

# エージェントコミュニティを利用したP2P型情報検索方式の提案

## Agent Community based Peer-to-Peer Information Retrieval

松野 大輔  
Daisuke Matsuno

九州大学大学院システム情報科学府  
Graduate School of Information Science and Electrical Engineering, Kyushu University  
matsuno@al.is.kyushu-u.ac.jp

峯 恒憲  
Tsunenori Mine

九州大学大学院システム情報科学府  
Faculty of Information Science and Electrical Engineering, Kyushu University  
mine@al.is.kyushu-u.ac.jp, <http://www-al.is.kyushu-u.ac.jp/~mine>

高橋 健一  
Ken'ichi Takahashi

九州大学大学院システム情報科学府  
Graduate School of Information Science and Electrical Engineering, Kyushu University  
tkenichi@al.is.kyushu-u.ac.jp

雨宮 聡史  
Satoshi Amamiya

(同上)  
roger@al.is.kyushu-u.ac.jp

雨宮 真人  
Makoto Amamiya

九州大学大学院システム情報科学府  
Faculty of Information Science and Electrical Engineering, Kyushu University  
amamiya@al.is.kyushu-u.ac.jp, <http://www-al.is.kyushu-u.ac.jp/~amamiya>

**keywords:** multi-agents, information retrieval, peer-to-peer, collaborative filtering,

### Summary

This paper proposes an agent community based information retrieval method, which uses agent communities to manage and look up information related to users. An agent works as a delegate of its user and searches for information that the user wants by communicating with other agents. The communication between agents is carried out in a peer-to-peer computing architecture.

In order to retrieve information related to a user query, an agent uses two histories: a query/retrieved document history and a query/sender agent history. The former is a list of pairs of query and retrieved documents, where the queries were sent by the agent itself. The latter is a list of pairs of query and sender agents and shows "who sent what query to the agent". This is useful to find a new information source.

Making use of the query/sender agent history is expected to cause a collaborative filtering effect, which gradually creates virtual agent communities, where agents with the same interests stay together. Our hypothesis is that a virtual agent community reduces communication loads to perform a search. As an agent receives more queries, then more links to new knowledge are achieved. From this behavior, a "give and take" (or positive feedback) effect for agents seems to emerge.

We implemented this method with Multi-Agents Kodama which has been developed in our laboratory, and conducted a preliminary experiment to test the hypotheses.

The empirical results showed that the method was much more efficient than a naive method employing 'broadcast' techniques only to look up a target agent.

### 1. はじめに

Web上の情報増加にともない、既存の集中型検索エンジンは情報のカバー率の低下や情報の鮮度の低下といった問題を抱えている。このような問題に対処するため、分散検索の研究 (e.g. [Callan 01]) が活発に行われている。

また、現在のインターネットユーザは、日常的に様々な情報の洪水に見舞われており、その中から自分にとって必要な情報だけを取り出す作業や、自分の必要な情報を膨大な Web の中から探す手間に追われている。そのような作業の手間を省くため、ユーザにとって必要な情報

だけを残す「情報フィルタリング」(e.g. [Lang 95]) やユーザの興味ある情報を推薦する「推薦システム」(e.g. [Schafer 99])、興味ある情報を探すために役立つ「専門家発見」(e.g. [Yimam-Seid 03] [Kanfer 97])、更には、他のユーザの評価情報を利用してユーザにとって興味ある情報を推薦する「協調フィルタリング」(e.g. [Goldberg 92, P. Resnick 94, Good 99, Sarwar 01]) などといった研究が盛んに行われている。

しかし、このような研究で開発されるシステムの多くは、サーバ・クライアント型のモデルに基づいており、情報の集中制御を行う際に生じるボトルネックに悩まされて

いる。そのため、情報の共有機能を、ピアツーピア (以下 P2P と略記) 型のモデルで実現するための研究も数多く行われてきている (e.g. [Stoica 01, Clarke 01, Gnutella 00, Napster 00])。しかし、そのモデル中の各ノードで行われる処理内容は画一的であることが多い。

このような背景から、本稿では、エージェントコミュニティを利用した P2P 型情報検索方式を提案する。この方式では、コミュニティに所属する他のエージェントから受けた検索結果履歴を元に、コミュニティ内での情報の在処の特定や、コミュニティ内で同じトピックに関心を持つエージェント同士のグルーピングを実現する。これにより、コミュニティ内で、必要な検索結果を得るために行う通信の量を、徐々に削減していくことができる。さらに、検索要求に多く答えられるエージェントほど、自分の求める情報源へのパスを増やすことができ、その結果、検索効率が向上するという「give and take」の効果の実現が期待できる。

まず、その第一ステップとして、情報の通信量の削減についての効果を確認するための簡単なシミュレーション実験を行った。その結果、本方式が仮定していた「通信量の削減効果」を確認できた。

以下、第 2 節では、本検索方式についての詳細を述べる。第 3 節で本検索方式を実現するために利用したマルチエージェント KODAMA [Zhong 02] について説明し、KODAMA 上で本検索システムがどのように動作するかについて述べる。第 4 節で実験結果について議論する。第 5 節で関連研究についてまとめ、本研究で提案する方式と、他研究との立場の違いを明らかにする。最後にまとめと今後の課題について述べる。

