

利用者の探索傾向を反映した P2P ネットワークの再構成と視覚化

羽多野 一磨[†] 田中 浩也^{††} 角谷 和俊^{††} 田中 克己^{††}

[†] 京都大学工学部情報学科 〒 606-8501 京都府京都市左京区吉田本町

^{††} 京都大学大学院情報学研究科 〒 606-8501 京都府京都市左京区吉田本町

E-mail: †{hatano,hirotanaka,sumiya,ktanaka}@dl.kuis.kyoto-u.ac.jp

あらまし 近年, Web ページは急速な勢いで増加しており, ユーザが手にすることの出来る情報量は非常に多くなった. しかし一方, その膨大な情報群から有用な情報を見つけ出す事が難しくなっている. このような中, ユーザの有用なページ発見を支援するシステムが望まれている. そこで本研究では, ユーザの Web 探索履歴を各々のユーザ間で共有し, それを基に P2P ネットワークを再構成することによって, 他のピアの Web 探索履歴を用いてユーザの求める情報への探索を行う際に参考に出来る探索履歴を提示する方式について提案する.

キーワード 情報探索支援, 協調, 視覚化, P2P ネットワーク

Reconstruction and Visualization Reflecting a User's Web Search Tendencies

Kazuma HATANO[†], Hiroya TANAKA^{††}, Kazutosi SUMIYA^{††}, and Katsumi TANAKA^{††}

[†] School of Informatics, Kyoto University Yosida-Honmati, Sakyo-ku, Kyoto, 606-8501 Japan

^{††} Graduate School of Informatics, Kyoto University Yosidahonmati, Sakyou-ku, Kyoto, 606-8501 Japan

E-mail: †{hatano,hirotanaka,sumiya,ktanaka}@dl.kuis.kyoto-u.ac.jp

Abstract Recently, amount of Web pages is increasing with rapid progress, and it becomes possible to get a lot of information. However, it is difficult to find out useful information from huge information. Under these circumstances, many people need support systems to discovery useful pages. In this paper, by sharing users' Web search history among each users, and reconstructing P2P network based on it, in case user searches for the required information, we propose a system which presents reference information on other peers, to refine their queries.

Key words information search support, cooperation, visualization, P2P network

1. 緒 論

近年のインターネットの発展には目を見張るものがあり, Web ページの数も年々膨大に膨れ上がっている. そこには非常に多くの情報が存在していると考えられるが, 一方ユーザが意図した情報を得ることが難しくなっている.

探索の多くはキーワード検索を基に行われている. しかし, 目的の情報を得るためのキーワードを正しく設定する必要があり, 目標が漠然としている場合やキーワードが思いつかない場合, 様々なキーワードを入力し試行錯誤する必要があった. また, キーワードを入れた後, どのようなページを見ればよいかを検索結果に表示された情報を読んで, 判断する必要がある.

そこで本研究ではユーザが目的を持って探索を行う際に, 目的ページの発見を容易にすることを目標としている. 多くのユーザが閲覧を行っているため, 類似した意図を持って探索を

行っているユーザが存在するはずである. これらのユーザらの探索履歴を利用して探索を支援する事を本研究の目的としている. 探索を支援するシステムは過去にも考案されているが, 従来のサーバ型では個人情報の扱いの問題から, 動的に構成される Web ページに関する情報を利用したりというのも困難であった. 更にプロキシサーバの様に全ての履歴をサーバに残すのは不安があった.

一方, 近年 Napster [4] や Gnutella [5] といった P2P ネットワークを利用したアプリケーションが注目されて久しい. P2P ネットワークは従来のサーバ・クライアント型のネットワークポロジとは違い, PC 同士がそれぞれに情報を提供・享受を行う. この様な点が匿名性に結びつき, 先の様な違法なファイル交換を行うアプリケーションが注目を浴びた.

しかし, このようなピア同士の直接の情報のやり取りというのは実際の人同士の情報のやり取りを考えると自然な形であ

る．現実世界においてもテレビなどの一方的に放送するようなメディアの情報量が多いものの人とのやり取りで得る情報量もまた非常に多い事は明らかである．この様な点からユーザ間の情報交換が重要である．

このようなことから本研究では P2P ネットワークのサーバレスでのユーザ間での直接的な情報のやり取りを利用した，探索履歴を共有・再利用するシステムを提案する．

2. 基本的事項と関連研究

2.1 基本的事項

2.1.1 P2P ネットワーク

近年脚光を浴びている技術に P2P(Peer-to-Peer) ネットワークがある．Peer は「対等の」という意味であり，従来のサーバクライアントモデルではサービスの提供をサーバのみが行っているものであるが，これと異なり全ての PC が対等な関係にあるネットワークである．

P2P ネットワークには Napster 等のように中央に補助的なサーバを用意するハイブリッド型のものと，Gnutella のようにそういったものを用意しないピア型のものが存在する．ピア型 P2P ネットワークを除いて，部分的にしるサーバがデータを一元的に管理する．そのため本研究で用いるのはピア型の P2P ネットワークモデルである．

