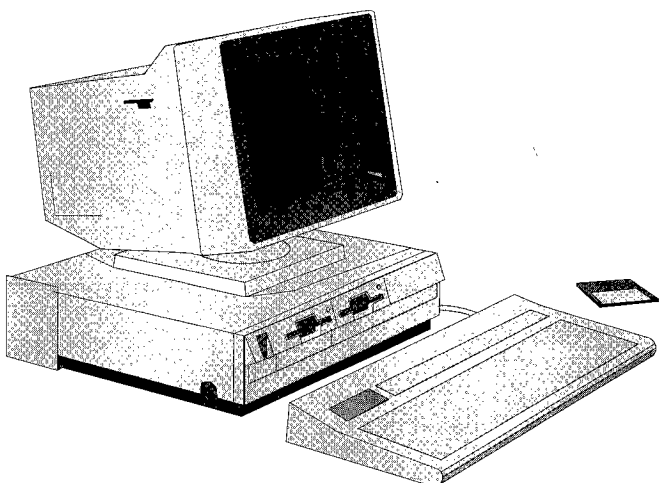


WWW上の 広告における ターゲティング手法

ラングハインリッヒ・マーク
神場知成
中村篤祥

NEC C&Cメディア研究所



急成長するWWW広告

1994年頃であったが、World-Wide Web (WWW)が急速に広がり始めた当初は、WWW上でどういうビジネスが開けるかという見通しはほとんどなかったように思う。その広がり原動力は、情報提供者側から見てサイト立ち上げ・運用が簡単であるということと、利用者側から見て使いやすく面白い、ということであって、ビジネス的に見れば、「儲かるのは、企業のホームページ作成を請け負うデザイン会社だけ」ともいわれていた。

それから約5年、状況は一変した。いわゆるポータルサイト(集客力のあるサービスにより、利用者アクセスが集中するサイト)をめぐる企業の提携・買収・合併など、WWWはビジネス革命の台風の目となった。今や、社員数百人のポータル系米国ベンチャー企業の株価時価総額が、日本の大企業のそれをはるかに上回るという事態である。ここで、ポータルサイトの主役は、検索サイト、コミュニティサイトなど時々刻々と入れ替わっているが、それらポータルサイトに共通して最大の収益源となっているのは、WWW広告である。WWWの場合はサイト相互をリンクすることでアクセス数に相乗効果があり、また、ページビュー(WWWページに対する利用者アクセス数)に応じて広告費が流れ込むという構図ができあがっている。前述のような企業提携等がすぐに収益に反映するわけである。

インターネット広告費は、統計によれば1998年に米国で約20億ドル、日本で約100億円とのことであるが¹⁾、この数字は今後急増すると筆者は予想している。広告費が増大するというよりも、インターネット広告が電子商取引(EC)と直結することで、企業の広告宣伝費だけでなく、それよりもはるかに大きいといわれる販売促進費がこの市場に流れ込むからである。すでに、「WWWページに広告を1回表示すると広告主が掲載者にX円払う」という広告費モデルではなく、「ページに表示した広告が、そのままオンラインで商品購入に結びつくと、手数料として売上金額の一定割合を払う」という手数料モデルが広がりつつある。

さて、本記事の目的は、このようにビジネスとして急成長しつつあるWWW広告の裏側で、技術的にどういうことが進みつつあるかを紹介することである。技術のポイントは、一言でいうと「ターゲティング」である。テレビや新聞・雑誌のようなマスメディアと異なるWWWの特徴を活かし、利用者1人1人に対する的確なターゲティングをした広告を配信するためにどのような工夫がされているか、以下では、機械学習技術、プライバシー保護技術などの観点を含めて解説する。WWWページに対する限られた数の利用者、あるいはアクセスに対して適切なターゲティングをすることは、広告主にとっても、サ



