

グループ生成エージェントによる情報検索方式

金子祐真[†] 澤本潤[†] 小泉寿男[‡]
 sin@itlab.k.dendai.ac.jp[†] sawamoto@jcom.home.ne.jp[†] koizumi@k.dendai.ac.jp[‡]

近年、インターネットのブロードバンド化に伴いデジタルコンテンツの多様化、大容量化が進み、個人の持つコンテンツを共有のファイルとして利用する研究が進められている。個人間コンテンツ共有システムには、ファイルの送受信をする際にファイルを持つユーザをどのように特定するのかの課題がある。本稿は、同じようなファイルを持つユーザ間でグループを生成させ、コンテンツ共有システム全体としての能率的な検索方式を提案する。コンテンツ共有システムは、検索やグループの生成の動作によりコンテンツ共有システム全体として複雑な動作になる。そこで、自律的プログラムであるエージェント技術を用いることでコンテンツ共有システムの動作を実現する。

A information retrieval system by group generation agent

Yushin Kaneko[†] Jun Sawamoto[†] Hisao Koizumi[‡]

In recent years, due to diversification and availability of large-scale digital contents with the progress of broadband services of the Internet, researches for utilizing contents, which an individual user has, as a shared files are advancing. In case a file is transmitted and received, there exists a difficulty of how to identify a user who holds the file in the shared file system. This paper proposes an efficient content retrieval method forming groups of users who hold similar contents files in the shared file system. The operations inside the shared file system become complex by individual users' activities of group formation and retrieval of files. Then, using the agent technology, which is an autonomous program, solves the complexity of operations of the shared file system.

1. はじめに

通信インフラの整備や、携帯・情報端末等の情報機器の発達によりインターネットを用いる情報の流通が一般化してきた。それに伴い多種多様で大容量のコンテンツが増えてきている。それらを受け、近年では個人間コンテンツ共有システムについての研究が盛んになってきている[1][2]。

コンテンツ共有システムとは、ユーザが欲しいコンテンツ(以後、コンテンツ=データ)を持つユーザを特定し、そのユーザと直接データを取得する方式である。これらの方式では、ユーザとデータを管理するサーバを設置する方式、またユーザがある程度のユーザを知っており、パケツリレー式にデータの検索を行っていく方式[3]がある。

しかし、管理サーバを用いる方法ではユーザやデータを管理しきれないという問題がある。また、パケツリレー式ではユーザを跨いでデータの検索を行う為、ユーザが n 人存在し、あるデータを持つユーザがそのうちの一人であったならば、最大で n 人にデータの有無を一人ずつ確認しなければならない。 n 人が n 人ともデータの有無を確認するならば、最大で $(n-1)^2$ 回のやり取りが必要となる。この動作をネットワーク上で行うと膨大なやり取りが発生してしまい、ユーザが特定しにくい。また、ユーザにアクセスするにもそのユーザを知っていなければならないという問題がある。現在、情報検索ではエージェント技術がサーチエンジンの補助機能や Web における情報収集ツールとして活用されている[4]。しかし、

これらはユーザの情報収集の補助が目的である。しかし、エージェント技術では、ユーザの持つデータから嗜好を示すベクトルを生成して嗜好の同じユーザを発見する研究や、エージェントによるグループ協調動作が研究されている[5]。

本研究では、エージェントによるグループを用いた協調動作に着目し、エージェントによるユーザ間情報検索システムを考える。本稿では、管理システムを要求しない個人間のコンテンツ共有システムを考える。本稿では、単にパケツリレー式にデータの検索を伝達するのではなく、ユーザの持つデータを Keyword で管理し、同じ Keyword を持つユーザをグループとして扱うことで検索能率を上げる方式を提案する。特定の管理システムを用いないことで、汎用性とシステムの安定をはかる。本方式ではユーザが独自に検索とグループ生成を行う、個々の動作に自律的プログラムであるエージェントを用いる。その協調動作によりコンテンツ共有システムの複雑な分散処理を実現し、本方式の実現を目指す。