ピア型の P2P ネットワークでは隣接したピアとしか直接情報をやり取りできない．このため，より遠くに情報を伝達するために「ホップ」を行う．「ホップする」とは情報を受け取ったピアが更にその隣接ピアへと送ることであり，あるデータが今まで通過してきたピアの数をホップ数と呼ぶ．また，データがいくつホップ可能な数を TTL と呼び，この数が 1 より大きければこのデータは更にホップする事が可能である

2.1.2 類似度

ユーザの特徴ベクトルを使ってユーザ間の類似度を測るためには特徴ベクトル間の類似度を定義する必要がある．特徴ベクトル間の相関値を求める手法は様々であるが，ここではベクトル間の余弦をそれとする手法を用いる．この手法では F, Q の類似度 $sim(x, y)$ は以下の様に計算される．

$$sim(F, Q) = \cos(F, Q) \quad (1)$$

$$= \frac{F \cdot Q}{\|F\| \|Q\|} \quad (2)$$

2.2 関連研究

原田ら [3] は検索システムに入れたキーワードと，その検索結果からユーザによって選択された URL の相関を求めることによってユーザの検索キーワードの決定を支援している．この研究は全てのユーザの検索結果を用いて関連語を抽出する方式であり，ユーザが現在何を目的としているかを考慮していない点本研究と異なる．

上田ら [6] は探索行為を共有し，視覚化する手法を提案している．事前にユーザのグループが作成されている事，グループが協調して意思決定を行うという目的を前提としている点で本研究と異なる．本研究では見知らぬ人との協調，個人の情報探索の支援を目標としている．

3. 基本概念

3.1 探索の定義

本研究での探索とは簡単には図 1 のように行われるものとする．ユーザが探索の目的を持つと，その目的からキーワードを決め検索エンジンで検索を行う．検索の結果表示されたページ一覧から目的に近いページを見つけると，それを閲覧しそこにある情報を吟味する．満足すれば探索行為はそこで終了するが，満足する情報を得られなかった場合，別のページを閲覧したり，キーワードを再検討したりすることによって，目的に向かう．

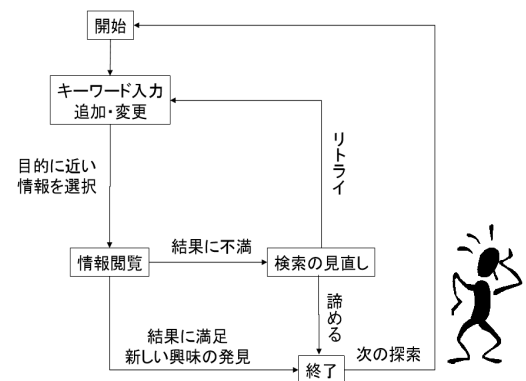


図 1 従来の探索行動

また，本論文で用いる幾つかの用語を以下のように定義する．

- 検索キーワード：ユーザが検索エンジンに投入したキーワード群
- 探索意図：ユーザが得ようとしている情報(群)
- 探索行為：ある目的を持ってユーザが検索し，その検索結果を閲覧すること
- 探索履歴：一連の探索行為の履歴

3.2 システム全体像

本研究のシステムは Web 上の探索を P2P ネットワークを用いて支援するものである．ユーザが Web ブラウジングを行っている情報をもとに P2P ネットワークを再構成していき，それまでの探索趣向が類似しているユーザを隣接ピアとしていく．これによって隣接したユーザの持っている探索履歴(あるいは現在の探索状況)を元にユーザの探索を支援する．図 2 はその様子を示している．本研究のシステムの利用により得られる利点は二点存在する．

- (1) 探索の効率化
- (2) 新しい事柄の発見

1 はある目的を決めて探索を行っている際に，うまく目的の情報にたどり着けないような場合の支援である．このような場合，試行錯誤を行っている最中にネットワークが再構成されて，似たような行為を行っているユーザと隣接することになる．このユーザの探索状況をユーザが参考にすることによって目的の達成を容易にする事が出来る．

例えば，筆者が自己組織化のサンプルソースがないのかを探

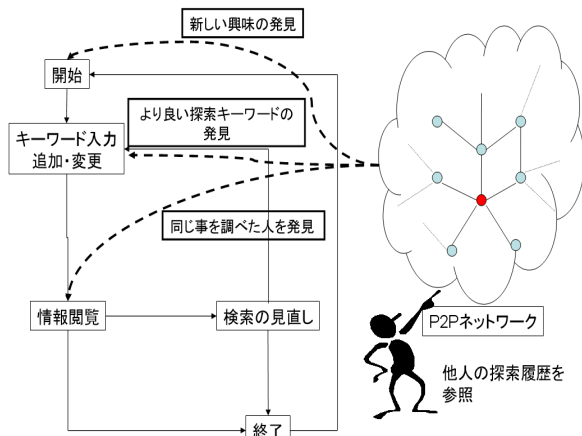


図 2 本研究の役割

探索したとき、「自己組織化 ソース」「SOM ソース」「自己組織化 プログラム」等をキーワードとして検索を行ってみたが、これを見つける事は出来なかった。このような場合に似たような探索を行っている人の様子を見て適切なキーワードを設定できる、もしくは目的の情報にたどり着けるのではないかと考えている。

