

P2P ネットワークにおける動的なカテゴリ生成を考慮した検索手法

小林 由依[†] 渡辺 俊貴^{††} 神崎 映光^{††}

義久 智樹^{††} 原 隆浩^{††} 西尾章治郎^{††}

[†] 大阪大学工学部 電子情報エネルギー工学科

^{††} 大阪大学大学院情報科学研究科 マルチメディア工学専攻

E-mail: [†]kobayashi.yukie@ise.eng.osaka-u.ac.jp,

^{††}{watanabe.toshiki,kanzaki,yoshihisa,hara,nishio}@ist.osaka-u.ac.jp

あらまし P2P ネットワークでは、検索効率を向上させるために、内容が類似したデータを同一のカテゴリに分類する研究が行われている。多くの場合、カテゴリは予め設定されており、容易に変更できない。そのため、ニュースやブログといった複数のカテゴリに分類されるデータ群にアクセスが集中すると、該当するすべてのカテゴリにクエリを伝播させる必要があり、検索効率が低下する。そこで本稿では、カテゴリを動的に生成する検索手法を提案する。提案手法では、予め複数のカテゴリ（静的カテゴリ）が設定された P2P ネットワークにおいて、複数の静的カテゴリに分類されている話題のデータ群へのアクセスが増加した場合に、それらのデータからなる新たなカテゴリ（動的カテゴリ）を生成する。これにより、検索時に発生するメッセージ数を削減する。

キーワード P2P, 検索, カテゴリ生成, アクセス頻度

A Search Method Considering Dynamic Category Construction in Peer-to-Peer Networks

Yukie KOBAYASHI[†], Toshiki WATANABE^{††}, Akimitsu KANZAKI^{††}, Tomoki YOSHIHISA^{††},
Takahiro HARA^{††}, and Shojiro NISHIO^{††}

[†] Division of Electronic, Information and Energy Eng., School of Eng., Osaka University

^{††} Dept. of Multimedia Eng., Graduate School of Information Science and Technology, Osaka University

E-mail: [†]kobayashi.yukie@ise.eng.osaka-u.ac.jp,

^{††}{watanabe.toshiki,kanzaki,yoshihisa,hara,nishio}@ist.osaka-u.ac.jp

Abstract In order to improve the efficiency of information retrieval in P2P networks, there has been many researches on categorizing data items based on their similarities. However, since these researches adopt predetermined categories, it is difficult to change the categorization policy. Therefore, when the access frequency for particular data items that are classified into multiple categories becomes large, the efficiency of data search may be deteriorated since the query has to be propagated to all relevant categories. In this paper, we propose a search method that can dynamically construct categories. Our method searches data items base on predetermined categories (static categories). In addition, our method constructs new categories (dynamic categories) according to the increase of access frequency for data items. This approach can reduce the number of query messages exchanged for searching data.

Key words P2P, search, category generation, accessibility

1. はじめに

近年、計算機の普及や高性能化に伴い、P2P ネットワークが注目されている。P2P ネットワークでは各端末（ピア）が互い

にサービスを提供し合うため、スケーラビリティや耐障害性に優れている。一方、各ピアがどのようなデータを所持しているかを把握することが難しく、検索の際には、フラッディングなどの無作為な検索方法が用いられることが一般的であり、検索

に必要なメッセージ数が増大してしまう[4]．

検索効率を向上させる方法として、類似したデータ群をカテゴリとよぶ集合として管理し、同一カテゴリ内のデータを所持する複数のピアからなるネットワークを構成する研究が行われている[1]～[3],[5]．これらの手法では、同一のカテゴリに分類された類似した内容のデータを所持するピア同士が少ないホップ数で接続されるようネットワークを構築する．検索の際には、目的のデータを所持している可能性が高いピアに優先的にクエリを転送することで、少ないメッセージ数で必要なデータを獲得できる．しかし、これらの研究で想定しているカテゴリは、ネットワーク全体の共通な情報として予め静的に設定されていることが多く、動的に変更できない．そのため、類似した内容にも関わらず、異なるカテゴリに分類されているデータ群へのアクセスが集中した場合、該当するすべてのカテゴリに頻繁にクエリを送る必要がある．たとえば、ニュースやブログにおいて、北京五輪に関する記事は一時的にアクセスが集中する可能性がある．北京五輪に関する記事は、スポーツのカテゴリだけでなく、政治や経済のカテゴリにも分類される可能性があるため、検索時には該当するすべてのカテゴリにクエリを送る必要があり、検索時の負荷が増大する．

