

CGMによる話題連動型広告配信システムの開発

増澤 晃^{*1} 南野 謙一^{*1} 渡邊 慶和^{*1}

^{*1} 岩手県立大学ソフトウェア情報学部

近年、ブログやSNS等が流行している。それらは、新聞や雑誌等の大手メディアと対比し、消費者が自ら情報を発信するメディアとしてCGM (Consumer Generated Media) と呼ばれている。CGMで広く取り上げられている話題は、商品を購入するための有用な情報源となっている。そこで本研究では、アフィリエイト広告を対象とし、CGMから話題を抽出して関連する商品広告の効果的な配信を行う話題連動型広告配信システムを提案する。本システムでは、広告掲載サイトのコンテンツ、CGMの話題、アフィリエイト商品広告とのマッチングを動的に行う。本システムは、利用者の購買履歴等の個人情報取得しそれをもとに個別に広告を配信するのではなく、CGMから多くの消費者が興味のある話題を取得しそれをもとに広告を配信する。本稿ではシステム開発と評価実験について述べる。

Development of Hot topic Match Advertising Delivery System from CGM

Akira Masuzawa^{*1} Kenich Minamino^{*1} Yoshikazu Watanabe^{*1}

^{*1} Faculty of Software and Information Science, Iwate Prefectural University

Recently, Blog and SNS (Social Network Services) are popular, and are called CGM (Consumer-Generated Media), compared with the mass media (newspaper, magazine, etc.). Hot topics from CGM are useful information for people to purchase goods. We present a Hot topic Match Advertising Delivery System which extract hot topics from CGM and deliver affiliate advertisings matched them dynamically. The system do not acquire personal information such as history of purchasing, but extract hot topics from CGM in which many people are interested. This paper describes the system development and an evaluation experiment of the system.

1. はじめに

近年、ブログ (Weblog) や SNS (Social Network Services) 等が流行している。それらは、新聞や雑誌等の大手メディアと対比し、消費者が自ら情報を発信するメディアとしてCGM (Consumer Generated Media) と呼ばれている³⁾。総務省の統計によると平成18年3月時点でブログ登録者数は868万人、SNS登録者数は716万人であり、昨年と比較するとほぼ2倍もの急激な伸び率を示している¹⁾。さらにクチコミサイト等が急増しそれらを踏まえたCGMで広く取

り上げられている話題は、商品を購入するための有用な情報源となっている¹⁾。

また、インターネット広告収入を企業の主利益とする企業が登場してきている。日本で最大級のSNSである(株)mixiでは全利益の2/3以上がインターネット広告による収入である⁴⁾。そのような企業などの登場が現すようにインターネット広告の手法は様々で市場は拡大している。消費者購買行動につながりCGMを広告媒体と捉え注目されているのがECサイト等に直結されている成果報酬型広告であるアフィリエイト

イト広告である。

そこで本研究では、アフィリエイト広告を対象とし CGM から話題を抽出して関連する商品広告の効果的な配信を行う話題連動型広告配信システムを提案する。本システムでは、広告掲載サイトのコンテンツ、CGM の話題、アフィリエイト商品広告とのマッチングを動的に行う。本システムは、利用者の購買履歴等の個人情報を取得しそれをもとに個別に広告を配信するのではなく、CGM から多くの消費者が興味のある話題を取得しそれをもとに広告を配信する。

まず2章では話題連動型広告を新たなインターネット広告手法として提案し、3章で話題連動型広告配信システムについて広告配信の流れやシステム構成を示しプロトタイプシステムの開発について述べる。4章で評価実験・実験結果として広告効果・性能評価を述べ、5章で本研究を通じての考察について述べる。

2. 話題連動型広告

CGM と購買活動に関する調査によると、ブログや SNS に代表される CGM 上での書き込み・クチコミを参考に商品・サービス購入をした経験は調査対象全体の約3割、購入検討まで含めると5割以上にものぼり、消費者の「生の声」を得る場として、ブログや SNS が大きな影響を与えているという結果が出ている²⁾。本研究では、このような CGM 上での書き込み・クチコミを参考にした購買活動を促進させる新たな話題連動型広告を提案する。話題連動型広告では、CGM 上で多くの消費者が取り上げている共通の書き込み・クチコミを話題として捉え取得し関連する商品広告を配信する。

