論文・事例研究

外部ソースを活用したウェブ・マーケティングのための分析フレームワークの提案

矢野　順子，加藤　元英，末永　高志，生田目　崇

1. はじめに

2006年の総務省推計によると，インターネットの利用人口は8,754万人にのぼり，インターネットは情報インフラとしてだけではなく，いまや生活インフラとして欠かすことのできないものとなっている[9]．また，消費生活にとってインターネットが果たす役割は，その黎明期では情報探索・収集のための利用が主であったが，近年では，商品購買などの生活活動のサポート，さらには消費者自らが情報発信する場へと大きく変化しており，またそれらが互いに結びついている．生活行動の中でも特に購買行動については，インターネットによる影響が大きい[10]．インターネットは，商品購買チャネルとしてだけではなく，商品に関する情報収集チャネルとしての役割も大きく，消費者の購買行動における情報処理プロセスに大きく作用している．

従来，企業のウェブ・マーケティング戦略は，ポータルサイト上の「バナー広告」を用いた消費者へのアピール，自社サイトへの誘導などが中心であった．しかし，1998年にサービスを開始したGoogleのような高度な情報検索サービス事業者の登場により，SEO（Search Engine Optimaization）対策や検索連動型広告への対応など，インターネット上の消費者行動・志向にあわせるような戦略にシフトしている．また最近では，ブログ（ウェブログ）やSNS（Social Networking Service）を利用し，消費者自らが商品評価や趣味・嗜好など様々な情報を発信する仕組みも一般化している．これらの情報を他の消費者が利用する，消費者間でのコミュニケーションも活発になっており，いわゆる「バズ・マーケティング」も注目されている[2]．

こういったインターネット環境の発展の中で，企業が消費者に対して自社の製品・サービスを効果的にアピールし，収益につなげるアクションを行うためには，企業自身が発信する情報だけではなく，CGM（Consumer Generated Media）による消費者発信情報にも目を向け，それらを企業のマーケティング活動に活用していくという視点も重要になってくると考えられる．

そこで本研究では，インターネット上における企業サイトの外部のデータ（以下，外部ソース）をウェブ・マーケティング戦略に活用するための分析フレームワークを提案する．また，提案したフレームワークに実データを適用し，分析結果を示す．

2. ウェブ閲覧行動に関する既存の研究と本研究の動機付け

近年では，消費者が製品・サービスの情報の入手元として，インターネットが新聞・雑誌を抜き第一位となっており，製品・サービスを認知し，情報収集するチャネルとして広く利用されている[8]．また，インターネット上で消費者が発信する情報は，企業が発信する情報と異なり製品・サービスに関してニュートラルな立場からの情報である．これらは検索サイトにより手軽に検索でき，ウェブのトラックバック機能などによって消費者間の積極的なコミュニケーションを誘発し，口コミの新たなチャネルとなる．このような消費者発信情報は従来の企業プロモーションと異なり，消費者自身の声が直接反映され，実世界の口コミよりも早く広範囲に届く．

Eコマースを行う企業におけるウェブ・マーケティングの目的のひとつは，自社製品・サービスに関する情報を消費者へ届け，自社のウェブ・サイトへ誘導し

やの　じゅんこ，かとう　もとひで，すえなが　たかし
㈱NTTデータ　技術開発本部
〒135-8671 江東区豊洲3-3-9
なまため　たかし
専修大学　商学部
〒214-8580 川崎市多摩区東三田2-1-1
受付07.7.13　採択07.11.15

て，購買（コンバージョン）に結びつけることである．製品・サービスに関する情報は自社が発信する情報だけでなく，上記に論じたような消費者発信情報などの外部ソースも数多くあり，企業にとってはこれらの外部ソースは無視できないものになりつつある．外部ソースの情報を効果的に利活用するスキームは，企業のマーケティング戦略にとって新たな機会となりうることからも，研究，実務両面でも関心が高まってきている．

