

ブログ記事の反響特性分析

みやた あきひろ かわしま はるみ
宮田 章裕 / 川島 晴美
 ふじむら こう
藤村 考

NTTサイバーソリューション研究所

本稿では、「読者からどのような反響を呼んでいるか」という観点でブログ記事を評価する反響特性分析という技術を紹介します。

ブログについて

ブログは国内だけでも1 000万件以上開設されていると推定されており、商品のクチコミやサービスの体験談、時事問題に対する意見やプログラミングのテクニックなどがブログ記事として日々Web上に投稿されています。また、大半のブログにはコメントやトラックバックという機能が用意されており、これを利用してブログ記事の作成者と読者がコミュニケーションを行うことがしばしばあります。一般的なブログ記事の構造を図1に示します。

コメントとは、ブログ記事の読者が感想や意見などを作者に伝えるために

送信するメッセージであり、送信されたコメントは一般的に記事本文のすぐ下に表示されて、誰でも閲覧できます。

トラックバックとは、ブログ記事Aの内容をブログ記事Bが引用・参照したとき、記事Bの作者が内容の関連性を記事Aの作者や読者に伝えるために、記事Bから記事Aに送信するものです（ただし、この原則に従っていないトラックバックも少なくありません）。つまり、多くの記事から引用・参照される記事にはたくさんのトラックバックが送信されることになります。

ブログ記事検索の問題点

冒頭で述べたように、ブログ記事の

内容は多岐にわたります。そして、これらのブログ記事から、目的に合ったものを見つけ出す技術がブログ記事検索です。しかし現状のブログ記事検索サービスでは、「読者から反響を呼んでいる記事」を発見することが困難です。

あるブログ記事が読者から反響を呼んでいるかどうか推定するためには、その記事が外部のページからどのくらいリンクされているか分析する方法がまず考えられます。実際、ブログ記事ではない一般のWebページの検索には、外部のページから多くリンクされているページに高いスコアを与えるというスコアリング方法がよく用いられます。しかし、外部のページからリンクされているブログ記事は全体のわずか1%程度であるという報告があります⁽¹⁾。ブログサイト単位ではリンクされているケースもありますが、ブログサイト内の個々の記事へのリンクは非常に少ないのが現状です。つまり、一般のWebページ検索で用いられているようなリンクベースのスコアリング方法はブログ記事検索には不向きであるといえます。

一方、ブログ記事のアクセス数を用いて読者からの反響を推測する手段も考えられます。しかし、ブログ記事のアクセス数を、そのブログサービスを運

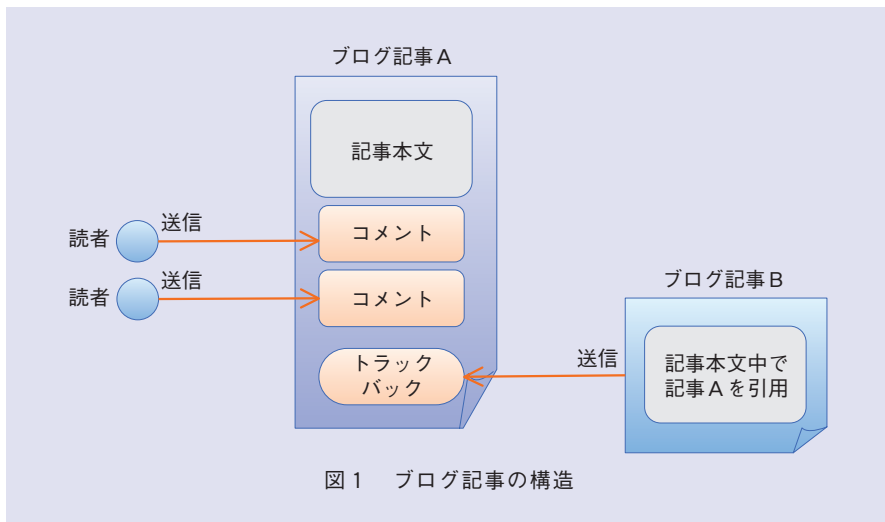


図1 ブログ記事の構造

営している事業者以外が取得することは一般に困難です。国内の大手ブログサービス事業者だけでも十数社が乱立している現状では、ブログ記事のアクセス数を網羅的に取得することは事実上不可能でしょう。

アプローチ：反響特性分析

前述の問題を解決するために、私たちは反響特性分析という技術を提案しています⁽²⁾。これは、読者の行動情報を分析することで、ブログ記事が読者からどのような反響を呼んでいるか評価する技術です。現時点では、記事に対して反響を示す行動として一般的であるコメント、トラックバックを分析します。

