

Social Bokkmarking Induced Active Page Ranking

武藤 克弥

富山県立大学 電子・情報工学科

April 30, 2021

背景

近年ソーシャルブックマーキングサービス (Social Bookmarking Service:SBM) によって、ブックマークされた数が Web ページに書かれている情報の有益性を評価する指標となってきた。しかし、時代遅れの古い Web サイトであっても、ブックマーク数が多いと有益な情報とみなされてしまうという問題があった。

目的

本研究ではソーシャルブックマークのタイムスタンプシーケンスに着目することで、今現在の Web ページの価値をモデル化し、従来提案されたページランキングメソッドを改善する。

SBM について

- 興味を持った Web サイトにタグを用いた注釈を与えて分類、共有するサービス

$$b = (p, u, t, A), A = \{a_1, a_2, \dots, a_n\} \quad (1)$$

b:ブックマーク情報, p:ページ, u:ユーザ, t:ブックマーク日時
A:注釈として与えられたタグ集合

従来のブックマークサービスとの改善点

- 従来サービス: Web ページの価値は用語の頻出度、クリックログなどで決まった
→読み手の評価、好みが含まれていなかった
- SBM: ユーザの趣味嗜好・利便性・その他個人的評価ポイントが反映され、SBM ユーザの推奨ページ、SBM 数がページの評価指標となった

以前提案したランキングメソッド (1)

4/13

ページランキングとは

元々は Google が開発. 他サイトからの参照リンクされた量などをもとにその Web サイトの重要度を数値評価してランク付けする.

→前回の研究では web サイトの評価基準としてブックマークを採用したランキング手法 (S-BITS) を提案した

前回の提案手法" S-BITS" について

- ページとユーザ間の 2 部グラフを対象にしたページランキングメソッド
- 誰が何をブックマークしたのかに重点をおき、より専門性がある人がブックマークしたサイトほど、信頼性があり、評価が高い

以前提案したランキングメソッド (2)

5/13

S-BITS アルゴリズム (1)

時間重みの活性化関数を導入

$$TimeWeight(t) = T_0 e^{-\log 2 (t' - t) / t_{(1/2)}} \quad (2)$$

t: ブックマークのタイムスタンプ, t' 評価した日のタイムスタンプ, $t_{(1/2)}$ 半減期, T_0 : 上記関数の上限値

S-BITS アルゴリズム (2)

ユーザの hub 度スコア $u_score_i^k$, ページの信頼度スコア $p_score_i^k$ を導入

$$p_score_i^k = \sum_{b_{ji} \in B} \varsigma_{\lambda}(act_{p_i}) u_score_j^{k-1}$$

$$u_score_i^k = \sum_{b_{ij} \in B} \varsigma_{\lambda}(act_{p_j}) p_score_j^{k-1}$$

以前提案したランキングメソッド (3)

6/13

S-BITS

$$p_score^0 = \{1, 1, 1, \dots, 1\}$$

$$u_score^0 = \{1, 1, 1, \dots, 1\}$$

$$k = 1$$

Repeat

foreach $p_i \in P$

$$p_score_i^k = \sum_{b_{ji} \in B} u_score_j^{k-1}$$

foreach $u_i \in U$

$$u_score_i^k = \sum_{b_{ij} \in B} p_score_j^{k-1}$$

normalize(p_score^k)

normalize(u_score^k)

until $|p_score^k - p_score^{k-1}|_1 < \epsilon_p$

and $|u_score^k - u_score^{k-1}|_1 < \epsilon_u$

return p_score^k and u_score^k

end

図 1: S-BITS アルゴリズム

はじめに

SBM とは

前提案手法

提案手法

性能実験

欠点

- ブックマークは過去の分も残る
- 古い情報でもブックマーク数が多ければ検索結果の上に出てきてしまう
- 現在でも有益で新鮮な情報を持つページを高評価にしたい

改善点

活性度を導入した新たなランキングメソッド (S-BITS*) を提案

以前提案したランキングメソッド (2)

8/13

活性度について

活性度が高いとき：ブックマークされる頻度が一貫して高い Web ページ

活性度が低いとき：ブックマーク頻度が少ない，一時期だけ大量にブックマークされていたページ

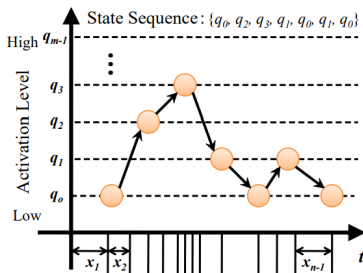


図 2: 活性度遷移のイメージ

活性度について 2

状態 q_i は確率密度関数 $f_i(x_i)$ に従う

$$f_i(x_i) = \alpha_i e^{(-)\alpha_i x_i}, \alpha_i = \hat{g} \beta^i \quad (3)$$

\hat{g} : 平均ブックマーク生成率, β^i : 分解能設定用パラメータ

$$\hat{g} = \frac{1}{\frac{1}{2}N} \sum_{k=\frac{1}{4}N}^{\frac{3}{4}N} \chi_k$$

活性度について3

遷移コストを導入し、ある活性状態 q_i から別の活性状態 q_j に移り
にくくする

→たまたまブックマーク頻度が頻繁になったり過疎化するのを防ぐ

$\tau(j, l)$: j から l への遷移コスト, γ : 状態遷移の起こりやすさ
SD: ブックマーク間隔の標準偏差, N : 今までのブックマーク数

$$\tau(j, l) = |l - j| \gamma \log N \log SD \quad \gamma > 0$$

S-BITS の改善アルゴリズム S-BITS*

S-BITS の式にそのページの活性度, $act(p)$ を導入し, 先ほどのアルゴリズムに適応させる

$$p_score_i^k = \sum_{b_{ji} \in B} s_{\lambda}(act(p_i) + 1) u_score_j^{k-1}$$

$$u_score_i^k = \sum_{b_{ij} \in B} s_{\lambda}(act(p_j) + 1) p_score_j^{k-1}$$

性能実験

$T_0 = 1.0$, $t_{(1/2)} = 30, 60, 120$ とし, それぞれ
S-BITS(30), S-BITS(60), S-BITS(120) として S-BITS* との比較で検
証する. $\beta = 4$, $\gamma = 10$, q_i を -5 から 5 の 11 段階とする.
また, 被験者を 20 人とし, 10 つの検索項目のうち 5 つだけランダム
で検索してもらい, 合計 100 ケースで検証する

iphone, java, ruby, php, web design,
How to write thesis[†], English Learning[†],
Recipe^{††}, Diet^{††}, Project Management^{††}

図 3: 検索キーワード

実験結果

Freshness(情報の新鮮度), informativeness(情報の有益性) とともに S-BITS が他の S=BITS*を上回る結果となった。

