

3U-7

パーソナル化を用いた WEB サイト推薦システム

松本 和博[†] 與良 光一郎[†] 内田 理[†] 中西 祥八郎[†]

Web Site Recommendation System Using Personalization

Kazuhiro Matsumoto[†] Koichiro Yora[†] Osamu Uchida[†] Shohachiro Nakanishi[†]

1. はじめに

近年、ブロードバンドの普及に伴い幅広い層にインターネットが浸透している。また、WEBサイトの数も、爆発的に増加している。これら膨大な数の WEB サイトから必要な情報を得るため、インターネット利用者は、主に情報検索サービスを利用して情報検索を行う。しかし、的確な条件を入力しなければ、真に必要なとされる情報が得られない場合がある。また、多くの情報検索サービスは静的な情報検索であり、ユーザによって必要な情報が異なるにもかかわらず、誰が検索を行っても同一の検索結果が表示される。結果として、現在の情報検索サービスは、検索に費やす時間が多く掛かり、ユーザの負担が大きいのが実情である。また、ユーザが欲しい情報や閲覧したいサイトまでたどり着かない事も多い。

そこで本研究では、パーソナライゼーションを用いた WEB サイト推薦システムの構築を試みる。本システムの利用によって、ユーザは検索を行うことなく自分の趣味・嗜好に適したサイトの閲覧が可能となる。

2. WEB サイト推薦システム

2.1 現在の情報検索

現在、情報検索サービスを利用する際には、目的の情報を得るために、ある程度の知識が必要とされる。検索されてきた情報を確認する必要があるからである。しかも、そのほかのサイトとも見比べてその情報の真偽をも確認しなければならない。このようにして個人での情報検索は情報の検索だけでなく、似たようなサイトの情報との検証を行いながら、どのサイトが一番自分の目的に合うのかを見極める必要がある。また、検索されてきた情報が大量であることも稀ではなく、目的の情報が後半にあった場合に、それを見つけることは困難である。これらの改善方法として、最も代表的な検索サイトである Google [1]では、PageRank [2]と呼ばれ

るページの優先度付けに基づいて検索結果を表示している。PageRank は、「多くの良質なページからリンクされているページは、やはり良質なページである」という考え方に基づいている。しかし、これではリンクを拒絶しているサイトがいくら優秀であったとしても上位に反映されることはなく、下位に位置付けられてしまう。Google ではさらに、あらかじめ興味のある分野を登録しておくことで検索をパーソナル化するサービスの提供を開始している。このサービスは、興味があるカテゴリーに属しているサイトを上位に表示させる方法である。しかし、それぞれの情報がどのカテゴリーに属すかなどが曖昧であり、ユーザ毎に認識が違っても稀ではない。

2.2 パーソナル化を用いた WEB サイト推薦

本研究では、パーソナル化を用いた WEB サイト推薦システムの構築を試みた。システムの概要図を図1に示す。本システムを利用する際、各ユーザは、まず自分で WEB サイトを登録する。登録する際は、WEB サイトの URL と評価を入力する。システムは、保存しているユーザ毎の WEB サイトのデータを比較し、各ユーザに自分以外のユーザが所持している WEB サイトのうち、比較対象のユーザがある一定以上の評価をしている WEB サイトを推薦する。

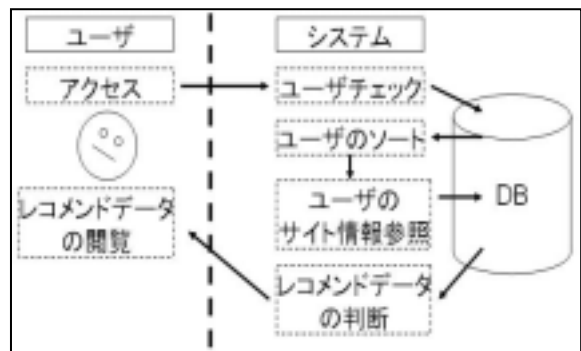


図 1. システム概要図

本システムにおけるサイトの評価は 5 点満点とし 5 点が最も好ましいサイトということにする。具体的には、以下のルールを用いて WEB サイトを推薦する。

[†]東海大学電子情報学部情報科学科
〒259-1292 神奈川県平塚市北金目 1117
School of Information Technology and Electronics,
Tokai University

