

モバイルエージェント系における 分散制約充足

沼澤 政信 栗原 正仁 能登 正人

本論文では、モバイルエージェント系における分散制約充足について、いつ・どのような目的でエージェントの移動が考えられるのか基本的な考察をする。その中から、実ネットワーク中を流れる総通信量を抑制する目的で移動する場合に焦点を当て、簡単な計算モデルを構成し、計算機シミュレーションにより、このモデルに基づく移動が総通信量に与える効果について定量的な評価をする。分散制約充足アルゴリズムとしては分散 breakout アルゴリズムを仮定し、シミュレーションにより総通信量を測定した。その結果、以下のことがわかった。

1. 総通信量は、一般にリンク数の増加とともに急激に増加し、ピークに達した後急激に減少する。その後、緩やかに増加した後に緩やかに減少する。
2. エージェントサイズが、上限として定義した s_n よりも十分小さいときには、移動したほうが総通信量が概ね削減できる。逆に、エージェントサイズが十分大きいときは、移動により総通信量はかえって増加する。

Distributed Constraint Satisfaction for Mobile Agents
NUMAZAWA Masanobu, 小樽商科大学商学部社会情報
学科, Department of Information and Management
Science, Faculty of Commerce, Otaru University of
Commerce

KURIHARA Masahito, 北海道工業大学工学部情報
デザイン学科, Department of Information Design,
Faculty of Engineering, Hokkaido Institute of
Technology

NOTO Masato, 神奈川大学工学部電気電子情報工学科,
Department of Electrical, Electronics and Informa-
tion Engineering, Faculty of Engineering, Kanagawa
University

3. 総通信量がピークとなる困難な問題に対しては、エージェントサイズがある程度大きくても総通信量を抑制できる。

1 はじめに

近年のネットワーク技術の急速な発達により、コンピュータの利用形態は、単体で使用するというよりも、インターネットや LAN などのネットワークを介して使用する環境になった。そのような分散環境に対応した分散型ソフトウェアを柔軟に構築するためにエージェント技術が注目され、さらに最近ではそれを発展させたモバイルエージェントと呼ばれるホスト間を移動するエージェント技術が有効であると考えられている [1][2][3]。モバイルエージェントは、途中のタスクの内容を保持したまま別のホストに移動してタスクを継続することができるため、ネットワーク上のコンピュータ資源を有効利用することが可能である。よく知られているモバイルエージェントシステムとしては、General Magic 社による Telescript [4]、IBM による Aglets [5]、東芝による Plangent [6] などがある。

このような状況に対応して人工知能の分野においても、知的エージェントの実現基盤の一つとして知られている制約充足技術が、分散制約充足技術の研究に移行してきている [7]。そこでさらに、分散型ソフトウェアの構築を考慮し、分散制約充足の計算モデルもモバイルエージェントに対応するように拡張することは自然な研究の流れである。しかし、そのような研究例はまだみられない。エージェントが移動しようがしまいが、いわゆる位置透過性があるシステムにおい

では、エージェント間のメッセージ交換の媒体が異なるだけで、メッセージの内容や交換のプロトコルが異なるわけではないので、抽象的あるいは論理的な意味では両者に差異はない。しかし、実装レベルでは、実ネットワーク中を流れる総通信量、延いては、実行効率の面で大きな違いとなりうる。

本論文ではそのような研究の第一歩として、まず、いつ・どのような目的でエージェントの移動が考えられるのか基本的な考察をし、その中から、実ネットワーク中を流れる総通信量を抑制する目的で移動するケースに焦点を当て、簡単な計算モデルを構成する。そして、計算機シミュレーションにより、このモデルに基づく移動が総通信量に与える効果について定量的な評価をする。

2 章では、モバイルエージェントと分散制約充足について概説し、その特徴を述べる。3 章では、考える種々の移動方式を簡単に考察した後、通信量を移動のトリガとする計算モデルを導入し、その評価のためのシミュレーション実験について述べる。4 章では、実験結果に基づき、この計算モデルの特徴と有効性を議論する。5 章では、まとめと今後の課題を述べる。

2 モバイルエージェントと分散制約充足

2.1 モバイルエージェントシステム

モバイルエージェントは、コンピュータ間の自律的移動能力を持つプログラムであり、エージェント自体が利用者の代理人としてネットワーク上を自律的に移動しながら特定のタスクを遂行する。そのため、利用者はネットワークに逐一接続して作業する必要がない。通常のエージェントでも、それぞれのホスト間で遠隔通信機能を使ってホスト間でのメッセージ通信が可能であるが、ネットワークが不安定で間欠的であったり、低速もしくは費用の高い通信路を使用しなければいけないといったケースが考えられるなど問題点も多い。なお、モバイルエージェントは AI 技術の関連で議論されている他のエージェント技術とは異なり、必ずしもインテリジェンスを持たないことが多いが、本研究で扱うモバイルエージェントは、分散制約充足という限定ながらも AI の基礎的な計算処理を能動的に実行し、移動先もそれ自身により選択し、移動する

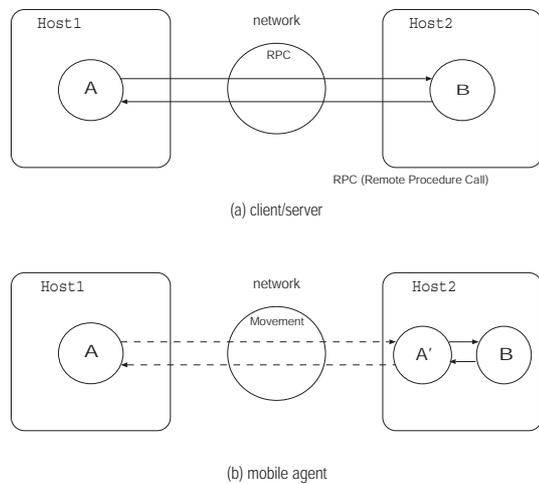


図 1 処理形態の違い

ことができるように設計されている。

モバイルエージェントを用いてコード転送し、それぞれのホストでその転送されたコードを実行するためには、セキュリティ保護、障害対策、トラフィック増加、管理の手間の増加といった問題が複雑化する。しかしながら、不安定で間欠的な通信路を常時使用する必要がなく、通信コストの削減が可能となるなどメリットが大きい。その理由は、従来のクライアント/サーバ方式とモバイルエージェント方式の二つの処理形態を比べることにより、容易に理解することができる。例えば、二つの異なるホストにいるエージェント間で 1000 回の要求を発行し、それに対する応答があるとき、クライアント/サーバ方式では、1000 回の要求と応答が Host1 と Host2 間のネットワーク上を往復することになる(図 1(a))。一方、モバイルエージェント方式(図 1(b))では、エージェントが 1 往復するだけで 1000 回の要求と応答は、Host2 上でローカルに行うだけで済む。この 1000 回が、非常に費用の高い通信路であったり、要求の回数が 1 万回、10 万回と増えていった場合を想定すると、移動して処理を行う効果は高いと想像できる。

また、この処理形態の違いにより、もう一つ重要な結果が生じる。クライアント/サーバ方式では、毎回の要求と応答がネットワーク上を行き来するため、ホスト間のコネクションは、常時維持しなければならない

いのに対して、モバイルエージェント方式では、エージェントが移動するときだけコネクションを確立すればよい。これは、電話回線でモデムを利用する場合の通信や、無線を利用しているためコネクションが不安定な携帯電話、移動体通信などの通信形態にとっては非常に有利である。

2.2 分散制約充足

制約充足問題 (Constraint Satisfaction Problem: CSP) [8] は、離散値をとるいくつかの変数に割り当てられる値の組合せのうち、与えられた制約をすべて満たす組合せを発見する探索問題である。分散制約充足問題 [9][10] は、制約充足問題の変数と制約が複数のエージェントに分散された問題とみなすことができ、さまざまな問題が分散制約充足問題として定式化されている [11][12][13][14]。制約充足問題では、初期状態において一つのアルゴリズムが全体の問題を見渡すことができるのに対し、分散制約充足問題の特徴は、初期状態はもちろんのこと、任意の時点においてそれぞれのエージェントは全体の状態を知らないことである。このため、各エージェントがどのような手順で、どのような情報を交換し合えば、全体として制約を満たす解が発見できるかが問題となる。

制約充足問題を解くアルゴリズムは、厳密解法である木探索アルゴリズムおよび近似解法である反復改善型アルゴリズムの二つに大きく分類される [15]。前者は整合性をとりながら部分解を拡張して探索木をたどり完全な解を求める。後者は変数に制約を満足しない値を割り当てた状態から制約違反を局所的に改善していくものであり、横尾、平山 [16] によって提案された反復改善型分散制約充足アルゴリズム (分散 breakout) が効果的な状態空間探索法の一つとして注目されている。

3 モバイルエージェント間の分散制約充足

本章では、モバイルエージェントと分散制約充足の二つの技術を結びつけてモバイルエージェント間での制約充足方式を提案する。はじめに、考えうるいくつかの移動方式を列挙し、次に、そのうちの一つである通信量に基づく移動方式を詳細化した計算モデル

を構築した後、それを評価するためのシミュレーション実験の条件を設定する。

3.1 種々の移動方式

現在、モバイルエージェントを分散制約充足に導入した例はまだないので、研究の第一歩として、いつ・どのような目的で移動したらよいかを検討する必要がある。以下に、タイミングと移動の目的として考えられる基本的な移動方式を述べる。

1. 時間による移動

一定時間、他のエージェントから何のメッセージ応答もないときに移動する。現在エージェントが存在しているホストが過負荷である、または何らかの理由でプラットフォームの状態が不具合であるかもしれないと考えられるときに移動するため、本移動方式は、ネットワークで問題を解くときのロバスト性を高める効果が期待できると考えられる。

2. 準局所最適時の移動

エージェントが準局所最適状態になったときに移動する。準局所最適の状態を抜け出すには多くのメッセージ交換が必要になる場合がある。本移動方式は、そのような場合、その負荷を軽減し得る可能性を持つ。

