

仲介者を用いた人のネットワーク拡張手法の提案

Proposal of the extension method for personal network by using mediator

濱崎 雅弘^{*1} 武田 英明^{*1*2}
Masahiro HAMASAKI Hideaki TAKEDA

^{*1}奈良先端科学技術大学院大学 情報科学研究科
Graduate School of Information Science, Nara Institute of Science and Technology

^{*2}国立情報学研究所
National Institute of Informatics

In this paper, we propose the matchmaker model as a method to extend personal networks. Personal networks is useful for information gathering, but in order to be a effective method, it is necessary to extend and optimize it. Using the matchmaker model, we can increase a new friend who is expected to share interest. Iteration of matchmaking is used to optimize personal networks. We simulate the matchmaker model with a real and virtual data and compare the results by the matchmaker model with those by the central server model. The matchmaker model can reach almost the same results obtained by the server model with each type of data.

1. はじめに

同じ関心を持った人同士の情報交換はインターネット上で極めて活発に行われている。現在運営されているメーリングリストやオンライン掲示板は数多くあり、例えば日本国内で無料メーリングリストを提供している FreeML^{*1}では2000年12月の時点で登録者数500万人にのぼり、月間6000万通ものメールが配信されている。巨大掲示板2ちゃんねる^{*2}では300以上の話題別カテゴリーが存在し、一日のページビューは700万を上回ると言われている[インターネット協会 01]。様々な情報共有サービスも提供されており、WWWブックマークを共有するBlink.com^{*3}では2001年3月の時点で全世界では100万人程度の会員を有し、収集したURL数は2400万になる。このように、多くの人々がネットワーク上の「場」に多くの情報を集積し共有することで有益な情報を手に入れている。インターネットにおける情報収集手段として情報共有の有効性は一般に認知されていると言える。

しかし、メーリングリストや掲示板といった「場」を提供する情報共有にも問題点がある。例えば、場に多くの人が集まりすぎて相互依存の情報共有関係が破綻してしまうことや、自分の関心に合った場を探索する必要があることが挙げられる。そこで、情報共有の形態として、このような場を使うのではなく、個人間の共有の連なりによって全体での共有を行うという情報共有の形態が考えられる[吉田仙 01][吉田匡志 01]。

このような共有には、ネットワークを構成する個人間の関係が適切である必要がある[竹内 01]。しかし、各個人が自発的に作っただけの情報共有ネットワークでは、全体として必ずしも最適な関係であるとは限らない。ゆえに作られた情報共有ネットワークを元に、さらにネットワークを拡張する方法が必要となる。この際、拡張は個人間の関係を元に自律分散的に行われることが望ましい。なぜならネットワークの拡張を第三者であるサーバなどに委ねた場合、情報隠蔽や情報提供の心理的負担などの点で問題が生じる可能性が考えられるためである。

このような背景を元に、我々は個人間の情報共有ネットワークを自律分散的に拡張する手法である仲介者モデルを提案す

る。仲介者モデルは、各ノードを中心として自分に直接リンクが張られているノード同士の中でリンクを張るべき関係を見出すという方法である。本論文では、仲介者モデルの有効性を中央集権モデルとの比較によって検証する。

2. 仲介者モデル

仲介者モデルとは、共通の友人が仲介者となって、新しい友人を得るという方法をアルゴリズム化したものである。前提となるネットワークは次のとおりである。

- 任意のノードとノードの間には関係の近さを計算することができる
- 任意のノード間にはリンクを張ることができる

このような条件のもと作られるネットワークは関係の近いノード間でリンクが張られているかどうかは保証されない。このとき、このようなネットワークから関係の近いもの同士のネットワークを作り出すことが必要である。そこで、分散的に目標のネットワークを形成するアルゴリズムとして以下の仲介者モデルを提案する。

仲介者モデルは、各ノードを中心として自分に直接リンクが張られているノード同士の中でリンクを張るべき関係を見出すという方法である。各ノードのこの動作を繰り返すことで、漸近的に目標とするネットワークに到達することが期待される。各ノードでは自分につながっているノードの集合からノードの組を作り関係の近さを計算することで、関係の近いノードの組を抽出する。