2. コンテンツ共有に関する課題

特定の管理システムを要求しない個人間のコンテンツ共有システムを適用する典型的な環境は、以下の条件を満たすものとする。

- 1) 不特定多数のユーザが存在する
- 2) 各ユーザがデータ(コンテンツ)を持つ
- 3) ユーザ・データを管理するサーバは存在しない

4) データを検索する際に、各ユーザは既知のユーザへと検索をバケツリレー式に伝達する構成を図 1 に示す。

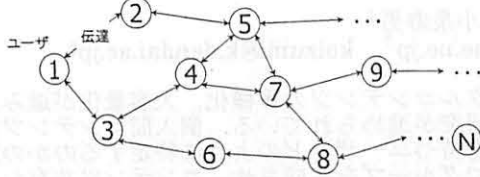


図 1 管理システムの無いユーザ間の連携

図 1 において、ユーザ①があるデータが欲しい時、ユーザ①はまず既知のユーザであるユーザ②、ユーザ③に対してデータに関する検索条件を伝達する。これを受けてユーザ②、ユーザ③は既知のユーザであるユーザ⑤、ユーザ④、ユーザ⑥にデータに関する検索条件を伝達していきます。伝達される検索条件にはユーザ①が設定した規定回数が記述されており、ユーザ一人がデータに関する検索条件を受け取るたびに規定回数を減少させ、規定回数が尽きるとそこでデータに関する検索条件の伝達は止まる。

ここで、ユーザが①～⑨まで存在していても、規定回数分の伝達しか行われなため、ユーザ⑨までたどり着くとは限らず、検索対象のデータを持つユーザを特定に至らない場合がある。また既知のユーザに対する伝達では、伝達を受信していないユーザにのみ効率よく伝達されていくとは限らず、ユーザ①が設定した規定回数により、検索対象のデータを持つユーザを特定するまでにまたぐユーザ数（ステップ数）が多くなる。これらのコンテンツ共有の環境に対して、以下の課題を解決する。

- 1) 検索をユーザ間で伝達する際のステップ数の軽減する
- 2) 検索対象のデータを持つユーザを特定する

これらの課題を解決し、限られたステップ数内での効率的な検索を実現する。また、本システムを適用することでネットワークに分散するデータを分類し、大きな分類から小さな分類を特定していく。

3. グループ生成エージェントによる情報検索方式

3.1 概要

本方式では、ユーザの持つデータをもとに嗜好の似通ったユーザでグループを生成し、そのグループを検索に利用することでステップ数を軽減し、検索対象のデータを持つユーザを特定する。検索の伝達はステップ数により制御されているので限られたステップ数内での効率の良い特定を目指す。これらの処理はエージェント技術を用いて実現する。本方式では、エージェントを「ルールとデータを持ち、ルールに基づきデータを収集し、収集したデータに基づきルールを作成する

という自律性を持たせたプログラム」とする。エージェントを用い本システムを自動化することで、ユーザに負担をかけずに恒常的なシステムを構築する。グループを生成・管理はエージェントが行う。

本方式のグループを用いたエージェントの連携を図 2 に示す。エージェントはユーザとユーザの持つデータを管理する。

図のようにグループの生成とデータの検索条件の伝達に注目し、“検索条件の伝達に対する応答を利用しグループを生成する”、“グループを利用して検索のステップ数を減少させる”の二つを相互に行うことでユーザ間に有機的な関係を構築する。

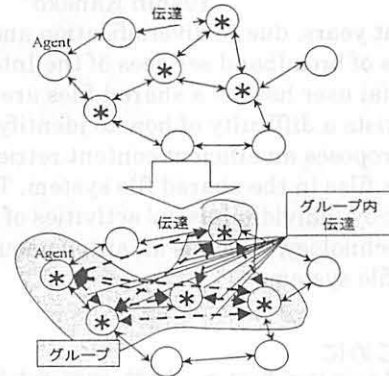


図 2 グループを用いた Agent の連携

3.2 グループの構成

グループはエージェントの持つデータに合わせて構築する。まず、エージェントは自身の持つデータに対してユーザから与えられた Keyword を用い管理する。Keyword はデータを端的に示す名詞とする。図 3 のように各エージェントは自身の持つデータを示す Keyword がほぼ同じエージェントとグループを生成する。

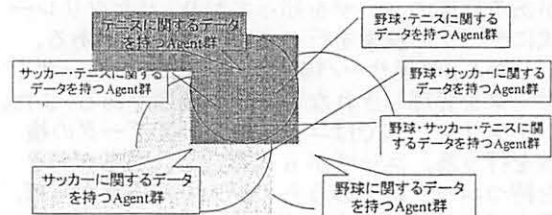


図 3 Keyword によるグループ生成

データに関する検索条件の伝達に対する応答によりグループを生成するという点と不特定多数のエージェントに対して検索条件を伝達するという点において、存在する全てのエージェントから同じ Keyword を持つエージェントを全て抽出してグループを生成させるのは不可能である。よって、図 4 のように同じ Keyword を持つエージェントのグループは一つではなく複数のグル

ープを生成させる。エージェントが所属するべきグループはエージェント間のステップ数で割り振られる。

エージェントは Keyword の数だけグループに所属する。また、同じ Keyword を持つグループが複数存在する。エージェントはグループに所属する他エージェントのメンバー情報を持つ。

