

プッシュ型とプル型を統合したパーソナライズ 情報配信システムの提案と実装

神場 知成[†] 坂上 秀和[†] 古関 義幸[†]

インターネット上の情報配信で、WWW アクセスを自動化した新しい形態はプッシュ型と呼ばれることが多く、これと比較した場合、従来の WWW ブラウザによるアクセス形態はプル型と呼ばれる。また、以前から使われてきた電子メール型の配信もある。プッシュ型とプル型、および電子メール型はそれぞれ利点欠点があり、ユーザがその時々に応じて使い分けられることが望ましい。さて一方、アクセス可能な情報量が非常に多くなってくれば、プッシュ型、プル型、電子メール型にかかわらず、各個人ごとに重みづけされた情報提供、つまりパーソナライズされた情報提供を受けることが有効になるが、これまでのパーソナライズシステムはプッシュ型、プル型、電子メール型といずれかのアクセスに限定されており、アクセス方法にかかわらず同じパーソナル化サービスを受けられるようなものはなかった。本論文では、プッシュ型の WWW アクセス、プル型の WWW アクセス、電子メール型をきっかけにした WWW コンテンツ配信という 3 つの形態において、同じユーザプロファイル、同じ手法を用いて自動パーソナライズした情報を配信するシステムを提案し、その実装方法を示す。

Proposal and Implementation of a Personalized Information System that Integrates Push-style and Pull-style WWW Access

TOMONARI KAMBA,[†] HIDEKAZU SAKAGAMI[†] and YOSHIYUKI KOSEKI[†]

The push-style is a new method to deliver information on the Internet, whereas conventional access using a WWW browser is called the pull-style. People often use email in addition to these. Since each of the push-style, pull-style, and email has advantages and disadvantages, users should be able to choose any of them according to their conditions. In addition, when the users get a large amount of information, personalization is useful for filtering information regardless of the access style. However, on the conventional systems, users could get personalized information either with push-style or pull-style, but not with both. In this paper, we will propose and show the implementation of a personalized information system that provides the same personalization mechanism with push-style WWW access, pull-style, and email.

1. はじめに

インターネット、特に World Wide Web (以下 WWW) はマルチメディア文書提供のプラットフォームとしてますます広がりつつある。これにともない、インターネットのユーザが情報を入手する方法も多様化しつつある。以前はほとんどの情報を電子メールで入手していたのに対し、ブラウザによる WWW アクセスが可能になり、さらに最近では WWW 上の情報が自動的にユーザ端末に配信・表示される、いわゆるプッシュ型 (自動配信型) のアクセスも可能になった^{1),2)}。

1 人のユーザが状況によっていろいろなアクセス方法を使い分ける可能性も高いので、どのような方法でも同じ情報が見えることが望ましい。たとえば、社内では専用線で WWW につながっているのでプッシュ型で画面上にたえず最新情報を表示し、自宅ではダイヤルアップ回線なのでプル型、つまり WWW ブラウザによるアクセスで情報を閲覧したいという場合などがあるだろう。

さて一方、情報が増えてくるとユーザはその中から本当に自分がほしいものを取り出すことが非常に困難になるため、プッシュ型、プル型にかかわらずパーソナライゼーション (情報を個人の興味に基づいてカスタマイズし、ユーザがほしい情報を容易に見つけられるようにすること。以下、動詞的に使う場合は「パー

[†] NEC C&C メディア研究所

C&C Media Research Laboratories, NEC Corporation

ソナライズ」と記す。)が重要になる。一般に、パーソナライズするためにはシステムがユーザの興味傾向を何らかのフォーマットで記憶していることが必要であるが(以下、このように形式的表現で記憶されたユーザの興味傾向をユーザプロフィールと呼ぶ)、興味は変動するし単純なキーワードで表現できない場合も多いから、ユーザが自分の興味をいちいちシステムに登録するのはわずらわしい作業である。このため筆者らはユーザの操作履歴に基づきシステムが自動的にユーザの興味傾向を獲得し、それを反映してパーソナライズした画面表示を行うシステムの研究を行ってきた。その手法をインターネット上のパーソナル電子新聞に適用し^{3),4)}、その効果を確認している⁵⁾。

本論文は、上記のパーソナル電子新聞を拡張して実装した、自動パーソナライズ可能でかつ、プッシュ型、プル型、および電子メール型という3つの利用方法を統合したパーソナライズ情報配信システムに関するものである。従来、同じコンテンツをユーザが上記3つの配信形態を適当に併用して受けとることは可能であったが、それらを同じユーザプロフィールでパーソナライズできるものはなかった。また、後述するように「自動パーソナライゼーション」を行う場合はその問題はさらに複雑になる。筆者らが開発したシステムは、ユーザがこの3つの配信形態のうちのどれを用いても同じユーザプロフィールでパーソナライズした情報が表示され、また、どの形態を用いてもシステムが自動的にユーザの操作履歴を参照して、つねに最新の興味傾向を反映するようにユーザプロフィールの更新を行う。ユーザから見ると、その時々自分の都合の良い方法で情報を閲覧することができ、つねにユーザプロフィールおよびパーソナライゼーションの方法に一貫性が保たれている。以下ではプッシュ型とプル型に関する既存サービス・システムの動向、パーソナライゼーション技術の動向、筆者らが実装したシステムにおけるプッシュ型、プル型、電子メール型の実現方法などについて述べる。なお、一般にWWW上でパーソナライゼーションを行うシステムにおいて、各ユーザのプロファイルはサーバ側に置かれる場合とクライアント側に置かれる場合とがある。それぞれの場合の得失はすでに議論されているが⁶⁾、本論文で述べるシステムではユーザプロフィールをサーバ側に保存しており、以下の議論も同様の場合を前提としている。

