

# ブログ上の広告情報を用いたアフィリエイト分析支援システム

Analysis System of Affiliate Activities over Blog Sites

石井 聡一\*<sup>1</sup>

Soichi Ishii

福原 知宏\*<sup>2</sup>

Tomohiro Fukuhara

増田 英孝\*<sup>1</sup>

Hidetaka Masuda

中川 裕志\*<sup>3</sup>

Hiroshi Nakagawa

\*<sup>1</sup>東京電機大学未来科学部

School of Science and Technology for Future Life, Tokyo Denki University

\*<sup>2</sup>産業技術総合研究所

Advanced Industrial Science and Technology

\*<sup>3</sup>東京大学 情報基盤センター

Information Technology Center, The University of Tokyo

In this paper, we report analysis results of affiliate marketing on the Web. Recently, affiliate programs such as Rakuten affiliate and Amazon associates program became popular on various web sites including personal blog sites. By using affiliate programs, bloggers and webmasters can easily add advertisements on their blog article and web pages, and people who access to these pages can find products related to contents those web pages. Our aim to find the state of affiliate marketing on the web; how many people incorporate affiliate programs into their web sites, what kind of products they sell and how much do they sell those products.

## 1. はじめに

近年、ブログやホームページ上に企業の広告を掲載する成功報酬型広告(アフィリエイトプログラム)が盛んに行われている。個人が手軽に Web 上で紹介したい商品の情報を掲載できるようになり、今までのようなメーカや販売店からの商品情報だけでなく、利用者の商品レビューなどを簡単に得られるようになった。一方で、アフィリエイトプログラムにより、個人の Web サイトまでも広告媒体として広く認知され、広告収入を目的としたスパム(アフィリエイトスパム)が問題となっている。

我々は、アフィリエイトスパムや人気アフィリエイトサイトの分析を行い、どのような特徴があるのかを調べ、その結果をアフィリエイトサイト開設時の参考にする事や、商品レビューを書くアフィリエイトを推薦するシステムの開発を目的とする。本論文ではアフィリエイト利用者(アフィリエイト)の広告掲載活動分析を目的として、アフィリエイト識別番号(アフィリエイト ID)に着目した分析アプローチを提案する。また、先行研究[原 09][Wang 07]での Web ページや Web サイトを基準とした分析手法では分析できない複数のブログサービスにまたがったアフィリエイトの分析や、掲載商品を調査するアフィリエイト分析支援システムを開発する。本分析では、アフィリエイト ID 単位での分析となるので、Web サイト単位では特定することが難しかった複数のブログを管理するアフィリエイトの商品傾向や、商品ジャンル、価格帯も調べることが可能となる。

本論文の構成は次の通りである。2. で先行研究の分析手法と我々の提案する分析アプローチを示す。3. でシステムの説明を行い。4. で実際にシステムを利用した分析事例を紹介する。5. で考察を述べ、6. で本論文のまとめと今後の課題について述べる。

## 2. 先行研究の分析手法と本研究の分析手法

### 2.1 先行研究

アフィリエイトスパムのフィルタリングに関して、原[原 09]らの研究では 2 つの分析が行われている。(1) ブログ内に含まれるアフィリエイトリンク数の集計、(2) スпамブログと ASP(Affiliate Service Provider)\*<sup>1</sup>との関係性の調査である。原らの分析結果では、1 つのブログ内に含まれるアフィリエイトリンク数が多いほど、スパムブログ(Splog)である傾向が高く、アフィリエイトブログの半数がスパムブログであると報告している。また、ASP22 社の中でスパムブログ割合が 90%以上の 5 社の存在を報告している。

Wang[Wang 07]らの研究では、スパムサイトから広告主に至るまでのトラフィック分析を行っている。アフィリエイトスパムのネットワークが、第 1 層 Doorway、第 2 層 Redirection Domain、第 3 層 Aggregators(プロバイダ)、第 4 層 Syndicator(ASP)、第 5 層 Advertiser(広告主)の 5 階層に渡っており、Doorway と Advertiser が多く存在し、Redirection Domain と Syndicator が少なく、Aggregators がさらに少数であることを示した。それが 2 つの漏斗を上下逆さまにして繋げたような形となることから Spam Double-Funnel モデルと名付けている。この各階層において頻出するドメインの統計分析を行っている。