図-1 バナー広告を表示したページの例

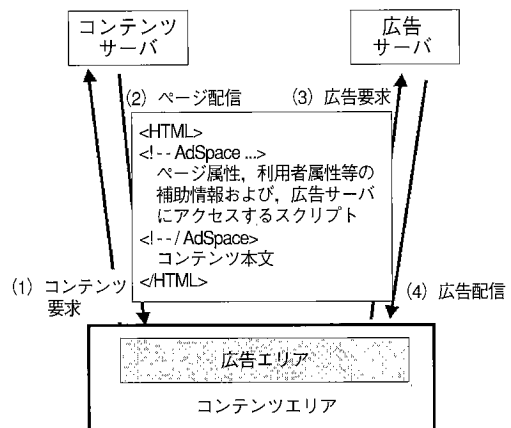


図-2 広告表示のためのサーバ構成

イト運営者にとっても重要なことである。もちろん利用者にとっても、テレビのCMなどと同様、無料あるいは安価なWWW閲覧にとって広告の挿入が必須のものであるならば、まったく興味のない広告に表示スペースを埋められてしまうよりも、興味があり、場合によってはコンテンツそのものよりも有用な情報が得られる可能性のある適切な広告が見たいであろう。なお、一般にインターネット広告というときWWW広告に次いで電子メール広告も重要であるが、本稿では対象をWWW広告に限定していることをあらかじめ了承されたい。

WWW 広告配信方法の基本

WWW 広告にもいくつかの形態がある。最もよく目にするのは、ページの横方向に伸びた旗型の広告であろう。これは通常、バナー広告と呼ばれている(図-1)。その他、1行程度のテキスト広告などもある。以下では特にバナー広告を対象として記述するが、技術的には他の広告でもほぼ同じである。

図-2に、広告を挿入したページを配信するためのサーバ構成の典型例を示す。図中に記したように、ページ表示は以下のプロセスで行われる。

- (1) 利用者がコンテンツを要求(ページアドレス指定)。
- (2) コンテンツページ配信。配信されるHTMLファイルの中には、広告サーバにアクセスするためのスクリプトなどが埋め込まれている。
- (3) スクリプトに基づき、ブラウザ(クライアント)が広告サーバにアクセス。
- (4) 広告が配信され、表示エリアに埋め込まれる。

上記の(2)~(4)は自動的に行われる。ハイパーテキストというWWWの原理からいって、このプロセスは特に変わったものではないが、ここで注目すべきは次の2点である。

- 上記のプロセス(2)において、適切な広告を表示するために必要な属性が広告サーバに引き渡される。
- 広告サーバは、コンテンツサーバとはまったく別のサイトが管理・運営している場合が多い。

ある程度大規模なサイトならば、広告表示のためには、表示スケジュール設定、表示実績管理、実績レポート機能などさまざまな機能が必要となるので、上記のように独立した広告サーバが設けられている場合が多く、最近では広告サーバ運営に特化した会社が多数のWWWサイトへの広告配信を請け負っているケースも多い。そして、広告サーバの特に重要な機能が「ターゲティング」であり、そのために必要な情報がコンテンツサーバから広告サーバに引き渡されるわけである(広告サーバ独自に取得する情報もある)。

以下では、まず広告の契約形態について述べた後、広告ターゲティングが現在どのような手法で行われているか、また、今後どのような手法が考えられているかについて述べる。

広告契約の形態

まずここで、簡単に広告契約の形態について述べる。一見ビジネス上の手続きだけの問題のようだが、ターゲティングの技術を論じる上でもこの問題を避けることはできない。広告配信をする際には、広告配信者と広告主との間でさまざまな契約の形態があり、ターゲティングもその制約の下で行わなければならないからである。

典型的な契約形態として次のようなものがある。

- **表示回数保証**：広告のページビューを保証する方法。たとえばバナー広告を10万回表示することを保証するなど。期間を限定して一定回数以上の表示を保証する場合や、契約回数が達成されるまで表示する場合などがある。

ターゲティング条件	例	典型的利用技術
ブラウザ, OS 等	利用者ブラウザ等に応じて異なる広告表示	一般のブラウザのHTTPメッセージで判定可能
利用者ドメイン	企業ドメイン経由とプロバイダ経由のアクセスで異なる広告を表示	サーバ側で時刻管理
表示時刻	特定の広告を深夜のみ表示	Cookieにより同一ブラウザからのアクセスを判定
フリークエンシ	同一利用者(同一ブラウザ)からのアクセスに一定回数以上同じ広告を見せないように制限	CGI アクセスの引数からキーワードを判定
検索キーワード	特定キーワードで検索をかけた利用者へのみ、特定の広告を表示	