3. メタレベルからの移動要求による移動

アプリケーションに依存して、自分が求めるエージェントに連絡をとり移動する。本移動方式は、タスクとして分散制約充足だけを行う研究レベルのエージェントシステムには付加されるものではない。しかし、このタスク機能を一部として包含する現実のアプリケーションにおいては「メタレベル」での要求によって移動する必要が生ずることもあり得ると考えられる。

4. 計算負荷による移動

自分が存在しているホストマシンの負荷が所定の値を超えたときに移動する。本移動方式では、資源の効率的な利用が期待できると考えられる。

5. 通信量による移動

メッセージの累積量が所定の量を超えたときに、一方のエージェントが他方のエージェントの

いるホストマシンへ移動する。本移動方式は、非常に汎用的な単純な方法であり、通信コストの削減が期待できると考えられる。

以下では、「通信量による移動」に基づく考え方で具体的に計算モデルを構築し、基盤となる分散制約充足アルゴリズムとして分散 breakout アルゴリズム [16] を採用してシミュレーションを実施し、移動が通信量に与える効果について評価する。

3.2 通信量をしきい値とする移動方式

任意の分散制約充足アルゴリズムが与えられたとする。そのアルゴリズムは、必ずエージェントが他のエージェントとメッセージ交換しながら制約充足のための計算を実行しているようにモデル化されていると仮定する。我々は、各エージェント i が現在自分のいるホストと異なるホストにいるエージェント j とメッセージ交換 (送信, 受信) する部分のプログラムコードに手を入れ、エージェントにこれまでの通信量の累積を把握させる。そして、それがあるしきい値を超えれば、エージェント i が他のホストに移動して、その後本来の計算を続行するような計算モデルを構築したい。そのような制限の下でもさらにいくつかのバリエーションが考えられるが、本論文ではこれから述べるような最も単純なもの一つを考察の対象としたい。

エージェント間で交換するメッセージ 1 つあたりのサイズ (メッセージサイズ) は簡単のため固定されていると仮定し、それを 1 と正規化して通信量の単位とする。エージェント i について、

q_{ij} : エージェント j とのメッセージ累積量

Q_i : 総メッセージ累積量

C_i : 総通信量

t_i : しきい値

とそれぞれ表し、初期値の設定は、 $q_{ij} \leftarrow 0$, $Q_i \leftarrow 0$, $C_i \leftarrow 0$, $t_i \leftarrow$ エージェントサイズ とする。エージェントサイズは、エージェントの大きさ (すなわち、プログラムコードおよびエージェントの状態を表すデータの量) であり、エージェントが一回移動するのにかかる通信量でもある。簡単のために、これも固定されていると仮定する。移動条件の初期しきい値は、エ

ージェントが移動するための初期値である。直観的に、メッセージ通信量がエージェントサイズより相対的に小さいうちは移動することによってかえって大きな通信量の増加をまねくので、しきい値の考える最小の値という意味で初期値をエージェントサイズと等しくした。以上の準備のもとで、エージェント i がエージェント j とメッセージ交換したときに実行すべき処理は以下の手順に従う。

1. $q_{ij} \leftarrow q_{ij} +$ メッセージサイズ;
2. $Q_i \leftarrow Q_i +$ メッセージサイズ;
3. $C_i \leftarrow C_i +$ メッセージサイズ;
4. もし、 $Q_i > t_i$ ならば以下の処理 (4.1 ~ 4.4) をする;
 - 4.1 $j \leftarrow \max_j \{q_{ij}\}$ を与える j ;
 - 4.2 エージェント j のいるホストへ移動する;
 - 4.3 $C_i \leftarrow C_i +$ エージェントサイズ;
 - 4.4 $t_i \leftarrow 2t_i$;
5. 分散制約充足アルゴリズムの処理;

最初に、しきい値を満たす前は、各エージェント間でメッセージ交換を行いながら、メッセージサイズを q_{ij} , Q_i , C_i に累積していく。その後、メッセージ累積量がしきい値を超えたときに、エージェントは、メッセージ交換したすべてのエージェントの中で個々のメッセージ量が最も多いエージェントのいるホストに移動する。このとき、総通信量にエージェントサイズを累積する。なお、メッセージ量の最大値をもつエージェントが複数の場合は、その中から非決定的に選択し移動先を決める。しきい値は移動するたびに 2 倍にしていく。これは、データ構造の分野で配列やハッシュ表のサイズを動的に拡張する際にも用いられる考え方の一つとほぼ同様で、単純ながらも自然な考えである。

本研究では、基盤となる分散制約充足アルゴリズムとして分散 breakout アルゴリズムを採用するが、システムを実ネットワーク上に実装するのではなく、性能の事前評価を目的とし、1 台のマシン上でのシミュ

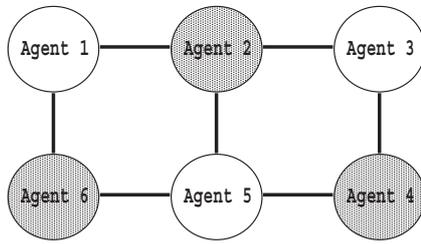


図 2 分散グラフ色塗り問題の例

レーション・システムとして実現している。

3.3 シミュレーション実験

分散制約充足問題の代表的な例題として分散グラフ色塗り問題がある。グラフ色塗り問題とは、任意のグラフと色数が与えられ、グラフ上の隣り合うノードを異なる色で塗るものである。このグラフのノードをエージェントとみなすことで、分散グラフ色塗り問題のエージェントネットワークが得られる。色数 2, エージェント数 6, リンク数 7 の分散グラフ色塗り問題の例および解を図 2 に示す。

3 色の分散グラフ色塗り問題をランダムに発生させ、メッセージサイズを 1 として表 1 に示す条件で、通信量に関する移動のシミュレーション実験を行う。

表 1 実験内容

	エージェント数	リンク数	エージェントサイズ
実験 1	30	30 ~ 270	10, 70, 300, ∞
実験 2	60	60 ~ 1100	50, 300, 500, ∞
実験 3	90	90 ~ 2600	100, 300, 500, ∞
実験 4	120	150 ~ 4500	100, 400, 600, ∞
実験 5	150	150 ~ 7000	100, 500, 900, ∞

実験では、リンク密度 0~100%の範囲でリンクの本数をパラメータとして設定し、各エージェント数 (30~150) に対してリンク数を変化させていき、各エージェントサイズでの通信量をそれぞれ求める。リンク数ごとに 100 回ずつ問題を解き、そのときの平均値をとる。なお、この問題では、リンク密度 25%前後付近の値のところでもっとも難しい問題が存在することが知られている [17][18]。便宜上、エージェントを移動させない (固定した) ケースをエージェントサイ

ズ ∞ で表す。

4 実験結果と考察

シミュレーション実験の結果として実験 1, 実験 3 および実験 5 (エージェント数が 30, 90 および 150) のときの総通信量のグラフをそれぞれ図 3 ~ 図 5 に示す。

はじめに、エージェント数が十分多い実験 (図 4, 図 5) において、比較の基準となるエージェントサイズが ∞ (すなわち移動しないケース) のグラフの特徴を確認する。リンク数が増加していくと、ある狭い範囲で急激に総通信量が増加し、そのピークを超えると急激に減少している。これはいわゆる「相転移」などと呼ばれる現象としてよく知られている [19]。その後、総通信量は緩やかに増加に転じた後、再度、緩やかに減少する。

次に、このグラフを他の (移動する場合の) エージェントサイズのケースと比べてみる。エージェントサイズが十分小さい場合には、リンク数にかかわらず、移動しない (固定した) ときの総通信量よりも移動したときの総通信量の方が概ね削減できていることがわかる。逆に、直観的にも明らかなことだが、エージェントサイズがあまりにも大きい場合は、移動することによりその移動に必要な大きな通信量が加算されるため、総通信量は移動しないときよりもかえって大きくなってしまふ。さらに、重要なこととして、リンク数が 250 (図 4) 付近 (図 5 では 400 付近) の最も難しい問題による通信量のピークが、いずれのエージェントサイズの場合も、 ∞ のケースより低く抑えられていることがわかる。これは、今回設定した移動基準が、過大な通信量を抑制するように働くことによるものと考えられる。

エージェントサイズがある程度小さければ移動の効果があり、逆にある程度を超えて大きければ移動は逆効果となることがわかった。そこで、移動が効果的でありうるようなエージェントサイズの上限を求めてみる。エージェント数に加えてリンク数が既知として与えられたときは、エージェントサイズをパラメータとして今回のような実験をして、移動しないケースと比較すれば上限は自明な方法で実験的に求めら

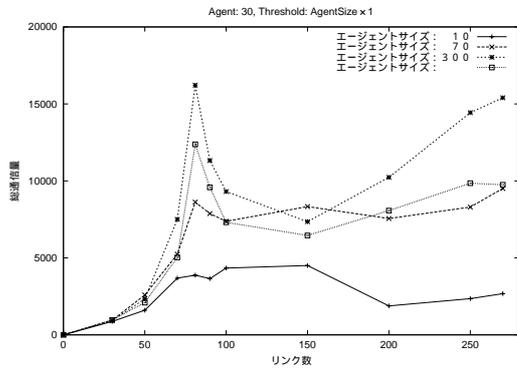


図 3 実験 1 (エージェント数 : 30)

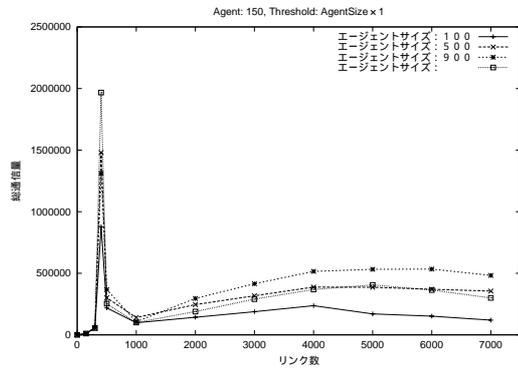


図 5 実験 5 (エージェント数 : 150)