我々は[濱崎 02]において、個人間の情報推薦のために共通話題ネットワークを用いることを提案し、さらに個人間の関係を測る指標を提案した。仲介者モデルは、これらの手法によって、それぞれ個人間で情報共有関係が築かれている状態において適用可能である。仲介者となるノードは、自分と相手との関係だけでなく、すでに自分と関係を築いている相手同士の関係も計算する。そして、その中から関係が深いと思われる二者の関係を、その両者に新しい個人間ネットワークとして提案する。実際に関係を結ぶかどうかは、提案された両者に委ねられる。このような仲介を繰り返すことで不完全な関係ネットワークから適切な関係によるネットワークを自律分散的に生成する。本手法では基本的に既にある関係を元に新しい関係を構築する。この、既に関係を結んだ相手が仲介者として存在すると

連絡先: 奈良先端科学技術大学院大学 情報科学研究科, 〒630-0101 奈良県生駒市高山町 8916-5, Tel:0743-72-5085, Fax:0743-72-5269, E-Mail:masa-ha@is.aist-nara.ac.jp

*1 <http://www.freeml.com/>

*2 <http://www.2ch.net/>

*3 <http://www.blink.com/>

というのは、新しい相手と関係を結ぶにあたっての心理的負荷を軽減する効果があると期待できる。この利点は、個人間の共有をベースとした情報共有形態の方針に沿うものと言える。

3. シミュレーション実験

3.1 実験概要

仲介者モデルによるネットワークの再構成が有効かどうかを、簡易なシミュレーションによって検証する。シミュレーションでは、仲介者となるノードが自身と接続しているノード間の関係を計算可能である。ここでの戦略は、仲介者から伝えられた新しいノードとの関連度情報をもとに、各ノードが自分とより高い関連度を持つノードとパスを張ろうとする単純なものである。本実験では、この戦略でどの程度最適な状態へ近づくことができるかを調べる。

シミュレーションの概要を図 1 に示す。まず初期パスを設定する。次にパスの張り替えを行うノードを選択する。そして張り替えを行うノードが、新たに張ることの出来るパスの中から追加するパスを、すでに張られているパスから削除するパスを選択し、張り替えを行う。張り替え後に収束判定を行うが、パスの張り替えには条件があり、全てのノードが条件を満たさなくなれば収束と判定する。収束していればシミュレーションを終了し、そうでなければ再度ノード選択から繰り返す。

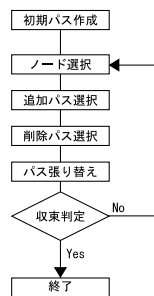


図 1: シミュレーション概要

シミュレーション開始時にノード間をつないでいる初期パスはランダムで設定される。ただし、必ず各ノード共一つはパスを持つようにする。パスを張り替えるノードもランダムで決定されるが、パスの張り替えは常により良いパス（より関連度が高いノードとのパス）を選ぶので、それが出来ないノードは対象とならない。つまり新たに張ることの出来る（仲介者によって提案される）パスの関連度が、すでにあるパスのどれよりも低いノードは対象とならない。なお、パスの関連度とは、そのパスがつなぐノードとノードの関連度を指す。

追加するパスは新たに張ることの出来るパスの中で関連度が最も高いものを選び、削除するパスは、すでに張っているパスの中で関連度が最も低いものを選ぶ。本シミュレーションでは一回の張り替えで、ある一つのノードが一つのパスを追加し、同時に一つのパスを削除するので、全体としてパス数は常に一定に保たれる。

図 2 はパスの張り替えの例を示した図である。ノード A は、すでに B, C, D, F の四つのノードとパスを張っていたとする。ノード A が追加可能なパスは、ノード B および F が仲介者となって紹介してくれたノード E, G とのパスである。このパスを張るに当たっての判断基準として、両ノードを知っている仲介者が関連度を与えている。まずノード A はすでにあるパスのなかで最も値の低いものを探す。この場合ノード B と

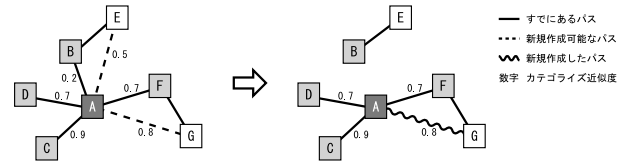


図 2: パスの張替え

のパスが 0.2 で最も低い。次に新しく張ることができるパスで最も値の高いものを探す。この場合ノード G とのパスとなる。新しくノード G と張るパスの方がノード B とのパスよりも高い値を持っているので、ノード A はパスの張り替えを行う。具体的にはノード B とのパスを削除し、ノード G とのパスを作成する。なお、パスの追加削除はパスの一端のノードの判断によって行われ、パスの両端のノード双方の判断で行われない。そのため、場合によってはどこからもパスのない孤立ノードが生まれる可能性もある。本シミュレーションでは、いったん孤立したノードが復帰する手立てはない。

