構造の分析, 情報通信学会誌, Vol.15, No.2, pp.69-79, 1997.
- 101 [下方 01a] 下方 拓：WWW アクセスにおける探索行動圏, 情報通信学会学会誌, Vol.19, No.3, pp.75-88, 2002.

- 102 [下方 01b] 下方 拓：コンテンツトランザクションによる WWW 探索行動パターン分析, 平成 13 年度情報通信学会年報, pp.21-34, 2002.
- 103 [国領 97a] 国領二郎：サイバースペースの経済空間としての特性, 情報処理, Vol.38, No.9, pp.764-771, 1997.
- 104 [国領 97b] 国領二郎：ネットワーク上の顧客間インタラクション, マルチメディア社会のシステムの諸相, 日科技連, 1997.
- 105 [小林 96] 小林四一：コンピュータ・コミュニケーションにおける電子コミュニティの構造分析, 筑波大学平成 7 年度修士論文, 1996.
- 106 [後藤 93] 後藤滋樹, 野島久雄：人間社会の情報流通における三段階構造の分析, 人工知能学会誌, Vol.8, 3, pp.348-356, 1993.
- 107 [サイモン 96] サイモン, A.ハーバート：システムの科学, パーソナルメディア, pp.66-70, 1996. (原著：Simon, A., H.:*The Science of the Artificial(Third Edition)*, MIT Press, 1996)
- 108 [佐藤 98] 佐藤俊雄：マーケティング地理学, 同文館, 1998.
- 109 [佐藤 99] 佐藤進也, 風間一洋, 清水奨, 神林隆：WWW トラフィック解析による情報生産者と情報消費者の数量的特徴づけ, 情報処理学会論文誌, Vol.40, No.7, pp. 2851-2860, 1999.
- 110 [柴山 01] 柴山悦哉：ブラウザのための可視化とナビゲーション支援, 人工知能学会誌, Vol.16, No.4, pp.509-514, 2001.
- 111 [商務省 00] 米国商務省：デジタルエコノミー2000 米国商務省レポート, 東洋経済新報社, 2000.
- 112 [末永編 98] 末永俊郎, 安藤清志編：現代社会心理学, 東京大学出版会, 1998.
- 113 [スキエナ 92] スキエナ, S. : *Mathematica 組み合わせ論とグラフ理論*, トッパン, 1992. (原著：Skiena, S. S. :*Implementing Discrete Mathematics: Combinatorics and Graph Theory with Mathematica*, Addison-Wesley Publishing, 1990.)
- 114 [総務省 02] 総務省編：平成 14 年版情報通信白書, 2002.
- 115 [高橋 99] 高橋正道, 北山聡, 金子郁容：ネットワーク・コミュニティにおける組織ウェアネスの計量と可視化, 情報処理学会論文誌, Vol.40, No.11, pp.3988-3999, 1999.
- 116 [多田 73] 多田和夫：探索理論, 日科技連出版社, 1973.
- 117 [大和総研 97] 株式会社大和総研：インターネット接続プロバイダの動向, 1996, 1997, <http://www.dir.co.jp>.
- 118 [東大・野村 96] 東京大学社会情報研究所・野村総合研究所：インターネットとパソコン通信に関するアンケート, 1996. <http://www.nri.co.jp/nri/news/960830.html>.
- 119 [豊島 00] 豊島雅和：Web 重力モデルに関する実証研究, 経営情報学会 2000 年度春季全国研究発表大会発表要旨集, pp.70-73, 2000.
- 120 [土木学会 95] 土木学会：非集計行動モデルの理論と実際, p51, 丸善, 1995.
- 121 [中井 97] 中井豊：インターネットにおけるホームページの成長現象とメカニズムについて, 情報通信学会誌, Vol.16, No.1, pp.59-66, 1997.