2. 関連技術の動向と本論文で扱う課題

2.1 プッシュ型、プル型、電子メール型

最近、インターネット上のサービスにおいてプッシュ

型とプル型という用語が使われるようになった。プル型は従来のWWWアクセス方法であり、WWWブラウザ上でユーザがURL(Uniform Resource Locator)のアドレスを入力したりハイパーテキストのリンクを選択したりしてコンテンツを見る手法である。ユーザがこれらの操作をすると、ブラウザからWWWサーバに対してHTTP(Hypertext Transfer Protocol)に基づくリクエストが送信され、WWWサーバがHTML(Hypertext Markup Language)で記述されたコンテンツを返す。ブラウザはHTMLを解釈して文書の表示を行う。

これに対しプッシュ型とは、コンテンツの自動配信である。ユーザから見ると、いったん配信のリクエストをしてしまえば、その後は適当なタイミングで自動的に自分の端末にサーバからコンテンツが配信される。技術的には、現在主流のプッシュ型は、端末上のクライアントソフトが自動的にかつ定期的にサーバにアクセスして文書をダウンロードする手法をとっている¹⁾。つまり実装上はプル型であるが、それが自動的に行われるために、ユーザが毎回能動的に情報源へのアクセスを指定する必要がない。プッシュ型の中にも、このような自動プル型ではなく情報が発生した時点でサーバからクライアントに情報を送信するという、本当の意味でのプッシュ(リアルプッシュと呼ばれることが多い)システムもあるが(<http://www.tibco.com/>等で参照可能)、本論文ではそれに関しては詳述せず、特にことわりがない限りプッシュ型とは自動プル型を指すものとする。また、プッシュ型では端末上へのコンテンツ表示方法として、スクリーンセーバやテロップ(画面の一部に電光表示をすること)が使われることが多い。つまり、プッシュ型では単にコンテンツを自動配信だけでなく、絶えず画面上に最新表示を行って情報がユーザの目に入りやすいようなユーザインタフェースが採用される場合が多い。プッシュ型の代表的な例としては、Pointcast Network(<http://www.pointcast.com/>で参照可能)がある。

上述のようなプル型とプッシュ型に加え、技術的にはかなり以前からあるが最近特に情報配信に広く使われるようになったものとして、電子メール型がある。例としてインプレス社のインターネットウォッチは、定期購読契約をしていれば毎日新しい情報が送信されてくる。このような電子メール型の情報配信は、サーバから積極的に情報が端末に送られてくるという意味では、前述のリアルプッシュ型の重要な一例である(ただし、「リアルプッシュ」という用語は、正確にはマルチキャスト技術を用いた同報通信に対して使われる

ことが多く、電子メールのようにユニキャスト技術に基づくものに対しては使われない場合が多いようである。このあたりの用語の利用法は完全には定まっていない。

2.2 パーソナライゼーション

ここではいうパーソナライゼーションとは、情報を個人向けにカスタマイズして提供することである。個人個人にとっての必要性や興味に応じて情報をフィルタリングしたり、フィルタリングした結果の情報をさらに重要度に応じて見せ方を変えたりする。これらは個人の日常作業をある程度自動的に支援するエージェント技術の一例である⁷⁾。

パーソナライゼーションは、インターネットの普及にとともないますます重要性を増してきた。個人がネットワーク上にある莫大な量の情報にアクセスできるようになった反面、あまりにも情報量が多くその中から本当に必要な情報を見つけ出すことが難しいという現象が生じてきたためである。検索エンジン（たとえば <http://altavista.digital.com/> で利用可能）は手に入りたい情報が明確に定まっている場合は有効であるが、たとえば新聞を毎日読む場合のように「特定の情報を探しているわけではないが、適当に興味を引かれる情報を読む」という目的には不向きである。パーソナライゼーションはそのような目的に適している。

パーソナライゼーションにおける技術的な課題としては、大きく分けて2つある。1つはいかにしてユーザプロフィールをシステムが獲得するかという問題であり、もう1つはどのように情報をパーソナライズしてユーザに提示するかという問題である。ユーザプロフィールの獲得に関して最も単純な方法は、ユーザが自らキーワード等を登録するものである⁸⁾。この方法では、正確な興味を表現するためにはユーザの負担が大きいので、それを軽減する方法として、ユーザが文書に対する自分の興味を点数化して与えると、システムがそれをもとにそのユーザの興味を学習するものや⁹⁾、ユーザが各文書を読むのに費やしている時間からそのユーザの興味傾向を推定する試みなどがある¹⁰⁾。また、それらの問題点を踏まえたくて筆者らのパーソナル電子新聞（プル型）においては、ユーザのマウスクリックなど記事購読操作の履歴から自動的に興味傾向を抽出する手法を検討した⁵⁾。

