

グループ支援型 Web 探索におけるナビゲーションのための既閲覧ページ群の同期化提示

Synchronized Browsing of Shared Navigation Histories for Group-Based Web Exploration

伊豆 陸[▽] 中島 伸介[◇]
田中 克己[△]

Atsushi IZU Shinsuke NAKAJIMA
Katsumi TANAKA

Web 情報検索は、単に個人のための情報収集にとどまるものではない。Web コンテンツをグループで共有する状況が考えられる。他のユーザから、推薦された Web コンテンツを共有する状況である。この時、コンテンツが選ばれた背景を知る事は重要である。背景とは、ユーザの Web 検索目的や、類似 Web ページの閲覧数、閲覧した内容の範囲等の情報である。この背景を理解する事によって、ユーザの閲覧がいかに目的を達成しているか、また、推薦されたコンテンツが、いかに閲覧した中から精選されているかについて知る事ができるからである。そこで初めて、推薦ページに対する信頼度を推定する事ができ、他のユーザの探索のために効果的に利用する事が可能になる。本論文では、Web ページが推薦されるまでの履歴から Web 探索の達成度や精選度を計算する事で、ユーザの探索履歴を評価し、その指標を用い、他人の既閲覧ページ群の中から、ユーザの探索を補足する有用なページを発見、提示する手法について提案する。

Recently, Web information retrieval is not only for single user but also for multiple users. Namely, it is the case that a user refer to Web contents retrieved and recommended by other users. In such case, it is important for users to understand the background (navigation history) until finding out the contents. Because users can judge whether the recommended contents are selected carefully and investigated extensively if we understand the background. And when they can understand the background, they can estimate the recommended contents with the background. In this paper, we propose a way to estimate user's recommendation contents through Web exploration using selectivity and achievement. Moreover, we propose a method to discover and provide useful contents for other users.

1. はじめに

WWW の急激な発達に伴い、Web での情報検索は身近なものに

[▽] 学生会員 京都大学大学院情報学研究科修士課程

izu@dl.kuis.kyoto-u.ac.jp

[◇] 正会員 独立行政法人 情報通信研究機構 (NICT)

メディアインタラクシオン G

snakajima@nict.go.jp

[△] 正会員 京都大学大学院情報学研究科

tanaka@dl.kuis.kyoto-u.ac.jp

なりつつある。職場や教育現場でのグループワークにおいて、また Web ショッピングや旅行先決定などで、他人から教えられた Web 情報を参考にする場面は多い。つまり、個人が Web 探索により得たコンテンツをグループで共有するという Web 利用法が求められている。

本論文では、ユーザが Web 探索により得たコンテンツをグループ内に推薦するモデルを想定している。これを現状のメールやチャット等の方法で行おうとすると、ユーザの選択理由や推薦を行うまでの Web 探索範囲が明確でなく、推薦情報の価値がわかりにくい。つまり、グループでの探索という事を考えた場合、他のユーザによって推薦されたページを見ても、自分の探索の手助けにはなり難い点が問題であると考えられる。

協調作業を支援する従来のシステムとしては、いわゆるグループウェアがあるが、Web 探索を対象としたものではない。また、各ユーザのブックマークを共有するような Web サイトも存在するが、タイトルと URL のみを保持するブックマークの共有ではグループを支援する事は難しい。また、ユーザへのコンテンツ提案手法として、ソーシャルフィルタリング [1] や様々な推薦システム [2] もある。しかし推薦される Web ページの関連分野は自分で探さなければならないという問題点がある。また、Web ページ自体を評価するものであって Web 探索行為そのものを評価したものではない。

本論文では、探索結果のみではなく、探索結果に至るプロセスやその行為を含めて保存するコンテキストブックマーク [3] の手法を拡張して用い、ユーザの Web 探索履歴を評価する。探索履歴を利用する事により、ユーザの探索目的や探索範囲を把握し、探索履歴と共にグループ内で共有する。その評価情報を用い、ユーザの探索に同期して、有用な Web ページを共有データ内から検索、提示することで、ユーザの探索を支援する事を目的とする。

2. 閲覧履歴の利用

2.1 グループ支援型 Web 探索

本論文では個人の Web 探索による情報をグループ内で共有する環境をグループ支援型 Web 探索 [4] と呼んでいる。グループ旅行の行き先を決定するために、各人が Web 探索を行うといった状況である。つまり、本論文においては、グループは特定多数を対象としている。グループは、同じ目的を持って探索を行っているユーザ群から明示的に定義する。