2 は全体的な探索趣向が似ているユーザが集まることによって実現される。こういった一つのことに関して非常に多くの探索を行っているユーザは、似たようなユーザを集めることによって、よりコアな情報がその近接ユーザ間で伝達されることが考えられる。

例えば阪神ファンであるユーザは阪神に関する探索をよく行っていると考えられる。この時期では契約更改に関する情報やキャンプ情報などが情報として主に求められていると考えられるが、春の選抜高校野球に来年阪神に入団しそうな選手を探索している探索履歴を発見するかもしれない。このような場合、ユーザにとって新たな情報の発見となり、更にこの探索履歴を利用することにより、効率的に探索が可能であると考えられる。

本研究の手順は、以下の 3 つに分けられる。

- (1) 探索意図の抽出
- (2) P2P ネットワークの作成とその再構成
- (3) ユーザへの提示

(1) の探索目的の導出はユーザがどんな目的を持って探索を行っているかを推定する部分である。探索を行う際、ユーザは検索システムに目的を達成できると考えられるキーワード群を入力する。また、検索の結果出てきたページ群のうちから自分の思い描いていた目的に沿ったページだと思われるページを選択する。このような行為からユーザの目的を推定する部分である。

(2) では、(1) で得られた目的を元にユーザ間の類似度にしたがってネットワークを構成していく。この際の目的としては類似度の高いユーザ同士をつなぐ事である。これにより、ユーザはこのネットワーク上で簡単に自分の探索を支援する情報にアクセスすることが出来る。

(3) では、ユーザにアウトプットが提示される。ユーザは

これを利用することにより、有意であると思われるページとそこから得られる情報を簡単に知る事が出来る。

4. 探索履歴の抽出

4.1 探索履歴

ユーザが探索のために検索システムにキーワードを入力する際、漠然と目的を持っていると考えられる。これを探索目的と呼ぶとする。ユーザの探索目的は流動的であるが本研究では簡単のため、検索キーワードを投入してから他の検索キーワードを投入するまでを一連の探索行為とする。

キーワードはユーザが目的を達成するために決定されるが、曖昧な目標を持っている場合にはこれを正しく設定する事は難しく、これがユーザの目的を端的に表しているとは限らない。

検索キーワードを投入してから、次の検索キーワードを投入するまでのスパンを一まとまりの探索行為であるとして、この際ユーザの探索行動により得られる情報に関して以下の様に定義する。

- キーワード Q_i :
ユーザが検索システムに投入したキーワード群
- $answer(Q_i)$:
 Q_i で検索した結果得られたページ群
- $browse(Q_i)$:
 $answer(Q_i)$ の内、ユーザが閲覧したページ群
- $save(Q_i)$:
 $browse(Q_i)$ の内、ユーザが有用だと判断したページ群更にページ群の平均特徴ベクトルを抽出する関数として $feature()$ を定義する。ユーザ意図を得るために以下のように定義する ($TF(x)$ は x の TF 値を抽出する関数)

$$fea(Q_i) = Q_i \quad (3)$$

$$fea(answer(Q_i)) = TF(answer(Q_i)) - Q_i \quad (4)$$

$$fea(browse(Q_i)) = TF(browse(Q_i)) - fea(Q_i) - fea(answer(Q_i)) \quad (5)$$

$$fea(save(Q_i)) = TF(save(Q_i)) - fea(Q_i) - fea(answer(Q_i)) - fea(browse(Q_i)) \quad (6)$$

具体的にページの特徴語を抽出する手法について述べる。ユーザが閲覧しているページのどの単語に興味を持っているかを抽出する研究は多く行われている。最も基本的な手法は TF 法である。TF 法ではページの頻出語を特徴語としている。頻出語との共起の偏りに注目して抽出する CF 法 [1] 等も挙げられる。更にこれを発展させユーザの閲覧ページの頻出語との共起を計算することによって興味語を抽出する IRM 法 [2] も提案されている。

本研究では探索傾向の抽出ということで検索キーワードという最も端的にユーザの興味を表す語を得ることが出来る。IRM 法ではユーザの閲覧履歴から頻出語を計算し、これをユーザの身近語としていた。これに対して本研究では検索キーワードをこの身近語として、これとの共起の偏りを計算することによ

てユーザが興味をもっている語を抽出する．

4.2 修正意図

通常，ユーザの探索行為は前後で相関があることが多いと考えられる．例えば，ユーザの設定したキーワードが不十分で絞り込みをかけるために新たに検索を行う場合や，逆に絞り込みすぎてキーワードを緩めたりする場合，また，検索結果を見て何かを思いつきそれを検索する場合などである．そこで以下のような意図関連度 rel_i を導入し，キーワードの修正意図を推定する．

$$rel_i = \frac{sim(Q_i + F(browse(Q_i)), Q_{i+1})}{\sqrt{|Q_i + F(browse(Q_i))| |Q_{i+1}|}} \quad (7)$$

$(F() = feature())$

rel_i の値によって，ユーザの修正意図を以下の三つに分類する．この値が高ければ高いほど前後の探索意図が関連しているため，ユーザの意図としてはこの rel_i の高い順に以下の様に推測できる．

