

平成22年度 修士研究論文

P2P ファイル配信環境における学習型ルーティング
方式によるコンテンツ検索手法

指導教官

松尾 啓志 教授

津邑 公暁 准教授

名古屋工業大学大学院
創成シミュレーション工学専攻
平成21年度入学 21413509番

伊藤 雄一郎

目次

第1章	はじめに	1
第2章	Peer-to-Peer モデル	3
2.1	クライアントサーバモデル	3
2.2	P2P ネットワークモデル	5
2.2.1	Pure P2P	6
2.2.2	Hybrid P2P	8
第3章	関連研究	10
3.1	構造型 P2P システムを利用したコンテンツ配信	10
3.1.1	Chord	11
3.2	非構造型 P2P システムにおける検索	11
3.2.1	検索手法	11
3.3	セマンティック P2P ネットワーク	13
3.3.1	RIs	14
3.3.2	SONs	15
3.3.3	Tellagate	17
3.3.4	コンテンツ類似度に基づいた P2P ネットワークの自己組織化	17
第4章	従来手法	20
4.1	従来手法の概要	20
4.2	従来手法における前提	21

4.3	従来手法における処理の手順	21
4.4	従来手法による問題点	27
第5章	提案手法	29
5.1	提案手法の概要	29
5.2	想定される環境	30
5.3	提案手法の詳細	31
5.3.1	ピアにおける情報の取得	31
5.3.2	ピアの選別	31
5.3.3	隣接ピアの組み換え	32
5.3.4	検索時の動作	32
5.3.5	検索メッセージ転送テーブルの更新	35
第6章	評価	39
6.1	本研究における比較手法	39
6.2	シミュレーションのモデル	39
6.3	該当ジャンル保持ピア数・ヒット率	41
6.4	メッセージ数	45
6.5	全ピア数を変化に伴う該当ジャンル保持ピア数の評価	49
6.6	提案手法における改善点	50
第7章	まとめ	51
	謝辞	54

第1章

はじめに

P2P(peer-to-peer) 技術に関する様々な研究が行われている。P2P 技術は、近年では Skype などの電話サービス、ストリーミング配信といったコンテンツ配信サービスに応用されている。その一方で、ファイル共有サービスに用いられることが多く、数多くのファイル共有アプリケーションが提案されてきた。P2P ファイル共有サービスとして、Gnutella[1]、Freenet[2]、BitTorrent[3]、Napster[4]、WINMX[5]、Winny[6] といった様々なアプリケーションが提案されている。一般に、P2P ファイル共有サービスは、ユーザは自身が欲しい音楽ファイルなどのコンテンツを検索し、そのコンテンツを受け取るという仕組みであり、ユーザはサービスを利用する度に、希望するコンテンツを検索する。しかし、非構造型の P2P ネットワークでは、そのネットワークに参加するピアの情報などを一元管理するサーバは存在しない。このため、コンテンツの検索において、フラッディングのようなバケツリレー方式で検索メッセージを転送する方法を用いることが一般的である。非構造型 P2P ファイル共有サービスの一例である Gnutella では、オーバーレイネットワーク上においてピアがランダムに配置されるため、このような P2P ネットワークでは、各ピアのユーザが検索したコンテンツを、制限された検索範囲内で発見できる確率は比較的少ない。ユーザが検索したコンテンツを取得する頻度を改善するために、ファイル共有サービスでは検索効率の向上は重要な課題である。このような目的のために、ユーザの嗜好性を考慮して P2P ネットワークを構築し、その中でコンテンツを検索し、検索効率を向上を図るセマンティック P2P

ネットワークの研究が行われている。本研究で提案する手法では、複数の方針に基づいてピアとの隣接関係を決定することにより、オーバーレイネットワークを構築する。また、強化学習を用いて、検索効率の向上が期待されるピアへの経路を学習する。これらにより、希望するジャンルを保持するピアに検索メッセージが転送される機会が増え、検索効率が改善する。提案手法とセマンティック P2P ネットワークの従来手法を検索効率、メッセージ数などの指標を比較し提案手法の有用性を示す。本論文の以降の構成を示す。第 2 章では、コンテンツ配信システムとして用いられる P2P ネットワークについて説明し、第 3 章では、その検索効率の向上を目的とする関連研究について説明する。第 4 章では、それらの従来手法の一つである、ユーザの嗜好性を考慮する検索効率を向上させる手法について説明し、第 5 章では提案手法について説明していく。第 6 章で、従来手法と提案手法を評価し、提案手法の有用性を示す。第 7 章では結言を述べる。

第2章

Peer-to-Peerモデル

本章では，クライアントサーバモデル，P2P ネットワークに基づくコンテンツ配信システムに関する研究の背景を説明する。

2.1 クライアントサーバモデル

現在，インターネットにおけるサービスで主に用いられているモデルは，クライアントサーバモデルである。クライアントサーバモデルについて図2.1に示す。クライアントサーバモデルでは，Web ページ等のコンテンツサービスを提供するサーバと，コンテンツサービスをリクエストするクライアントから構成されている。サーバはクライアントからのリクエストに備え，常に待機している。クライアントからサーバに対して，コンテンツのリクエストがあった場合，サーバはクライアントに対してコンテンツを提供する。このように，クライアントサーバモデルでは，クライアントとサーバの機能は明確に区別されている。また，クライアントの数は，サーバの数に比べ圧倒的に多く，そのため，サーバは多くのクライアントが要求するコンテンツリクエストに応じなければならない。この結果，サーバの負荷は大きくなってしまふ。さらに，近年では，ADSL (Asymmetric Digital Subscriber Line)，FTTH(Fiber To The Home) といった，常時接続，高速回線が急速に普及したことや，インターネットの利用者が増加しているため，クライアントからサーバへのリクエストが飛躍的に増加することが予想され，よりサーバの負荷が大きくなる，サーバ周辺のトラヒックも増大するといっ

た状況になりやすい。このため配信遅延等、各クライアントに提供できるサービス品質が低下してしまう恐れがある。また、単一故障点（SPF：Single Point of Failure）の問題もある。停電などのアクシデントにより、サーバに障害が発生した場合、該当サーバにリクエストを行うすべてのクライアントはサービスを受け取ることができなくなってしまう。

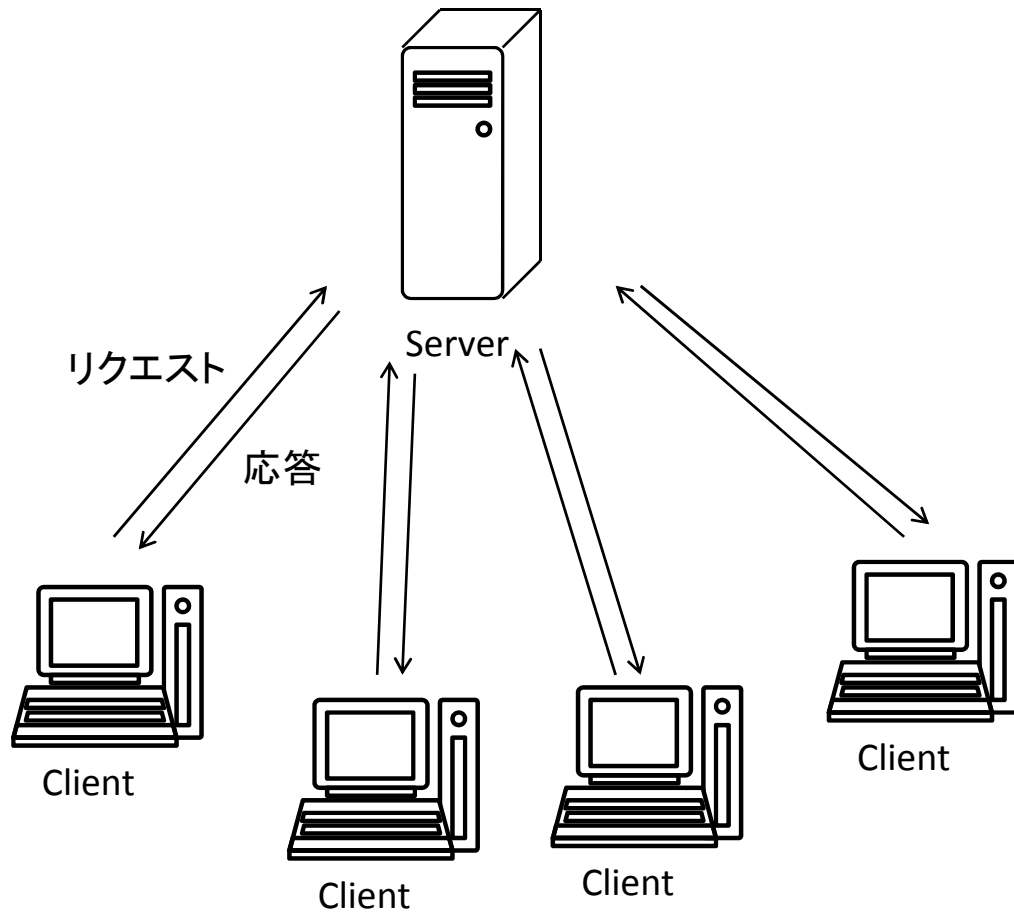


図 2.1: クライアントサーバモデル

2.2 P2P ネットワークモデル

P2P(Peer-to-Peer) ネットワークは、クライアントサーバモデルの問題点を改善するネットワークとして利用される。P2P ネットワークでは、システム全体の中心となるようなノードを基本的に持たず、各ノードはすべて互いに対等な立場とされる。すべてのノードは、コンテンツサービスをリクエストするクライアントにもなり、リクエストに応えるサーバとしても機能する。したがって、どのピアもサーバに成りうるために、クライアントサーバモデルにおける、サーバへの負荷集中、トラフィック集中という問題や、単一故障点の問題の改善が期待できる。P2P ネットワークの特徴として以下の点が挙げられる。

耐障害性

クライアントサーバモデルでは、サーバに障害が発生したり、その周辺のネットワークに障害が発生した場合、そのサーバからのサービスは提供できなくなってしまい、各クライアントはサービスを受けることができなくなってしまう。この一方、P2P ネットワークの場合、各ノードはサーバにも成りうるといった点や、特定のノードに障害が発生した場合でも、別のルート、別のノードを通してコンテンツを提供できるといった点があるため、単一故障点という問題は起きにくくなり、比較的安定して、コンテンツサービスを提供できる環境であるといえる。

スケーラビリティ

P2P ネットワークでは、先に述べたとおり、各ノードはいつでもサーバに成りうる。したがって、どのノードもコンテンツサービスを提供できるモデルである。このため、クライアント数が増加する、クライアントからのリクエスト数が増加した場合でも、複数のノードがリクエストに回答できるようなモデルであるため、コンテンツ配信負荷や、トラフィックをネットワーク全体に分散することが期待できる。したがって、クラ

クライアント数が増加した場合でも，サービス性能は変わらず，クライアント数，サービスリクエスト数の増加に対するスケーラビリティが高いといえる．

アドホック (ad-hoc) 性

P2P ネットワークでは，サービスコンテンツの要求，それに対する応答というプロセスにおいて，サーバなど特別なノードを介する必要がない．したがって，サーバによる一元的な管理が必要ないため，管理コストを考えなくてよい．

P2P ネットワークでは，先にも述べたとおりデータはネットワーク内に分散して配置され，また，各ノードの持つデータを一括して管理しているノードは存在しない．コンテンツ配信サービスにおいては，クライアントは希望するコンテンツを検索する際には，速やかなコンテンツ発見が求められるが，フラッディングなど場当たりの検索方法，バケツリレー方式の検索方法を用いる場合，コンテンツ発見までに時間がかかる，検索メッセージによる多くのトラヒックが発生するという問題がある．したがって，希望するコンテンツを効率的に検索する仕組みが必要となる．このような，コンテンツ配信サービスを想定し，大きく分けて，PureP2P 型の P2P ネットワークモデル，Hybrid P2P 型の P2P ネットワークモデルの 2 つが挙げられている．

2.2.1 Pure P2P

PureP2P 型ネットワークの概念図を図 2.2 に示す．PureP2P 型ネットワークの大きな特徴は，ネットワークに参加しているノードの管理や，そのノードが保持しているコンテンツの所在情報などを管理する役割を果たすノードが存在せず，完全に全ピアが対等な立場であることである．ネットワークに参加しているすべてのノードはお互いに，サーバにもクライアントにもなり，相互にコンテンツのサービスのリクエスト，応答を行っている．仮に障害などにより，あるノードの機能が停止したとしても，システム全体としては，停止することなく稼働し続けることができるという利点を持って

いる。しかし、PureP2P 型ネットワークでは、どのノードが何のコンテンツを持っているかという情報は一括して管理されていないため、コンテンツの検索時には各ノードは、どのノードにコンテンツをリクエストすれば良いのかが分からないというのも特徴である。ここで、PureP2P 型ネットワークにおける、コンテンツ検索方法について説明する。一般的に PureP2P 型ネットワークでは、コンテンツの検索には、フラッディングという手法が用いられる。コンテンツのリクエストを行うノードは、隣接するすべてのノードに対して、検索メッセージを送信する。検索メッセージを受け取ったノードは、該当するコンテンツを持っているかどうかを確認し、持っている場合には、検索メッセージを送信した送信元ノードへコンテンツ発見メッセージを返信しその後、ノード間で直接通信を行い、コンテンツを提供する。持っていない場合には、さらなる隣接ノードに検索メッセージを転送していく。この際、際限なくメッセージが転送されつづけるのを防ぐために、検索メッセージには TTL 値 (Time To Live) が設定されていることが一般的である。検索メッセージが一度転送される度、TTL 値を 1 ずつ減らしていき、TTL 値が 0 になった場合、転送を止める。PureP2P 型ネットワークでは、先に述べたとおり、どのノードが何のコンテンツを保持しているかという情報は一元管理されておらず、通常、周辺のノードの保持コンテンツの情報については、情報がないか、もしくは局所的な情報であるケースが多い。そのため、検索時には、上に挙げたフラッディングのような検索方法に頼らざるを得ない。フラッディングでは、バケツリレー式の手当たり次第な検索方法であるため、希望するコンテンツがネットワーク上にはあるが、発見できないという問題がある。このような場合、検索メッセージは転送され続けることによって、トラフィックが増加してしまう。トラフィックの増加を防ぐために TTL 値が設けられているが、手当たり次第に検索メッセージを転送することによる、トラフィックの増加に関しては完全に防ぐことができない。このため、検索効率性の向上といった課題が挙げられる。

