

P2P ネットワークにおける コンテンツのグループ化に関する研究

佐々木 拓也^{†1} 澤本 潤^{†2}
加藤 貴司^{†2} 和田 雄次^{†3}

近年，情報通信技術の発達にともない情報の多様化，多在化が急速に進んでいる．情報の利用方法は利用者によって様々である．そういったなか，P2P ネットワークを利用した情報の共有への注目が高まり，多方面で利用されている．P2P 上のオーバーレイネットワークとして代表的なものが分散ハッシュテーブルだが，検索の効率性に欠けているという問題点もある．本稿では，検索の効率性を向上させるため，利用者の検索履歴と検索頻度を利用したコンテンツのグループ化手法とそのグループに基づく効率の良いコンテンツ検索法を提案し，シミュレーションによる評価を行った．

A Research of Contents Grouping on P2P Network

TAKUYA SASAKI,^{†1} JUN SAWAMOTO,^{†2} TAKASHI KATOH^{†2}
and YUJI WADA^{†3}

In recent years, the diversification and ubiquitousness of information is rapidly advancing due to the development of the information and communication technology. As for the usage of the information, it depends on the user and the attention of sharing of information using P2P network rises, and it is becoming to be used in many fields. The distributed hash table (DHT) is one of the typical overlay networks on P2P. However, there is a problem of lacking of flexibility in the retrieval on DHT. In this paper, to improve the convenience of the contents retrieval, we propose an efficient content retrieval method based on the grouping technique of the contents using the content access history and frequency and the evaluation by the simulation is performed.

1. はじめに

近年，情報通信技術の発達にともない情報の多様化，多在化が急速に進んでいる．たとえば，ショッピングモールなどの小売店では取り扱っている商品情報，在庫情報，キャンペーン情報を管理している．バス会社は市街地におけるバスの運行時間，運賃などの情報を管理し，各停留所の掲示板を利用しバス利用者へ情報を表示している．Web 販売システムでは商品情報，在庫情報の管理だけでなくインターネットを利用し商品の購入が可能など，情報を必要としている利用者によって使い方も様々である．

そういったなか，Peer-to-Peer (以下 P2P) ネットワークを利用した情報の共有への注目が高まり様々な方面で利用されている．P2P ネットワークを利用したアプリケーションの例として，Napster¹⁾，Gnutella²⁾ などがある．クライアントが接続したサーバから一方的にサービスの提供を受けるクライアントサーバモデルのシステムとは異なり，P2P 型システムでは，コンピュータどうしが相互に接続し平等な関係で直接，情報やサービスをやりとりすることによって負荷が分散される，耐故障性が高い，などの利点がある．

P2P ネットワークは検索メッセージの転送方式によって，非構造化オーバーレイネットワーク，構造化オーバーレイネットワークの 2 種類に分類することができる．非構造化オーバーレイネットワークはキーに対応する相手を探索するために，自ノードと隣接しているノードに対し検索メッセージを送信する．受信したノードがキーに対応したコンテンツを持っている場合は返答するが，持っていない場合はさらに隣接されたノードに対し検索メッセージを転送することになる．メッセージをホップするたびに転送メッセージ数が増えるため，ネットワークトラフィックが増えてしまう．一方，構造化オーバーレイネットワークは検索メッセージを転送する際の転送先を選択する方法をあらかじめ構造的に決めており，キーを保持しているノードが分かるようにしている方式である．代表的な方式として分散ハッシュテーブル (Distributed Hash Table) や Skip Graphs³⁾ などがある．

分散ハッシュテーブルは，各ノードに割り当てられたアドレス (ノードの ID) とコンテンツのハッシュ値を空間に写像し，その空間を複数のピアで分割管理することで，特定ピ

^{†1} 岩手県立大学大学院ソフトウェア情報学研究科
Graduate School of Software and Information Science, Iwate Prefectural University
^{†2} 岩手県立大学ソフトウェア情報学部
Faculty of Software and Information Science, Iwate Prefectural University
^{†3} 東京電機大学情報環境学部
Department of Information Environment, Tokyo Denki University

アに負荷が集中することなく大規模なコンテンツ探索を実現する。こうすることで、アドホック性とスケーラビリティの両立を目指している。そのため、分散ハッシュテーブルではフラッシング方式に比べ非常に少ないメッセージ数で検索を行うことができるが検索時に情報の識別子を指定しなければならないため、1度に複数のコンテンツを取得したい場合や、キーワードの部分一致などによる検索には不向きであるといわれている。分散ハッシュテーブルの代表例として、Chord⁴⁾、Kademlia⁵⁾などが存在しており、今日においても分散ハッシュテーブルのアルゴリズムの開発が行われている。すでに分散ハッシュテーブルを用いた構造化されたP2Pネットワークにおける、検索システムの研究と評価は種々行われている⁶⁾。分散ハッシュテーブルにおいてはコンテンツの柔軟な検索を行うことを目的とした階層分類化構造を導入した分散ハッシュテーブルの研究も事例として存在する⁷⁾。