- 122 [中村 01] 中村知子 : WWW における情報検索行動の研究, 筑波大学大学院修士論文, 2001.
- 123 [長尾 00] 長尾真, 松山隆司, 杉本晃宏, 佐藤理史, 麻生英樹 : 情報の組織化, マルチメディア情報学 2, 岩波書店, 2000.
- 124 [永谷 02] 永谷敬三 : 入門 情報の経済学, 東洋経済新報社, 2002.
- 125 [成生 94] 成生達彦 : 流通の経済理論-情報・系列・戦略-, 名古屋大学出版会, 1994.
- 126 [丹羽 94] 丹羽芳樹・新田義彦 : 単語ベクトルを用いた多義語の意味推定, 自然言語処理 102-7, 1994.
- 127 [野島 00] 野島久雄, 新垣紀子 : 現実空間とバーチャル空間のナビゲーション, NTT R&D, Vol.49, No.5, pp.242-251, 2000.
- 128 [原 96] 原広司 : 空間の基礎概念と<記号場>~空間の比較社会学に向けて~, 時間と空間の社会学, pp.15-74, 岩波書店, 1996.
- 129 [藤原 93] 藤原章正 : 交通機関選択モデル構築における選好意識データの信頼性に関する研究, 広島大学博士論文, pp.67-71, 1993.
- 130 [藤本 96] 藤元健太郎 : サイバー空間におけるマーケティングの新潮流, 電子商取引のマーケティング戦略, pp.69-94, ダイヤモンド社, 1996.
- 131 [マクルーハン 87] マクルーハン, M. : メディア論, みすず書房, 1987.
- 132 [三友 95] 三友仁志 : 通話の経済分析, 日本評論社, 1995.
- 133 [森川 92] 森川高行・Moshe Ben-Akiva : RP データと SP データを同時に用いた非集計行動モデルの推定方法, 交通工学, Vol.27, No.4, pp.21-30, 1992.
- 134 [安田 01] 安田雪 : 実践ネットワーク分析 関係を解く理論と技法, 新曜社, 2001.
- 135 [山西 96] 山西健司 : 確率的コンプレキシティと学習理論, オペレーションズ・リサーチ, Vol.41, No.7, pp.379-385, 1996.
- 136 [若林 02] 若林幹夫 : 「情報都市」は存在するか?, 情報都市論, pp.223-262, NTT出版, 2002.

付録

電子メールで送付したアンケートを以下に示す。

インターネットアクセスの利用状況に関する調査協力のおねがい（全23問）

下方（げほう）と申します。

私は一会社員であるとともに、現在、社会人大学院において「インターネットローカルアクセスに関するユーザの選択行動の研究」をテーマに研究活動をおこなっています。この課題について、調査活動としてアンケートをおこなうことを計画しました。どうか、ご協力ください。

（会社の仕事とは一切関係ありません。自費でおこなっている活動です。）

この調査は、ご自宅からインターネットへアクセスする際に、どのようなアクセス回線を用いてどのようなインターネットサービスプロバイダーを利用しているかということに関して、利用者の方々に利用意識をうかがい、望ましいアクセスサービス案について研究するための資料とするものです。

ご回答は、各質問の最後にある回答欄のかっこ内に数字を記入してください。また、住所などの具体名は直接該当欄にご記入ください。お答えいただいた結果は本研究以外の目的に使用されることは決してありません。

アンケートの集計・解析の結果については機会をみてフィードバックしていきたいとおもいます。ご回答は10月11日（金）までにご返送ください。

どうかよろしくお願ひします。

また、この調査にご興味のある方もいらっしゃいましたら下記アドレスまでご連絡ください。

下方 拓（げほう たく）

筑波大学大学院経営システム科学専攻

E-mail : gehou@gssm.otsuka.tsukuba.ac.jp

（この調査は仕事とは一切関係ありませんが、私の職場は以下のとおり）

N T T 関東支社企画部

TEL.03-3244-7522 FAX.03-3244-7526

E-mail : takugeho@knt.vyc.ntt.jp

・まず、あなたのプロフィールについてお聞きします。

問1 現住所（市町村まで）と自宅電話番号の市外局番をご記入ください。

県：

市（郡）：

町：

市外局番：

問2 あなたの性別を教えてください。

1. 男

2. 女
回答欄()

問3 あなたの年齢を教えてください。

回答欄
()歳

問4 あなたの職種はなんですか。次のうちから1つ選んでお答えください。

1. 技術・研究開発
2. 営業
3. 企画・総務・経理
4. 学生
5. 自営業
6. その他()
回答欄()