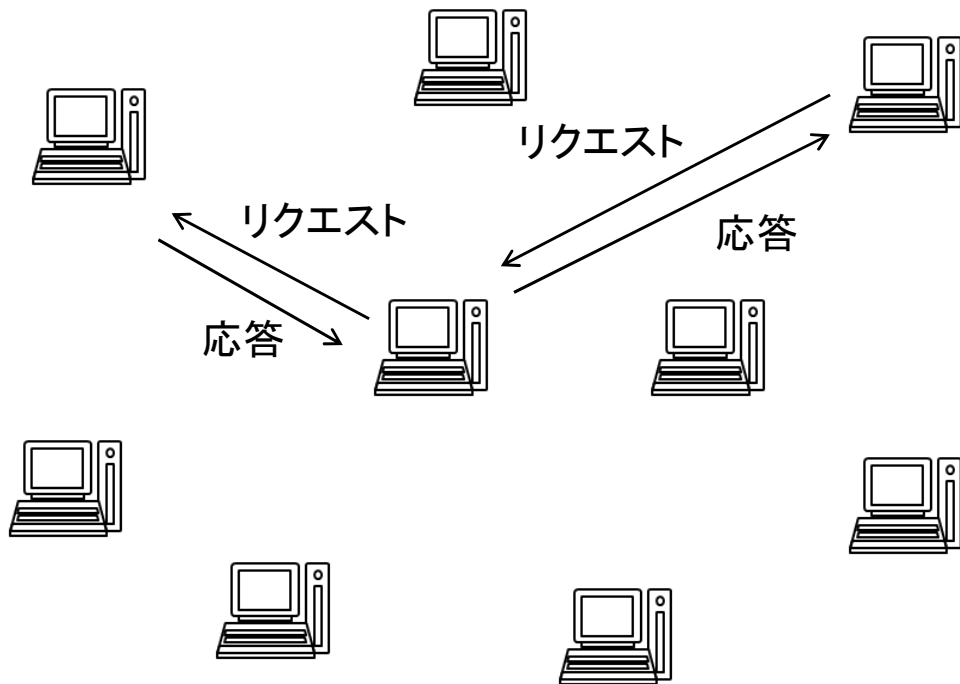


図 2.2: PureP2P の例

2.2.2 Hybrid P2P

Hybrid P2P 型ネットワークでは、先に述べた、Pure P2P 型の抱える、ノードが一元管理されていないことにより発生する問題を対処できる手法である。Hybrid P2P 型ネットワークでは、各ノードの持つコンテンツのインデックス情報を管理するサーバノードが存在する。データを保持する各ノードは特定のサーバノードに情報を登録している。このネットワークモデルにおける、コンテンツ検索時の動作を説明する。まず、希望するコンテンツをリクエストする検索元ノードは、サーバノードに対して情報の検索を依頼する。サーバノードは、登録された情報を利用し、リクエストされたコンテンツを保持しているノードを調べ、保持しているノード情報を検索元ノードに伝える。その後、検索元ノードは、得られた情報を元に、ノード間で直接通信を行い、希望するコンテンツを得る。このように、Pure P2P 型ネットワークモデル、クライアントサーバモデルの両者の特徴を併せ持つ。

Hybrid P2P はコンテンツの取得時には P2P モデルの特徴を利用し、インデックス情

報を取得する際に、クライアントサーバの特徴を利用する。このため、クライアントサーバモデルと比較して、サーバの負荷を削減することができる。しかし、インデクス情報を取得する際には、クライアントサーバモデルであるため、単一故障性の問題を Hybrid P2P 型ネットワークモデルは抱えている。

第3章

関連研究

P2P ネットワーク上におけるコンテンツ配信では，ユーザが望むコンテンツを効率よく検索できることが必須の課題である．また，同時に，検索時にかかるコストをいかに少なくするかということも考える必要がある．つまり，検索効率を向上しつつも，検索自体にかかるコスト（検索時間，検索メッセージ等）を低く抑えることが理想的である．P2P ネットワークにおいて，検索効率の向上，検索コストの削減についての研究は数多く行われている．本章では，関連研究として検索効率の向上手法，検索コストの削減についての研究について説明する．

3.1 構造型 P2P システムを利用したコンテンツ配信

構造型 P2P システムを利用したコンテンツ配信モデルでは，分散ハッシュテーブル（DHT:Distributed Hash Table）を用いて，ネットワークトポロジを決定したり，データの検索に必要なキーを利用し検索を行っている．この構造型 P2P システムを利用したモデルの例としては，Chord，CAN(Content-Addressable Network)，Pastry などが例としてあげられる．一般的に，構造型 P2P システムを利用したモデルでは，ピアの識別子にハッシュ関数を適用してハッシュ値を算出し，これをハッシュ空間の 1 つに対応づける．また，各ノードが保持する各データのキー情報も同様にデータ名などからハッシュ値を算出し，ハッシュ空間内で，一番近いピアにそのキー情報を格納する．

データの検索時には、希望するデータのハッシュ値を算出し、ハッシュテーブルを参照し、最も近いノードに検索メッセージを送信することにより、希望するデータについての発見率を向上している。このように、構造化することにより、検索効率を向上している。ここでは、構造型 P2P システムの代表例である Chord について説明する。

3.1.1 Chord

Chord[8] では、円上の形をした一次元の分散ハッシュテーブルになっている。各ノードは、Fingar Table という経路表を保持しており、データの検索には、この Fingar Table が用いられる。各ノードは、ハッシュ空間における位置が円上に並ぶハッシュ空間の大きさの $1/2$ 、 $1/4$ 、 $1/8$ といったように、2 の累乗分の 1 だけ離れているピア、およびハッシュ空間内の位置が自身より大きいものの中で最も近いノードの IP アドレス等の識別子を、fingar table に保持している。検索時には、fingar table に情報のあるノードのうち、最も近いノードに検索メッセージを送信する。検索メッセージを受け取ったノードは、さらに自身の fingar table を元に次のノードに検索メッセージを転送するというように、目的のノードに辿り着くまで転送が繰り返される。各ノードが、ハッシュ空間の $1/2$ 先、 $1/4$ 先、 $1/8$ 先のノード情報を管理していることから、検索負荷を $O(\log N)$ に抑えることができる。

3.2 非構造型 P2P システムにおける検索

非構造型 P2P システムでは、構造型 P2P システムで用いられている、分散ハッシュテーブルのような、構造化された仕組みを持たない。本節では、非構造型 P2P システムにおける、検索効率の向上についての研究について説明する。

3.2.1 検索手法

非構造型 P2P システムでは、コンテンツの検索にはフラッディングという手法が使われていることを前章で説明した。フラッディングを用いた検索では、集中管理が必

要なく、スケーラビリティに優れたネットワークを構築することができる一方で、検索メッセージの増大によるトラヒックの増大や、コンテンツを発見するまでに時間がかかるといった点が問題となる。また、フラッディングではTTLの値を大きく設定すると、検索メッセージによるトラヒックが指数的に増大してしまうという点や、同じピアにクエリが重複して届く確率が高くなり、無駄なトラヒックが発生してしまう点が問題となる。このようなフラッディング特有の問題に関して、次の手法が提案されている。

エキスパンディングリング

初期値は小さいTTL値を設定しフラッディングによる検索を行い、希望するデータが見つからなかった場合、少しずつTTL値を増加し、再びフラッディングを行う。エキスパンディングでは、過剰にTTL値が高くなるのを抑え、検索メッセージによるトラヒックを抑える。

Random Walk

Random walkの概念図を図3.1に示す。検索メッセージを受け取ったノードは、自身のノードが希望するデータを持っているかどうかを調べ、持っていなかった場合には、隣接ピアを1つ選択しそのノードに検索メッセージを転送する。これにより、フラッディングでは検索メッセージが転送されるごとに、指数的にトラヒックが増大していたが、Random Walkでは、検索メッセージ数を抑えることができる。その一方で、希望するデータが発見されるまでが何ホップもノードを経由するために、発見までに時間がかかるという問題点もある。

k-Walker Random Walk

k-Walker Random Walkでは、フラッディングのように全ての隣接ノードに検索メッセージを転送するのではなく、隣接ノードのうち、転送するノードをk個に限定する。

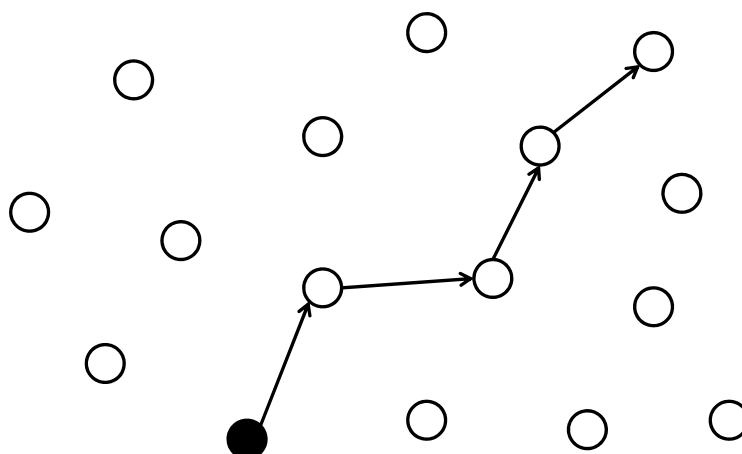


図 3.1: Random Walk

検索メッセージを受け取ったノードは、自身のノードが希望するコンテンツを保持しているかを調べ、保持していなかった場合、隣接するノードよりランダムに k 個選択し、転送する。この手法により、フラッディングに比べれば検索メッセージの増大を抑えることができるが、以前として、手当たり次第の検索であるため検索メッセージが多いという問題は残る。Random Walk は、 $k = 1$ の時の手法に相当する。

3.3 セマンティック P2P ネットワーク

ユーザがコンテンツを検索する際、何のコンテンツを検索するかは、ユーザの嗜好性に基づいて決定されることが多く、またその嗜好性は変化しにくい。「音楽が好き」という嗜好を持ったユーザは、積極的に音楽に関するコンテンツを探ることが多いであろう。このような背景から、検索効率の向上を目的として、ユーザの嗜好性に基づいて検索メッセージ送信先を決定したり、トポロジを組み替えるといった手法により、データ検索効率の向上を目的とした研究も多くなされている [12],[13]。本研究では、ユーザの嗜好性に着目したセマンティック P2P ネットワークを研究対象としている。本節ではユーザの嗜好性を利用した、セマンティック P2P ネットワークの関連研究について説明する。

3.3.1 RIs

データの検索時に、希望するデータの発見が最も期待できるノードに検索メッセージを送る RIs(Routing Indices)[9] という研究がある。RIs の概念図を図 3.2 に示す。RIs では、各ノードは隣接するノードのもつデータ情報をトピック（例えばテキストデータの場合、Language、Theory などのトピックに分類できる）ごとに分類し、各トピックに属する情報数を Routing Indices として管理している。各ノードは、検索メッセージを発信する場合、またノードが検索メッセージを受け取り転送する場合、Routing Indices を参照する。Routing Indices をもとに登録されている各ノードから得られる応答数の期待値を計算し、その期待値の最も大きいと思われるピアへクエリを送信する。実際に、RIs の例を図 3.2 に示す。ピア A が、Language のトピックについてのデータを検索したい場合、ピア A は自身の Routing Indices を参照し、最も Language のトピックについての応答数の期待値が高い、ピア D に検索メッセージが送信される。ピア A からのメッセージを受け取ったピア D は自身の Routing Indices を参照し同様に、ピア D はピア E に送信する。