- 連続：同目的を持っている
- 連想：目的間に関連がある
- 新規：まったく関連性がない

「連続」は同じ（あるいはほとんど同じ）目的をユーザが持っている状態である．目的が同じ場合はキーワード間の関連度は非常に高くなるのが通常であると考えられる．目的が同じである場合ユーザが新たにキーワードを設定する事は少ないためである．複数のキーワードの内，いくつかを増減したり，類似語に置き換えたりすることで行うのが通常であるためである．

連続探索の場合，前後の探索行為は一連の行為とみなす事が出来る．例えば探索行為 $i \sim j$ が一連の行為である場合，一連の探索意図 $seq_intention_{i,j}$ は以下のように表せる．

$$seq_intention_{i,j} = (Q_i, \sum_{i \leq k \leq j} feature(browse(Q_k))) \quad (8)$$

また， $i \sim j$ の探索履歴を一まとめに扱う．

「連想」とはユーザが検索結果のページを閲覧することにより何らかの新たなアイデアを得た場合のことである．このため $feature(Q_i)$ と Q_{i+1} の関連性が高いはずであるため， rel_i はある程度高くなると考えられる．

「新規」はユーザの目的が一新される場合である．ユーザが何か新しいことを思い出したりする場合はこれにあたり，この場合 rel_i が低いと考えられる．

4.3 ユーザの特徴ベクトル

ユーザの特徴ベクトルは 4.1 節で述べた探索意図と 4.2 節で述べた修正意図から導かれる．特徴ベクトルを v ，探索意図を int ，関連度を rel とすると以下の様に導かれる．

$$\begin{aligned} v_{now} &= (Q, obj) \\ &= intention_{old} * rel + (Q_{now}, 0) \end{aligned} \quad (9)$$

次の探索行為が始まるときユーザの探索意図が加味され，

$$intention_{now} = intention_{old} * rel + intention_i \quad (10)$$

が計算される．

本研究では次章で示すようにユーザ間の類似度を計算する必要がある．ユーザの探索意図は (a, b) のベクトルのベクトルの形で表されている．よってこの様な場合は，

$$c = a + b \quad (11)$$

とする．つまりユーザ間の類似度 $sim(A, B)$ は以下の様にして計算される．

$$sim(A, B) = \cos(Q_A + obj_A, Q_B + obj_B) \quad (12)$$

5. P2P ネットワークの再構成

本研究ではユーザ間の情報共有のために P2P ネットワークを用いる．目的としては探索意図の近いユーザ同士を直接つなぐことによって情報の共有をよりスムーズにし，多くのデータを転送出来るようにすることである．

通常，ピアとはネットワークを構成するノードのような意味で使われるが，本研究ではピアは探索を行っているユーザに一对一で対応する．つまり本研究ではユーザ同士のネットワークを再構成するともいうことが出来る．

また P2P ネットワークを用いる理由として，Web の閲覧履歴のような個人的な情報はサーバ・クライアントモデルのような一元的に管理されるべきではなく，個人間のやり取りを主にするのが自然な形であるという点がある．これにより，ユーザの情報共有に対する敷居が下がり，ネットワーク全体としてはより情報量が上がると思われる．更に本研究の提案手法においては類似度の高いユーザ同士での情報共有が主に行われるため，ユーザにとってはより自然なやり取りであると考えられる．

P2P の特徴として人と人が直接情報をやり取りできるという点が挙げられる．P2P 上では一種の匿名性が得られる．また，情報を発信する側と情報を受け取る側が二分されるわけではないので，こういった情報を用いるのに適していると考えられる．

ピア型の P2P ネットワークである Gnutella を見るとネットワークはユーザ間のデータ転送を効率よく行うためにピア間の接続を決められていた．本研究ではユーザ間の興味の類似に注目して構成する．しかし，類似度の高いユーザ同士はファイル転送量も増加するため，適切なネットワークが構築されると考えている．

5.1 方法

再構成の手法について詳しく説明する．

まず，ピア（本研究では一つのピアは一人のユーザに対応する）は以下のような情報を持っている．

- 特徴ベクトル
- 隣接ピア（と，それとの類似度）
- 参照ピア
- 接続要求発行数
- 接続最大数

「特徴ベクトル」とは，ユーザの探索履歴を用いて抽出された

ユーザの探索の特徴を表すベクトルである。

「隣接ピア」はユーザが P2P ネットワーク上で実際に接続しているピアのことである。ユーザは自分の特徴ベクトルを更新する際、この隣接ピアへその情報を送るため、隣接ピアとの類似度は常に最新になっている事が保障される。

「参照ピア」はユーザが特徴ベクトルを送るために存在している。参照ピアとしているピアに特徴ベクトルを送る事はあるが、参照ピアとしているピアからそれ以外の情報が送られてくる事はない。また、参照ピアからこのための接続を要求された場合は必ずそれを受け入れなければならない。

「接続要求発行数」は「接続最大数を超えないために保持している情報である。P2P ネットワークではユーザがデータをホップさせるとい特徴があり、隣接ピアの数が増えると流れるデータ量は指数関数的に増大していく。このため、ユーザの隣接ピア数を一定以下にするために「接続最大数」を指定する。後述のアルゴリズム中には記述していないが、FIFO の原理で参照ピア数はこれを保つようになっている。

