

# 情報掲載サイトのコンテンツ順序決定評価モデル

○森 正人 倉橋 節也 (筑波大学)

## An Evaluation Model for Order-decision Methods of Contents in Information Publication Sites

\* M. Mori and S. Kurahashi (University of Tsukuba)

**Abstract**— Web sites called an information publication site provide information to people looking for a job, a home and so forth. Companies that want to display their products or services on the sites pay advertisement rates for administration companies of the sites, such as a job offer site and a real estate site.

In this study, we consider how to decide the display order of products for publication on list pages of contents in information publication sites. The order should be desirable not only for advertisers, but likewise for users.

First, we show a result of a regression analysis between display order and page views. Second, we define User Popularity Degree using the analysis. Finally, we employ agent-based modelling to simulate the websites in which web contents are defined as agents bidding advertisement rates for getting the optimised or effective display orders autonomously.

**Key Words:** information publication site, display order, User Popularity Degree, agent-based modelling, bid

## 1 はじめに

### 1.1 情報掲載サイトとは

情報掲載サイトと呼ばれるWebサイトがある。求人や不動産サイトが代表的である。採用意欲のある企業や不動産会社等の情報コンテンツを持つ企業が、情報集合サイト運営企業に広告料金を支払い、そのコンテンツを掲載するものである。本研究は、情報掲載サイトのコンテンツ一覧ページにおいて、ユーザーと広告主にとって望ましく、サイト運営社にもメリットのあるコンテンツ順序決定方法を探求するものである。

これらのサイトは、ある一定期間の掲載を申し込む形で商品を提供している。1ヶ月や3ヶ月などの契約期間だけ、広告主の情報コンテンツが掲載される。いわば広告掲載である。この商品はパッケージ化されており、表示できる最大文字数や写真数、アピール項目の設定数に、広告主による差はない。価格も同一である。ただし、プレミアムな高額商品と簡易な低額商品という差はあり、高額商品の方が多くの情報や写真を掲載できる。バナー掲載等のオプション商品もあるが、オプションに触れることは本筋ではないため省略する。

サイト運営社は多数の広告主から集めた掲載料金をもとに大規模なマーケティング活動を行い、ユーザーを集めている。各広告主が個別にマーケティング予算を費やすよりも費用対効果のよい集客を提供しているのである。このコンテンツ掲載広告型のビジネスは、業界では情報掲載ビジネスと呼ばれている。そして、そのWebサイトを情報掲載サイトと呼ぶ。

### 1.2 不動産情報サイトの特徴

情報掲載サイトのビジネスモデルは、求人情報サイトも不動産情報サイトも基本的に同じである。以下では不動産情報サイトを考えることにする。

居住物件を探しているユーザーがWeb検索やインターネット広告等の何らかの手法で不動産情報サイトを見つけ、アクセスする。そして自分が希望する物件の住所や価格、間取り等により絞り込みを行う。その状態で検索をかけると、絞り込んだ条件に合致する物件一覧が表示される。一覧ページには物件の名称や写真、価格、最寄り駅、詳細情報へのリンク等が表示されている。間取り図や設備等の詳細な情報を確認したい物

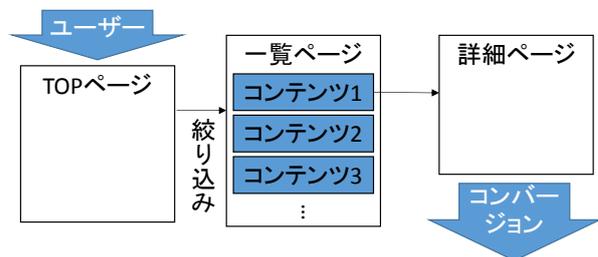


Fig. 1: Real estate information site

件があれば、そのリンクをクリックする。詳細情報のページを確認した後、資料請求や内覧したい場合は、個人情報を送信するページへと遷移する。この個人情報の送信が完了することをコンバージョンと言う (Fig. 1参照)。広告主が最も欲しいのは、このコンバージョンである。次に求めるのは、ユーザーがコンテンツを閲覧した数を増やすことによる物件認知度向上である。