このように、従来のP2Pネットワークにおける構造化オーバーレイネットワークとして普及している分散ハッシュテーブルは、完全一致検索しか行えないといった欠点が存在する。完全一致のみの検索は利用者にとって目的のコンテンツを探すのが難しく効率の良い検索が行えない。そこで、本稿では検索効率が不十分であるP2Pネットワークを利用した検索を行う際に、検索効率を向上させることを目的に、利用者の検索履歴と検索頻度を利用したコンテンツのグループ化手法とそのグループに基づく検索数の削減による効率の良いコンテンツグループ検索法を提案し、シミュレーションによって提案した手法の機能評価を行う。

2. P2P ネットワークにおけるコンテンツのグループ化

2.1 P2P ネットワークを利用したコンテンツ検索

P2Pネットワークを利用したコンテンツ検索は、サーバクライアント型の検索に比べ中央サーバを利用していない。P2Pネットワークは中央サーバに代わって、各ノードどうしが連携しコンテンツ情報を管理している。P2Pネットワーク利用者はキーを手がかりにキーに対応するコンテンツを検索し、コンテンツを保持しているノードと通信を行いコンテンツ情報の取得を行う。キーとコンテンツ情報をペアとした情報をインデックス情報といい、このインデックス情報をどのように管理するかが重要となる。

構造化ネットワークの主な手法である分散ハッシュテーブルは、このインデックス情報のキーにハッシュをかけ、ハッシュ値を利用しインデックス情報を管理している。文字が1文字異なっているだけでハッシュ値はまったく別の値になってしまうため、分散ハッシュテーブルでは完全一致検索しか行うことができないという欠点が存在している。

一方、非構造化ネットワークでは隣接したノードへ検索メッセージを転送して検索を行う

ため、完全一致検索、範囲検索、部分検索といった様々な検索手法を利用することが可能であるが、検索クエリが増えることによって、大きなネットワークトラフィックが生じてしまう。

2.2 コンテンツのグループ化

2.2.1 概要

本稿ではP2Pネットワークにおけるコンテンツのグループ化を、利用者による検索の連続性によって行う手法について提案する。

コンテンツがグループ化されていない場合、利用者が欲しいコンテンツの情報ごとにコンテンツの検索を行う必要がある。一方コンテンツがグループ化されている場合、あるコンテンツの検索を利用者が行った際に、そのコンテンツが属しているグループのコンテンツ情報を同時に取得することが可能となるため1回の検索で複数のコンテンツ情報を取得することができる。グループ化されているコンテンツ群は、多数の利用者の連続したコンテンツの検索によって行われているものであるため、同じ嗜好を持った利用者にとっても有用なコンテンツグループである可能性が高いと考えられる。コンテンツの主な例として料理の材料や、ファミリーレストランのメニューなどといった消費材など複数回検索されることが想定されるコンテンツなどを想定している。

2.2.2 関連研究

本項ではコンテンツのグループ化に関する関連研究について述べる。

対象コンテンツがWebページの場合、文献8)ではWebページの特徴ベクトルから類似度を判定しグループ化を行っている。文献9)ではWebページのメタデータを利用したグループ化を検討、文献10)ではWebのリンク関係から重要なページを探している。

Webコミュニティのグループ化に関する研究においても、Webマイニングの技術を用いて、Web上の情報だけから特定のコミュニティの人間関係を自動的に抽出¹¹⁾、参照の共起性からコミュニティを発見する¹²⁾など様々な研究が行われている。

モバイルエージェントを利用し同じ嗜好を持った人の仲介を行い、嗜好の関連性によるグループ化¹³⁾を行う研究も存在している。

以上のように、グループ化の研究はグループ化の対象が違っても多数存在している。本研究のグループ化を行う対象はP2Pネットワーク内のノードが保持しているコンテンツである。またP2Pネットワークのノードは頻繁に参加、離脱を行うためにその対応が必要である。本稿での提案方式では、利用者の検索のたびに自コンテンツと関連性の高いコンテンツグループの更新が行われるので、ノードの参加や離脱など変化に自然に対応できている

といえる。また、検索が行われるたびにグループの更新を行うことは、新しいコンテンツの発見や、リアルタイムに利用者の嗜好やネット上での流行などの傾向を反映できるといった特徴にもなっている。

2.3 コンテンツのグループ化における関連性強度

コンテンツのグループ化は P2P ネットワーク内のコンテンツ情報を保持しているノードに対し、利用者の検索の連続性を基に行っている。検索の連続性とは 2 つのコンテンツが利用者によって連続して検索されているかどうかを示している。利用者が複数のコンテンツを連続して検索した場合、それらのコンテンツは関連性強度が強くなる。関連性強度とは、2 つのコンテンツがどれだけ利用者にとって連続的な検索の価値があるかの強度のことである。関連性強度を表す式を式 (1) に示す。ここで、 $K_{s,i}$ はコンテンツ s と i の関連性強度である。 $v_{s,i}^n$ はノード n におけるコンテンツ s とコンテンツ i の連続検索の有無について示している。ノード n においてコンテンツ s と連続して検索されたコンテンツ i が存在する場合は 1 となり、存在しない場合は 0 となる。つまり関連性強度は利用者が連続して検索した場合に 1 ずつ増加していく。