収束判定はパスを張り替えるノードがあるかどうかで判定する。つまり、全てのノードがパスを張り替えるノードとして選択可能でなくなった場合（全てのノードが、一回のパス張り替えによって、今あるパスよりも高い関連度を持つパスを張ることができなくなった場合）、収束したとみなす。

3.2 評価基準

パスの関連度を全て計算できれば指定された数のパスでもっとも良い物（関連度の高いパス）を選ぶことができる。仲介者を用いたパスの張り替えだけで、どこまでこれに近づくかという点が問題となるので、被覆率と到達率という二種類の値を評価に用いる。

被覆率とは、パスの関連度を全て計算できた場合に作成できる、指定された数のパスでもっとも良い（関連度の高い）パス集合をどの程度再現できているかを示す割合である。大して到達率とは、各時点で張られているパスの関連度の平均値を最高値（中央集権モデルで求めた最適な関係のネットワークにおけるパスの関連度の平均値）で割った値である。

4. 実験結果

4.1 実データを用いた場合

[濱崎 02] では、個人間の関係を測るために WWW ブックマークを用いた。この実験データを用いて仲介者モデルのシミュレーションを行う。ノードを作るためのデータには被験者 12 人から集めた WWW ブックマークを用い、ノード間の関連度はカテゴリ化近似値によって求める。シミュレーションの設定はノード数を 12 ノード、パス数を 12, 16, 20, 24 の 4 種類とした。これは一つのノードが平均 2~4 のパスを持つことを想定している。各パス数に対し、それぞれ 10 種類の初期パスを乱数を用いて作成する。そして各初期パスに対してシミュレーションを 100 試行を行い、その平均値を結果とした。

この実験設定に基づいて行った実験の結果を以下に示す。図 3 は、収束するまでに平均何ターン要したかを示している。一つのノードが一つのパスを追加し、一つのパスを削除するのが 1 ターンであるから、10 ターンで収束ということは、収束までの過程で 10 個のパスが新たに生成され、10 個のパスが削除されたということである。

収束ターン数は、パス数が増えるにつき増加する傾向にあるが、その増加量は減少傾向にあるため、パス数が多くなって

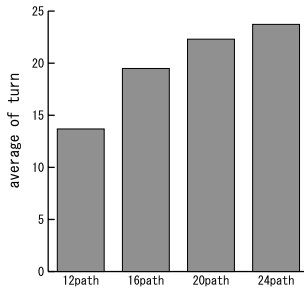


図 3: 平均収束ターン数

も指数的に増大するような問題はないと思われる。

図 4 は、初期状態と収束状態の被覆率および到達率を比較したグラフである。図 4-a は被覆率の比較であるが、パス数で初期値に差が現れているのは、初期パスの作成条件として各ノードが必ず一つはパスを持つという制約があるためと思われる。収束状態で差が現れているのは、パスが多いほど仲介者となるノードが多くなるので、生成可能なパスの数も多くなる。つまりパスの変化に対して柔軟になるといえる、それが収束に時間を要する理由でもあるが、より最適な状態へと遷移する可能性も高いといえる。図 4-b は到達率の比較であるが、被覆率で見られたようなパス数による差があまり現れていない。初期状態ではランダムでパスが生成されるので到達率には差が現れていないと考えられる。

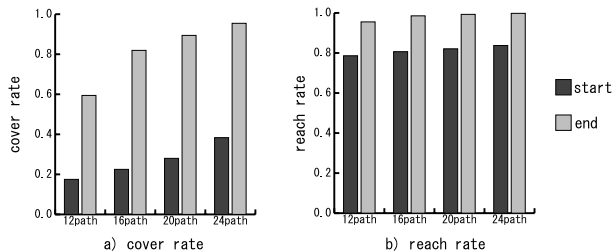


図 4: 評価値の初期値と収束値の比較

図 5 と図 6 は被覆率および到達率の、ターンごとの変化をみたものである。グラフ中の縦線は平均収束ターンを示している。図 5 を見ると変化量という点ではパス数による差はあまり大きくないが、収束値で差が大きく現れていることがわかる。図 6 では変化量による違いはあるが、収束値の差は比較的小さくなっている。また、ある程度の到達率を目標とした場合、例えば 0.9 だとすると、どのようなパス数でも 10 ターン程度でネットワークの再構成ができることがわかる。

