

ニュースフィルタリングエージェント における協調

中嶋 卓雄*, 稲益康晴*, 河島健一†, 中村 良三*

*熊本大学 工学部 †熊本日日新聞

コンピュータネットワークの発展にともない、情報量が増加するにつれてユーザの要求にそったフィルタリングシステムが求められている。一方情報公開ツールとして WWW が広まり、新聞業界も WWW に情報を公開しつつある。本稿では、熊本日日新聞社が発行する新聞を html 形式に変換した電子新聞において、ユーザが望む情報を選択するニュースフィルタリングエージェントの協調動作の設計を示す。さらに、時間の経過により記事の重要度が低下するなどの電子新聞に固有な問題についても考察する。

1 はじめに

コンピュータネットワークの発展にともない、メールやニュースなどから各種の情報を容易に獲得することが可能となり、さらにユーザインタフェースの発展により、コンピュータを専門としない一般ユーザ層が広範囲に拡大している。

しかし、全般的な情報量が増加するにつれて、その操作性が多様化しており、さらに現存するグラフィカルなインターフェースについてもシステムに依存する傾向が強くなり、汎用性に欠けている。そのため、大量で多様な情報をユーザの要求にそって処理できる汎用的で操作性に優れた情報フィルタリングシステムが求められている。[1] その解決策として人間の作業を代行し自律的に活動するエージェントシステムを利用して、ユーザの意志に沿って電子メールの

整理や NetNews のフィルタリングを行なうシステムが提案されている。[2],[3]。

一方、インターネット上でマルチメディア情報を公開するツールとして WWW が広まっており、個人や組織の紹介の他にも公共機関や図書館が保持している情報を WWW を利用して公開する試みが各地で行なわれている。今まで情報メディアと呼ばれていた新聞・マスコミ業界も WWW に情報を公開しつつあり、商用ネットワークの普及により活発化すると思われる。

熊本地域では、地域の情報公開の一環として熊本日日新聞社が発行する新聞を html 形式に変換し、電子新聞として限定的に公開した [4]。しかし、将来的に新聞各社が同様のシステムを開発した場合、共通記事が多くなり個性がなくなる可能性がある。そこで、新たな内容・構成をもつ新聞として、新聞記事データベースが保有する大量の情報から効率良く個人の望むデータを獲得し、ユーザ固有の新聞を構成することにより、“あなたの紙面”を提供するシステムの構築を考えた。

本稿では、熊本日日新聞社が発行する新聞を

Cooperation in News Filtering Agent
Takuo Nakashima*, Yasuharu Inamasa*, Kenichi Kawashima†, Ryozo Nakamura*

*Faculty of Engineering, Kumamoto University,

†Kumamoto Daily News Company,

html形式に変換した電子新聞において、ユーザが望む情報を選択・提供するニュースフィルタリングエージェントの設計を示す。また、新聞の主題により分類された記事データベースから、主題を表すカテゴリー単位にエージェントを構成し、エージェント間の通信を通してコミュニティを形成するモデルを提案する。さらに、時間の経過により記事の重要度が低下するなどの電子新聞に固有な問題についても考察する。

2 基本モデル

2.1 対象とする情報

情報は利用する形態から次の2つに分類できる。[5]

- フロー情報
電子メール、NetNewsのように刻々と更新される情報。
- ストック情報
技術文献のように蓄積されている情報。

従来はストック情報に対する検索など研究されてきたが、近年、フロー情報に対するデータベース化、検索などの研究も活発化してきている。ストック情報は情報化の段階で推敲され洗練されているため、そのまま蓄積しても問題はないが、フロー情報の場合、電子メールのように会話を文章化した内容に近いこともあり、情報の圧縮、削除などの処理を行い、情報を整理する必要がある。

今回、研究の対象としている情報は新聞であり、フロー情報といえるが、事実および評論などの解説文は蓄積されるストック情報と同様に扱う必要がある。

2.2 新聞記事データベース

これまでの文章情報データベースは、原資料の書誌的なデータ、すなわち論文のタイトル、著者、掲載誌名などを収録して原資料の所在を示すレファレンス型のデータベースであった

が、新聞記事データベースは記事の全文を収録しており、法令・判例などのフルテキストデータベースと並んでファクト型データベースと呼ばれる。そのため記事の主題を記事の全文から抽出する必要があり、また、様々なデザインが施された見出し、カラー写真なども含まれており、従来のデータベースとは異なる視点でシステム化する必要があるだろう。

今回、データベースを構築するための記事は

- 編集責任者から構成される会議で合意され、分類されたカテゴリー
- 見出し
- 本文

から構成されるとする。

この記事から以下の要素を抽出する。

1. カテゴリー
新聞記事の最も大きな枠組み。
2. キーワード
記事の主題を表す単語。
3. 重要度
キーワードが記事中でどの程度重要な単語であるかを示す指標。0から1までの数値で表す。