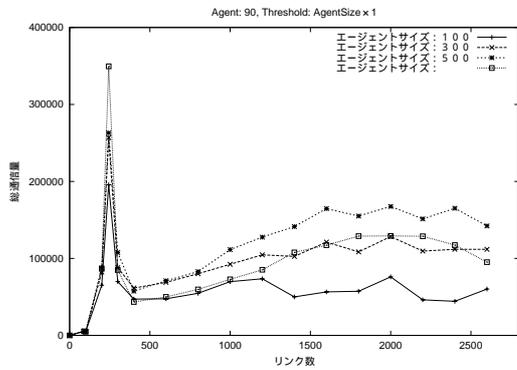


図 4 実験 3 (エージェント数 : 90)

ただし、 $L = n(n-1)/2$ はリンク数の上限値である。今回の実験から求めた $D_{n,s}$ の概算値の一部を表 2 に示す。

表 2 $D_{n,s}$ の計算結果

s	エージェント数 n				
	30	60	90	120	150
10	-4220	-21800	-61500	—	—
50	-1520	-14500	-51400	-124000	—
100	1180	-7270	-38600	-107000	—
200	3230	5640	-16700	-71900	-94200
300	—	14000	2730	-35100	-59400
400	—	18500	18400	-5380	-28900
500	—	—	29500	26200	17300

れる。しかし、分散環境においては一般にエージェントは問題の全体を知らないで、エージェント数のみは知っているとして、リンク数を知らない場合の上限を定義してその概数を試算しておくことが興味深い。そのようなヒューリスティクスはあまり厳密に論じても意味が薄いので、本論文では非常に単純なものだけを以下に述べる。

はじめに、エージェント数が n 、エージェントサイズが s のときの総通信量をリンク数 ℓ の関数として $C_{n,s}(\ell)$ で表す(図 3 ~ 図 5 の一本一本の折れ線がそのような関数の実例を表している)。そこで、 $s = \infty$ の関数との平均的な差を以下の式で定める。

$$D_{n,s} = \frac{1}{L} \sum_{\ell=1}^L [C_{n,s}(\ell) - C_{n,\infty}(\ell)]$$

表 2 からわかるように、エージェント数を固定したとき、エージェントサイズ s が十分小さいときは $D_{n,s} < 0$ であることから、移動が効果的であるといえる。逆に、 s が十分大きいときは $D_{n,s} > 0$ であることから、移動は効果的でないと見える。そこで、先ほど論じた上限として、 $D_{n,s} < 0$ を満たす最大の s の値を採用する。その値は n 毎に定まることから、 s_n と表すことにする。表 2 から補間によって求めた s_n の近似値を表 3 および図 6 に示す。

表 3 s_n の近似値

n	30	60	90	120	150
s_n	75	156	286	417	463

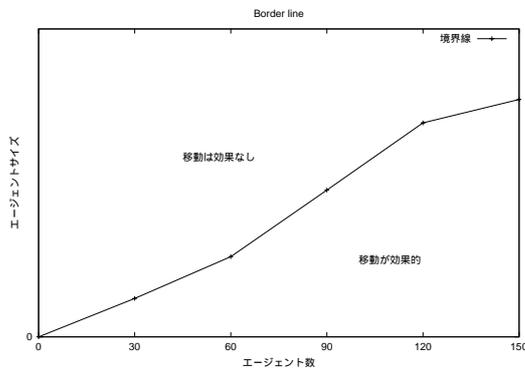


図 6 移動の境界線

きわめて大まかな指針として、リンク数が未知の場合、メッセージサイズを 1 としたときのエージェントサイズが、エージェント数のおよそ 3 倍程度以内のときは、移動による通信量の削減が期待できる。エージェントを実装するクラスファイルを各ホストに事前に置いておくなど、実装上の工夫でこの条件を満たせば、通信量の削減という理由だけで、エージェントを移動させる正当な理由となり得る。

ただし、エージェントサイズがこの指針より大きいときに本方式が役に立たないというわけではない。リンク数が既知で、特に、計算量的に「難しい問題」を発生させるようなリンク数のときには、多少エージェントサイズが大きくても、それを大きく上回るメッセージ量を削減できるので効果が大きい。また、3.1 節で述べたような種々の理由で移動そのものが効果を発揮することもあり得る。

5 おわりに

本研究では、モバイルエージェント系における分散制約充足について、いつ・どのような目的でエージェントの移動が考えられるのか基本的な考察をし、その中から、総通信量を抑制する目的で移動するケースに焦点を当て、簡単な計算モデルを構成した。次に、分散制約充足アルゴリズムとして分散 breakout アルゴリズムを仮定し、シミュレーションにより総通信量を測定した。その結果、以下のことが確認できた。

1. 総通信量は、一般にリンク数の増加とともに急

激に増加し、ピークに達した後急激に減少する。その後、緩やかに増加した後に緩やかに減少する。

2. エージェントサイズが、上限として定義した s_n よりも十分小さいときには、移動したほうが総通信量が概ね削減できる。逆にエージェントサイズが十分大きいときは、移動により総通信量はかえって増加する。
3. 総通信量がピークとなる困難な問題に対しては、エージェントサイズがある程度大きくても総通信量を抑制できる。

今後の課題は以下のとおりである。

- 分散 breakout アルゴリズム以外の分散アルゴリズムの場合について検討する。
- 3.1 節で示したような通信量以外の移動方法について検討する。
- 1 つのホストに多数のエージェントが集中すると、CPU タイムが各エージェントに分配されて 1 エージェント当りの計算スピードが減少する問題があるので、それを解決するモデルを考えたい。つまり、上記の 2 つの検討事項を踏まえた上で、複合的要因により移動を決定する方法の検討を行う。
- 実際のモバイルエージェント系に実装を行い、シミュレーションとの比較を行う。

謝辞 本研究の一部は、文部科学省科学研究費(課題番号 12780246)の補助によって行われた。

参考文献

- [1] 本位田真一, 飯島正, 大須賀昭彦: エージェント技術, 共立出版 (1999).
- [2] 長尾確編著: エージェントテクノロジー最前線, 共立出版 (2000).
- [3] 新谷虎松, 大園忠親, 福田直樹: モバイルエージェントの応用-マルチエージェントシステムのためのモビリティの利用-, 人工知能学会誌, Vol. 16, No. 4, pp. 488-493 (2001).
- [4] White, J. E.: *Mobile Agents, Software Agents* (Bradshaw, J. M.(ed.)), AAAI Press/The MIT Press, chapter 19, pp. 437-472 (1997).
- [5] Lange, D. B. and Ohshima, M.: *Programming and Deploying Java Mobile Agents with Aglets*, Addison-Wesley (1998).

- [6] Ohsuga, A., Nagai, Y. and Irie, Y.: Plangent: An Approach to Making Mobile Agents Intelligent, *IEEE Internet Computing*, Vol. 1, No. 4, pp. 50-57 (1997).
- [7] 横尾真, 平山勝敏: CSP の新しい展開: 分散/動的/不完全 CSP, *人工知能学会誌*, Vol. 12, No. 3, pp. 33-41 (1997).
- [8] Mackworth, A. K.: Constraint Satisfaction, *Encyclopedia of Artificial Intelligence* (Shapiro, S. C.(ed.)), Wiley-Interscience Publication, New York, pp. 285-293 (1992).
- [9] 平山勝敏, 横尾真: 分散不完全制約充足問題, *人工知能学会誌*, Vol. 14, No. 4, pp. 60-69 (1999).
- [10] Yokoo, M. and Hirayama, K.: Algorithms for Distributed Constraint Satisfaction: A Review, *Autonomous Agents and Multi-Agent Systems*, Vol. 3, No. 2, pp. 198-212 (2000).
- [11] Conry, S. E., Kuwabara, K., Lesser, V. R. and Meyer, R. A.: Multistage Negotiation for Distributed Constraint Satisfaction, *IEEE Transactions on Systems, Man and Cybernetics*, Vol. 21, No. 6, pp. 1462-1477 (1991).
- [12] Huhns, M. N. and Bridgeland, D. M.: Multi-agent Truth Maintenance, *IEEE Transactions on Systems, Man and Cybernetics*, Vol. 21, No. 6, pp. 1437-1445 (1991).
- [13] Lesser, V. R. and Corkill, D. D.: The Distributed Vehicle Monitoring Testbed: A Tool for Investigating Distributed Problem Solving Networks, *AI Magazine*, Vol. 4, No. 3, pp. 15-33 (1983).
- [14] Sycara, K. P., Roth, S., Sandeh, N. and Fox, M.: Distributed Constrained Heuristic Search, *IEEE Transactions on Systems, Man and Cybernetics*, Vol. 21, No. 6, pp. 1446-1461 (1991).
- [15] Morris, P.: The Breakout Method for Escaping from Local Minimum, *Proc. of the Eleventh National Conference on Artificial Intelligence*, pp. 40-45 (1993).
- [16] 横尾真, 平山勝敏: 分散 breakout : 反復改善型分散制約充足アルゴリズム, *情報処理学会論文誌*, Vol. 39, No. 6, pp. 1889-1897 (1998).
- [17] Cheeseman, P., Kanefsky, B. and Taylor, W.: Where the Really Hard Problems Are, *Proc. of 12th International Joint Conference on Artificial Intelligence*, pp. 331-337 (1991).
- [18] Minton, S., Johnston, M. D., Philips, A. B. and Laird, P.: Minimizing Conflicts: A Heuristic Repair Method for Constraint Satisfaction and Scheduling Problems, *Artificial Intelligence*, Vol. 58, pp. 161-205 (1992).
- [19] Hogg, T., Huberman, B. A. and Williams, C. P.: Phase Transitions and the Search Problem, *Artificial Intelligence*, Vol. 81, pp. 1-15 (1996).

ユーザ選好に基づく分散コンテンツキャッシュポリシー

中塚 康介 八槇 博史

Abstract.