インターネット上の消費者行動分析においては，サイトの構成を分析したり，また，ウェブ閲覧履歴データ（以下，ウェブ・ログ・データ）のパターンを分析し，以降に閲覧されるであろうページを予測するような研究が多く行われている．データとの親和性よりマルコフ・モデルや離散選択モデルが使われた研究例が複数ある[1][3][6]．ただし，これらの研究では，消費者のサイト内行動に限定されており，サイト到着からコンバージョンもしくは離脱までの行動をモデル化してはいるが，どのような情報を元にサイト訪問されたかという自社サイト外のデータや行動については考慮されていないものがほとんどであり，サイト訪問の目的や訪問者の情報検索志向を考慮したものではない．

また，外部ソースの活用については，インターネット上の文章を効率的かつ大量に収集し，評判分析などのように文脈を解析し集計する試みがなされている[12]．実務面では，ウェブ・ログ・データは，配信もしくはバナー広告を効果的に行うための行動ターゲティング広告戦略において用いられている[15][16]．行動ターゲティング広告は，閲覧履歴や閲覧者属性などを分析することで，閲覧者の志向にあった広告配信もしくは広告表示を行おうというものである．

他には，クローラー技術などを利用して，ブログやニュース記事などからリアルタイムで情報収集を行いながら，評判分析やキーワードの時系列集計を行っているサービス例もある[13][14]．

これらは，外部ソース情報を効率的なSEOやリスティング広告などの自社サイト外への情報発信を念頭においたものであり，外部ソース情報と企業サイト情報との結びつきを考慮したものではない．外部ソースの重要性の高まりとともに，各閲覧者がどのような情報を元に訪問し，さらにどのようにサイト内を閲覧しながらコンバージョン（もしくは離脱）に結びつくかという，相互の関係を考慮することは，今後のウェブ・マーケティング戦略に新たな方法論を提供することができる．

3. 提案するウェブ・マーケティング分析フレームワーク

消費者行動の分野で情報処理プロセスをモデル化する研究は，認知科学を基礎として広く行われてきた．一般には，消費者の購買行動に関する情報処理プロセスは，「商品の認知」→「知覚/学習/整理」→「購買」といった過程を経る[7]．

実市場では，消費者は広告やチラシ，テレビコマーシャルなどの外部刺激により商品を認知しながら，商品に関する情報収集・集積を経ながら比較・検討して，購買の意思決定をするというプロセスを踏む．実社会ではこれらに関してそれぞれ別々の手段で行われることが多かった．インターネットの登場は，この一連の作業をインターネット上で効率的かつ効果的に行うことのできる場を創生した．

図1のように，「認知」，「知覚/学習/整理」のプロセスではブログやSNSといった口コミサイトも利用されている．また，上述の通り，多くの消費者は，情報検索して情報収集をする受動的な行動だけではなく，購入した商品の情報や評価を自身のブログなどで能動的・積極的に発信するようになってきている．

本研究では，購買プロセスの認知過程における消費者自身が発信する外部ソースに着目し，ウェブ閲覧に関する外部ソース利活用のための分析フレームワークとして図2に示すようなフローを提案する．基本的な流れは「収集」「分析」「評価」の3つのステップからなる．