そこで、本稿では、予め複数のカテゴリ（静的カテゴリ）が設定され、カテゴリごとにネットワークが構築された P2P ネットワークにおいて、動的なカテゴリ生成を考慮した検索手法を提案する．提案手法では、複数のカテゴリに分類されるデータ群に対するアクセスの増加を検知し、それらのデータ群に適する新たなカテゴリ（動的情カテゴリ）を生成する．その後、新たに発生したクエリが動的情カテゴリを生成する要因となったクエリ群を類似している場合、その動的情カテゴリにのみクエリを伝播させることで、検索に必要なメッセージ数を削減する．

以下では、2 章で関連研究について述べ、3 章で本研究における想定環境について説明する．4 章で提案手法について説明し、5 章でシミュレーション実験による性能評価を述べる．最後に 6 章で本稿のまとめと今後の課題について述べる．

2. 関連研究

P2P ネットワークにおける検索効率を向上させるために、類似したデータをもつピア同士を少ないホップ数で接続する方法が数多く研究されている．例えば文献[3]では、データを意味的に配置するベクトル空間を設定して等分割することでデータの分類法を定義し、これらの分類法に基づいてデータをカテゴリに分類し、同一のカテゴリに分類されたデータをもつピア同士を少ないホップ数で接続する手法が提案されている．また、文献[1]では、データの意味的な内容やデータが作成された年代を考慮して定義されたデータの分類法に基づいてデータをカテゴリに分類し、同一のカテゴリに分類されたデータをもつピア同士を少ないホップ数で接続する手法が提案されている．これらの手法では、検索時に検索対象であるカテゴリ内にのみクエリを転送すればよいため、少ないメッセージ数で適切なデータを多く獲得できる．このように、検索効率を向上させるために、データをカテゴリなどに分類することは有効である．しかし、

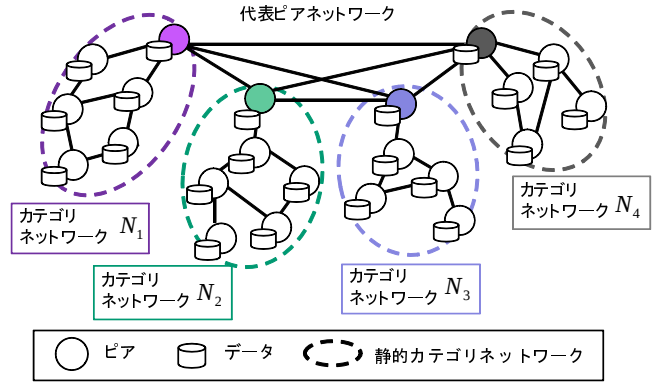


図1 ネットワークの構成例

Fig. 1 An example of our assumed network.

従来手法でデータをカテゴリに分類するためには、予めデータの分類法を定義しておく必要がある．そのため、新たに発生した話題のデータ群が、予め定義された分類法に従って適切に分類できず、複数のカテゴリに配置されてしまう場合には、クエリの転送先が適切に判断できず、検索効率が低下してしまう．一方、提案手法では、新たなデータの発生やアクセス頻度の上昇に応じて動的にカテゴリを生成するため、高い検索効率を維持できる．

3. 想定環境

本章では、本研究における想定環境について説明する．

3.1 ネットワーク構成

各データにはそれぞれデータの特徴を表すベクトル（特徴ベクトル）が付与されているものとする．この特徴ベクトルは、データに含まれる単語の出現頻度などを基に算出される．また、予めネットワーク全体の共通の情報として一定の規則に従って類似した内容をもつデータ群をひとつの集合として表す n 個の静的カテゴリ ($C_1, \dots, C_n$) が設定されており、各データは類似度の最も高い静的カテゴリに分類される．同一の静的カテゴリに分類されるデータを所持しているピアで構築されるネットワークを静的カテゴリネットワーク ($N_1, \dots, N_n$) と呼ぶ．ネットワーク構成例を図1に示す．各静的カテゴリネットワークには代表ピアが存在し、代表ピア同士は互いに接続しているものとする．本研究では、代表ピアで構築されているネットワークを代表ピアネットワークと呼ぶ．

