

意味情報と嗜好情報に基づく P2P システムの提案と実装

大場 正博 中沢 実 服部 進実

金沢工業大学工学部情報工学科

e-mail:{masa,nakazawa,hattori}@infor.kanazawa-it.ac.jp

近年、パーソナルユースでのマシン性能の向上やネットワークの高速化に伴い、情報配信手法の一つとして、P2P が注目を集めている。この技術によってピアを独立させ、フラットなネットワークを構築することは可能になったが、情報の取捨選択という面においては、未だ手探りの状況が続いている。

本稿では、ブローカレスな P2P ネットワークを構築し、自らが情報配信者となる手法を提案する。その際、利用ユーザの嗜好情報に着目し、個人嗜好に基づいてパーソナル化を行なうことでピアを自立させ、それらによる応答の連鎖によって、メタネットワークを構築する。このメタネットワークを利用して情報探索を行なうことによって、間接的に情報配信を行なっていく。これによって、P2P ネットワークにおける情報の取捨選択をサポートする

Proposal and construction of P2P System

based on semantic and preference information.

MASAHIRO OBA, MINORU NAKAZAWA and SHIMMI HATTORI

Department of Information Engineering, Kanazawa Institute of Technology

e-mail:{masa,nakazawa,hattori}@infor.kanazawa-it.ac.jp

Recently, the performance of network and personal computer has been improved. Therefore, P2P has become attractive to the one technique of information distribution. By this technology, it is possible to construct the flat network to be the independence of the peer.

This paper proposes the technique of peer based information distribution, implementing broker-less P2P network. This peer has personalized characteristic using liking information of the user. A meta-network is constructed by the chain of a response of these peers, searching information on this meta-network.

1. はじめに

近年、ネットワーク環境の整備と個人所有端末のスペックの向上に比例して、インターネットは飛躍的に普及し、それと意識せずとも日常生活に欠かせない存在へと入りつつある。約 10 年前に発表された WWW 技術によって、ユーザはブラウザを介し、世界中に分散された情報を活用することが可能になった。同時に、自らが WEB サイトを構築することにより、情報配信者になる術を手

にした。しかし、それらの情報を探索するためには、ポータルサイトに代表される一種のデータベースへとアクセスする必要がある。運営者によって収集・蓄積された断片的な情報を頼りに、ユーザは自らが欲する情報を選別せねばならず、それは膨大な「情報」という名の樹海に地図もなく踏み込むのに等しい。またそのデータの信頼性、リアルタイム性といった点においてもブローカに依存することになる。

本研究は、年々肥大化するネットワークと膨大な量の情報において、ユーザの意思によるコンテンツの取舍選択を可能とするインフラストラクチャを提案するものである。次世代インターネットのインフラストラクチャとしても期待されるP2P技術を用いて、ピアの自立によるブローカレス情報配信とブローカレス探索手法を提供する。探索には嗜好情報と意味情報を用い、探索要求に対して自立的に反応・応答し、それらの連鎖によって分散されたピアにおいて、ピア、情報、サービスなどの意図的な探索を可能にする。

2. P2P

P2Pは次世代ネットワークに大きく影響を与えるキーアーキテクチャのひとつである。これまでの低速、低キャパシティのネットワークからブロードバンド化が進み、それに伴いユーザが要求する情報の質も量も数年前とは比較にならないほど高品質、大容量化が進んでいる。さらにパーソナルコンピュータの低価格/高性能化が進み、インターネットがとる形態も以前の単なるデータ交換、情報閲覧にとどまらず、コミュニケーション手段やビジネス展開など、新たなツールとしての転機を迎えたといってもよい。