情報をパーソナライズして提示する方法に関しては、ユーザ興味に応じて不要な情報を間引くもの（フィルタリング）や、情報の提示順序を変えてユーザが興味ある情報を発見しやすくするもの（ソーティング）などがある。ただし、フィルタリングにおける情報の間

引きは、ソーティング結果である順序づけされた情報の、上位のものだけを選択した結果と考えれば、両者の間に本質的な差はない。筆者らのパーソナル電子新聞においては、各記事の重要度に基づいて紙面レイアウトを自動編集するという手法をとった^{3)~5)}。

一般にパーソナライゼーション技術はプル型のサービスと比較してプッシュ型のサービスではさらに重要になる。プル型のサービスは、基本的には「ユーザが能動的に情報を取りに行く」というモデルに基づいていたのに対し、プッシュ型では情報がユーザの端末に自動的に配信されてくる。適切なパーソナライゼーションが行わなければ、ユーザの端末には莫大な量の無駄な情報が送信され、ネットワークの負荷、端末上のディスクスペースなどを圧迫するとともに、配信された情報の中からユーザが必要なものを探すのが著しく困難になる。

2.3 本論文で扱う課題

本論文で扱うのは、同じ情報をプッシュ型、プル型、さらに電子メール型というインターネット上の3つの基本的なサービスで提供する際に、ユーザがどのサービスを受けるかにかかわらず、共通の手法・共通のユーザプロフィールに基づく自動パーソナライゼーションを実現する手法である。つまり、上記3つのサービスにおける自動パーソナライゼーション統合の手法である。自動パーソナライゼーションではなくもっと単純なパーソナライゼーション、たとえばキーワード登録によるパーソナライゼーションを行っている場合には、これらのサービスを統合することは特に難しいものではない。キーワードが登録されていれば、プッシュ型、プル型、電子メール型いずれのサービスで情報を提供する場合でも、まずサーバ側でもとの情報をキーワードでフィルタリングし（たとえば特定のキーワードにマッチする情報だけに絞り込む）、フィルタリング後の情報だけを送ればよい。しかし、自動パーソナライゼーションでは問題は複雑である。自動パーソナライゼーションでは、（ユーザが意識しているかしていないかにかかわらず）ユーザからのフィードバックによりユーザプロフィールを自動更新する必要があるが、ユーザ側の情報閲覧方法が異なれば当然、フィードバックの方法も異なる（図1）。したがって、それぞれの受信方法においてどのようにユーザ操作のフィードバックし、それをどうやって共通のユーザプロフィールに反映するかが問題となる。

一般的な傾向として、今後インターネット上のサービスにおいてパーソナライゼーションは次第に自動化されていくであろう。サービスの種類が増加していく

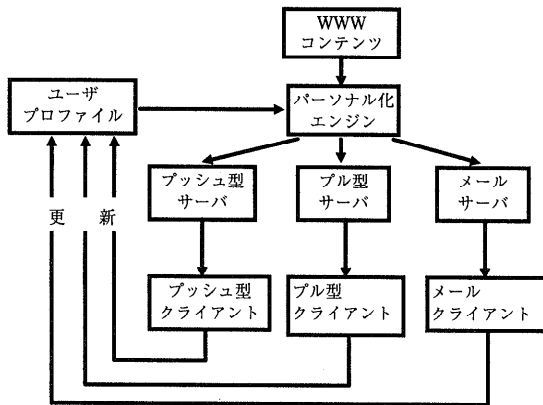


図1 プッシュ型, プル型, 電子メールの統合

Fig. 1 Integration of push-style, pull-style access and email.

際に、いちいちユーザが自分でユーザープロフィールを登録するのはわずらわしいからである。同じユーザープロフィールを複数のサービスで共用するという方法もさほどユーザの手間の軽減にはならない。なぜなら、一般に異なるサービスは異なる会社などによって運営されていてユーザープロフィールの共用はしたくない、あるいはできない場合が多いという商業上の問題だけでなく（このような問題は標準化が強制的に進められれば解決する）、たとえば「新聞情報提供」「特許情報提供」「音楽情報提供」というようなコンテンツが異なるサービスにアクセスする際に、ユーザが求めている情報の内容はまったく異なるからである。新聞提供サービスを受ける際に「天気」というキーワードを登録したとしても、音楽情報提供サービスを受ける際にはまったく役立たない。このような理由から、それぞれのサービスの中で、できる限りユーザに負担をかけないでパーソナライゼーションをすることが重要である。したがって、異なる配信形態における自動パーソナライゼーションの統合は非常に重要である。

3. パーソナル電子新聞におけるプッシュ型、プル型および電子メール型の統合

本章ではプル型、プッシュ型、電子メール型それぞれのサービスにおいていかに統一した自動パーソナライゼーションを実現したかを述べる。

3.1 プル型のパーソナル電子新聞

前述したように、筆者らはすでに WWW 上のパーソナル電子新聞を実現している。これは、ユーザがブラウザで WWW サーバにアクセスした際に表示される電子新聞をパーソナライズするもので、プル型のサービスである。基本的には以下のような方法でパー