はじめに，Step 1ではウェブ・ログ・データに残された外部サイト情報について，インターネットから

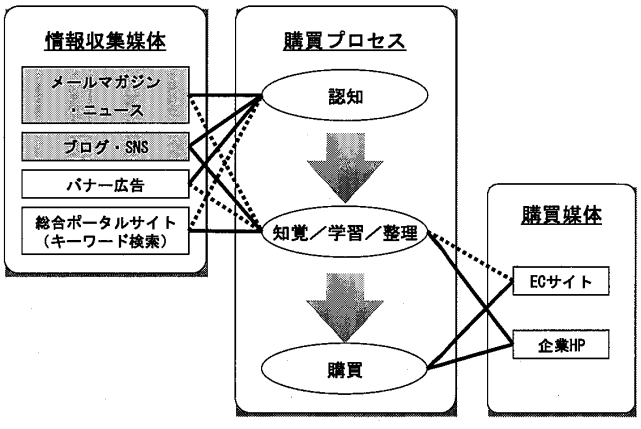

図1 購買プロセス

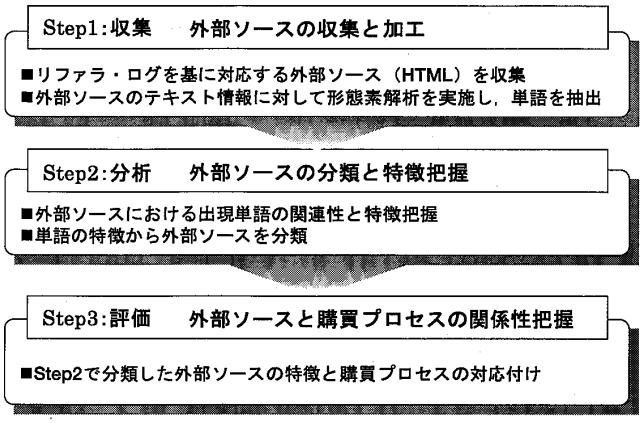

図2　分析フロー

htmlファイルを収集し，単語ごとに出現頻度などを算出する．次に，Step 2では，Step 1で算出した単語頻度をもとに外部ソースの特徴を把握する．そしてStep 3では，消費者の購買意識の変化や，実際のコンバージョンに結びつくような外部サイトの特徴，およびその関係性を把握することを目的としている．

4. 分析データ

本稿で使用するデータは，経営科学系研究部会連合協議会主催の平成18年度データ解析コンペティションで提供された，あるソフトウェア商品紹介サイトの「ウェブ・アクセス・ログ・データ」である．

本稿で使用するデータの詳細を以下に示す．

- データ期間：2006年1月1日～6月30日の6カ月間
- データ項目：セッションID，参照元（リファラ）URL，参照先（リクエスト）URL

なお，総セッション数は22,276件であり，総ページビュー数は56,353件であった．

分析データは，当該サイト内の各ページへの訪問履歴がセッションID情報などと同時に得られているため，セッションIDごとにウェブ・ログのリクエストURLを追跡することにより，サイト内における閲覧履歴を把握することができる．また，リファラURLには，どのサイトからどのような方法でこのサイトに訪問したかという情報が含まれており，閲覧履歴をたどるデータとなる．同一セッションIDにおいて，最初にリクエストされたページはランディングページと呼ばれ，そのページにおけるリファラは，どのサイトから訪問したのかという外部ソースの情報である．さらに，この項目には，例えば検索サイトでどのようなキーワードで検索したかといった情報も含まれる．したがって，この項目を利用することで，訪問者がどのサイトから訪問したかを把握することができる．

5. 外部ソースの収集と加工

Step 1の目的は，サイト訪問者がアクセスするに至った外部ソースのデータを収集し，分析するためのデータ形式に加工することである．具体的には，前節で述べたようにランディングページにおけるリファラURLを辿ることで，そのページ・ファイルを収集し，記述されているテキスト情報を単語単位に分解する．以下，外部ソース収集の際の条件について示したあと，外部ソースに含まれるテキスト・データの分析結果を述べる．

5.1 外部ソースの収集

リファラ情報をもとに，リファラURLのファイル・データを収集する．リファラに含まれているのは，

- 自社サイト
- 総合ポータルサイトのキーワード検索結果からのリンク
- 検索連動型広告からのリンク
- ニュース記事などパブリシティ情報内でのリンク
- ブログなどのCGM内でのリンク
- メールマガジン記事内のリンク
- その他

のような情報である．また，各ブラウザに登録されている「お気に入り」からの訪問や，URLを直接打ち込むような場合は，リファラ項目には何も記載されない．

本研究で着目するのは，購買プロセスの認知プロセスにおいて利用された消費者自身が発信する外部ソースである．そこで上記のような外部ソースのうち「自社サイト（本ソフトウェア制作会社のHP)」，「総合ポータルサイト」，「ブログ内の設置された検索エンジン」，「メール」は分析の対象外とした．

その結果，ユニークなサイトとして，227件のhtmlファイルを収集した[1]．収集した227件のhtmlファイルをリファラとしたセッションは651件であった．図3は，横軸をリファラURL，たて軸をリファ

[1] 実際にはさらに多くのURL情報が存在したが，ページ削除，リンク切れ，コンテンツ更新のように本分析との関わりが認められなかったものについても分析対象外とした．