カテゴリの特徴を表すベクトルをカテゴリベクトルと呼ぶ．全ての静的カテゴリのカテゴリベクトルは各ピアにおいて既知とする．これは、ピアがネットワークに参加する際に既に参加しているピアからその情報を受信することで実現できる．一つのピアは複数のデータを所持しており、自身が所持しているデータが分類される全ての静的カテゴリの静的カテゴリネットワークに参加する．また、ピアは自身が所属している静的カテゴリネットワークの代表ピアの情報（IP アドレスなど）を把握しているものとする．

3.2 検索方法

検索の際には、ユーザが入力したキーワードから、そのキーワードの特徴を表すクエリベクトルが生成され、クエリとして送信される．このクエリベクトルは、シソーラス等を用いてそ

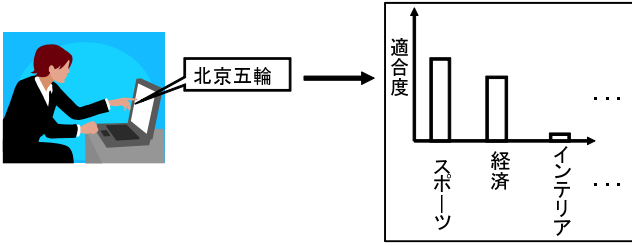


図 2 クエリベクトル

Fig. 2 Query vector.

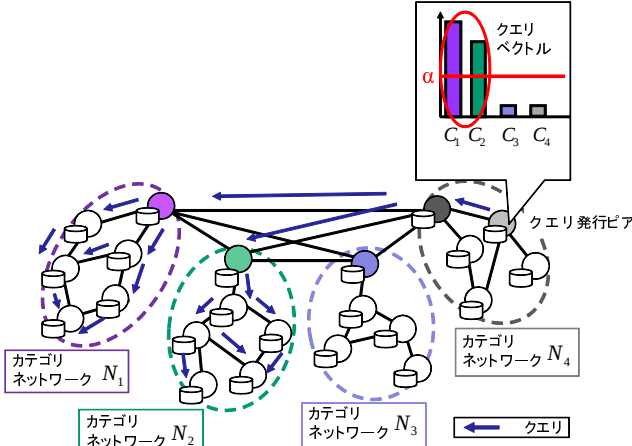


図 3 検索方法

Fig. 3 Search method.

のキーワードが各静的カテゴリとどの程度関連しているか（適合度）を算出することで生成される。例えば、図 2 のように、ユーザが「北京五輪」と入力した場合、関連が高い「スポーツ」や「経済」の静的カテゴリとの適合度が高く、それ以外の静的カテゴリとの適合度が低いクエリベクトルが生成される。

クエリを発行するピア（クエリ発行ピア）は、クエリベクトルと各静的カテゴリとの適合度が、式 (1) を用いて算出した α より大きくなる全ての静的カテゴリネットワークの代表ピアにクエリを転送し、それらの静的カテゴリネットワークにクエリを伝播させる。

$$\alpha = \frac{\sum_{i=1}^n q_i}{n}. \quad (1)$$

式 (1) において、 n は全静的カテゴリ数、 q_i はクエリベクトルのカテゴリ C_i との適合度を表す。クエリ発行ピアは、最初に自身が所属する静的カテゴリネットワークの代表ピアにクエリを送信する。その際、クエリ発行ピアが複数の静的カテゴリネットワークに所属する場合には、所属する静的カテゴリネットワークの代表ピアから一つをランダムに選択する。クエリを受信した代表ピアは、クエリの転送先に指定された静的カテゴリネットワークの代表ピアにクエリを転送する。クエリを受信した各静的カテゴリネットワークの代表ピアは、自身の受け持つ静的カテゴリネットワーク内にクエリを伝播させる。検索時の動作例を図 3 に示す。図 3 において、クエリ発行ピアは、静的カテゴリ C_1 および C_2 との適合度が高いクエリを発行している。そのクエリは静的カテゴリネットワーク N_1, N_2 の代表ピアに転送され、その後、そのカテゴリネットワーク内の各ピアに伝播される。

クエリグループベクトル	受信クエリ数	関連カテゴリ	タイムスタンプ
クエリグループ1 	20	C_1, C_2	13
クエリグループ2 	5	C_2, C_3	16

図 4 クエリグループ情報

Fig. 4 Query group information.