前述したように、情報掲載ビジネスは広告掲載型である。ここ数年はコンバージョンに対して課金する成果課金モデルのサイトも出現しているが、本研究は広告掲載型サイトについてのものであることを明記する。

### 1.3 掲載順序

Web サイトではページの上部の方が下部よりも面積当たりの価値が高い。一般的な Web サイトを考えて欲しい。ユーザーに対して最初に表示されるのは、ページの最上部である。画面をスクロールすると、徐々にページ下部が表示される。そのため、画面の上の方が下部よりもユーザーの目に留まりやすい。これは Web サイトを管理運営する者にとっては共通認識であるが、一般の方でも想像に易いのではないだろうか。

Fig. 1にあるように、情報掲載サイトのコンバージョンは、詳細ページを経なければ達成されない。また、特別な機能を除けば、詳細ページへは一覧ページを経る必要がある。つまり、一覧ページはコンバージョンへとつながる重要な経路となっている。その一覧ページにおいても、ページ上部に表示される方が、下部に表示されるよりもユーザーの目に留まりやすい。また、コンテンツの数が多くて1ページに収まらない場合は複数ページにわたって一覧が続くことになるが、1ページ目の方が2ページ目以降よりもユーザーの目に留ま

りやすい。つまり、掲載位置が上部であり、掲載ページが前である方が、広告効果は高いということになる。

当然であるが、広告主はより上部に、より前のページに掲載して欲しい。だが、順序というものは1から最後までついてしまうから、全部が最上位というわけにはいかない。そのため、1位のコンテンツと最下位のコンテンツでは、その広告効果には大きな差が出る。同一価格の同一商品を購入した広告主からすれば、幸運に上位掲載されて喜ぶこともあれば不満を持つこともある。この運任せの広告効果をなるべく平等にするため、サイト運営側は掲載順序の工夫をしている。

#### 1.4 情報掲載サイトのコンテンツ掲載順序

本節で挙げる掲載順序は初期設定の状態である。ユーザーによる並べ替え機能を使わない状態であることに注意して欲しい。

情報掲載サイトの掲載順序において、最もよく見られるのは新着順と呼ばれるものである。掲載を開始した日が直近のものが上位に並び、以前からあるものは下位に沈んでいく。例えば、今月から掲載されたものの方が先月から掲載され続けているものよりも上位に掲載される。どのコンテンツも新着である期間はあるため、上位表示される機会を均等にするという面では公平である。しかし、同じ時期に掲載開始したコンテンツ同士では、上位と下位の不公平感は解消されない。また、ユーザーの希望条件にどれくらい近いかどうかは考慮されていない。

新着順と並んでよく見られるのが、商品価格順である。1.1 節で述べたように商品はパッケージ化されており、高額商品か低額商品かという価格差と掲載期間しか違いはない。この高額商品を上位に掲載し、低額商品は下位に掲載するというのが、価格順の掲載順序である。広告主は経済的な理由から納得できる順序である。だが価格区分が少ない以上、多くの広告主が同じ価格であり、運任せである状況には変わらない。また、ユーザーの視点が欠けている点は新着順と変わらない。

おすすめ順という掲載順序もある。これはサイト運営側が推薦する順であり、恣意的な要素が大きい。ブランド価値が高いコンテンツを上位に並べて華やかさを出すサイトがあるかもしれない。ロジック非公開の機械学習を行っていることもある。推薦順序を決定するロジックが公開されていない以上、評価は難しい。

最近では減っているが、五十音順を採用しているサイトもある。広告主の企業名や物件名等による五十音である。当然ながら不満を持つ広告主は多い上、ユーザー視点も入っていない。

ランダムに掲載順序を入れ替えているサイトも存在しているだろう。だが、広告主の広告効果が運任せであることは変わらないし、ユーザー視点も欠けていると言わざるをえない。