本研究では、アフィリエイト広告を対象とし話題連動型広告を配信する。アフィリエイト広告とは、Web 利用者が Web サイト（広告掲載サイト）の商品広告をクリックし EC サイトで商品を購入すると、EC サイトから広告掲載サイトに報酬が支払われる広告である。話題連動型広告によって消費者の購買活動を促進させるこ

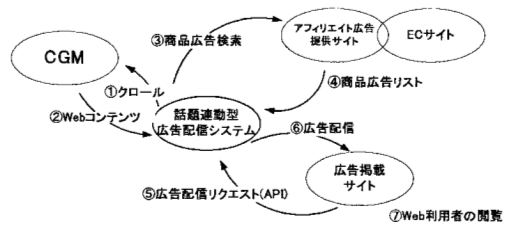


図1 広告配信の流れ

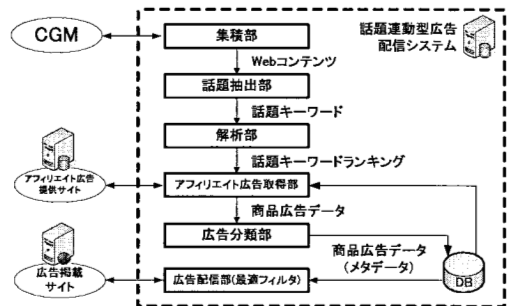


図2 システム構成

とにより、広告掲載サイトおよび EC サイトは利益を得ることができる。

3. 話題連動型広告配信システム

3.1 広告配信の流れ

本システムでは CGM から抽出した話題と、アフィリエイト広告提供サイト (EC サイト) の API を用いて取得した商品広告との動的なマッチングを行うことにより、商品広告リストを生成する。そして、これをもとに広告掲載サイトに対して広告配信リクエストが可能な API を公開する。この API では商品広告のジャンルを指定することができる。本システムでは、商品広告ジャンルが指定された広告配信リクエスト (API) を受け取ると、前述の生成した商品広告リストと動的なマッチングを行い、広告配信を行う。広告配信の流れを図1に示す。広告掲載サイトから API を介した本システムへの広告配信リクエストは、固定であり1度記述するだけでよいが、本システムから配信される商品広告は、CGM の話題が変動することにより動的に変化することになる。Web 利用者が広告掲載サイトの商品広告をクリックした場合の動作は、

通常のアフィリエイト広告と同様に、EC サイトに誘導されることになる。

3.2 システム構成

システム構成を図2に示す。本システムは6つのモジュールから構成される。各モジュールの動作について述べる。まず、CGM から Web コンテンツ（書き込み・クチコミ）をクロールし集積する（集積部）。話題をキーワードにより抽出し（話題抽出部）、話題キーワードのランキングを生成する（解析部）。それらをもとにアフィリエイト広告提供サイト（EC サイト）の API を介して関連する商品広告データを取得する（アフィリエイト広告取得部）。それらの商品広告をジャンルごとにグループ化する（広告分類部）。グループ化された商品広告は、商品広告データ（画像データ）そのものではなく、そのデータへのリンクを記述したデータ（商品広告メタデータ）としてデータベースで管理する。広告掲載サイトから API を介して広告配信リクエストを受け取ると、指定されたジャンルと商品広告リストとの動的なマッチング（最適フィルタ）を行い、広告配信を行う。配信される広告は、解析部の話題キーワードランキングの生成をトリガとして、動的に変化することになる。