クエリを受信したピアは、自身が所持するデータの特徴ベクトルとクエリベクトルの類似度 s を式 (2) を用いて計算する。

$$s = \frac{\langle Q, D \rangle}{\|Q\| \times \|D\|}. \quad (2)$$

式 (2) において、 Q, D はそれぞれクエリベクトル、およびデータの特徴ベクトルを表す。 s が定数 β 以上のデータを持つピアは、そのデータをクエリ発行ピアに返送する。

4. 提案手法

提案手法では、複数のカテゴリに分類されるデータ群へのアクセスの増加を検出し、動的カテゴリを生成する。本章では、上記それぞれの手順、および動的カテゴリを利用した検索方法について説明する。

4.1 アクセス増加の検出

複数の静的カテゴリに分類されるデータ群へのアクセスの増加を検出するために、各静的カテゴリネットワークの代表ピアは、複数のカテゴリネットワークに転送されたクエリの情報を記録する。そのため、各クエリ発行ピアは、3.2 節で説明した検索方法においてクエリを送信する際に、複数の静的カテゴリネットワークへ転送する必要があるクエリには、その旨を示す情報（複数カテゴリ転送フラグ）を付加する。

各静的カテゴリネットワークの代表ピアは、受信した複数カテゴリ転送フラグが付加されたクエリのうち、類似したクエリベクトルをもつものを図 4 のように一つのグループと見なし、その情報を保持する。このグループをクエリグループと呼ぶ。各クエリグループの情報には、グループの特徴を表すための各静的カテゴリとの適合度から構成されたベクトル（クエリグループベクトル）、受信クエリ数、グループに属するクエリが転送された静的カテゴリネットワークの静的カテゴリ（関連カテゴリ）、クエリグループの情報が最後に更新された時刻（タイムスタンプ）を含む。例えば、図 4 のクエリグループ 1 は、過去に類似したクエリを 20 回受け取り、この 20 回のクエリのうち一つ以上を静的カテゴリ C_1, C_2 の静的カテゴリネットワークに転送したことがあることを表している。また、このクエリグループ 1 は、時刻 13 において、最も最近に類似したクエリを受信し、クエリグループの情報を更新している。

4.1.1 クエリグループの作成と更新

複数カテゴリ転送フラグが付加されたクエリを受信した代表ピアは、自身の保持する各クエリグループのクエリグループベクトルとクエリベクトルとの類似度を式 (3) を用いて計算する。

$$sg_j = \frac{\langle Q, G_j \rangle}{\|Q\| \times \|G_j\|}. \quad (3)$$

式 (3) において Q はクエリベクトル, G_j はクエリグループ j のクエリグループベクトルを表す. G_j の各要素 $g_{j,k} (k = 1, \dots, n)$ はクエリグループ j と静的カテゴリ C_k との適合度を表す.

sg_j がシステムが設定した定数 γ 以上の場合, そのクエリグループベクトルとクエリベクトルは, 類似した内容のデータを表していると見なす. その後, 代表ピアは, 以下の手続きに従って自身のもつクエリグループの情報を更新する.

- ・類似度 sg が γ 以上のクエリグループが一つ存在する場合
クエリを受信した代表ピアは, 受信したクエリと類似したデータ群を表すクエリグループがすでに一つ存在していると判断し, 受信したクエリの情報をそのクエリグループに加える. 具体的には, 受信クエリ数の値を 1 増加させ, タイムスタンプを現在の時刻に更新する. また, 新たな関連カテゴリが存在する場合には, そのカテゴリの情報を追加する. さらに, クエリグループベクトル G の各要素 $g_{j,k}$ を式 (4) を用いて更新する.

$$g_{j,k} = \frac{g_{j,k}^* \times r_j + q_k}{r_j + 1}. \quad (k = 1, \dots, n) \quad (4)$$

式 (4) において $g_{j,k}^*$ は更新前の $g_{j,k}$ の値である. q_k は Q の静的カテゴリ C_k との適合度, r_j はクエリグループ j の受信クエリ数を表している.