表-1 広告ターゲティング手法の例

```
GET http://www.geocities.com/TimesSquare/
HTTP/1.0 : 現在アクセスしようとしているサイト
Accept: image/gif, image/x-bitmap, image/jpeg,
image/pjpeg, image/png, */* : 受信可能なファイル形式
Accept-Charset: iso-8859-1,*utf-8 : 文字コード
Accept-Language: ja, en : 言語
Host: www.geocities.com
Referer: http://dir.yahoo.com/Government/Chat/
: このサイトにたどりつく直前のサイト
User-Agent: Mozilla/4.0 (compatible; MSIE 5.0;
Windows NT); : ブラウザ側情報 (IE5.0, WindowsNT)
Cookie: GeoId=20219605885185780914 : Cookie情報
Proxy-Connection: Keep-Alive
```

図-3 IE4.0がHTTPヘッダとしてサーバに送付する情報の例(下線部は説明)

- **クリック回数保証**：広告がクリックされる回数を保証する方法。ユーザが広告をクリックした場合は、通常そこにリンクされているサイト(つまり広告主のページ)にジャンプするので、表示回数保証と比較すると広告効果が明示的に分かる特徴がある。

上記以外にも、特定キーワードの検索結果表示ページや、特定のページの広告表示枠を買い取ってしまう「レンタルキーワード」、「ページ固定」などがある。

典型的なターゲティング手法

さて、広告サーバは、多かれ少なかれターゲティング機能を持っている。つまり、WWWページ内への広告表示に際し、単に決められたページに決められた広告を表示するというのではなく、利用者あるいはその他の条件に応じて選択的に広告を表示することである。ターゲティングの例を表-1に示す。

ここで問題になるのは、ターゲティングをするために、ユーザに関するどのような情報がサーバから取得可能かということである。前章で述べたように、このような情報はコンテンツサーバが取得して広告サーバに引き渡す場合も、広告サーバが独自に取得する場合もあるが、方法としては同じなのでここでは区別しない。サーバ側で簡単に取得可能な情報としては、以下のようなものがある。

- ブラウザ種類, OS
- 利用者ドメイン
- 検索キーワード
- フリークエンシ

ブラウザ種類, OSなどは、広く使われているブラウザ(Internet Explorer, Netscape Navigator等)であれば自動的にサーバに送るので、サーバ側ではそれを解釈すればよい。たとえば、ブラウザの広告(たとえばInternet Explorer)を打つ場合に、対抗するブラウザ(たとえば

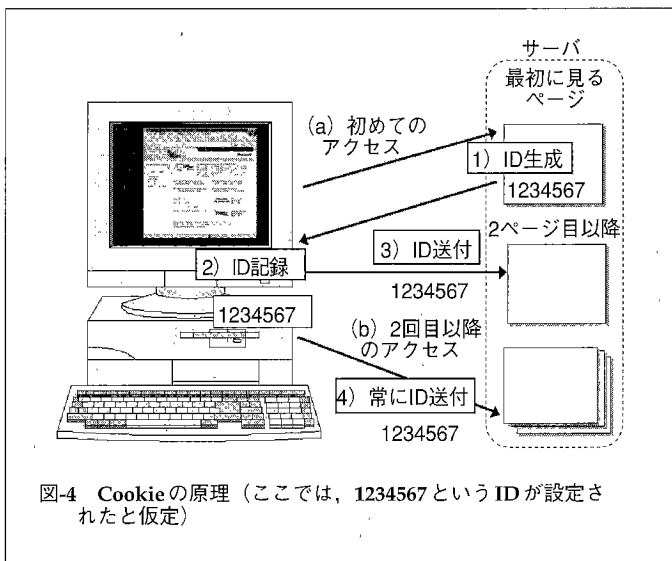
Netscape Navigator)を使っている人だけをターゲットにする場合などがあるだろう。図-3に、IE4.0がサーバに対して送付するHTTPヘッダの情報を示す。