## 2. エージェントコミュニティ方式の情報検索

### 2.1 エージェントコミュニティ方式の情報検索の意義

本方式は、P2P 型のネットワーク上に構成された階層的なエージェントコミュニティに所属するエージェント (Peer) 間で交された検索要求の履歴を利用して、効率良く自分が求める情報の在処を見つける一手法である。エージェントが所属するエージェントコミュニティは、現実の社会に対応した特定の組織のメンバー、あるいは同じ興味・関心を持つメンバーのエージェントが所属する枠組みを表す。

本検索方式が、このようなエージェントコミュニティを利用する理由は、以下の 4 つである。

**P2P 型の情報流通** P2P 型の情報流通アーキテクチャを採用することで、サーバクライアント型のような、情報要求のボトルネックを避ける。

**範囲を限定した情報の流通** コミュニティ単位で情報を扱うことにより、特定のグループに限定した情報の公開などといったフィルタリングを行う。

**通信量の削減** 本方式では、初めて提出される検索要求

に対しては、その検索要求をコミュニティ内の全エージェントに対してブロードキャストを行うことを仮定している。そのため、エージェントをコミュニティ単位に分けておくことで、ブロードキャストの範囲を限定する。

**検索要求受信ログを利用した検索** 他のエージェントから送られてくる検索要求を受信した履歴 (以下「検索要求受信ログ」と呼ぶ) を利用して、ユーザが求める情報を効率良く探すための自然な構造を与える。

### 2.2 本方式で利用するエージェント

ユーザ毎に用意されるエージェントには、ユーザインタフェースエージェント、検索エージェント、および履歴管理エージェントの 3 つがある。

以下に、各エージェントの機能を示す。

#### ● ユーザインタフェースエージェント

ユーザインタフェースエージェントの仕事は次の 3 つである。

- (1) ユーザから出された検索要求を受け取り、検索エージェントに対して、その関連情報の検索を依頼する。
- (2) 検索エージェントから返された関連情報をユーザに提示する。
- (3) ユーザからの検索結果に対するフィードバック (結果の選択・不選択や、印刷、ブックマークなどような暗黙のフィードバック) を、履歴管理エージェントに送る。各結果には、その結果を返してきた検索エージェントのアドレスが付随している。これにより、次回の検索の際に良好なフィードバックを受けた結果を返した検索エージェントが優先的に選ばれることが期待される\*1。

#### ● 検索エージェント

検索エージェントの仕事は次の 4 つである。

- (1) ユーザインタフェースエージェント、ポータルエージェント\*2、および他の検索エージェントから送られてきた検索要求に関連する情報の検索を行う。

検索要求に関連する情報の検索は、検索要求と、検索対象との間の類似度を計算し、一定閾値以上の類似度が認められた時、「YES (関連情報有)」のメッセージと、関連情報とを検索結果として返す。閾値に満たない場合には、「NO (関連情報無)」のメッセージのみを返す。検索対象は、ユーザが作成したコンテンツ情報を保持す

\*1 ただし、本稿では、この件に関して議論しない。

\*2 コミュニティ内の検索エージェントの名前と位置、特徴を管理するエージェントである。また、検索エージェントから受けたブロードキャスト依頼に基づき、コミュニティ内の全検索エージェントに、検索の依頼を行う。(詳しくは 3.2 節で述べる。)

るファイル（コンテンツファイル）、ユーザの検索結果履歴を保持するファイル（検索結果履歴ファイル）、および他の検索エージェントから送られてきた検索要求と、その検索エージェントのアドレスとを保持するファイル（検索要求受信ログファイル）である。各ファイルの詳細については、2・3 節で述べる。検索結果の返信は、検索を依頼してきたエージェントが、検索された情報の公開許可範囲に含まれている場合に行われる。

- (2) ユーザインタフェースエージェント以外のエージェントから検索要求が送られてきた場合、その検索依頼を出してきた検索エージェントのアドレス\*3 と検索要求とを、履歴管理エージェントに送る。
- (3) 関連情報の中に含まれている、情報元となる他の検索エージェントに対して、新しい情報を得るために、その検索式を送る\*4。
- (4) ユーザインタフェースエージェントから送られてきた検索要求に対し、検索結果が、ユーザの要求数を満たさない場合には、コミュニティ内のポータルエージェントに対して、検索要求を送り、検索を依頼する。

検索要求を受け取ったポータルエージェントは、（検索要求、検索依頼エージェント）のペアを、コミュニティメンバである他の検索エージェントに送る。

●履歴管理エージェント

ユーザが出した検索要求と、その検索結果（検索要求に関連する情報）の対を記録した「検索結果履歴ファイル」および、他のエージェントから送られてきた検索要求とそのエージェントのアドレスとを記録する「検索要求受信ログファイル」の管理を行う。