- ・類似度 sg が γ 以上のクエリグループが二つ以上存在する場合
類似度が γ 以上のクエリグループ同士は互いに類似したデータ群に対するクエリを表している可能性が高い. そこで, クエリを受信した代表ピアは, これらのクエリグループ同士の類似度を sgg を式 (5) を用いて計算する.

$$sgg = \frac{\langle G_j, G_h \rangle}{\|G_j\| \times \|G_h\|}. \quad (5)$$

式 (5) において G_j, G_h はそれぞれクエリグループ j, h のクエリグループベクトルを表す. 代表ピアは, この値が γ 以上となる場合, クエリグループを一つに統合する. 統合後のクエリグループのクエリグループベクトル G' の各要素 $g_{j,k}$ は, 式 (6) を用いて設定する.

$$g'_{j,k} = \frac{g_{j,k} \times r_j + g_{h,k} \times r_h}{r_j + r_h}. \quad (k = 1, \dots, n) \quad (6)$$

なお, 統合後のクエリグループの受信クエリ数は, クエリグループ j, h それぞれの受信クエリ数の加算値, 関連カテゴリは, j, h の関連カテゴリの和集合, タイムスタンプは現在の時刻に設定する.

クエリグループを統合した後, 再び受信したクエリベクトルと統合したクエリグループベクトルとの類似度を計算し, 類似度 sg が γ 以上であれば, 受信したクエリの情報をそのクエリグループに加えて更新する.

- ・類似度 sg が γ 以上のクエリグループが一つも存在しない場合
クエリを受信した代表ピアは, 受信したクエリと類似したデータ群に対するクエリグループが存在しないと判断し, 受信したクエリの情報を基に新たなクエリグループを作成する. 具体的には, 受信したクエリのクエリベクトルをクエリグループベクトルとし, 受信クエリ数を 1, タイムスタンプを現在の時刻にする. また, 関連カテゴリは, そのクエリが転送された静的カテゴリネットワークの静的カテゴリとする.

カテゴリ生成メッセージの情報

クエリグループベクトル	関連カテゴリ
	C_1, C_2

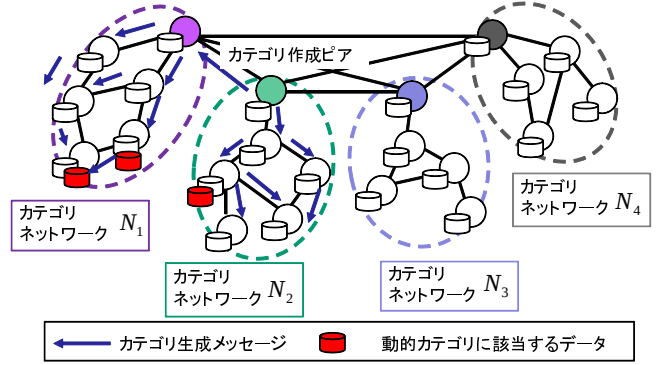


図 5 動的カテゴリに含まれるデータの検索

Fig. 5 Searching data included in dynamic category.

4.1.2 クエリグループの削除

一定期間類似したクエリを受け取らず, 更新されなかったクエリグループは, アクセス数が少なく, 動的カテゴリを生成する必要がないと判断し, 代表ピアはその情報を削除する. 具体的には, 最後に更新された時刻であるタイムスタンプから一定時間 T 以上経過しているクエリグループが存在する場合, そのクエリグループの情報を削除する.

4.2 動的カテゴリの生成

各静的カテゴリネットワークの代表ピアは, 一つのクエリグループに含まれる受信クエリ数が一定数 M に達した場合, そのクエリグループへのアクセスが増加したと判断し, 動的カテゴリを生成する.

動的カテゴリを作成することを決定した代表ピア (カテゴリ作成ピア) は, 図 5 のように動的カテゴリを作成するクエリグループのクエリグループベクトルを付加したメッセージ (カテゴリ生成メッセージ) を, 自身の静的カテゴリネットワーク内に伝播させる. また同時に, そのクエリグループの関連カテゴリの静的カテゴリネットワークの代表ピアにもカテゴリ生成メッセージを送信し, その静的カテゴリネットワーク内にも伝播させる. カテゴリ生成メッセージを受信したピアは, 通常の検索と同様, 式 (2) を用いて, カテゴリ生成メッセージに付加されたベクトルと自身が所持するデータの特徴ベクトルとの類似度を計算する. 類似度が定数 δ 以上であれば, そのデータは動的カテゴリに含まれると判断し, ピアは自身の ID をカテゴリ作成ピアに返送する.