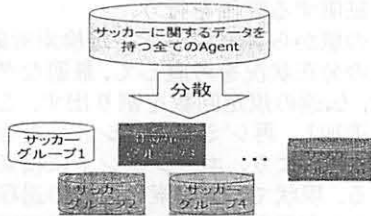


図 4 エージェント管理による複数のグループ

3.3 グループ生成と検索

本方式は、グループを用いて検索の伝達を行うことにより、ステップ数の軽減とあるデータを持つユーザの特定を行う。グループが生成されていない場合にはパケツリレー式に検索を伝達し、あるデータを持つユーザの特定とグループの生成を行う。エージェントはデータに関する検索条件を規定回数分だけエージェントに対して伝達し、検索条件に該当するデータを持つエージェントは検索条件の発信源のエージェントに対して、図 5 のように直接検索結果を送信する。この時、検索条件に対する応答結果からグループが生成される。また、検索条件の伝達時にデータを持つエージェントを経由した場合は検索条件にヒットしたエージェントが記載されて伝達が行われていく。検索はグループの更新となり、検索頻度の高い Keyword を持つグループの更新頻度は高く、検索頻度が少なければグループの更新頻度も下がる。

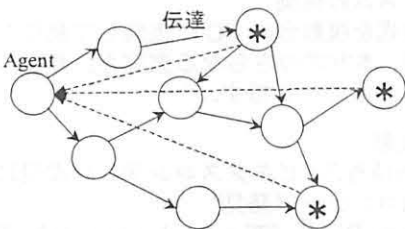


図 5 検索条件の伝達と結果の送信

規定回数のステップ数で伝達が行われて図 6 のようにグループを生成した後、同様の検索条件がエージェント間に流れた場合、グループのエージェントがその条件を受け取ると図 6 のようにエージェントは所属するグループの属性との関連を確認する。属性が同じならばメンバーに対して図 7 のように検索条件を送信する。このとき、検索条件を受信したエージェントはグループのリーダーであるファシリテータ[6]となり、データの収集と返信を行う。ファシリテータは図 8

のようにグループのリーダーとして、グループ内の各エージェントを一時的に統率・管理する。

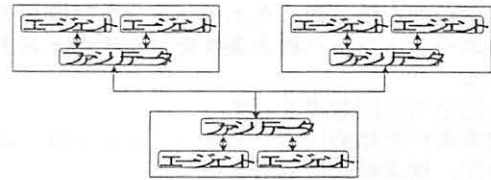


図 6 ファシリテータを用いたメッセージ交換

本方式ではグループに所属するエージェントは、グループ内の各エージェントについての情報とグループの持つデータを端的に示す Keyword 群(属性)を持つ。このようにグループにはリーダーは存在しないが、メンバーの一人が検索条件を受け取ると図 7 のように一時的にファシリテータとなりグループ内でのデータ収集と返信を担う。これにより、検索条件の伝達を削減する。

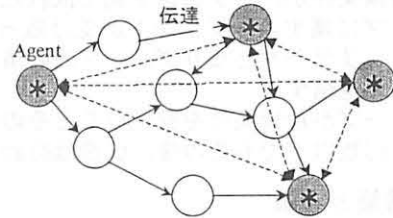


図 7 検索結果に基づくグループ生成

まず、グループに所属するエージェントが受信した場合、自身のグループとの比較を行う。グループに対して関連のある検索条件ならば、図 8 のように条件を受け取ったエージェントがファシリテータとなりグループ全体から検索条件にあうデータを探し、検索条件発信者に結果を送信する。このとき、グループ内のエージェントはファシリテータから要求を受けると、ファシリテータとのみ通信し、その後、間を置かず同検索条件発信者から同検索条件を受信してもファシリテータにならず、また既知のエージェントに対しての送信を行わない。このように独立したエージェントの連携によりシステムを構築することで有機的に動作と機能を提供する。