また「接続要求発行数」「参照ピア数と隣接ピア数の和」は接続最大数を超えないようにする。前者は隣接ピア数が接続最大数を超えないようにするために必要であり、後者は情報の送信量を一定以下に抑えるためである。

以下、アルゴリズムを説明する。

1. P2P ネットワークに初めて接続するユーザは適当な初期ノードを発見し、それを参照ピアとする。

2. ユーザは一定確率、または自分の特徴ベクトルが変化した際に隣接ユーザに $TTL = k$ を定めて特徴データを送る。図 3 はその様子を表している。

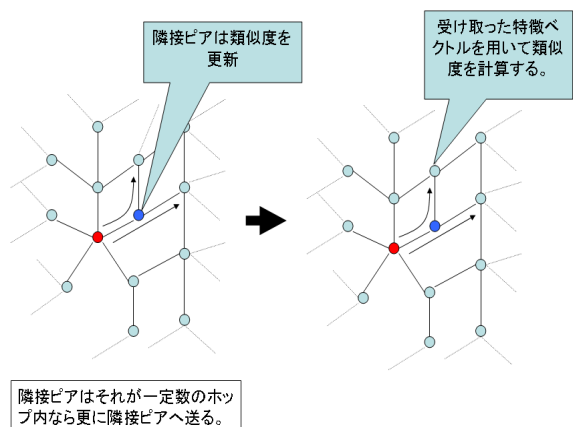


図 3 特徴データの分散

3. ユーザは以下の情報を受け取り、それぞれについて対応する。

- (1) 特徴データ:特徴データの発信元, 特徴ベクトル
- (2) 接続要求:接続要求の発信元とその特徴ベクトル
- (3) 切断要求:切断要求の発信元
- (4) 接続受理通達:接続受理通達の発信元
- (5) 接続拒否通達:接続拒否通達の発信元

1. 特徴データを受けた時

特徴ベクトルをうけとったユーザは隣接ユーザの類似度と比較する

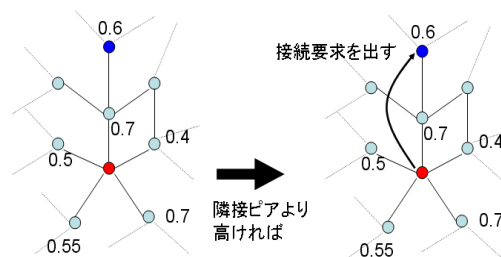


図 4 特徴データと接続要求

特徴ベクトルの送信元のピアが隣接ピアかどうかを判断する。図 4 は隣接ピアより類似度が高かった場合の処理を表している。

● 隣接ピアであった場合
隣接ピアとの類似度を更新する。

● 隣接ピアでなかった場合
そのピアとの類似度を計算し、隣接ピアの類似度と比較する。
- 接続数が接続最大数に満たない場合
接続要求を特徴データの送信元に送信する。

- 特徴データの特徴ベクトルとの類似度がいずれかの隣接ピアよりも高く、かつ接続要求発行数が接続最大数を超えていない場合

送信元のピアへ接続要求を出し、接続要求発行数を 1 増やす。

- 特徴データの特徴ベクトルとの類似度が全ての隣接ピアよりも低い場合

TTL の値によって隣接ピアへ転送する。TTL=0 の場合はそれ以上転送しない

2. 接続要求を受けた時接続要求を受けた場合、隣接ピアの類似度と比較して、それを受けようかどうかを判断する。図 5 はその様子を簡単に示している。

● 接続要求発行数が接続最大数である場合
接続拒否を接続要求元に送信する。

● 接続数が接続最大数に満たない場合
接続受理を接続要求元に送信する。

● 全ての隣接ピアより要求ピアの類似度が低い場合
接続拒否通達を接続要求元に送信する。

● いずれかのピアよりも高い場合
接続受理通達を接続要求元に送信する。また、これにより接続最大数を超えた場合、最も類似度の低いピアを参照ピアに加え、切断要求を送信する。

3. 切断要求を受けた時

切断要求元のピアが隣接ピアに含まれている場合はこれを参照ピアのリストに追加する。参照ピア数と隣接ピア数が接続最大数を超えていた場合参照ピアの内、もっとも以前に加えられたピアを削除する。

4. 接続受理通達を受けた時

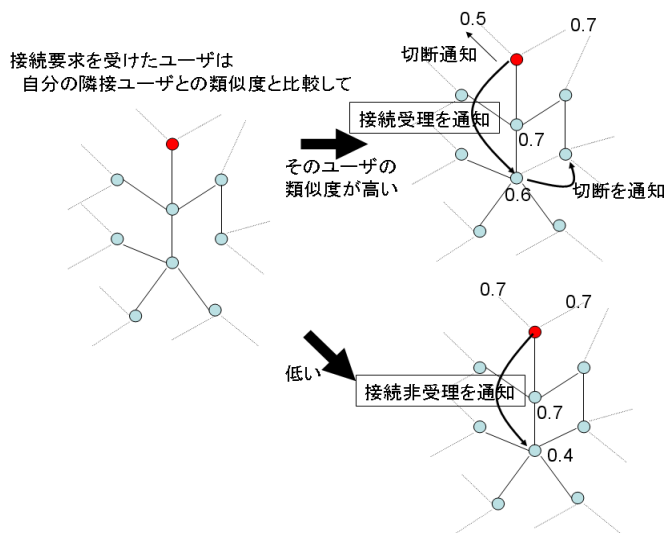


図 5 接続要求への応答

隣接ピアにこの通達元が含まれていなければこれを隣接ピアに追加し、接続要求発行数を 1 減じる。また、隣接ピア数が接続最大数を超える場合は最も類似度の低いピアを切断する。