利用者ドメイン(地域)も上記と同様、サーバ側で容易に取得可能である。ただし、利用者からのアクセスがプロキシ経由である場合、そのプロキシからアクセスをしたように見えるので利用者自身がどのような地域からアクセスしたかという正確な情報は消えてしまう。それでも、大雑把なドメインは分かるので、たとえば日本からプロキシ経由で海外のサイトにアクセスした際に日本語の広告が表示される、というようなことは経験された方も多であろう。

検索キーワードというのも容易に取得可能な情報の典型例である。つまり、たとえば「自動車」というキーワードで検索があった場合に、単にその検索結果を表示するだけでなく自動車の広告を表示するというように利用できる。冒頭でも述べたように検索サイトというのは多数のアクセスが集中するポータルサイトの1つであるが、ほとんどのサイトにおいてこのようなターゲティングが行われている。

フリークエンシというのは、ここでは広告をユーザが見た回数を意味している。これは、Cookieを用いたターゲティングとして分かりやすい例なのでここで挙げる。Cookieというのは、すでにご存知の方も多いと思うが、「同一ブラウザからのアクセスを同定する」ことができる。つまり、あるブラウザからあるサイトに初めてアクセスがあった場合に、固有のIDをそのブラウザに割り当て(ブラウザが動作するマシンのディスクに対し、サーバからIDを書き込み)、次回以降そのブラウザから同じサーバにアクセスする場合には、そのIDが自動的に送付される仕組みである(図-4)。

フリークエンシコントロールと呼ばれる手法では、このCookieの仕組みを使い、同一ブラウザに対して同一広



告が一定回数以上表示されることを抑制する。これは、「同じ広告を一定回数以上表示すると、その広告に対するクリック率(ユーザがクリックする確率)は急速に減少する(DoubleClick社による)」というデータに基づいている。

以上、いくつかの広告ターゲティング手法を示した。実際の広告配信においては、これらを組み合わせて利用することになる。

自動ターゲティング

前章で紹介したように、広告のターゲティングにはさまざまな手法を利用することができるが、実際にこれらの組合せによる最適なターゲティングをすべて手作業で行うことはきわめて困難な作業である。なるべく効果が大きくなるようにパラメータの値を決めようとしても、最初は勘に頼らざるを得ない。そして効果を観測しながら、徐々にチューニングすることになる。広告主をこのような面倒な作業から解放するために、最近では、効果を観測しながら自動的にターゲティングを行うシステムが開発され始めた。ここでは、そのようなシステムの例として、DoubleClick社 (<http://www.doubleclick.net/>)の DART (Dynamic Advertising Reporting and Targeting)、Aptex Software社 (<http://www.aptex.com/>)の Select Cast、さらに筆者らが開発している ADWIZの手法について紹介する。

(1) DART

このシステムは、ClickBoostersと呼ぶ技術によってクリック率の最大化を試みる。ClickBoostersは、広告主が設定したパラメータの値を満たす各ページに対しバナーのクリック率を監視し、クリック率が高いページを識別し、それらのページに重点的にバナーを掲載する。DoubleClick社によれば、この技術によりクリック率を10～40%上げることができる。

(2) SelectCast

このシステムでは、テキストに対するコンテンツマイニング技術を個人嗜好抽出に適用することで、クリック率を最大化する。まずCookieを使ってユーザの振る舞い(クリック、検索、ページビュー等)を観察し、それを基に各ユーザのプロファイルを作成する。次にユーザプロファイルは同じような興味を持つユーザのグループに分類され、バナー広告表示時点では、ユーザが属するグループでクリック率が高いものが選択される。ここで用いられているコンテンツマイニング技術の特徴は、コンテキストベクトル(Context Vector)による表現をニューラルネットワークで学習することである。コンテキストベクトルの学習に際しては、ユーザの振る舞いのシーケンスから、類似性を測るのに適した空間軸を抽出し、その軸を使って表現することでコンパクトなベクトルを作成する。コンテキストベクトルの空間軸抽出として、ニューラルネットワークを使う方法のほか、特異値分解(SVD, Singular Value Decomposition)を使う方法もある。