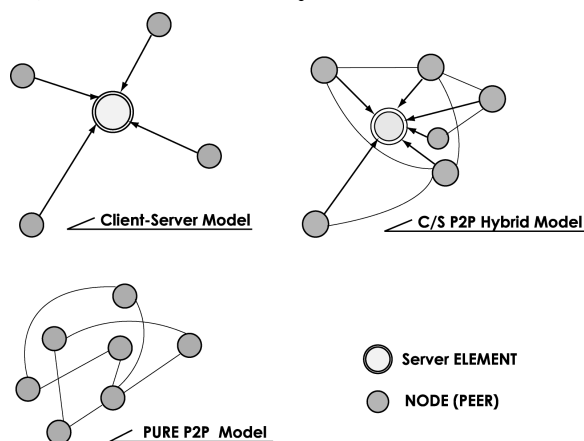


図 1 Network Model

2.1. 提案

現在、主なP2Pアプリケーションで利用されているのは、ハイブリッドタイプ(図1:右上)が主流であり、サーバにおけるメタデータ集積と、そこか

らの情報の検索という、従来から利用されてきた方法によって情報探索が行われている。この手法を用いた場合、情報提供者は本質的には情報配信者になることができず、あくまでネットワークからみたりソースの一部となってしまう。また、P2Pアーキテクチャ自体もデファクトスタンダードとなるプロトコルや、ネットワークアーキテクチャが確定せず、非常に不安定な状態といえる。そのような状況であっても、P2Pに対する期待や需要は日増しに高くなってきており、ネットワークやクライアント環境が整備されていくにつれて、新たな情報探索手法が必要不可欠になってくることは前述の通りである。

以上のようなP2Pの側面、利点や問題点、現状を踏まえ、ピュアなP2Pアーキテクチャを指向するインフラストラクチャに必要不可欠な、ブローカレス情報探索手法の提案を本研究の目的とする。

現在普及しているP2Pソフトウェアのピアは、ひとつのノードにすぎない。そこでピアに対して、利用者およびピア自体の意味情報を付加し、ピアとしての挙動と、その上で動作するサービスにおいて自律的に動作することで、ユーザの行動(探索、サービス実行等)をサポートする。

つまり、本研究は今後のインターネットインフラストラクチャを担うと考えられるP2Pアーキテクチャを用いて、ピアやピアが提供するサービス、配信するコンテンツに対して、ユーザの意思(嗜好、要求)を反映した探索を可能にする手法を提案する。つまり、各ピアが自立し、ブローカレス配信とブローカレス探索手法を提供することで、P2Pネットワークにおいて各ノードとのネゴシエーションを独自で行い、それを連鎖させることによる次世代インターネットインフラを構築するものである。

3. ブローカレスP2Pシステム

本稿で提案するシステムは、ピュアP2Pネットワーク(図1:左下)において、ピアがブローカレス情報探索/配信を行い、情報配信者としてネットワークに参加するための手法を提供するものである。

3.1. システム概要

今回提案するシステムは、JXTA¹を基本インフラストラクチャに用い、その上で稼動するサービスとしてブローカレス情報探索/配信を構築するものである。

ブローカレス情報探索を行うためには、従来のC/S型でいうデータベースが存在しないために、ピアが持つ情報を他のピアが参照できなくてはならない。そのためのアプローチとして、P2Pネットワークに対し、メタデータを流通させ、その情報を分散させることによって、広域なネットワークをカバーする手法もあるが、この場合本質的にはC/S型モデルを分散させているのと同じで、情報のリアルタイム性や正確性といった点で疑問点が生じる。また、それらの分散されたメタデータを協調させるとなると、インターネットという莫大な資源を考えれば容易なことではない。そういった理由からも、P2Pネットワーク上にデータ流通基盤としてのメタネットワークを構築するのではなく、ピアとピアのネゴシエーションの連鎖によって、動的なメタネットワークを構成する手法を提案する。

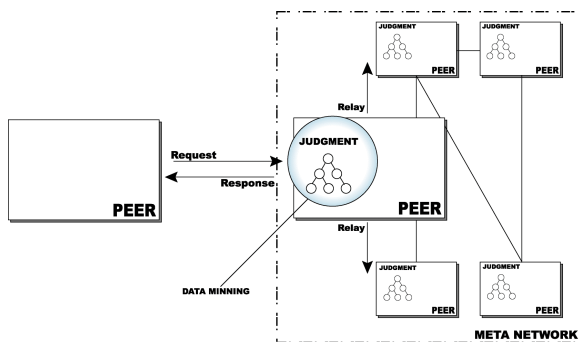


図 2 P2P Meta-network.

ピアとピアのネゴシエーションとは、探索要求という刺激に対する自発的な応答のことである。この応答において、ピアが独自の判断によって、レスポンスや探索要求のフォワード、ドロップといったルーティングを行っていく。このピアの応答の連鎖によって、ブローカレス情報探索を可能

とし、それこそがブローカレス情報配信手法となる。

3.2. 探索への応答

ブローカレス情報配信を可能にするため、ピアは探索要求に対し自発的に応答を行う。この応答には、探索要求に対するレスポンス、探索要求を他のピアにフォワードする探索要求のルーティングがある。