・ 現状のインターネットの利用環境についてお聞きします。

問5 コンピュータをはじめて何年になりますか。

回答欄
()年

問6 現在ご自宅で利用のコンピュータはなんですか。次のうちから一つ選んでお答えください。

1. DOS/Vマシン
2. NEC98シリーズ
3. Macintosh
4. Unixマシン
回答欄()

問7 ご自宅からインターネットを利用されていますか。

1. はい
2. いいえ
回答欄()

(「いいえ」を選択された方は「将来のインターネットの利用についてお尋ねします」以降の問20に進んでください。)

問8 現在自宅からご利用される場合のインターネットサービスプロバイダ名を記入してください。(複数回答可)

1. :
2. :
3. :

問9 現在主に利用されているインターネットサービスプロバイダーの料金制度はどのようになっていますか。制度と金額を教えてください。

-
1. 定額制
 2. 完全従量性
 3. 定額 + 超過料金

回答欄

()

金額

定額分 月あたり()時間まで()円

超過分()分あたり()円

問10 現在ご利用のインターネットサービスプロバイダーのアクセスポイントはお自宅からどれくらいの位置にありますか。次のなかから一つ選んでください。

-
1. 同一市内(同一MA)
 2. 隣接した市外(隣接MA)
 3. 上記以外

回答欄()

問11 現在ご利用のインターネットサービスプロバイダーの品質について満足ですか。

-
1. はい
 2. いいえ

回答欄()

問12 問11で「いいえ」と答えた方にお尋ねします。不満な点はなんですか。次のうちから選んでください。(複数回答可)

-
1. 料金が低い。
 2. 接続スピードがおそい
 3. 話中回数が多い
 4. プロバイダーからの十分なサポートが得られない。
 5. その他()

回答欄()()()()

問13 インターネットサービスプロバイダーへのアクセス回線はなにをご利用ですか。次のなかから一つ選んでお答えください。

-
1. アナログ回線
 2. ISDN
 3. 専用線
 4. その他

回答欄()

問14 アクセス回線で割引サービスを利用されていますか。次のうちから選んでお答えください。(複数回答可)

-
1. 定額型(テレホーダイなど)
 2. 割引型(テレチョイスなど)
 3. その他

回答欄() ()

問15 ご自宅からインターネットを利用される主な時間帯を教えてください。次のうちから選んでお答えください。(複数回答可)

-
1. 平日昼間(午前8時から午後5時)
 2. 平日夜間(午後5時から午後11時)
 3. 平日深夜(午後11時から翌日午前8時)
 4. 土日休日昼間(午前8時から午後5時)
 5. 土日休日夜間(午後5時から午後11時)
 6. 土日休日深夜(午後11時から翌日午前8時)

回答欄() () ()

問16 一回あたりのおおよその接続時間を教えてください。

()時間()分

問17 一週間での接続回数を教えてください。

()回

問18 もっとも多く利用されるサービスはなんですか。次のうちから1つ選んでお答えください。

-
1. 電子メール
 2. WWW
 3. FTP
 4. その他()

回答欄()

問19 ご自分で公開ホームページをおもちですか。

-
1. はい
 2. いいえ

回答欄()

・ 将来のインターネット接続サービスの利用についてお聞きします。

問20 次のアクセス回線のなかから、今後自宅からのアクセスで利用したいとおもふものを上位3つを選んで、回答欄に番号を記入してください。

番号	回線種類 (速度)	導入費用	月額基本 使用料	従量料金(単位) 備考
1	アナログ回線 + テレホーダイ (28.8kbps)	0円	3550円	0円 通常のアナログ回線に比べ深夜の利用時間月12時間以上でお得
2	アナログ回線 (28.8kbps)		0円	1750円 10円(3分) 別途通話料必要
3	I N S 6 4 + INSテレホーダイ (64kbps)	2000円	5230円	0円 通常のI N Sと比べて深夜の利用時間月16時間以上でお得
4	I N S 6 4 (64kbps)		2000円	2830円 10円(3分) 別途通話料必要
5	デジタルアクセス64 (廉価版専用線64kbps)	90000円	28000円	0円 帯域保証あり
6	デジタルアクセス128 (廉価版専用線128kbps)		90000円	38000円 0円 帯域保証あり
7	OCN128 (近距離専用線として使用)		(未)	20000円 0円 帯域保証なし
8	CATV (10Mbps)	なし	30000円	0円 (アクセスとISPを区別不能) プロバイダーに加入の必要なし。