一方検索利用されなくなったコンテンツグループは時間が経過することで関連性強度が弱まりグループ化が解消される。関連性強度の減少率については表 1 の「各パラメータの説明」において説明している。

$$K_{s,i} = \sum_{n \in \text{node}} (v_{s,i}^n) \quad (1)$$

3. 構 成

3.1 ノード構成

本稿では一般のオーバーレイ P2P ネットワークのノードに検索履歴リストとグループリストを保持させる。ノード間の通信に検索履歴リストとグループリストのやりとりを追加させる。図 1 は、P2P ノードに「検索履歴リスト」、「グループリスト」を持たせた状態を示している。

以下図 1 について説明する。

3.1.1 コンテンツ情報

ノードが保持するコンテンツファイルである。また本稿ではコンテンツの関連性によるグループ化を検証することを目的としているため、1 ノードにつき 1 コンテンツのみ保持させている。すなわち、ノード ID = コンテンツ ID とする。また、同様にコンテンツのサイズ



図 1 ノード構成のイメージ

Fig. 1 Image of P2P node construction.

や実際の内容、および取得時間についても詳細を必要としないので、ここでは詳細を省略したりパラメータとしての考慮を行ったりしていない。ノードがコンテンツ情報を複数持つ場合、コンテンツ数分のグループリストを保持させるなど拡張は容易であると考えられる。

3.1.2 分散ハッシュテーブル

分散ハッシュテーブルでは、各ノードのインデックス情報を持っている。基本的には分散ハッシュテーブルはハッシュを用いてインデックス情報を分散して保持することになるので、コンテンツ情報を保持しているノードの場所を知っているノードはハッシュ値によって左右される。また分散ハッシュテーブルの種類にもよるが各ノードが保持できるインデックス情報は 20 個など限られている。インデックス情報をホップすることで目的のコンテンツを持ったノードの情報を取得することになる。

3.1.3 検索履歴リスト

ノードの検索履歴リストには、利用者の 1 サイクル分の検索履歴を保持する。今回の検索手法では利用者がコンテンツ情報を、分散ハッシュテーブルを利用し検索した際にヒットしたノード ID を検索履歴リストに保持する。利用者が目的とするコンテンツを検索し終了した時点で、検索履歴リストにたまっているヒットしたノードに対してコンテンツの取得要求を出す。この連続的に検索する単位を 1 サイクルと呼んでいる。利用者の検索履歴はコンテンツ情報を検索し、コンテンツを取得するときに利用され、コンテンツの取得が完了した時点でクリアされる。ここでいうノード ID とはノードの IP アドレスやホスト名のことである。

3.1.4 グループリスト

自ノードと関連性の強いノードを保持する。関連性が強いということの判断は 2.3 節で述

べた関連性強度と「3.3 節グループ化におけるパラメータ」で説明している関連性強度基準値（閾値）を基に判断している。リストはノード ID とノードとの強度を保持し、強度の強さによって降順にソートされている。ノードの強度は時間が経つにつれて弱まっていく。この関連性の強いノードは、自ノードが検索された際に自ノードのコンテンツ情報と合わせて同時に利用者ノードへ送信される。

3.2 検索履歴と検索頻度を利用したコンテンツのグループ化手法

検索履歴と検索頻度を利用したコンテンツのグループ化におけるノード間の通信について以下に具体的に記述する。図 2 はノード間の通信イメージを示している。以下、① から⑤ の操作の前に、分散ハッシュを利用した A ノードからの一連の検索命令は実行され、希望するコンテンツの所在は検索履歴リストに保持されているものとする。

- ① A ノードは検索履歴リスト内の 1 つのノードを B ノードとし、その B ノードのコンテンツを要求する。
- ② A ノードは自ノードの検索履歴リストを B ノードへ送信する。A ノードの検索履歴リストを受信した B ノードは、自ノードのグループリストへ格納する。グループリストに同じコンテンツ ID などのコンテンツ情報が存在していた場合は、関連性強度を 1 増加させる。
- ③ B ノードは A ノードから要求されていた自ノードのコンテンツ情報を A ノードへ送信する。
- ④ さらに、B ノードは自ノードのグループリストに存在しているコンテンツ情報の中で、

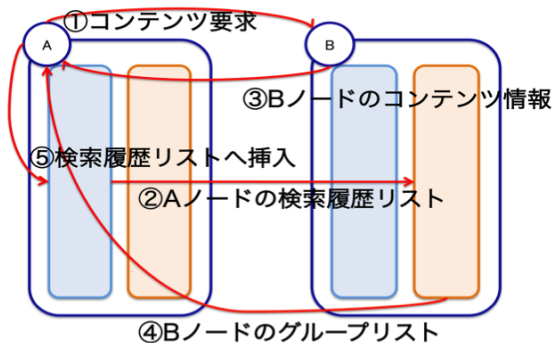


図 2 コンテンツのグループ化におけるノード間通信のイメージ
Fig. 2 Node communication process for the grouping of contents.