3.3 プロトタイプシステム

プロトタイプシステムは OS : CentOS linux, Web サーバ : Apache2.0.52 , PHP5.1.6

(eAccelerator0.9.5), DBMS : MySQL5.0.22, 形態素解析エンジン : Mecab0.93 を用いて開発した。プロトタイプシステムでは、集積部、話題抽出部の実装を行わず、CGM を対象とした検索エンジンやブログ集約サイトが API, RSS を介して提供する話題キーワードを取得することとした。これにより短期間での開発が可能となっている。ただし、必ずしも評判の良い商品キーワードを取得できるものとはなっていない。このため、プロトタイプシステムの広告分類部では、商品名である話題キーワードと商品名以外の話題キーワードに分類する。商品名以外の話題キーワードの場合には、話題キーワードに関連のある商品の中から売れている商品の広告を自動的に選定し配信することにより対応する。

プロトタイプシステムの動作例を示す。まず、CGM 話題抽出サイトの 5 サイトから話題キーワードを定期的に収集する。5 サイトから 1 度に収集されるキーワードは、合計約 100 個である。ただし、各サイトの話題キーワードの更新時間は同一でないため、収集できる最新のキーワードは一定数ではない。次に、ランキングの重み付けや同義語処理により総合ランキングを生成する。12月31日0時の例では、家庭用ゲーム機「Wii」、テレビドラマ「のだめカンタービレ」、元イラク大統領「サダムフセン」が上位にある総合ランキングが生成された。



図3 広告掲載サイトにおける話題連動型広告例

総合ランキングの話題キーワードから、アフィリエイト広告提供サイトである Amazon ECS 4.0 の API に基づき商品広告メタデータを取得する。Amazon ECS 4.0 は REST (REpresentational State Transfer) リクエストであり、http リクエストの URL に各種の検索パラメータを設定できる。しかしながら、Amazon ECS 4.0 では、例えば「最優秀歌唱賞」等の話題キーワードのように商品広告を取得できない場合もある。広告分類部では、売れている商品や商品画像の有無を考慮し、取得した全ての商品広告に対してジャンルごとにグループ化する。広告分類部の処理の例として、話題キーワード「のだめカンタービレ」を挙げると、このキーワードは商品名ではないため、関連した商品広告を取得し、音楽 CD などにジャンルにグループ化する。このとき、画像がない商品広告の分類、売り上げごとのソートが行われる。

広告掲載サイトから受け付けている広告配信リクエスト (API) は、商品カテゴリ (Amazon ECS 4.0 で取得できる商品広告カテゴリを使用している) を事前に指定することができる。News サイトであればドキュメンタリー商品を優先することや、公序良俗に反する商品 (アダルト商品の除外など) を除外する等の指定を行うことができる。これにより簡易な広告掲載サイトのコンテンツ情報とのマッチングを実現している。

解析部の話題キーワードランキングの生成をトリガとして配信される広告は動的に変化する仕組みとなっている。CGM 話題抽出サイトの話題キーワードの更新時間を考慮し、時間間隔を 1 時間としている。トリガとなる話題キーワードの総合ランキングの生成を 1 時間毎に行い、広告を生成している。広告掲載サイトからの広告リクエスト (API) が広告表示スペースの理由から商品広告 3 つ分の表示のため、話題キーワードランキング 1 位~9 位までを 3 分割し、20 分毎に分割配信を行っている。

4. 評価実験

4.1 実験方法

話題連動型広告配信システムを開発し広告効果評価および性能評価を行う。評価実験の対象として広告掲載サイトに、(株)IBC 岩手放送のホームページにご協力を頂いた。この広告掲載サイトはアクセス数が月間 100 万 PV 以上であり広告掲載サイトとしての価値が高く、また、高度なアクセス解析が導入されており広告表示回数を取得でき、既に導入している Amazon アソシエイトサービス、コンテンツ連動型広告との比較が可能である。評価実験における広告配信例を図 3 に示す。

広告効果評価としては、eCPM (広告 1000 回表示またはアクセス 1000 ページビューあたりの収益額) \cdot imp (1 クリックあたりの収益額) \cdot Click rate (広告がクリックされた回数を、広告が表示された回数で割ったもの) の 3 つを評価基準としている。既存の静的 Amazon 商品リンク広告と、既存のコンテンツ連動型広告 (Google AdSense) の 2 種類と比較する。