本システムとしては、クライアント端末を持つユーザグループと、情報を蓄積、共有するサーバマシンから構成する。各ユーザは Web 閲覧を従来のように行い、他のユーザにも薦めたいと思った情報 (Web ページ) をサーバに送信する。また、その時同時に、履歴情報やクリックしたアンカー情報など、解析に必要な情報もサーバに送る。解析に必要な情報の詳細は後述する。サーバは送信された情報を保存、解析し、グループ内で利用できる環境にするものである。

2.2 閲覧履歴の利用価値

まず具体例として、北海道のスキー場のページが推薦されている場合を考える。スキー場を探しているユーザが、その推薦ページを見た時に、その推薦ページだけで行きたいスキー場を決定するという事は想定し難い。推薦されたスキー場が、他所に比べてどのように良いのかが、明記されていない限り判断しにくいからである。結果、ユーザはその推薦ページ以外のページも見たいと考えても、どこをさらに調べれば効率的なのかは不明である。ユーザは結局、ページの推薦者

と似たような Web 探索を行う事になる。

この問題を解消するため、推薦者の Web 探索の評価情報から、他のユーザに提示すべきページを発見する事で Web 探索を効率化する事が本論文の目的である。そこで、ユーザの Web 探索を評価する尺度として達成度、精選度を以下に定義する。

- (1) 達成度：ユーザの探索目的に沿った Web ページをどれくらい多く閲覧したかを表す指標。
- (2) 精選度：どの程度他の類似 Web ページと吟味されて推薦ページは選ばれたのかを表す指標。

ここで、達成度と精選度を算出するに当たって、ユーザの目的を知る必要が出てくる。そこで、ユーザの目的を表す単語を閲覧履歴から抽出し特徴ベクトル化することを考える。同時に閲覧ページ内の単語も抽出し、特徴ベクトル化する事を考える。

3. 閲覧履歴のトラッキング

3.1 特徴ベクトルの抽出

目的を推定するために、ユーザの閲覧行動から目的を表すと考えられるキーワードを抽出し、特徴ベクトルを作成する。以下では、その手法について説明する。

対象とする履歴としては、ユーザが Web 閲覧を始めてから、検索サイトを利用したり、リンクナビゲーションを行ったりしながら、辿り着いたページを推薦するまでの Web ページ群である。Web 閲覧で用いられる検索キーワードとユーザがクリックしたアンカーの文字列からユーザの閲覧目的を抽出する。

ユーザがクリックし、ナビゲーションが行なわれたリンクアンカーの文字列から形態素解析[5]により単語を抽出し、その単語と出現頻度を要素として取り出し、閲覧においてクリックされたすべてのクリックアンカーについて収集する。また、検索キーワードに出現頻度としての重みをつけ、単語集合に含める。これらの単語集合から単語と出現頻度を要素とした特徴ベクトルを作成する。この特徴ベクトルを"トラッキング特徴ベクトル"と呼び、ユーザが推薦を行うまでの、一連の探索における探索目的を表すものとする。

各閲覧 Web ページについても、トラッキング特徴ベクトルと同様に各閲覧 Web ページの文章から単語を抽出し、その単語と出現頻度を要素として"ページ特徴ベクトル"を作成する。(図1参照)

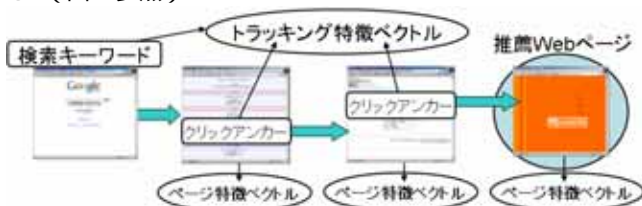


図1 各種特徴ベクトル
Fig.1 Feature Vector

3.2 達成度

前節で、各ユーザが推薦 Web ページを探し出すまでの目的を推定する方法を述べた。また、2.2 節の定義から、ユーザが閲覧目的に沿ったページを多数閲覧した場合に、その探索の達成度は高い。そこで、本論文では、トラッキング特徴ベクトルと、各ページ特徴ベクトルの類似度を用いる。つまり、この類似度の高い Web ページを多く見ている Web 探索は、達成度が高いと考える。よって、達成度(Achievement)を表す