2.3 ユーザの興味の表現方法

ユーザが持つ趣味・嗜好をフィルタリング情報と呼び、

1. カテゴリー
新聞記事を分類する最も大きな枠組み。
2. キーワード
ユーザが興味を持っている新聞記事中の単語。キーワードは一つまたは複数のカテゴリーに含まれる。
3. 関心度
カテゴリーとキーワードから構成されるペアにユーザがどの程度興味を持っているかを示す指標。0から1までの数値で表す。

の3つの要素の集合から構成する。

また、フィルタリング情報は

- ユーザ単位に保存される。
- ユーザ個人の興味の変化を表現するため、ユーザのアクセス頻度に影響を受け、参照されない記事については、時間の経過により影響力が弱くなる。
- 時代の流れを反映させるため、記事および他のユーザから影響を受ける。

とする。

2.4 ニュースフィルタリングシステム

WWW を使いニュースフィルタリングシステムを構成するため、サーバ・クライアント構成のシステムとする。また、現状では http に十分な機能が実現されていないため、サーバ側にシステムを中心を置く。

- ユーザインターフェース
ユーザインターフェースとしては、Mosaic などの WWW クライアントを使用する。
- マルチエージェント

ユーザのフィルタリング情報を管理するユーザエージェント、カテゴリーレベルの記事の動向および複数のユーザの嗜好などを管理しユーザエージェントにフィードバックをかけるコミュニティエージェント、記事の主題として連日取り上げられるテーマを表すニュースエージェント、記事データベースを検索し記事をニュースエージェントに対応付ける記事データベースエージェントなどから構成する。

記事データベース

記事データベースを構築する場合、検索する方法を考慮に入れる必要がある。記事データベースを検索する方法としては、統制語による方法と自然語による方法がある。統制語検索で

は、事前に人手で用語が管理されているため確実に検索結果を得ることができるが、管理作業の困難さや、新しい用語や用法が頻繁に発生する可能性があるため、自然語検索を用いることを前提としてデータベースを構築する。

まず、記事の内容を表すキーワードを抽出しデータベースを構成する手順を以下に示す。またその概略を図1に示す。

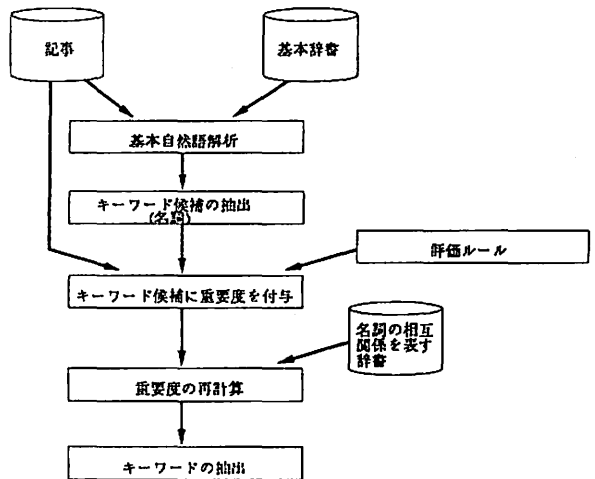


図1:キーワード抽出の概略

1. キーワード候補の選択
基本辞書に基づき、接頭語、接尾語、複合語処理などを行い、記事の内容を表すキーワードの候補となる名詞を選択する。
2. 選択された候補へ重要度を付与
選択された名詞がその記事の中でどの程度重要かを示す重要度を評価ルールに従って付与する。重要度としては、0から1までの数値を考え、評価ルールとしては、
 - (a) 見出しに使われている
 - (b) 主格として使われている
 - (c) 出現回数などを考えている。

- 候補相互間のつながりの強さにより重要度を再計算
名詞の同義語、関連語、上位語、下位語および相互接続の強さを表す共起関係を考慮して重要度を再計算する。

- キーワードの抽出
重要度の値が大きい順に、一定数または一定の値以上の候補をキーワードとして抽出する。

3 ニュースフィルタリング

電子メールや NetNews に対するフィルタリングと異なり、新聞記事に対するフィルタリングには、新聞からその日の話題を拾い、日々変化する時代の流れをつかむ作業について考慮する必要がある。

以下では、設計段階であるマルチエージェントシステムとその協調作業について述べる。

3.1 ユーザエージェント

本システムでは WWW により電子新聞を公開することを前提として設計しており、インターフェースとして WWW のクライアントを採用し、ネットワーク上は図 2 のようにサーバ・クライアントモデルを採用する。