性能評価としては、話題の変動に応じた広告変動数、新語含有率、商品画像取得、商品広告取得時間についての評価を行う。

4.2 実験結果

4.2.1 広告効果

広告効果評価の結果を図 4 に示す。本研究の広告が他の広告と比較しクリック率が高いことが分かった。静的 Amazon 商品リンクとは商品広告を固定掲載することを指し、通常のアフィリエイト広告の運用形態である。コンテンツ連動型広告は収入の形態がクリック保障型広告であるため imp が高いという特徴がある。広告の統一指標に用いられる eCPM では本研究の広告が高いことが分かった。長期運用によって eCPM は低下する可能性も有りうる。これら結果から、本システムの広告は他の広告よりも Click rate (クリック率) が高く消費者の購買活動に結びついていることが分かった。CGM から話題に関連した広告に加えて、話題の変動に応

7日間	静的 Amazon 商品リンク広告	コンテンツ 連動型広告	本システム
広告表示回数	13,989	13,989	13,989
クリック数	11	33	69
click rate	0.02	0.07	0.15
imp	0.00	29.10	21.91
eCPM	0.00	20.73	32.63

図4 広告効果

28日間	静的 Amazon 商品リンク広告	コンテンツ 連動型広告	本システム
広告表示回数	73281	73281	73281
クリック数	63	203	313
click rate	0.09	0.28	0.43
imp	0.00	33.45	8.56
eCPM	0.00	92.67	36.54

図5 広告効果

じて商品広告が変化することが、本研究の広告の特徴であり、他の広告との違いとなっている。商品広告を動的に変化させることにより、閲覧者に対するインプレッション効果を高めることができ、広告効果に影響を与えている。

4.2.2 性能評価

(1) 広告変動数

話題キーワードの変動に応じた広告の動的な変動を分析し、広告変動数を評価する。図6に話題キーワードの1日毎の変動数を示す。この図はシステムで生成した話題キーワードの総合ランキングの上位10個がどの程度入れ替わっているかを1日毎に示したものである。1位のキーワードの順位の変動ではなく、ランキング全体集合10個の中でどの程度変動したのかを示している。結果として1日毎の話題キーワード変動数の平均は10個中5.23であった。

次に広告掲載サイトにおける商品広告の1日毎の変動数を図7に示す。広告変動数の平均は10個中4.96個であった。以上から示すように1日に半分以上動的に話題が変化し、広告を動的に生成することができた。

広告変動数と広告クリック率の変化を分析する。広告クリック率の変化を図8に示す。図6

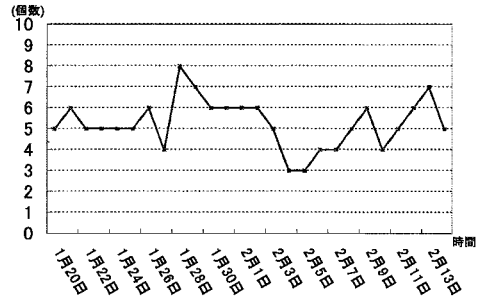


図6 話題キーワード変動数/1日

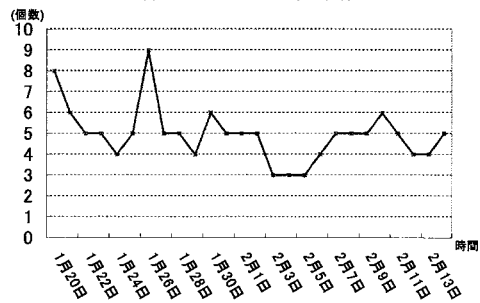


図7 広告変動数/1日

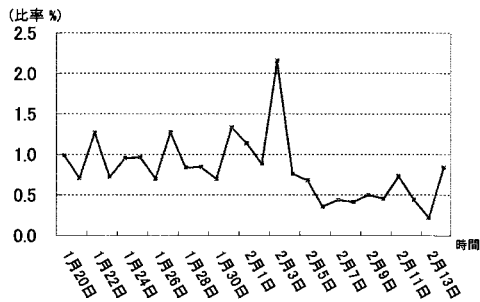


図8 クリック率/1日

から図8のグラフの定義域(期間)は同じである。固定掲載されたアフィリエイト広告ではクリック率が大きく変化することはなく徐々に低下することが問題となっている。これは閲覧者数が限られており、2回以上の表示で広告の効果が減衰していくために起こる。本システムでは話題キーワードにより広告を変動させることによって同じ広告閲覧者に対しても興味を起こさせることにより高いクリック率を得ることができている。

