

## Web 閲覧者の視点を考慮した付箋アノテーション間の リンク構造に基づく情報推薦モデルの提案

佐野 博之<sup>†</sup> 伊藤 太樹<sup>†</sup> 柿元 宏晃<sup>†</sup> 平田 紀史<sup>†</sup>  
白松 俊<sup>†</sup> 大園 忠親<sup>†</sup> 新谷 虎松<sup>†</sup>  
名古屋工業大学大学院 工学研究科 情報工学専攻<sup>†</sup>

### 1 はじめに

本研究室では“付箋アノテーションシステム”と呼ばれる、Web ページ上のコンテンツに付箋アノテーションを貼ることを可能にするシステムを開発した [1]。これによりユーザが指し示したいコンテンツをブックマークやソーシャルタグよりも正確に指し示すことが可能になる。ブックマークやソーシャルタグは閲覧者にとって Web ページを特定する手段としては利用可能であるが、任意のコンテンツに対するリンクを実現するには元の Web ページを編集する必要があった<sup>1</sup>。また、本システムを利用することで、付箋付きのコンテンツ間に存在する関連をユーザ間で共有するために付箋に対して双方向リンクを作成することが可能となっている。

しかしユーザによって作成された双方向リンクの数が多くなると、ユーザにとってどのリンクが有用なのか分かりづらくなり、ユーザに混乱を与えるという課題が存在していた。本稿ではこの課題を解決するために、リンク構造に基づいた双方向リンク、および Web コンテンツの推薦を行うための手法について検討する。

### 2 既存の情報推薦モデル

情報推薦システムを実現するために用いられる手法には、コンテンツベースのフィルタリングと協調フィルタリングの 2 つが存在する。コンテンツベースのフィルタリングでは、推薦する情報の内容に着目して情報のフィルタリングを行う。協調フィルタリングでは、ユーザの嗜好情報に着目して情報のフィルタリングを行う。インターネットの大手通販サイトである Amazon.com<sup>2</sup>では、商品の推薦を行うために協調フィルタリングを用いている。

Recommendation Model using Hyperlink Structure of Web Page  
Stickies Considering Web Users' Perspectives

Hiroyuki SANO, Taiki ITO, Hiroaki KAKIMOTO, Norifumi HIRATA, Shun Shiramatsu, Tadachika OZONO, and Toramatsu SHINTANI

Dept. of Computer Science and Engineering, Graduate School of Engineering Nagoya Institute of Technology, Gokiso, Showa-ku, Nagoya, 466-8555 JAPAN

<sup>1</sup>アンカータグを利用することで Web ページ内の特定のコンテンツへのリンクを実現可能であるが、対象 Web ページを編集する必要がある。

<sup>2</sup><http://www.amazon.com/>

近年ではソーシャルブックマークの情報を利用し、Web 上で情報推薦を行う研究が盛んに行われている。具体的には、ユーザがブックマークした Web ページ群からユーザのプロファイルを作成する、Web ページに対して与えたタグをその Web ページを特徴づけるキーワードとみなしてコンテンツベースのフィルタリングを行う、などの手法が用いられている。

### 3 リンク構造に基づく情報推薦モデル

World Wide Web では Web ページが単独で存在するのではなく、Web ページ間がハイパーリンクで結びついていることが重要である。Web における既存のハイパーリンク構造は Web ページ製作者の意図に基づいて作成されたものであり、Web 閲覧者の意図とは異なるものである。すなわち、Web 閲覧者が既存の Web ページ間に自分の視点でリンク構造を作成可能になった場合、Web 閲覧者ごとによって異なるリンク構造が発生することが期待できる。本稿ではそのリンク構造を用いて推薦を行うための手法について検討する。

#### 3.1 Web ページ閲覧者の視点によるリンク構造

Web のハイパーリンク構造は Web ページ製作者の意図に基づき作成されており、Web 閲覧者の視点は考慮されていない。Web ページに対してリンクを埋め込む場合には対象となる Web ページを編集する必要があり、それを行うためには Web ページの書き込み権限が必要である。したがって Web ページ閲覧者が既存の Web ページに対してリンクを埋め込むことは不可能である。

付箋アノテーションシステムが提供する付箋を用いることで、ユーザは付箋間に自由にリンクを作成できる。すなわち、付箋アノテーションによって、疑似的ではあるが既存の Web ページ間に対してユーザが自由にリンクを作成できる。本システムを用いることで、Web 閲覧者が独自の視点に基づいたハイパーリンク構造を容易に表現することが可能となる。