2・3 コンテンツファイル、検索結果履歴ファイル、および検索要求受信ログファイル

検索エージェントが検索に利用するファイルは、ユーザが作成した情報（コンテンツ）を保持するコンテンツファイル、ユーザの過去の検索結果履歴を保持する検索結果履歴ファイル、および他の検索エージェントから送られてきた検索要求の受信履歴を保持する検索要求受信ログファイルの3つである（表 1 参照。）。コンテンツファイルは、(title, text, original, range) の4つ組からなり、それぞれ、コンテンツのタイトル、コンテンツの内容である本文、コンテンツを作成したユーザの検索エージェントのアドレス、および、情報の流通範囲である。検索

結果履歴ファイルは、(query, contents, from) の3つ組からなり、それぞれ、検索エージェントが、検索のために送信した検索要求、その検索要求に関連する情報として取得したコンテンツ、およびそのコンテンツを返信してきた検索エージェントのアドレスである。contents に記録されるコンテンツの形式は、先に述べたコンテンツファイルの形式と同じである。検索要求受信ログファイルは、検索要求 query と、それを送ってきたエージェントのアドレス from の2つ組 (query, from) からなる。

以下にコンテンツファイルの一部を示す。ただし、original には、復号化されたエージェントのアドレスを示す。

title	text	original	range
Netscape 非 公 式 FAQ 日本 語版	HTML タグを除いた本文のテキスト（詳細は省略）	com_Netscape@	ALL

また、検索要求受信ログファイルの一部を以下に示す。同様に、from には、復号化されたエージェントのアドレスを示す。

query	from
電報	root.p2p.com_電報@
治療	root.p2p.sic_C 型肝炎@
喘息	root.p2p.sic_ぜんそく@
人	root.p2p.sic_アダルトチルドレン@
事	root.p2p.sic_アルツハイマー病@
病	root.p2p.sic_クロイツフェルト・ヤコブ病@
夢	root.p2p.sic_ディーラー@
乳がん	root.p2p.sic_乳癌@
蹄	root.p2p.sic_口蹄疫@

2・4 検索要求配送先決定手順

ユーザが出した検索要求は、ユーザインタフェースエージェントが受け取り、検索エージェントが解釈できる形式（以下、クエリ）に変形された後、検索エージェントに渡される。

クエリを受け取った検索エージェントは、まず「1. 検索結果履歴を用いた検索」を行い、それで要求した検索件数（検索結果受信数）に達しない場合、「2. 検索要求受信ログを用いた検索」を行う。それでも検索結果受信数に達しない場合、「3. ブロードキャストによる検索」を行う。以上の手順を、図 1 を例として説明する。

1. 検索結果履歴を用いた検索 「検索結果履歴」の中にユーザから受け取ったクエリと一致するものか類似するものがある場合、その検索結果履歴の from および、original のアドレスの検索エージェントにクエリを送信する。例えば、検索エージェント A が、クエリ「無線 LAN」を出す場合、検索結果履歴を見て、同一の query 項目の from と original に登録されている検索エージェント D と E に、そのクエリを送る（図 1 (1)）。ここで、クエリの類似性は、適当な類似度計算式（例えば、クエリ中の単語の含有

\*3 検索エージェントのアドレスは、プライバシー保護の理由から暗号化されており、そのアドレスから、直接、エージェント名の推測はできない。

\*4 実際には、情報の更新日時が、現時点から一定時間以上、過ぎている場合などに限るが、本稿ではこの点は議論しない。

表 1 本検索方式で使用するコンテンツ、検索結果履歴、検索要求受信ログの各ファイル\*5

コンテンツ	title	コンテンツのタイトル
	text	コンテンツ本文
	original	このコンテンツを最初に作成・発信した検索エージェントのアドレス
	range	流通範囲 (ALL, Community, Agent)
検索結果履歴	query	送信したクエリ
	from	この検索結果を返信した検索エージェントのアドレス
	contents	検索により取得したコンテンツ (上の段のコンテンツの形式に従う)
検索要求受信ログ	query	受信したクエリ
	from	このクエリを送信してきた検索エージェントのアドレス