ソナライズを行っている。なお、ここでは自動パーソナライゼーションの統合を説明するのに必要な最小限の情報だけを記すのでシステムの詳細およびその評価結果は他文献を参照されたい^{3)~5)}。

- ユーザは、電子新聞の利用開始時にユーザ登録を行い、ユーザ ID とパスワードを設定する。
- サーバ内では、すべての記事の特徴ベクトルが自動的に抽出される。特徴ベクトルは、記事内に出現する全単語の中から一般的に頻出し記事の特徴には関係ないと思われる単語（あらかじめ不要語辞書に登録されている）を除いたものである。
- ユーザの興味傾向を示すユーザープロフィールは、単語と重みのペアを複数集めた集合である。
- サーバには、ユーザープロフィールの学習エンジンと記事スコアの子測エンジンがある¹¹⁾。学習エンジンは、複数の記事に対してユーザが点数づけ（0～100点）を行った場合にそのユーザのプロフィールを学習するもので、子測エンジンは新しい記事がある場合に、その記事に対するユーザの興味を程度を子測（0～100点）するものである。これらの学習と子測は記事の特徴ベクトルをもとに行われる。
- システムは、ユーザからアクセスがあると、サーバ内にあるすべての記事の特徴ベクトルとユーザープロフィールを参照して、上記子測エンジンがすべての記事のスコアを計算する。次にレイアウトエンジンがそのスコアを反映した紙面レイアウトを作成する。

- 表示された記事（記事タイトルだけ、あるいは記事本文の一部しか見えていない）をユーザがクリックして別ウインドウに表示した場合には、ユーザがその記事にある程度興味を持ったものとして、その記事の子測スコアに一定の点数を加算し（現在のシステムにおける加算点数は10点）、それを学習エンジンに渡す。ユーザが何も操作をしなかった記事はユーザは興味を持たなかったものと見做し、記事スコアから一定の点数を減算して（現在のシステムでは5点）学習エンジンに渡す。このようにして、ユーザープロフィールの中では、ユーザがよく読む記事に共通に含まれる単語の重みが徐々に増加し、逆に読まない記事に含まれる単語の重みは徐々に減少する。ユーザが数日間本システムで記事を読んでいると、システム内にはユーザの興味傾向を反映したユーザープロフィールが形成される。

上記のパーソナル電子新聞は、読売新聞社が記事を提供し NEC がパーソナライゼーションを行うシステムとして1998年1月現在、インターネット上（<http://pnews.cplaza.ne.jp/>）で公開されている。画面の1例を図2に示す。

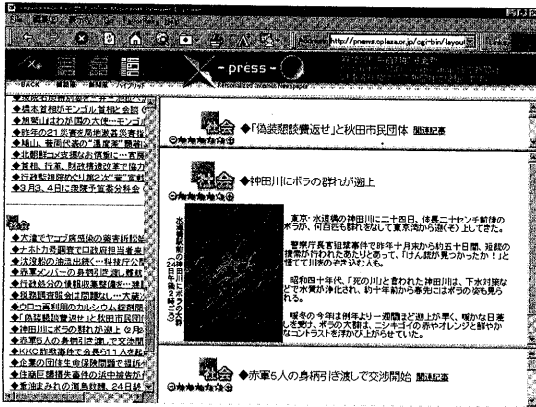


図2 プル型のパーソナル電子新聞の画面例

Fig. 2 Screen shot of a pull-style personalized newspaper.

3.2 プッシュ型パーソナル電子新聞 ANATAGONOMY/SS

本節で述べるプッシュ型のパーソナル電子新聞では、前節で述べたプル型のシステムと同じ学習エンジン、予測エンジンを用いてパーソナライズを行う。しかしユーザの利用方法は異なる。プル型では、記事を読みたいときには毎回ユーザがブラウザ上で WWW サーバへのアクセスを指定しなければならないのに対し、プッシュ型ではユーザがいったん記事購読を登録してしまえば、記事は定期的かつ自動的にクライアント端末上にダウンロードされる。記事ダウンロードおよび画面上への記事表示を行うのは、クライアント上で動作するアプリケーションプログラムであり、スクリーンサーバとしても機能する。以下にプッシュ型パーソナル電子新聞 ANATAGONOMY/SS の概要を述べる。これに関してもここでは本稿の主題を述べるのに必要な最小限の情報を記し、詳細は稿を改めて述べる予定である^{12),13)}。

3.2.1 システムの構成

システム構成を図3に示す。サーバ側では前節で述べたプル型のシステムと同じ学習エンジン、予測エンジン（総称してパーソナライゼーションエンジン）を利用する。システム構成上プル型との最大の相違は、クライアント側にあるプログラムが定期的にサーバにアクセスして記事ファイルをダウンロードし、それを順次自動的に画面上に表示することである。記事ダウンロード等を担当するモジュールをレシーバと呼び、画面表示等を担当するモジュールをビューアと呼ぶ。なお、記事ダウンロードのタイミング（何時間、あるいは何分おきにアクセスをするか）はユーザが設定可能である。

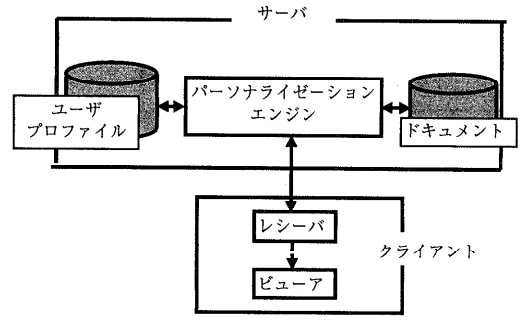


図3 プッシュビューアのシステム構成

Fig. 3 A system architecture of our push-viewer.