カテゴリ作成ピアは, 動的カテゴリに参加するピアの ID を収集し, それぞれの隣接関係をランダムに設定する. その際, 動的カテゴリに参加するピアから代表ピアを選出し, 全ての関連カテゴリの静的カテゴリネットワークの代表ピアと接続させる. その後, 設定したピアの隣接関係に関する情報を再度クエリグループ内の全ての関連カテゴリの静的カテゴリネットワークに伝播させる. この情報を受け取った動的カテゴリに参加するピアは, 図 4.2 のように, その隣接関係の情報に基づいて,

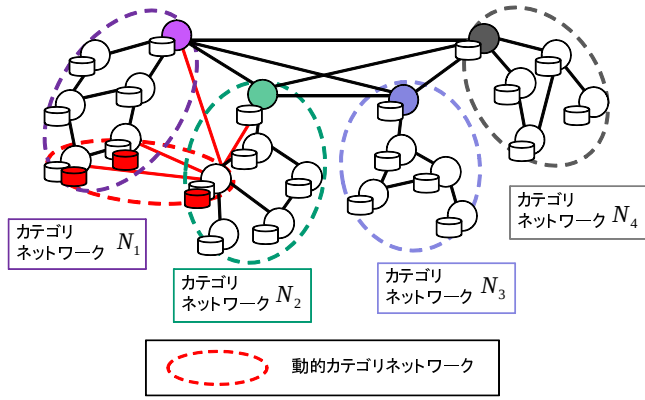


図 6 動的カテゴリネットワークの構築

Fig. 6 Construction of a dynamic category network.

動的カテゴリネットワークを構築する。

動的カテゴリの情報は、全ての関連カテゴリの静的カテゴリネットワークの代表ピアが保持する。この動的カテゴリの情報は、動的カテゴリを生成したクエリグループのクエリグループベクトル、および関連カテゴリで構成される。

4.3 動的カテゴリを利用した検索

各静的カテゴリネットワークの代表ピアは、クエリを受信した際、そのクエリベクトルと自身が所持する動的カテゴリのクエリグループベクトルが類似している場合、その動的カテゴリの動的カテゴリネットワークにのみクエリを転送することで、検索時のメッセージを削減する。

具体的には、静的カテゴリネットワークの代表ピアは、複数カテゴリ転送フラグが付加されたクエリを受信した際に、自身が所持する動的カテゴリのクエリグループベクトルとクエリベクトルとの類似度を式 (3) を用いて計算する。この類似度が自身が代表する静的カテゴリベクトルとクエリベクトルとの適合度より高い場合は、対応する動的カテゴリネットワークの代表ピアにクエリを転送し、動的カテゴリネットワーク内に伝播させる。

5. 性能評価

本章では、提案手法の有効性を検証するために行ったシミュレーション実験の結果を示す。実験では、P2P ネットワークモデルにおけるデータ検索を想定し、静的カテゴリネットワークのみのネットワークにおける検索方法との性能比較を行った。

5.1 シミュレーション環境

シミュレーション実験では、静的カテゴリ数を 20、ネットワーク上に存在するデータ数を 7,500 個と設定し、そのうち 5,000 個のデータは、一つの静的カテゴリに対してのみ適合度が高いものとした。具体的には、各データの特徴ベクトルを、ランダムに選択した一つの静的カテゴリとの適合度を $[0,100]$ の範囲でランダムに設定し、それ以外のカテゴリとの適合度を 0 とした。また、残りの 2,500 個のデータは、複数の静的カテゴリとの適合度が高いデータとした。具体的には類似した特徴ベクトルをもつ 50 個のデータを一種類のデータ群とし、50 種類のデータ群を作成した。各データ群の特徴ベクトルは以下の手順で生成した。

表 1 パラメータ設定

Table 1 Parameter setting.

パラメータ	意味	値
β	クエリに該当する該当データの類似度の閾値	0.85
γ	クエリに該当するクエリグループベクトルの類似度の閾値	0.85
δ	動的クラスタに該当するデータの類似度の閾値	0.75
M	動的クラスタを生成するクエリ数の閾値 (個)	20
T	クエリグループを削除するまでのタイムスロット	70

(1) 適合度の高い静的カテゴリとして、2~5 個のカテゴリをランダムに選択し、それらのカテゴリとの適合度の値を $[30,80]$ の範囲でランダムに設定する。

(2) 手順 (1) で選択されたカテゴリの適合度の値に $[-20,20]$ の範囲のランダムな値を加えて特徴ベクトルを作成する。手順 (2) の操作を 50 回繰り返すことで、一つのデータ群の 50 個のデータの特徴ベクトルを作成する。