ラURLからのセッション数としたグラフである．この図に示すように，全体の227件のうち上位の約10%で半分以上のセッションをカバーしていることがわかる．そのうち最も件数の多かったのは，191のセッションのリファラとなっていた外部ソースである．このリファラのページには，分析対象のデータのソフトウェア制作会社とは直接の関係はないものの，本サイトで提供するツールに関する詳細な紹介が記載されていた．

5.2 外部ソース・テキスト・データの分析

収集した外部ソース・データからテキスト・データを抽出し，形態素解析[4][2]を行って数詞を除く2文字以上の名詞を抽出した．抽出された単語のtf (term frequency) 値を算出し，それらが各外部ソースに含まれる度合いを分析する．そのために，単語抽出対象ファイルjに単語候補iが出現する数$tf_{i,j}$から，単語ごとのtf値の合計(tf_i)[5]と，単語候補iが含まれるファイル数であるdf (document frequency) 値(df_i)[5]を求めた．具体的な算出方法は以下の通りである．

$$tf_i = \sum_j tf_{i,j} \tag{1}$$

$$df_i = \sum_j 1_{\{tf_{i,j} \neq 0\}} \tag{2}$$

ただし，df_iの値が非常に高い単語は，ほとんどのソースに含まれていることを示しているため，サイトの特徴を抽出する目的には不向きである．そこで，df_iの値が抽出した外部ソース・ファイル数のほぼ半数を超えている上位8単語[3]は分析から除いた．そして，単語ごとにtf_iの値を集計し，値の高い60単語を分析に使用した．特徴抽出に用いる60単語を表1に示す．表1より，全体として「ツール」，「ホームページ」，「エンジン」といったインターネットやソフトウェアに関する単語が多いことがわかる．一方，「アフィリエイト」，「収入」などインターネットを利用した情報提供者側に関する単語も多く見られた．

6. 外部ソースの分類と特徴把握

Step 2ではStep 1の分析の結果，抽出された単語に着目し，各々の外部ソースの特徴把握を行う．上述したように外部ソースはその形態においてもさまざまであるが，内容についてもそれぞれのサイト運営者の方針によりさまざまである．例えばブログの中にも，日常の出来事を日記形式で記述するような生活者サイトや，ある特定の分野に関する技術情報を提供するようなサイトがある．したがって，企業にとって，すべてのリファラが有効なものというわけではなく，どのような外部ソースが企業にとって有効であるかを判断しなければならない．そこで，外部ソースの類型化を行う．

外部ソースの類型化を行うために，前節で抽出したソース別の単語別の頻度を用いた主成分分析を用いて，外部ソースに共通する特徴を分析する．そして，主成分分析により抽出された特徴およびそれを表す主成分負荷量を基に，クラスター分析を行い，外部ソースを分類する．

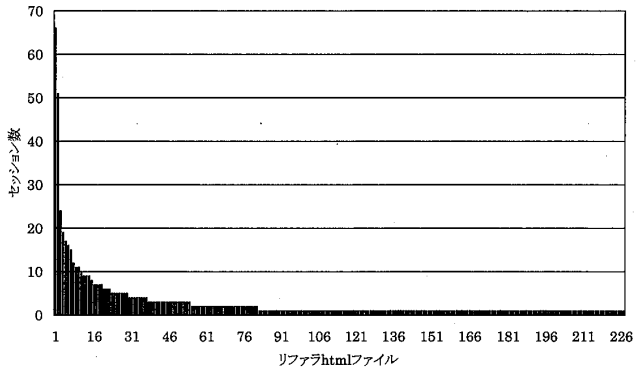

図3 リファラURL別セッション数

表1 tf値の高い60単語

単語	tf値	単語	tf値	単語	tf値
記事	492	販売	253	分析	175
リンク	443	ランキング	253	キーワード	172
広告	408	表示	246	企業	172
ツール	403	提供	245	ビジネス	172
ホームページ	403	制作	244	Web	171
レンタル	386	簡単	242	シビラ	169
ページ	371	可能	233	運営	159
アフィリエイト	348	効果	232	状況	158
登録	343	会社	227	通信	157
ネット	342	自分	227	人気	153
機能	342	銘柄	217	必要	153
作成	340	SEO	213	クリック	152
利用	337	インターネット	207	WEB	151
メール	329	収入	191	リンク集	151
ログ	314	対応	189	更新	151
エンジン	307	紹介	188	時間	150
アップ	303	投資	186	ヤツ	146
対策	297	注目	182	報酬	146
携帯	259	ショッピング	180	ソフト	145
サーバー	254	ユーザー	176	マーケティング	142