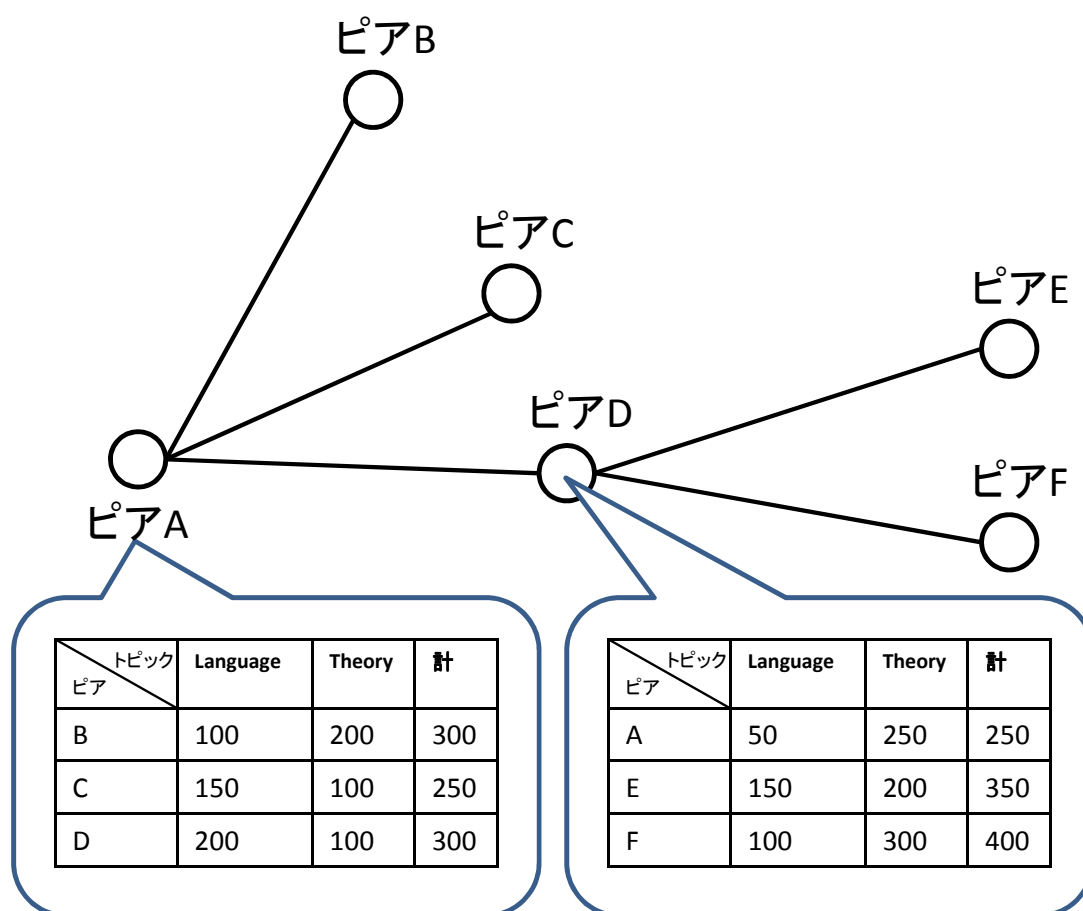


図 3.2: RIs

3.3.2 SONs

データの検索時に、各ノードが持つコンテンツのジャンルを考慮し検索することにより、検索効率を向上させる SONs (Semantic Overlay Networks) [10] という研究がある。SONs の概念図について図 3.3 に示す。SONs では、データ情報のファイル名とその情報に対応するジャンルを記したデータベースが存在する。各ノードではデータベースを参照することにより、データ情報に対応するジャンルを一意に決定する。また、ジャンルは root, style, substyle という順の 3 層に、階層化された SONs を形成する。

音楽ファイルを扱うケースを例に挙げると、root という階層は音楽ファイル全体に相

当する．また，その一階層下の style には，ROCK，CLASSIC，JAZZ などのジャンルが位置する．さらにその一階層下である substyle には，Visual ROCK，PUNK ROCK といったさらに細分化されたジャンルが位置する．ノードが参加する際，あるジャンルに属するファイル数が一定の閾値以上の数を保有している場合に，そのジャンルに相当する SON に参加する．

また，ノードがファイルを検索する時には，検索するファイルに相当すると思われるジャンルを決定し検索する．そして，下の階層の SON から順に検索をしていく．例えば，PUNK ROCK に属するファイルを検索する場合には，まず，PUNK ROCK の substyle の SON 内に転送され，その SON 内でフラッディングされる．仮にその SON 内に見つからなかった場合には，その一階層上の ROCK という style の SON にクエリが転送される．さらに，SON 内で見つからなかった場合には，root の SON にクエリが転送される．このように，ファイルのジャンルを考慮し検索することにより，検索効率の向上を図っている．

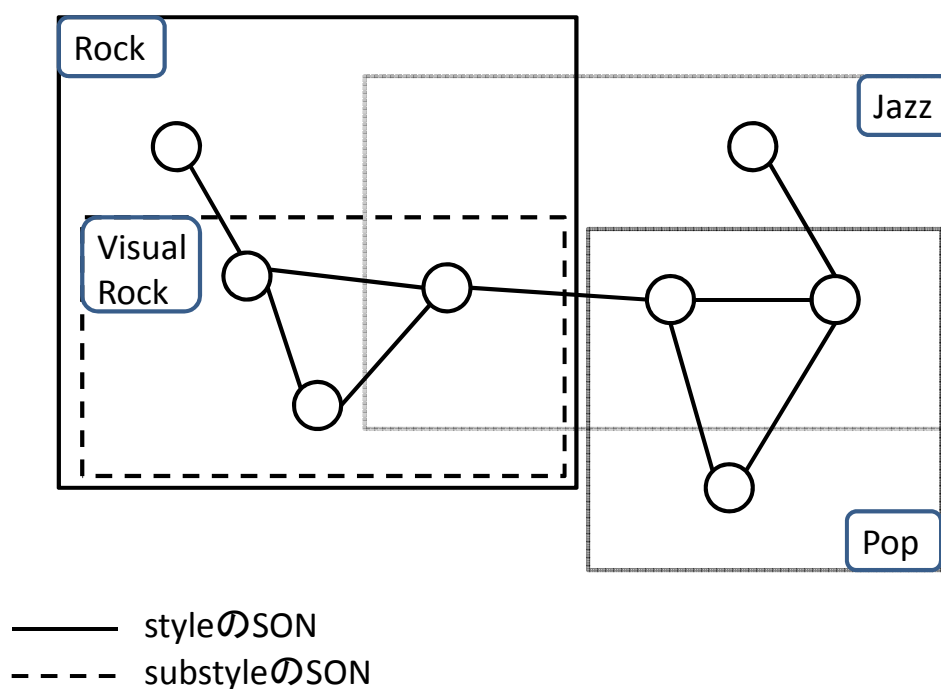


図 3.3: SONs

3.3.3 Tellagate

Tellagate[11] は、保持するファイルや検索するファイルの嗜好が近いノード同士をオーバーレイネットワーク上で近くに配置する。このような、P2P ネットワークの自己組織化を行うことにより、検索効率を向上している。Tellagate が扱うファイルはテキストデータのみを対象としている。

Tellagate では、ノードの持つデータ情報は digest にまとめられ管理されている。また、各ノードは自身から 1 ホップ目にあたるノードを 1st neighbor、さらに 1 ホップ先の 2 ホップ目にあたるノードを 2st neighbor と定義している。データの検索には、QRP with Firework というプロトコルを用いて検索が行われる。検索メッセージを送信、転送するノードは digest を参照し、検索するキーワードが含まれている 1st neighbor にあたるノードに検索メッセージを転送する。自己組織化の際には、定期的に 2nd neighbor の全てのノードの digest に対して嗜好の近さを計算し、嗜好の最も近いと思われるノードとリンク接続を行う (CSOA:Community Self-Organization Algorithm)。

以上の操作により、嗜好の近いノード同士が、オーバーレイネットワーク上で近くに配置されるように自己組織化が進み、検索効率の向上を図っている。

3.3.4 コンテンツ類似度に基づいた P2P ネットワークの自己組織化

Pure P2P 型ネットワークにおける情報流通ネットワークにおいて、高い検索効率を実現するオーバーレイネットワークトポロジの構築法が、遠藤らにより提案されている [13]。通例、大規模な Pure P2P 型ネットワークでは、すべてのノードに検索メッセージを伝達するのは難しい。また、従来では、コンテンツを検索する際には無作為に検索メッセージを転送していくモデルだったために、検索メッセージによるトラヒックの増大が問題であった。そこで、遠藤らは、各ノードが保有するコンテンツの類似性に基づいてネットワークの論理的な接続トポロジを変更することにより、検索効率の向上を図った。

この手法では、各ノードの保有コンテンツの類似性測るために特徴ベクトルを利用

している。また、ノード間の接続に関しては、ノードの満足度を用いて接続を決定している。特徴ベクトルとは、ノードが保有するコンテンツから抽出された特徴ベクトルであり、いわば「ユーザの嗜好性を表す」ベクトルとも言い換えることもできる。特徴ベクトルは、単語の出現回数を並べたベクトルに次元圧縮を施す、Latent Semantic Indexing 等を利用して保有するコンテンツより、抽出することができる。つまり、特徴空間上での距離が近いノード同士は、特徴ベクトルも近く、ユーザ同士の嗜好が似通っているといえる。このため、このノードは似通ったコンテンツを保持していることを期待できる。また、満足度とは「ノードに接続していることの有効性」と言い換えることができる。つまり、満足度が高いノードほど検索効率が良いノードということになる、この満足度は、以下の指標により算出される。

- 特徴空間上で、近くのノードと接続しているほど、ノードの満足度は高い
- 特徴空間上で、様々な方向のノードと接続しているほど、ノードの満足度は高い

特徴空間上での近さはユーザ同士嗜好の近さを示しており、方向はどのような近さであるかという尺度となる。

それでは、実際にこの手法の処理の手順を説明する。

1. 広告パケットのフラッディング

各ノードは定期的に、自身の嗜好を表す特徴ベクトル、IP アドレスからなる広告パケットをフラッディングする。広告パケットにはTTL値が設定されており、広告パケットを受け取ったノードは、TTLが0になるまで他のノードへと転送していく。これにより、自身の特徴ベクトルを周辺に知らせる。

2. トポロジの変更

広告パケットを受け取ったノードは、現在の自身のノード満足度を上昇させるノー

ドであれば、接続許可メッセージを広告パケット送信元ノードに送信する。広告パケット送信元のノードは、接続許可メッセージを送信したノードのうち、自身のノード満足度が最大となるノードと接続する。

以上の手法により、様々な尺度から図った嗜好について、嗜好の近いノードを隣接ノードとすることにより検索効率の向上を行っている。

次の章では、本研究の対象としているセマンティック P2P ネットワークを用いた既存手法について説明する。

第4章

従来手法

前章では、検索効率の向上の手法としてコンテンツの意味情報を利用したセマンティック P2P ネットワークの研究について説明した。セマンティック P2P ネットワークでは、ノードの保持するコンテンツ情報から嗜好性の近さを算出することによって、検索メッセージの転送先を決定したり、トポロジを変更する自己組織化を行い検索効率の向上をしていた。本章では、本研究での従来手法 [14][15] について説明し、問題点を挙げる。

4.1 従来手法の概要

P2P ネットワークにおけるコンテンツ配信においてコンテンツの検索効率の向上は重要な課題である。従来手法では、各ノードが保有するジャンルからユーザの嗜好性を測ることができると考え、ノードが保有するコンテンツが類似しているノードと隣接関係を結ぶ。各ファイルは、ジャンルという種類に分られており、各ノードは様々なジャンルに属するコンテンツを保持している。そして各ノードは、各ジャンル属するコンテンツ保有数からユーザの嗜好性を測る。また、ノードの間の RTT(Round Trip Time) も考慮しつつ隣接関係を結ぶことによって、ファイルの検索効率の向上とダウンロード時間の短縮の両方を実現している。

4.2 従来手法における前提

従来手法において、P2P ネットワーク上に存在するコンテンツは正確にジャンル分けすることができる。ここでのジャンルとは、アニメ、音楽、映画といった例が挙げられる。また一方で、コンテンツは、アニメ作品、音楽作品、映画作品の1つ1つと言える。例えば、A というタイトルのアニメ作品と B というタイトルのアニメ作品はコンテンツであり、いずれも、アニメというジャンルに仕分けられる。

コンテンツのネットワーク登録時には、ジャンルの選択や、ユーザによるタグづけ等によって全てのコンテンツは一意にジャンル分けされているとする。また、コンテンツのダウンロード時間は単純化のため RTT に対して正の相関を持っており、RTT が大きくなればダウンロード時間も大きくなるとする。

4.3 従来手法における処理の手順

従来手法の手順について説明していく。従来手法では、まず各ノードが Gnutella 式のフラッディングによってピア情報を取得し、その情報をピアリストというテーブルに格納する。その後、ピアリストに格納されたピアを RTT とピアの持つコンテンツ情報から算出される嗜好の近さの 2 段階でランクづけをし選別を行う。この選別によって、順位づけられたピアのうち、上位 R 個のピアを検索メッセージの送信先、転送先として決定する。この手法により、検索効率の向上を図っている。それでは、従来手法における詳細の手順について説明する。

ピア情報取得

各ノードは、ピア情報取得用メッセージを用いて、周辺のピア情報を得る。ピア情報とは、アドレスやポート番号、ピアの保持コンテンツのジャンル比率情報のことである。この情報を得るために発信されるのが、ピア情報取得用メッセージである。また、ジャンル比率情報とは、ジャンルごとのピアの保有する全保有コンテンツに対する割合の情報のことである。ピア情報取得用メッセージがピアから発信されると、メッ

セージは Gnutella 方式のフラッディングにより送信・転送される。ピア情報取得用メッセージを受け取ったピアは、自身のピア情報を、ピア情報取得用メッセージのたどった経路を逆にたどり送信していく。この時、このメッセージを中継したピアは、そのメッセージに含まれている種々のピア情報のデータを受けとることができる。ピア情報を取得できたピアは、その情報をピアリストに格納されていく。また、ピアリストにはピアリスト容量という上限（値 P）があり、仮にこのピアリスト容量を超える場合には FIFO (First In First Out) によりリプレイスされる。

ピアの選別

各ピアは、ピア情報を取得したのちに、ピアリストにあるピアとの RTT を取得し、さらに、ピアリストにあるピアとの嗜好性の近さを算出する。嗜好性の近さはピアとピアとの間で保有コンテンツのジャンルの重複率の和によって得られる関数によって算出される。ピアとピアとの嗜好の近さは以下の関数により求められる。

以下は、ピア i とピア j の嗜好の近さ $S(i, j)$ を求める関数である。

$$S(i, j) = \sum_{m \in G} \min\{c_i(m), c_j(m)\}$$

ここで、 G はジャンルの集合を意味する。また、 $c_i(m)$ とは、ピア i が持つコンテンツについて、ジャンル m に属するコンテンツの全保有コンテンツに対する割合を表し、 $\min\{a, b\}$ は、整数 a, b のうち、小さな値をとる。

ここで、嗜好の近さを求める例を図 4.1 に示す。

図 4.1 では、ピア A とピア B の嗜好の近さを評価値 $S(\text{ピア A}, \text{ピア B})$ として計算している。ピア A とピア B と共通して保持しているジャンルは、音楽、学問、スポーツである。共通したジャンルについて、それぞれのジャンル比率を比較し、ジャンル比率の小さいほうの値を評価値として加算していく。この例では、音楽というジャンルについて、ピア A はジャンル比率 0.3、ピア B はジャンル比率 0.1 であるため、0.1 を評価値として加算する。また同様に、学問というジャンルについては 0.25 を、スポー