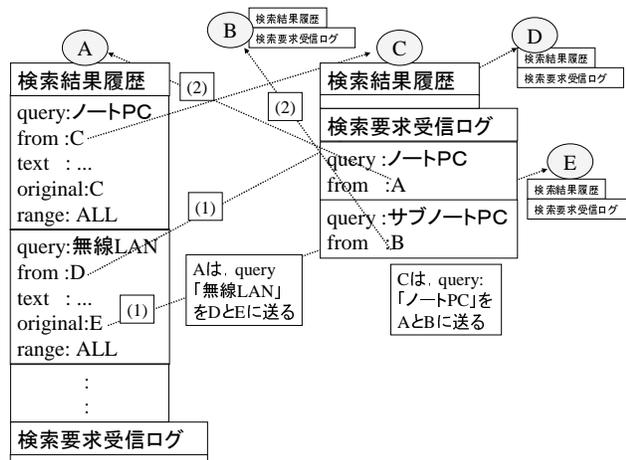


図 1 本検索手法の検索手順

率)によって求める。

2. 検索要求受信ログを用いた検索 ユーザから受け取ったクエリと同じか類似したクエリが「検索要求受信ログ」にある場合、そのクエリを送ってきたエージェントは、既に検索を完了させて、このクエリに関する情報を保持しているものと見なし、そのエージェントにクエリを送信する。例えば、検索エージェント C がクエリ「ノート PC」を出す場合、検索要求受信ログを見て、そこの query 項目で一致もしくは類似した from のエージェント A および B に、そのクエリを送信する (図 1 (2))。
3. ブロードキャストによる検索 上記 1., 2. の手順で、クエリの送信先エージェントがある一定の個数 (検索結果受信数) 見つからなかった場合、そのクエリに対する関連情報の検索をポータルエージェントに依頼する。ポータルエージェントは、同じコミュニティに所属する検索エージェントに、クエリをブロードキャストする。
3. のブロードキャストによる検索では、ポータルエージェントからクエリを受信した検索エージェントは、まず、履歴管理エージェントに、このクエリと、クエリを送ってきた検索エージェントのアドレスを送る。検索結果履歴管理エージェントは、それらを、検索要求受信口

グに登録する。これにより、自分が同じクエリで検索する際に、検索に利用すべきサイトを特定することができる (2. の検索)。次いで、所持しているコンテンツおよび、検索結果履歴に対して、クエリの関連情報を検索し、その結果をポータルエージェントへ返信する。

## 2.5 検索要求受信ログ利用の効果

前節の 1. および 2. のように、検索結果履歴と検索要求受信ログ双方を検索に利用することで、3. のブロードキャストの回数を削減することができ、効率的な検索が実現できる。また、検索結果に対しては、ユーザの判断が加えられ、その判断結果も、検索結果履歴ファイルの中に加わることとなる。

積極的に情報を作成しているユーザの検索エージェント程、クエリに関連する情報を、検索結果として返すことができ、それゆえ、他の検索エージェントからクエリを受信する回数が増える。従って、これら他の検索エージェントからの検索要求が、検索要求受信ログに溜っていくこととなる。このとき、同一のクエリや、類似のクエリ、意味的に関連したクエリなどを何度も送ってくる検索エージェントは、自分にとっても興味があり、かつ、自分の持たない情報を保持している可能性があると考えられる。検索要求受信ログを利用することで、容易にこのような検索エージェントに、クエリを送り、必要な情報を得ることが可能となる。

従って、類似した、あるいは関連した情報ならびに興味を持つ検索エージェントどうしの方が、そのような情報を持たない検索エージェントとよりも、クエリの送受信が多くなると予想される。これは、すなわち、同じトピックに関心を持つユーザの検索エージェント同士の間でグルーピングが行われ易くなることを意味する。

しかしながら、検索要求受信ログを利用した検索は、必ずしも成功するとは限らない。たとえば、自分の与えた情報しか、持っていない場合も可能性としてあるからである。このような場合の対処として、「検索要求受信ログを利用して検索に失敗した場合には、検索要求受信ログの中から、このクエリと、失敗を返信してきたエージェントのアドレス情報とを削除する」などの方法が考えら

れるが、詳細な議論は、別稿で行いたい。

### 3. KODAMA による本方式の実装

#### 3.1 KODAMA

本方式を我々の研究室で研究・開発しているマルチエージェントシステム KODAMA (Kyushu university Open & Distributed Autonomous Multi-Agent) [Zhong 02] を用いて実装した。

KODAMA システムは階層的コミュニティ構造を取るマルチエージェントシステムである。すべての KODAMA エージェントはコミュニティに所属し、各コミュニティは階層構造を成す。このような構造を取る理由は、階層構造が分散ソフトウェアシステムを含む複雑な問題に取り組むのに適しているためである。各コミュニティには、そのコミュニティを代表するポータルエージェントが一つだけ存在する。ポータルエージェントはコミュニティ内のエージェントの管理と、コミュニティ間での通信を仲介する役割を担当する。更に、ポータルエージェントが上位コミュニティのメンバとなることにより、コミュニティの階層構造を実現する。また、KODAMA には、ACZ (Agent Communication Zone) と呼ばれるエージェントのための物理的通信環境を支援する機構が用意されている。ACZ は、エージェントの論理的なコミュニケーション世界と、物理的なネットワーク通信環境世界とをつなぐ役割を担うミドルウェアであり、エージェントの論理世界から、物理的ネットワーク世界を完全に切り放す。これにより、各エージェントは、物理的なアドレスを意識することなく、他のエージェントと通信を行うことができる。

KODAMA エージェントは、カーネルユニットとアプリケーションユニットから構成される。カーネルユニットは、各エージェントから共通に利用される基本モジュールからなる。たとえば、コミュニティを構成するための機能やエージェント間通信を実現するための機能が含まれる。アプリケーションユニットは、個々のエージェントの性能や特徴を決める機能を実現するモジュールであり、物理的には、plug-in モジュールの集合として構成される。この plug-in モジュールを記述することで、アプリケーション固有の動作をエージェントに行わせることができる。

#### 3.2 本検索方式の動作

検索エージェントは、KODAMA エージェントのアプリケーションユニットに本検索方式の動作を記述した plug-in モジュールを組み込むことで作成した。また、ユーザインタフェースエージェントならびに履歴管理エージェントについても、それぞれの機能動作を実現するよう記述した plug-in モジュールを組み込み実現した。検索エージェントならびに履歴管理エージェントは、同一の検索