To deliver multimedia contents efficiently, a content delivery network (CDN) is used in the Internet. CDN has caches in points-of-presence of internet service providers and stores multimedia contents actively. In CDN, it is important to choose contents to cache, because the caching policy is important to effectiveness.

In this paper, we formulate the preference of end-users and describe the mechanism for helping constructing caching policy. Our approach is based on market mechanism where producer agents and consumer agents negotiate and trade resource. We believe that it not only reaches effective resource allocation but also matches with business model for CDN.

1 はじめに

現在, web が普及し, その上でマルチメディアコンテンツが流通しているが, 音声や動画などのマルチメディアコンテンツは, 帯域幅などのネットワーク資源を多く使用し, ネットワークや web サーバに大きな負荷をもたらすことが問題となっている. この問題の解決のために, インターネットサービスプロバイダ

構内にキャッシュを設置し, コンテンツのキャッシュを行うことによって, ネットワークや web サーバへの負荷を減少させるコンテンツデリバリネットワーク (Content Delivery Network; CDN) が利用されている.

コンテンツデリバリネットワークでは, 各キャッシュにどのようなコンテンツをキャッシュするかは, コンテンツ配信者の決定に委ねられている. 現在コンテンツデリバリネットワークは大規模なコンサートのライブ中継や, 社内教育などの一定の利用が見込まれるコンテンツに対して適用されており効果を上げているが, キャッシュコンテンツの選択にユーザの個別の選好が集約できず, エンドユーザの要求に応じてキャッシュするコンテンツを選択・配信し, コストを回収するモデルが求められている.

本稿では, この問題に対して, エンドユーザの潜在的な選好 (preference) のモデルを構築し, 収集したエンドユーザの選好を反映しながらコンテンツ配信に対する無駄のないネットワーク資源の利用, 及び, ネットワーク事業者, コンテンツ配信者, エンドユーザのコストの配分への応用を考えたキャッシュポリシーを提案する. これにより, 効率的なネットワーク資源の利用をはかるとともに, コンテンツ配信におけるビジネスモデルへの応用が考えられコンテンツ配信に貢献するものであると考えている.

本稿で提案するモデルは, これまで研究・開発が行われている分散キャッシュ技術により構築されたコンテンツデリバリネットワークの中で, ネットワーク事業者, コンテンツ配信者, 及び, エンドユーザの各

Kosuke NAKATSUKA, 京都大学 大学院 情報学研究科 社会情報学専攻, Department of Social Informatics, Kyoto University

Hirofumi YAMAKI, 京都大学 大学院 情報学研究科 社会情報学専攻, Department of Social Informatics, Kyoto University

主体のコスト負担，特にエンドユーザに対する課金を含めて，キャッシュの置くコンテンツの選択を支援することを目標としており，従来のキャッシュ技術と共に，コンテンツデリバリネットワークの効率的，かつ，持続的な運用を図るものである．

2 コンテンツデリバリネットワーク

コンテンツデリバリネットワークは，図 1 を一例とする構造を持つ仮想的なネットワークである．

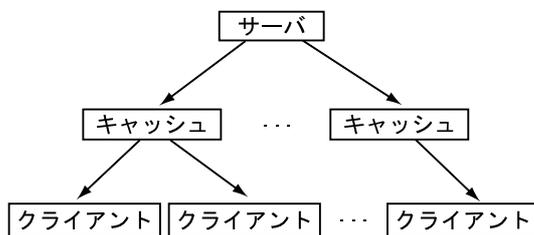


図 1 CDN のアーキテクチャ

図 1 において，サーバはテキスト文書，画像，音声，動画などのコンテンツが保存されるホストであり，クライアントは，コンテンツを利用するエンドユーザを表している．

従来のクライアント/サーバモデルでは，サーバにクライアントが直接要求を出してデータを入手するものであり，サーバの負荷が高くなる，また，ネットワークにより大きな帯域幅を要するなどの問題があった．

これに対して，CDN では，サーバとクライアントの間にキャッシュを導入することにより，負荷をサーバとキャッシュで分散し，必要な帯域幅を減少させている．CDN は，以下のように動作する．

1. インターネットサービスプロバイダの POP^{†1} にキャッシュを設置する．
2. キャッシュは，サーバからデータをコピーし，クライアントからの要求に備える．どのデータをコピーするかは，管理者が設定として与える．
3. クライアントからの要求はサーバに対して行わ

†1 Point-of-Presence，サービスプロバイダへのアクセスポイントを意味する

れるが，実際にはその要求は，経路上に存在するキャッシュが途中で横取り (intercept) する．要求を横取りしたキャッシュは，“どこからのアクセスか”，“何を要求しているか”を調べ，ユーザから近く負荷の低いサーバへの要求振り分けや，データがキャッシュされている場合は，サーバにかわり，そのデータの送信を行う．

クライアントからの要求は経路上で横取りされるため，web ブラウザなどを変更する必要はなく，透過的に CDN を用いることができる．

従来 CDN に類似する技術としてプロキシ型のキャッシュサーバが用いられてきたが，ユーザからの要求によってキャッシュがはじめて行われ，キャッシュを行うポリシーをサービス提供者が設定することができないという問題や，

- 要求は多くのコンテンツに対して行われ，一部の人気のあるコンテンツのみをキャッシュすることは，ある程度有効ではあるが，十分ではない
- web ページのアクセス頻度とページサイズには弱い相関しかない
- web ページのアクセス頻度と更新頻度には弱い相関しかない

といった問題^[1]から，ポリシー設定が困難であり，また，有効性にも限界がある．これに対して，コンテンツデリバリネットワークでは，サーバとクライアントの間にキャッシュを導入し，このキャッシュにサーバから特定のコンテンツを前もってコピーしておき，クライアントからの要求に備えることで，負荷を分散し，配信に必要な資源を減少させている．

2.1 CDN の利点

CDN は，インターネットサービスプロバイダによる配信など，コンテンツ配信者が明らかであり，コンテンツ利用者が限られている，あるいは，コンテンツ利用が十分に見込まれる場合のように，単一のポリシーが定義できる場合の負荷分散・高速配信に有効である．このような条件を満たすものとして，各地に支社のある企業での企業内教育や，コンサートのライブ配信など，特定のイベントにおけるストリーミング配信などがあり，負荷分散が実現されている．

2.2 CDN における問題

CDN では、どのコンテンツをどのキャッシュにコピーするかを決定する必要がある。現在の CDN では、キャッシュ技術については、[2] にあげられるようなキャッシュアルゴリズムが適用され、効果を上げている。一方で、こういったコンテンツをコピーするかのポリシーについては、CDN プロバイダがコンテンツプロバイダと契約し、その契約に基づいてコンテンツを優先的にキャッシュするようになっており、各エンドユーザの要求に応じてコンテンツを配信し、配信についてのコンテンツやネットワークのコストを回収するモデルとはなっていない。このため、

- コンテンツプロバイダにとっては、コンテンツプロバイダが提供するコンテンツに対し、エンドユーザの利用を見通して、CDN を運用する必要があること、
- CDN プロバイダにとっては、ネットワークや計算機等の性能の向上に対して配信設備を維持し、継続して事業を行う必要があること、
- エンドユーザにとっては、高速なコンテンツ配信が望ましい

といった各主体の要求を満たすようなビジネスモデルの構築とそれに対応したキャッシュポリシーが必要となっている。この問題に対し、エンドユーザの嗜好を集約し

- エンドユーザ、コンテンツプロバイダ、ネットワークプロバイダの各主体にとって効率的なコンテンツ選択
- 各主体に対するコスト負担の決定

の 2 点が可能なポリシーの構築を目標として、市場機構による最適化手法を適用したキャッシュコンテンツ決定手法を提案する。

3 キャッシュコンテンツ決定プロトコル

キャッシュポリシーは、エンドユーザとキャッシュ間で効用情報をやりとりするプロトコル、及び、得られた効用情報をもとに適切なコンテンツを決定するアルゴリズムからなる。本章では、キャッシュポリシーを運用するためのエンドユーザとキャッシュの間のプロトコル記述する。

エンドユーザは、図 2 に示すように最寄りのインターネットサービスプロバイダの POP に設置されたキャッシュに対して接続されており、サーバに対する要求は、このキャッシュを経由するものとする。

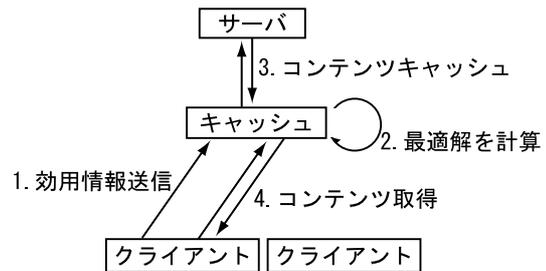


図 2 エンドユーザ・キャッシュ間プロトコル

以下、キャッシュコンテンツ決定のプロトコルを記述する。

1. エンドユーザはどのコンテンツがどの程度望ましいのかを表す効用情報を効用関数として送信する。効用関数はコンテンツに対して十分小さく、コンテンツ配信そのものに影響を与えないと考えられる。
2. キャッシュは、キャッシュの CPU 能力・ストレージ・ネットワークの帯域幅などの制約条件の中で集められた効用関数を最大化するようなコンテンツ集合を計算する。
3. キャッシュは得られたコンテンツ集合のコンテンツをキャッシュプロトコルに従い取得する。
4. エンドユーザはコンテンツに対する要求を行いコンテンツを取得する。

以上のプロトコルにより、クライアントの効用情報が収集され、最適なコンテンツ集合が決定される。

4 キャッシュ・エンドユーザ間のキャッシュ決定アルゴリズム

前節のように集められた効用情報は、映画よりも音楽が望ましいといったコンテンツ種別毎の嗜好が記述されているが、ストレージや帯域幅など資源に制約がある中で、このような嗜好を反映した資源割当てが可能な手法に経済学の知見を応用した市場機構による