ジャンル	ジャンル比率
音楽	0.3
映画	0.35
学問	0.25
スポーツ	0.1
アニメ	/

ジャンル	ジャンル比率
音楽	0.1
映画	/
学問	0.4
スポーツ	0.2
アニメ	0.3

$$S(\text{ピアA}, \text{ピアB}) = 0.1 + 0.25 + 0.1$$

図 4.1: 嗜好の近さを求める例

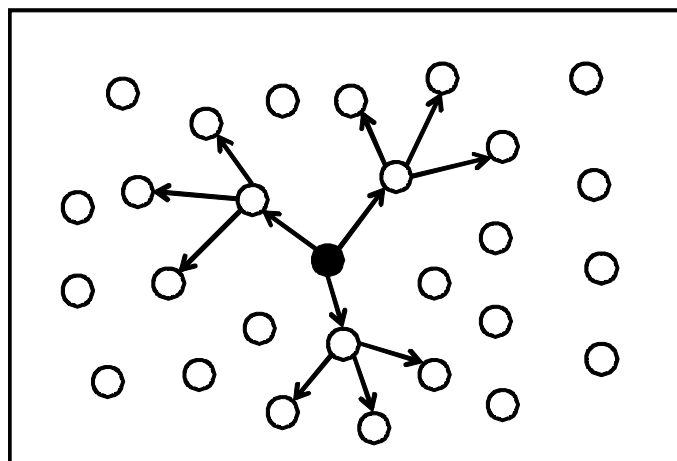
ツというジャンルについては、0.1 を評価値に加算する。このようにして、ピア A とピア B 間の評価値は $0.1 + 0.25 + 0.1 = 0.45$ となる。以上の操作によって、各ピアは、自身意外のピアの RTT と自身との嗜好の近さを求める。

その後、ピアリストに格納されているピアを 2 段階でランクづけをし、選別をする。仮にピアリスト容量を P 個であるとする。まず選別第 1 段階では、RTT の大きいピア Q 個 ($Q < P$) を転送先候補から除外する。次の第 2 段階では、第 1 段階で除外され残されたピアのうち、上記式から得られた、嗜好の近い上位 R ($R < Q$) 個のピアを新たなコンテンツ検索メッセージ発信、転送先と決定する。

隣接ピア組み換え

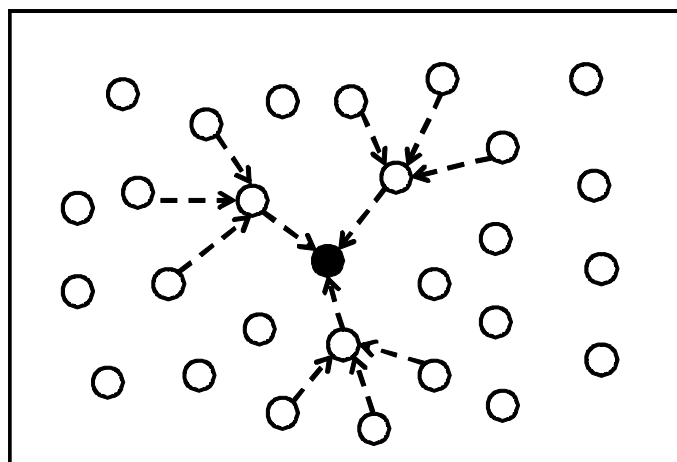
上記のように、ピア情報取得、ピアの選別の 2 ステップにより、選別されたピア R 個のピアを新たに、検索メッセージを発信、転送先として決定し、ノードはその R 個のピアと隣接関係を結ぶ。以前の転送先のピアについてはリンクを切断する。このようにして、各ノードは常に転送先数を一定にするようにする。

このピア情報取得，ピアの選別，隣接ピア組み換えの一連の動作により，コンテンツの検索・ダウンロードというプロセスにおいて，ダウンロード時間を短縮し，検索効率の向上を実現する．また，この従来手法ではこの一連の動作は，周期 T ごとに行うとしている．



—————→ ピア情報取得用メッセージ

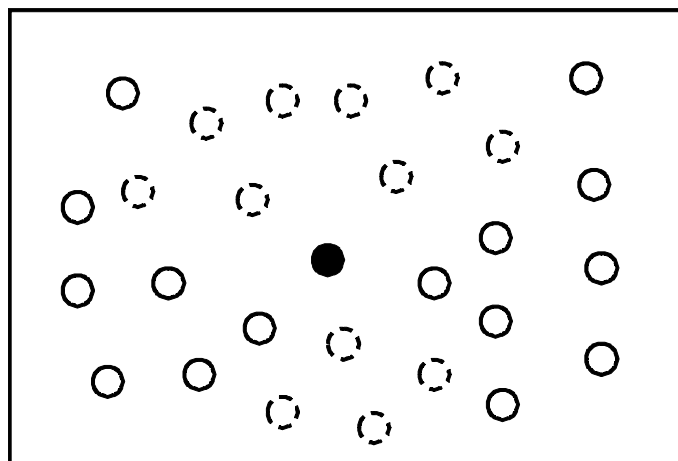
(a) ピア情報メッセージ送信



- - - - -→ ピア情報が格納された応答メッセージ

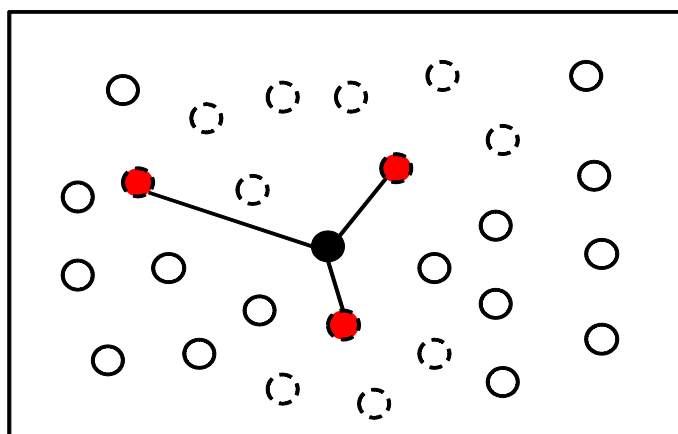
(b) 応答メッセージ

図 4.2: 転送先候補ピアの選定 1



(○) ピアリストに格納されたピア

(a) ピアリストに格納



(○) ピアリストに格納されたピア

(●) 検索メッセージ転送先候補

(b) 転送先ピア候補の決定

図 4.3: 転送先ピア候補の選定 2

4.4 従来手法による問題点

従来手法では、ピアの選別をする際、ピアとピアとの間で保有コンテンツのジャンルの重複率の和によりピアとピアの嗜好の近さを算出している。つまり、ピアが保持しているコンテンツ数が多いジャンルについて優先した検索メッセージ候補先を決定することになる。そのため、ピアが保持しているコンテンツ数が少ないジャンルについて検索効率は上がらないという問題点が挙げられる。問題点を示す例を図 4.4 挙げ説明していく。

図 4.4 では、ピア A がピア B、ピア C のどちらを転送先候補として決定するのかという場面の例である。従来手法に従って、ピア A は、ピア B もしくはピア C との間の嗜好の近さを算出している。

ピア B との嗜好の近さは、評価値 0.5 (音楽のジャンルによるもの) として評価され、また、ピア C との嗜好の近さは、0.4 (0.1 + 0.2 + 0.1 … 音楽、ゲーム、映画のジャンルによるもの) として評価される。したがって評価値の高い、ピア B を転送先として決定する。ピア A にとって、音楽というジャンルについては検索する効率は向上する期待は上がったが、ゲーム、映画というジャンルについては、ピア B は保持しておらず、このジャンルについての検索効率は上がらない。また、従来手法においては、2 ホップ目以降の検索メッセージ転送についても、転送先の転送先決定方法に依存している。このため、ピア C のほうが、ピア A が保持しているジャンルの種類を持っている (音楽、ゲーム、映画) ので、2 ホップ目以降で、音楽というジャンルしか共通していないピア B よりも嗜好の近いピアに転送される期待ができる。したがって、ジャンル比が低いジャンルのコンテンツの検索についても検索効率を向上する余地が残されている。

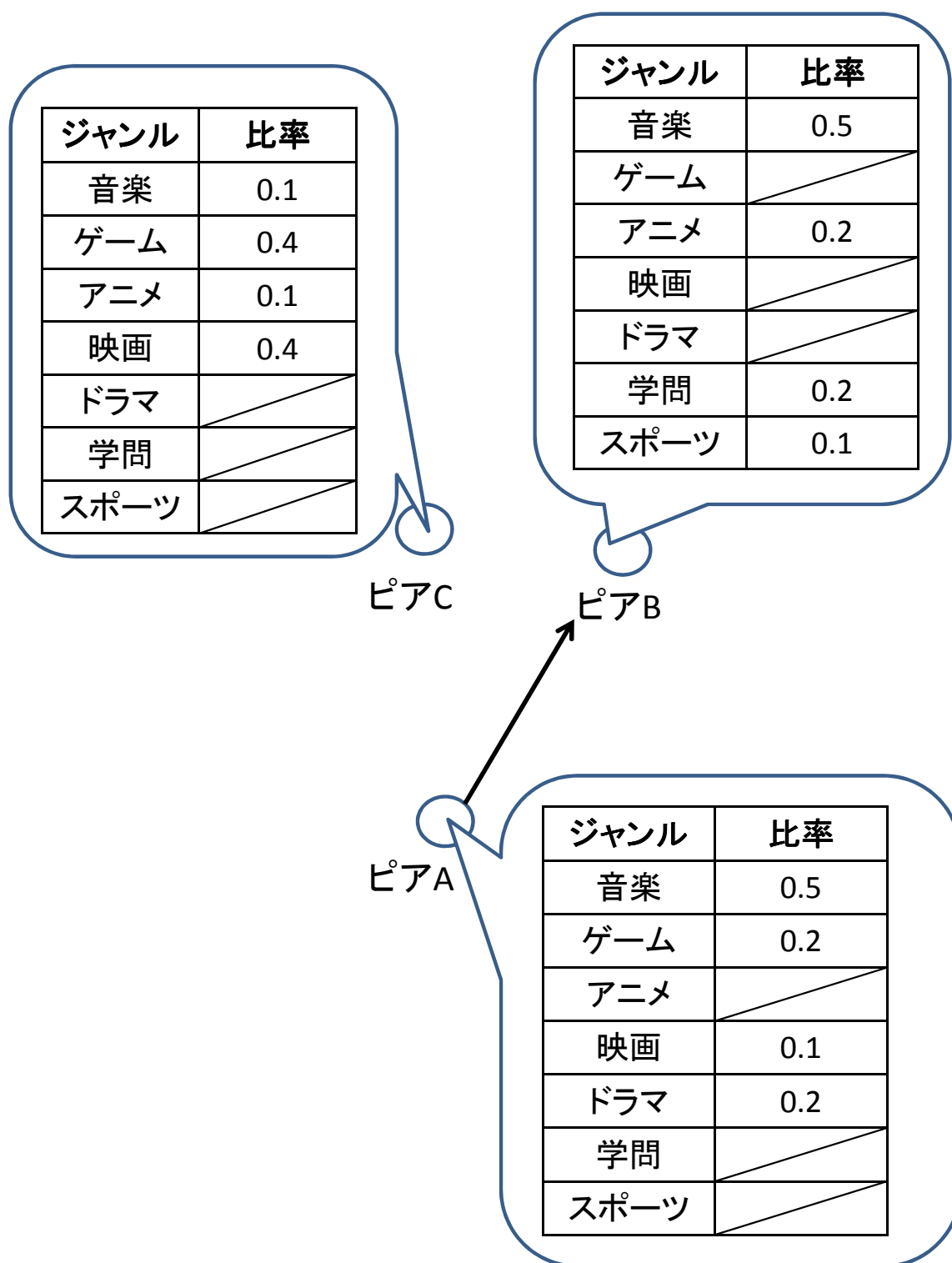


図 4.4: 従来手法による問題点

第5章

提案手法

本章では、ユーザの嗜好を考慮し検索効率を向上させる手法を提案する。従来手法では、コンテンツの保持比率の高いジャンルが主に優先された検索メッセージ転送先決定方法であったため、比較的コンテンツの保持比率の低いジャンルについての検索効率が悪いという改善点があった。本手法では、コンテンツの保持比率に関わらず検索効率の向上をする手法を提案する。

5.1 提案手法の概要

本提案手法は、コンテンツの保持比率に関わらず検索効率を向上させる手法である。コンテンツ配信を想定した P2P ネットワークにおいて、希望するコンテンツの属するジャンルを保持しているピアに検索メッセージが発信・転送されるようにする。各ピアは、ピア情報取得メッセージを送信し、ピア情報取得メッセージを受け取ったピアは、自身のピア情報（IP アドレス、ピアの保持コンテンツのジャンル比率情報）を返信することによって、自身以外のピア情報を得る。このピア情報を用いて、ピアはコンテンツの検索メッセージ転送先候補を決定する。従来手法では、コンテンツ保持比率の高いジャンルが優先される決定方法であったが、この従来手法で決定された検索メッセージ転送先候補に加えて、ある程度コンテンツ保持比率の低いジャンルも優先されるような決定方法を用いて、検索メッセージ転送先候補を用意する。検索時は、各ノードの保持するジャンルごとの転送先決定テーブルという確率テーブルを用いて確率

的に検索メッセージを転送する。

また，コンテンツ検索メッセージを送信時にピギーバック方式を用いて，ジャンル検索メッセージも併せて送信する。ジャンル検索メッセージを受け取った，該当ジャンルを保持しているピアは，検索メッセージ送信元ピアに返信する。検索元ピアは，強化学習 [20] を用いて該当ジャンルの転送先決定テーブルを更新する。この一連の動作により，希望するコンテンツを保持していると可能性の高いピアに検索メッセージを発信・転送するため，検索効率の向上を実現することができる。