式は、トラッキング特徴ベクトル T と、各閲覧 Web ページのページ特徴ベクトル P の類似度の和を取って、次のようになる。

$$Achievement = \sum_n Sim(T, P)$$

類似度計算にはコサイン類似度を用いる。

3.3 精選度

2.2 節より、ユーザが推薦ページを類似ページと吟味するほど精選度は高い。つまり、ユーザが推薦する Web ページを探し出すまでに、類似したページを多く見てから選んだ場合に、精選度は高いと考えられる。そこで、これらの客観的な類似度をみるために、推薦 Web ページとその他の各閲覧 Web ページそれぞれとのページ特徴ベクトルの類似度を求める。この類似度の和が大きいほど精選度は高いといえる。よって、達成度と同様に考えて、精選度(Selectivity)を表す式は、推薦ページのページ特徴ベクトル rP と、各閲覧 Web ページのページ特徴ベクトル P を用いて、次のようになる。

$$Selectivity = \sum_n Sim(rP, P)$$

この達成度と精選度の二つの基準によりユーザの Web 閲覧を評価する。

4. 既閲覧ページの同期化提示

前節までで、ユーザの目的を推定し、閲覧履歴を評価した。本節では、本論文の目的である、他のユーザにグループ内の既閲覧ページを提示する手法について述べる。ユーザが閲覧を行うにつれて、その閲覧目的の変遷を読み取り、補足ページをリアルタイムで提示する手法についてである。

4.1 ユーザの行動に沿った Web ページの提示

本論文では、他のユーザの閲覧履歴から効果的な Web ページを検索し、提示する。この時、時間と共に変遷していくユーザの探索目的を探索行動から推定する事で、何を提示すべきかの判断を行う。本節では、ユーザの探索行動によって提示するページを変える手法について述べる。ユーザの行動によって以下の3種類のフェーズに分けられる。

1. 内容の似通っていないページをいろいろ見ている時
この時は、ユーザはなかなか目的が絞りにくいと考えるため、ユーザの探索目的に合致した、ユーザの探索の指針となるようなページを発見するよう促したい。そこで、ユーザの目的にあったいろいろなページを提示する"発散"フェーズとする。
2. 類似ページを続けて見ている時
この時、ユーザの探索は気に入った一つの内容のページを発見し、他の似た様なページを見て精選しようとしている状態といえる。よって、一つ目の推薦がなされるまでは、その精選する探索を支援する事が効果的であると思われる。そこで、より探索の精度を上げるために、現在閲覧しているページ群に対して、類似度の高いページを提示する"収束"フェーズとする。
3. ある程度ユーザの考える範囲は調べた時

推薦ページを送った後などにこの状況になる。ユーザが今まで関心がなかった部分であるが、グループ内の探索としては閲覧すべき内容の部分の提示する事で、新たな分野の関心を促す事ができると考えられる。そこで、今まで探索していなかった内容のページを提示する"補完"フェーズとする。

4.2 フェーズの決定

前節述べたフェーズのそれぞれの決定方法について述べる。決定に当たっては精選度(Selectivity)と達成度(Achievement)を利用した S-A グラフを導入する。

図2に探索履歴の精選度と達成度についてあらわすS-Aグラフの実例を示した。ここで、探索中のユーザはまだ推薦を行っていないため、現在の閲覧ページを仮想的に推薦したと仮定した時の精選度と達成度を、ユーザの閲覧するWebページが変わるごとに算出する。そこで、ユーザが現在のページを見るまでのWeb探索に対して、履歴Webページを2次元平面上に配置したもので、履歴ページを赤い点で表している。

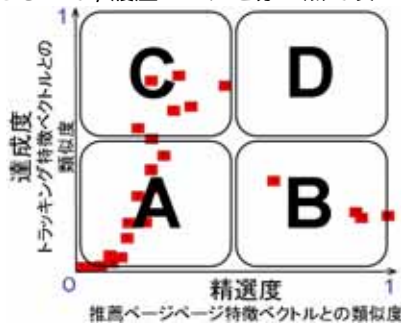


図2 S-A グラフ
Fig.2 S-A Graph

S-A グラフの横軸は閲覧中のWebページと各履歴Webページとの類似度、縦軸はWeb探索におけるトラッキング特徴ベクトルと各履歴Webページの特徴ベクトルの類似度である。このグラフ内の点の分布が何処に偏っているかを見る事により、ある程度ユーザのWeb探索状況を推定する事ができる。つまり、S-A グラフを作成する事により、ユーザのWeb閲覧探索を達成度、精選度の面から評価する事ができ、このグラフを利用することでユーザの行動モデルを推測できる。

ここで、前節で定義した分類に沿って、ユーザの行動をS-Aグラフに当てはめて考える。

1. 発散フェーズ

ユーザの探索行動の初期で、あまり多様なページを閲覧していないため達成度は低く、ひとつの内容に絞ったの取捨選択もなされておらず、精選度も低い状態であると思われる。つまりS-Aグラフの図のAの部分に多くのページが分布している状態である。この時はユーザの目的に関連したページを多く見せてCの部分に分布が移るよう支援する。