これらの結果から、最適なネットワークにどれだけ似ているものを作るかという点では、パス数が少ない場合に問題があることがわかる。だが実際に情報共有を考えた場合、適切な関係に基づくネットワーク、つまり関連度の高いノード同士が結びついたネットワーク、を作ることが重要であるため、仲介者モデルによるネットワークの再構成はパス数が少ない場合でも十分な効果が得られているといえる。

4.2 仮想データを用いた場合

実データを用いた実験では、シミュレーションのノード数が収集したブックマークデータ数に制限される。そこで仮想的にデータを作成し、ノード数が多い場合を想定したシミュレ-

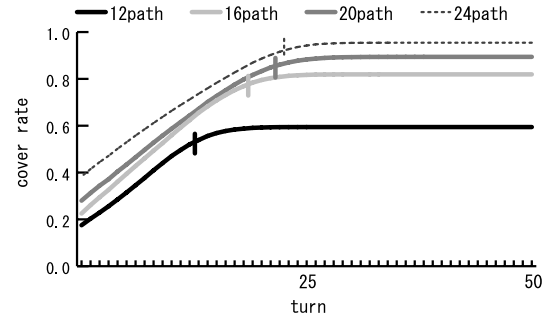


図 5: 被覆率の変化 (実データ)

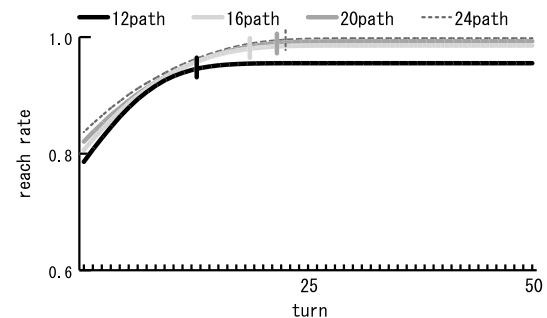


図 6: 到達率の変化 (実データ)

シオン実験を行う。

シミュレーションに用いるデータはノードとノード間の関連度で構成される。仮想データでは、ノードは乱数で作った 10 次元の特徴ベクトルを用い、ノード間の関連度は特徴ベクトルの余弦とした。仮想ノードを構成する 10 次元の特徴ベクトルの要素は 0.0 ~ 1.0 までの乱数を用いて作る。ただし、実データのノード間における関連度の分散値を調べたところ、同じコミュニティであるノード間と、そうでない場合とで関連度の分散値の差が現れた。この結果を踏まえ、ある程度ノード間の関係にも偏りを出すために特徴ベクトルの各要素はそれぞれ 50% の確率で 0.0 であるとした。

実験設定は、ノード数を 12 ~ 120 の 5 種類、パス数をそれぞれノード数の 2 倍 (一つのノードが平均 4 つのパスを持つ) とした。その他の条件は実データを用いた場合と同じである。

実験結果を以下に示す。図中の縦線は平均収束ターンを示している。図 7 は被覆率の変化を示している。ノード数が多くなるほど被覆率は低くなっており、その差は大きい。図 8 は関連度の到達率の変化を示している。同じくノード数の増加と共に到達率は低くなっているが、被覆率と比べて差は小さいことがわかる。図 9 は関連度の平均値の変化を示している。他の二つのグラフと異なり、ノード数が増加すると値も増加している。

パスの被覆率が 0.5 ~ 0.9 付近にばらつくのに対し、関連度の到達率は (今回調べたデータでは) 全て 0.9 以上の結果となっている。さらに、ノード数の増加によって到達率は下がるが、ノード間関連度の平均値そのものはノード数が増加すると共に増加している。これらのことから、ノード間の関連度という点ではノード数の増減による影響は小さいことがわかる。つまり、ノード数が増加することにより、中央集権モデルと似たネットワークを作ることは難しくなるが、ノード間の関係において同等のネットワークを構築することは可能であるとい

