

F-049

ブログ空間における情報伝搬パターンの分析

An analysis of Information Diffusion Patterns in Blogspace

伊藤 政志*1
Masashi Ito

木村 昌弘*2
Masahiro Kimura

斉藤 和巳*3
Kazumi Saito

1. はじめに

近年、ブログは Web 上で誰もが容易に情報を発信できるメディアとして急速に普及してきている。ブログ空間は重要なコミュニケーションの場として注目され、ブログ空間における情報拡散現象の分析が始まっている[1, 2]。一方、ネットワーク上で情報拡散の数理モデルに従って、情報の拡散を最大にする情報源ノード群を抽出する研究[3]や、汚染情報の拡散を最小にする削除リンク群を抽出する研究[4]など、ネットワーク上での情報拡散に関する研究も行われている。

ブログ空間における情報拡散現象には、ブログロールのような社会ネットワークを介したロコミ的な情報伝搬現象が存在すると考えられる。一方、ニュースサイトなどの主要なメディアから配信された情報を、他人のブログ記事をほとんど見ることなく自分のブログにおいて言及することにより、ブログ空間において情報が拡散していくという場合もあると考えられる。本論文では、Adar と Adamic の研究[1]に従い、ブログ記事に記述された URL 情報の拡散によりブログ空間での情報拡散パターンを調べ、ブログロールを通じてのロコミ的な情報の拡散パターンと、主要メディアを情報源とするような情報の拡散パターンとの違いについて分析する。

2. 分析データ

分析には Doblog*1 のデータ*2 を利用した。ブログ記事数は 1,540,077、ブログ数 (ユーザー数) は 52,525、ブログロール数 (ブックマーク数) は 115,552 であった。

3. 分析方法

ブログ記事群に記述されていた相違なる URL の全体

$$\{u_n; n=1, \dots, N\}$$

を同定し、URL u_n が記述されていたブログ記事とその時刻のペアの全体

$$\{(b_{n,k}, t_{n,k}); k=1, \dots, K_n\}$$

を抽出した。ただし、

$$K_n > 1, t_{n,1} < \dots < t_{n,K_n}$$

とする。このとき、

$$N = 64430, \sum_{n=1}^N K_n = 264845$$

*1 龍谷大学 大学院理工学研究科 電子情報学専攻

*2 龍谷大学 理工学部 電子情報学科

*3 静岡県立大学 経営情報学部

であった。

次に、URL u_n が記述されていたブログ間のブログロール全体を抽出し、その中で u_n 情報の伝搬に関与した可能性があるブログロール群を同定した (図1参照)。ブログロールを通じてのロコミ的な情報の拡散パターンと、ニュースサイトなどの主要なメディアを情報源とするような情報の拡散パターンとの違いを調べるために、ブログロールのみを通じて伝搬した可能性がある URL の全体 U_{br} と、主要なニュースサイト内の URL の全体 U_{mv} を抽出した。

$$|U_{br}| = 2520, |U_{mv}| = 749$$

であった。表1に、抽出した主要ニュースサイトを示す。

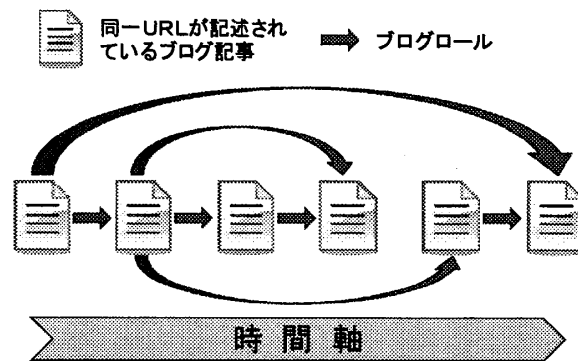


図1：同一URLが記述されたブログ記事の時系列と情報伝搬に関与した可能性があるブログロール

表1：抽出したニュースサイト

headlines.yahoo.co.jp	www.chugoku-np.co.jp
www.asahi.com	sports.yahoo.co.jp
www.yomiuri.co.jp	japanese.chosun.com
www.nikkansports.com	news.goo.ne.jp
www.itmedia.co.jp	www.jsgoal.jp
www.mainichi-msn.co.jp	news.www.infoseek.co.jp
sportsnavi.yahoo.co.jp	internet.watch.impress.co.jp
www.sanspo.com	www.hmv.co.jp
www.sankei.co.jp	www.daily.co.jp
www.nikkei.co.jp	pc.watch.impress.co.jp
www.sponichi.co.jp	www.nhk.or.jp
www.zakzak.co.jp	

※1 © (株) NTT データ。http://www.doblog.com.

