

協調的な情報を用いた検索ナビゲーション手法 Search navigation method using collaborative information

庄司功樹[†]
Koki Shoji

小林亜樹[‡]
Aki Kobayashi

1. はじめに

Web 上にある膨大な情報の中から目的に沿った有用な情報をより少ない労力で検索できることが求められている。その解決手法の1つに目的に応じてより適した検索サービスを利用することがあるが、1人の検索者があらゆるサービスについて精通している状態は現実的ではない。また検索システムに検索クエリ文を入力して要求を伝える現在の一般的な検索スタイルにおいては、適切なクエリ文を考えつく労力がシステムの利用者にとって負担が大きい。このため、入力されたクエリ文と意味的に適合する結果、あるいは万人受けするような情報を提示するだけでは、利用者の検索意図に十分に答えられず満足させることが難しい。そこで満足度の高い情報に辿り着けるよう、検索行動にかかる労力を減らしてかつ個人のニーズに合わせたアクセス支援を行う推薦システムが有効と考える。推薦方式は、情報の中身に立ち入るコンテンツベースフィルタリングと立ち入らない協調フィルタリングに大分されており、同一情報が利用者へ与える満足度が検索目的によって異なるとの立場から後者を利用する。

本稿では、検索者が目的の情報を得るのにクエリ文を検討・入力し、提示された結果を評価する一連の行動を繰り返して行っているのに着目し、過去に同様あるいは類似した目的を持った検索者が満足したページや検索サービスを推薦することで検索行動にかかる労力を事前に減らし、満足度を高める支援手法を提案する。

関連研究には、ソーシャルブックマークサイトにつけられたタグをクラスタリングしてユーザとの関連性を求めて推薦ページを決定する [1] やアクセスログからユーザの行動を予測してページを推薦する [2] などがある。

2. 提案手法

2.1. モデル化

異なる検索者(ユーザ)が同一の検索クエリ文を入力した際、満足した Web リソースが同一であれば、両者の検索目的が類似していると仮定する。ユーザが検索時に検索システムに提示して欲しい Web ページや検索サービスをクエリ文と URL の組で表現すると、同じ組が多いユーザ間では、検索行動における嗜好が同じ傾向にあると考えることができる。この組をアイテムとし、全クエリ文についてまとめた表以降、関連空間対応表と呼び、これはすなわちユーザの検索行動における嗜好情報である。

嗜好情報の取得には、実際に提案システムを利用して1つのクエリ文に対し、1つの URL をユーザに明示的に関連付けてもらい、プロフィールの1アイテムとしてサーバ上に保存する。プロフィール内のアイテムが1つ以上あるユーザをプロフィールユーザと呼び、現に推薦を受けようとしているプロフィールユーザを特に区別してアクティブユーザと呼ぶことにする。

アクティブユーザがクエリ文を入力すると、まず検索行動の嗜好傾向が類似するプロフィールユーザをプロフィール情報から特定する。このようなユーザを協調ユーザと呼ぶと、続いて協調ユーザが過去に同一クエリ文を入力して満足した情報、つまり過去に関連づけた検索サービスや Web ページをアクティブユーザに推薦して満足度の高い検索行動の実現を期待する。

アクティブユーザと同一アイテムをもっとも多くプロフィールにもつユーザを1次近傍協調ユーザと、次いで多いユーザを2次近傍協調ユーザのように呼称していく。本稿における検索クエリ文とは、空白で連結されている1個以上の検索

語であると想定する。また、検索語は検索に用いる文字列のことを意味し、クエリ文内での検索語の順序に重みを付加しない(例えば、“旅行 京都”と“京都 旅行”を同一クエリ文と見なす)。さらに、ある Web ページに対してどのクエリ文が適切であるかは、実際に入力するユーザにとって認識が異なるため、クエリ文と関連ページの意味合いの適合性は考慮しない。たとえトリッキーなクエリ文が関連付けられたとしても排除せず、あくまでユーザの主観に任せる。

2.2. 協調ユーザの特定

あるユーザとそれに対する協調ユーザ、またそれらの類似度を求めるために定式化を行う。検索語の種類を M とすると、検索語集合 T と検索クエリ文集合 Q は、以下のように定義する。

$$T = \{t_m \mid 1 \leq m \leq M\} \quad (1)$$

$$Q = \{q_j \mid q_j \subseteq T, 1 \leq j \leq J\} \\ = \{\{t_1\}, \{t_2\}, \dots, \{t_M\}, \{t_1, t_2\}, \{t_1, t_3\}, \\ \dots, \{t_{M-1}, t_M\}, \dots, \{t_1, t_2, \dots, t_M\}\} \quad (2)$$