探索等に送信するメッセージには、探索要求のルーティングに用いる「ネットワーク情報」、要求元の嗜好情報や、ピアの情報が含まれる「意味情報」、「キーワード」の3種類の情報がXMLにおいて記述されている。

探索要求メッセージをピアが受け取ると、そのピア自身の特徴を示す複数の決定木や、両者の嗜好情報の相関、単語の結びつきをグラフ理論で表現した、意味情報ネットワークによるマッチングによって、自分自身が探索要求に該当するかを高速に判断し、条件を満たすならばレスポンスを返す。その後、要求メッセージの生存期間が許す間、ピア自身が知る他のピアに対して、探索要求の転送を行う。このあらかじめ作成された決定木を用いることで、複雑な意味情報の処理、データマイニングによる恩恵を高速に受けることができる。

決定木の作成は各ピアが、バックグラウンドプロセスとして定期的に行う。対象となるデータは、ピア利用者の嗜好を含む個人情報、履歴情報、意味情報ネットワークである。これらの情報をもとに、データマイニング手法によりピアの特徴を検出し、嗜好情報とのマッチングを高速に行う決定木を作成する。

3.3. ピア

ピアとは、P2Pネットワークに接続されているすべてのノードを指す。ネットワークを構成する要素がピアであり、ピアこそがP2Pネットワークそのものを表す。つまり、ピアが存在する時点で、そのピアは一つのP2Pネットワークを形成することを示す。

ピアの構成は、図3のように3層に分類され、インフラストラクチャを含み最下層にあたる

¹ projectJXTA が提供する P2P プロトコル群 (<http://www.jxta.org>)

Network 層に JXTA を利用する。ピアの発見や、ピアまでのルーティング、メッセージの送信といった P2P システムにおけるコア部分を JXTA に委任することで、JXTA 上のサービスとしてブローカレス探索を提供することができる。このことにより、P2P コアアーキテクチャが大きく変化しようとも、その流れを追従するのは JXTA だけでよく、探索モデルというコアサービスをコアインフラストラクチャから切り離して設計することが可能になる。

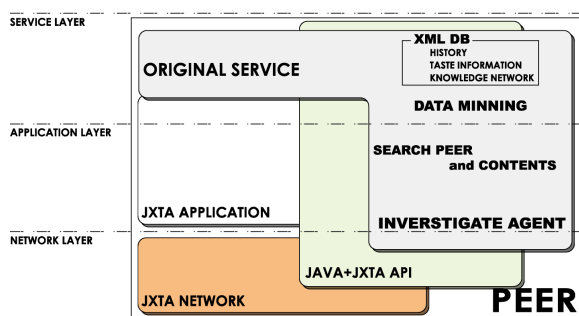


図 3 PEER

3.4. 嗜好情報と知識ネットワーク

ピアの個性を表現するために、ピア利用者の興味を示す個人嗜好情報と、嗜好情報のパラメータやキーワードの繋がりを知識ネットワークとして保持する。これらは、履歴情報やメタデータの解析によって、そのピア独自の再構築を繰り返し、ピアの自立の中心となるデータを作成する。

ピアの嗜好情報には、静的嗜好データと動的嗜好データがある。静的嗜好データは、各ピアの利用者に対して行なうアンケートによって取得する。これにより利用者の意思による確実な嗜好情報をピアに反映することができる。動的嗜好データは、利用者の行動履歴をログデータとして常に取得し、そのログデータを解析することで得ることができる。これによって、ユーザが無意識に感じている嗜好を取得することができる。静的嗜好データは、ユーザ自身が入力しているためデータの信憑性が高い。動的嗜好データは、確実性を求めるものではなく、静的嗜好データでは取得することのできない、利用者の潜在的な思考を補完するデータで

ある。

嗜好情報は、関連性の強いもの同士を相互に連結し、意味情報ネットワークを構成している。

この意味ネットワークを用いて、嗜好の相関を求めることによって、ジャンルを超えた嗜好の相関を求めることができる。

3.5. データマイニングによるピアの自立

ピアの自立とは、ブローカレス情報探索が行なわれた際、探索要求メッセージを自動的に処理することを指す。そのとき、前述の嗜好情報や意味情報ネットワークをピア自身が理解できるようにすることで、ピアが独自の解釈を行うことができる。これによって、ブローカレス情報配信を可能にすると共に、他ピアとの差別化が計れるため、ピアの探索要求にも柔軟に対応できる。

ピアの探索では、ユーザ任意のキーワードと、嗜好情報を利用することができる。キーワードを入力しない場合でも、ピアの嗜好情報を基にして、他ピアとの嗜好の相関を求めることにより、ピアの探索が可能になる。