齊藤[齊藤 09]らの研究では、ブログ中の商品紹介記事を収集し、アフィリエイトの有無によって、言語表現に異なる特徴が現れるか検討を行っている。この結果、アフィリエイトのある記事では、話題が対象商品に集中しやすく、アフィリエイトのない記事では、対象商品から離れた話題に変遷しやすい。また、アフィリエイトのある記事ほど、書き手は評価軸を複数持ち、強調表現を用いていると報告している。

### 2.2 本研究におけるアプローチ

図 1 に我々が提案する分析手法を示す。本研究では、次に説明するアフィリエイト ID 特定分析に基づき、アフィリエイト空間の分析を行う。アフィリエイト ID 特定分析とは、ASP か

連絡先: 石井 聡一, 東京電機大学未来科学部情報メディア学科,  
〒101-8457 東京都千代田区神田錦町 2-2, 03-5280-3281,  
ishii@csl.im.dendai.ac.jp

\*<sup>1</sup> アフィリエイトプログラム提供会社。広告主と広告掲載主との間にいる中間業者である。

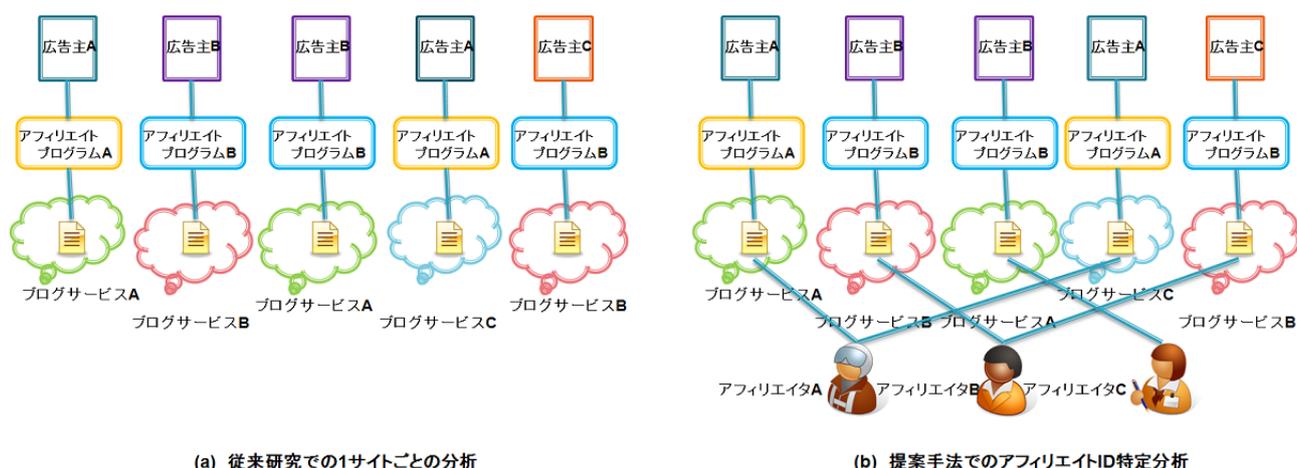


図 1: 先行研究の分析手法とアフィリエイト ID 特定分析手法

ら登録ユーザ各々に付与される独自の ID(アフィリエイト ID) をブログ記事等から抽出し、アフィリエイト ID を単位として分析する分析手法である。

先行研究では、1つのウェブページやウェブサイトを分析対象とし、そのウェブページのドメインや利用するアフィリエイトプログラム、掲載する広告主といった縦の関係のトラフィック分析(ブログサイト毎の独立した分析)を行っていた(図 1(a))。これに対し、我々の提案するアフィリエイト ID 特定分析では、先行研究のトラフィック分析に加え、アフィリエイト記事に含まれるアフィリエイト ID を用いて情報発信者(アフィリエイト)を特定する(図 1(b))。これにより、どのアフィリエイトがどのブログで広告活動を行っているのかという情報を得ることができ、アフィリエイトスパムに見られる同一のアフィリエイト ID を用いた複数のブログサイトにまたがるスパムを特定できる。この結果、従来の分析方法では得られなかった横の関係の分析(アフィリエイト ID からブログサービスが異なるブログサイトであっても、同じユーザのブログとして分析)も可能となる。