5. 接続拒否通達を受けた時

接続要求発行数を 1 減じる。

5.2 アルゴリズムの停止性の証明

本節では先に示したアルゴリズムが、ピアの特徴ベクトルが一定である場合に停止することを証明する。まず見通しをよくするため、隣接ピアを切断して異なるピアへと接続を変更を試みるための条件は以下のとおりである。

- 隣接ピアより類似度の高いピアを発見した場合

この条件を満たすため、隣接ピアの類似度はより高くなる。

このため切断される事がない限り、隣接ピアとの類似度が高くなっていく事は保障されている。また、切断するための条件は以下のとおりである。

- 隣接ピアより類似度の高いピアとの接続が確立し、隣接数が接続最大数を超えた場合

つまり切断した側はより類似度が高くなっている。

ところで特徴ベクトルは変化しないため、ユーザ間の類似度は一定である。よって「より類似度の高いピアを発見する」事は切断されない限り、有限回（ユーザ数を K とした場合、最大 $K - 1$ 回）しかおこりえない。また、切断するための条件は「より類似度の高いピアを発見する」事であるため、切断されない限り有限回である。

ここでピア A の持つ隣接ピアとの類似度のうち、もっとも低いものを S_A とする。この時、類似度は数値であるため、ある一瞬において $S_1 \leq S_2 \leq \dots \leq S_N$ （但し、 N はユーザ数）と並べる事が可能である。このためこの時点でユーザ N が切断される事はない。

ここで次の二つの場合が考えられる。

- (1) S_N の隣接ピア数が接続最大数の場合

- (2) S_N の隣接ピア数が接続最大数未満の場合

1 の場合、このユーザ N が切断される事がないため、次の最大値 $S_{N_{next}}$ は

$$S_N \leq S_{N_{next}} \quad (13)$$

が成り立つ事が分かる。

2 の場合、ユーザ N は更に接続数を増やすことから、その接続先ピアとの類似度は S_N より小さい事がある。こういった場合、ユーザ N は他ユーザに切断されたという経緯が考えられる。つまりその際の最低類似度 $S_{N_{old}}$ はその時の最大値ではない事が分かる。つまりこの時、接続最大数を持つユーザのうち、最大の S を持つユーザを M とすると、 S_M より大きな類似度のユーザとは接続できない事が分かる。最低類似度の最大値の推移において、考えるのは接続最大数を持つユーザのみである事が分かる。

またユーザ間の類似度が一定であり、有限であることから最低類似度の最大値も有限であることは明らかである。つまりこの最大値 S_N は有限回の再構成のうちに一定の最大値に達する事が分かる。よってこの最大値を取るユーザは切断される事がなく、残りの $N - 1$ のユーザにおいても同じ事が言えるため、このアルゴリズムは有限回で停止する。

5.3 予備実験

このアルゴリズムを適用してシミュレータを作成し、このアルゴリズムによりどの程度、類似度の高いユーザを集められるかを確かめる。このシミュレータ上での処理は、二つのフェーズに分かれる。概要は以下の通りである。

- ユーザが行動
- ネットワーク上のパケットの移動

本シミュレータでユーザの行動は全て一まとめに行われる。それによる時間経過は再現されていない。また、ネットワーク上に流れたパケットは一定時間後に目的のピアへ到達する。この際、ネットワークのトラフィックによる到達の遅延なども起こらないとしている。

ユーザエージェントの特徴ベクトルとしては適当な次元用意し、そこからランダムに選択することにより生成した。この特徴ベクトルは一定確率で変化し、また、一定確率で隣接ピアへ特徴データを送信する。

この実験を行った環境を以下に示す。

- OS:WindowsXP
- CPU:Pentium4 3.06GHz
- Memory:1.50GB
- 開発環境:VisualStudio.net 2003/C #

5.4 予備実験の結果と考察

5.3 節で作成したシミュレータを用いて実験を行った。ユーザエージェントの数は 1000、最大接続数は 5、TTL=2、特徴ベクトルは 10 分に一度の確率で変化し、隣接ピアへ特徴データを送信するのは 30 秒に一度の確率、一ステップを 10ms として計算を行った。探索行為は前後に関連している可能性が高いため、特徴ベクトルは今までのものからランダムな個数をその