ピアの自立を行なう上で、嗜好情報を用いたデータマイニングを行なう。これによって、ピアの嗜好情報や履歴情報からピアの特徴を抽出し、単なるマッチングでは表現できなかった検索手法を提供することが可能になる。今回提案するシステムでは、ピアの特徴を検出するのに用いる方法として、特徴検出、相関、ログ解析がある。これらによってピアが保持するあらゆる情報が分析され、探索要求への応答を判断する、重要な材料になる。

3.5.1. ピアの特徴検出

嗜好を用いたピアの探索では、処理の高速化のため、最初にピアの特徴情報によるマッチングを行なう。特徴情報とは、ピアの嗜好を分析した結果、顕著に現れた特徴を示す情報である。特徴検出では、静的嗜好データから平均、分散を求め、そのピアが何に興味があるかを判断する。この特徴情報が類似した場合、両者の類似性は著しく高いと判断し、その後のあらゆる嗜好の相関を試すまでもなく、応答処理を行なう。

3.5.2. 相関を用いた探索

検出された特徴によるマッチングが失敗した場合、両者の思考の相関を求める。相関はピア同士の嗜好の類似度という観点から行われる。

相関を取る手法として、静的嗜好データの相関、意味情報ネットワークを利用した嗜好の相関、前者にキーワードによる刺激を加えた相関、特徴情報に意味情報ネットワークを反映させた相関の4種類を提供する。

3.5.3. キーワードを用いた探索

3.5.3.1. コンテンツ情報の把握

これまでに述べた相関を用いた探索は、嗜好情報によりピアを探索する方法であった。次に、嗜好ではなく、キーワードを用いたピア及びコンテンツの探索手法を提案する。

キーワードを用いた探索では、ピアに存在するコンテンツが重要となる。コンテンツとは、JXTA に依存する事無く、ピアが公開しているリソースを指す。これらのコンテンツに対しメタデータを付加することによって、ピアにコンテンツを認識させることが可能になる。

メタデータは、コンテンツの特徴を示す事柄が、テキスト情報である XML 形式で記述される。このメタデータをテキストマイニング手法の一つである、形態素解析を用いて分析することで、メタデータ中に存在するキーワードを抽出ことができ、さらに、特定の単語の出現率や、単語の組み合わせなどが把握できる。これにより、コンテンツの形式を超えて一様にリソースを扱うことが可能になる。また、このメタデータによって、従来のような全文検索的なコンテンツの探索も可能になる。

3.5.3.2. 意味情報ネットワークの動的再構築

キーワードを用いたピアやコンテンツの検索では、複数のキーワードを指定する事ができる。大抵の場合、入力された複数のキーワード間には、何らかの因果関係があると予測できる。このキーワードをログデータとして保管し、バックグラウンドで解析を行い、その結果を意味情報ネットワークに反映することで、新たな単語を学習していく。

履歴データの解析方法として、データマイニン

グ手法の一つであるマーケットバスケット分析を用いる。これにより、キーワードの依存関係を分析し、意味情報ネットワークへの動的追加、再構築を可能にする。

3.6. ブローカレス探索の流れ

3.6.1. ネットワークアーキテクチャ

本システムでは、P2P インフラストラクチャ上においてブローカレス探索モデルを適用するために、JXTA プロトコルが提供するネットワークアーキテクチャに加えて、探索要求を P2P ネットワーク上に展開させるために、独自のルーティングを行なっている。JXTA では、あらゆるメッセージ通信が XML 形式で取り扱われている。そのため、その下位エレメントとして独自の XML を挿入する事が可能である。その構造を利用して、サービス独自のメッセージを、JXTA プロトコルの上位プロトコルの位置づけで利用することが可能になる。これにより、ピアを P2P ネットワークという概念から開放し、あくまでサービス同士の通信として扱うことができる。

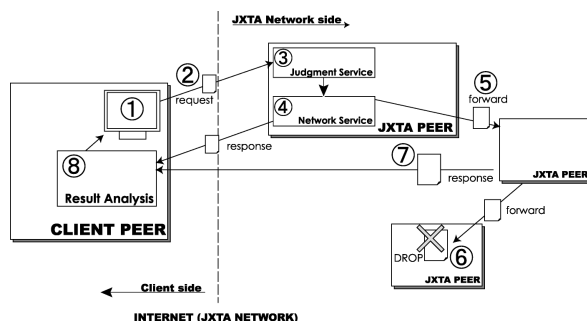


図 4 flow of broker-less Search

- i. 探索要求モジュール (図 4: ,)
- ii. 探索応答モジュール (図 4:)
- iii. メッセージ制御モジュール (図 4: , ,)
- iv. 探索応答待ち受けモジュール (図 4:)