このとき、単純にコメント数やトラックバック数を数えるだけでは、記事が呼んでいる反響を精度よく測定することができません。例えば、図2(a)、(b)の記事はともに10個のコメントを受信していますが、読者の反響の様子が異なるのがお分かりだと思います。図2(a)の記事には10人の読者が1個ずつコメントを送信しています。このように大勢から反響を得る記事には、多くの読者が関心を示すような有益な情報が書かれている場合が多いでしょう。

一方、図2(b)の記事には2人の読者が5個ずつコメントを送信しています。このような記事には、特定の人だけが興味を持つような情報や、知人にしか理解できない内輪話が書かれている場合が多いでしょう。このように、コメント数が同じでも反響の様子が同じであるとは限りません。つまり、単純にコメント数やトラックバック数を

数えるだけでは記事が呼んでいる反響の詳細を識別することができないのです。

そこで反響特性分析では、表に示すようなコメントやトラックバックに関する特徴量を利用して、記事が呼んでいる反響の広さ、強さ、速さ、長さを測定します。

(1) 反響の広さ

図3に示すのは、反響の広さの概念です。広さとは記事に関心を示した読者の幅を表しており、コメント送信者数、トラックバック送信者数を用いて測定します。大勢の人が関心を示した

記事(例：多くの第三者にとって有益な情報を含む記事)は、広さスコアが高くなります。一方、特定の人だけが関心を示した記事(例：マニア向けの情報や内輪話から成る記事)は、広さスコアが低くなります。

(2) 反響の強さ

図4に示すのは、反響の強さの概念です。強さとは記事に対する読者の関心の度合いを表しており、平均コメント送信数を用いて測定します。各読者の関心の度合いが高い記事(例：何度もコメントを送信して議論したくなるような物議を醸す記事)は、強さスコ

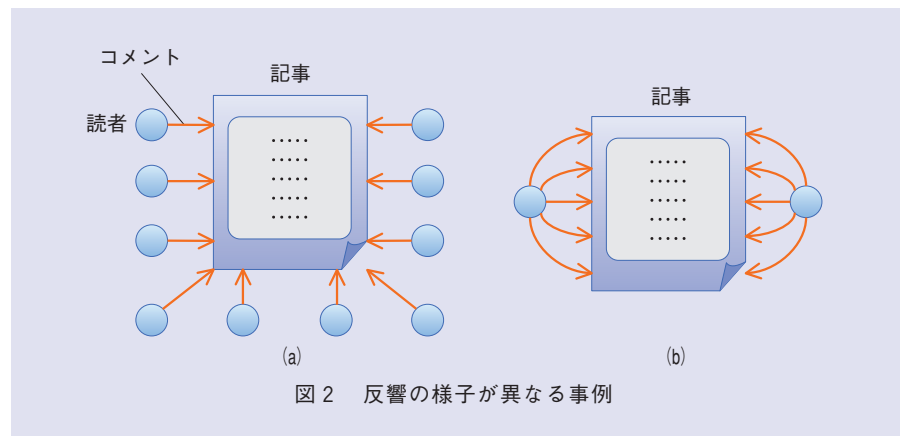


図2 反響の様子が異なる事例

表 コメント・トラックバックの特徴量

特徴量	説明
コメント送信者数	コメントを送信した人数
平均コメント送信数	コメント送信者1人当りの平均コメント送信数
初コメント受信経過時間	記事が投稿されてから最初のコメントを受信するまでに経過した時間
最終コメント受信経過時間	記事が投稿されてから最終のコメントを受信するまでに経過した時間
平均コメント間隔	コメント受信時間間隔の平均値
トラックバック送信者数	トラックバックを送信した人数
初トラックバック受信経過時間	記事が投稿されてから最初のトラックバックを受信するまでに経過した時間
最終トラックバック受信経過時間	記事が投稿されてから最終のトラックバックを受信するまでに経過した時間
平均トラックバック間隔	トラックバック受信時間間隔の平均値

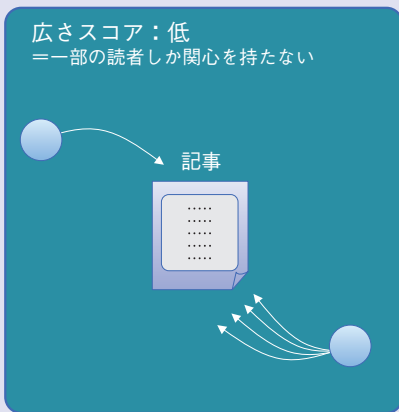
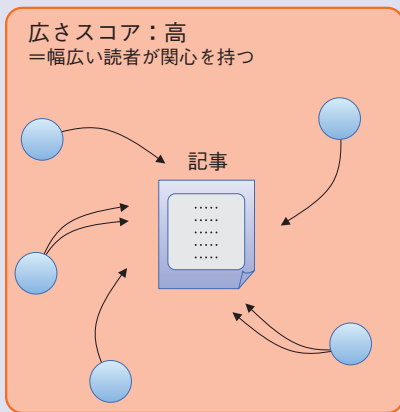