5.2 想定される環境

提案手法では，従来手法と同様に，P2P ネットワーク上に存在するコンテンツは正確にジャンル分けすることができ，ジャンルの選択やユーザによるタグ付け等の手段により一意にジャンル分けすることができるとする。またネットワークの初期トポロジはBA モデル [16] を想定している。

BA モデル

BA モデルは，Barabasi と Albert により提案された，スケールフリーであるグラフを生成するモデルである。BA モデルは以下のアルゴリズムでグラフを生成していく。

1. m 個のノードからなる完全グラフ K_m からネットワークの増大がスタートする。
2. 新しいノードを 1 個追加した場合，そのノードから既に存在している m 個のノードにリンクを張る。この時，リンクが張られる確率はそれぞれのノードのその時点での次数 k に比例する。
3. 2. を，ノードが所定の数になるまで繰り返していく。

5.3 提案手法の詳細

提案手法では、主に2つのステップに分けることができる。1つは検索メッセージ転送先候補ピアを決定するための動作として、周辺のピア情報の取得、ピアの選別、隣接ピアの組み換えの一連の動作(図5.1, 図5.2)である。2つ目は、コンテンツ検索メッセージを送信した際の強化学習を用いた転送先決定テーブルの更新の一連の動作である。本節では、順に、提案手法の詳細について説明していく。

5.3.1 ピアにおける情報の取得

従来手法と同様に、各ピアはピア情報取得メッセージを Gnutella 方式のフラッディングを用いて送信し、ピア情報取得メッセージを受けとったピアは、ピア情報(IP アドレス、ピアの保持コンテンツのジャンル比率情報等)を返信する。この操作により、各ピアは他のピア情報を受けとることができる。また、各ピアはピアの識別子やそのピアが保持するコンテンツのジャンル比率情報が格納されている、ピアリストを保持しており、ピア情報を得ることができたピアのピア情報をピアリストに格納していく。また、ピアリストには容量があるとする。

5.3.2 ピアの選別

ピア情報取得フェーズによってピアリストに格納されたピアから選別を行い、検索メッセージ転送先候補を決定する。従来手法では、ピアリストにある全てのピアとの間で保有コンテンツのジャンルの重複率の和によって、嗜好性の近さを評価値として算出し、ピアリストの中で評価値の高い上位 R 個のピアを検索メッセージ転送先としていた。本手法では、従来手法によって決定した R 個の検索メッセージ転送候補ピアに加えて、別の決定方法に基づいて検索メッセージ転送先候補ピアをさらに R 個選定する。つまり提案手法では、2種類の決定方法(以下決定方法1, 決定方法2)により、転送先候補ピアを合計 $2R$ 個用意している。決定方法1では、従来手法同様に前章で挙げた評価関数を用いて、ピアリストにある全てのピアとピアの嗜好性の近さを評

価値として算出し，評価値の上位 R 個を転送先候補ピアとして決定する．決定方法 2 では，ピアリストにある評価値のある程度高い上位 n ピアからランダムに R 個のピアを転送先候補ピアとして決定する．上位 R ピアに限定せずに，ある程度緩やかに決定することにより，従来手法のようなコンテンツのジャンル比率の高いジャンルについての検索効率を優先されていた決定方法に比べ，ある程度，ジャンル比率が低いジャンルについても優先されるように選定されるようになる．

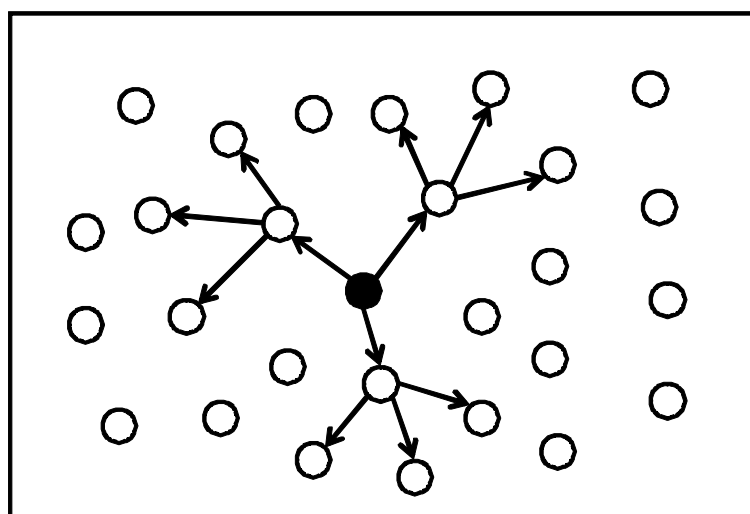
5.3.3 隣接ピアの組み換え

ピアの選別によって決定した， $2R$ 個の検索メッセージ転送先候補ピア全てと接続し，隣接関係を築く．これまで接続していたピアとの隣接関係は全て解消し，常に検索メッセージの転送先候補ピア数は $2R$ 個になるようにする．

5.3.4 検索時の動作

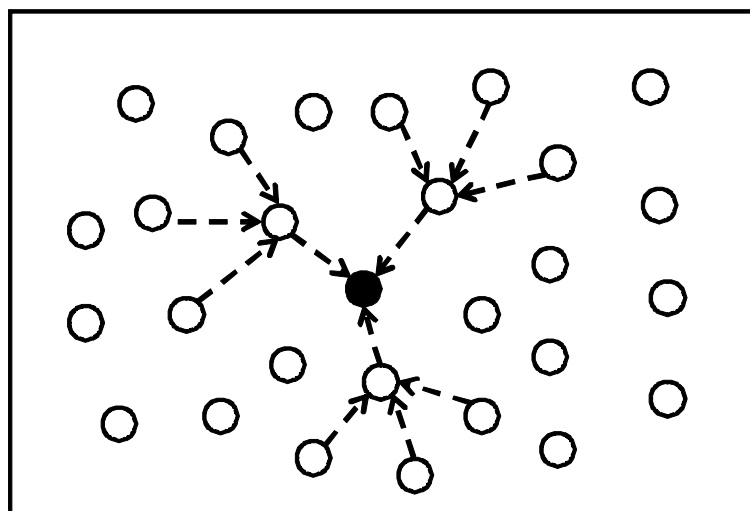
各ピアはコンテンツを検索する際，確率的ルーティングを用いる．各ピアはそれぞれ，ジャンルごとの転送先決定テーブルを持っている (図 5.3)．転送先決定テーブルとは，あるジャンルについて検索メッセージを発信または転送する際に，どの転送先候補ピアがどの程度の確率で選択されるかという確率値が記されているテーブルのことである．仮に，ジャンルが 10 種類存在したとすると，10 個の確率テーブルを持つことになる．

各ピアは検索メッセージを送信する際には，検索されたジャンルに対応する転送先決定テーブルを参照し，転送先として R 個のピアを選択し検索メッセージを転送する (図 5.4)．検索メッセージには，TTL 値が設定されており，ピアは検索メッセージを受けると TTL の値を 1 減じて，TTL が 0 にならない限り転送する．また，転送の際もまた，検索コンテンツに対応するジャンルについての自身の転送先決定テーブルを参照し，転送先ピアを決定し検索メッセージを転送する．コンテンツを発見時には，コンテンツ発見の通知メッセージを検索メッセージが届いた経路を遡り，検索元ノード



————→ ピア情報取得用メッセージ

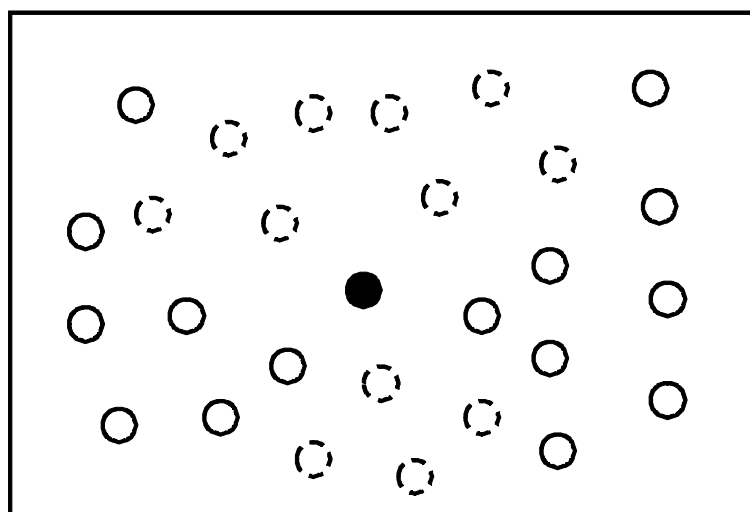
(a) ピア情報メッセージ送信



-----→ ピア情報が格納された応答メッセージ

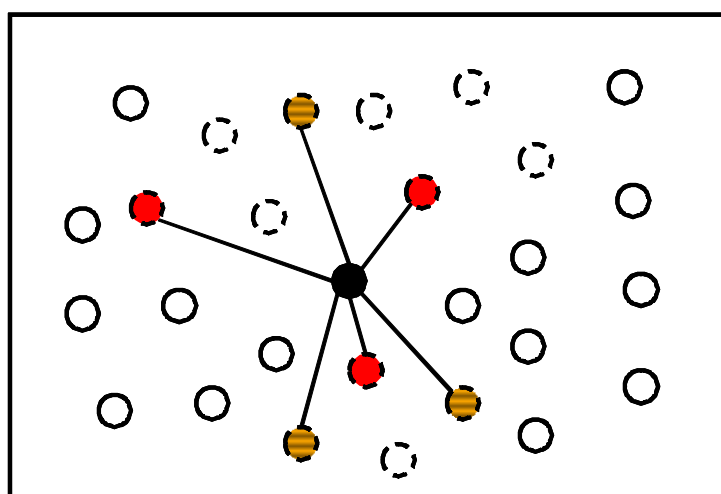
(b) 応答メッセージ

図 5.1: 転送先候補ピアの選定 1



(○) ピアリストに格納されたピア

(a) ピアリストに格納



(○) ピアリストに格納されたピア

(●) 決定方法1による検索メッセージ転送先候補

(●) 決定方法2による検索メッセージ転送先候補

(b) 転送先ピア候補の決定

図 5.2: 転送先ピア候補の選定 2

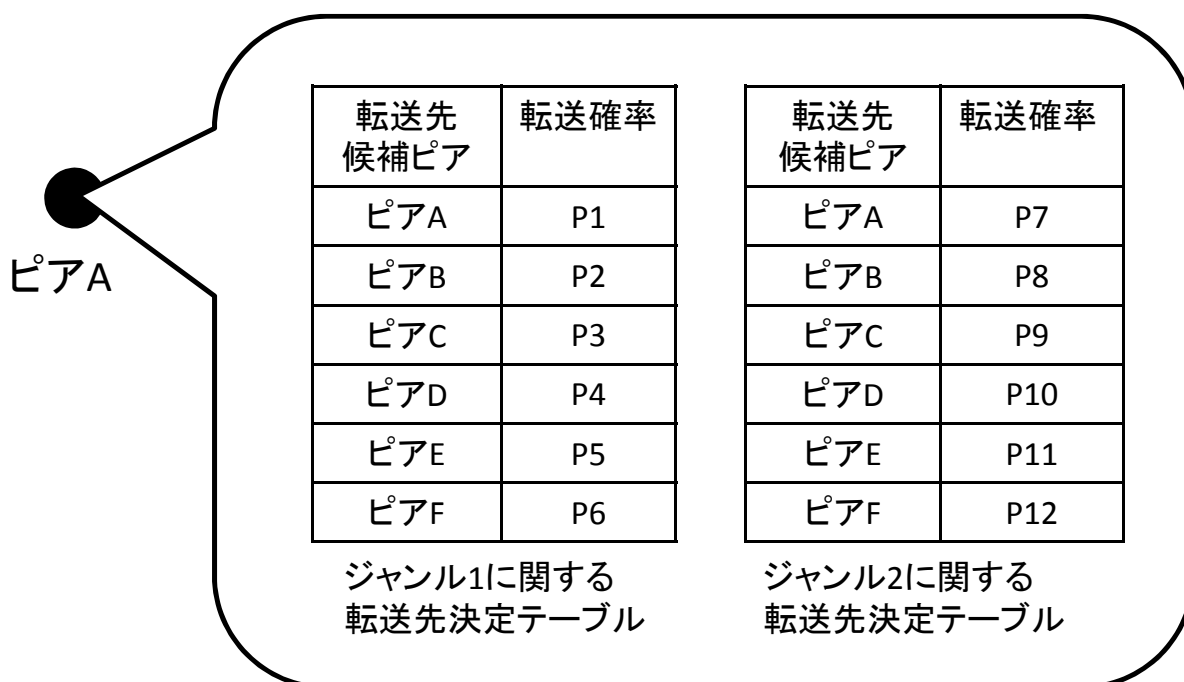


図 5.3: ピアと転送先決定テーブル

に返信される。

5.3.5 検索メッセージ転送テーブルの更新

本提案手法では、検索メッセージを送信する際にピギーバック方式を用いて、ジャンル情報について検索するメッセージを送信し、その返信に基づき、強化学習により転送先決定テーブルを更新する方式をとっている。各ピアは希望するコンテンツを検索する際には、検索メッセージに検索するコンテンツ情報とそのコンテンツが属するジャンルの情報を格納し検索メッセージを発信する。前節の検索時の動作において、このメッセージはコンテンツが見つかるまで転送されていくが、この過程において希望するジャンルを保持する最初のピアは、検索元ノードに向け報酬値の情報を記載した報酬メッセージを返信する。検索元ピアは、強化学習を用いて転送先決定テーブルを更新するが、その際の更新にはこの報酬値が用いられる。この強化学習の概念は、AntRouting[17],[18],[19] という強化学習を基本としたルーティングアルゴリズムを参

