

社会的関係と行動類似性に着目した ブログ空間におけるユーザ訪問行動の予測

北野 由幸*1

Yoshiyuki Kitano

熊野 雅仁*1

Masahito Kumano

木村 昌弘*1

Masahiro Kimura

1. はじめに

本研究では、ブログ空間におけるユーザの訪問行動の有効性のある予測法の構築を目指している。ユーザが将来どのブログを訪問するかについては、社会的関係のある他のユーザと同じ行動をする場合、もしくはユーザ同士の過去の行動が類似している場合という二つの観点から考えられる。そこで前者を社会的関係、後者を行動類似性のある関係と見なし、これら二つの関係に着目した、ユーザのブログへの訪問行動の予測結果を調べる。

2. ユーザ訪問行動予測

ブログへの訪問履歴に基づいて、ブログユーザの訪問行動を予測する。訪問行動を予測するため、本研究では Crandall ら [1] が着目した社会的関係と行動類似性という二つの観点に基づき、ユーザ A と繋がりのあるユーザ B が、あるブログ α を見たなら、ユーザ A もブログ α を見ると仮定し、二つの観点による予測結果を比較する。

2.1 社会的関係に基づいた予測

社会的関係とは繋がりが強い関係、つまり影響を受けやすい関係にあるユーザ同士のことである。本研究では、ブログロールのリンクに繋がりのあるユーザ同士を社会的関係と見なし、ブログロールを通じて繋がった他のブログユーザの数を N とする。このとき、社会的関係に基づいた予測では、ブログユーザ A が、社会的関係にある他のブログユーザ K 人がブログ α を見た後、ブログ α を閲覧するかについて予測する。

2.2 行動類似性に基づいた予測

本研究では、類似行動するユーザ同士が、出会う前から少しずつ似た行動をしていると考え、ブログユーザ A は、類似行動を行う他のブログユーザ K 人がブログ α を見た後、ブログ α を閲覧するかについて予測する。この行動類似性に基づいたユーザ訪問行動の予測法では、ブログへの訪問履歴を利用し、ユーザごとに、類似度が高いユーザ N 人を抽出する。N 人としたのは二つの行動予測のパラメータを合わせるためである。また、ユーザ A と類似行動関係にあるユーザ K 人がブログ α を見た後、ユーザ A がそのブログ α を見るかを予測する。訪問行動の類似度はコサイン類似度と Jaccard 係数を用いて算出した。ただし、実験上 Jaccard 係数よりもコサイン類似度が精度的に良い傾向を示すため、紙面の都合上、コサイン類似度の結果を掲載する。

3. 実験

3.1 実験データ

実験データとして Doblog*2 を使用した。今回使用するデータはユーザ数 52,525、ブログロール数 115,552、アクセス数 12,542,581 である。

*1 龍谷大学 理工学部 電子情報学科

*2 株式会社 NTT データ <http://www.doblog.com>.
(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。

3.2 実験の設定

実験では、期間 t1 にユーザの繋がりを調査し、期間 t2 で予測を行う。期間 t1 を 2004 年 1 月から 6 月、期間 t2 は 2004 年 7 月のそれぞれ 1, 2, 3 週目とする。ただし、本稿の予測実験では、メジャーブログへのアクセスのみを考える。また、訪問回数に基づいて標準偏差を算出し、平均値より標準偏差以上の訪問回数があるブログをメジャーブログとした。期間 t1 においてユーザ A と繋がりのあるユーザ K 人がブログ α を見た後、期間 t2 においてユーザ A がそのブログ α を見る確率 P(K) *3 を求め、f(K) を計算した。

$$f(K) = \sum_{k=1}^K P(K)$$

3.3 実験結果と考察

図 1 と図 2 のグラフは、社会的関係及び行動類似性に基づいた予測の結果であり、それぞれ期間 t2 の 1, 2, 3 週目の予測結果を示している。いずれのグラフにおいても、最も精度が優位にあるのは 2 週目であることが分かる。また、社会的関係と行動類似性、共に精度の良かった 2 週目を比較したものが図 3 である。図 3 から、行動類似性に基づいた予測の方が良い精度であることが分かる。以上の結果から、社会的関係に基づいた予測と行動の類似性に基づいた予測、共に精度が良いのは 2 週目の予測であり、また、3 週目になると精度が極端に落ちるといった傾向が示された。

4. まとめ

実験結果で一番精度が良いものはいずれも 2 週目であった。しかし、1 週目と 2 週目の間の期間で実験を行った場合、より精度の高い結果が得ら

*3 $P(K) = U(K) / V(K)$ である。ここに V(K) はユーザと繋がりのある N 人の内 K 人が、期間 t1 にメジャーブログを見た回数、U(K) は期間 t2 にユーザがそのメジャーブログを見た回数である。

れる可能性がある。今後は、t1, t2 の期間を変更し、社会的関係に基づく予測をプログロールではなくコメントやトラックバックなどについても、実験を行い、また他の類似度を用いた実験や、予測精度を向上させる手法の探究も行う予定である。

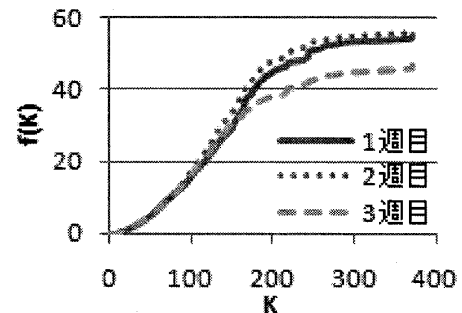


図 1 : 社会的関係に基づいた予測

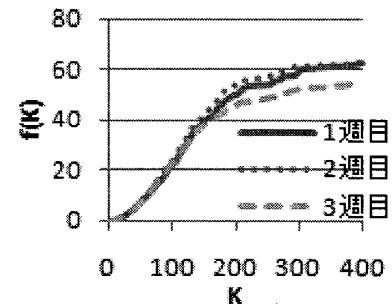


図 2 : 行動類似性に基づいた予測

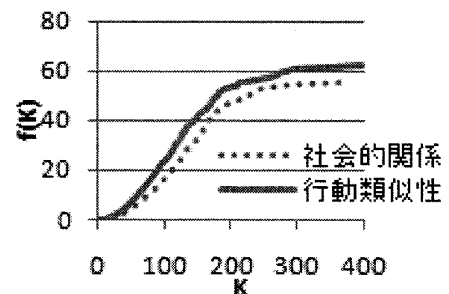


図 3 : 社会的関係と行動類似性の 2 週目の評価
謝辞

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] D. Crandall, D. Cosley, D. Huttenlocher, J. Kleinberg and S. Suri. Feedback Effects between Similarity and Social Influence in Online Communities. KDD2008.