SelectCastはInfoseekのUltramatchという最適バナー選択技術にも利用されている。Aptex社によれば、この技術によりキーワードやトピック連動のバナー選択法に比べ、最大50%クリック率を上げることができるとのことである。

(3) ADWIZ

現在、筆者等が開発中の広告サーバADWIZも、学習機能によりクリック率の向上を行うが、次の2つの特徴を持っている²⁾。

- Cookieを利用しない。これは、後述するようにプライバシー上の問題が少ない。
- 表示回数やクリック回数に関する制約条件を考慮しながらクリック率最大化を行う。

1番目の点は、個人と広告との相性を最適化するのではなく、広告を表示するページと広告との相性や、(検索サイトにおける広告の場合)検索キーワードと広告との相性を最適化することを意味している。2番目の点は、過去の実績から予測したページの表示回数(あるいはキーワードの入力回数)と各バナーの契約掲載回数を制約条件とした上でのクリック率最大化問題を解くことで実現している。

以下では、ADWIZの自動ターゲティング手法について少し詳しく説明する。

• 輸送問題への還元

ADWIZでは、契約を満たしながら契約期間中バランスよくバナーを掲載するために、表示スケジューリングの問題を次のように定式化した。

【問題】 m 個のバナー $A(1), \dots, A(m)$ を、 n 個のページ $P(1), \dots, P(n)$ に掲載とする。契約回数から計算した各バナーの1日あたりの掲載回数を $a(1), \dots, a(m)$ 、各ページの1日あたりの表示回数を $p(1), \dots, p(n)$ とする。また、ページ $P(j)$ に広告 $A(i)$ を掲載する回数を $d(i, j)$ とする。ペー

バナー広告クリック数最大化問題

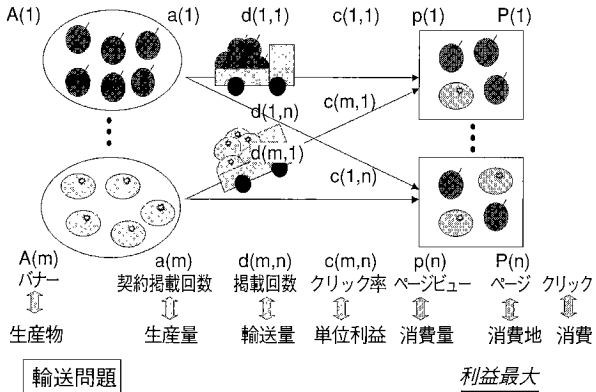


図-5 クリック率最大化問題とその輸送問題との関係

ジP(j)に広告A(i)を掲載した場合のクリック率c(i, j)がすべてのi, jに対して分かっているときに、総クリック数が最大になるように各i, jに対する掲載回数d(i, j)を求めよ(ただし、上記において $1 \leq i \leq m, 1 \leq j \leq n$ である)。

この問題は、生産量a(i)の生産物A(i)を消費量p(j)の消費地P(j)に輸送するときの単位当たりの利益がc(i, j)の場合、全体の利益が最大になるように各A(i)に対するP(j)への輸送量を求める問題と同値である(図-5参照)。これはHitchcock型輸送問題と呼ばれる線形計画法の問題であり、効率的なシンプレックス法による解法が知られている3)。

クリック率推定誤差の問題

上の定式化では、各ページの1日あたりのアクセス数と、「ページ-バナー」のペアに対するクリック率(つまり、あるページに、あるバナーを表示した時のクリック率)は分かっているものと仮定した。しかし、実際にはこれらの値は過去の実績からダイナミックに推定しなければならない。バナーA(i)をページP(j)に掲載した場合のクリック率c(i, j)が実際には高いにもかかわらず、立ち上がりで運悪くほとんどクリックされなかった場合、クリック率が低いと推定してページP(j)にバナーA(i)を掲載する確率d(i, j)を0にしてしまうと、いつまでたってもc(i, j)が実際には大きいということが学習できず、全体のクリック率を最大化することができない。また、逆に実際にクリック率c(i, j)が低いバナーA(i)とページP(j)の組合せに対し、たまたま運が悪かっただけかもしれないと考えてd(i, j)をずっと大きくしておく、全体のクリック率はいつまでたっても上がらないことになる。これは機械学習分野では、「知識の獲得と利用のトレードオフ」と呼ばれ、盛んに研究されている問題である。