- (1) DB から自分以外のユーザを探し出し、WEB サイトのデータ数で降順にソートをする。
- (2) 情報が一致しているユーザを優先させるため、自分と相手が所持しているWEB サイトの一致率を計算し、降順でソートする。図2のように優先順位を決める。ここでは上位にあるユーザからA, B, C...とする。
- (3) 推薦元となるユーザの優先順位が決定したら、最初の推薦元のユーザ A と推薦を受けるユーザとでデータの共通していないデータを取り出す。そのときに推薦元となるユーザが WEB サイトに対して閾値以上の評価をつけている場合にそのデータを推薦する。同様にして、次の推薦元のユーザ B と推薦を受けるユーザとユーザ A から推薦を受けたWEB サイトデータとで、共通していないWEB サイトのデータを取り出す。以下同様にしてユーザ C との推薦はユーザのデータとユーザ A, B から受けた推薦の情報とでデータの抽出を行う。

dandhiの場合			
hirakazu	サイ	hirakazu	一
omatsu	イト	ishihara	致
ishihara	数	moto	率
moto	ソ	omatsu	ソ
pooh	ン	pooh	ソ
yoshimura	ト	totu	ト
totu	ト	yoshimura	ト

図2. ユーザの優先順位

WEB サイトを推薦するにあたり、各ユーザに適したサイトを推薦することが望ましい。今回、本システムを評価するため、次のような実験を行った(被験者は8名とした)。まず、著者がランダムに選出した50サイトから各被験者の趣味・嗜好が特に合うサイトとして30サイトを選出してもらい、5点満点で評価してもらう。その後、残りの20サイトがどのように推薦されてくるかを考察した。

表1に、自分以外のユーザ7名から推薦されてきたWEBサイトの平均評価を示す。

2.3 WEB サイト推薦の優先度

このシステムで問題になってくるのが、WEBサイトの推薦順位である。今回の方法で上位にくる推薦サイトは一致度の高いユーザが所持しているサイトである。さらにその中で優先されるWEBサイトの推薦データは、一致度が高いユーザが4点以上をつけているサイトである。

表1. ユーザごとの推薦後の評価(平均点)

	dandhi	hirakazu	ishihara	moto	omatsu	pooh	totu	yoshimura
dandhi		2.48897	3.33333	2.77778	2.85714	3.2	2	3.33333
hirakazu	2		1.6	2.11111	2.2	2.42857	2.28571	2.4
ishihara	2.4	2		2	1.57143	2.4	2.28571	2.33333
moto	2	2.2	2		2	2.14286	2.28571	2.25
omatsu	2.2	3.4	3	3.33333		3.4	2.85714	2.77778
pooh	3	3	3.14286	3.25	2.66667		3.2	3.33333
totu	1.75	1.88889	2	1.85714	1.57143	1.25		1.42857
yoshimura	2.25	2	3.4	3.16667	2.6	3	3	

3. システムの動作結果と考察

パーソナル化を用いることで、ユーザ毎に適したサイトを提供することができたといえるが、不必要なデータも含まれていた。今回の実験では、自分以外のユーザが3点以上の点数をつけたサイトを推薦した場合、再現率は100%であった。登録ユーザ数を増やすことによって、4点以上あるいは5点といった高得点サイトのみを提供でも高い再現率を維持することが可能であると思われる。表2に相手が4点以上つけたサイトを推薦した場合の再現率を示す。ただし推薦されてくる20のサイトのうち高評価をしているサイトがないユーザは除いている。

表2. 高評価をつける予定のサイト再現率

ユーザ名	20サイトのうち高評価をつけるサイト数	他のユーザから推薦された高評価サイト数	再現率
dandhi	6	6	100%
ishihara	2	1	50%
omatsu	8	7	87.50%
pooh	5	4	80%
yoshimura	6	5	83.30%

4. まとめと展望

本研究によって、ユーザは検索を行うことなく、自分の趣味嗜好にあったWEBサイトを閲覧することが可能となった。今後の課題としては、より効率的な推薦アルゴリズムの検討や、類似研究([3]など)との比較などがあげられる。

参考文献

- [1]Google: <http://www.google.com/>
- [2] L. Page et al.: The PageRank Citation Ranking: Bringing Order to the Web, <http://www-db.stanford.edu/~backrub/pageranksub.ps>
- [3] 根元 潤, 遠山 元道: 閲覧履歴に基づく情報検索の相互支援, DEWS2004, 3-B-02, 2004.