ネットワークには、500 個のピアが参加し、各ピアはそれぞれ 15 個のデータをもつものとした。各ピアは、各データが分類される静的カテゴリの静的カテゴリネットワークに参加する。ここで各静的カテゴリネットワーク内での隣接ピア数は Power-Law Random Graph [4] に従って決定するものとした。具体的には、隣接ピア数の降順でピアを並べ、 i 番目のピアの隣接ピア数 d_i を $\lfloor 15 \times i^{-0.28} \rfloor$ とした。

また、1 タイムスロットごとに、ランダムに選択された一つのピアがクエリを発行する。発行するクエリのクエリベクトルは、ランダムに選択した一つの静的カテゴリとの適合度を $[30,80]$ の範囲で設定するものとした。ただし、本実験では、複数の静的カテゴリに分類されるデータ群へのアクセスが一時的に集中する環境を実現するため、定期的 (250 タイムスロットごと) に複数の静的カテゴリに分類されるデータ群へのアクセスを増加させた。具体的には、上述した複数の静的カテゴリに分類される 50 種のデータ群の内の一つが頻繁にアクセスされる期間を、各データ群ごとにそれぞれ 150 タイムスロットずつ設けた。複数の静的カテゴリに分類されるデータ群へのアクセスが増加している期間中に発行するクエリは、上述したデータの作成手順 (1) で作成した特徴ベクトルの各静的カテゴリに対して、 $[-20,20]$ の範囲のランダムな値を加えることで作成した。その他の閾値に関する各種のパラメータは表 1 に示すように設定した。

また、実際に返送されたデータの中で、検索要求に該当するデータの割合を表す適合率は、全ての検索において 1 となるように設定した。以上のシミュレーション環境において、15,000 タイムスロットのシミュレーション実験を行い、以下の項目を評価した。

- 総メッセージ数

検索時に発行されるメッセージ数、および、提案手法において動的カテゴリネットワークを生成する際に必要なメッセージ数の合計。ただし、シミュレーション開始時からの累計とする。

- 再現率

各クエリにおけるネットワーク上に存在する該当データのう

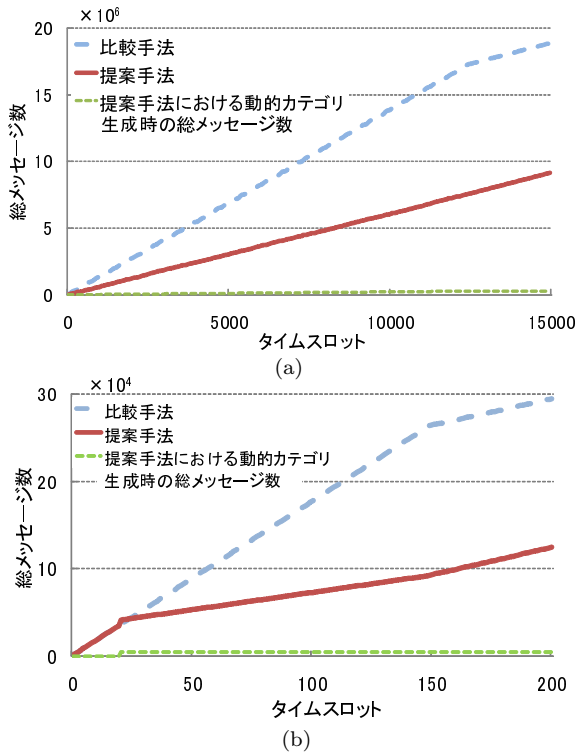


図 7 総メッセージ数

Fig. 7 The number of messages.

ち、実際に返送されたデータの割合。

5.2 評価結果

5.2.1 総メッセージ数

評価結果を図 7(a), 図 7(b) に示す。図 7(b) は、可視性を高めるため、図 7(a) においてタイムスロットが 1~200 の部分を抜粋している。図の横軸はタイムスロットを表し、縦軸は総メッセージ数を表す。

