

WWWにおけるユーザの情報収集アクティビティの解析

北海道大学大学院 ○ 米田 貴雄, 吉井 伸一郎, 古川 正志

本研究では、情報収集・編纂ツールであるソーシャルブックマークサービスを対象に、ユーザが収集した情報をリアルタイムに観測するシステムを構築した。ついで、情報収集アクティビティに基づいた解析により、情報収集者としてのユーザの能力を評価する手法を提案する。

1. はじめに

近年、WWWを利用した情報収集法では、他のユーザが収集・編纂した情報から有用な情報を入手するソーシャルブックマークサービスが注目を浴びている。このような情報収集形態においてはパーソナルネットワークの有用性が示されており^{1),2)}、効率の良い情報収集を行うためには情報収集能力の高いユーザを情報源に持つことが効果的であると考えられる。しかしながら情報収集能力の高いユーザを特定するためには、情報の流通経路やユーザのアクティビティを考慮する必要がある³⁾。

本研究では、Web上での情報共有ツールであるソーシャルブックマークサービスを対象とし、ユーザ間のパーソナルネットワーク及びユーザの情報収集アクティビティに基づいた情報収集能力の評価手法を提案する。

2. ソーシャルブックマークサービス

ソーシャルブックマークサービス (Social Bookmark Service, SBS) とは自分が閲覧した Web ページをオンラインでブックマークすることが可能な Web サービスである。SBS の機能には、自分が収集したブックマークを不特定多数のユーザに公開する機能があり、これにより SBS ユーザは他のユーザが収集した情報 (ブックマーク) を容易に入手可能である。ブックマークは人間のフィルタリングがかかった有益な情報を含むので、情報検索に応用した研究が行われている^{4),5)}。

2.1. SBSにおけるパーソナルネットワーク

del.icio.us やはてなブックマークに代表される SBS は、お気に入り機能を持つ。ユーザはお気に入りユーザを指定することで、お気に入りユーザが新しく収集したブックマークを定期的にチェックできる (図 1)。すなわち、お気に入りにユーザが情報源となる。本稿では、ユーザ間のお気に入り関係をパーソナルネットワークとして扱う。このネットワークを $G(V, E)$ と定義する。ここで、ユーザ集合を頂点集合 $V = \{v_i\}$ 、ユーザ間のお気に入りの関係を有向辺集合 $E = \{e_{ij}\}$ とする。辺集合では、情報が流れる方向に有向辺が張られる。

3. 提案手法

ネットワークのリンク構造を解析しノードの重要性を評価する手法として Brin らの PageRank⁶⁾ や Kleinberg の HITS アルゴリズム⁷⁾ が提案されている。

本稿ではこれらのリンク構造解析に加え、ユーザのアクティビティに基づいた評価手法を提案する。

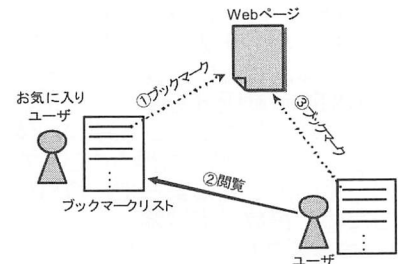


図 1. お気に入りユーザを利用した情報収集

3.1. ユーザの情報収集アクティビティの定義

ユーザ v_i のお気に入りユーザ v_j に対する感度 (Sensitivity) の指標を $S(v_i, v_j)$ と定義する。また、ユーザ v_i のファン (ユーザ v_i をお気に入りにしている) ユーザ v_j への影響度 (Influence) を示す指標を $I(v_i, v_j)$ と定義する。さらに、ユーザ v_i の良質な情報へのブックマークの早さを表す指標を $\alpha(v_i)$ と定義する。良質な情報とは多くのユーザにブックマークされている Web ページである。これらは次式で定義される。

$$S(v_i, v_j) = \frac{\|B(v_j, v_i)\| - \|B(v_i, v_j)\|}{\|B(v_i) \cap B(v_j)\|} \quad (1)$$

$$I(v_i, v_j) = \frac{\|B(v_i, v_j)\| - \|B(v_j, v_i)\|}{\|B(v_i) \cap B(v_j)\|} \quad (2)$$

$$\alpha(v_i) = \sum_{b \in B(v_i)} \left\{ \frac{\text{pop}(b)}{\text{rank}(v_i, b)} - 1 \right\} \quad (3)$$

ここで、各変数は以下のように定義される。

$B(v_i)$: 頂点 v_i のブックマーク集合

$B(v_i, v_j)$: 頂点 v_i, v_j の共通ブックマーク集合

(ただし、 v_i が先にブックマークしている)

$\text{pop}(b)$: ブックマーク b を持つユーザ数

$\text{rank}(v_i, b)$: ブックマーク b でのユーザ v_i のランク

S はお気に入りユーザを情報源としてブックマークする場合に高くなり、 I はファンユーザが同じ情報をブックマークする場合に高くなる。 α は人気度の高いブックマークを早期にブックマークしているほど高い値を与えられ、ブックマークが遅い場合は低い値が与えられる。これらの指標 S, I 及び α の値は、ユーザのブックマークによる情報収集アクティビティにより変動する。

3.2. 情報収集力評価アルゴリズム

本稿では優良な情報収集者の定義を「良質な情報を発見している探索者」、または、「優良な情報源から情報を入手している編集者」とする。そして、以下の仮

定に基づき, 各ユーザの情報探索能力(search)と情報編集能力(edit)を式(4),(5)のように定義する.