図6で話題の変動が大きい具体例を示す。2007年1月28日は映画のタイトルやテレビ番

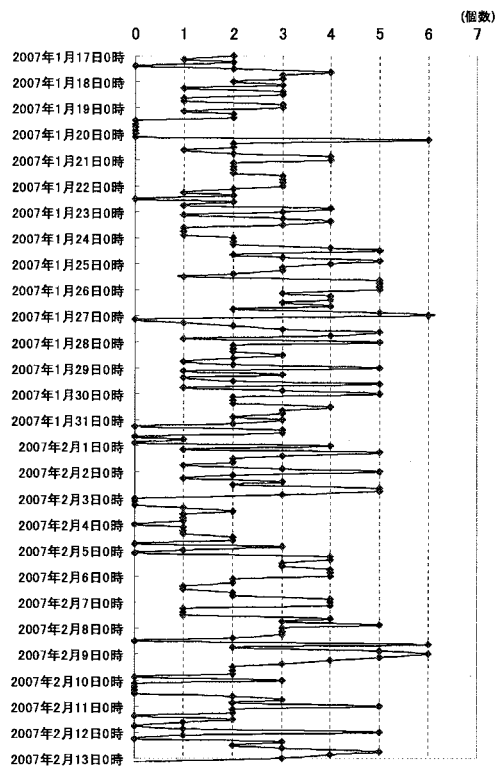


図9 話題キーワード変動数/3時間

組やニュースで取り上げられた「どろろ、大日本人、wii、芸能人タイプ診断、ベリサタ、国旗制定記念日、帝人事件、花より男子」が前日から入れ替わった話題キーワードであった。

3時間毎の話題キーワードの変化を図9に示す。図6の1日毎と比較すると、話題の変化が大きいこと分かる。CGMの話題が頻繁に変動しており、広告を変動させるトリガとして有効であることが分かる。

話題変動数の平均と広告変動数の平均とが異なる原因として、話題キーワードに関連する広告が見つからない場合や不適切な場合がある。新語であるキーワードの場合や、アフィリエイト広告提供サイトの精度の問題がある。これらについては以下に述べる。

(2) 新語含有率

新語とは辞書に掲載されていない単語のことであり、アフィリエイト広告提供サイトから商品広告を取得する際に、広告が見つからないと

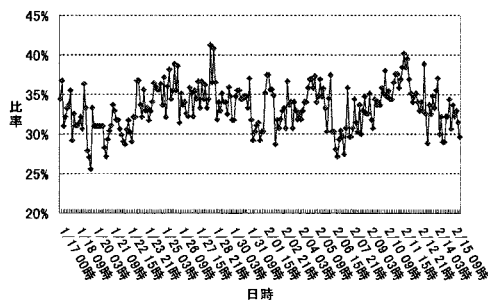


図10 新語含有率/3時間毎

いう問題を引き起こす。この新語含有率を分析し、システムに及ぼす影響を評価する。新語の検出には形態素解析エンジンMecabの未知語検出機能を利用した。未知語とは辞書に未登録な単語であり、新語と同様のものである。辞書はIPA辞書であるipadic-2.7.0-20060707を使用した。話題キーワードの総合ランキング全体のうち新語である割合の平均値は30.26%であった。図10に示したように変化があまり起きず一定の割合で新語が発生していることが示された。話題キーワードが新語の場合には、その新語の同義語、広義語や関連語から広告を検索する必要があり処理速度が遅くなる。プロトタイプシステムでは、1時間毎に広告を取得しているため、問題とはなっていないが、取得間隔を短くすると問題となってくる。