このように、情報掲載サイトのコンテンツ一覧ページの掲載順序は新着順や商品価格順であることが多く、広告主やユーザーにとって必ずしも望ましいとは言えない。本研究ではABM (Agent-Based Model) によるシミュレーションを行い、この順序決定方法を考察する。コンテンツをエージェントとし、エージェントが自動入札を行うことで掲載順序を決定する。コンテンツのユーザー人気度や類似度も考慮し、ユーザーと広告主に望ましい順序決定方法を見いだすことを目的とする。

## 2 先行研究

掲載順序が重要だとされるサービスで広く知られているもののひとつに、Web 検索サービスがある。日本においては、Yahoo! JAPAN や Google の利用者が多い。Search Engine Optimization (SEO) と呼ばれる検索エンジン最適化とは、検索結果一覧ページにおける上位掲載を目的とする施策である。それは、上位表示された方がユーザーの目に留まりやすく、クリックされる確率が高いからである。

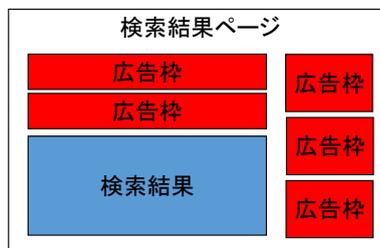


Fig. 2: Search result page

この検索結果一覧ページには、広告表示枠が存在することが多い (Fig. 2 参照)。これは検索連動広告と呼ばれ、広告主がキーワードに対して入札を行うものである。ユーザーが検索すると、その検索ワードに入札を行った広告が表示され、クリックされると広告主に対して課金される。検索連動広告は複数表示される仕組みになっており、その表示順序は入札価格やクリック率 (CTR) によって決定されると言われている。この検索連動広告の入札においても、広告主は上位掲載を求めている。櫻井ら<sup>1)</sup>は、検索連動広告のアルゴリズムについての研究を行っている。通常の第二価格秘密入札では正直入札が最良の策であるが、検索連動広告に使用されている一般化セカンダリプライスオークション (GSP) では、過少申告によって期待効用が増加する可能性があることを示している。

入札によって表示順序を決定する手法は、別の Web サービスでも使用されている。AdNetwork と呼ばれる広告配信ネットワークがその一例である。AdNetwork は、広告掲載によって収益を得たい Web サイトを多数集めてネットワークと化し、ネットワーク全体を広告表示の対象とする。広告を出稿する側は AdNetwork 運営会社と契約すれば、個別サイトに広告掲載の申込を行わなくても、ネットワーク全体に対する広告掲載が可能となる (Fig. 3 参照)。

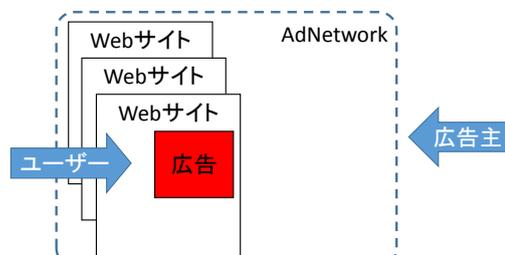


Fig. 3: AdNetwork

Mookerjee ら<sup>2)</sup>は、米国の AdNetwork 運営企業である Chitika 社のサービスについて分析している。Chitika 社は 10 万以上の Web サイトからなる巨大ネットワー

クを確立しており、この広告はクリック課金である。Chitika 社は、サイト運営社が自サイトに掲載する広告の平均クリック率 (CTR) の閾値を設定し、それを下回らないよう広告配信するサービスを提供した。クリック課金の広告の CTR が低いということは、その広告が生む収益は少ないということである。また、ユーザーが興味ない広告にスペースを割くことは、媒体価値の毀損である。高い平均 CTR を実現するためには、その広告に興味がある人に対してしか広告を表示しないことになってしまう。そうすると表示回数が少なくなり、クリック数も少なくなる。収益も小さくなる。

このサービスは、収益の責任を AdNetwork 運営側からサイト運営側へと移した。多くのユーザーを持つ有名サイトよりも、小さなロングテールの合計ユーザー数の方が多現在のインターネット環境において、個別サイトへの広告最適化は困難であり、サイト運営側が選択するこのモデルは非常に興味深い。