関連性強度の基準値を上回っているコンテンツを A ノードへ送信する。

- ⑤ A ノードは B ノードのコンテンツ情報を検索履歴リストへ追加する。B ノードから受信したグループリストを利用し、A ノードはグループ化されたコンテンツを要求する（この機能を推薦機能と呼ぶ）。

検索履歴リスト内のすべての B ノードからのコンテンツの取得を確認した時点で A ノードは自ノードの検索履歴リストを消去（クリア）する。以上の一連の利用者の動作を、ここでは 1 サイクルと呼ぶ。

3.3 グループ化におけるパラメータ

ここでは本アルゴリズムに関連するパラメータについて説明する。

利用者の検索履歴数の最大保持数、各コンテンツが持つ自コンテンツと関連性が高いコンテンツ情報を保持するグループリストの最大保持数などのリストのサイズパラメータがあげられる。また、一定の時間が経つことによるグループ内コンテンツとの関連性の減少率、各コンテンツとの関連性がある値以上であれば関連性があると判断する基準値といったグループを形成するうえで関連性に関するパラメータが重要となる。これらのパラメータについて表 1 において説明する。

4. 性能評価

4.1 実験

4.1.1 実験概要

コンテンツのグループ化の実験はオーバーレイ構築ツールキットである「OverlayWeaver」内のツールである分散環境エミュレータを利用し行った¹⁴⁾。エミュレータを起動したサーバは、多数のノードを生成し管理することが可能である。エミュレータでは各ノードの生成から検索、コンテンツ取得などといったノードの動きを指定したシナリオファイルを用意しておく。エミュレータは起動時にシナリオファイルを読み込み、逐次処理を行っていく仕組みになっている。本実験ではグループ化を行う際に関連してくるパラメータの設定値ごとにエミュレーションを行い、結果を出力する。その出力されたデータを基に評価値を算出して比較、検証を行っている。実験時に設定値を変更する各パラメータについては 4.1.2 項で説明する。

シナリオファイルの流れは、図 3 のとおりになっている。検索対象ノードはあらかじめ初期設定で仮グループを設定しておく。本実験では、3 章で提案したアルゴリズムの基本的機能の検証を目的とする。そのため、前述したとおりノードには実在するコンテンツ情報を持たせているわけではない。そこで本実験では生成したノードが保持するコンテンツを初期

表 1 グループ化におけるパラメータ
Table 1 System parameters for contents grouping.

パラメータ名	説明
検索履歴リスト最大数	利用者の検索の際に、ヒットしたコンテンツの情報を保持する。この値が大きくなるほど、利用者の多くの検索履歴を保持することが可能となる。本アルゴリズムでは、利用者の1サイクルの平均的な検索数を考慮してリストの最大数を設定する必要がある。ここでは、関連性を持って連続的に検索する対象数を3~5程度と仮定した。
グループリスト最大数	コンテンツと関連性の強いコンテンツ情報を保持する。この値が大きくなることによって自ノードが保持しているコンテンツと関連性が高いコンテンツの数が大きくなる。逆に少なくなるとグループ内のコンテンツ数が制限される。ここでは、関連性が強いいわゆる推薦可能な対象数として10以下程度を仮定した。
関連性強度減少率	一定時間(例えば1日)がたった際に減少する関連性強度の率を表す。この値を大きくすることでグループ内のコンテンツ間の関連性が急速に減少しやすくなる。逆に小さくなることでコンテンツ間の関連性が長く持続することになる。妥当な値は、対象や環境にも関連してくるので確定することは困難と考え、5%~50%程度の範囲を指定することとする。
関連性強度基準値	コンテンツ間の関連性の有無を決定する閾値を表す。この値大きくすることで、より関連性の強いコンテンツとのグループ化を可能とする。逆に小さくすることで、関連性の低いものもグループ化として認識されることになる。予備実験などを行なうことによりその範囲を1~3程度とした。

- ① 指定した軌道ノード数文のノード生成を行う
- ② 起動ノードでオーバーレイネットワークを形成する
- ③ コンテンツ保持ノードは分散ハッシュテーブルに登録を行う
- ④ 各ノードがキーワードによるコンテンツ検索を行う
- ⑤ 各ノードは検索したコンテンツのデータ取得を行う
- ⑥ コンテンツ保持ノードの関連性強度減少を行う

図 3 シナリオファイルの一連の流れ
Fig. 3 Typical flow of the scenario file.