3.6.2. 探索要求・応答の手順

ブローカレス探索を行なう一連の手順は、図 4 のようになる。基本的な流れとしては、探索要求者が JXTA ネットワークに対し、探索要求メッセージを発行し、それを受け取った各々のピアが、メッセージにより応答処理を判断し、応答メッセージの送信、もしくは他ピアへの転送処理を行な

う。探索要求モジュールでは、サービス上から JXTA プロトコルの PDP,PBP を利用して、P2P ネットワークに探索要求の配信を行なう。

探索応答モジュールは、C/S 型でいうサーバにあたり、探索要求メッセージを受け取るための告知を発行している。この告知のパイプを通じて、探索要求メッセージを受信する。受け取ったメッセージが探索要求であった場合、探索要求手法に応じて、探索要求を満たしているか判断し、条件を満たす場合、探索結果を XML 形式で挿入し、メッセージ制御モジュールに転送を依頼する。また、条件を満たす、満たさないに関わらず、探索要求メッセージをメッセージ制御モジュールに転送する。

メッセージ制御モジュールは、ブローカレス探索サービスのメッセージ転送を制御する部分にあたる。処理内容としては、探索応答メッセージの送信、探索要求メッセージのマルチキャスト配信、探索要求メッセージのドロップの 3 種類になる。メッセージには生存期間を示す TTL の値が設定されており、メッセージを転送させる毎に減少させ、メッセージの生存チェックを行なう。探索要求の場合、先ずメッセージに経路履歴として自身のピアの ID を追加する。その後、探索要求モジュール同様に配信先をリストアップする。その際、探索要求の重複を避けるために、経路履歴内に記述されているピアをリストから削除し、要求の配信を行なう(図 4:)。応答メッセージの転送の場合、探索要求元のピアに対して、JXTA プロトコルの ERP を用いて直接転送を試みる(図 4:)。経路が見つからない等の理由で直接送信できない場合、メッセージが転送されてきたピアの履歴に基づき、JXTA プロトコルの PBP、ERP を用いて探索要求元への転送を依頼し中継していく。ここで送信先がなくなった場合、TTL の値に関わらずメッセージをドロップする。

探索応答待ち受けモジュールでは、探索要求に対する応答を収集する。探索要求の内容によって、応答メッセージは変化するが、XML 形式で扱うことで、フォーマットを標準化する。また、プロ

ーカレス情報探索においては、あくまでネットワークを経由しての非同期なセッションとなるために、時間の経過と共にレスポンス情報が収集されることになる。この点においても XML 形式で格納することによって、応答メッセージの動的反映や分析が容易に行なえる。

4. まとめ

本稿では、P2P ネットワークにおいて、自らが情報配信者になる手段を提供し、利用者自身の意図による情報探索を行なうブローカレス情報探索手法を提案した。本システムでは、サーバを介入させないピュア P2P を前提とした。それは、情報配信者になるという、インターネットの本質的な部分を考慮した場合、ブローカを挟まない形こそが最善のソリューションを提供できると判断したからである。そのため、ピアの自立によって間接的なブローカレス情報配信を可能とし、ピュア P2P モデルにおいてブローカレス情報探索と配信を実現した。

謝辞

本研究は、通信・放送機構の地域提案型研究開発制度の支援を受けて実施された。ここに記して謝意を表す。

参考文献

- [1] Michael J.A.Berry and Gordon Linoff , *Datamining Techniques For marketing , Sales and Customer support*, 1999/9/24
- [2] 星合隆成,柴田弘 “御用聞き型情報提案のための自律分散照合環境アーキテクチャとその性能評価”,(D-I) Vol.J83-D-I No.9 pp.1001-1012,2000,9
- [3] 星合隆成,“意味情報ネットワークアーキテクチャ SION” http://www.jnutella.org/jnudev/jws_0522/jws_hoshiai.lzh
- [4] XML/SGML サロン,“標準 XML 完全解説”,株式会社技術評論社,1999,5
- [5] 西山和雄,“P2P ネットワークによるコミュニティ嗜好検索システム” http://www.jnutella.org/jnudev/jws_0522/jws_nishiyama.lzh