資源割当てがある [3]。市場機構による資源割当ては、ストレージや帯域幅など資源が限られた状況でも、利用者間の要求の衝突を調停しながら利用者の選好に応じた資源の分配を可能にする最適化手法であり、ネットワークの QoS 制御などに適用されている [4]。

コンテンツデリバリネットワークでは、ネットワーク事業者、コンテンツ提供者、エンドユーザの各主体が存在し、どの主体がどのように運用コストを負担するのも問題となるが、配信ポリシーに価格機構を用いた市場機構による資源割当てを応用することにより、コストの問題を系に含めて最適化問題として記述することができ、また、流通するコンテンツの価値を考えた応用も可能であると考えられる。

以下、市場機構分散資源割当てに基づいて、エンドユーザの効用関数の記述、および、コンテンツ決定アルゴリズムについて記述する。

4.1 市場モデル

消費者エージェント・生産者エージェントと消費者エージェントの支払う対価、生産者エージェントの生産するキャッシュサービスとの関係を図 3 に示す。

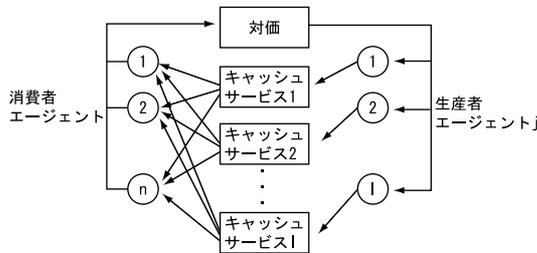


図 3 CDN における市場モデル

図 3 において、左側の円は消費者エージェントを示し、右側の円は生産者エージェントを示す。また、図中の長方形は各エージェント間で流通する財を表す。

流通するコンテンツは、直接提供されるのではなく、ネットワークやストレージなどを含めて、コンテンツキャッシュサービスとしてエンドユーザに提供される。消費者エージェントは、対応するエンドユーザの選好情報をエンドユーザから取得し、キャッシュサービスに対する需要を決定する。生産者エージェン

トは、複数あるキャッシュのうちの 1 つであり、このキャッシュが、エンドユーザの対価を用いてストレージやネットワークなどキャッシュの各資源からコンテンツキャッシュサービスを生成する。

以下、エンドユーザを代表する消費者エージェントと、キャッシュを代表する生産者エージェントの行動について記述する。

4.2 消費者エージェント

エンドユーザの効用関数の例として、CES 型の効用関数をあげる。ユーザ数を n とし、エンドユーザ k ($1 \leq k \leq n$) に対応して消費者エージェント k がエンドユーザの効用情報を持つとする。消費者エージェント k の CES 効用関数 u_k は以下のように表される。

$$u_k(x_{1k}, \dots, x_{lk}, e_k) = \left(\sum_{i=1}^l a_{ik} x_{ik}^{\frac{\rho}{\rho-1}} + a_{ek} e_k^{\frac{\rho}{\rho-1}} \right)^{\frac{\rho-1}{\rho}} \quad (1)$$

式 (1) において、 x_{ik} は、消費者エージェント k に割り当てられる資源 i の量、 l は資源の種類の数、 e_k はキャッシュサービス利用のために用いる対価としての初期財を表す。

すなわち、ネットワーク上には映画や音楽などの l 種のコンテンツがあり、その各コンテンツの中で、エンドユーザが x_{ik} だけの個別のコンテンツを要求すると理解できる。 a_{ik} は資源 i が消費者エージェント k にとってどの程度価値があるのかを示し、 ρ はある一つの資源が他の資源をどの程度代替するかを示す。

消費者エージェントは、消費者エージェントが初期に持つ予算を制約条件としてこのような効用関数を最大化するように行動する。すなわち、 p_i を資源 i に付けられた価格^{†2}とすると、

$$\begin{aligned} \max u_k(x_{1k}, \dots, x_{lk}, e_k) \\ \text{s.t. } \sum_{i=1}^l p_i x_{ik} \leq e_k \end{aligned} \quad (2)$$

の最大化問題の解としてエンドユーザの選好を得ることができる。

式 (2) のためのエンドユーザの効用情報の取得は、キャッシュコンテンツ決定のプロセスを行う前に、ブラウザやマルチメディアプレイヤー上のカテゴリ選択などで行うことが可能であると考えている。

^{†2} 提供されるキャッシュサービスの価格にあたる

4.3 生産者エージェント

エンドユーザの選好を効用関数で表わし、前述のプロトコルによってキャッシュに集められると、キャッシュはこれらの効用関数の総和として表記されるような社会的効用を最大化する。

キャッシュを代表する生産者エージェントは、生産関数であらわされる技術内でキャッシュの資源からコンテンツキャッシュを生産する。各コンテンツ i に対しては、 $\sum_{k=1}^n x_{ik}$ だけの需要があり、 $\sum_{k=1}^n p_i x_{ik}$ だけの財が割当てられる。キャッシュは、 $\sum_{i=1}^l p_i \sum_{k=1}^n x_{ik}$ を用いて、キャッシュコンテンツとストレージを購入し、キャッシュサービスを構築する。キャッシュサービスは、各コンテンツ種別 i 毎に $\sum_{k=1}^n x_{ik}$ だけの需要がなされ、この範囲内で、個別のコンテンツを購入し、キャッシュに置く。

各コンテンツ i に対するキャッシュ j の行動は、コンテンツ価格を p_i^c 、ストレージ価格を p^s 、コンテンツの購入量、ストレージの購入量、そこから生産されるキャッシュサービスをそれぞれ、 c_{ij} 、 s_{ij} 、 y_{ij} とし、式 (3) の最大化問題の解として与えられる。

$$\max \sum_{i=1}^l (p_i y_{ij} - (p_i^c c_{ij} + p^s s_{ij})) \quad (3)$$

$$y_{ij} = f_j(c_{ij}, s_{ij})$$

ここで、 $f(c_{ij}, s_{ij})$ は、コンテンツとそれを提供するためのストレージ等の資源からキャッシュサービスを生産する生産関数となっている。

式 (3) の均衡解を y_{ij}^* とする。 y_{ij}^* を与える 2 財 c_{ij} 、 s_{ij} の組みを (c_{ij}^*, s_{ij}^*) で表わすと、 c_{ij}^* 、 s_{ij}^* は、式 (4) を満たす。

$$\begin{cases} \frac{p_i^s}{p_i} = \frac{\partial}{\partial s_{ij}} f(c_{ij}^*, s_{ij}^*) \\ \frac{p_i^c}{p_i} = \frac{\partial}{\partial c_{ij}} f(c_{ij}^*, s_{ij}^*) \end{cases} \quad (4)$$

式 (4) を満たす解 (c_{ij}^*, s_{ij}^*) が得られても、キャッシュは必ずしもサービスを提供するとは限らない。価格の調整が終了し、最終的な均衡 (c_{ij}^*, s_{ij}^*) が得られた時、式 (5) を満たしている必要がある。

$$\sum_{i=1}^l (p_i^c c_{ij} + p^s s_{ij}) \leq \sum_{i=1}^l p_i \sum_{k=1}^n x_{ik} \quad (5)$$

式 (5) は、キャッシュがキャッシュサービスを提供

することにより、利益が得られることを示しており、この式が満たされない場合、キャッシュサービスのみを独立して運用することは、損失が生じることから不可能であると言える。

また、ストレージやネットワーク等のリソースにおいて大容量なものを用意することが困難であるなどの収穫逨減の性質を示すのに対し、コンテンツは従量課金あるいは固定の料金体系となりうるため、生産関数が不連続、あるいは、収穫不変となりうる。

図 4 に、 x 、 y 軸をキャッシュサービスを生産するために必要なストレージなどのリソース、及び、コンテンツをとり、 z 軸に生産されるキャッシュサービスをとり、このような狭義凸関数とならない生産関数を例示する。

図 4(a) は、コンテンツの費用が定数で、一定額支払えば制限なくコンテンツを利用できる場合であり、図 4(b) は、コンテンツの費用は従量制で、利用するコンテンツ量に対して線型にコストが増加する場合を示している。このようなコンテンツに特有の性質により、式 (4) より得られる均衡解は、大域的な最適解ではなく局所的な最適解に留まる、あるいは、利潤に上限がなくなり式 (3) を満たす解が存在しない可能性がある。

以上のように、生産者エージェントは、

1. キャッシュサービスを提供して正の利潤を得るための条件
 2. 最適なコンテンツ集合が得られるための条件
- を考えて実装する必要があり、今後この条件を明らかにしていく予定である。

4.4 価格調整機構

各生産者エージェント・消費者エージェントは対応する財について最大化問題を解き、その解を競売人エージェントに入札する。競売人エージェントは、生産者エージェントと同じくキャッシュシステム内のソフトウェアエージェントであり、各資源・コンテンツの需給がつりあうように価格を調整する。

競売人エージェントは、 $\sum_{k=1}^n x_{ik}$ と、あるキャッシュ j での y_{ij} が表明された場合に、関数 θ_i を単調増加な関数として、新しい価格を式 (6) に従い変更し、

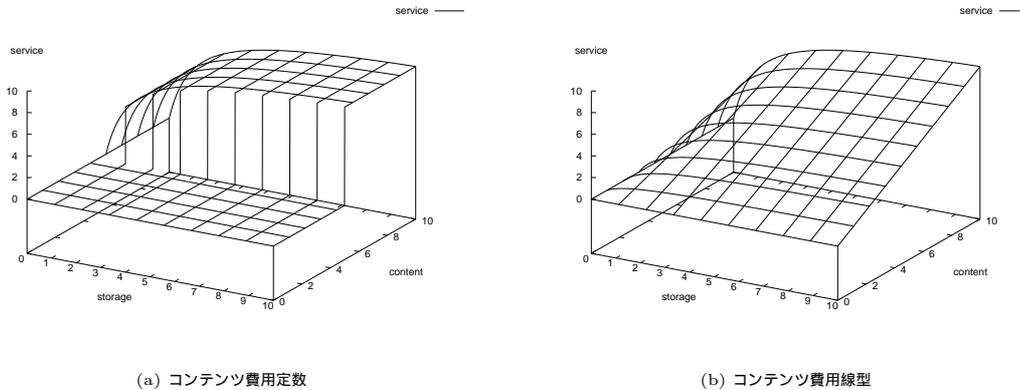


図 4 収穫逓減でない生産関数

キャッシュ j において $\sum_{k=1}^n x_{ik} - y_{ij} = 0$ となる均衡解 $p^* = (p_1^*, \dots, p_l^*)$ を探す。この時の y_{ij} が、コンテンツ種別 i に対するキャッシュサービスの最適供給量となる。

$$\forall i, \frac{dp_i}{dt} = \theta_i \left(\sum_{k=1}^n x_{ik} - y_{ij} \right) \quad (6)$$