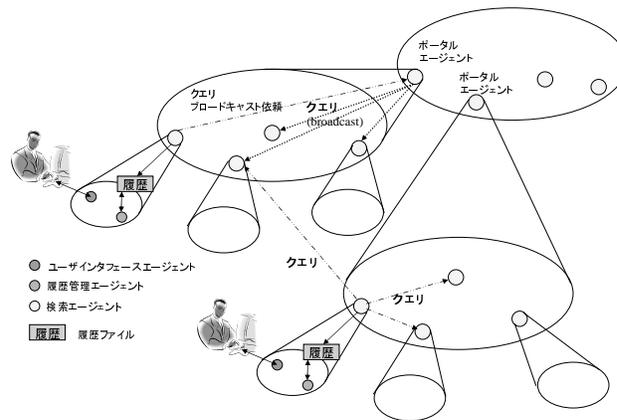


図 2 各エージェントとコミュニティ構造

結果履歴ファイルと検索要求受信ログファイルにアクセスを行う。この検索エージェントが任意のコミュニティに配置されることで、システムは稼働を開始する。ユーザインタフェースエージェントならびに履歴管理エージェントは、検索エージェントをポータルとするコミュニティに所属する (図 2 参照)。これは、検索エージェントが所属するコミュニティ側から見れば、そのコミュニティの直ぐ下位のコミュニティである。ユーザインタフェースエージェントおよび履歴管理エージェントは、それらの所属するコミュニティのポータルにあたる検索エージェントとのみ通信を行う。

本検索方式では、2.4 節の手順 1. および 2. では、クエリを相手先のエージェントに直接送信する。一方、手順 3. では、コミュニティ内の全検索エージェントに対して、クエリのブロードキャストが行われる。

コミュニティメンバへのクエリのブロードキャストは、ポータルエージェントが行う。ポータルエージェントは、コミュニティ内の検索エージェントから「ブロードキャスト依頼」と「検索結果受信数」を受け取ると、まずコミュニティ内の全エージェントに、受け取ったクエリと、そのクエリに対する返答が可能か否かを尋ねるメッセージを送る。このブロードキャストメッセージは、一定数段 (例えば、2 ~ 3 段) 下の下位コミュニティに対しても同時に行われる。メッセージを受け取った検索エージェント (ただし、依頼を出した当のエージェント以外) は、この問に対し、YES か NO でポータルエージェントに返答する。ポータルエージェントから受け取ったメッセージを、自分がポータルとなっている下位コミュニティの検索エージェントにブロードキャストした場合には、ブロードキャストした検索エージェントの YES の返答をまとめて、ポータルエージェントに渡す。ポータルエージェントは、コミュニティ内のエージェントからの YES の返答数が「検索結果受信数」に達した時点で、YES と返答してきたエージェントのアドレスリストを、ブロードキャスト依頼をしてきた検索エージェントに送る。ブ

ロードキャスト依頼をした検索エージェントは、このアドレスリストを使って、そのアドレスの検索エージェントにクエリを送信し、結果を受け取る。仮に、検索エージェントからの YES の返答数が「検索結果受信数」に達しない場合には、コミュニティのポータルエージェントが所属する上位コミュニティのポータルエージェントに対して、検索依頼を行う。ただし、その際にも、コミュニティ間の段数は、一定段数以内とする。

## 4. 実験

### 4.1 実験の目的

本実験の目的は、本検索方式で採用した検索要求受信ログを利用する効果を調べることである。その効果を調べるために、2.4節の手順1と手順3を利用する手法（検索結果履歴ファイルを利用する検索手法）と、手順3のみを利用する手法（常にクエリをブロードキャストする検索手法）との比較を行う。

調査対象項目は、各検索エージェントの平均受信メッセージ数、ならびに獲得できた検索結果の平均数である。ここで、検索エージェントの平均受信メッセージ数は、クエリに関連する情報を求めるのに役に立たない不要なメッセージ数の減少度合いを調べるためであり、検索結果の平均数は、求める関連情報を漏れなく得られたかの度合いを調べるために求める。

### 4.2 準備

本実験では、検索に利用するデータとして、Yahoo! JAPAN[Yahoo 03]に登録されているWebサイトのコンテンツを利用した。Yahoo! JAPANでは、登録されたWebサイトがカテゴリ分けされているが、このカテゴリ名を検索エージェント名とし、検索エージェントのコンテンツ（2.3節参照）として、そのカテゴリに登録されているサイトから収集したWebページを利用した。

ここで、使用したカテゴリは、大きく分けて、動物、スポーツ、コンピュータ、医療、金融の5つに分類されるものである。実験では、各カテゴリから、登録数の多い順に20個（計100個）を選び、利用した。

検索エージェントが利用するクエリは、そのエージェントのコンテンツから、出現頻度の高い名詞10個を自動抽出して作成した。この時、名詞の抽出には、形態素解析システム茶筌[松本 03]を用いた。

今回の実験では、検索エージェントは、1つのコミュニティ上に全てを配置した。従って、今回の実験では、コミュニティ階層間のメッセージ通信は行われない。

クエリ $q$ と検索対象ファイル $doc$ との類似度値 $Sim(q, doc)$ の計算は、 $q$ と $doc$ の各単語ベクトル $\vec{q}$ と $\vec{doc}$ の内積を求めることで行った。この時、各ベクトル $\vec{q}, \vec{doc}$ の要素の値は、それぞれの中で出現する単語の頻度である。関連