* 導入費用は現在アナログ電話回線を所有していることを前提に工事費や加入料の合計を提示

* 深夜時間は午後11時から翌朝午前8時まで

* デジタルアクセス：平成8年10月以降のNTTの廉価版専用線サービス。障害監視機能や自動バックアップ機能を省いてコストダウンを図ったもの。ルータなどの費用は除外。

* O C N : Open Computer Network. 来春までに開始予定のNTTのネットワークサービス。ベストエフォート型のサービスで帯域保証はない。

*CATV：ここでは、CATVの通常の放送サービスとは別に、インターネットへのアクセスサービスのみの提供が可能なケースを想定している。そのため、アクセスのみの経費は0円となっている。10Mbpsの通信速度も可能。追加支出で通常の放送サービスの視聴も可能。別途ケーブルモデムが必要。

 回答欄

1. 最も好ましい()
2. 2番目に好ましい()
3. 3番目に好ましい()

問21 次のインターネットサービスプロバイダー（ISP）等のダイヤルアップIP接続サービス（今後の提供予定分も含む）から、今後自宅から利用したいとおもつものを上位から3つ選んで、回答欄に番号を記入してください。

 （注）

・提供形態としては、ISP（全国型）、ISP（地域密着型）、CATV、OCNの4つがあります。ISPはInternet Service Providerの略。ISP（全国型）は全国にアクセスポイントを設置しているもの。ISP（地域密着型）はある県や市などに限定的に1から3程度のアクセスポイントを設置しているもの。

・ブランドは大手と中小の2種類。

・料金制度は従量性と定額性の2種類。料金は従量制は3分あたりの金額、定額制は1ヶ月あたりの金額を提示。

・アクセスポイントまでの距離は市内と市外の2種類。電話料金に影響します。

・アクセス1回線あたりの加入制限のありなしは話中頻度に影響します。

・ISDN対応は64kbpsと128kbpsの2種類があります。

・バックボーンの色度は256kbpsと1.5Mbpsの2種類、接続後の回線スループットに影響します。

・サポートは平日昼間と24時間の2種類があります。

 ・ダイヤルアップIP接続

番号	形態	ブランド	料金制度	料金	アクセス ポイント までの距離	アクセス1回 線あたりの加 入者数制限	ISDN 対 応	バック ボーン の速度	サポート
1	ISP（全国型）	大手 市外	従量制 加入制限なし	¥10（3分）			128kbps	1.5Mbps	24時間
2	ISP（全国型）	中小 市外	定額制 加入制限あり	¥3,000（1ヶ月）			64kbps	256kbps	24時間

3	I S P (全国型)	大手 市内	定額制 加入制限なし	¥5,000 (1ヶ月) 128kbps	256kbps	平日昼間
4	I S P (全国型)	中小 市内	従量制 加入制限あり	¥50 (3分) 64kbps	1.5Mbps	平日昼間
5	I S P (地域密着型)	大手 市内	定額制 加入制限あり	¥1,500 (1ヶ月) 64kbps	256kbps	平日昼間
6	I S P (地域密着型)	中小 市内	従量制 加入制限なし	¥30 (3分) 128kbps	1.5Mbps	平日昼間
7	I S P (地域密着型)	大手 市外	従量制 加入制限あり	¥50 (3分) 64kbps	1.5Mbps	24時間
8	I S P (地域密着型)	中小 市外	定額制 加入制限なし	¥5,000 (1ヶ月) 128kbps	256kbps	24時間
9	C A T V	中小 市内	従量制 加入制限なし	¥10 (3分) 64kbps	256kbps	24時間
10	C A T V	大手 市内	定額制 加入制限あり	¥3,000 (1ヶ月) 128kbps	1.5Mbps	24時間
11	C A T V	中小 市外	定額制 加入制限なし	¥5,000 (1ヶ月) 64kbps	1.5Mbps	平日昼間
12	C A T V	大手 市外	従量制 加入制限あり	¥50 (3分) 128kbps	256kbps	平日昼間
13	O C N	市外	定額制 加入制限あり	¥1,500 (1ヶ月) 128kbps	1.5Mbps	平日昼間
14	O C N	市外	従量制 加入制限なし	¥30 (3分) 64kbps	256kbps	平日昼間
15	O C N	市内	従量制 加入制限あり	¥50 (3分) 128kbps	256kbps	24時間
16	O C N	市内	定額制 加入制限なし	¥5,000 (1ヶ月) 64kbps	1.5Mbps	24時間

