

## プッシュ型サービスにおける配信情報のパーソナル化への一検討\*

4U-2

竹内格 森保健治†

NTTソフトウェア研究所‡

## 1 はじめに

現在、プッシュ型情報配信サービスが注目されている。その将来を考えると、情報プロバイダやチャンネル数の増加などが予想され、ユーザは膨大な数の情報に埋もれてしまい、本当に必要な情報を得られなくなる恐れがある。しかし、各ユーザごとに必要な情報だけを配信するには大変な処理負荷と通信コストを要する。

本報告ではこのようなプッシュ型情報配信サービスにおける情報洪水を回避するために、処理負荷と通信コストを考慮しつつユーザニーズにマッチした情報配信を行なう方法についての検討結果を報告する。

## 2 問題

PointCast Network[1]やBackWeb[2]などに代表されるプッシュ型情報配信サービス（以下、プッシュ型サービスと略す）とは、必要な情報ジャンルのチャンネルを選択しておけば、自動的に欲しい広告や情報群がユーザの手元に届くサービスである。今後プッシュ型サービスは、情報プロバイダの増加とチャンネル数の増加、そしてチャンネル内情報数の増加が予想される。このままでは、その量の膨大さのためにユーザにとって本当に必要な情報が埋没してしまう恐れがあり、ユーザに配信する情報に対して何らかのフィルタリングを行なう必要がある。

ところでユーザニーズにかなった情報を配信するために、アンケートを用いて対処しようとするものがみられる。ただしユーザの興味や関心は時間とともに変化していくため、ユーザニーズに対応していくには定期的にアンケートを行なわねばならない（例：HYPER NET [3]）。しかもユーザニーズを把握するため、膨大な数のユーザに対してアンケートを頻繁に実施したり、各ユーザのアンケート結果を管理していく必要があり大変な処理負荷と通信コストを必要とする。またユーザにとっても何度もアンケートに回答していくのは煩わしい。

他方、膨大な数のユーザごとに異なった情報配信を行なう方法も、大変な処理負荷と通信コストを要し、各ユーザへの反応が遅れてしまうなどの問題がある。

ここで配信情報をユーザニーズにマッチさせるために解決すべき問題は次の二つにまとめられる：

## 1. 処理 / 通信コストを考慮したユーザニーズの自動収集

## 2. 処理 / 通信コストを考慮しつつユーザニーズにマッチした情報配信

本稿ではこの二つの問題に対して、配信情報のパーソナル化方式を提案する。特にローカルネットを対象としている。

## 3 配信情報のパーソナル化方式

通常ローカルネットは、社内ネットなど業種や作業内容のまとまった団体ごとに設置するので、比較的ユーザニーズの似通った集団が使用することになり、類似した記事を配信することになる。そこでキャッシュサーバなどの中間サーバで記事をキャッシュすることにより、効率的な記事配信が期待できる。同時に配信の通信コストの軽減も可能となる。最近のプッシュ型サービスでもこのような中間サーバを用いたシステム構成が一般的になりつつある。この節ではこの中間サーバにおいて、ユーザニーズへの対応を効率的に行なう配信方式を提案する（図1）。以下では広告などの情報コンテンツを総称して「記事」と呼んでいる。

## 3.1 システム構成

- IPサーバ：情報プロバイダ（IP）により運営され、ローカルサーバにチャンネル単位で記事を配信する。
- ローカルサーバ：各ローカルネットごとに運営され、各クライアントに記事を配信する。
- 閲覧クライアント：各ユーザはローカルサーバから配信された記事を閲覧クライアントを使って閲覧する。

## 3.2 記事配信方式

- 閲覧クライアント：ユーザの閲覧状況を観測し、ユーザの興味が高いと思われる各記事の閲覧履歴を「ユーザ閲覧ログ」として記録しておく。また適切なタイミングでユーザ閲覧ログをローカルサーバに転送する。
- ローカルサーバ：IPサーバからチャンネルを経て配信された記事をストックする。この時各記事ごとに属性キーワード群を抽出し、「記事プロファイル」として記録しておく。また各ユーザのユーザ閲覧ログと記事プロファイルから、「ユーザプロファイル」を作成しておく。各ユーザへの記事配信時に、ユーザプロファイルと記事プロファイルを比較することにより、記事を選択して配信する。さらに全ユーザプロファイルを集計してローカルネットでの記事の消費状況をまとめ、「グループプロファイル」とする。そしてグループプロファイルをもとにIPサーバから記事配信を受ける。