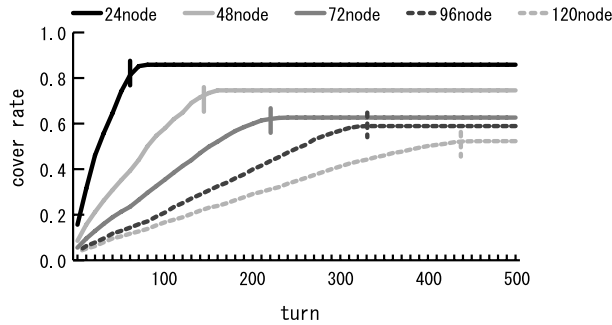


図 7: 被覆率の変化 (仮想データ)

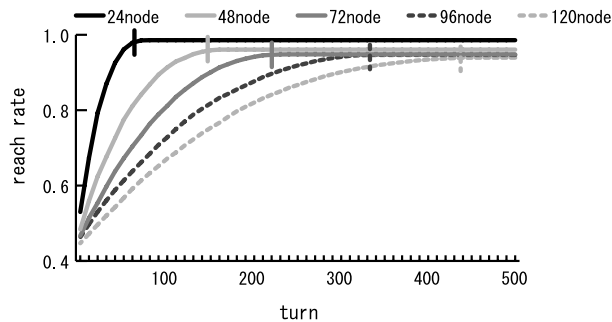


図 8: 到達率の変化 (仮想データ)

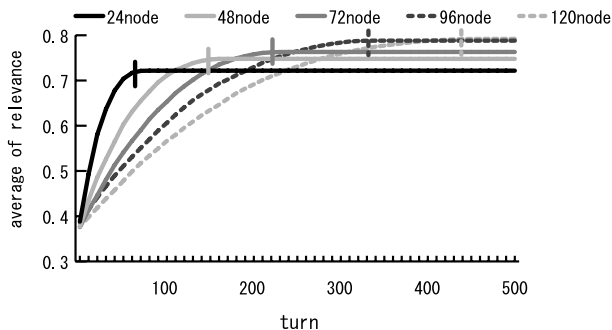


図 9: ノード間関連度の平均値の変化 (仮想データ)

える。

5. まとめと展望

本稿では、情報共有のネットワークを拡張する方法として仲介者モデルを提案し、シミュレーションによる中央集権モデルとの比較実験により有効性を検証した。仲介者モデルでは、個々のノードが自分に隣接したノードに関する情報のみでネットワークを再構築するため、中央集権モデルのように確実に最適解を知ることとはできない。しかし、実データと仮想データを用いたシミュレーション実験の結果、各ノードが単純なルール（よりよいノードとパスを張る）にのっとって仲介を繰り返すことで、中央集権モデルと比べて遜色のない程度に、不完全な関係ネットワークから適切な関係によるネットワークを生成可能であることがわかった。これはノード数、パス数が変化しても可能である。

先述したが、我々は [濱崎 02] において WWW ブックマークを用いて共通話題ネットワークを生成し、個人間の情報共有関係を構築する手法を提案した。今後は、この仲介者モデルを用いて、より大きな共有ネットワークシステムの構築に取り組みたい。

参考文献

- [インターネット協会 01] インターネット協会: インターネット白書 2001. インプレス (2001)
- [濱崎 02] 濱崎雅弘, 武田英明, 松塚健, 谷口雄一郎, 河野恭之, 木戸出正継: Bookmark からの共通話題ネットワークの発見手法の提案とその評価. 人工知能学会誌, Vol.17, No.3, pp.276-284 (掲載予定)
- [吉田仙 01] 吉田仙, 亀井剛次, 服部文夫: Shine: ネットワークコミュニティ支援システムのエージェント指向フレームワーク. 信学技報 AI2001, Vol.101, No.211, pp.9-16 (2001)
- [吉田匡志 01] 吉田匡志, 伊藤雄介, 沼尾正行: 口コミによる分散型情報収集システム. Workshop on Multi-Agent and Cooperative Computation (2001)
- [竹内 01] 竹内亨, 鎌原淳三, 下條真司, 宮原秀夫: ユーザの関連性に基づいた情報伝播モデル. 第 12 回データ工学ワークショップ (2001)