2. 収束フェーズ

ひとつの内容に絞り、類似ページを見ようとしているため、精選度が高くなってきた状態であると考えられる。つまり、ACに分布していたページがBDの方に移行してきた状態である。この時は、閲覧ページとの類似ページを見せて、より精選度を上げる(BDでもより右の方へ移る)よう支援する。

3. 補完フェーズ

ある程度自分の中の目的について調べ終わっており、達成度、精選度が共に高い状態であると考えられる。ある程度Dに分布してきた状態である。この時、ユーザが考えるページはある程度閲覧し終わったと思われるので、グループとしての目的ではあるが、ユーザの探索目的ではなかった内容のページを提示する。

4.3 提示ページの検索方法

本節では、各フェーズでの補完ページの検索方法について述べる。検索対象は、グループ内ユーザによってサーバに蓄積されている推薦ページと、それに付随する履歴ページ群である。この一つの推薦ページと、それに対応する履歴ページ群を履歴セットと呼ぶ。

フェーズに関係なく共通する検索方法として、まず履歴セットを検索し、履歴セットの中から適当なWebページを検索

する。ここで、履歴セットの検索方法を説明する前に、達成度の定義を二つに拡張する。2.2節で述べたように、達成度はユーザが考える目的に対して、どの程度の範囲調べているかを表す指針として考えてきた。本節からは、これまで考えてきた達成度を個人達成度とする。これに対して、グループとしての閲覧目的の内、各ユーザがどの範囲を閲覧したか、という指針を導入する。個人の目的に対する個人達成度に対応して、これはグループ達成度と呼べるものである。

[グループ達成度]

3.1節で、各ユーザの探索における個々の探索目的を特徴ベクトルによって表す方法を述べた。ここでグループ全体としての目的を定義する。

本論文では、グループは同じ目的を持ったユーザ群と明示的に定義してあるので、多数のユーザから同じ目的を持つユーザグループを決める必要はない。そこで、グループ内の各ユーザのトラッキング特徴ベクトルを足し合わせる事によって、グループとしての目的を表す。各ユーザのトラッキング特徴ベクトルの次元を合わせ、その要素を足し合わせた特徴ベクトルを"グループ特徴ベクトル"とし、グループとしての目的を表すものとする。よって、各ユーザの探索におけるグループ達成度(Group Achievement)は、グループ特徴ベクトルGと、探索履歴の各閲覧Webページのページ特徴ベクトルPを用いて、個人達成度と同様に次のように定義する事ができる。

$$GroupAchievement = Sim(G, P)$$

4.4 フェーズにおける検索方法

本節で各フェーズにおける検索方法を述べる。

4.4.1 発散フェーズ

図3において、Web空間内のページの分布を模式的に表した。実際のWebはこのように連続的に分布していないが、模式的にWebページの集合であると考えられる。図で、ユーザが考える情報範囲とは、ユーザが自分の探索目的に近いと推定されるWebページ集合を表す範囲である。また、グループにとっての情報範囲とは、グループの各ユーザが考える探索範囲の和である。この図内で、近くにあるページほど類似している。また、より広範に渡るページ集合を見ているほど、目的を達成していると考えられる。



図3 発散フェーズ

Fig.3 Divergence Phase

このフェーズでは、ユーザはまとまりのない範囲を閲覧しているため、ユーザの目的に適合するページを提示する事で発散させる方向性を与える事を考える。図において、閲覧範囲と提示すべきページの関係を表している。具体的な手法としては、トラッキング特徴ベクトル同士の類似度を用いて検索する。この類似度が高い履歴セットは、行動を推定しているユーザと似通った目的で探索が行われた履歴であると考えられるからである。その履歴セットから、ユーザのトラッキングベクトルとの類似度が高いページから提示する

4.4.2 収束フェーズ

前節の図と同様に図4に示した。図にあるように、ユーザ

が類似度の高いページ同士ばかりを見て、探索が収束している時のフェーズである。この時は、現在閲覧しているページとの類似ページを提示する事でユーザの探索の精選度をより上げる事を考える。この時、ユーザの閲覧するページと類似した内容で、多くの類似ページを見た精選度の高い履歴セットから検索をしたい。そこで、検索方法としては、まず、トラッキング特徴ベクトル同士の類似度がある閾値以上の履歴セット集合の中で、精選度の最も高い履歴セットを検索する。その履歴セットから現在閲覧しているページとの類似ページ群を提示する。