ここで、 Q は T のべき集合 $\mathfrak{P}(T)$ 内から空集合の元を除いた集合と同じになるので、クエリ文の種類 $J = |\mathfrak{P}(T)| - 1 = 2^M - 1$ となる。また Q のうち、 n 元クエリ文集合 r_n を元とする集合を R とすると、その元の数 N は1クエリ文内でユーザが入力できる検索語の上限数を意味し、最大で M (制限無しの場合) となる。さらに r_n は式 (4) のようにクエリ文 q で表現でき、その数は M 個から n 個を選ぶ組み合わせなので $|r_n| = {}_M C_n$ となる。

$$R = \{r_n \mid r_n \subset Q, 1 \leq n \leq N\} \quad (3)$$

$$r_n = \{q_e \mid q_e \in Q, \sum_{k=1}^n |r_k| - |r_n| + 1 \leq e \leq \sum_{k=1}^n |r_k|\} \quad (4)$$

全プロフィールユーザ数を P 、Web リソースの URL の数を K として、それぞれの集合 U と L を次のように定義し、プロフィールユーザ u_p の関連空間対応表内で検索クエリ文 q_j に対応する Web リソースの URL を l_j^p と表す。

$$U = \{u_p \mid 1 \leq p \leq P\} \quad (5)$$

$$L = \{l_k \mid 1 \leq k \leq K\} \quad (6)$$

l_j^p の内容は u_p が q_j に対して指定した場合は、実際に指定した URL (例えば <http://www.example.co.jp/> など)、そうでない場合は関連付けが行ってないことを示す定数 NULL とする。

$l_j^p = \begin{cases} \text{指定した URL} & (\text{ユーザに関連づけられている場合}) \\ \text{NULL} & (\text{ユーザに関連付けていない場合}) \end{cases}$

アクティブユーザ u_a と他のプロフィールユーザ $u_p (p \neq a)$ が実際に関連付けているページ URL のアイテム (プロフィール内アイテム) の集合を L_a, L_p で表すと、互いに同一クエリ文に共通して関連付けている同一 URL のアイテムの集合は $L_a \cap L_p$ で、そのアイテム数を関数 s_i で求めることができる。

$$s_i(u_a, u_p) = |L_a \cap L_p| = |\{l_y^a \mid l_y^a \neq \text{NULL}, l_y^a = l_y^p\}| \quad (7)$$

アクティブユーザの f 次近傍協調ユーザ集合を C_f とすると、その協調ユーザ数は関数 c_u で求まる。ただし、その時に比較

[†]工学院大学大学院工学研究科電気・電子工学専攻

[‡]工学院大学工学部情報通信工学科

する対象ユーザの選択範囲は、 U を全体集合とする集合 W_f に従う。1次近傍協調ユーザ集合を求める場合、アクティブユーザを除いた残りのユーザが比較対象ユーザになる。このユーザから、さらに1次近傍協調ユーザを除くと、2次近傍協調ユーザを求める際の比較対象ユーザ集合となる。

$$C_f = \arg \max_{u_s \in W_f} \text{si}(u_a, u_s) \quad (8)$$

$$\text{cu}(u_a, f) = |C_f| \quad (9)$$

$$W_f = U - \{u_a\} - \sum_{k=1}^f C_k - C_f \quad (10)$$

u_a とその協調ユーザ u_p の類似度は、 $|L_a|$, $|L_p|$, $|L_a \cap L_p|$, $|C_f|$ などを用いて類似関数 $\text{ur}(u_a, u_p)$ で求めるものとする。類似度の指標としては、Jaccard 係数や Simpson 係数などが知られている。

3. 提案システム

3.1. 概要

ユーザから明示的に興味情報、すなわちクエリ文と関連づけしたページ URL 情報を取得するのにアンケートをとる手法も考えられるが、負担軽減と実用性を考慮立場で関連付けの一連動作をウェブブラウザ上で行えるようにする。関連付け動作と結果の推薦は、ユーザがクエリ文を入力してページを評価する一連の行動を繰り返し行う従来の検索スタイルの流れを汲み、その詳細を 3.2 で説明する。