* C A T V : 今後提供される C A T V のインターネットのアクセスサービス。
料金にはアクセス部分 (通常の電話通話料) も含まれる。通信速度は LAN 型ダイヤルアップの場合は数 Mbps が可能。

* O C N : OCN へのダイヤルアップ接続のネットワーク部分のみの想定料金。

回答欄

1. 最も好ましい ()
2. 2 番目に好ましい ()
3. 3 番目に好ましい ()

問22 次のインターネット接続サービスの専用線 I P サービスの中から利用したいとおもうものを上位から 3 つ選んで、回答欄に番号を記入してください。

(注)

- ・各属性は前問同様です。
- ・アクセスポイントまでの距離は15kmと30kmの専用線の料金区分で区別。30kmになるとほぼ2倍のアクセス料金が別途かかると考えてください。CATVとOCNにはありません。

・専用線 I P 接続

番号	形態	接続速度	ブランド	月額料金
		アクセス ポイント 距離	バック ボーン の速度	サポート
1	I S P (全国型)	128kbps 15km	大手	¥30,000 1.5Mbps 24時間
2	I S P (全国型)	128kbps 30km	中小	¥100,000 256kbps 24時間
3	I S P (全国型)	128kbps 15km	大手	¥200,000 256kbps 平日昼間
4	I S P (全国型)	128kbps 15km	中小	¥200,000 1.5Mbps 平日昼間
5	I S P (地域密着型)	128kbps 15km	大手	¥30,000 256kbps 平日昼間
6	I S P (地域密着型)	128kbps 15km	中小	¥100,000 1.5Mbps 平日昼間
7	I S P (地域密着型)	128kbps 30km	大手	¥200,000 1.5Mbps 24時間
8	I S P (地域密着型)	128kbps 30km	中小	¥200,000 256kbps 24時間
9	C A T V	10Mbps	中小	¥30,000 256kbps 24時間
10	C A T V	10Mbps	大手	¥100,000 1.5Mbps 24時間
11	C A T V	10Mbps	中小	¥200,000 1.5Mbps 平日昼間
12	C A T V	10Mbps	大手	¥200,000 256kbps 平日昼間
13	O C N	128kbps	中小	¥30,000

		1.5Mbps 平日昼間	
14	OCN	128kbps 256kbps	大手 平日昼間 ¥100,000
15	OCN	128kbps 256kbps	中小 24時間 ¥200,000
16	OCN	128kbps	大手 1.5Mbps 24時間 ¥200,000

ただし、月額料金にはルータ代、DSU代は含まれていません。

回答欄

1. 最も好ましい()
2. 2番目に好ましい()
3. 3番目に好ましい()

問23 その他、今後のインターネット接続について望むことや感想など、あればなんでも結構ですからご記入ください。

記入欄

ご協力ありがとうございました。

全研究業績リスト

査読付き学術論文

- 下方 拓, 「インターネットアクセスにおけるユーザーの選択行動の分析」, 情報通信学会誌, Vol.15, No.2, pp.69-79, 1997年9月.
- 下方 拓, 「WWW アクセスにおける探索行動圏」, 情報通信学会誌, Vol.19, No.3, pp.75-88, 2002年1月.
- 下方 拓, 「コンテンツトランザクションによる WWW 探索行動パターン分析」, 情報通信学会年報 13, pp.21-34, 2002年3月.

学会・全国大会

- 下方 拓, 「インターネットアクセスにおけるユーザーの選択行動の分析」, 第14回情報通信学会大会, 1997年6月.
- 下方 拓, 「WWW アクセスにおけるインターネット探索行動圏モデル」, 第18回情報通信学会大会, 2001年6月.
- 下方 拓, 「コンテンツトランザクションによる WWW 探索行動パターン分析」, 経営情報学会 2001年春季全国研究発表大会予稿集, pp.245-248, 2001年6月.
- Taku Geho and Takao Terano : Finding Web Navigation Structures by Cost-based User Behaviors, *Proc. of The 6th SANKEN International Symposium*, pp. 65-66, March, 2003.