設定で関連性が高いという設定のもと、仮グループを設定しておく。その設定においてアルゴリズム上変更可能なパラメータの設定値を変化させることで、グループ化がどの程度行われるかを検証する。1日の初回の検索はランダムで行う。2回目以降の検索に関しては、検索傾向度で設定した値を基に仮グループ内から検索を行うか、他の仮グループから検索を行うかを判別していく。また仮グループ内のコンテンツ群に、人気コンテンツを設定し、各ノードが検索を行う際は人気度の重み付けによって検索にある程度の偏りを持たせるように設定している。人気度とは、現実社会においてもコンテンツによって利用者からの検索頻度に差が生じることがあるが、そういったことを反映するために初期設定において仮グループ内のコンテンツに重み付けを行っていることを意味する。

実験では、図3内で示した④、⑤のコンテンツ検索、取得の一連の流れを各ノードごとに実験時パラメータの検索回数期待値を平均とする正規分布による回数分繰り返して行うことで1日分の動作とする。本実験はこの1日分の検索・取得を30日分を行っている。各ノードが1日の流れをひととおり終えたところで、関連性強度減少の処理を行い(⑥)、各ノードの情報をデータとして出力している。つまり実験をひととおり終えた段階で30日分のデータの出力が行われている状態になる。

4.1.2 パラメータ

グループ化に関連するパラメータを表2に示す。

ノード数、検索傾向度、検索回数期待値はシナリオファイルを利用して変更を行う。検索履歴リスト保持数、グループリスト保持数、関連性強度減少率、関連性強度基準値、推薦有無のパラメータに関してはオーバーレイネットワークを変更する。

ノード数は、起動ノード数と其中でコンテンツを保持しているノード数とを示している。本実験では、起動している全ノードがコンテンツを保持しているとは限らないとしている。現実でも情報取得者が必ず情報提供をしているとは限らないためである。そのため、コンテンツ保持ノード数は1,000ノードと固定し検索ノード数の変更を行い両者の割合を変更させ実験を行っている。

4.1.3 各実験のパラメータ設定

各実験において設定したパラメータを表3に示す。今回の実験では、実験PCの性能上、ノードを生成しシミュレーションできる起動ノード数の上限が10,000ノードであったので、その値を起動ノード数とし、それを基準にコンテンツ保持ノード数を設定した。現社会において1つのデータベースに対し何人もの人がアクセスしている。最低でもデータ数と利用者数の比率は1:10以上だと考えられるため起動ノード数10,000に対し1,000のコンテン

表 2 グループ化に関連する各パラメータの説明
Table 2 Explanation of parameters concerning grouping.

パラメータ名	説明
起動ノード数	実験時に起動する全ノード数.
コンテンツ保持ノード数	コンテンツを保持するノード数.
検索履歴リスト最大数	検索を実行する利用者ノードが保持する検索履歴リストの最大エントリー数.
グループリスト最大数	自ノードのコンテンツと関連性の高いノード情報を保持するグループリストの最大エントリー数.
関連性強度減少率	1 日経過した際に各ノードとの関連性強度が減少する割合. 減少率が X%, 現在の関連性強度を S とした場合, 減少後の関連性強度は $(S \cdot X/100)$ となる.
関連性強度基準値	グループ化が行われているかを判断するための関連性強度の閾値.
検索傾向度	各利用者が 1 日に検索するコンテンツの傾向度. この値により検索を行う際に予め設定した仮グループ内のコンテンツの検索, 取得を行う傾向を設定する.
検索回数期待値	各利用者が 1 日に検索を実施する回数の期待値.
推薦機能の有無	検索した際に, 検索対象ノードのグループリストを利用者へ推薦し, そのグループリストを利用し検索を行うかどうかの有無.

表 3 各実験のパラメータ設定
Table 3 Parameters set for the experiment.

パラメータ名	パラメータ値
コンテンツ保持ノード数	1000 ノード
起動ノード	10000 ノード
検索履歴リスト最大数	5
グループリスト最大数	10
関連性強度減少率	5%, 10%, 30%, 50%
関連性強度基準値	1, 2, 3
検索傾向度	20%, 50%, 80%
検索回数期待値	3, 6, 9
推薦機能の有無	アリ, ナシ.

ツを保持しているノードを設定した.

関連性強度減少率は, 3.3 節において説明したように, 確定が困難であるが, 最大 50% (1 日経過することにより関連性強度が半減するという極端なケース) とし, 漸次減少させることによりその効果を確認する設定とした. 関連性強度基準値については, 予備実験により今回のパラメータ設定環境での妥当性は, 3 となった. 確認のため, 本実験において 1~3 の値についてシミュレーションを行った. 検索傾向度については, コンテンツ利用のリピータの割合を示すパラメータであるが, 今回の実験のようにジャンルを固定した実験環境での妥当性を確認することが困難であるため, 20%~80%の範囲のサンプル値として設定した. 検索回数期待値は, 一般ユーザによるインターネット検索回数が 1 日 3 回程度といわれていることから, 平均 3 回程度が妥当と考えられるが実験での進行の加速度合いを見るため, 値 6, 9 についても確認した.