競売人エージェントは決定された新しい価格を各消費者エージェント・生産者エージェントに通知する。

4.5 計算手順

消費者エージェント・生産者エージェントはその時点の価格における最大化問題の解を計算し、競売人エージェントにその解を伝える。競売人エージェントは、キャッシュサービスの最適供給量を得られる均衡価格を計算し、消費者エージェント・生産者エージェントに伝える。この過程を収束するまで繰り返し、得られたキャッシュ集合に対応するように各コンテンツをキャッシュする。図 5 にこのアルゴリズムを示す。

市場が適切な性質を持つ場合、上述のプロセスは収束し、需要・供給量は均衡することが知られている。また、均衡点での需要・供給量は実際に割り当てが可能である、すなわち、実行可能な解であり、パレート最適性を示すことが知られている。

パレート最適性は、あるエージェントがそれより有利な資源配分を得る時は、必ず他のエージェントが不利になる性質であり、パレート最適な割り当ては無駄

の無い割り当てがなされていることを表している。

5 おわりに

本稿では、コンテンツデリバリネットワークにおけるコンテンツのキャッシュポリシーについて記述した。

マルチメディアコンテンツが容易に作成され公開されるようになり、サーバの負荷、また、ネットワークトラフィックが増大している現在、コンテンツデリバリネットワークはこの問題に対して有効な解決策と言える。より多くのコンテンツに対してコンテンツデリバリネットワークを適用しエンドユーザにより速いサービスを提供するためには、どのコンテンツをキャッシュするのかの指針を定めることが必要となると考えられる。

本稿で取り上げた手法により、ユーザの要求に適合し、経済機構を背景としたコストモデルを持つキャッシュの指針が与えられ、コンテンツデリバリネットワークに貢献するものであると考えている。今後、

1. 式 (3) において、収穫逓減でない性質の市場における最適性の検討
2. 本稿のモデルに対応したシミュレーションによる評価
3. ネットワーク事業者、コンテンツ配信者、及び、エンドユーザの各主体へのコスト負担を含めて、ビジネスモデルを考慮したモデルへの改善

の 3 点を行い、モデルの有効性を確認していく予定である。

```

begin Auctioneer Agent
  begin loop
    for each Producer Agent, Consumer Agent
      send Price Vector
    end of for
    for each Producer Agent
      receive Production matrix
    end of for
    for each Consumer Agent
      receive Demand matrix
    end of for
    if balance ( Production matrix, Demand matrix) exit
    Price Vector =  $\theta$ ( Price Vector, Demand matrix, Production matrix)
  end of loop
end of Auctioneer Agent

begin Producer Agent
  begin loop
    receive Price Vector from Auctioneer Agent
    Production matrix = Calculate production ( Price Vector )
    send Production matrix to Auctioneer Agent
  end of loop
end of Producer Agent

begin Consumer Agent
  begin loop
    receive Price Vector from Auctioneer Agent
    Demand matrix = Calculate demand ( Price Vector )
    send Demand matrix to Auctioneer Agent
  end of loop
end of Consumer Agent

```

図 5 各エージェントのアルゴリズム

謝辞

本研究の機会を与えて下さり、また、的確な御指導を賜りました京都大学大学院情報学研究科社会情報学専攻の石田 亨 教授、NEC USA の原 良憲 様、京都大学大学院情報学研究科社会情報学専攻の亀田 弘行 教授に深く感謝申し上げます。

参考文献

- [1] Breslau, L., Cao, P., Fan, L., Phillips, G., and Shenker, S.: Web Caching and Zipf-like Distribu-
- tions: Evidence and Implications, 1998.
- [2] Karger, D. R., Lehman, E., Leighton, F. T., Panigrahy, R., Levine, M. S., and Lewin, D.: Consistent Hashing and Random Trees: Distributed Caching Protocols for Relieving Hot Spots on the World Wide Web, *ACM Symposium on Theory of Computing*, 1997, pp. 654-663.
- [3] Wellman, M.: A Market-Oriented Programming Environment and its Application to Distributed Multicommodity Flow Problems, *Journal of Artificial Intelligence Research*, Vol. 1(1993), pp. 1-23.
- [4] Yamaki, H., Wellman, M. P., and Ishida, T.: A market-based approach to allocating QoS for multimedia applications, *Proceedings of the First International Conference on Multi-Agent Systems*(Lesser, V.(ed.)), MIT Press, 1995.

口コミによる分散型情報収集システム

WAVE を起こそう—Word-of-mouth-Assisting Virtual Environment

吉田 匡志 伊藤 雄介 沼尾 正行

Abstract

情報処理装置として出発したコンピューターは、近年、コミュニケーションメディアとしての役割を果たすようになってきた。しかし、コミュニケーション上のトラブルなどやコミュニティ参加への心理的不安が活発な情報交換を妨げる原因となっている。また、氾濫した情報の中から有用な情報を獲得するためには大変な労力を必要とする。そこで、本稿では、口コミを電子コミュニティ上に再現することで効率の良い情報収集や円滑なコミュニケーションを支援するシステム、WAVE (Word-of-mouth-Assisting Virtual Environment) を提案する。また、このシステムを実装し実験した結果、いくつか興味深い現象を観察した。

Recently, computers play an important role not only in knowledge processing but also as communication media. However, they often cause troubles in communication, since it is hard for us to select only useful pieces of information. To overcome this difficulty, this paper proposes a new system, WAVE (Word-of-mouth-Assisting Virtual Environment), that helps us to communicate and to collect information by word of mouth, and shows some interested phenomenon in the experiment.

1 はじめに

現在、WWW や電子メールといった電子メディアによるコミュニケーションが大変賑わいを見せている。情報処理装置として出発したコンピューターは、通信技術やヒューマンインターフェースなどの発達により、コミュニケーションメディアとしての役割を果たすようになってきた。

しかしその一方で、フレーミング^{†1}などのコミュニケーション上のトラブルや、コミュニティ参加への心理的不安といったものが実世界と比べて顕著に表れ、活発な情報交換を妨げる原因となっており、コンピューターを介したコミュニケーションと実世界におけるコミュニケーションがどう違うのかといった議論 [1][2] が盛んに行われている [3]。

また、計算機ネットワーク上には様々な情報が氾濫し、ユーザーにとって有用な情報を獲得するのに大変な労力を必要とする。情報を発見するための最も身近なツールとして検索エンジンがあるが、最もカバー率の大きいとされる Alta Vista でもせいぜい4割から6割程度であると言われており、広大なWWW空間に氾濫した情報をすべて把握するのは非常に困難である。それにどのような情報が欲しいかユーザー自身が明確に分からない場合も多い。

そこで、情報の内容を解析しユーザーが文章中のどの部分に対して興味を持っているか推定することで同様の部分を持つ情報を有用な情報として提供する

東京工業大学 大学院情報理工学専攻 計算工学専攻
{myoshida, blanke, numao}@nm.cs.titech.ac.jp

^{†1} コミュニケーションの中で現れる、ののしり、侮辱、強調された表現

Content based filtering [4][5] や、評価傾向の似ている他ユーザーの持つ情報を参考にして有用な情報を提供する Colaborative filtering [6][7] などの情報フィルタリングの研究が盛んに行われてきている。しかし、これらは数値的な計算だけで人の心の動きや人間関係を解釈しようとする傾向が強く、機械的な精度を上げて限界がある。

これに対して、大量の意見を要約して質の高い情報を提供するマスメディアと自由に情報発信が可能でインタラクティブ性を持つ電子メディアを融合した新しいメディア [8] や、コミュニティの可視化 [9][10][11] など、人の心の動きや人間関係を考慮しコミュニティの形成を支援する研究もある。コミュニティの形成を支援することは情報収集の支援にもつながる。しかし、情報フィルタリングのようにユーザーにとって有用な情報を収集するといったことまでは考慮されていない。

そこで、本稿では、口コミを電子コミュニティ上に再現することで、効率のよい情報収集や円滑なコミュニケーションを支援するシステムを提案する。

2 口コミの特徴

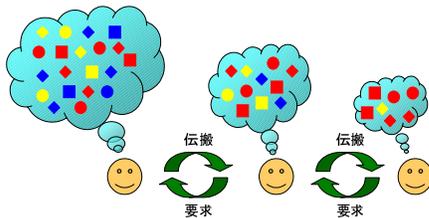


図 1 口コミのイメージ

口コミは、不特定多数の人々に均一な情報を伝えるマスコミと異なり、個人間で双方向的に情報の伝達が行われ [12]、図 1 のように、氾濫する情報の中から多くの人々の評価と伝播を経ることで有用な情報だけが人々の間に広まっていく。口コミによる情報伝播では、相手の興味のありそうな情報であるかを判断して伝えることや、個人の情報収集では見逃してしまうような重要な情報を伝えることができる。さらに、情報を受け取る側は情報提供者に対して、信頼できる人が

や、その話題に関して専門性を持っている人であるかといった評価をあらかじめ持っている。また、本人の経験を元に重要な部分が強調されるので、情報の質が高くなる。よって容易に情報の価値判断ができる。

しかしその反面、「平均化」(伝達要素の減少)や「強調化」(少数要素の強調)、「同化」(予期的枠組みへ一貫する方向への内容変化)などの情報内容の変容が起こりやすく、うわさや流言など発生しやすい。また、人間関係のつながりにより情報が伝播していくので、強いインパクトのある流言でもない限りは狭い範囲に限定されがちである。