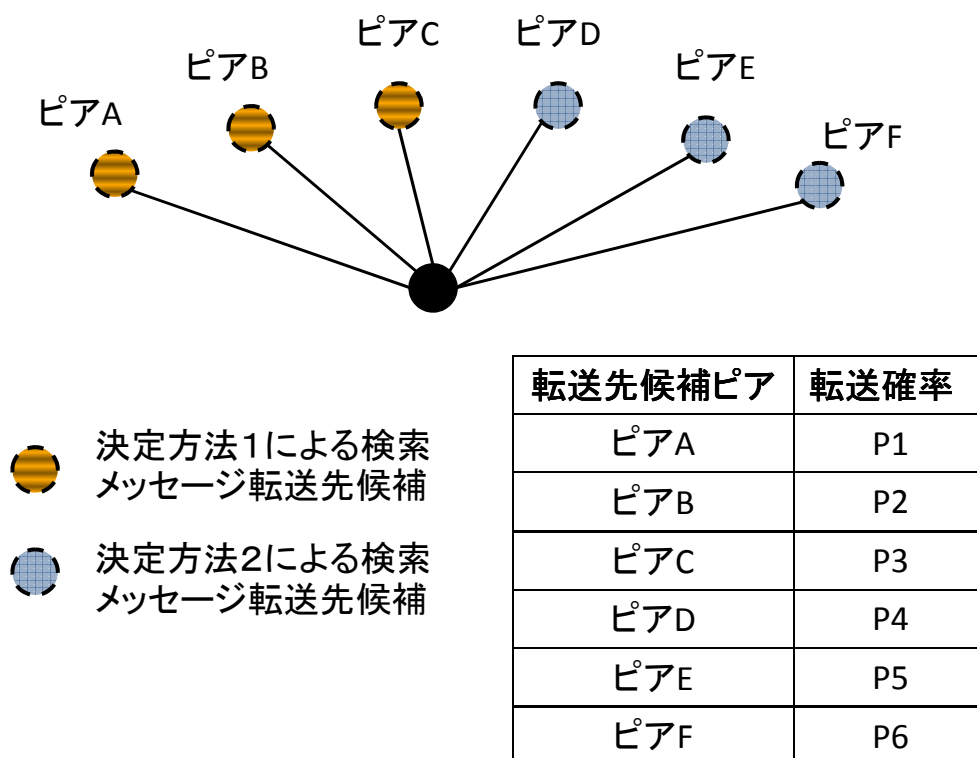


図 5.4: 転送先決定時の例

考にしている．コンテンツ検索時に，希望するジャンルを保持しているノードからのメッセージを受信することにより，受信したメッセージと逆向きの経路を，有効な経路として学習する．報酬値（Reward）は以下の条件によって決定される．

$$Reward = F \times W^{rank}$$

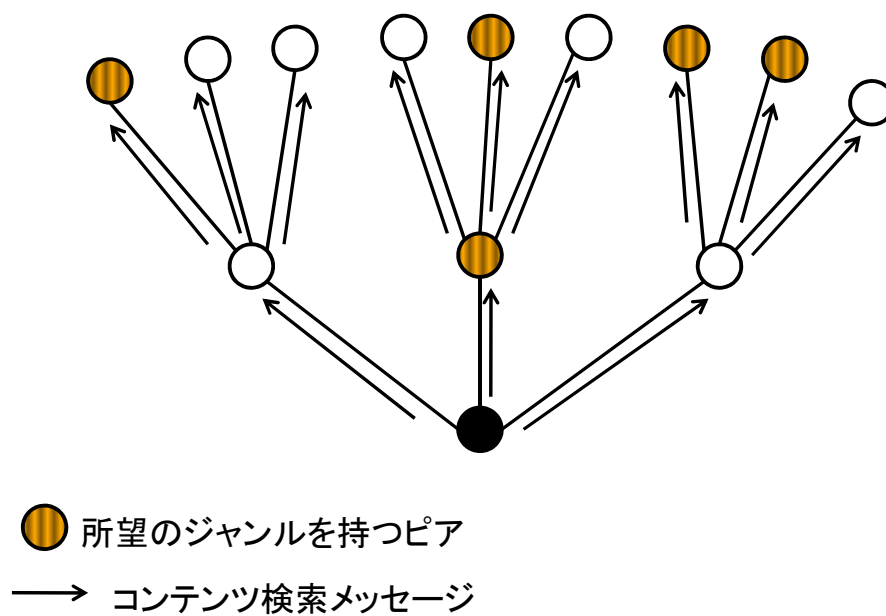
上記式において F は発見した場合は一律の値になるような強化値， $W(0 < W \leq 1)$ は減少値， $rank$ は，希望するジャンルを保持するピアにおけるジャンル比率の順位である．希望するジャンルを持っているピアにおいて，該当するジャンルのコンテンツ保持比率がそのピアにおいて最も大きければ， $rank = 1$ となり，2 番目に大きければ $rank = 2$ となる．発見時の強化値に，減少 W ，コンテンツ保持比率の順位 $rank$ からなる値をかけることにより，よりコンテンツ保持比率の高いピアから得られる報酬値が高くなる．また，この報酬メッセージは，検索メッセージが辿ってきた経路を遡り送

られていくが，その途中にあるピアも報酬メッセージから報酬値を獲得する．そして，転送先決定テーブルを更新し，報酬値を一定値 $N(N < 1/\text{転送先決定数 } R)$ で乗算し，少なくした上で，検索元ピアと接続するリンクへの報酬メッセージを送信する．これにより，より近く，より該当ジャンルのコンテンツ保持比率が大きいピアからの報酬値が大きくなるようになる．また，各ピアは得られた報酬値を元に転送先決定テーブルを更新する．この時，確率テーブルは以下の式により更新される．

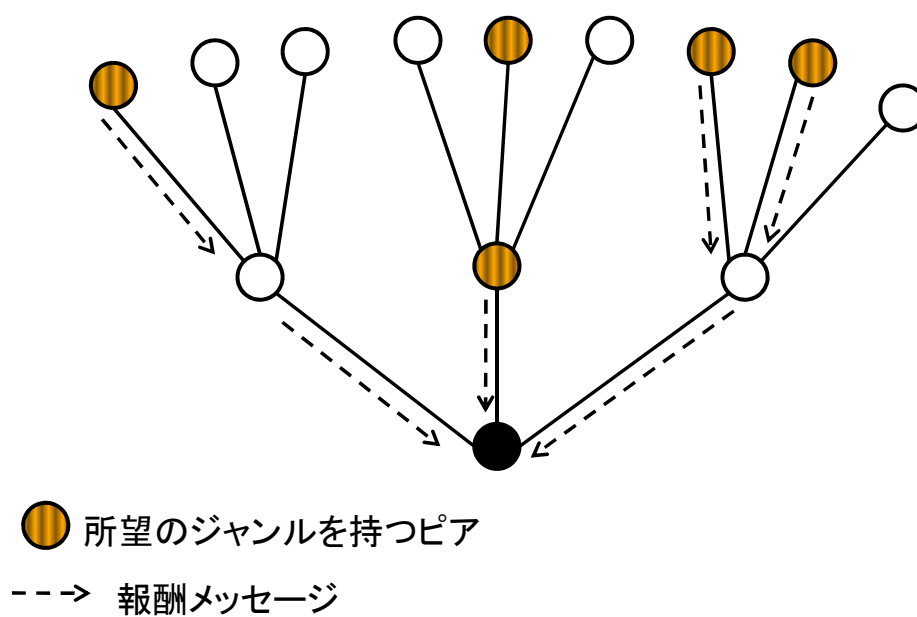
$$P_{genre}(z) = \begin{cases} \frac{P_{genre}(z) + \Delta p}{1 + \Delta p} & (z = x) \\ \frac{P_{genre}(z)}{1 + \Delta p} & (\text{上記以外}) \\ (z \in neighbor, x \text{ は報酬メッセージを送ったピア}) \end{cases}$$

$$\Delta p = F \times W^{rank} \times N^{hop}$$

上記式において， Δp は，転送先決定テーブルの更新に用いられる報酬値となる．この式は，先に説明したとおりより近く，より該当ジャンルのコンテンツ保持比率が大きいピアからの報酬値が大きく割り当てられる． hop は，該当ジャンル発見ピアから，自身のピアまでのホップ数である．このように，報酬メッセージを送ってきたピアは，希望するジャンルを保持しており，そのピアにそのジャンルに関して検索メッセージを送信したほうが検索効率の向上が期待できる．また，オーバーレイネットワーク上において近いピアのほうがより報酬値が高くなるためにより，該当ジャンルをより近くで保持しているピアに検索メッセージが転送されやすくなる．また，ピアの選別のステージでは保持しているジャンル情報を元に転送先候補を決定したため，希望するジャンルを保持しているピアの隣接ピアもまた希望するジャンルを保持している確率が高い．したがって，希望するジャンルを保持しているピアに検索メッセージを転送することにより，検索効率の向上を期待できる．



(a) コンテンツ検索



(b) 希望するジャンルの発見による応答

図 5.5: コンテンツ検索・報酬値の獲得の手順

第6章

評価

本章では，提案手法，従来手法の両手法の評価を行い提案手法の有用性を検証する．まず，比較手法，シミュレーションモデルについて説明した後に，該当ジャンル保持ピア数，ヒット率，メッセージ数，スケーラビリティの評価を行い，考察を行う．

6.1 本研究における比較手法

本章では比較する比較手法は従来手法としている．比較手法では，第4章で説明したように，ピア情報取得，ピアの選別，隣接ピア組み換えという3ステップを行っているが，ピアの選別時において，RTTによる選抜はしていない．これは，本研究ではセマンティック P2P ネットワークにおいて，検索効率のみを取り上げており，ダウンロード時間による影響を無くし，手法を比較したいためである．

また，トポロジの変更前のネットワークの初期状態は，BA モデルであり，さらに，ピア情報取得，ピアの選別，隣接ピア組み換えの動作を各ピアが1回行うものとする．

6.2 シミュレーションのモデル

提案手法，比較手法の両手法を評価する上での前提としている評価条件について説明する．各評価で特別に説明していない限り，本節で説明するパラメータで評価を行っている．

まず、両提案手法における初期状態のネットワークトポロジーは、BA モデルを構築している。BA モデル構築において、ネットワークに参加するピアの接続先ピア数は、コンテンツ検索時に検索メッセージを転送するピア数と同数とした。BA モデル構築過程において、初期配置されるピアは、BA モデルのトポロジー構築後にランダムに接続先を選択し直すものとした。また、ピア情報取得メッセージ送信時の TTL は 4 とした。

次に、各ピアが保持するコンテンツについて説明する。両方式において、各ピアは全部で 20 種類のジャンルから 5 ジャンルを選びその 5 ジャンルについてのコンテンツを保有する。またコンテンツ数は 100,000 とし、各ジャンルに割り当てられるコンテンツ数は均等になる（つまり 1 ジャンルあたり 5000 コンテンツ。）ピアが持つ各ジャンルの保有コンテンツ数の比率は、Zipf 則に従う。各ピアは、検索をするコンテンツは自身が興味をもつ 5 ジャンルに対してのみ検索を行うものとする。また、各ピアが検索するジャンルの割合は、保持するジャンル比に従った割合になる。さらに、コンテンツの人気度について Zipf 則に従う偏りで重み付けされているものとする。

Zipf 則とは、出現頻度順に並べたとき k 番目に大きい要素が全体に占める割合が、

$$f_k = xk^{-\alpha} \quad (6.1)$$

に比例するとされる経験則である。 f は k 番目の出現頻度の大きさの割合である。 x は f の合計値が 1 になるような定数である。本節以降、評価パラメータとして、Zipf 則における偏りの度合いである、 α の値を変化させて、偏りの状況を変化させている。次節以降において、Zipf 則による変化させるパラメータは、先に説明した、

- 各ピアが保持するジャンルごとのコンテンツの配分
- ジャンルの人気度

の 2 種類である。また、ピア情報取得時の TTL、コンテンツ検索時の TTL は共に 4 とし、ピアリスト容量は 200 ピアとした。検索メッセージの転送先ピア数を 4 ピアに固定した。また、学習時に用いられる強化値は 0.15、減少値は 0.95 とした。

初期トポロジ	BA モデル
ピア数	20000
ジャンル数	20
コンテンツ数	100000
1ジャンルあたりのコンテンツ数	5000
ピア容量リスト	200
ピア情報取得メッセージの TTL	4
コンテンツ検索メッセージの TTL	4
検索メッセージ転送ピア数	4
コンテンツの人気度	Zipf 分布

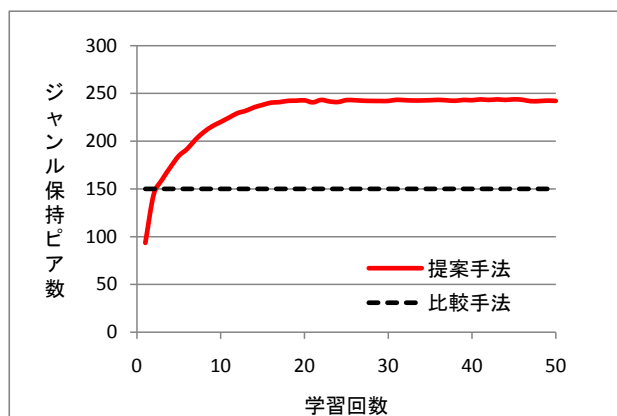
表 6.1: 主な評価パラメータ値

6.3 該当ジャンル保持ピア数・ヒット率

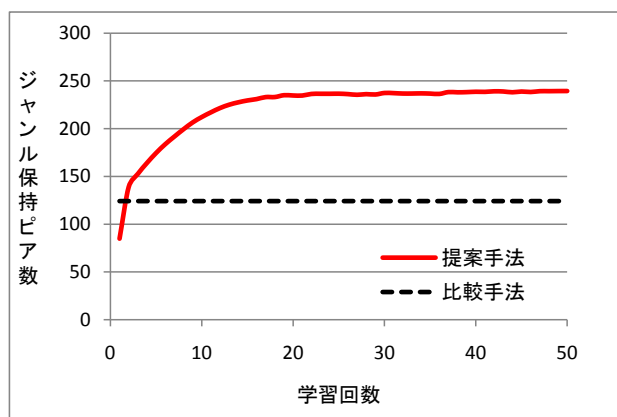
本節では、提案手法と比較手法と該当ジャンル保持ピア数を比較する。該当ジャンル保持ピア数とは、検索元ピアがコンテンツを検索した場合、検索メッセージが届く範囲における希望するコンテンツの属するジャンルを保持しているピア数を表す。希望するコンテンツの属するジャンルを保持するピアが多ければ多いほど、希望するコンテンツが発見され易くなる為、該当ジャンル保持ピア数が多いほうが高い発見を期待できる。ここでは、パラメータを変化させて様々な状況についての評価を行う。行う評価は、