4.2 評価

4.2.1 評価方法

本稿の実験では, 4.1.3 項で述べたとおり各パラメータを設定した. このパラメータ値を変化させノードのグループ化がどのように変化しているかを検証している. グループ化の性能を評価するために, 以下の評価指標を設け各実験において評価値を算出することによって評価の目安とした.

4.2.2 評価指標

4.2.2.1 グループ化しているコンテンツ数

コンテンツを保持しているノードの中で初期設定において関連性のあるコンテンツの仮グループに設定したとおりのグループ化を行っているノード数. コンテンツを保持しているノード数は 1,000 ノードである.

4.2.2.2 各グループのコンテンツ平均数

グループ化されているノードが保持しているグループコンテンツの平均数. グループリスト最大数は 10 であるので, 最大保持数は 10 となる.

4.2.2.3 精度

グループ化の精度を式 (2) に示す. グループ化されているノードの中で, 初期設定の際に設定したコンテンツの仮グループどおりにグループ化が行われているノードの割合である. 分子は, 正常にグループ化されているノードのグループリスト最大数に対してグループリスト内コンテンツ数の割合の和である.

S : 精度, k : 仮グループ設定どおりにグループ化されたコンテンツ数, n : 仮設定グループ

表 4 実験パラメータ

Table 4 Parameters set for the experiment.

ノード数	減少率	基準値	傾向度	期待値	推薦
10000	5%	3	80%	3	ナシ
10000	5%	3	50%	3	ナシ
10000	5%	3	20%	3	ナシ
10000	5%	3	80%	3	アリ
10000	5%	3	50%	3	アリ
10000	5%	3	20%	3	アリ

数, m : グループ化されたグループ数, l : グループリスト最大数

$$s = \frac{\sum_1^n \binom{k}{l}}{m} \tag{2}$$

4.2.2.4 再現率

グループ化されているコンテンツのグループ再現率を式 (3) に示す. コンテンツを保持しているノードのうち初期設定において関連性のあるコンテンツの仮グループに設定したとおりのグループ化が行われているノードの割合である. 分子は精度の場合と同様である. r : 再現率, k : 仮グループ設定どおりにグループ化されたコンテンツ数, n : 仮設定グループ数, l : グループリスト最大数

$$r = \frac{\sum_1^n \binom{k}{l}}{n} \tag{3}$$

4.3 実験結果と考察

4.3.1 実験結果

本実験では, 各パラメータを変更し実験を行った. ここでは, その中でも大きな特徴があったと考えられる表 4 に示す実験パラメータでの評価結果と考察について述べる.

(1) 検索傾向度 80%の場合

図 4 は検索傾向度が 80%の場合の精度, 再現率を示した実験結果である. この実験では初期設定であらかじめ設定していた仮のグループ内の検索傾向度が 80%となっていることが

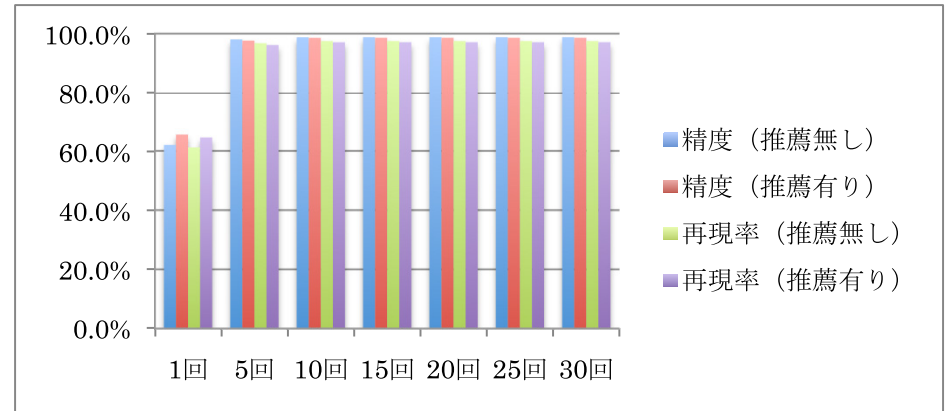


図 4 実験結果 1
Fig.4 Experimental result 1.

ら, 各ユーザが 1 サイクルの検索において似た性質のコンテンツを頻繁に検索していることになる. そのため, 推薦機能がないケースにおいても非常に高い精度, 再現率となっている.