※2 (株) ホットリンクと (株) NTT データの共同事業契約に基づき、(株) ホットリンクより提供。2003年10月から2005年6月のデータを利用。

4. 分析結果

図2は、URL u_n が記述されていたブログ記事数 K_n に対して、ブログロール数の比率 $R_n = M_n / (K_n - 1)$ をプロットしている。ここに、 M_n は URL u_n が記述されていたブログ間のブログロールの総数である。「十字印」はブログロールのみを通じて伝搬した可能性がある URL $u_n \in U_{br}$ における結果を、「丸印」はニュースサイト内の URL $u_n \in U_{mv}$ における結果を、それぞれ表わしている。図2において直線 $R=1$ に注目することにより、ほとんどのニュースサイトの URL 情報は、ブログロールを介さずに広がったことが観察される。また、ニュースサイト内の URL $u_n \in U_{mv}$ に比べて、ブログロールのみを通じて伝搬した可能性がある URL $u_n \in U_{br}$ には、それを記述していたブログ記事数の非常に多いものが存在することも見て取れる。

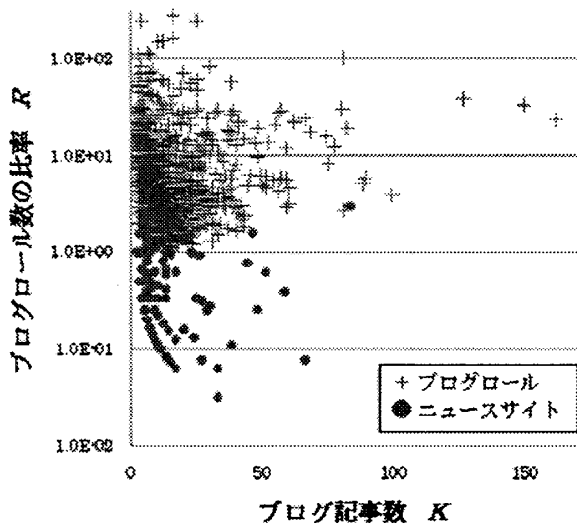


図2：同一 URL が記述されている記事数とそれらのブログ間のブログロール数の比率

図3は、URL u_n がブログ記事に記述されていた期間 (URL u_n の存在期間)

$$T_n = t_{n,K_n} - t_{n,1}$$

の分布を、ブログロールのみを通じて伝搬した可能性がある URL 全体 U_{br} に対して、およびニュースサイト内の URL 全体 U_{mv} に対して表示している。「十字印」は U_{br} における分布を、「丸印」は U_{mv} における分布を、それぞれ表わしている。図3より、どちらの分布もべき乗則的な性質が観察される。また、 U_{br} の存在期間分布の方が U_{mv} の存在期間分布に比べてべき指数の値が小さいことも観察される。したがって、ニュースサイト内の URL 情報に比べて、ブログロールのみを通じて伝搬した可能性がある URL 情報の方が、存在期間の長いものの割合が高いという傾向があると考えられる。すなわち、ブログロールを通じて口コミ的に伝搬するような情報は、ニュースサイトなどの主

要なメディアを情報源とするような情報と比べて、長い期間存在する傾向があると考えられる。

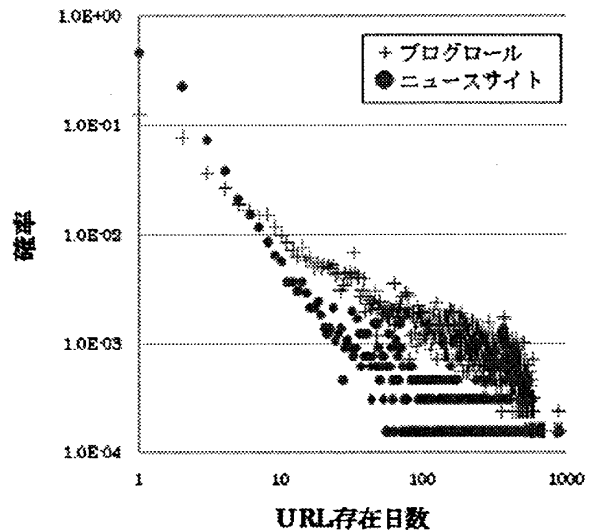


図3：URL 存在期間分布

5. まとめ

ブログ空間での情報拡散を、ブログ記事に記述されている URL に注目し分析した。特に、ブログロールを通じて伝搬するような口コミ的な情報と、ニュースサイトなどの主要メディアを情報源とするような情報との違いを分析した。前者は後者に比べて、非常に多くのブログに拡散するものが存在すること、また、長い期間をかけて拡散する傾向があることを示した。

謝辞

Doblog データは (株) NTT データおよび (株) ホットリンクより提供を受けた。記して感謝致します。

本研究は、科学研究費補助金基盤研究(C) (No.20500147) の補助を受けた。

参考文献

- [1] Adar, E., and Adamic, L. A.: Tracking information epidemics in blogspace, *Proceedings of the 2005 IEEE/WIC/ACM International Conference on Web Intelligence (WI'05)*, pp.207-214 (2005).
- [2] Furukawa, T., Matsuo, Y., Ohmukai, I., Uchiyama, K., and Ishizuka, M.: Analyzing reading behavior by blog mining, *Proceedings of the 22nd AAAI Conference on Artificial Intelligence (AAAI-07)*, pp. 1353-1358 (2007).
- [3] Kempe, D., Kleinberg, J., and Tardos, E.: Maximizing the spread of influence through a social network, *Proceedings of the 9th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD-2003)*, pp. 137-146 (2003).
- [4] Kimura, M., Saito, K., and Motoda, H.: Minimizing the spread of contamination by blocking links in a network, *Proceedings of the 23rd AAAI Conference on Artificial Intelligence (AAAI-08)*, pp. 1175-1180 (2008).