[2] 与えられた文を形態素・語の並びに分解し，それぞれの形態素・語の品詞などを決定する．

[3] 実際には$df_i > 112$として「サイト」，「アクセス」，「無料」，「情報」，「検索」，「解析」，「ページ」，「サービス」を除外した．

表2　主成分負荷量

単語	第1主成分負荷量	第2主成分負荷量
銘柄	−0.615	−0.381
注目	−0.448	−0.264
投資	−0.295	−0.165
携帯	−0.222	0.038
ネット	−0.161	0.029
ランキング	−0.147	−0.012
提供	−0.126	0.254
機能	−0.125	0.278
メール	−0.122	0.229
記事	−0.107	0.208
対応	−0.102	0.105
ビジネス	−0.101	−0.020
ツール	−0.098	0.220
ソフト	−0.092	0.178
レンタル	−0.089	0.184
分析	−0.089	0.137
ページ	−0.087	0.067
利用	−0.086	0.177
運営	−0.084	0.082
販売	−0.082	0.183
サーバー	−0.082	0.102
作成	−0.082	0.177
登録	−0.081	0.115
対策	−0.079	0.106
ログ	−0.071	0.048
エンジン	−0.071	0.138
ホームページ	−0.069	0.174
広告	−0.068	0.166
インターネット	−0.067	0.126
アップ	−0.067	0.094

6.1　主成分分析

外部ソースにおける出現単語の関連性と特徴を把握するために，前節で抽出された60単語について，各外部ソースごとの単語同時出現頻度をデータとして，主成分分析を行った．分析の結果算出された主成分負荷量を表2に示す．表2の単語は，第1主成分負荷量がすべてマイナスになったため，絶対値の大きいものから30単語を抽出した．抽出された単語の中には，第2主成分負荷量の上位14単語が含まれていた．

主成分負荷量から，第1主成分は「銘柄」，「注目」，「投資」の単語に対する負荷量が低く「評価・推奨・利益追求に関する軸」と解釈された．また，第2主成分は「機能」，「提供」，「メール」の主成分負荷量が上位となり，「サイト運営・分析に関する軸」と解釈された（図4）．

各主成分の寄与率と累積寄与率が表3であり，この表のように第10主成分までで累積寄与率は81.1%となったため，以降の分析では第10主成分までの値を用いて外部ソースの類型化を行う．

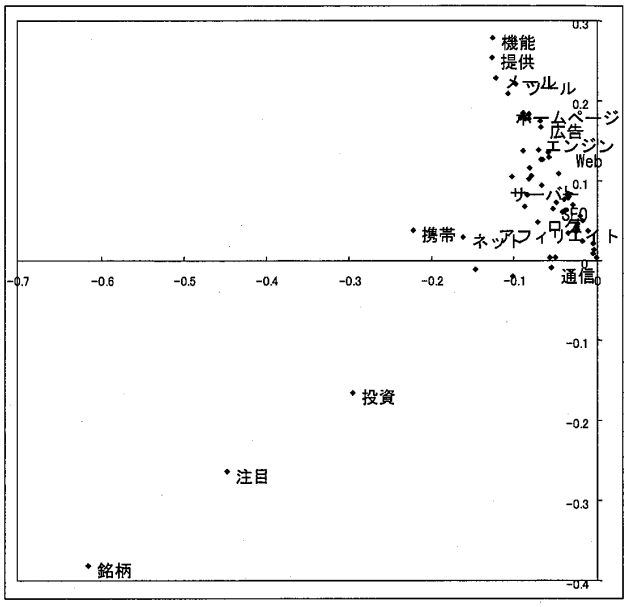

図4　第1主成分と第2主成分

表3　主成分の累積寄与率

主成分	標準偏差	寄与率	累積寄与率
第1主成分	13.53	22.8%	22.8%
第2主成分	12.31	18.8%	41.6%
第3主成分	8.05	8.1%	49.7%
第4主成分	7.23	6.5%	56.2%
第5主成分	6.74	5.7%	61.8%
第6主成分	6.32	5.0%	66.8%
第7主成分	5.88	4.3%	71.1%
第8主成分	5.70	4.0%	75.2%
第9主成分	5.01	3.1%	78.3%
第10主成分	4.73	2.8%	81.1%