情報として選択する基準は、 $Sim(\vec{q}, \vec{doc}) \geq 1$ とした\*6。

### 4.3 実験結果

まず、検索結果受信数を、3個、5個、7個、または10個とし時の、クエリ投入回数（検索回数）が増加していく際の平均受信メッセージ数の変化を調べた。図3には、検索結果受信数が、5個と10個の場合についての、各手法の比較結果を示す。

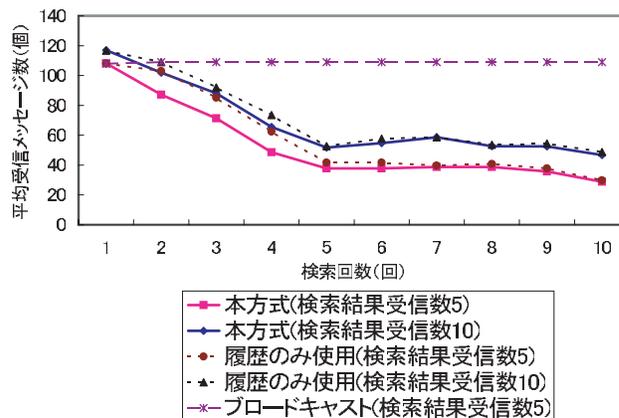


図3 各検索結果受信数における平均受信メッセージ数の比較

本方式も検索履歴を用いる手法も、検索の回数を重ねるにつれ、平均受信メッセージ数が減少していくことを確認した。これは最初のほうの検索では、ブロードキャストを行わなければならないが、その後、検索を重ねるにつれて、検索結果履歴ファイルや検索要求受信ログファイルを利用することで、検索結果受信数を満たすクエリ送信先を見つけることができるようになり、ブロードキャストを行う回数が減るためである。また、検索回数の早い回（5回以内）では、検索結果履歴を優先的に利用したとしても、検索要求受信ログの利用が有効となることもわかった。なお、本方式と検索結果履歴を用いる手法との間の差が小さいのは、本方式が、検索要求受信ログを優先的に利用していないことにも関係すると思われる。これについては、スペースの関係から、別稿で議論したい。

次に、表2に各検索結果受信数に対する検索結果の平均獲得数の変化を示す。ブロードキャストによる方法とは、検索結果受信数が3個のときは、多少、獲得できる検索結果に差が出ているものの、検索結果受信数が5個、7個、10個のときは、ほとんど差が出ていない。すなわち、検索結果履歴ファイルと検索要求受信ログを利用することで、探索のためのメッセージ数を減らしながら、必要な情報を獲得できていることを意味している。なお、本方式と、検索結果履歴を利用する方法との間にも、検索結果の獲得効果に大きな差は見られなかった。

\*6 つまり、クエリ中の単語を1つでも含むファイルを、関連情報を持つファイルとして検索する。

ただし、これについては、検索エージェント数や、クエリ、クエリと検索対象間の類似度計算式なども依存するため、更に調査を行う必要がある。

表 2 各検索エージェントの検索結果の平均獲得数

検索結果受信数	3	5	7	10
本方式	489.9	714.5	772.8	826.0
検索履歴を用いる手法	487.4	701.3	752.8	816.9
ブロードキャスト	604.5	735.1	781.5	829.8

最後に、各クエリ毎の検索に対して、検索エージェント間で交されるメッセージ 1 個あたりの検索結果の獲得数の比較を図 4 に示す。

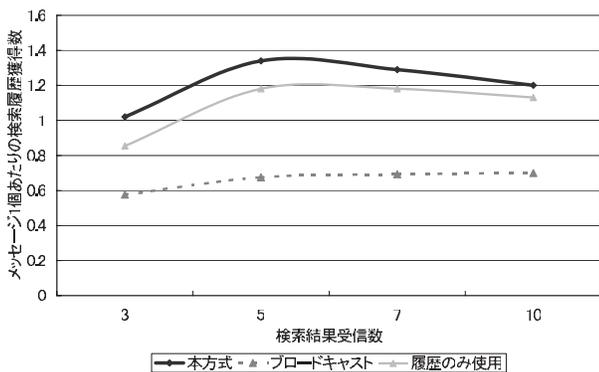


図 4 検索効率

図 4 より本方式のほうが、明らかに検索効率が良いことがわかる。今回の実験では本検索方式の検索結果受信数が 5 個の場合が、最も検索効率がよく、ブロードキャストによる検索の 2 倍の効率が得られている。その後、検索結果受信数が増えるにつれ、効率は緩やかに悪くなっている。ブロードキャストによる検索は検索結果受信数が 3 個から 5 個に増えるにつれ、効率が良くなっているが、その後検索結果受信数を増やしてもほとんど変化していない。これは、検索エージェント数が、高々 100 個だからであると思われる。