表示課金型の AdNetwork において、入札戦略についての研究を行ったのが Ghosh ら<sup>3)</sup>である。ある入札機会において落札できなかったとき、次回の入札価格を2倍に増加させる戦略、自分が出せる最高価格で入札する戦略、入札価格を指数的に増加させていく戦略を用意した。1回の広告表示に対して支払える単価設定が低いと、全体の広告表示回数に占める獲得目標シェアが高くなるにつれて、2倍増加戦略の落札機会が少なくなることを示した。また、最高価格入札と指数増加戦略では、目標表示単価が低くても獲得単価が高額になり過ぎることを示した。

AdNetwork の環境を ABM (Agent-Based Modeling) で構築し、入札戦略の違いによる結果をシミュレーションしたのは鍋田ら<sup>4)</sup>である。Real-Time Bidding (RTB) というオークション取引において、どのような戦略をとれば、より少ない金額で多くの商品を落札できるのか考察している。RTB とは、ユーザーが Web サイトに訪れた際、そのユーザーのクリック履歴や広告主の業種等の情報を加味して自動入札を行うもので、多くの AdNetwork で採用されている。鍋田らは5つの入札戦略を設定し、4種類の実験を行っている。

### 3 データ

本研究の目的は、情報掲載サイトのコンテンツ一覧ページにおいて、ユーザーと広告主にとって望ましく、サイト運営側にもメリットのある掲載順序を研究することである。1章で説明したコンテンツをエージェントとし、ABM のモデルを構築して入札による順序決定手法を検証する。データは以下の2種類用意する。

#### 3.1 物件データ

どのコンテンツがどの掲載順位を得ているかを取得するため、Web スクレイピングのプログラムを作成した。簡易的な Web クローラーである。そのクローラーを一日おきに実行し、ある不動産サイトの東京都の物件情報を取得した。そのデータの一部が Fig. 4 である。物件 ID、物件名称、住所、最寄り駅、戸数、価格、間取り等を取得している。また、クローラーはコンテンツ掲載順序が上のものから順番に取得しているため、このデータは掲載順位順となっている。

#### 3.2 ページビュー (PV) データ

次に、ページビュー (PV) 数のサンプルデータを用意した。物件 ID 単位で日次の数字である。

98945	ウェリス有	東京都港	東京外	ロ	57戸	-	4億1,456.43	LDK183	1戸		
99022	サンクレイ	東京都江	総武線	上	26戸	第1期	2,998万円	1LDK+S	6戸		
89460	G/CLASS	東京都豊	山手線	上	43戸	(管理1期)	2次1,498.8万円	3LDK+4L	6戸5戸		
87951	パークホ	東京都大	東急池上	上	51戸	第1期	5次16,088万円	3LDK	69.7未定1戸		
88630	王子飛鳥	東京都北	京浜東北	上	285戸	(55)	レジデンス	4,238万円	2LDK~4L	未定120戸	
94782	ヴェレーナ	東京都板	都営三田	上	60戸	(※)	第1期	2次3,698万円	3LDK	65.7	6戸4戸
85087	アデウム	東京都江	京成本線	上	41戸	(管理1期)	2次	2,590万円	1LDK+3L	3戸6戸	
93429	イニシア	東京都文	東京外	ロ	34戸	(うち)	2期	1次	1,459.8万円	2LDK+3L	未定4戸
89196	シティイン	東京都品	京浜東北	上	60戸	第2期	第3	2,440万円	Studio+1L	14戸未定	
91895	イニシア	東京都葛	京成金町	上	48戸	(うち)	第3期	1次	2,900万円	2LDK+S	4戸2戸3戸
81439	セントラル	東京都文	東京外	ロ	50戸	#NAME?	5,780万円	2LDK	54	6未定2戸	
80668	ウィーク	東京都世	小田急小	256戸	(17)	くら	カー	4,920.1万円	3LDK~4L	9戸4戸未	

Fig. 4: Web scraping data

### 3.3 データの加工

3.1節で取得した物件データから、掲載順位を知ることができる。3.2節のPV数データと合体させることで、何番目に掲載されていた日に、どのくらいのPV数を獲得したかを導くことができる。これにより、上位の掲載順位を得ていたにも関わらずPV数が少ない物件は、ユーザーにとって人気がないと言うことができる。