(3) 広告画像取得

インターネット広告では画像によるインプレッション効果が期待できるため、アフィリエイト商品広告データに画像情報があるかどうかは重要な問題である。プロトタイプシステムでは話題キーワードに対して複数の商品広告を取得した場合、それらを売上順にソートし商品広告を選定している。このとき、商品広告に画像データがなければ排除している。商品広告に画像データがない割合を図11に示す。画像データがない商品広告データは平均値13.41%存在する。下限5%～上限25%の割合で画像データがなかったが、すべての商品広告に画像データがないということはない。

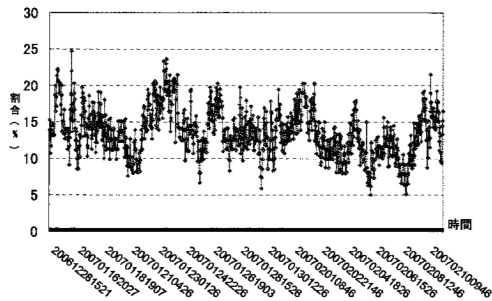


図 11 アフィリエイト広告画像なしの割合

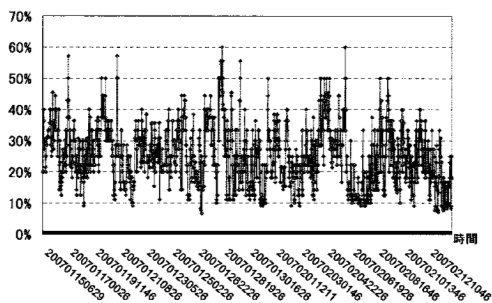


図 12 画像なし新語含有率

売上順にソートした商品広告において、1 番目の商品広告では 13.41%で画像データがなかったが、分析対象をすべての商品広告データとする場合、平均値 28.32%の割合で画像データがなかった。これらから分かるように売上順位高ければ高いほど商品画像が存在することが示された。

商品広告データに画像データがない原因を分析するため、それらの商品広告を取得するために検索した話題キーワードの新語含有率を調査した。方法は前節の (2) 新語含有率と同じ手法を取る。図 12 に示したように結果として画像データが無い商品広告は 25.5%の割合で新語であることが示された。この図において局所的に 60%の割合で新語である日時も確認でき、新語が商品広告データの有無に関係あることが強く示された。これらのデータは Amazon ECS 4.0 を用いた場合であり、アフィリエイト商品広告提供サイトに依存している数値である。

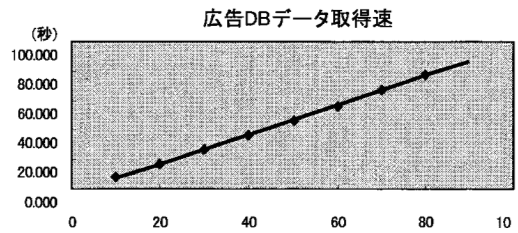


図 13 商品広告の取得時間

(4) 商品広告取得時間

本システムではリアルタイム性を重視し広告配信を行うことを目的としているが、アフィリエイト広告提供サイトからの商品広告の取得に最も時間がかかり、リアルタイム性に影響を与えていることが分かった。アフィリエイト広告提供サイトからの商品広告の取得時間を図 13 に示す。システムの実行速度はアフィリエイト広告提供サイトからの商品広告の取得時間に依存していた。このため、プロトタイプシステムではアフィリエイト広告を検索するキーワード数を調整しシステム実行速度を向上させている。

次にプロトタイプシステムの可能な配信数を示す。CPU 負荷などから広告配信数の理想値を算出すると 100 クエリ毎秒であった。評価実験での広告配信は 1 サイトであり問題がなかったが、プロトタイプシステムのネットワーク速度を測定し広告配信ファイルサイズから算出すると 1000 クエリ毎秒であったため、実運用に向けて負荷分散により配信可能数を向上する必要がある。