検索エージェント数を 1,000 個、10,000 個と増やして行った時に、検索結果受信数と、検索結果獲得数との変化については、今後確かめる必要がある。また、類似度計算では、今回は、検索結果獲得数を重視して、非常に単純な式を用いたが、今後、よりユーザの興味に適合する情報を優先して検索するための実験を行う必要がある。

## 5. 関連研究

分散情報検索は、複数の検索システムの中から、まず、検索要求を投げる検索システムをいくつか選択し、選択された検索システムから返された検索結果を統合して、ユーザに提示する検索方式であり、主に、検索システム側がサーバとなり、ユーザ側システムがクライアントとなる

サーバ・クライアント方式で実現される。分散サーバの各情報源の内容を示すものは、resource description(情報源特徴)と呼ばれ、これを得ることで、これまで数多く提案されてきた内容に基づくデータベース選択アルゴリズムを利用することができる [Callan 01]。この情報源特徴を得る方法として、検索要求に基づく標本抽出法 [Callan 99] が提案されている。この手法は、複数の検索要求を、各分散データベースに投げ、その返された内容に基づき、データベースの特徴を調べる方法である。これに対して、本方式はエージェントコミュニティを利用した P2P 型の検索方式であり、他のユーザエージェント(データベースを持つ)から送られてきた検索要求を基に、自分の求める情報を持つエージェントの特定を試みる。この方法は、「何度も検索要求を送ってくるエージェントほど、自分の持つ情報源に対して興味を持つことであるため、逆に、そのようなエージェントは、自分にとって興味深い別の情報を持っている可能性が高い。」というアイデアに基づいている。

P2P 型の検索方式としては、Freenet[Clarke 01] や Chord[Stoica 01], Gnutella[Gnutella 00], Napster [Napster 00] など多数が提案されている。Freenet や Chord は、純粋な P2P 型で、Gnutella のようなブロードキャストは行わず、Napster のような中央集権型サーバに頼ることもない。Freenet は分散された匿名での情報共有と情報検索機能を提供する。また、Chord は匿名機能は有しないが、効率の良いノード検出プロトコルを提供する。これらは、各ノードが持つ情報の内容については触れず、その情報に付与されたキーワードだけを基に、ノードの探索を行う。

Freenet や Chord と同様に、本方式も、効率良い高精度な検索を実現することを目的とするが、階層的エージェントコミュニティの枠組みをもつ、マルチエージェントシステム KODAMA [Zhong 02, 高橋 03] を用いて効率の良いコミュニティ内・間のエージェント間通信を実現しようとする点と、各エージェントが持つデータの内容に対する検索を行う点、ならびに、各エージェントに送られてきた検索要求の履歴を基に、自分にとって最も適した情報を持つエージェントを検出しようとする点異なる。

自分が興味ある項目について他のユーザの過去の意見を基に、自分にとって興味ある新たな情報を見つける(推薦してもらう)手法は、協調フィルタリング (Collaborative Filtering) (e.g. [Goldberg 92, P. Resnick 94, Shardanand 95, Balabanovic 97, Herlocker 99, Melville 01, Sarwar 01]) として知られている。協調フィルタリングには、データの希薄さ (sparsity) と、最初の評価者の問題 (first-rater problems) という特有の問題が存在するが [Good 99], 情報の内容自体についての解析を行う情報フィルタリング (e.g. [Lang 95]) などと一緒に用いることで、その問題を解決する試みが行われている (e.g.

[Balabanovic 97]). 情報フィルタリングもしくは協調フィルタリングを利用して、情報の推薦を行うシステムは、推薦システム (Recommender System) (e.g. [Schafer 99]) として知られる。その推薦システムでは、ユーザからの要求を待たずに、積極的に新たな情報をユーザに対して推薦する PUSH 型と、ユーザの要求を待って推薦を行う PULL 型に分類することができる。しかしながら、解析を行う際に取られる手法のほとんどは、全てのデータを集めて解析を行う集中型である。

一方、本方式は、エージェントコミュニティを利用した分散管理手法を取り、各エージェントに送られてきた検索要求の履歴を基に、自分にとって最も適した情報を持つエージェントを知らせる推薦システムと見ることができる。

## 6. 終わりに

本稿ではエージェントを用いたコミュニティベースの P2P 型情報検索方式を提案した。本検索方式は、検索結果履歴と検索要求受信ログを検索に用いることによって、効率のよい検索を実現する。実験の結果、本方式が仮定していた「通信量の削減」と「検索効率の向上」の効果があることを確認した。しかしながら、今回の実験で利用したエージェントコミュニティは一つであるので、ブロードキャストの範囲を制限するというコミュニティ利用の利点が示されていない。そこで、複数のエージェントコミュニティを利用して、最も効率の良い検索を実施する階層化の方法やコミュニティの形成方法についての検討が必要である。また、各検索エージェントの「give and take」の効果についても、早急にその評価実験を行う必要がある。

その他の課題として、ユーザの興味に適合する情報を優先して検索する手法についての検討、コミュニティ単位・エージェント単位で情報の流通範囲を限定する機能の検討などがあげられる。