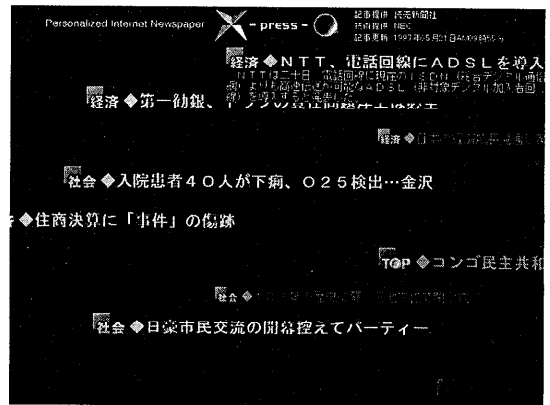


図4 ANATAGONOMY/SS 画面例

Fig. 4 Screen shot of ANATAGONOMY/SS.

3.2.2 パーソナライズした表示の作成方法

ユーザは ANATAGONOMY/SS 利用の登録時に、プル型のパーソナル電子新聞で設定したのと同じユーザ ID およびパスワードを登録し、クライアント上のレシーバはこれを記憶している。レシーバがサーバにアクセスする際にはこのユーザ ID とパスワードをサーバに渡し、サーバ側ではそのときに表示すべきすべての記事に対してユーザの興味を予測し、スコアづけを行う（0~100点）。ビューアはユーザが自分にとって興味のある記事を簡単に見つけることができるように、各記事のスコアを反映した表示を作成する。画面例を図4に示す。複数のレイアウト方法があるが、この図に示した例ではスコアの高い記事ほど画面上方に表示されている。記事タイトルは画面の右から左側に向かって常時流れており、ユーザが何も操作をしなくてもすべての記事タイトルが表示されるようになっている。

3.2.3 自動パーソナライゼーションのためのフィードバック方法

画面に記事が表示されているときに、ユーザが記事タイトルをクリックすると記事本文が表示される。ANATAGONOMY/SSではユーザがこの操作を行った場合にはユーザがその記事に興味を持ったものと判断し、サーバに対してそれを通知している。サーバ上ではプル型の場合と同様、予測エンジンが予測したその記事のスコアに対し、一定の点数（現在は10点）を加算して学習を行う。ユーザが操作しなかった記事に対しては予測スコアの減算（現在は5点）を行う。これらの加算、減算の点数は現在のところプル型の場合と同じに設定している。プル型とプッシュ型との大きな相違点は、これらのユーザ操作がサーバ上にあるユーザプロファイルに反映されるタイミングである。プル型では、ブラウザ上でユーザが操作をするとただちにサーバ側に通知され、サーバではそれをもとにユーザプロファイルの変更をするのに対し、プッシュ型ではまず操作履歴を蓄積し、その次にサーバに記事更新のためにアクセスするときに、操作履歴を反映した記事の擬似スコアを通知する。記事の擬似スコアとは、予測エンジンが予測した記事スコアに対して、ユーザ操作を反映した点数加算、減算を行ったものである。たとえば、記事に対するユーザの興味を50点と予測エンジンが予測し、それをユーザが画面上でクリックした場合には、50+10点、つまり60点が擬似スコアとなる。この場合に、クライアント・サーバ間での通信と、サーバ、クライアント上での処理は以下の順序で行われる。

- (1) クライアントからサーバに対し HTTP によるアクセスをして、すべての記事の擬似スコアを通知。
- (2) サーバ内で、受け取った擬似スコアを学習エンジンに渡し、ユーザプロファイルを更新。これにより、ユーザが読んだ記事に含まれていた単語に関しては、ユーザプロファイルの中で重みが増加する。
- (3) 新しいユーザプロファイルに基づき、サーバに新しく登録された記事のスコアを計算し、HTTP によるレスポンスの中にすべての記事の ID とスコア、さらにすべての記事のタイトル（つまり新聞記事のタイトル）を記述する。
- (4) クライアントでは、受け取った記事スコアに基づき画面上での記事タイトル表示位置を決定し、それを画面に表示する。デフォルトの設定では、記事スコアが高いものほど画面上部に表示する。
- (5) クライアントでは、上記表示を行うとともにサーバに対してすべての記事の本文を要求する。

(6) サーバから、複数の記事の本文を1つのファイルにまとめたものをクライアントに送付する。ここで記事本文を1つにまとめているのは通信が何度も生じて遅延が生じることを避けるためである。ただし、1回の通信によって転送されるデータ量が多くなりすぎないように、記事数が多すぎる場合には複数回に分けて送る。