図3 反響の「広さ」

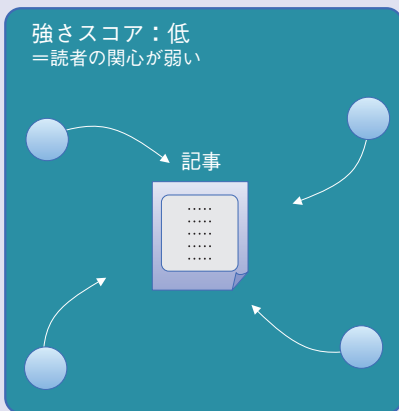
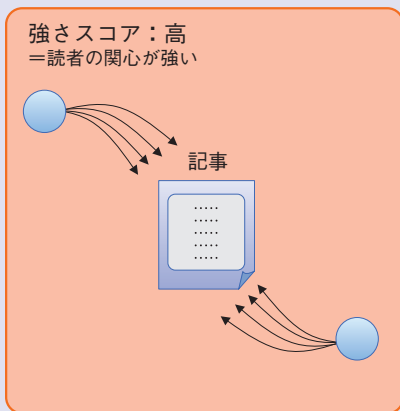


図4 反響の「強さ」

クバック受信経過時間を用いて測定します。読者からの関心を集め続けた期間が長い記事（例：資料的な内容の記事）は、長さスコアが高くなります。一方、読者からの関心を集めた期間が短い記事（例：時事ニュースに関する記事）は、長さスコアが低くなります。

サービスイメージ

反響特性分析技術を用いると「読者から反響を呼んでいるか」という観点でブログ記事を評価できるので、現状のブログ記事検索サービスをより豊かなものにすることができます。

例えば「ガソリン税」というキーワードでブログ記事検索をしてみてください。検索結果には、ガソリン税に関する愚痴をこぼしているだけのブログ記事や、ニュースサイトからガソリン税関連の記事を切り貼りしているだけのブログ記事が入り混じっていると思います。この検索結果の記事集合に対して反響特性分析を行い、広さスコア・強さスコアが高い記事を抽出すると、「ガソリン税について大勢の人がコメントで議論を交わし合っている記事」が発見しやすくなります。速さスコアを考慮すれば議論のテンポが速い記事、遅い記事を区別して探すこともできます。

資料として役立つようなブログ記事を検索する場合にも反響特性分析は有効です。丁寧に書かれた映画レビュー記事や、重大ニュースに関して深い考察を行っている記事は多くの人から長期間にわたって関心を集め続けることが多いので、これらが発見したいときは広さスコア・長さスコアが高い記事