5. 考察

5.1 話題キーワードの意味解析

話題キーワードから総合ランキングを生成するとき、意味解析レベルでの話題キーワードをクラスタリングする必要がある。例えば、テレビドラマであれば、番組名・出演者・役名・脚本家・原作者・監督・主題歌・撮影場所などのテレビドラマの数限りない属性をまとめることによって、重複のない正確なランキングを生成することが可能となる。商品広告を取得する

場合には、どのキーワードを用いるかで広告効果に違いが出るため、キーワードの適切な選択方法も必要となる。ただし、CGM で話題なのは登場人物なのか原作者なのか、どちらかは判断しがたく容易にまとめすぎることが問題となる場合もある。

プロトタイプシステムでは、集積部、話題抽出部の実装を行わず、CGM を対象とした検索エンジンやブログ集約サイトが API、RSS を介して提供する話題キーワードを取得することとしたが、集積部、話題抽出部を設計、実装し、これらの問題を解決する必要がある。

5.2 広告掲載サイトのコンテンツへの適合

本研究の評価実験では、話題キーワードの総合ランキングには手を加えず、商品広告のみを広告掲載サイトのコンテンツに適合させている。しかしながら、話題キーワードと広告掲載サイトのコンテンツが合っていない場合には、適切な商品広告の選定は必ずしもできるとは言えない。このため、広告掲載サイトのコンテンツによって総合ランキングを、コンテンツに適したキーワードの中から生成し、それから商品広告を選定し配信することが可能である。これにより、どの程度広告効果が得られるか調査する必要がある。

5.3 広告掲載サイトによる広告効果の違い

評価実験では話題連動型広告の配信先として PV 数が確保できるテレビ局のトップページを対象とした。しかし、本システムの配信する広告が最も効果的である Web ページは、EC サイト、ポータルサイト、ブログ、SNS、ニュースサイトなどの掲載媒体の中でどれになるのか、広告効果を測定する必要がある。

EC サイトにおいては、他のサイトとは異なり、その EC サイトの商品を広告として表示し、話題の商品に消費者を誘導することができるため、キーワードを用いたサイト内商品検索よりも効果が期待できる。また、ブログや SNS については、CGM に興味を持って閲覧している Web 利用者を対象としているため、話題に対す

る興味は他のサイトの閲覧者よりも高いことが期待できる。このため、広告とともに話題のブログや SNS に誘導し、広告効果を高めることができる。

6. おわりに

本研究では、アフィリエイト広告を対象とし、CGM から話題を抽出して関連する商品広告の効果的な配信を行う話題連動型広告配信システムを提案した。評価実験の結果、広告効果として他の広告である固定掲載アフィリエイト広告やコンテンツ連動型広告よりも広告効果を得られた。性能評価を通じて話題キーワードから商品広告をリアルタイムに生成・配信することが出来た。

今後の課題として、(1) 話題キーワードの意味解析を目的とした集積部、話題抽出部を設計・実装する。(2) 様々な広告掲載サイトに対して広告配信実験を行う。(3) 応用として、話題に連動した広告やニュースなどの旬な話題を仕事に生かす営業ツールの開発を行うことが上げられる。

謝辞

本研究を進めるにあたり、実験にご協力頂いた (株) IBC 岩手放送メディア企画部に深く感謝致します。

参考文献

- 1) 総務省、報道資料平成 18 年 4 月 13 日、ブログ及び SNS の登録者数、
http://www.soumu.go.jp/s-news/2006/060413_2.html
(最終閲覧日 2007.2.5)
- 2) (株)サイバーエージェント、プレスリリース
<http://www.cyberagent.co.jp/news/press/2006/1225.html>,
(最終閲覧日 2006.12.25)
- 3) 石田和成、社会環境の変化と CGM、情報処理学会 情報システムと社会環境研究会研究報告 2006-IS-97PP.37-44 (2006)
- 4) RBB TODAY (ブロードバンド情報サイト)
<http://www.rbbtoday.com/news/20070130/38068.html>,
(最終閲覧日 2007.2.5)