アクティブユーザの要求にあわせて、システムは該当する検索 API から受け取った検索結果と協調ユーザから求めた推薦先を1つのページに合成してランキング形式でユーザに提示し、関連づけしてもらう。ユーザが関連付けたページを次回と同クエリ文に対する応答結果ランキングの最上位として先頭に提示する。ユーザは検索システムを自分専用カスタマイズした感覚を持てるようになり、可読性の向上とアクセスしやすさから、積極的に関連付けてもらう動機になると期待する。実際にユーザが利用する画面構成については、3.3 で説明する。

3.2. アクティブユーザの検索行動

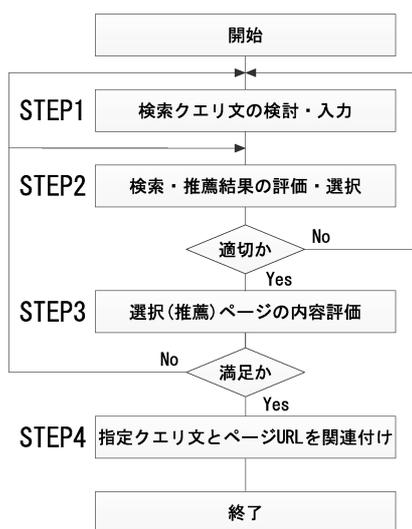


図 1: アクティブユーザの行動プロセス

アクティブユーザの一連の検索行動を開始から終了まで4ステップに分け、図1に示す。

STEP1 ユーザが検討したクエリ文を提案システムに入力する。また、対象の情報を検索できるサービスを1つ以上選択できる。デフォルトでは、Google の Web 検索が自動的に選択される。

STEP2 提示された検索結果のリスト内から、閲覧したいページがあれば選択し、なければSTEP1に戻りクエリ文を再検討する。

STEP3 選択したページの内容を評価し、満足しないならばSTEP4へ、満足したならばSTEP1またはSTEP2に戻って閲覧したいページを再要求する。

STEP4 検索行動の開始から入力したクエリ文の履歴と、閲覧したページ履歴あるいは利用した検索サービス履歴から、クエリ文と満足したページのURLを関連付けて1つの情報要求における一連の検索行動の終了とする。

3.3. インターフェイス

システム設計にあたっての図2: 画面モックを構成する各部分について説明する。まず図2(a)の検索画面にて、ユーザが①検索フォームにクエリ文を入力し、②検索サービスの一覧から利用するサービスを指定できる。③ユーザの関連付け先にはアクティブユーザ自身が同クエリ文に対して関連付けているページへのリンク、④推薦結果には協調ユーザから求めた推薦先ページへのリンクが表示される。⑤検索結果では指定した検索サービスから同クエリ文で検索した結果がサービスごとにグルーピングされた形で表示される。ユーザがこれらのリンクリストから閲覧したいページのリンクをマウスクリックで選択すると、図2(b)関連付け画面へと遷移し、⑩指定したページの内容を閲覧できる。内容に満足したならば、⑥関連付け領域内で⑦検索クエリ文履歴と⑧ページ閲覧履歴の中から選択したクエリ文とページURL(検索サービスのURLも含む)が⑨現在の関連付け状態に表示され、関連付けボタンでシステムに伝えることができる。関連付け先が検索サービスの場合、そのサービスでの検索結果がグルーピングされた形で③にて推薦される。関係を解除するには、関連解除ボタンを利用すればプロフィール内の該当アイテムを削除できる。



図 2: 画面モック

4. まとめ

本稿では検索者にクエリ文と提示して欲しいページを関連付けさせた情報を用い、類似する検索行動を取る過去の検索者の情報から推薦先を決めて満足度の高いページへのナビゲーション手法を提案した。今後、検索者の満足度をさらに高める方向としてクエリ文にあわせて表示画面を提示することも考えられる。また、推薦精度の向上のための類似度を検討する余地が残されている。

参考文献

- [1] 丹羽智史, 土肥拓生, 本位田真一: Folksonomy マイニングに基づく Web ページ推薦システム, 情報処理学会論文誌, Vol.47, No.5, pp.1382-1392(2006)
- [2] 山元理絵, 小林大, 吉原朋宏ほか: アクセスログに基づく Web ページ推薦における LCS の利用とその解析, 情報処理学会論文誌, Vol.48, No.11, pp.38-48(2007)