我々は、アフィリエイト ID 特定分析に基づき、アフィリエイトスパムの分析を行った[石井 09]。その結果、アフィリエイトが管理するブログサイト数においてスパム、非スパムで顕著な差を示し、スパムでは複数のブログサイトを管理するアフィリエイトが 67.9%、非スパムでは 4.5%であることを報告した。

本論文で提案する分析アプローチでは、1つのアフィリエイト ID を 1 ユーザとみなす。アフィリエイトでは通常、継続的に実績をあげていると紹介料率<sup>\*2</sup>が上がるなどの優遇措置がある[伊藤 07]。このことからユーザは 1 つのアフィリエイトサービスにつき 1 つのアフィリエイト ID を取得している可能性が高いと考えられる。例外として、Amazon アソシエイト(以下 Amazon)では、1 つの ID につき複数のトラッキング ID を利用することができ、複数のアフィリエイト ID の収益を 1 つのアフィリエイト ID にまとめることが可能である。

### 3. アフィリエイト分析支援システム

アフィリエイトの活動内容分析を目的として、アフィリエイト ID 特定分析を用いたアフィリエイト検索を提案する。利用者はアフィリエイト ID, URL, 商品名をクエリとして、ブログサイト、掲載商品を検索することが可能である。この検索により、そのアフィリエイト ID が存在するブログサイト、今まで扱ってきた掲載商品、掲載商品カテゴリ、ジャンルを得ることができる。これによって、アフィリエイトがいつどんな商品をどのように紹介していたのかが分かる。例えば、アフィリエイトスパムが大量に記事を投稿している場合、その記事にあるアフィリエイトは全て同じ商品であるのか、または、ブログサイトによって掲載する商品を変えているのかなどがわかり、アフィリエイトスパムの活動内容の分析が可能である。

#### 3.1 分析データ

分析データとして、ブログサービス運営会社 8 社<sup>\*3</sup>から取得したブログ記事を利用する。提案システムはブログサービスから配信される RSS ファイルを定期的に取得し、そこから新着記事の URL を抽出し、ブログ記事を収集する。クロールは毎日 1 時, 12 時, 23 時台の記事を取得している。また、取得した記事から HTML 内にあるリンク, HTML タグ情報などを用いて、ASP13 社の<sup>\*4</sup>アフィリエイトプログラムの記事数を集計し、その中でも記事数の多かった上位 3 社として、Amazon アソシエイト(以下 Amazon), Google AdSense(以下 Google), 楽天アフィリエイト(以下楽天)のアフィリエイトの選定を行った。今回の分析では、商品分析は Amazon だけとし、商品情報取得のために Amazon を利用する記事から Amazon 独自の商品識別番号 ASIN(Amazon Standard Identification Number)を取得し、AWS(Amazon Web Services)に問い合わせ商品情報を取得する。

#### 3.2 システム構成

図 2 にシステム構成を示す。1 台の計算機を用いて、Ubuntu 9.04, MySQL 5.0.75, Apache 2.2.11 を利用している。提案システムは Web から毎日ブログ記事を収集し、そこからアフィ

\*2 アフィリエイトから商品が購入された場合に、広告掲載主に販売額の何%を支払うかを定める値。

\*3 Ameba, fc2, JUGEM, livedoor, Seesaa, Yahoo!, ココログ, ウェブブログ

\*4 A8.net, AccessTrade, Amazon アソシエイト, DMM, Google AdSense, infocart, infostore, infostyle, LinkShare, Store-mix, TrafficGate, ValueCommerce, 楽天アフィリエイト