6.2　クラスター分析

主成分分析により得られた特徴から外部ソースを分類するために，前節の第10主成分までの主成分得点を用いてクラスター分析を行った．なお，アルゴリズムはK-means法を用い，解釈のしやすさからクラスタ数は3とした．得られた各クラスタの主成分得点の平均値を表4に示す．表より，クラスタ1は第1主成分の傾向が非常に強く「利益追求の情報提供，評価推奨を行っているサイト」，クラスタ2は「ウェブサイトの運営や分析に興味があるサイト」，クラスタ3は「その他」のクラスタとなった．またクラスタごとに分類された外部ソースの特徴をみると，クラスタ1は，ウェブ・アクセス解析に関する最新動向やツールの紹介ページが多く，クラスタ2にはホームページ制作会社の企業HPなどが多く含まれていた．

表4　クラスタ別主成分得点平均値

	クラスタ 1	クラスタ 2	クラスタ 3
第 1 主成分	−26.474	−0.441	3.921
第 2 主成分	16.781	2.393	−3.479
第 3 主成分	−1.937	0.762	−0.057
第 4 主成分	−1.498	5.029	−2.060
第 5 主成分	−5.018	1.166	0.127
第 6 主成分	−0.500	−1.001	0.490
第 7 主成分	0.485	−2.516	1.113
第 8 主成分	1.098	−0.343	0.001
第 9 主成分	2.062	1.458	−1.014
第 10 主成分	−0.206	0.089	−0.046

表6　クラスタ別の閲覧率

購買プロセス	クラスタ 1	クラスタ 2	クラスタ 3	全体
一般利用	0.0%	0.0%	0.8%	2.1%
積極的情報収集	11.1%	17.7%	25.2%	38.0%
商品購買	16.7%	5.6%	6.7%	6.7%

表7　クラスタ別のパートナーシップの閲覧率

参照先サイト分類	クラスタ 1	クラスタ 2	クラスタ 3	全体
パートナーシップ	0.0%	4.0%	3.1%	1.6%

表5　購買プロセスと参照先サイトの分類

購買プロセス（参照目的レベル）	参照先サイト分類	参照先サイト内容
一般利用	アクセス・ログ解析	コラム，メルマガ，用語集
積極的情報収集	アクセス解析ツール	キャンペーン，機能，問い合わせ，FAQ
商品購買（コンバージョン）	申込	申込ページ，申込の流れ，キャンペーン
その他	会社概要	TOP，会社概要，他社製品紹介，プレスリリース
その他	パートナーシップ	アフィリエイト募集，代理店募集

7. 外部ソースと購買プロセスの関係性把握

Step 3では，Step 2において分類した外部ソースとユーザの購買プロセスを対応付け，コンバージョンに至るサイト訪問者が，どのような外部ソースを参照しているのかについて分析する．

商品販売サイトにおけるコンバージョンは，一般には実際に購入申し込みをした場合を指すが，分析に使用したデータは専門性の高いソフトウェアに関するサイトのアクセス・ログであり，購入申し込みにまでいたった件数は非常に少なかった．そこで，商品購買申し込みページ，もしくは試用キャンペーン申し込みページ[4]へアクセスした場合をコンバージョンと定義した．

サイトに含まれるページは100以上あり，多くはカテゴリ別に構造化されているが，グローバルナビゲーションを有しているため，異なるカテゴリ間のページについてもリンクが張られている．ただし，このリンクはそれぞれのカテゴリのトップページについて張られている．

分析に使用したデータのサイトは，当該のソフトウェアに関する情報のみでなく，関連する項目や，コラム，用語集，会社概要などといった，当該商品とは直接的な関係の薄いページも多くある．そこで，各ページ情報を掲載内容から，表5のように，5つのカテゴリに分類した．