仮定 1: 良い編集者 (ファンユーザ) に影響を与えているユーザは良い探索者である.

仮定 2: 良い探索者 (お気に入りユーザ) に対し感度が高いユーザは良い編集者である.

$$search(v_i) = \sum_{v_j \in \Gamma_{fan}(v_i)} \{I(v_i, v_j) \times edit(v_j)\} \quad (4)$$

$$edit(v_i) = \sum_{v_j \in \Gamma_{favorite}(v_i)} \{S(v_j, v_i) \times search(v_j)\} \quad (5)$$

ここで, 各変数は以下のように定義される.

$\Gamma_{fan}(v)$: 頂点 v のファンユーザ群

$\Gamma_{favorite}(v)$: 頂点 v のお気に入りユーザ群

評価値 search 及び edit は式(4), (5)を反復することにより計算される. 以下にアルゴリズムの詳細を示す.

1. あるユーザ間のお気に入り関係より, パーソナルネットワーク $G(V, E)$ を作成する.
2. 各有向辺 $e_{ij} \in E$ に対し, S 及び I を計算する.
3. 式 (3) により評価値 edit を更新
4. 式 (4) により評価値 search を更新
5. 終了条件に達するまで 3, 4 を繰り返す.

4. 実験

提案する情報収集力評価アルゴリズムによりユーザの情報収集能力の評価を行う. 解析対象はオンライン SBS である del.ici.us とする. Web クローリングシステムを構築し, ユーザのお気に入りリストとブックマークリストを取得する.

情報収集力評価アルゴリズムに従い, ユーザ間のパーソナルネットワークを構築する. ついで, ユーザの感度 S 及び影響度 I を計算する. そして, 式(4),(5)の反復により, search 及び edit を計算する. また, 比較対象として α を計算する.

4.1. データセット

2006 年 7 月 13 日~28 日の期間, del.icio.us ドメインをクローリングした. ユーザ数 2,117, ブックマーク数 1,712,982 のデータを使用する. このとき, 1 ユーザ当たり平均ブックマーク数は約 810 であった.

4.2. 実験結果

取得したユーザのお気に入りリストを使用し, 頂点数 2,117, 有向辺数 6,952 のネットワークを構成した. ただし, クローリングをしていないユーザは除いた.

表 1 に優良情報収集者発見アルゴリズムによる search と edit の上位 5 ユーザの値とそのときの α をそれぞれ示す. 全ユーザの α の平均は 1247.98 であった.

5. 考察

表 1 より, search が高いユーザは α が高くことがわかる. すなわち, search が高いユーザは人気度の高いブックマークを早期に発見している傾向にあると言える. 一方, edit が高いユーザは α が低い. そこで, edit の上位 5 ユーザが持つお気に入りユーザの search を調べてみた

表 1. 提案手法による search と edit の値

rank	search	α	rank	edit	α
1	0.0248	7615.5	1	0.0287	2959.0
2	0.0241	11633.8	2	0.0257	651.8
3	0.0132	9735.7	3	0.0229	600.3
4	0.0128	4902.6	4	0.0213	246.9
5	0.0118	8857.6	5	0.0195	713.8

ところ, 多くのユーザの search が高いことを確認した. つまり, edit が高いユーザは情報探索能力が高いユーザを情報源とし情報収集している傾向にあると言える.

以上より, 情報探索能力, 情報編集能力が高いユーザを特定するには, 提案手法が有効であると考えられる.

6. まとめ

本研究では, ソーシャルブックマークにおいて情報収集能力の高いユーザを特定することを目的とし, ユーザの情報収集能力を評価する手法を提案した. 提案手法ではパーソナルネットワーク及びユーザのアクティビティに基づき, ユーザの情報探索者能力 search と情報編集者能力 edit を評価した. その結果, search が高いユーザは早期に情報を発見する能力が高いことを示し, edit が高いユーザは情報探索能力が高いユーザを情報源としていることを示した.

参考文献

- (1) 米田,吉井,古川: 情報発信アクティビティのネットワーク解析, 2006 年度精密工学会(投稿中)
- (2) 濱崎雅弘, 武田英明, 大向一輝, 市瀬龍太郎: パーソナルネットワークを利用したコミュニティシステムの提案と分析, 人工知能学会論文誌(2004).
- (3) Adar, E.: Adamic L.A., "Tracking Information Epidemics in Blogspace," to appear in the 2005 Conference on Web Intelligence (2005).
- (4) Li, W.-s., Vu, Q., Agrawal, D., Hara, Y., and Takano, H.: PowerBookmarks: A System for Personalizable Web Information Organization, Sharing, and Management, Proceedings of The 8th International World Wide Web Conference (1999)
- (5) 濱崎, 武田, 松塚, 谷口, 河野, 木戸出: Bookmark からの共通話題ネットワークの発見手法の提案とその評価, Vol. 17, No. 3, pp. 276-284 (2002)
- (6) S. Brin and L. Page: The Anatomy of the Large-scale Hypertextual Web Search Engine, Proceedings of 7th International World Wide Web Conference (1998)
- (7) J. M. Kleinberg, "Authoritative sources in hyperlinked environment," Journal of the ACM, Vol. 46, No. 5(1999)