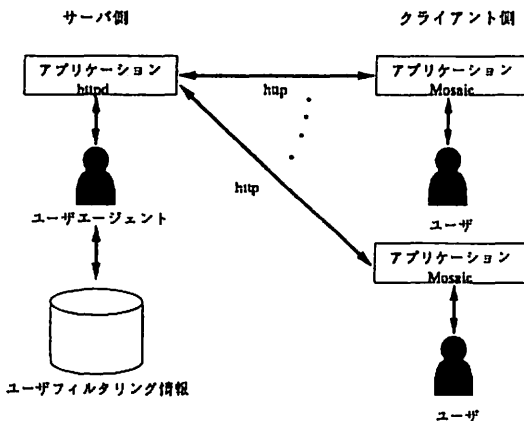


図 2:サーバ・クライアントモデル

クライアント側ではユーザが WWW のクライアントアプリケーションを起動し、そのウィンドウ上でデータの入力やアクセスを行う。サーバ側では、ユーザエージェントが複数のユーザのフィルタリング情報を管理する。ユーザエージェントは以下の動作をする。

1. フィルタリング情報の登録

ユーザが最初にシステムを利用する場合、フィルタリング情報をクライアントアプリケーション経由でユーザエージェントがデータとして登録する。

2. コミュニティエージェントへの問い合わせ

ユーザ固有のフィルタリング情報に基づきコミュニティエージェントに関連する記事を問い合わせ、優先度をつけ html ファイルを生成してユーザに公開する。

3. コミュニティエージェントからのトレンド情報の獲得

社会の流れを反映させるために、他のユーザおよび一般的に流行しているトレンド情報についてコミュニティエージェントから情報を獲得し、ユーザへ公開するとき、その情報を反映させる。

コミュニティエージェントへの問い合わせ概要を図 3 に示す。

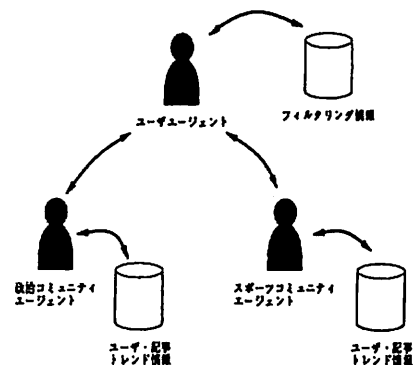


図 3:コミュニティエージェントへの問い合わせ

4. ユーザアクセス情報のフィードバック

図4に示すように、ログ監視エージェントによってアプリケーションから得られたアクセス情報をユーザ毎のユーザログ情報として管理する。ユーザエージェントはログ監視エージェントの起動・終了を制御し該当するアクセス情報を引きだし、その内容をフィルタリング情報にフィードバックさせる。

5. フィルタリング情報の更新

時間の経過とともに個人の趣味が変化することを反映させるため、参照されない情報については関心度を低下させる。

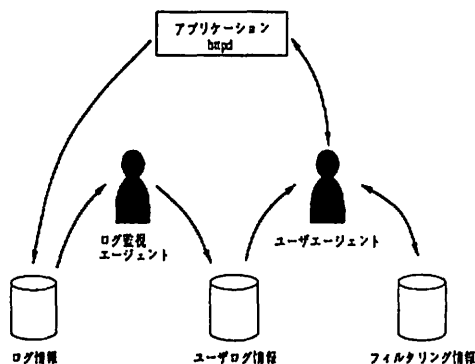


図4:アクセス情報のフィードバック

3.2 コミュニティエージェント

複数のニュースエージェントを新聞の大分類に相当するカテゴリ単位に統合して同類の概念で複合したエージェントをコミュニティエージェントとする。ニュースエージェントは複数のコミュニティエージェントに属することが可能である。

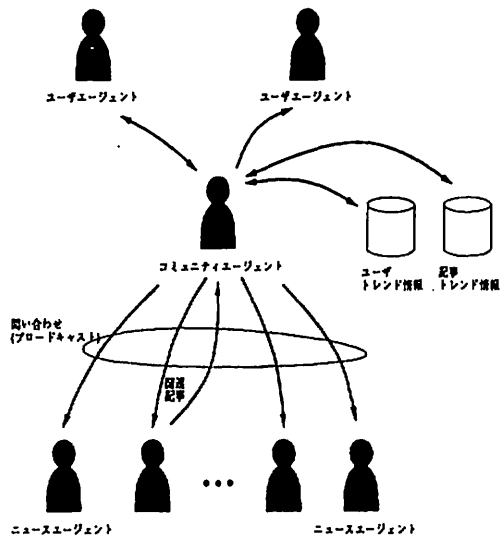


図5:コミュニティエージェントの協調動作

コミュニティエージェントは図5に示すような協調的な動作をする。具体的には以下の動作をする。

1. エージェント間の協調動作

コミュニティエージェントはユーザエージェントから問い合わせ内容を管理下にあるすべてのニュースエージェントに対して broadcast して、該当するニュースエージェントから返答をもらい、その内容をユーザエージェントに返す。