- (a) コンテンツの保持比率は偏りがあり、ジャンルの人気度は一律である場合
- (b) コンテンツの保持比率は一様に近く、ジャンルの人気度は一律である場合
- (c) コンテンツの保持比率は偏りがあり、ジャンルの人気度も偏りがある場合

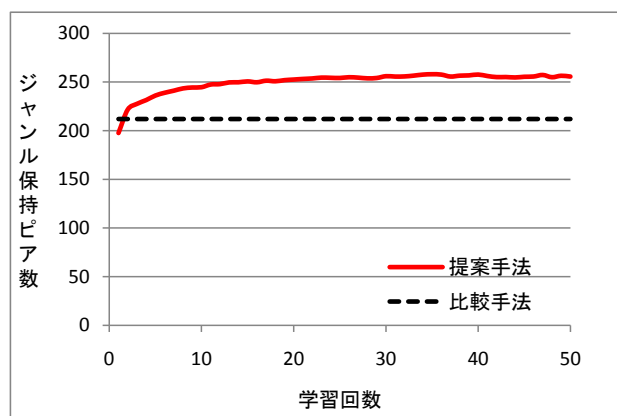
の3種類の状況について評価を行った。コンテンツの保持比率とジャンルの人気度に用いられる Zipf 分布の式 6.1 における α の値は 1、分布が一様に近い場合の分布は α の値は 4 となっている。また、提案手法では強化学習を行っている。各ピアが保持する全ジャンルについて 1 回ずつ学習するごとに、ジャンル保持ピア数の変化を調べている。該当ジャンル保持ピア数の比較を図 6.1 に示す。



(a) コンテンツの保持比率は偏りがあり、ジャンルの人気度は一律である場合



(b) コンテンツの保持比率は一様に近く、ジャンルの人気度は一律である場合



(c) コンテンツの保持比率は偏りがあり、ジャンルの人気度も偏りがある場合

図 6.1: 該当ジャンル数

図 6.1 に共通した傾向として、提案手法では、学習回数を重ねるごとにジャンル保持ピア数が大きくなっていることが分かる。この理由として次の 2 点が挙げられる。比較手法においては、転送先決定数は 4 つとされていたが、提案手法では転送先決定先候補を比較手法によって決定された転送先候補 4 つに加え、別の決定方法により決定された転送先 4 つ (合計 8 つ) を保持しており、ピアが保持するジャンルごとの有利不利の差異が少なくなったことが 1 点目の理由である。また、強化学習方式により、ジャンル情報を基に転送先を学習し、ジャンルごとに転送先を決定している。検索するジャンルごとに決定テーブルを用いて転送先を変えていることによって、各ジャンルごとにジャンル保持ピア数が増加すると考えられるピアを転送先を選択しているという点があるが、2 点目の理由である。

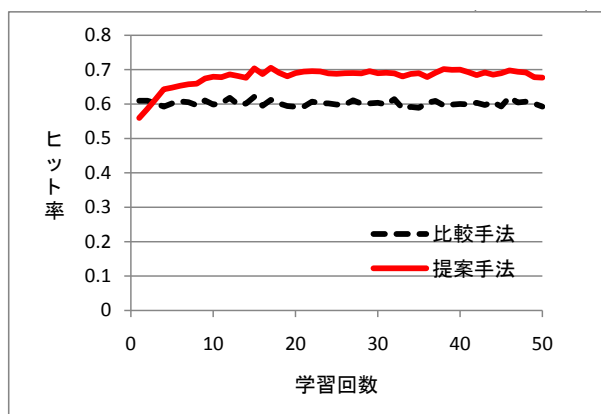
次に、学習回数が少ない段階においては比較手法の方が、該当ジャンル保持ピア数が大きくなっていることが分かる。提案手法では、初期状態として確率テーブルは均等に初期化されている。このため、学習が進むまでは提案手法はより適切な検索メッセージ転送先が選択されておらず既存手法のほうが該当ジャンル保持ピア数が大きい。

次に、図 6.1(c) では、図 6.1(a)(b) に比べて、提案手法において強化学習による、ジャンル保持ピア数の上昇数は少なく、さらに、両手法において初期状態におけるジャンル保持ピア数も大きい。これは、多くのピアが同じようなジャンルを保持しているためである。このことから、本提案手法は各ピアが色々なジャンルのコンテンツを持っている状況において、特に有効であると言える。

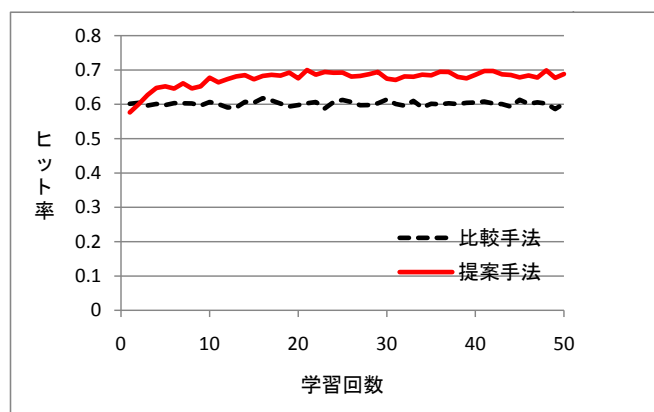
また、同じ条件下でピアがコンテンツを検索した場合のヒット率 (HitRate) も評価した (図 6.2)。

ヒット率は、以下の式にて評価する。

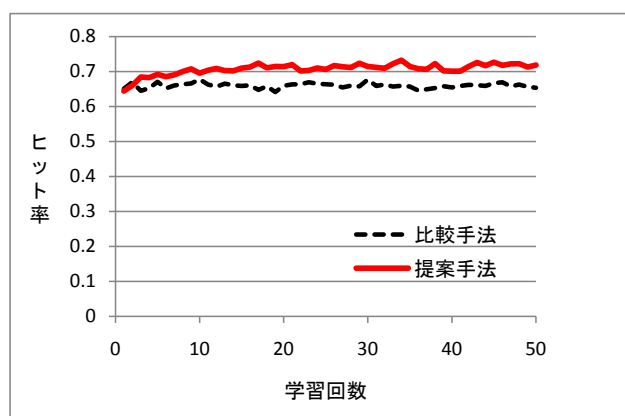
$$HitRate = \frac{\text{コンテンツの発見回数}}{\text{コンテンツの検索回数}} \quad (6.2)$$



(a) コンテンツの保持比率は偏りがあり、ジャンルの人気度は一律である場合



(b) コンテンツの保持比率は一樣に近く、ジャンルの人気度は一律である場合



(c) コンテンツの保持比率は偏りがあり、ジャンルの人気度も偏りがある場合

図 6.2: ヒット率

ヒット率を見ると、両手法共に、図 6.1 の結果に似たグラフの傾向にあった。コンテンツ検索はそのコンテンツが属するジャンルを保持しているノードでのみ希望するコンテンツを発見できる可能性がある。従って、該当するジャンル保持ノードが検索範囲に多いほど、ヒット率が上がるために、図 6.1 の結果に似た傾向になるといえる。

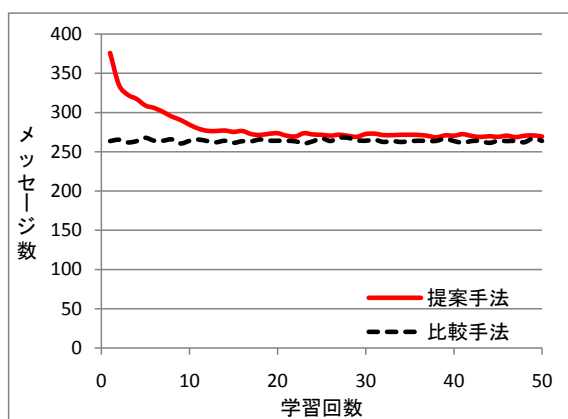
6.4 メッセージ数

次に、該当ジャンル発見までのメッセージ数を提案手法と比較手法で比較する。メッセージ数については、図 6.3, 図 6.4(a) ~ (c) の 3 種類についての評価を行っているが、(a) ~ (c) の条件については前節の評価と同じである。

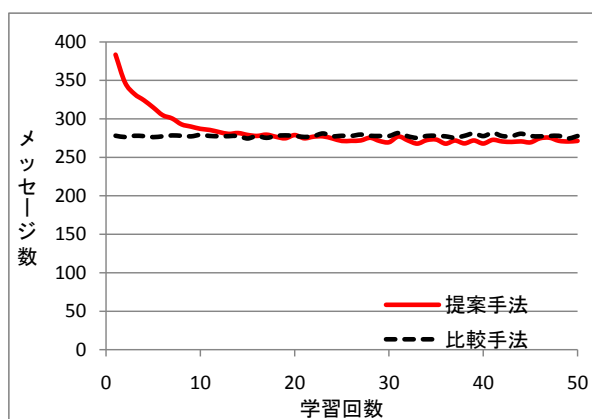
両手法において、各ピアはコンテンツを検索する際に、検索メッセージを転送している。提案手法では、この検索メッセージにピギーバック方式によりジャンル検索メッセージも併せて送信し、コンテンツが属するジャンルを保持するピアも検索している。学習回数増加に伴うメッセージ数の変化を図 6.3 に示す。

図 6.3 に共通する傾向として、学習回数が少ない段階では、既存手法に比べて提案手法では多くのメッセージ数が発生している。その後、学習回数を重ねるにつれ、そのメッセージ数は減少し一定値に収束する。ここでは、発生し得るメッセージは、コンテンツ検索メッセージ、希望するコンテンツが属するジャンルを保持するピアを発見した場合にそのピアから返信される報酬メッセージの 2 種類が挙げられる。ここで、メッセージの内訳を示すために提案手法における学習回数増加に伴う報酬メッセージ数の推移を図 6.4 に示す。

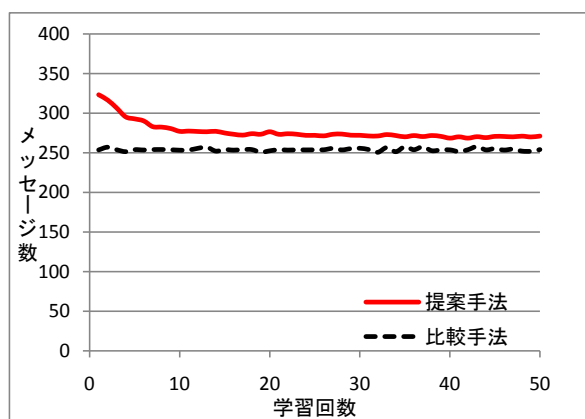
図 6.4 のすべての図において、学習回数が少ない段階では多くの報酬メッセージが発生していることが分かる。この理由について、説明する。提案手法では、検索メッセージはフラディング方式で転送され、ホップ数を重ねるごとに検索メッセージを受け取るピアは指数的に多くなる。これに伴いネットワーク全体に発生する検索メッセージ数も増大する。したがって、検索メッセージの返信として送信される報酬メッセージも増大するという点が 1 つ目の理由である。検索元より希望するジャンルを保持して