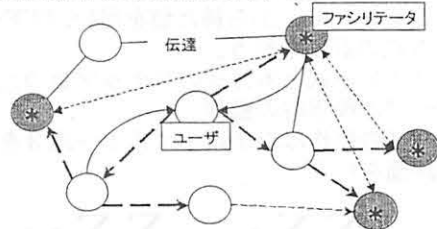


図 8 グループ生成後の検索

3.4 エージェントのアルゴリズム

エージェントはユーザの代理として、データの管理と検索の伝達、グループの生成を行う。検索

の伝達とグループ生成には以下のルールを適用する。

- 1) 検索条件を受け取ったエージェントは既知の他エージェントへ検索条件をマルチキャストする
- 2) 上記を規定回数繰り返す
- 3) 検索条件を始めに出したエージェントは一定時間、検索結果の返信を待つ
- 4) 一定時間内に受け取った返信結果を基にグループ生成を行う
- 5) グループは特定のステップ数内・検索条件を満たすデータを持つエージェントメンバーで構成される
- 6) 検索結果を受信したエージェントは5)の条件を満たすエージェントにグループのメンバーとなるエージェントとそれらが管理するデータに関する情報を与える
- 7) 以後の検索では、グループを示す属性に関係する検索条件がエージェント間で流れたとき、グループに属すエージェントが受け取った場合、ファシリテータとなりグループ内の情報をまとめて返信する
- 8) グループが検索条件を受け取るとその条件は規定回数以下でもその後、伝達は行わない

4. 構築と評価

4.1 シミュレーションによる検証

現在、シミュレーションプログラムの構築、本方式の検証を行っている。シミュレーションプログラムをJava言語であるJ2SDK1.4.2を用い構築・測定中である。

シミュレーションでは配列一つを一つのエージェントとし、ランダムにKeywordを二次元配列に配置する。検索では各配列内に格納されたKeywordを拾い、配列データからグループ生成を行うプログラムである。測定対象は各規定回数での検索における特定のKeywordを内包した配列の取得状況、グループの構成エージェント数(配列の個数)とする。また、シミュレーション結果としてランダムに配置されたエージェント(配列)に対して、検索の伝達を示すツリーを表示させ、グループの有無における差異を測定する。

シミュレーションから得た値を用い、以下の項目で本方式の評価を行う。

- 1) バケツリレー式と本方式のステップ数の比較
- 2) グループの検索時の動作
- 3) グループの動作は以下に示す式1～式3を用いて評価を行う。

$$\frac{\sum_{i=1}^N z_i}{N-1} \dots \text{式1} \quad \frac{\sum_{i=1}^N \sum_{j=1}^N z_{ij}}{N(N-1)} \dots \text{式2} \quad \frac{\sum_{i=1}^N \sum_{j=1}^N z_i \cdot z_{ji}}{N(N-1)} \dots \text{式3}$$

評価方法としてあるエージェント*i*がネットワークにおいて接続先がどのくらい存在し、中心的

役割を果たしているかを示す中心性(式1)、与えられたネットワークにおけるあるエージェント*i*からあるエージェント*j*までの接続経路がどのくらい存在するかを示す密度(式2)、ネットワーク中にあるエージェント*i*からあるエージェント*j*までの双方向の接続経路がどのくらい存在するかを示す結束性(式3)を算出して、ステップ数の減少を証明する評価を行う。

これらの値からエージェントが検索対象とするデータの分布状況を考慮して、最適なグループの生成数、伝達の規定回数を割り出す。これらのルールを追加し、再シミュレーションを行う予定である。これにより、エージェントに検索範囲を判断させる。現状では、検索の伝達の過程において、データの有無を記載していき、規定回数の終端のエージェントにおいて、規定回数を増やして検索の伝達を続行させる方式を検討中である

4.2 PC多数台使用の実装・評価

シミュレーションで本方式の動作を確認したので、実際にエージェントプログラムを用い実装する。まず、5台PCにエージェントを配置し、検索条件の伝達と検索結果の返信、それに伴うグループの生成を確認する。次に、段階的にPCの台数を増やして評価を行う。

5. まとめ

管理システムを必要としないグループ生成による情報検索方式のコンテンツ共有システムを提案した。今後の作業・課題は以下の通りである。

- 1) シミュレーションより、グループに所属するエージェント数、規定回数の算出とグループ生成、検索条件伝達の動作の確認・シミュレーション結果から得られた値と評価を基にルールを見直し、状況に応じたルールの構築
- 2) エージェントによるルールの取捨選択のアルゴリズムの構築
- 3) 本方式を複数台のPCに実装して動作を確認
今後、本方式のさらなる充実と、実際のネットワークへ対応を目指す。

参考文献

- [1] 中内清秀, "ユビキタスコンテンツ環境における分散コンテンツ発見"
- [2] Peter. Biddle, "The Darknet and the Future of Content Distribution," <http://msl1.mit.edu/ESD10/docs/darknet5.pdf>
- [3] 加藤剛志他, "ユニバーサルP2PプラットフォームにおけるシングルホップP2Pマルチキャストの検討," DICOMO 2004, July 2004
- [4] 國頭吾郎他, "モバイルエージェントを用いた連携検索システムの実装に関する検討,"
- [5] L. Foner, "A Multi-Agent Referral System for Matchmaking," PAAM '96, April 1996.
- [6] 木内東他, "マルチエージェントシステムの基礎と応用", コロナ社, 2002年