ADWIZでは、(1)「バナーA(i)-ページP(j)」に対するクリック率c(i, j)の推定値の信頼度が低い場合、その組

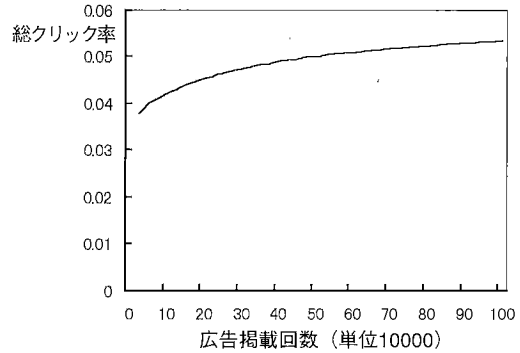


図-6 ADWIZによる、時間的な総クリック率向上

合せに対する表示確率d(i, j)が小さくなりすぎないように制御する、(2)クリック率の推定値としてGittins Index(不確定な将来の事象ほど重みを下げて計算したクリック率の期待値)を使う、(3)線形計画問題を解く際に精度のやや悪い近似値を使う、などの方法によりこの問題に対処している。

図-6に、「あるページに、ある広告を表示した時にユーザがクリックする」という事象および「ページの表示」という事象がいずれも確率的に決まると仮定した場合の、ADWIZによる総クリック率向上効果シミュレーション結果を示す。

プライバシーの問題

ADWIZの項でも述べたが、今後広告ターゲティングの問題を考えるにあたり、個人のプライバシーを重視するという側面は避けて通れないであろう。筆者は、現段階でターゲティングにおける個人情報利用は次の5段階に分けて考えるのが適切であると考えている。

- 段階1：まったく利用せず
- 段階2：ブラウザから送付される属性を、その場限りで利用
- 段階3：サイト内でのユーザ行動を追跡して利用(セッション内だけのCookieの利用)
- 段階4：同一ユーザを永続的に追跡(永続的なCookieの利用)
- 段階5：ユーザが自ら登録した個人情報、興味情報等を利用

段階1の場合、広告表示ページと広告との相性だけを手がかりにターゲティングすることになる。つまり、たとえばドライブ情報のWWWページに自動車の広告を表示する、といったことがこれに相当する。

段階2の場合、ブラウザタイプ、OS、さらに検索サ

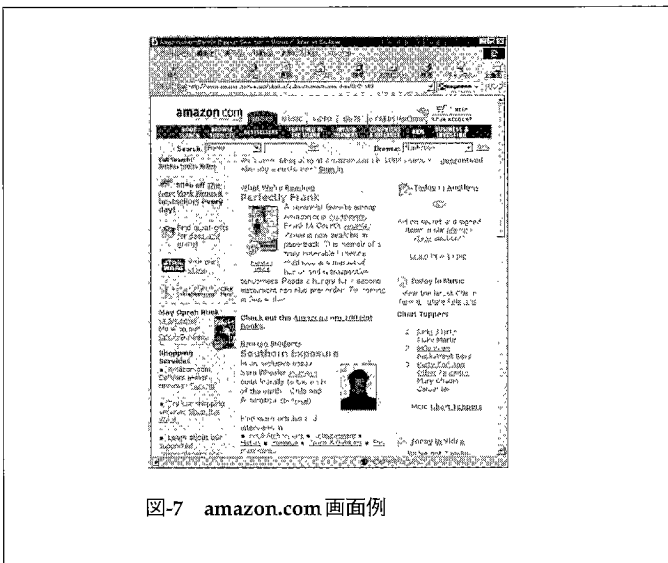


図-7 amazon.com 画面例

イトにおける検索キーワードなどを利用したターゲティングができる。前述のADWIZはここまでを利用している。

段階3の場合は、Cookieを利用するという点で、プライバシーの観点からは一段踏み込んだ形となるので、やや注意が必要である。しかし、これによりたとえば「ページAとページBを見たユーザに、ある広告を表示する」というような制御や、前述のフリークエンシコントロールなどができる。