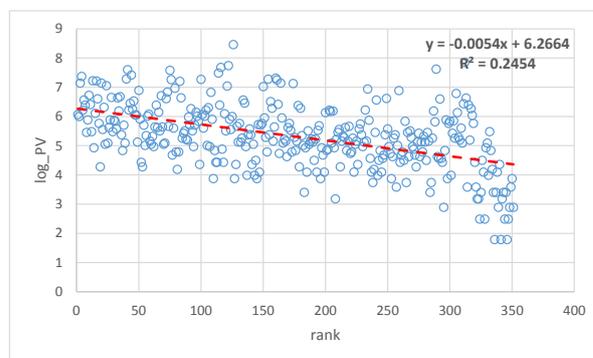


Fig. 5: Scatter plot of PV

Fig. 5は、ある日の掲載順位 (rank) とサンプルPV数をプロットした散布図である。rankが大きくなることは、掲載順位が下がることを意味する。掲載順位が下がるとPV数が減っていることが分かる。回帰直線とその式も Fig. 5 に示している。この直線よりPV数が多い、つまり上部にある物件はユーザー人気度が高いと言えるだろう。

取得したデータには、Fig. 4のように物件データも含まれている。このデータをもとにクラスター分けし、物件をいくつかのカテゴリに分類しておく。この作成は今後取り組む予定である。

### 4 シミュレーションモデル

本研究は、先行研究のモデルとは以下の点で異なる。

まず、財が多数ある点である。AdNetworkの研究で言及されている財はディスプレイ広告である。これはWebサイトの1ページに1つしか表示枠がない。そのため、落札できる広告主は1社しかない。また、検索連動広告も表示枠数に上限がある。それに対して、情報掲載サイトは全ての広告を表示するため、購入された広告の数だけ財が存在する。全ての広告がいずれかの順位を得て、財を獲得することになる。

次に、ユーザー人気度である。AdNetworkは個別ユーザーのクリック履歴から入札価格を判断している。1

人のユーザーのクリック履歴が十分な数の蓄積をされていとはならず、誤差が出やすいと言わざるを得ない。本モデルでは、Fig. 5 で示した回帰直線を用いてユーザー人気度を算出する。例えば rank10 であれば、回帰値は  $y = -0.0054 \times 10 + 6.2664$  で 6.2124 となる。実 PV がこの何倍であるかを求め、それをユーザー人気度とする。この手法を用いることで、コンテンツがユーザー全体にどの程度人気があるのか知ることができる。

本研究で考察する入札モデルは、Fig. 6 のような形で動くものとする。まずユーザーのアクセスが発生する。すると入札リクエストが発生する。そして、コンテンツエージェントが入札価格を決定して入札する。

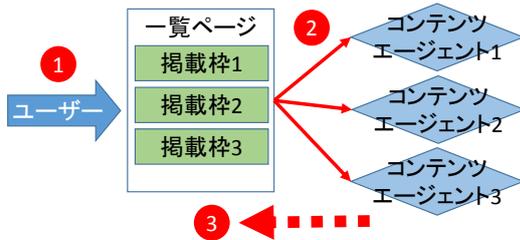


Fig. 6: Flow of the bid

このシミュレーション環境の構築は今後の課題である。現段階では簡易的なモデルを作成した。エージェント数は 10、入札発生回数は 10 とした。その 10 回の入札で使用できる予算はエージェント番号 1 から 6 が 1,000、7 から 10 が 500 とした。高額商品と低額商品をイメージしている。各入札において、入札価格順に掲載順位を与えるものとし、掲載順位 1 位には報酬 10、以降 1 ずつ減らして 10 位が 1 とした。各エージェントの予算と入札方法をまとめたのが Table 1 である。