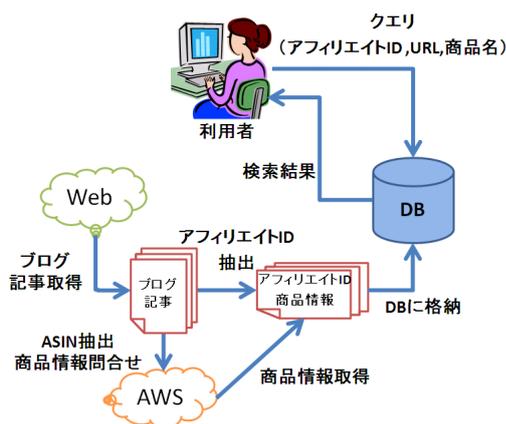


図 2: システム構成

商品数	最高値	最低値	平均価格
204	6980円	291円	1134円

商品ジャンル	カウント数	ジャンル
ヘルス&ビューティー	35	ヘルス&ビューティー-Refinements
ジャンル別	31	ジャンル別 雑誌
科学-テクノロジー	18	本ジャンル別 科学-テクノロジー-科学-テクノロジー-全般
旅行ガイド	16	本ジャンル別 旅行ガイド-旅行ガイド-全般
ホーム&キッチン	15	本ジャンル別 ホーム&キッチン-Home Brands
おもちゃ&玩具	12	おもちゃ&玩具-Refinements
スポーツ&アウトドア	11	本ジャンル別 スポーツ&アウトドア-スポーツ&アウトドア-全般
エンターテインメント	10	本ジャンル別 エンターテインメント-エンターテインメント-全般
ファッション	10	本ジャンル別 ファッション-ファッション-全般
動物&ペット	9	本ジャンル別 動物&ペット-動物&ペット-全般

ブログURL	記事URL	商品ID	商品名	記事更新日	記事取得日	ASP
http://www.amazon.co.jp/	http://www.amazon.co.jp/	B002764164	新装版 コイロカール ベイブズ (講談社文庫)	2009-12-19	2009-12-20	Amazon
http://www.amazon.co.jp/	http://www.amazon.co.jp/	B000JA7080	古稀入門百科辞典 増補の字引 (1959年)	2009-12-19	2009-12-20	Amazon

図 3: アフィリエイト ID 検索結果画面

リエイト ID と ASIN の抽出, 問合せを行いデータベースに格納している。2010年3月時点で, アフィリエイト記事約40万件, 商品件数約12万件のデータが蓄積されている。

#### 4. アフィリエイト分析支援システムを用いた分析事例

アフィリエイト分析支援システムを実際に利用した結果を, システムの出力画面と共に説明する。本システムの強みは, 先行研究と同じく1つのブログサイトごとの分析に加え, アフィリエイト ID 単位 (複数のブログサイトをまとめた) での分析も可能な点である。図3にアフィリエイト ID をクエリとした時の出力画面を示す。出力画面はメニューバー, 価格帯テーブル, 商品ジャンルテーブル, 記事一覧テーブルの4つのブロックに分けられる。メニューバー部分ではブログサイト一覧がドロップダウンメニューになっており, クエリのアフィリエイト ID が管理しているブログサイトを一覧で表示し, ブログサイトを基準とした分析結果を表示するページにリンクされている。価格帯テーブルでは, クエリのアフィリエイト ID が管理するブログサイト全てでの商品数, 最高値, 最低値, 平均価格を表示している。商品ジャンルテーブルでは, アフィリエイト ID が扱っている商品ジャンル上位10件と商品カテゴリ (例: 本, 音楽, DVD) から商品ジャンル (例: ミステリー, J-POP, アクション) までをパスとして表示している。記事一覧では, ブログ URL, 記事 URL, 商品 ID, 商品名, 記事更新日, 記事取得日, ASP を表示しており, どの記事にどんな商品が紹介されているのかがわかる。また, 列毎にソート機能が付いているので, 記事更新日でソートすることによって, 同時刻帯に複数のブログサイトに記事を投稿しているアフィリエイトの分析が可能となる。

##### 4.1 スпамアフィリエイトと非スパムアフィリエイトの商品比較

本システムでは, 個々のアフィリエイト ID の分析だけでなく, 複数のアフィリエイト ID の分析結果をまとめることによって, スпамと非スパム間や人気アフィリエイトサイトと一般アフィリエイトサイト間において, どのような特徴で差があるのかを分析することも可能である。本節では分析事例として, スпамアフィリエイトと非スパムアフィリエイトの商品比較を行う。