アが高くなります。一方、各読者の関心の度合いが低い記事（例：議論の余地がない記事）は、強さスコアが低くなります。

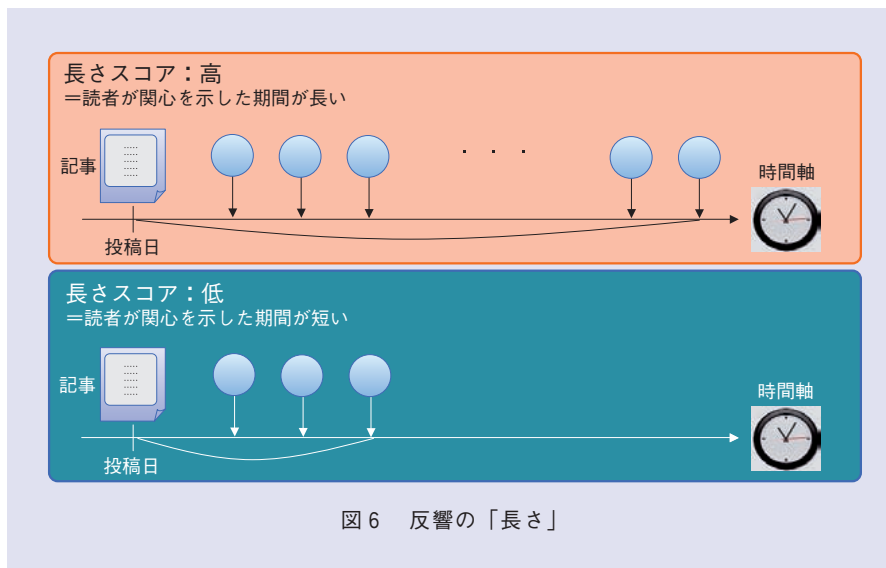
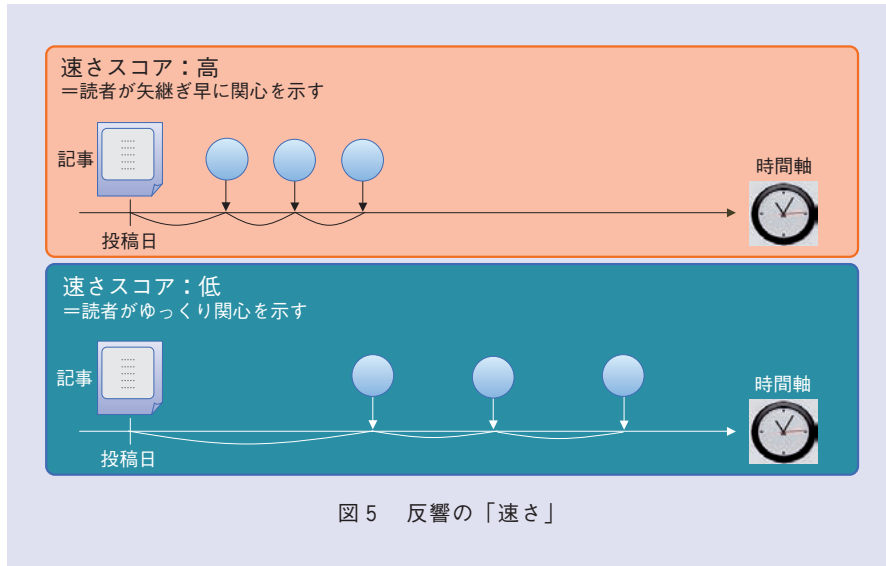
(3) 反響の速さ

図5に示すのは、反響の速さの概念です。速さとは記事に対して読者が関心を示す行為を行う速度を表しており、初コメント受信経過時間、平均コメント間隔、初トラックバック受信経過時間、平均トラックバック間隔を用いて測定します。読者が関心を示した速度

が大きい記事（例：旬のニュースのように読者が矢継ぎ早に関心を示す記事）は、速さスコアが高くなります。一方、読者が関心を示した速度が小さい記事（例：急いで反応する必要がないような日常的な内容の記事）は、速さスコアが低くなります。

(4) 反響の長さ

図6に示すのは、反響の長さの概念です。長さとは記事に対して読者が関心を示し続ける期間を表しており、最終コメント受信経過時間、最終トラッ



を抽出します。

今後の取り組み

今後、本技術を用いたブログ記事検索が、どのような分野・目的において有効であるか検証を行う方針です。すでに実施した評価実験では、「多くの人が参考している映画レビュー記事」や「読者への影響力が強いクチコミ記事」の発見に見込みがあることを確認

しています。

また、本稿ではブログ記事に対する反響の広さ・強さ・速さ・長さに関して解説しましたが、他にもコメントの文字数や、画像・絵文字の数など考慮すべき点は数多くあります。今後はこれらの特徴量も用いて反響特性分析の精度を向上させるとともに、SNS (Social Networking Service) や電子掲示板などブログ以外のコンテ

にも適用していく方針です。

参考文献

- (1) K. Fujimura, T. Inoue, and M. Sugisaki : "The EigenRumor Algorithm for Ranking Blogs," WWW 2005 2nd Annual Workshop on the Weblogging Ecosystem, Chiba, Japan, May 2005.
- (2) 宮田・松岡・岡野・山田・石打・荒川・加藤 : "反響特性分析を利用したブログ記事検索手法," 情処学論, Vol.48, No.12, pp.4041-4050, 2007.
- (3) A. Miyata, H. Kawashima, and H. Okuda : "A Communication-based Approach for Detecting Influential Blog Entries," IADIS International Conference WWW/Internet 2008, Freiburg, Germany, Oct. 2008.



(左から) 川島 晴美/ 宮田 章裕/
藤村 考

皆様の心に響く情報を発見する技術の実現を目指し、私たちはこれからも努力をしていきます。

◆問い合わせ先

NTTサイバーソリューション研究所
メディアコンピューティングプロジェクト
TEL 046-859-3916
FAX 046-859-5552
E-mail miyata.akihiro@lab.ntt.co.jp