口コミによる情報伝達については、いろいろな研究がされており [13]、その一つに社会ネットワーク分析^{†2} [14] の研究分野がある。ここでは、親友など頻繁に対面接触する緊密な人間関係を「強い紐帯(ちゅうたい)」と呼び、まれにしか対面接触をしない薄い人間関係を「弱い紐帯」と呼ぶ。また、広がる人間関係の網の中で、派閥のように互いに直接結びつきあっている人間関係(強い紐帯)の集合を「クリーク」と言う。そして、クリークの間を結びつける弱い人間関係(弱い紐帯)が「ブリッジ」である。

一般に、口コミはクリークの中で活発に行われ、人間関係の紐帯が強いほうが影響力があり、信頼性が高く有効である。また、専門性を認知されたり信頼性が高い方がより説得的であるなどの結果も出ている [15]。しかし、転職時に弱い紐帯を通じて情報を得た人の方が転職後の満足度が高く、弱い紐帯は強い紐帯よりも有効であったという興味深い結果も報告されている [16]。クリーク内では情報伝達が早いですが、同じような興味を持った人々が集まっているので似たような話題についての情報交換がされやすい。新しい情報は、ブリッジを通じてクリークへと導入される。

これらの口コミの諸性質は、電子コミュニティ上においても当てはまると考えられる。それと同時に、電子メディア特有の性質が実世界の口コミが持っていた

^{†2} 集団成員間のコミュニケーションの構造を見出す分析手法。誰と誰がどれくらいコミュニケーションをとったのか、データを採り、これをグラフ化し、グラフ理論を用いて、中心性、密度といった指標を計算することによって、クリークやブリッジといった構造的役割を発見する。

欠点を補うと考えられる。電子メディアにおける情報の伝播はオリジナルのコピーの転送や URL のリンク情報の伝達といった形で行われる。このため実世界における情報伝播よりも正確な情報伝播が行われ、情報の変容が起こりにくい[17]。また、口コミは狭い範囲に限定されがちであるが、電話により口コミの伝達速度が急激に上昇し、その伝達範囲を大きく広げた[13]ことから、電子コミュニティではさらに伝達速度が上昇し、その伝達範囲も広がる。

3 WAVE のしくみ

3.1 口コミ成立の要件

本稿で提案するシステム、WAVE^{†3}(Word-of-mouth-Assisting Virtual Environment) は、電子コミュニティ上において口コミを再現することで、効率の良い情報収集や円滑なコミュニケーション支援を行う。過去に、口コミの考え方を導入し検索エンジンの分散管理を行った研究[18]があるが、口コミ成立の要件や実現すべき目標が本稿と異なる。なお本稿で意図する口コミとは以下のようなものである。

- 情報のプッシュとプル統合
- 信頼関係にある人同士での情報交換が可能
- 評価情報を伴う情報伝達

この要件を満たすことで、2 章で説明した性質を持って情報が伝播していく。また、口コミの成立を支援するためにユーザーインターフェースにも工夫を施した。以下では WAVE のしくみとその特徴について図 2 に沿って説明していく。

3.2 情報の発信

各ユーザーは、自分が知っている新しい情報を発信することができる。図 3 のように、自然言語による情報だけでなく、Web ページや画像データの URL 情報を付加することができたり、情報を閲覧するユーザーが情報の内容を判断しやすいように、情報を簡単に分類するためのカテゴリを割り当てるようになって

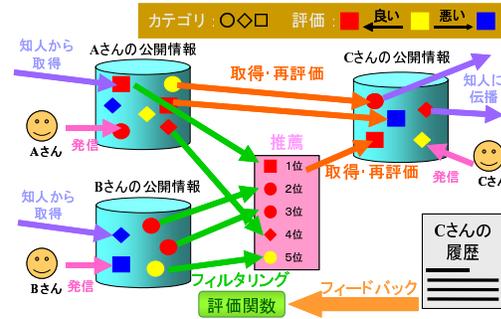


図 2 WAVE における口コミの再現

いる。また、情報に対して 1~5 (1 が一番悪く、5 が一番良い) の評価値を与える。発信した情報は、自分の情報として他ユーザーに公開され、WWW やメーリングリストのように、自分なりの情報を多数の人々に向けて自由に発信することが可能である。

3.3 情報の公開

図 2 のように、自分が発信した情報や他ユーザーから取得した情報は、自分の情報として蓄積される。また、自分の情報は他ユーザーに公開され、隣接するユーザーへ自動的に情報が流れていく。例えば、図 2 の C さんが知人の A さんや B さんの持つ WAVE 専用のアドレス ("ユーザー名@ホスト名:ポート番号" の形) を図 4 の画面で "お友達リスト" としてシステムに登録すれば、C さんは A さんや B さんに隣接するユーザーとなって、A さんや B さんから自動的に流れてきた情報を簡単に閲覧することができる。WAVE では、情報をもらいたいユーザーをお友達リストに登録したり、逆に、自分の情報を伝えたい人に自分のアドレスをお友達リストに追加してもらうように通知したりすることができる。

これにより、情報のプルとプッシュを統合することができる。また、ユーザー同士の人間関係が強く現れ、相手の専門分野や信頼性を判断することができるので、WWW などに欠落していた情報源の信頼性を高めることができる。さらに、自分の信頼する隣接ユーザーとのやり取りがほとんどなので、安心してコミュニティに参加できる。

^{†3} WAVE には、口コミが波のように伝播していくというイメージと、このシステムが、WAVE を起こし、世界中の人々に使ってもらえるようなコミュニケーションメディアとなしてほしいとの願いがこめられている。

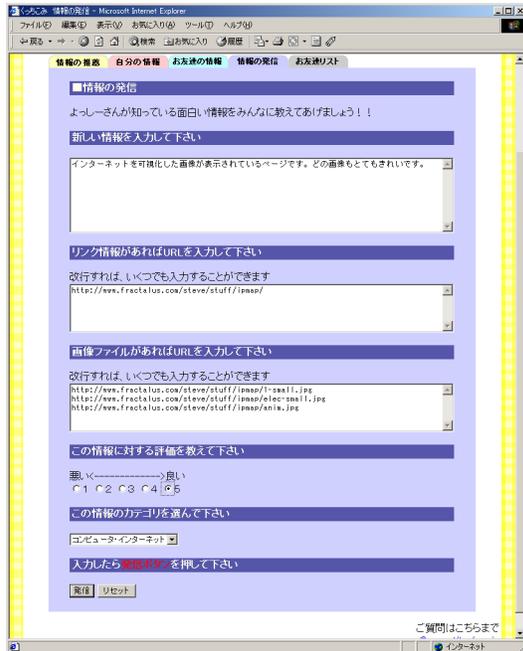


図 3 情報の発信画面



図 4 お友達リスト画面

3.4 情報の評価と取得

図 2 で、C さんが A さんから流れてきた情報の中から興味を持った情報があれば、図 5 の画面のようにその情報に関する詳しいデータを見ることができる。このとき、その情報に新しい評価をつけたり、自然言語による付加情報としてコメントを与えることができる。この画面では、自然言語による情報だけでなく画像やハイパーリンク、情報提供者による評価・コメントなどが表示されるのでその情報が有用かどうかの判断を支援するようになっている。C さんが新しい評価やコメントを与えると、A さんの情報は自動的に取得され C さん自身の情報として公開される。そして、それがまた他のユーザーに伝播していく。

このように WAVE では情報の公開、閲覧、評価、取得をシームレスに行うことができるので、情報のアップロードや情報の存在を人々に知らせるため手間を軽減し、情報のプッシュを促進する。また、BBS やメーリングリストにおける ROM(Read Only Member) が持つような積極的な参加への心理的抵抗感も軽減する。このことは、情報伝播において重要な役割を果たすブリッジの存在を維持することにつながる。

多くの人間を経由して情報の評価と取得が繰り返し行われるうちに段階的にフィルタリングが行われ、有用な情報だけがネットワーク上に広まる。相手を侮辱するような情報やチェーンメールのような情報を発信しても周りから取得されず自分の情報としてとどまり広がりにくい。また、同じような嗜好を持つ人間はクリークとして集まりやすいので情報の流れは指向性を持つ。そしてブリッジを介して情報がひとたび自分の属するクリークに到達すれば、自分もその情報を得られる。WAVE では、人間のコミュニティそのものが分散化された情報収集システムとして機能し、グローバルな情報交換ネットワークを形成されるといえる。

3.5 情報の推薦

”お友達リスト”に登録したユーザーの数が増加したり、ひとりのユーザーが公開する情報の数が増加したりすることによって、流れてくる情報をすべて閲覧するのは負担となってくることが予想される。

そこで、図 2 のようにユーザーの閲覧履歴等をもとに動的に作成した評価関数により、公開されている情報の中から有用であると思われる情報の一覧(図 6)



図 5 情報の詳細画面



図 6 情報の推薦画面

を提示し情報の閲覧にかかる負担を軽減する。なお、評価関数は以下の項目を基準にして作成した。以下の項目の値が大きくなると、評価関数の値が大きくなり、ユーザーにとって有用な情報として推薦される。

- 情報提供者が与えた情報に対する評価
口コミによって流れる情報は、情報提供者によってすでに1~5(1が一番悪く、5が一番良い)の評価値が与えられている。情報提供者の評価値が高ければ、他のユーザーにとっても評価の高い情報であると考えられる。
- 情報提供者に対するユーザーの評価
ある人の情報を見たり取得する回数が多い時、その人に対して高い評価をしているので、その人から流れてきた情報の評価も高いと考えられる。このとき、コンピューターに詳しい人からはコンピューターに関する情報を取得しがちであるというように、ユーザーが取得する情報の分野は情報提供者によって偏りがあり、提供者の評価には提供者の持つ専門性も考慮する必要がある。そこで、すべての情報提供者の中で、今までに、その情報提供者の情報をどれくらい閲覧・取得