2009年12月に取得した1ヶ月分のアフィリエイト記事43,686

件の中からアフィリエイト ID あたりのアフィリエイトリンク数 (広告数) を集計した。アフィリエイトリンク数上位の Amazon のアフィリエイト ID をスパムユーザ候補として, 第一著者が目視でブログサイトを確認し, アフィリエイト ID 毎にスパム判定を行い, スпамアフィリエイト ID50件, 非スパムアフィリエイト ID (以下ハムアフィリエイト ID)50件を選定した。判定の基準は原 [原 09] らの論文を参考にし, 以下の2つとした。

1. アフィリエイト型スパムブログ (例: アフィリエイトリンク, 商品写真のみ)
2. 自動生成型スパムブログ (例: 他サイトからのコピー&ペースト, マルチポスト)

その結果, スпамアフィリエイト ID50件から331のブログサイトを, ハムアフィリエイト ID50件から52のブログサイトを確認した。最大で, 1つのアフィリエイト ID で183のブログサイトを管理していた。スパム, ハムで商品数を合計したところ, スпамアフィリエイト ID で6,833件, ハムアフィリエイト ID で3,923件となった。図4, 5にスパムアフィリエイト ID, ハムアフィリエイト ID それぞれでの商品カテゴリと価格帯の図を示す。括弧で囲まれた項目は件数が0であることを示す。図4のスパムアフィリエイト ID の図に着目すると, 商品カテゴリが18, 価格帯が200円から250,000円までに渡っていることがわかる。一方, 図5のハムアフィリエイト ID の図に着目すると, 商品カテゴリが13, 価格帯が400円から100,000円までに渡っていることがわかる。このことからスパムアフィリエイト IDの方が, 商品カテゴリ, 価格帯とも幅が広いことが確認できる。

#### 5. 考察

本分析手法を用いたアフィリエイト分析支援システムで行った分析により, スпамアフィリエイトの方が, 商品カテゴリ, 価格帯の幅が広いという結果を確認した。また, 1つのアフィリエイト ID で183のブログサイトを管理しているアフィリエイトを確認した。個々のスパムアフィリエイト ID に着目すると, 同時刻帯に複数のブログサイトに記事を投稿し, ブログサイトごとに商品傾向を変えているアフィリエイトを確認した。今後, 定量的な調査を行いアフィリエイトスパムに共通した広告掲載活動があるのか分析を行う必要がある。