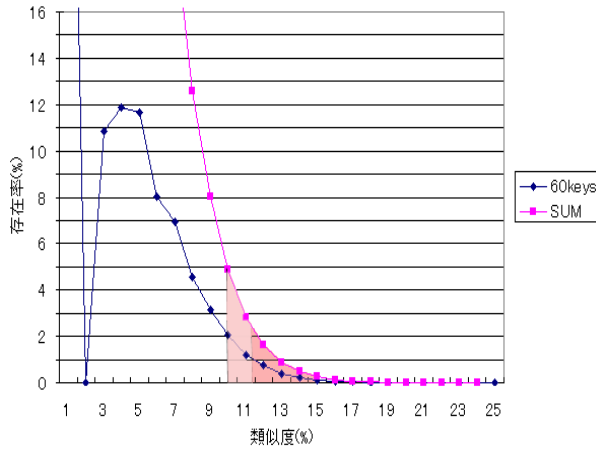


図 6 1 ユーザに対する類似度別の存在率

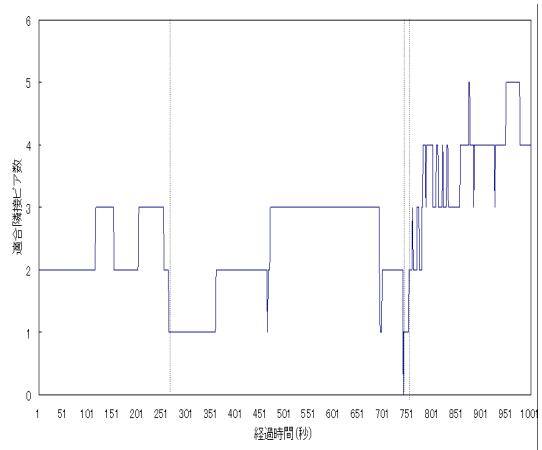


図 9 一つのピアの隣接適合ピア数

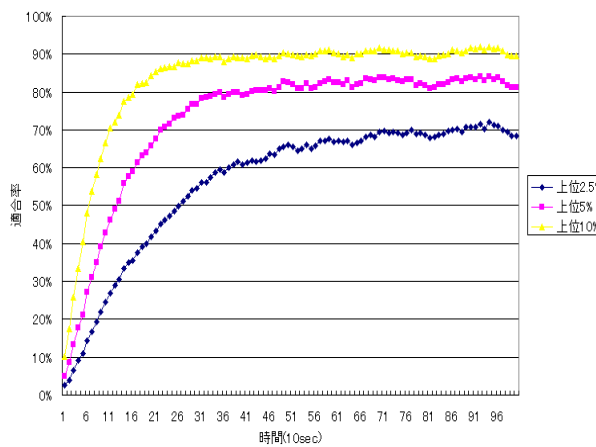


図 7 近接ピアの適合率

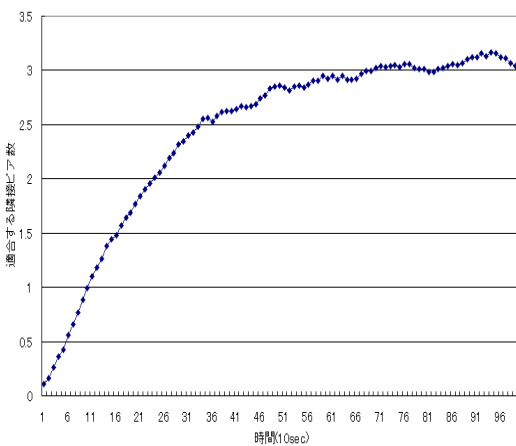


図 8 1 ユーザに対する類似度別の存在率

まま引継ぎ更に複数のベクトルを加えるようにしている．データの転送にかかる時間は 50ms として設定している．初期値としてはランダムに 4 つのピアと接続している状態を与えた．

図 6 は前節で示したようにランダムに特徴ベクトルを生成した時の類似度別の存在率をグラフ表示している．再構成がうまく働いているかの評価の基準として，あるユーザにとって類似度の上位 n % に含まれる確率を計算した．図 6 に引かれている

縦線は右からそれぞれ 2.5 %，5 % の境界線を示している．図 7 はこの確率の推移をグラフにしたものである．

また，図 8 はユーザの隣接するピアのうち上位 2.5 % に含まれる率である．これによりユーザにとってかなり類似度の高いユーザが隣接ピアに集まってきていることが分かる．このアルゴリズムではそれぞれのピアが局所的に最適にしていいため，最適解は保障されないが，自分の情報を全ピアへと流さず，局所的にだけ流すことによって，隣接ピアに類似度が高いピアを持つてくる事が出来ていることがわかる．

図 9 はあるピアに注目した時の，隣接ピアの内上位 2.5 % に含まれる数である．グラフ中縦に引かれた破線はユーザの特徴ベクトルが変化した事を表している．これによると有意だと思われるピアを発見するのにそれほど時間を要していないことが分かる．特に 750 秒前後では二回の変化があるが，すぐに有用な隣接ピアを発見するに至っている．ただし，250 秒の辺りを見ると分かるように，あまり特徴ベクトルが変化していない時もある．これが本当に有意かどうかは，実際の探索履歴を用いて実験する必要があると思われる．