図4 収束フェーズ

Fig.4 Convergence Phase

4.4.3 補完フェーズ

図5に示した。図にあるように、ユーザはある程度、自分の目的から考えられる範囲を探索したと思われるフェーズである。この時は、ユーザが考えていた目的からは少し外れているかもしれないが、グループの目的には入っている部分をユーザに提示する事により、ユーザの探索の視野を広げると共にグループの目的を補完する事ができると考えられる。

そこで、4.3.1節で定義したグループの目的を表すグループ特徴ベクトルを利用する。ユーザのトラッキング特徴ベクトルとの類似度が低く、かつグループ特徴ベクトルとの類似度が高い、トラッキング特徴ベクトルを持つ履歴セットを検索する。その中からより吟味されているページ群を提示するために、精選度の高いページ群を検索する。



図5 補完フェーズ

Fig.5 Complementary Phase

4.5 インターフェイス

前節で検索方法について述べたが、検索結果のページのユーザへの提示方法として、どういう順序で、どういうタイミングで提示するかというインターフェイスの問題がある。前節の定義から、検索結果は履歴セットとその中のページがランキング形式で抽出される事がわかる。そこで、検索結果のページをランキング順に提示する。

提示するページ数としては1ページとし、ユーザのブラウザの横に並べて表示する。これはユーザが自身の閲覧に対して補助的なページであり、閲覧と同時に参照できるページ数は、1ページであると考えたからである。

また、ユーザのWeb探索と提示されるページとの関係を理解できるように、S-Aグラフを同時に表示する。S-Aグラフにより、提示されたページがユーザの現在の探索に対してどういった位置づけであるのかを見る事ができると考えられる。この時、4.4節でも述べたとおり、達成度として、個人達成

度とグループ達成度があるため、補完フェーズでは2種類の達成度を軸としたS-Aグラフをそれぞれ表示する。

5. おわりに

本論文では、ユーザの探索を効率化する探索履歴の共有と提示方法について提案した。まず、個人のWeb探索の背景を知るために、グループで共有する際の評価方法として、精選度と達成度という指針を導入した。それら2つの指針を利用してS-Aグラフを作成し、ユーザの閲覧行動の様子を推定し、動的にユーザの行動に応じた補完ページを検索し、提示する方法について提案した。

今後の課題としては、評価実験を行いたい。評価方法としてはユーザの行動に対するフェーズの正確性、提示ページの有用性について検討したいと考えている。

[謝辞]

本論文の一部は、平成16年度文部科学省科学研究費特定領域研究(2)「Webの意味構造に基づく新しいWeb検索サービス方式に関する研究」(課題番号:16016247)、および21世紀COEプログラム「知識社会基盤構築のための情報学拠点形成」による。ここに記して謝意を表す。

[文献]

- [1] U. Shadanand, et al. Social Filtering: Algorithms for Automating 'Word of Mouth', CHI'95, pp.210-217, ACM Press.
- [2] P. Resnick ed. Recommender Systems, CACM Vol.40, No.3, pp.56-89, March 1997. Let's Browse: A Collaborative Web Browsing Agent International Conference on IUI (1999).
- [3] Shinsuke Nakajima, Satoshi Oyama, Kazutoshi Sumiya and Katsumi Tanaka: "Context-Dependent Web Bookmarks and Their Usage as Queries". Proc. of the 3rd International Conference on Web Information Systems Engineering, WISE2002, pp.333-340 (2002).
- [4] 伊豆 陸, 中島 伸介, 小山 聡, 角谷 和俊, 田中 克己: グループ型Web閲覧による探索アクティビティ情報の共有と利用. 第14回データ工学ワークショップ (DEWS2003), 2003年3月.
- [5] 奈良先端科学技術大学松本論文室 茶筌ホームページ: <http://chasen.aist-nara.ac.jp/index.html>

伊豆 陸 Atsushi IZU

京都大学大学院情報学研究科修士課程在学中。2003年京都大学工学部情報学科卒業。日本データベース学会学生会員

中島 伸介 Shinsuke NAKAJIMA

独立行政法人情報通信研究機構勤務。2004年京都大学大学院情報学研究科博士後期課程修了、博士(情報学)。日本データベース学会、情報処理学会、人工知能学会、環境システム計測制御学会各会員。

田中 克己 Katsumi TANAKA

京都大学大学院情報学研究科社会情報学教授。1976年京都大学大学院博士前期課程修了、工学博士。主にデータベース、マルチメディアコンテンツの処理の研究に従事。IEEE Computer Society, ACM, 人工知能学会, 日本ソフトウェア科学会, 情報処理学会, 日本データベース学会等各会員。