(7) クライアントでは記事本文を蓄積しておき、ユーザが記事タイトルをクリックした際には表示する。

以上のようにすることで、ユーザが記事本文を読もうとしたときに通信による遅延なく記事を読むことができるというプッシュ型の利点を生かしながら、ユーザ操作のフィードバックに基づく自動パーソナライズを実現している。なお、ここで述べたプッシュ型ビューアは1998年1月現在、フリーソフトとしてダウンロード可能（アドレスは <http://www.labs.nec.co.jp/freesoft/ANATAGONOMY/>）である。

3.3 電子メールによるパーソナル電子新聞配信

WWW が盛んに使われるようになって、新聞記事のように定期的に発生する情報は電子メールで受け取りたいというユーザは多い。たとえば、WWW ブラウザによるアクセスでは新聞記事を提供するサイトを毎回自ら指定して見にいかなければならないので、この作業がわずらわしいと感じる場合がある。また、前節で述べたようなプッシュ型のアクセスは、専用線アクセスのように端末がネットワークに接続されている時間が比較的長い場合に適している。その場合はユーザが特に意識なくともつねにデスクトップに新しい記事が配信されている。逆に1日に1回ネットワークにつながりようなユーザの場合には、プッシュ型といってもつかないときにクライアントがサーバにアクセスして新しい記事を取得してくるので、実質的にプル型と大きな差がない。これに対し、情報を手に入れるために主に電子メールを利用しているユーザの場合には、新聞記事がメールで配信されれば他のメッセージと同様に簡単に読むことができる。また、WWW ブラウザが使えないような状況でも電子メールだけは読めるという場合は多い。

しかし、電子メールによる記事配信に対して自動パーソナライズを採用することは難しい。メールはいったん配送されてしまえば、受信者がそれをどのように読んだかは通常送信者には分からないからである。このような問題点を解決し、電子メールによる記事配信に自動パーソナライゼーションを行うために、筆者らは以下のような方法をとった。

ユーザ ID

【国際】◆サミット宣言案、アフリカ積極支援で合意 (6月23日6:13)
http://foo.nec.co.jp/cgi-bin/article2_j.cgi?id=kamba&html=kok-id23id01.htm&scoreup=1&no_link=1&passid=08cf5ff1accdce1a583312a854a59

パスワード (暗号化したもの)

【国際】◆OIC首脳会議にフセイン大統領を正式招請 (6月22日20:42)
http://foo.nec.co.jp/cgi-bin/article2_j.cgi?id=kamba&html=kok-id22id23.htm&scoreup=1&no_link=1&passid=c33b35a1271a7b6e4f8bf6471d2

記事タイトル

【国際】◆大統領、自ら対口投資拡大の旗振り役に (6月22日20:42)
http://foo.nec.co.jp/cgi-bin/article2_j.cgi?id=kamba&html=kok-id22id22.htm&scoreup=1&no_link=1&passid=1ed06c2efc86e9fee94d0055f84f2

図5 送られる電子メールの例 (一部)
 Fig. 5 Part of an email message.

最近、電子メールの送受信を行うソフトウェア (以下、メーラーと呼ぶ) の多くは、メール中に WWW のアドレス、すなわち URL が記載されている場合、それをクリックすると直ちに WWW ブラウザを起動し、記載された URL を表示する。

筆者らは、記事配信サービスにおける電子メールには、記事タイトルと、その記事を見ることができる WWW 上のアドレスをペアにして記載することとした (図5 参照)。記事タイトルの表示順序は、前述のプル型アクセス、プッシュ型アクセスの場合と同じ予測エンジンが予測した、各記事のスコア順にソートされている。記事タイトルだけを送付し記事本文を送付しない理由であるが、實際上、電子メールを1日1回送付するとすれば新聞記事は多数あり、筆者らのシステムの場合でも100以上の記事がある。これらの記事に対してすべて本文を記載するとメールサイズが非常に大きくなり、通信のトラフィックに対する影響、ユーザの便宜を考えても好ましくない。ユーザは記事タイトルを見て興味を引かれれば、そこに記載された WWW アドレスをアクセスして本文を見るであろう。本システムの特徴は、ここに記載された WWW アドレスの中に各ユーザの ID とパスワードが記載されていることである。実際にはそれらが暗号化されている (図5 参照)。したがって、ユーザが興味を持った記事をメーラー上で選択すると、自動的に WWW サーバにアクセスし、選択された記事本文を表示するとともにそのユーザのプロファイルを参照して変更する。この場合に起動されるサーバ側のプログラム (WWW の用語では CGI: Common Gateway Interface) は、WWW ブラウザによるアクセスで記事選択を行って本文表示を行う場合と同じである。プロファイル変更の方法としてはプル型、プッシュ型の場合と同様、選

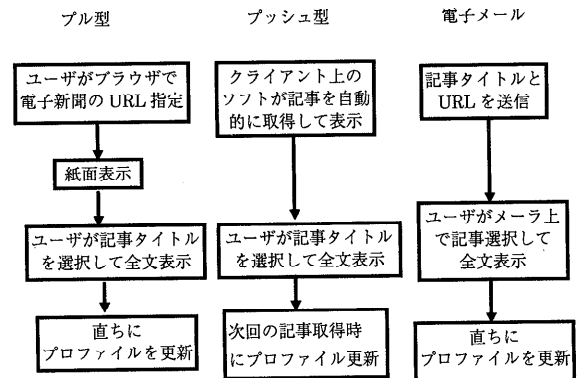


図6 それぞれの手法の動作の流れ
 Fig. 6 Procedures of each method.