2. トレンド情報の管理

複数のユーザからの問い合わせを管理し、頻繁に問い合わせのある情報をユーザトレンド情報として管理する。さらにニュースエージェントから得られた記事情報の中で重要度の高い単語として頻繁に用いられた情報を記事トレンド情報として管理する。

3.3 ニュースエージェント

ニュースエージェントは記事データベースから抽出された(カテゴリ、キーワード、重要

度)の集合に基づき、共通の内容を表す同類の複数の記事からキーワード集合とともに生成される。

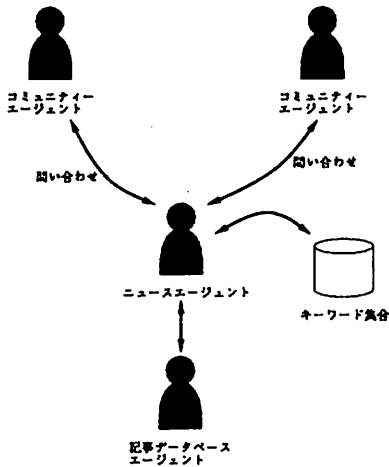


図 G: ニュースエージェントの協調動作

ニュースエージェントは図 G に示すような協調的な動作をする。具体的には以下の動作をする。

1. エージェント間の協調動作

複数のコミュニティエージェントから問い合わせられた内容がニュースエージェントのキーワード集合にヒットしたら、記事データベースを管理する記事データベースエージェントに問い合わせ、関連する記事を選択しコミュニティエージェントに結果を返す。

2. キーワード集合の管理

複数の記事の共通項となる単語および複数の単語を抽象化した単語などから構成され、あるニュースの内容を表すものとする。長期的には関連語、同義語辞書なども関係させて管理する必要がある。

3.4 記事データベースエージェント

ユーザのフィルタリング情報から適切な記事を選択するために、ユーザが表現するキーワー

ドと記事の主題を表すキーワードとの間の相互関係および階層関係を考慮する必要がある。

ニュースエージェントから渡されたフィルタリング情報を記事データベースエージェントが次のように記事を検索する。

- カテゴリーのマッチングを行なう
- キーワードのマッチングを行なう
- マッチした項目について重要度と関心度の積を計算する
- 積の総和を記事毎に計算する

1つの記事について得られた数値をニュースエージェントが統合し、最終的にコミュニティエージェントが複数のニュースエージェントの持つ重要性を加味して、優先順位を付けユーザに返答する。

4 電子新聞固有な問題

ストック情報に基づくデータベースと異なり記事データベースでは、

1. 時間の経過により記事の重要度が低下する。
2. 記者が執筆した記事は紙面の関係からすべて掲載されるとは限らない。そのため、記事に対して記者の認識と読者の認識に相違点が生じる可能性がある。

などに注意を払う必要がある。

上記の問題に対して、前者はフィルタリング情報へ反映させ、後者は、掲載されない記事に対する取り扱いとなるため、十分な検討が必要であろう。

5 おわりに

今回は新聞記事データベースに対する情報フィルタリングとして、WWWを利用したマルチエージェントによるフィルタリングを提案した。現在そのプロトタイプ版を作成中である。

今後は、マルチエージェント間の協調作業をシミュレーションを通して定量的に評価していきたい。また、個人情報の一部がコミュニティエージェントに共有される場合があるため、個人情報のプライバシーについても考慮していく予定である。

参考文献

- [1] 森田昌宏：情報フィルタリング技術の現状と展望，電子情報通信学会技報，

Vol.93, No.153, pp.49-56, 1993.

- [2] Pattie Maes, "Agents that Reduce Work and Information Overload" *Communications of the ACM*, Vol.37, No.7, 31-40, 1994.
- [3] Beerud Sheth, Pattie Maes, "Evolving Agents For Personalized Information Filtering" In *Proceedings of the Ninth Conference on Artificial Intelligence for Applications*. IEEE Computer Society Press, 345-352, 1993.
- [4] 柏谷佳樹，河北隆生，中嶋卓雄，河島健一：熊本日日新聞をベースにした電子新聞の試作。第9回熊本県産学官技術交流会資料，pp.103-104(1995)
- [5] 上田隆也，柴田昇吾，佐藤史郎，廣田誠，池田裕治，藤田稔：フロー制御を対象にした情報検索システム(1)~(4)情報処理学会第50回全国大会資料，4F-6-4F-9(1995)