#### 3.2 リンクを用いた推薦

本稿では以降、ユーザ  $X$  が付箋を貼り付けたコンテンツの集合を  $fusen(X)$  と表現する。2 人のユーザ  $A$ ,

入力：付箋を貼り付けたユーザ  $U_a$   
 付箋を貼り付けたコンテンツ  $C_m$   
 リンク先のコンテンツ  $C_n$   
 出力：推薦するコンテンツの配列  $Contents$

```

procedure recommend( $U_a, C_m, C_n$ )
     $Contents[]$ ; //配列変数の宣言
    for each  $U \in Users$  do
        if  $sim(U_a, U_x) > T_{user}$  then
            for each  $C \in fuser(U)$  do
                if  $link(U_x, C, C_n) == TRUE \ \&\&$ 
                     $sim(C_m, C) > T_{contents}$  then
                    push( $Contents, C$ );
            endif
        endif
    return  $Contents[]$ ;
    
```

図 1: 推薦するコンテンツの決定手法

B が共通のコンテンツに対して付箋を貼りつけた場合、すなわち  $fuser(A) \cap fuser(B) \neq \emptyset$  である時には、リンク構造とコンテンツ間の類似度に基づいた推薦を行う。

推薦するコンテンツを決定する関数を図 1 に示す。コンテンツ  $C_x, C_y$  との類似度を  $sim(C_x, C_y)$  とする。類似度  $sim(C_x, C_y)$  はベクトル空間モデルに基づき、コンテンツの特徴ベクトル  $\vec{D}_{C_x}$  と  $\vec{D}_{C_y}$  のなす余弦値によって算出する。特徴ベクトルの各要素は、コンテンツに含まれる単語の tf-idf 値に対応している。ただし貼り付け対象のコンテンツが画像などのテキストデータを含まないコンテンツの場合には、コンテンツ前後のテキストデータを利用する。これは画像などのコンテンツの前後には、そのコンテンツを説明するような文章が記述されているという経験則に基づいている。類似度は 0 以上 1 以下の値を取り、一般的にその値が大きいほどコンテンツが類似していることを示している。

本システムでは協調フィルタリングのためにユーザ間の類似度  $sim(U_1, U_2)$  も計算している [2]。ユーザ間の類似度計算は、先ほど述べたコンテンツ間の類似度と同様に、各ユーザ毎に管理されている付箋集合の特徴ベクトルから余弦値によって求める。余弦値が閾値  $T_{user}$  を超えた場合、それらのユーザが同じ嗜好を持つと仮定した。

$link(U, C_x, C_y)$  はコンテンツ  $C_x$  と  $C_y$  の間にユーザ  $U$  が作成したリンクが存在するかどうかを返す関数である。リンクが存在する場合には  $link(U, C_x, C_y)$  は  $TRUE$  を返し、存在しない場合には  $FALSE$  を返す。

本システムを用いて作成したリンクのイメージは図

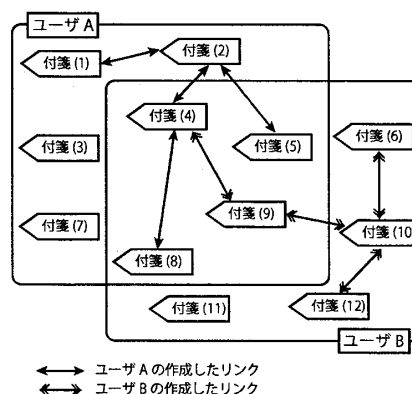


図 2: 付箋アノテーションが形成するリンク構造の例

2 のようになる。この図の付箋 ( $n$ ) はコンテンツ  $C_n$  に貼り付けられていることを意味する。ユーザ A がすでに  $C_2$  と  $C_4$ 、 $C_4$  と  $C_8$  間にリンクを生成済みの状態において、ユーザ B が  $C_4$  と  $C_9$  の間にリンクを生成したとする。その場合には、システムはユーザ B に対して  $C_4$  から  $C_2$  へのリンクを推薦するかどうかを判断するために、 $C_2$  と  $C_9$  の類似度を計算する。計算した類似度が閾値  $T_{contents}$  を超えた場合にそのリンクを推薦する。同様に、 $C_4$  から  $C_8$  へのリンクを推薦するかどうかの判定も行う。

#### 4 おわりに

既存の付箋アノテーションシステムの拡張として、ユーザが作成した付箋アノテーション間のハイパーリンク構造を用いた情報推薦モデルの提案を行った。ユーザは本システムを用いることで、既存の Web ページ間に対してリンクを作成することが可能となる。本稿では Web 閲覧者の視点によって作成されたリンク構造に着目した。協調フィルタリングとリンク構造に基づくコンテンツベースの推薦を行うための手法を提案した。これにより、システムがユーザに対して意味のない双方向リンクを掲示することがなくなり、ユーザにとって付箋アノテーションシステムがより使いやすいものになることが期待できる。

#### 参考文献

- [1] 佐野博之, 浅見昇平, 大園忠親, 新谷虎松, “Web エージェントを用いた Web コンテンツへの付箋アノテーションシステムの実現,” コンピュータソフトウェア Vol.26(2009), No.3, pp.69-77.
- [2] 加藤健太, 佐野博之, 大園忠親, 新谷虎松, “Web ページへの付箋アノテーションを用いたニュース記事閲覧支援システム,” 第 71 回情報処理学会全国大会論文集, 2009 年 3 月.