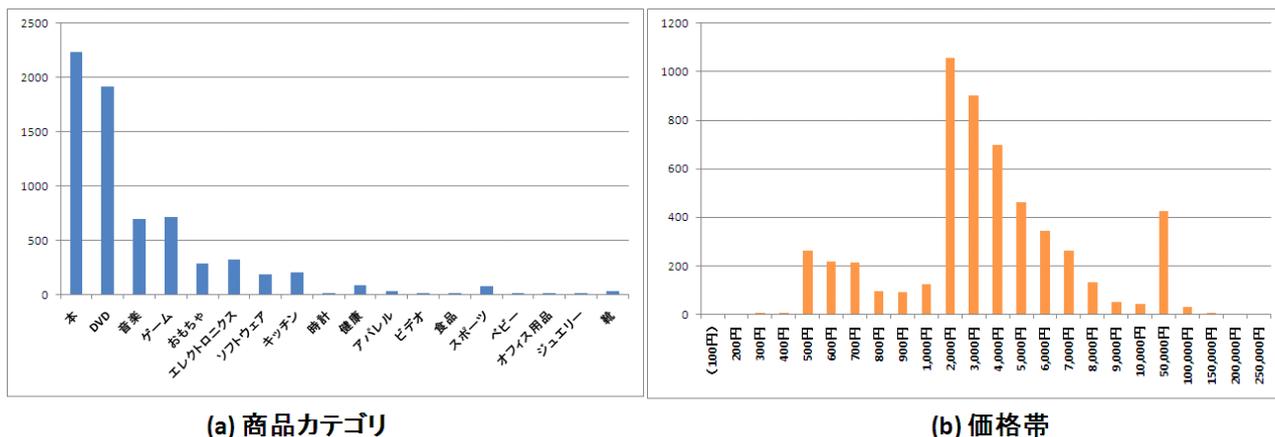


図 4: スпамアフィリエイト ID50 件での商品カテゴリと価格帯

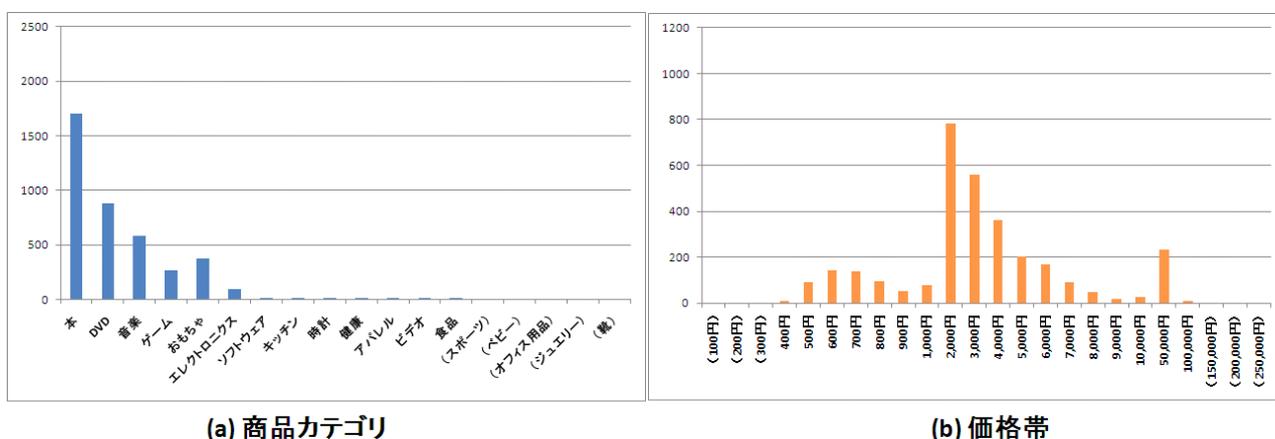


図 5: ハムアフィリエイト ID50 件での商品カテゴリと価格帯

## 6. おわりに

本論文では、個人のアフィリエイトが Web 上でどのような広告活動を行っているか分析するために、アフィリエイト ID を用いてアフィリエイトを特定するアフィリエイト ID 特定分析の提案とアフィリエイト分析支援システムの開発を行った。アフィリエイト ID に着目し、分析することによって複数のブログサイトにまたがるアフィリエイトの分析と掲載商品の分析が可能となった。今後の課題としては、アフィリエイトブログサイト間のリンク解析、広告掲載位置と本文内容を確認する為に、ブログ記事を本文と付随情報部分に分けての分析とブログ記事本文の分析が必要である。さらに特徴数を増やしていき、スパム、ハム間や、一般アフィリエイトサイト、人気アフィリエイトサイト間において、どのような特徴で差が出るのか分析を行っていき、アフィリエイトスパムのフィルタリングや良質なアフィリエイトサイトの推薦を行っていく。

## 参考文献

- [原 09] 原 正憲, 長谷 巧, 山本 匠, 山田 明, 西垣 正勝: スпамブログとアフィリエイトの関連性に爛する一考察, 情報処理学会論文誌, Vol.50, No.12, pp.3206-3210 (2009).
- [Wang 07] Y. M. Wang, M. Ma, Y. Niu, and H. Chen: Spam double-funnel: Connecting web spammers with advertisers, In Proceedings of the 16th WWW Conference (WWW2007), pp.291-300 (2007).
- [斉藤 09] 斉藤 香里, 村井 源, 往住 彰文: 心の状態と言語的特徴: ブログにおける商品紹介文の分析, 情報知識学会誌, Vol.19, No.2, pp.144-151 (2009).
- [石井 09] 石井 聡一, 芳中 隆幸, 福原 知宏, 増田 英孝, 中川 裕志: Web 上の広告情報を用いたアフィリエイトスパムの分析, 楽天研究開発シンポジウム 2009, pp.49-52 (2009).
- [伊藤 07] 伊藤 哲哉, 坂巻 隆之, 今村 仁彦, 鈴木 将司: アフィリエイトの達人養成講座, 翔泳社 (2007).