段階4の場合は、サービス提供者側が、ユーザを匿名ではあるが永続的に同定することが可能になるので、ユーザ側にとっての不安は段階3よりも増大するであろう。極端な話、Cookieによってサイト内の行動をずっと記録されていたユーザが、ある時、同じサイト内の何かのサービスで自分の名前等を登録したとすれば、そのユーザのサイト内行動はすべてサーバ側に分かってしまうわけである。

段階5は逆に、ユーザが同意の上で個人情報をサーバ側に登録しているので、問題はやや少ないであろう。このような仕組みは、サービス提供者と顧客とが同意の上で密な関係を築く場合などに見られる。たとえばBroad Vision社(<http://www.broadvision.com/>)のワントゥワンマーケティングサーバでは、顧客データベースと照合することで、特定個人に向けて広告ターゲティングをする仕組みがある。

なお、WWWサイトによるユーザ情報取得の仕組みに関しては、WWWに関する標準勧告を行う団体であるWWW Consortium (W3C)が1997年6月からPlatform for Privacy Preferences (P3P) Projectとワーキンググループを発足して検討を行っている。これは、WWWサイトがユーザ情報を取得する場合には、その内容や目的を明示してユーザの同意を得なければいけないという原則を実現するもので、サイトからの個人情報問合せ方法(Proposal)、クライアント側(ブラウザ)における個人情報開示原則(Preference)の記述法などが定められてい

る。1999年5月末現在、P3P仕様は最終勧告に至っていないが、勧告になった際には、前述の段階3～5における個人情報取得(Cookie利用も含む)は、P3Pの仕組みに則って行われることになるであろう。W3Cに関しては<http://www.w3.org/>、P3Pに関しては<http://www.w3.org/p3p/>を参照のこと。

今後の発展— EC との融合—

さて、今後の広告ターゲティングであるが、広告という枠内に閉じて表示回数やクリック数をカウントするという流れから、ECとの融合という流れへと急速に移るだろう。冒頭でも述べたように、特に米国ではビジネスモデルも広告料モデルから手数料モデルへと移り始めている。

この流れは、たとえばすでにWWWの枠を越えて知られるところとなったサイト、amazon.com(<http://www.amazon.com/>)(図-7)でも感じることができる。そこでは、たとえば本を購入すればその配達のためにユーザ登録をすることになるので、次回以降のアクセスでは自動的にお勧めの本などが表示される。お勧めの手がかりとしては、いわゆるソーシャルフィルタリング(興味が共通する人のグループを見つけ、そのグループの中で人気のある情報をシステムが相互に教える)や、匿名で投稿される書評が用いられている。勧められた本を利用者がクリックすれば直ちに購入手続きにつながる。ここで表示されている情報は狭義のWWW広告とは異なり、いわば書店で店員がアドバイスをしているようなものだが、今後はこれらの境目はほとんどなくなっていくだろう。インターネットの外ならば、たとえば新聞に掲載された広告、書評、店頭での店員のお勧めなどは明確に区別されるが、WWWの世界ではこれらが融合し、ECを促す手段となっていくと予想できる。

このような意味で、WWW広告はまだ端緒にすぎないばかりである。本稿で述べたようなターゲティング手法と、ECとの融合という流れによって、これまでの広告とは異なるモデルに基づく巨大ビジネスに成長していくであろう。

謝辞 日頃から本稿と関連する活動全般を指導してくださっているC&Cメディア研究所阪田史郎統括マネージャ、古関義幸研究部長に感謝いたします。特に広告サーバADWIZは、本稿の筆者に加え、古関義幸、安倍直樹、下島健彦、広屋修一氏の貢献によるところが大であります。

参考文献

- 1)日経マルチメディア99年4月号, pp.52-67 (1994).
- 2)Langheinrich, M., Nakamura, A., Abe, N., Kamba, T. and Koseki, Y.: Unintrusive Customization Techniques for Web Advertising, 8th International WWW Conference (1999).
- 3)古林: 線形計画法入門, 産業図書 (1980).

(平成11年6月2日受付)