(2) 検索傾向度 50%の場合

図 5 は検索傾向度 50%の場合の精度, 再現率を示した実験結果である. この実験は仮のグループ内の検索傾向度が 50%となっている. 50%の検索傾向度において推薦機能がない場合は検索傾向度 80%に比べて精度, 再現率ともに低下しているのに対し, 推薦機能がついている場合は, 検索傾向度 80%と比較しても精度, 再現率ともに大きく低下していないのが特徴である. この実験の結果から, ユーザが 1 回の検索サイクルにおいてリピートの割合が半分程度の検索をしていたとしても, その中から必要なコンテンツを選びグループ化を行うことができることが分かる.

(3) 検索傾向度 20%の場合

図 6 は検索傾向度 20%の場合の精度, 再現率を示した実験結果である. この実験では各ユーザの検索がきわめてランダムの場合となるため, 推薦を行う際に誤った推薦を行ってしまう. そのため, 効率的なグループ化を行うことができず, このように精度, 再現率ともに低い結果になったと考えられる.

4.3.2 考察

今回の評価は, ノード数を固定し本稿で提案した P2P ネットワークにおけるコンテンツ

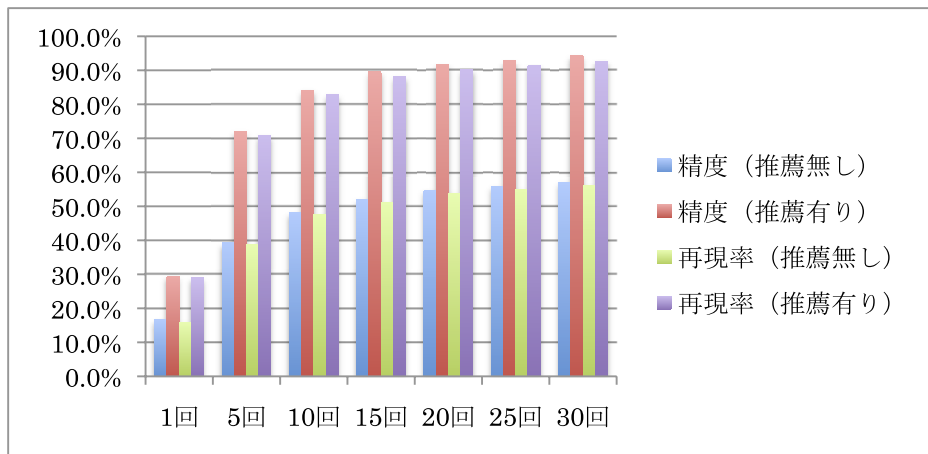


図 5 実験結果 2
Fig. 5 Experimental result 2.

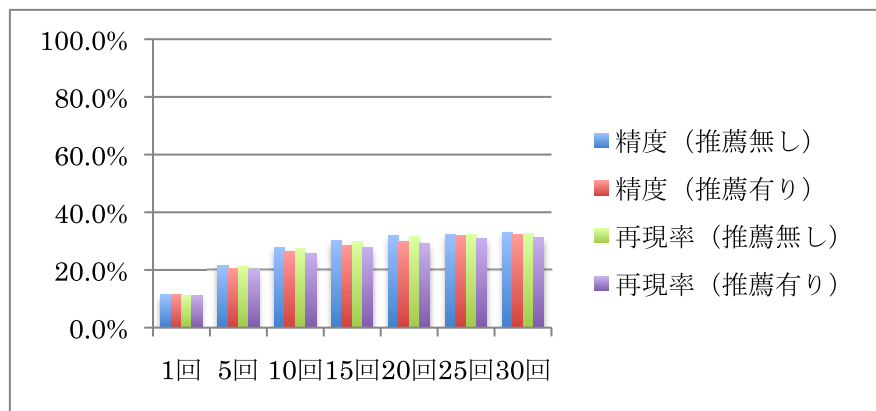


図 6 実験結果 3
Fig. 6 Experimental result 3.

のグループ化の基本的な機能が有効であるかを検証した。検索傾向度 50%のケースでも、簡単な推薦機能を付加することによりグループ化が有効に働くことが確認できた。今回示した

のは、関連性強度減少率が 5%、関連性強度基準値 3、の場合である。特に、減少率 5%は最も緩やかな設定値であるが、15 日では関連性強度が半減するペースであり今回のパラメータ設定値として妥当であったといえる。本グループ化機能は、コンテンツが利用者からコンテンツ要求を受けた際に動的にコンテンツのグループ化を利用者の検索状況に応じて行っている。そのため P2P ネットワークを対象とした動的なノードの変化（参加や離脱）に対応した機能であるといえる。また検索要求を行った際にそのコンテンツが属しているコンテンツのグループを提供することによって、結果的には検索数を減少させ効率的にコンテンツの検索・取得が行えるようになったといえる。

5. おわりに

本稿では、P2P ネットワークにおけるコンテンツのグループ化を提案しシミュレーションによる評価を行った。今後はコンテンツのグループ化を実装した検索システムの検討、実装を行うとともに、実際のデータを利用しグループ化が行われているのかを検証していく予定である。システムの実装にあたって、モバイルエージェントを利用した仮想化データベースの管理、推薦機能を実装したモバイルエージェント開発を順次行っていく予定である。これらの技術を利用し、P2P ネットワークを利用した検索システムを実装していく。