択された記事に対しては予測スコアに対して一定の点数を加算し、学習エンジンを起動する。

以上で述べた3つのサービス形態における、記事配信とユーザフィードバックの流れを図6にまとめる。

3.4 ささまざまな使い方の例

本節では、このようにプッシュ型、プル型、電子メール型で同様の自動パーソナライゼーションを提供するシステムにおいて、どのような利用法が考えられるかをケーススタディとして示す。

- ケース1: WWW ブラウザによるアクセスが主体のユーザ

最近のインターネットユーザに比較的多い、WWW ブラウザを利用してさまざまなサイトにアクセスすることを主な使い方とするユーザの場合である。プル型のアクセスなので WWW ブラウザ上で記事選択をすると、その操作がユーザプロファイルに自動的に反映される。

- ケース2: プッシュ型ビューアによるアクセスが主体のユーザ

たとえば企業内 LAN など、常時インターネットにアクセスしているユーザに適した利用法。クライアントプログラムが定期的 (たとえば1時間に1回) にサーバにアクセスし、新しい記事を取得する。ユーザがスクリーンセーバ上で適当に興味を引かれた記事をクリックして本文を読んでいると、その操作が次回のサーバへのアクセス時に自動的にサーバに通知されて、ユーザプロファイルに反映される。

- ケース3: 電子メールによるアクセスが主体のユーザ

電子メールを主な使い方とするユーザに適する。たとえば外出先などから WWW ブラウザの動作しない携帯端末で電子メールを読むと、記事タイトル一覧を読

むことができるので出来事の概要がわかる。その中で特に興味ある記事があった場合には、外出から戻った際などに WWW ブラウザでアクセスして読むことができる。

● ケース 4: プッシュ型アクセスとプル型アクセスの併用

たとえば、会社内では専用線でインターネットにつながっているの、随時記事更新が自動的に行われるプッシュ型で記事を読み、自宅に帰ったらダイヤルアップでインターネットプロバイダにつないで、WWW ブラウザでアクセスして読むというような場合があるであろう。このような場合に、プッシュ型とプル型とで異なるユーザプロファイルが利用されたり、どちらか一方の利用法でしか自動パーソナライゼーションが行われないとすると不便であるが、本論文で述べたシステムを利用することでそのような問題が解決される。

4. 今後の課題と議論

4.1 ユーザ操作のフィードバック方法評価

現在までのところ、プッシュ型、プル型、電子メール型による利用法でユーザが記事を読んだ場合に、どの方法でアクセスしている場合も同じ点数を記事に対して加算している。つまり、システムが 50 点と予測した記事をユーザが WWW ブラウザ上でクリックして読んだ場合も、プッシュ型ビューア上でクリックして読んだ場合も、メーラー上でクリックして読んだ場合も同じように 10 点が加算され、システム側ではそれに合わせて学習エンジンがユーザプロファイルを変更する。しかし、実際には「記事を選択して読む」という操作の価値はそれぞれの読み方によって異なると筆者は思う。なぜならば、それぞれのアクセス方法によって記事選択という操作の心理的負荷が異なることが推測できるからである。たとえばプル型の場合には、ユーザが記事選択をするとそのときに WWW サーバに対して通信が行われ、記事本文がサーバから取得されるから、少しの時間ではあるが遅れが生じる。これに対し、プッシュ型の場合には記事はすべてサーバから自動的に取得してクライアント端末上に蓄積されているので、ユーザが記事選択をすればただちに画面に表示される。したがって、プッシュ型の場合の方がおそらくユーザは「気軽に」記事をクリックして読んでみる可能性が高いであろう。

このように利用方法、表示方法でユーザの「記事選択という操作の価値」がどの程度異なるかは今後の実験、観察を通じて検討していく予定である。それにより、記事選択が行われたときの加算点数も変更する

可能性がある。具体的には、プル型ユーザが全体記事の中で選択して本文を読む記事の割合と、プッシュ型ユーザが本文を読む割合との相違などが手がかりとなるであろう。

4.2 パーソナライゼーションとプライバシー

自動パーソナライゼーションを行う際に、ユーザの操作や興味をどの程度まで細かく取得するかという問題は、プライバシーの問題と深くかかわっている。たとえばプッシュ型ビューア上でユーザ端末上のカーソル位置、それぞれの記事が画面に表示されていた時間等、すべての操作を自動的に記録してサーバ側に送付してパーソナライゼーションのための基礎情報とすることも可能である。しかし、ユーザがそのようなことを明確に意識せずにこれらの処理が行われる場合、プライバシーの問題が生じる可能性がある。最近では、このような個人情報とプライバシーとの関連については WWW に関する標準の推奨などを行っている WWW Consortium が検討を行っており、個人情報の記述、交換方法に関しても関連する企業から提案が行われているので (<http://www.firefly.net/OPS/>等で参照可能)、それらの基準を守ったうえで自動パーソナライゼーションの検討を進めていくべきであろう。