## 謝 辞

本研究の一部は、通信放送機構 (TAO) からの委託研究「相互接続時のセキュリティポリシーの管理技術に関する研究開発」の補助により行われた。

## ◇ 参 考 文 献 ◇

- [Balabanovic 97] Balabanovic, M. and Shoham, Y.: Content-Based, Collaborative Recommendation, *Communications of the ACM*, Vol. 40, No. 3 (1997)
- [Callan 99] Callan, J., Connell, M., and Du, A.: Automatic discovery of language models for text databases, *ACM SIGMOD*, pp. 479-490 (1999)
- [Callan 01] Callan, J. and Connell, M.: Query-Based Sampling of Text Databases, *ACM Transactions on Information Systems*, Vol. 19, No. 2, pp. 97-130 (2001)

- [Clarke 01] Clarke, I., Sandberg, O., Wiley, B., and Hong, T. W.: Freenet: A Distributed Anonymous Information Storage and Retrieval System, *Designing Privacy Enhancing Technologies: International Workshop on Design Issues in Anonymity and Unobservability*, <http://www.doc.ic.ac.uk/~twh1/academic/> (2001)
- [Gnutella 00] Gnutella, : <http://gnutella.wego.com/> (2000)
- [Goldberg 92] Goldberg, D., Nichols, D., Oki, B. M., and Terry, D.: Using collaborative filtering to weave an information tapestry, *Communications of the ACM*, Vol. 35, pp. 61-70 (1992)
- [Good 99] Good, N., Schafer, J. B., Konstan, J. A., Borchers, A., Sarwar, B. M., Herlocker, J. L., and Riedl, J.: Combining Collaborative Filtering with Personal Agents for Better Recommendations, in *AAAI/IAAI*, pp. 439-446 (1999)
- [Herlocker 99] Herlocker, J. L., Konstan, J. A., Borchers, A., and Riedl, J.: An Algorithmic Framework for Performing Collaborative Filtering, in *SIGIR99*, pp. 230-237 (1999)
- [Kanfer 97] Kanfer, A., Sweet, J., and Schlosser, A.: Humanizing the Net: Social Navigation With A "Know-Who" Email Agent, in *The 3rd Conference on Human Factors & The Web*, <http://www.ncsa.uiuc.edu/edu/trg> (1997)
- [Lang 95] Lang, K.: NewsWeeder: learning to filter netnews, in *Proceedings of the 12th International Conference on Machine Learning*, pp. 331-339, Morgan Kaufmann publishers Inc.: San Mateo, CA, USA (1995)
- [Melville 01] Melville, P., Mooney, R. J., and Nagarajan, R.: Content-Boosted Collaborative Filtering, in *SIGIR-2001 Workshop on Recommender Systems* (2001)
- [Napster 00] Napster, : <http://www.napster.com/> (2000)
- [P.Resnick 94] P.Resnick, , N.Iacovou, , M.Suchak, , P.Bergstrom, , and J.Riedl, : GroupLens: Open Architecture for Collaborative Filtering of Netnews (1994)
- [Sarwar 01] Sarwar, B., Karypis, G., Konstan, J., and Riedl, J.: Item-Based Collaborative Filtering Recommendation Algorithms, in *WWW10*, pp. 285-295 (2001)
- [Schafer 99] Schafer, J. B., Konstan, J. A., and Riedl, J.: Recommender systems in e-commerce, in *ACM Conference on Electronic Commerce*, pp. 158-166 (1999)
- [Shardanand 95] Shardanand, U. and Maes, P.: Social Information Filtering: Algorithms for Automating "Word of Mouth", in *Proceedings of ACM CHI'95 Conference on Human Factors in Computing Systems*, Vol. 1, pp. 210-217 (1995)
- [Stoica 01] Stoica, I., Morris, R., Karger, D., Kaashoek, M. F., and Balakrishnan, H.: Chord: A Scalable Peer-To-Peer Lookup Service for Internet Applications, in *Proceedings of the 2001 conference on applications, technologies, architectures, and protocols for computer communications*, pp. 149-160 (2001)
- [Yahoo 03] Yahoo, : <http://www.yahoo.co.jp/> (2003)
- [Yimam-Seid 03] Yimam-Seid, D. and Kobsa, A.: Expert Finding Systems for Organizations: Problem and Domain Analysis and the DEMOIR Approach, *Journal of Organizational Computing and Electronic Commerce*, Vol. 13, No. 1, pp. 1-24 (2003)
- [Zhong 02] Zhong, G., Amamiya, S., Takahashi, K., Mine, T., and Amamiya, M.: The Design and Application of KODAMA System, *IEICE Transactions INF.& SYST.*, Vol. E85-D, No. 04, pp. 637-646 (2002)
- [高橋 03] 高橋 健一, 雨宮 聡史, 鍾 国強, 松野 大輔, 峯 恒憲, 雨宮 真人: エージェント・プラットフォームを相互利用するためのメッセージ通信プロトコルの提案と実装評価, 電気学会論文誌 C(電子・情報・システム部門誌), Vol. 123-C, No. 8, pp. 1503-1511 (2003)
- [松本 03] 松本裕治, 北内啓, 山下達雄, 平野善隆, 松田寛, 高岡一馬, 浅原正幸: 日本語形態素解析システム「茶釜」ver. 2.3.1, <http://chasen.org/> (2003)