ここで，リファラ情報である外部ソースの特徴の違いにより，その後の消費者の購買プロセスに違いがあるかどうかについて分析する．まず，消費者の購買プロセスである「認知」「知覚/学習/整理」「購買」について，本稿ではサイトの参照目的レベルに応じて「一般利用」「積極的情報収集」「商品購買（コンバージョン）」に分類した．そして，表5のように，3つの購買プロセスと分類した5つの参照先サイトとを対応させた．なお，「会社概要」および「パートナーシップ」は，購買の意思決定に直接的に関わらないサイトと仮定し「その他」とした．「一般利用」レベルは，「アクセス・ログ解析」に関するサイトを閲覧した場合，「積極的情報収集」は，参照先が「アクセス解析ツール」に関するサイトの場合，「コンバージョン」レベルは参照先が「申込」の場合とそれぞれ定義した．

外部ソースの特徴と購買プロセスとの関係を把握するため，クラスタ別に分類したサイトカテゴリごとの閲覧率を算出した．ここでの閲覧率とは，各クラスタに含まれる外部ソースを通じて訪問したユーザが参照目的で分類されたサイトの閲覧をした割合を示したものである．その結果を表6に示す．表より，クラスタ

[4] 分析対象データの企業では，当該商品について1カ月無料利用のキャンペーンを展開していた期間があり，メールなどで周知していた．

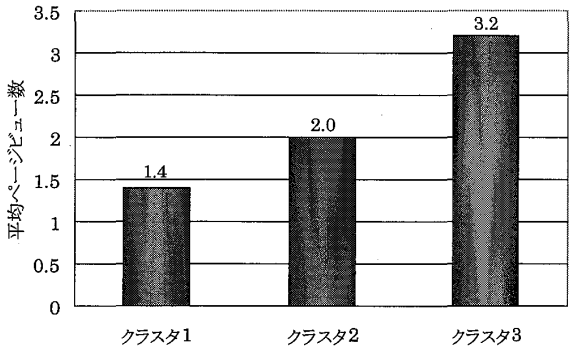

図5　平均ページビュー数

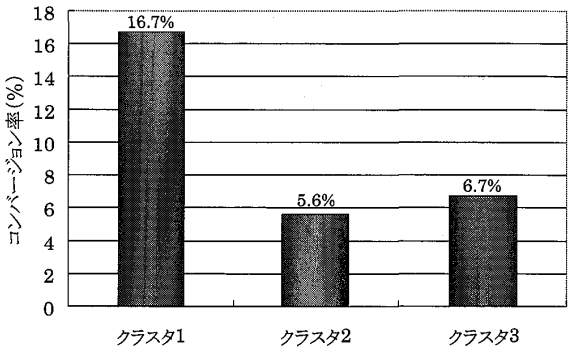

図6　コンバージョン率

1は，その他のクラスや全体に比べ「申込」サイトの閲覧率が非常に高いが，「アクセス・ログ解析」に関するサイトは閲覧していないことが分かる．これより，クラスタ1に含まれる外部ソースは，消費者に「一般利用」や「積極的情報収集」といった購買プロセスをあまり介させずに，効率的にコンバージョンに結びつける役割を果たしているといえるであろう．なお，全体の「積極的情報収集」の閲覧率が高いのは，そのページへのリンクがロボット型の検索エンジンによる検索結果に表示され，直接ツールのページにアクセスするような場合が多く含まれているためと考えられる．

一方，クラスタ2は，どの参照先サイトの閲覧率も高くはないが，購買プロセスと関係のない「パートナーシップ」に関する閲覧率が4%と全体の1.6%に比べ高かった（表7）．その理由としては，クラスタ2の外部ソースは，ウェブサイトの運営や分析に興味があるサイトであるため，このようなサイトから訪問するユーザは当該商品の消費者というよりも，商品を提供する企業である可能性が高いためと考えられる．