5. おわりに

WWW ブラウザによるプル型の WWW アクセス、プッシュビューアを利用したプッシュ型のアクセス、電子メール型による配信という 3 つの利用形態において、自動パーソナライゼーションを共通化したシステムを提案し、その実装方法を示した。具体的には新聞記事の配信に適用した。本システムによりユーザは、上記 3 つのどの方法で新聞記事を読んでも、自動的にその操作履歴がユーザプロファイルに反映し、パーソナライズした紙面を得られるようになる。今後は本システムの運用を通じ、パーソナライゼーションの精度をあげるためのパラメータ調整等を行っていく予定である。

謝辞 快く記事を提供して下さった読売新聞社殿に感謝いたします。本論文中に示された新聞記事本文の著作権はすべて読売新聞社に属します。

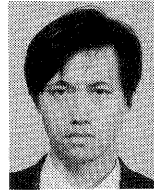
参考文献

- 1) Web キャストの時代がやってきた, *Internet ASCII*, Vol.2, No.5, pp.161-168 (1997).
- 2) ビル・ヴェンナース: Web ブラウザを超えた Java の世界, *Java World*, pp.130-134, IDG コミュニケーションズ (1997).

- 3) Kamba, T., Bharat, K. and Albers, M.C.: The Krakatoa Chronicle: An Interactive Personalized Newspaper on the Web, *World Wide Web Journal, Proc. 4th International World Wide Web Conference*, pp.159-170 (1995).
- 4) Kamba, T., Sakagami, H. and Koseki, Y.: ANATAGONOMY: A personalized newspaper on the WWW, *International Journal of Human-Computer Studies*, Vol.43, Special Issue on Innovative Applications of the WWW, pp.789-803 (June 1995).
- 5) Sakagami, H. and Kamba, T.: Learning personal preference on newspaper articles from user behaviors, *Proc. 6th International World Wide Web*, pp.291-299 (1997).
- 6) Kamba, T. and Frelechoux, L.: Where to locate user profiles of personalized applications, *Proc. INTERACT'97*, pp.555-562 (1997).
- 7) Maes, P.: Agents that reduce work and information overload, *Comm. ACM*, Vol.37, No.7, pp.30-40 (1994).
- 8) Yan, T.W. and Garcia-Molina, H.: SIFT - A Tool for Wide-Area Information Dissemination, *USENIX Technical Conference*, pp.177-186 (1995).
- 9) Lang, K.: Learning to Filter Netnews, *Proc. 12th International Conference Machine Learning (ICML'95)*, pp.331-339 (1995).
- 10) Morita, M. and Shinoda, Y.: Information Filtering Based on User Behavior Analysis and Best match Text Retrieval, *Proc. 17th Annual International ACM-SIGIR Conference on Research Development in Information Retrieval*, pp.272-281 (1994).
- 11) 中村, 馬見塚, 鳥羽, 安倍: プール変数実数多項式による嗜好関数の学習, 第52回情報処理学会全国大会論文集第1分冊, pp.55-56 (1996).
- 12) 坂上, 神場, 古関: プッシュ型ニュースビューアにおける自動パーソナライズ方式, 第55回情報処理学会全国大会論文集第4分冊, pp.345-346 (1997).
- 13) Sakagami, H., Sugiura, A., Kamba, T. and Koseki, Y.: Effective personalization of push-type systems, *Proc. 7th International World Wide Web Conference* (to appear in April) (1998).

(平成9年6月27日受付)

(平成10年3月6日採録)



神場 知成 (正会員)

1984年東京大学工学部電子工学科卒業。1986年同大学院修士課程修了, 同年 NEC 入社。入社以来, インターネットアプリケーション基盤, ビジュアルインタフェース, マルチメディア利用技術, ソフトウェアエージェント等の研究開発に従事。1994年から1995年にかけてジョージア工科大学 GVU センター客員研究員。1997年東京大学より博士(工学)。現在, NEC C&C メディア研究所主任。ACM SIGCHI 会員。



坂上 秀和 (正会員)

1989年東京大学工学部船用機械工学科卒。1991年同大学院工学部産業機械工学専攻修士課程修了。同年日本電気(株)入社。以来, NEC C&C メディア研究所にてマルチメディアクラスライブラリ, ビデオオンデマンドシステム, 分散ハイパーメディアシステム, インターネットマルチメディアシステム, オンラインパーソナライズシステムなどの研究開発に従事。第53回情報処理学会全国大会大会奨励賞受賞。日本ソフトウェア科学会会員。



古関 義幸 (正会員)

1979年東京工業大学工学部情報工学科卒業。1981年 University of California, Los Angeles (UCLA) において Master of Science in Computer Science 取得。同年日本電気(株)に入社。1987年 Stanford 大学 Computer Science 学科客員研究員。1993年東京工業大学より博士(工学)。現在, 同社 C&C メディア研究所研究課長。人工知能応用, 知識獲得, ビジュアルソフトウェア, インターネット情報サービスに関する研究開発に従事。人工知能学会, IEEE 各会員。