(a) コンテンツの保持比率は偏りがあり、ジャンルの人気度は一律である場合

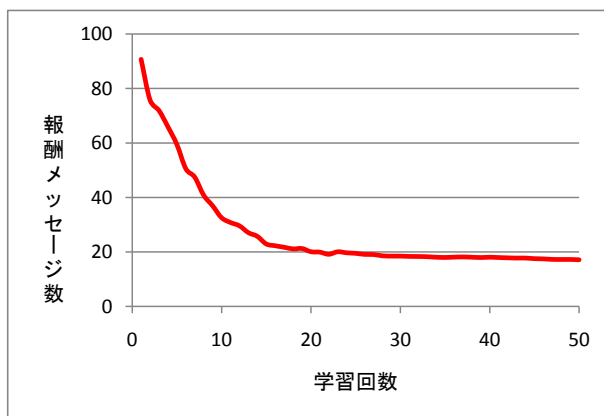


(b) コンテンツの保持比率は一様に近く、ジャンルの人気度は一律である場合

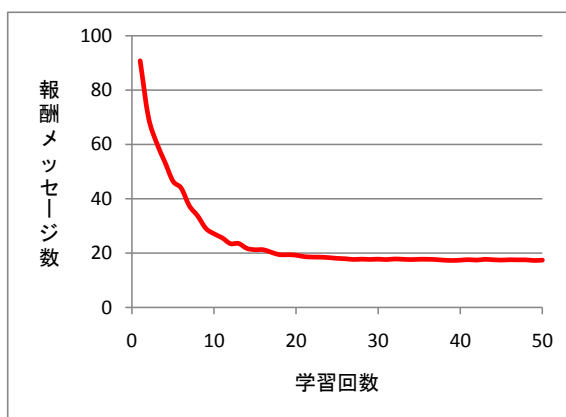


(c) コンテンツの保持比率は偏りがあり、ジャンルの人気度も偏りがある場合

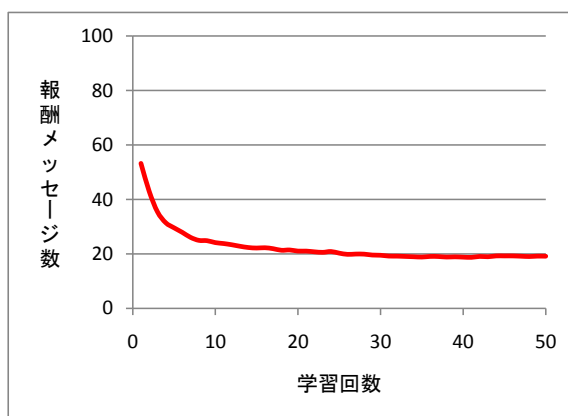
図 6.3: メッセージ数



(a) コンテンツの保持比率は偏りがあり、ジャンルの人気度は一律である場合



(b) コンテンツの保持比率は一様に近く、ジャンルの人気度は一律である場合



(c) コンテンツの保持比率は偏りがあり、ジャンルの人気度も偏りがある場合

図 6.4: 報酬メッセージ数

いるピアまで検索メッセージが届くホップ数が大きくなるにつれ、報酬メッセージが検索元ピアに届くまでのホップ数が多くなる。これによって、ピア間でやりとりするメッセージが多くなるという点が2つ目の理由である。この2点の理由より、学習回数が少ない段階では、検索元ノードから少ないホップ数で辿り着く希望するジャンルを保持しているピアを学習するまで、多くの報酬メッセージが発生してしまっている。このことから、図6.3のように、学習回数の少ない段階において、メッセージ数が多くなっている1つの要因として考えられる。また、図6.1のように、提案手法では学習回数の少ない段階では、比較手法に比べ、ジャンル保持ピア数が少ない。従って、希望するコンテンツが比較手法に比べ提案手法のほうが発見されにくく、検索メッセージが多く転送されてしまっているという点も要因の1つと考えられる。

学習回数が多くなってくると、図6.3(a)(b)(c)では、メッセージ数が学習回数が少ない段階に比べて減っていることが分かる。これは、学習が進んだことにより検索元ピアが、自身から少ないホップ数に位置する、希望するジャンルを保持しているピアに検索メッセージを送信するようになったことによって、報酬メッセージ数が少なくなったためと考えられる。また、学習によって、希望するジャンルを保持している、より多くのピアに検索メッセージが届くようになったことにより、検索効率が向上し検索メッセージが転送される頻度が減ったことによることも要因の1つと考えられる。

図6.3(a)(b)では、メッセージ数は比較手法に近い水準までに少なくなったが、6.3(c)では、十分学習した後も提案手法は比較手法よりもメッセージ数が多くなっていることが分かる。これは、学習が進みジャンル保持ピア数が多くなってもなお、報酬メッセージ数のメッセージ数分だけ、検索メッセージを減らなかつたためである。このことから、学習によりジャンル保持ピア数の増加が少ない場合には、メッセージ数が多くなってしまふことが言える。

6.5 全ピア数を変化に伴う該当ジャンル保持ピア数の評価

次に，提案方式のスケラビリティを調べるために，ネットワークに参加するピア数を変化させて，ジャンル保持ピア数を提案手法，比較手法で比較をした．ピア数は2000 から 20000 ピアまで 2000 ピア増加させ，ジャンル保持ピア数の変化を調べた．ここでは，コンテンツの保持比率は偏りがあり，ジャンルの人気度は一律とした．提案手法では，学習回数 50 回行った時点でのジャンル保持ピア数を計測している．結果を図 6.5 に示す．

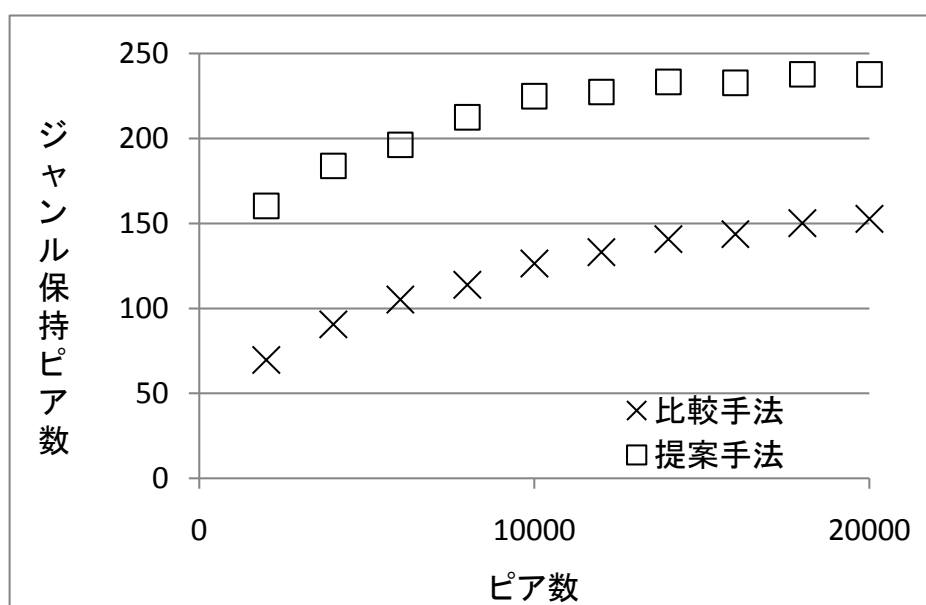


図 6.5: ジャンル保持ピア数の推移

ピア数増加に伴って，両手法とも増加し収束していっていることが分かる．比較手法に比べ提案手法ではジャンル保持ピア数は大きくなっている．この関係は，ピア数が変化しても変わらない．また，両手法に共通して，検索メッセージ TTL を固定値にしているにも関わらず，ピア数が減少すると共に少なくなっていることが言える．これは，参加するピア数の減少に伴い検索メッセージが同じピアに届けられる確率が高くなっているためである．

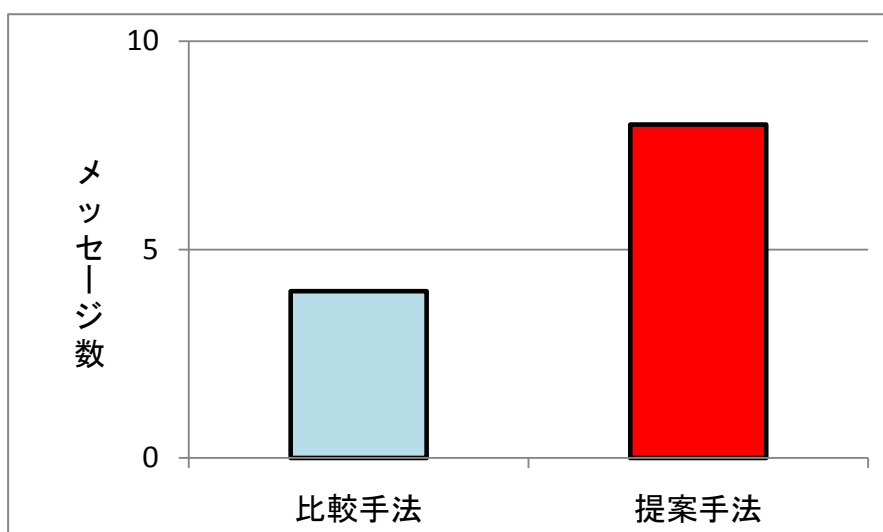


図 6.6: 状態確認メッセージ

6.6 提案手法における改善点

提案手法における改善点を挙げる．まず，学習回数が少ない段階において，提案手法では比較手法に比べてジャンル保持ピア数少ないという点や，メッセージ数が増大しているという点が挙げられる．これは，提案手法において初期状態で，転送先決定テーブルを均一に初期化していることや，学習回数が少ない段階では，希望するジャンルを保持するピアを学習するために多くのメッセージが転送されていることが要因と考えられる．次に，提案手法では，比較手法に比べて，転送先候補とするピアを多く保持している．ピアと接続している場合，リンクが切れているかというリンクの状態を確認する状態確認メッセージが定期的に送られている．図 6.6 で，1 回のチェックで送信される状態確認メッセージを比較する．決定先候補先ピアが多ければ状態確認メッセージも増大する．また，リンクの状態を確認する頻度が多いほど，提案手法においての状態確認メッセージの増加についても改善する必要がある．状態確認メッセージ数の削減も課題といえる．

第7章

まとめ

本研究では、ユーザの嗜好性を考慮して構築された P2P ネットワークにおいて、ピアの持つ様々なジャンルについての検索効率を改善する為に、検索メッセージを転送するピアの候補を複数の方針により決定し、その後、強化学習を用いて検索効率の向上が期待されるピアへの経路を学習する手法を提案した。

従来手法ではコンテンツのジャンルのうち優先される一部のみ高い検索効率を得られるが、提案手法では他の様々なジャンルについての検索効率も向上した。また、提案手法では、強化学習に用いられる報酬メッセージによるメッセージ数の増大を抑制するために、自身のピアより少ないホップ数で希望するジャンルを保持するピアへの経路を優先的に学習する。これにより、報酬メッセージ数は抑制される。

実験により提案手法と従来手法を評価し、提案手法は、学習後では、従来手法よりもコンテンツの検索効率が改善することを示した。また、提案手法の検索と学習におけるメッセージ数が、既存手法に近い水準まで抑制されることを示した。

今後の課題としてピア接続するリンク数の増加に伴うリンクの管理コストの増加を抑制すること、学習の初期状態において検索効率を改善するための手法の導入が挙げられる。

参考文献

- [1] "Gnutella" <http://gnutella.wego.com>
- [2] "Freenet" <http://freenetproject.org/>
- [3] "BitTorrent" <http://www.bittorrent.com/>
- [4] "Napster" <http://www.napster.jp/>
- [5] "WINMX" <http://ja.losts.net/WinMX>
- [6] 金子 勇 "Winny の技術" アスキー (2005)
- [7] "Skype" <http://www.skype.com/intl/ja/home/>
- [8] Stoica , I. , Morris, R. , Karger, D. , Kaashoek, F. and Balakrishnan, H.
: Chord; A Scalable Peer-To-Peer Lookup Service for Internet Applications,
Proc. 2001 ACM SIGCOMM Conference, pp. 149-160(2001).
- [9] A. Crespo and H. Garcia-Molina: "Routing Indices for Peer-to-Peer Systems, "
in Proc. the 22nd International Conference on Distributed Computing Systems
(2002).
- [10] A. Crespo and H. Garcia-Molina: "Semantic Overlay Networks for P2P Systems,
" Technical report in Computer Science Department, Stanford University (2002).
- [11] Kazuhiro Kojima: "Self-Organizable P2P Document Search Engine for Knowledge
Management, " in P2P Knowledge Management (2004)

- [12] 立野雅紘, 清水翔, 荒川豊, 山中直明 ”嗜好性を考慮した Peer-to-Peer オーバーレイネットワーク” 電子情報通信学会技術研究報告. IN, 情報ネットワーク 105(628), pp.143-146 (2006)
- [13] 遠藤慶一, 最田健一, 河原稔, 高橋豊 ”コンテンツの類似度に基づいた P2P ネットワークの自己組織化アルゴリズム” 電子情報通信学会論文誌 D Vol.J91-D No.6 pp.1494-1505 (2008)
- [14] 大林功実, 朝香卓也, 高橋達郎, 佐々木純 ”ピア-ピア間スループットを考慮したセマンティック P2P ネットワークトポロジー構築法” 電子情報通信学会論文誌. B, 通信 J91-B(1), pp.35-46 (2008)
- [15] 大林功実, 朝香卓也, 高橋達郎, 佐々木純 ”クリッピングと保持コンテンツ数の偏りを考慮したセマンティック P2P ネットワークの構築法” 電子情報通信学会論文誌. B, 通信 J93-B(2), pp.230-241 (2010)
- [16] Albert-Laszlo Barabasi and Reka Albert ”Emergence of Scaling in Random Networks” Science 286, 509-512, (1999)
- [17] D. Subramanian, P. Druschel and J. Chen: ”Ants and reinforcement learning: A case study in routing in dynamic networksb,” Proceedings of the Fifteenth International Joint Conference on Artificial Intelligence, 2,pp. 832b-839 (1997).
- [18] Hiroshi Matsuo , Kouichi Mori ”Accelerated Ants Routing in Dynamic Networks” in Proc. Intl. Conf. On Software Engineering, Artificial Intelligence, Networking and Parallel/Distributed Computing pp.333-339 (2001)
- [19] 加藤新吾, 松尾啓志 ”動的環境下における Profit Sharing” 電子情報通信学会論文誌. D-I, 情報・システム, I-情報処理 J84-D-I(7), 1067-1075 (2001)
- [20] 宮崎和光, 木村元, 小林重信 ”Profit Sharing に基づく強化学習の理論と応用” 人工知能誌, vol.14, no.5, pp.800-807, (1999)

謝辞

本研究のために多大な御尽力を頂き、日頃から研究方針等、明確なご指導、熱心な的確なアドバイスを賜り、また、ミーティングにて妥協のない徹底的な議論により研究の視野を広げていただきましたな名古屋工業大学の松尾啓志教授に深く感謝致します。大変学びにつながるが多かったです。ありがとうございます。また、研究過程において発表の仕方、研究の進め方にアドバイスをくださり報告の仕方、研究の組み立て方等を松井俊浩助教より多く教わったと思います。深く感謝致します。また、津邑公暁准教授には、研究についての御指導はもちろん、精神的なサポートや進路など、研究を進めるにあたり大変大きな存在でありました。深く感謝致します。また、本研究を進めるにあたり松尾・津邑研究室、齋藤研究室の学生の皆様の存在も大変大きいものでした。松尾・津邑研究室、齋藤研究室のメンバー 1人1人に心より感謝致します。