さらに，データ全体の1セッションあたりの平均ページビュー数を算出すると2.5であった．これに対して，外部ソースクラスタ別の平均ページビュー数をみると，クラスタ1は非常に少ないことがわかる（図5）．また前述したとおり，外部ソースクラスタ別にコンバージョンである「申込」ページに至った割合は，クラスタ1が最も高い（図6）．以上から，このサイトにおいてはクラスタ1に属する外部ソースを経由する訪問者が，他のクラスタよりも直接的にコンバージョンに結びついていることが分かった．この理由としては，クラスタ1の特徴として示したようなウェブ・アクセス解析に関する最新動向やツールの紹介ページやパートナーシップに関する記述がある「利益追求の情報提供，評価推奨を行っている」サイトから訪問するユーザは，これらの外部ソースを通じて，対象商品に関する紹介など購買に繋がる情報をしっかりと収集できているためであると考えられる．

8. おわりに

本稿では，アクセス・ログ解析の新たな視点として，外部ソースとウェブ・ログ・データを利用したウェブ・マーケティング戦略のための分析フレームワークを提案した．本稿の事例の場合，参照元の外部ソースの特徴によって，その後の購買行動，特にコンバージョン率が異なることを確認できた．これらの分析により，各企業は自社サイトへの誘導または商品購買の可能性の高い外部ソースの特徴を把握することができ，より効果的なウェブ・マーケティング戦略を構築できることが期待できる．例えば，コンバージョン向上に繋がる外部ソースへの積極的な情報掲載依頼を行うなど，CGMとの協力的な関係を築き上げることが可能となるであろう．

今後の展開としては，今回は用いなかったHTTP cookieの情報などを利用して特定できる訪問者へのアプローチ方法や，ランディング・ページとコンバージョンの関係についての考察が考えられる．さらに，今回は各ページの閲覧時間は考慮しなかったが，情報収集の面からは滞留時間は重要なデータであるので，これらを含めた分析フレームワークの構築も課題である．また，データが入手できるならば，ウェブのトラックバックやコメントなどの消費者間のコミュニケーションまで含めた，インターネット環境における効果的なマーケティング戦略についても分析を試みたい．

参考文献

[1] P. Baldi, P. Frasconi and P. Smyth, *Modeling the Internet and the Web : Probablistic Methods and*

Algorithms, Wiley (2003).
[2] 伊藤史,「CGM—消費者発信型メディア」, 毎日コミュニケーションズ (2007).
[3] A. L. Montgomery, S. Li, K. Srinivasan and J. C. Liechty, "Modeling Online Browsing and Path Analysis Using Clickstream Data," *Marketing Science*, Vol. 23, No. 4, pp. 579-595 (2003).
[4] 長尾真,「自然言語処理」, 岩波書店 (1996).
[5] G. Salton and M. J. McGill, *Introduction to Modern Information Retrieval*, McGraw-Hill (1983).
[6] 里村卓也, "ECサイトの閲覧・購買行動のモデル分析," 井上哲治, 日本マーケティング・サイエンス学会(編),「Webマーケティングの科学」, 千倉書房 (2007).
[7] 清水聰,「新しい消費者行動」, 千倉書房 (2000).
[8] 総務省,「インターネット白書平成13年度版」(2001).
[9] 総務省,「インターネット白書平成19年度版」(2007).
[10] 総務省情報通信政策局情報通信経済室,「我が国の社会生活におけるICT利用に関する調査」(2007).
[11] 田村直樹, "インターネットマーケティングの基礎と現状," オペレーションズ・リサーチ, Vol. 51, No. 12, pp. 723-728 (2006).
[12] 山西健司, "データ・テキストマイニングの最新情報," 応用数理, Vol. 12, No. 4, pp. 341-356 (2002).
[13] http://www.nec.co.jp/press/ja/0707/0201.html (NECプレスリリース: 企業向けブログ情報分析サービスを開始)
[14] http://biz.kizasi.jp/pdf/kizasiRS.pdf (きざしレポーティングサービス)
[15] http://pr.yahoo.co.jp/release/2007/0626a.html (Yahoo! JAPANプレスリリース: 行動ターゲティング広告のサービス拡充について)
[16] http://www.cci.co.jip/ir/press/pdf/press070606.pdf (サイバー・コミュニケーションプレスリリース:「ハイブリッドターゲティング」広告サービスの提供について)