\*A Study on the Personalized Information Broadcasting Architecture

†Kaku TAKE-UCHI, Kenji MORIYASU

‡NTT Software Laboratories

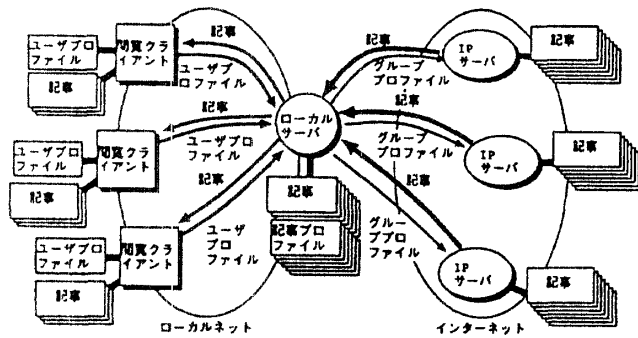


図1: 配信情報のパーソナリ化方式

#### 4 検討

2 節で挙げた二つの問題に対し、3 節で提案した情報配信方式でどのように対処されるかを整理する。

##### 1. 処理 / 通信コストを考慮したユーザニーズの自動収集

- ユーザの興味や関心は、記事の閲覧動向に反映していると考えられ、ユーザ閲覧ログからその時点でのユーザの興味や関心を抽出することができると思われる。またユーザ閲覧ログは閲覧クライアントにより、ユーザの記事閲覧作業と連携して自動的に作成されるので、ニーズの把握にアンケートへの回答などユーザに特別な作業を行なってもらわなければならない。
- 各ユーザプロファイルの作成、つまり各ユーザごとのニーズの把握はローカルサーバで行なうので、各ユーザのニーズ把握に関する作業コストがローカルネットごとに分散され、またローカルネット外への通信コストも軽減される。

##### 2. 処理 / 通信コストを考慮しつつユーザニーズにマッチした情報配信

- 各ユーザへのニーズへの対応はローカルサーバで行ない、各ユーザへの情報選択にかかる処理コストはローカルネットごとに分散される。また情報プロバイダは、各ローカルネットごとに適切な記事を配信するだけなので通信コストも軽減される。

ところで 2 節の IP サーバからローカルサーバへの配信は、どちらが配信の主休になるかにより次の二つの方法がある:

- ローカルサーバ主体: グループプロファイルに基づいて、各 IP サーバごとに適切なチャネルの選択をローカルサーバが行なう。この場合、ローカルサーバは各 IP サーバがどんな記事を提供するか前もって知る必要がある。
- IP サーバ主体: ローカルサーバから IP サーバにグループプロファイルを送り、IP サーバがグループプロファイルに基づいてチャネルの選択やチャネル内の記事の選択を行なう。

#### 5 技術課題

本稿で提案した情報配信方式の技術課題について述べる。本方式は情報フィルタリング技術 [4] と同様の技術課題が存在する。ここでは特に閲覧状況の観測、記事プロファイルの作成、記事の選択、記事の収束問題に関して述べる。

- 閲覧状況の観測: ユーザの閲覧状況には閲覧頻度と、どのように記事を消費したかの消費状況が考えられる。閲覧頻度はユーザが実際に閲覧した各記事の閲覧回数から観測できる。消費状況は記事の閲覧時間 (記事への注視時間) や、記事のプリントアウトを行なったかなどから観測できる。ただし、ユーザの興味は時間とともに変化していくので、古い閲覧状況への重みづけは減衰するような配慮が必要である。
- 記事プロファイルの作成: 属性キーワードの抽出は、記事の構造解析、形態素解析によるキーワード抽出、抽出したキーワードの頻度計算、属性キーワード群の選定を行なわれる。これらの処理の中で属性キーワード群の選定が最も困難であり、キーワード出現頻度、ヘッドラインやタイトル等に出て来るキーワードの尊重、あるいは記事作成者が意図したキーワードの尊重などを考慮する必要がある。
- 記事の選択: ユーザプロファイルと記事プロファイルがどれだけ重なるか、この重なり具合を定量化せねばならない。例えば双方の属性キーワード群の交わり部分がどれだけ存在するかで定量化する、双方で極めて頻度の高い属性キーワードがどれだけ交わり部分に含まれるかで定量化する方法などが考えられる。
- 配信記事の収束: 常に類似した記事ばかり配信される恐れがある。これを避けるには、たまにニーズの低い記事を配信したり、選択されない記事はヘッドラインのみを配信し必要に応じてプル型で対処する、すべて配信するがニーズの高いものから表示するなどが考えられる。

#### 6 おわりに

本稿では、近い将来生じるであろうブッシュ型情報配信サービスにおける情報洪水を回避するために、配信情報をユーザニーズに効率的にマッチさせる方式を提案した。今後はこの配信方式の実現方法について、5 節の技術課題を中心にさらに検討を進めていく予定である。

特に配信情報のユーザニーズへの適合を焦点にして述べてきたが、ブッシュ型情報配信には情報コンテンツ特性への対応、表示形態、ユーザ個別のネットワーク環境への対応など、数多く課題が存在する。これらについても検討を行なっていきたいと考えている。

#### 参考文献

- [1] "PointCast Network", <http://www.pointcast.com/>
- [2] "BackWeb", <http://www.backweb.com/>
- [3] "HYPER NET", <http://www.hypernet.co.jp/>
- [4] "情報フィルタリングシステム", 森田昌宏・速水治夫, 情報処理 (学会誌), Vol.37, No.8, pp751-758.