更に加えてユーザの特徴ベクトルが変化した際，類似度が上位 2.5 % に含まれるような有効なピアをひとつ発見するまでの時間を測定した．グラフには示していないが，およそ 5 秒と非常に短い時間である事が分かった．最長でも 10 秒程度の結果であった．この事から十分実用に耐える程度の時間と考えられる．

ただし，本研究は探索行為を対象にしており，4.3 節で示したように探索行為に前後の関連があると考えられるため，本シミュレーションの結果と実際の結果は異なっていると考えられる．よって今後，こういった事も含め実験が必要だと思われる．

6. 視覚化

このように構成されたネットワーク上では探索意図の近いユーザのホップ数が近くなっていると考えられる．つまりこれによりユーザの探索を支援する情報がユーザから近いホップ数のピア上に配置されている．そこでこれらの情報を視覚化する事によりユーザの探索を支援する．

まず，ユーザ間の類似度を用いてピア間の距離を二次元上に

表す。これにより自分との類似度を合わせて、そのユーザの探索の様子を参照することが可能になる。図 10 はその概念図である。中央にあるウィンドウがユーザがブラウジングを行うウィンドウである。その周辺に配置されているのが、隣接ピアのウィンドウであり、その探索状況をリアルタイムに表示する。更に 2 ホップ離れたユーザの検索キーワードをその周辺に表示する。



図 10 ユーザ提示方法

更にそのピアの持っている探索履歴を視覚的に表示することによって、更にその視認性を高める。これを表しているのが図 11 である。探索履歴間は 4.2 節に示した修正意図によって、関係性が表されている。これを利用して表示する。

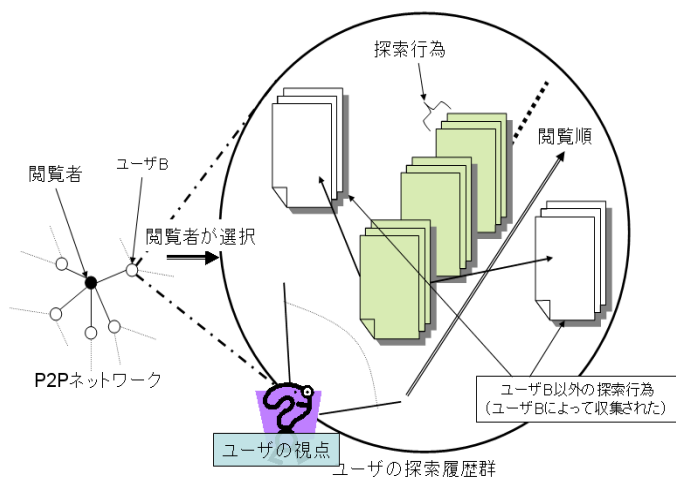


図 11 探索履歴参照の概念図

7. 結 論

本研究では P2P ネットワークを利用した探索支援方法を提案した。従来、ファイル共有にしか使われていなかった P2P ネットワークを用いてのユーザ支援の方法について示した。P2P ネットワークは対等に情報をやり取りするという点で情報共有に非常に適したネットワーク形態であると思われる。また、こ

の様なユーザ同士のやり取りによって、ユーザに近いピアを発見できる事も予備実験によって示した。

ユーザの意図の取得方法やユーザへの提示方法を含め、今後更に検討していく必要が有るとと思われる。本研究では give&take のやり取りを基調としたため、探索を行っている当該ユーザとほぼ同時期に同じような探索を行っていなければ、探索支援は出来ない。このような状況は非常にまれだと思われる。これはユーザにとって一方的に有用だと思われるユーザを集めれば解消されるが、そうすることによって P2P ネットワークにおけるピア間の接続を有向リンクとみなしてしまう。これは一方的な情報の流れを生み出してしまふことになってしまう。

また、Peer-to-Peer の接続形態はユーザ間のコミュニケーションに適した形態であると思われる。本研究では匿名的な情報のやり取りを重視したが、Web ブラウジングの状況を反映してネットワークを再構成することにより、似たページを見ているユーザ間のコミュニケーションを促進するようなモデルも考えられる。これはユーザの明示的なコミュニケーションを要するが、これによるメリットも大きい。

よって P2P ネットワークにおける情報共有の利点を生かす方法を今後より一層模索していく必要があると思われる。

文 献

- [1] 松尾豊, 石塚満, 語の共起の統計情報に基づく文書からのキーワード抽出アルゴリズム, 人工知能学会論文誌, Vol. 17, No. 3, pp. 217-223, 2002.
- [2] 松尾豊, 福田隼人, 石塚満, ユーザ個人の閲覧履歴からのキーワード抽出によるブラウジング支援, 人工知能学会論文誌, Vol.18, No.4, pp. 203-211 (2003.7)
- [3] 原田 昌紀, 清水 奨「WWW 検索システムにおける不特定多数の操作履歴の活用」情報処理学会研究会報告, 97-DPS-81-11, 1997. 6
- [4] Napster
<http://www.napster.com/>
- [5] Gnutella
<http://www.gnutella.com/>
- [6] 上田 正明, 中島 伸介, 角谷 和俊, 田中 克己, 探索アクティビティの共有と視覚化に基づく協調型情報探索, 第 13 回データ工学ワークショップ (DEWS2002) 論文集, A2-5, 2002 年 3 月