結果より、提案手法では、比較手法と比較して、総メッセージ数を大幅に削減できることがわかる。これは、比較手法では、複数の静的カテゴリとの適合度が高いデータ群へのクエリは適合度の高いすべての静的カテゴリネットワークに伝播される一方で、提案手法では、それらのデータ群を所持するピアからなる動的カテゴリネットワークを構築することで、必要なピアにのみクエリを伝播できるためである。例えば本実験では、タイムスロットが 0 から 150 の間、複数の静的カテゴリに分類されるデータ群へのアクセスが集中しており、比較手法と比べて提案手法におけるメッセージ数の増加量が抑えられている。また、提案手法では、動的カテゴリネットワークの生成にメッセージが発生しているが、この量は極めて小さいことがわかる。

5.2.2 再現率

評価結果を図 8 に示す。図の横軸はタイムスロットを表し、縦軸は再現率を表す。結果より、提案手法では、必要なデータの大部分を獲得できることがわかる。しかし一方で、再現率が 1 にならない検索が存在する。提案手法では、類似したクエリを一つのクエリグループにまとめる際に、各クエリのクエリベクトルを平均化してクエリグループベクトルを算出している。この場合、クエリグループに含まれる個々のクエリのクエリベクトルとそのクエリグループベクトルに違いが生じるため、個々のクエリに対して該当する全てのデータがその動的カテゴ

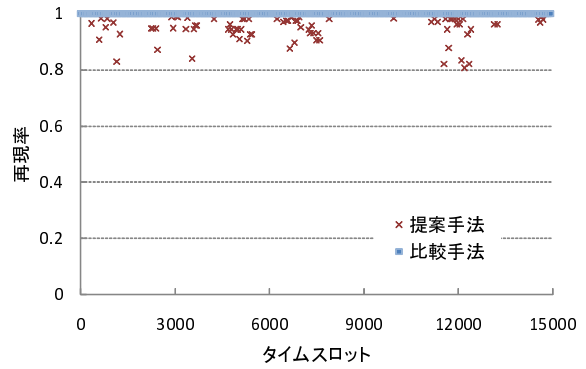


図 8 再現率

Fig. 8 Recall

リネットワーク内に配置されるとは限らない。そのため、動的カテゴリネットワークにのみクエリをフラッディングさせても必要なデータを獲得できない場合があり、再現率が低下している。再現率を完全に 1 にするためには、動的カテゴリに該当するデータの類似度の閾値 δ を低く設定し、動的カテゴリに該当するデータの条件を緩くする必要がある。しかし、 δ を低く設定すると動的カテゴリに該当するデータが多くなり、動的カテゴリネットワークの規模が大きくなるため、検索に必要なメッセージ数が増加してしまう。従って、検索の質と検索に必要なメッセージ数の関係を考慮し、 δ の値を適切に設定する必要がある。

6. おわりに

本稿では、予め複数のカテゴリが設定され、カテゴリごとにネットワークが構築された P2P ネットワークにおいて動的なカテゴリの生成を考慮した検索手法を提案した。提案手法では、複数のカテゴリに分類されるデータ群に対するアクセス数の増加を検知し、それらのデータ群に適する新たなカテゴリを生成することにより、検索に必要なメッセージ数を削減する。シミュレーション実験により、提案手法によって複数のカテゴリネットワークに配置されるデータ群に対するアクセス数が増加した場合でも、少ない検索メッセージ数で必要なデータを取得できていることを確認した。

今後は、動的カテゴリネットワークの維持にかかる負荷を考慮し、使用頻度が少なくなった動的カテゴリは削除する。

謝辞 本研究の一部は、特定領域研究 (18049050) の研究助成によるものである。ここに記して謝意を表す。

文 献

- [1] A. Crespo and H. Garcia-Molina, "Semantic overlay networks for P2P systems." Technical report, Computer Science Department, Stanford University, pp.23-32, 2002.
- [2] H. Jin, X. Ning, H. Chen, and Z. Yin, "Efficient query routing for information retrieval in semantic overlays." Proc. SAC 2006, pp. 1669-1673, Apr. 2006.
- [3] M. Li, W. C. Lee, and A. Sivasubramaniam, "Semantic small world: An overlay network for peer-to-peer search." Proc. ICNP 2004, pp. 228-238, Oct. 2004.
- [4] Q. Lv, P. Cao, E. Cohen, K. Li, and S. Shenker, "Search and replication in unstructured peer-to-peer networks." Proc. ICS 2002, pp. 84-95, June 2002.
- [5] C. Tang, Z. Xu, and S. Dwarkadas, "Peer-to-peer information retrieval using self-organizing semantic overlay networks." Proc. SIGCOMM 2003, pp. 175-186, Aug. 2003.