Table 1: Budget and bid method

No	予算	入札方法
エージェント 1	1,000	予算を 10 回の入札に均等配分
エージェント 2	1,000	均等配分価格に平均 0、標準偏差 1 の正規分布による乱数を乗算
エージェント 3	1,000	均等配分価格に平均 0.5 の指数分布による乱数を乗算
エージェント 4	1,000	均等配分価格に平均 1.0 の指数分布による乱数を乗算
エージェント 5	1,000	均等配分価格に平均 1.5 の指数分布による乱数を乗算
エージェント 6	1,000	ランダムに入札価格を決定
エージェント 7	500	予算を 10 回の入札に均等配分
エージェント 8	500	均等配分価格に平均 0、標準偏差 1 の正規分布による乱数を乗算
エージェント 9	500	均等配分価格に平均 1.0 の指数分布による乱数を乗算
エージェント 10	500	ランダムに入札価格を決定

この設定で 10 回実験を行ったところ、その報酬は Table 2 のようになった。右に列を進むごとに、1 回目の実験、2 回目の実験となり、値は 1 回の実験当たり 10 回の入札を行った累積報酬である。そして、右端に全 10 回の実験の合計を計算している。

これを見ると、いずれも均等価格での入札を行ったエージェント 1 の累積報酬が最大となっており、一番

の勝者ということになる。予算が少ないエージェント 7 から 10 の中でもエージェント 7 の累積報酬が多く、均等価格での入札が最も優れているようだ。

Table 2: Reward list

	1	2	3	4	5	6	7	8	9	10	合計
エージェント 1	77	82	80	82	83	78	80	81	79	79	801
エージェント 2	73	72	75	68	69	73	64	63	74	68	699
エージェント 3	52	65	52	55	58	62	50	73	52	53	572
エージェント 4	64	52	56	61	66	68	65	46	54	61	593
エージェント 5	60	50	64	65	57	64	70	55	59	63	607
エージェント 6	54	55	32	44	41	34	37	52	47	45	441
エージェント 7	53	57	54	54	60	54	63	59	54	59	567
エージェント 8	46	49	56	43	53	51	53	50	47	47	495
エージェント 9	44	58	51	44	49	47	47	46	50	52	488
エージェント 10	49	29	47	38	39	38	35	35	37	38	385

## 5 おわりに

現段階で構築しているのは簡易的なモデルである。今後の実験では 3 章で示したデータを取り込み、エージェントを増やす。入札方法も戦略的なものを構築したい。4 章のモデルにおいては、他者が明確な意図を持たない入札の場合、均等価格入札が有効であることを示した。だが、均等価格より 1 だけ大きい入札を行えば、均等価格入札にも勝利できるはずである。強化学習をする入札戦略の設定も試みたい。

掲載順序の決定方法は、入札価格の大小だけにはしない。コンテンツのユーザー人気度やカテゴリ類似度の要素を組み込む。ユーザーにとって望ましい順序を実現するためには、人気度の高いものを上位表示することになるだろう。だが、これでは人気度の低いコンテンツは下位に沈み続け、その広告主は不満を抱く。人気度が低くても人気コンテンツと類似度が高ければ上位表示する工夫や、予算を考慮した工夫を加えたい。

シミュレーション実験を行った後、本モデルで考察した掲載順序決定方法の結果を実際の掲載順序と比較し、ユーザーと広告主、情報掲載サイト運営社にとってより望ましいものとなっているかの評価を行う。この評価結果が有益なものとなれば、情報掲載ビジネスにおけるコンテンツ順序決定方法に新しい提案をすることができる。

## 参考文献

- 1) 櫻井, 横尾: キーワード広告におけるゲーム理論・オークション理論, 人工知能学会誌, 24(4) (通号 136), 472/480 (2009)
- 2) Mookerjee, Kumar, S.Mookerjee: To Show or not Show: Using User Profiling to Manage Internet Advertisement Campaigns at Chitika, Interfaces, Vol.42, No.5, 449/464 (2012)
- 3) Ghosh, Rubinstein, Vassilvitskii, Zinkevich: Adaptive bidding for display advertising, Proceedings of the 18th international conference on World wide web, ACM, 251/160 (2009)
- 4) 鍋田, 山本, 吉川, 寺野: WEB 広告における実時間オークションのシミュレーションモデル, 計測自動制御学会社会システム部会研究会資料, Vol. 5, 13/22 (2014)