謝辞 本研究は科研費（20500095）、若手研究（B）（21700084）の助成を受けたものである。

参考文献

- 1) Napster. <http://www.napster.jp/>
- 2) Gnutella. <http://www.gnutella.com>
- 3) Aspnes, J. and Shah, G.: Skip Graphs, *14th Annual ACM-SIAM Symposium on Discrete Algorithms*, pp.385–393 (2003).
- 4) Stoica, I., Morris, R., Karger, D., Kaashoek, M.F. and Balakrishnan, H.: Chord: A scalable peer-to-peer lookup service for internet applications, *Proc. ACM SIGCOMM '01 Conference*, San Diego California (Aug. 2001).
- 5) Maymounkov, P. and Mazières, D.: Kademlia: A Peer-to-peer Information System Based on the XOR Metric, *Proc. 1st International Workshop on Peer-to-Peer Systems (IPTPS)* (Mar. 2002).
- 6) Ramabhadran, S., Ratnasamy, S., Hellerstein, J.M. and Shenker, S.: Pre-fix Hash Tree: An indexing data structure over distributed hash tables, Technical Report, Intel Research (2004).

- 7) Wan, Y., Asaka, T. and Takahashi, T.: A Hybrid P2P Overlay Network for Non-strictly Hierarchically Categorized Content, *IEICE Trans. Communications*, doi:10.1093/ietcom/e91-b.11.3608, Vol.E91-B, No.11, pp.3608-3616 (2008).
- 8) Tajima, K., Mizuuchi, Y., Kitagawa, M. and Tanaka, K.: Cut as a Querying Unit for WWW, Netnews, and E-mail, *Proc. 9th ACM Conference on Hypertext and Hypermedia*, pp.235-244 (June 1998).
- 9) Ayan, N.F., Li, W.-S. and Kolak, O.: Automating Extraction of Logical Domains in a Web Site, *International Journal of Data and Knowledge Engineering*, Vol.43, No.2, pp.179-205, Elsevier Science (Nov. 2002).
- 10) Mizuuchi, Y. and Tajima, K.: Finding Context Paths for Web Pages, *Proc. 10th ACM Conference on Hypertext and Hypermedia*, pp.13-22 (Feb. 1999).
- 11) 金 英子, 松尾 豊, 石塚 満: Web 上の情報を用いた弱い社会的関係のネットワーク抽出手法, *IEICE Trans. Information and Systems (Japanese edition)*, Vol.J91-D, No.3, pp.709-722 (2008).
- 12) 村田剛志: 参照の共起性に基づく Web コミュニティの発見, *Trans. Japanese Society for Artificial Intelligence: AI Trans. Japanese Society for Artificial Intelligence*, Vol.16, pp.316-323 (2001).
- 13) Foner, L.N.: Political Artifacts and Personal Privacy: The Yenta Multi-Agent Distributed Matchmaking System (1999).
- 14) OverlayWeaver. <http://overlayweaver.sourceforge.net/index-j.html>

(平成 22 年 5 月 15 日受付)

(平成 22 年 11 月 5 日採録)



佐々木拓也

2010 年岩手県立大学ソフトウェア情報学部ソフトウェア情報学科卒業。同年より同大学院ソフトウェア情報学研究科博士前期課程。現在に至る。P2P ネットワークにおける、情報検索に興味を持つ。



澤本 潤 (正会員)

1975 年京都大学大学院工学研究科修士課程修了。同年株式会社三菱電機入社。2006 年より岩手県立大学ソフトウェア情報学部教授。博士 (工学)。基盤ソフトウェア, ソフトウェアアーキテクチャ, 組み込みソフトウェア, ユビキタスコンピューティングの研究教育に従事。IEEE, ACM, 電子情報通信学会, 電気学会各会員ほか。



加藤 貴司 (正会員)

1971 年生。2001 年東北大学大学院情報科学研究科博士後期課程修了。現在, 岩手県立大学ソフトウェア情報学研究科講師。博士 (情報科学)。マルチエージェントシステムにおけるエージェントの協調に関する研究に従事。人工知能学会, 電子情報通信学会各会員。



和田 雄次 (正会員)

東京電機大学情報環境学部情報環境学科教授。1976 年早稲田大学大学院理工学研究科修士課程修了。同年三菱電機株式会社入社。2002 年までデータベースシステム, OA システム, DSS, ユーザインタフェース, CAI, 映像画像処理システム, 危機管理システム等の研究開発に従事。1997 年静岡大学大学院電子科学研究科社会人博士課程修了。2003 年より現職。データベースシステム, データマイニング, E ラーニングの研究に従事。博士 (工学)。ACM, IEEE-CS, 電子情報通信学会, 日本データベース学会, 日本ソフトウェア科学会, 人工知能学会各会員。