したかや特に今回の情報と同じカテゴリの情報はどれくらい閲覧・取得したかクリック率を計算し、この値が大きいほど情報提供者の評価が高いとする。さらに時間が経つと、情報をいつも参照する人が変わることがあるので、情報提供者の情報を最後に見た日数が経過するにしたがって、その情報提供者の評価が低くなるようにする。

- ユーザーの嗜好
ユーザーが興味のあるカテゴリに関するものほど、情報に対する評価は高いと考えられる。そこで、今回の情報と同じカテゴリの情報に対して今まで与えてきた評価値の平均や、すべてのカテゴリの中で今回の情報と同じカテゴリの情報をどれくらい閲覧・取得しているかクリック率を計算し、これらの値が大きいほど情報の評価が高くなるようにする。
- 情報の鮮度
新しい情報ほどユーザーの評価は高いと考えられる。情報が取得・公開されてからの日数が経つたものほど情報の評価は下がるようにする。
- 情報が伝達した人数

多くの人の間を伝達した情報ほどユーザーの評価は高いと考えられる。伝達した人数が多いほど情報の評価が高くなるようにする。

WAVE における情報の推薦は、Collaborative filtering のように評価傾向の似ている他ユーザーの持つ情報を参考にして有用な情報を提供するのではなく、隣接ユーザー間で情報をプッシュする側とプルする側のマッチングを行う。現段階では本システムは、この情報はあの人に教えたいというように、情報ごとに特定のユーザーを指定してプッシュすることができないので、情報の推薦がこの機能を補う。

3.6 システムの分散化

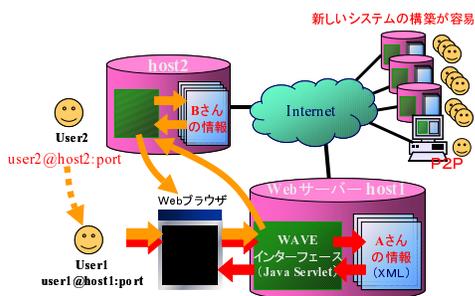


図 7 システムの分散化

図 7 のように、WAVE ではシステム自身の分散化も行っている。システムは Java サブレットを用いて実装されており、Web サーバー上で動作する。ユーザーは、Web ブラウザを介してユーザー登録をしたホストにアクセスするだけで、情報のやり取りをすることができる。

自分の情報は、ユーザー登録したホスト上に蓄積される。なお、この情報は、次世代の Web の標準言語となる XML により保存されている。これにより、既存の Web ページからの情報を有効利用することなどが将来的に可能とあり、WAVE は WWW を包括するメディアとなり得る。

分散化されたシステム同士での連携の仕組みは非常に単純である。図 7 を例に説明すると、User1 は、host1 で動作するシステムにユーザー登録しているので、host1 へアクセスすることで、情報のやりとり

を行うことができる。また、他のホスト host2 を利用している友人の User2 の情報が閲覧したいときは、登録した User2 の WAVE 専用のアドレスからそのユーザーが持つ情報の存在位置が一意に定まるので、User2 の情報が host1 に存在するかのように情報の存在位置を気にすることなく取り扱うことができる。

分散化によりシステムをどこでも自由に構築できるので、ネットワークの負荷分散や、CPU、ストレージなど計算資源の有効活用が可能である。また、ユーザー数が増加しても新しくシステムを構築すればよく、スケラビリティが高い。したがって、ユーザー数の増加が促進され、より効率的な情報収集を WAVE で行うことが可能になる。さらに、WAVE を Peer-to-Peer 方式^{†4}で実装すれば、一般的になりつつある個人のインターネットへの常時接続環境において広く使われると考えられる。

4 実験

WAVE を実装し、被験者 33 名で 20 日間運用した。その後、実験データを解析し、社会ネットワーク分析のためのグラフ可視化ツール KrackPlot [19] を用いて、図 8 のように、WAVE 上でのコミュニティにおけるグラフを作成した。

ノードはユーザーを表しており、情報を最初に発信した回数によって色分けされている。この場合、myoshida や blankey、roy、t-sugie などがオピニオンリーダーとなってコミュニティに話題を提供している。ノード間を結ぶアークは、あるユーザーから別のユーザーへ矢印の方向に情報が取得されたことを表している。線の太さは、情報を取得した回数である。ノードの入次数が多い場合、ユーザーは多くのユーザーを信頼し情報を獲得していると言える。逆にノードの出次数が多い場合は、多くのユーザーから信頼されるような情報を提供していると言える。

図の中心にあるノードからアークが放射上に広がっ

^{†4} クライアント・サーバー型のネットワークと違い、ネットワーク上のコンピュータのそれぞれが、サーバーでありクライアントとなる。よって、集中的に処理を行なうサーバーを設置することなく、各ネットワーククライアントが持つ資源をお互いに共有する事が可能である。

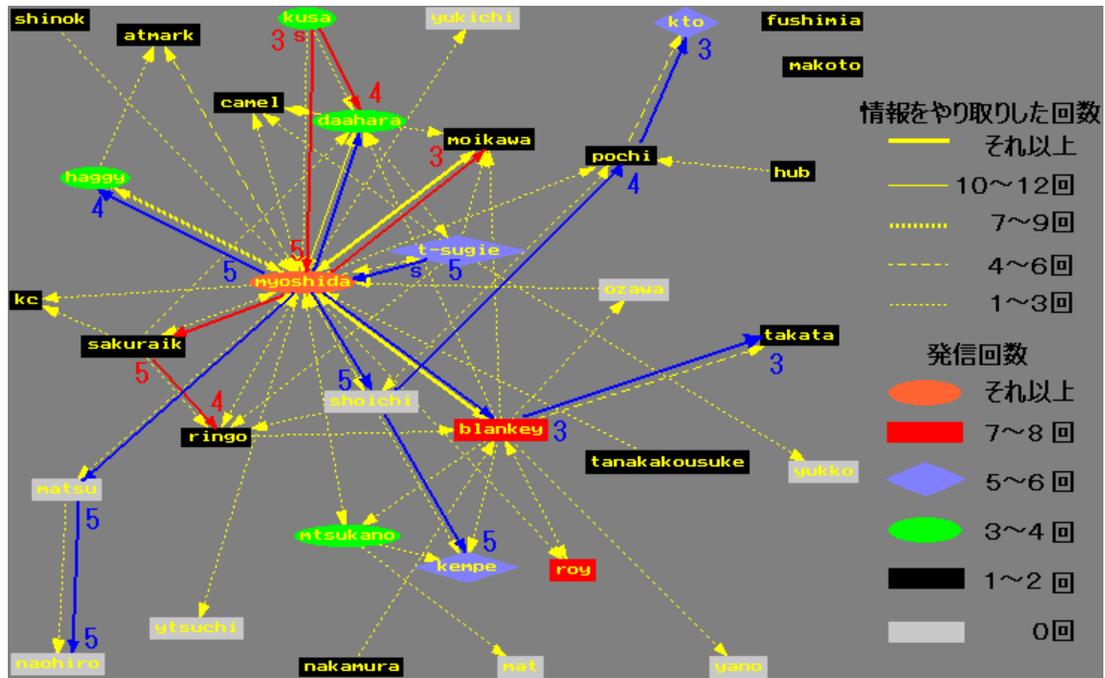


図 8 WAVE 上でのコミュニティ

ており、中心にいる myoshida や blankey の周辺で活発な情報のやりとりが行われていることが分かる。また、アークによって結ばれた三角形がいくつも確認できる。このような三者関係 (triad) [14] は小さなクリークで結びつきが強く、実世界におけるコミュニティを照らし合わせると、この三角形がいくつも結びついて大きなクリークを形成しているのが分かった。

さらに、情報が伝播していった例を 2 つ、それぞれ赤と青の線でトレースした。ノードの脇にかかれている数字は、ユーザーがその情報に対して与えた評価値 (1~5) である。ここで、情報が伝播するにしたがって徐々に評価が下がっていることが分かる。また、青い線の情報は、matsu、naohiro の辺りでは評価が高いが、blankey、takata や poti、kto では低くなっている。これは、集団内では似たような価値観を持った人が多いことや自分の評価を決定するときに情報提供者の評価を参考しているからではないかと考えられる。情報が伝播されるうちに評価が減少していくと、取得される価値がないと判断され伝播が途絶える。一方、その情報に興味をもった人たちの間には

広まっていく。

また、情報は人間関係に沿って流れ、myoshida、shoichi のように情報の経路が枝分かれしているところで、複数の集団 (クリーク) へ情報が伝播し、集団内では急速にその情報が広がるといった現象が起こっている。この場合、図から弱いつながりとしてブリッジの存在を確認するのは難しいが、そのようなユーザーは実世界において、サークルや研究室など異なる集団を結びつける人物であることが分かった。このような人物がゲートキーパー (gatekeeper) [20] として働いていると考えられる。ユーザーに使用してもらった感想を聞くと、システム上で流れてきたある情報について、同じ集団に属するユーザーの間で実世界において初めて話題になった時にすでに全員が知っていたなど、その現象をユーザー自身で実感することができたようだ。このように、2 章で説明した口コミの性質が観察され、WAVE の有効性を確認することができた。

5 まとめ

本稿では、口コミを電子コミュニティ上に再現することで、効率のよい情報収集や円滑なコミュニケーションを支援するシステムの提案を行った。また、実験データから作成した WAVE 上でのコミュニティのグラフから WAVE の有効性を確認することができた。

今後も引き続きシステムを運用し実験を行う。

現在、<http://www.nm.cs.titech.ac.jp:12581/wom/>で、システムを公開している。興味を持たれた方は実験に御協力頂きたい。また、本システムを配布し、複数のホストで動作させ、分散環境を実現したい。

その後、本稿で行ったようなコミュニティの可視化を行うことで、クリーク、ブリッジなどの構造的役割や情報の伝播の仕方などさらに詳しい分析を行った。ユーザーにアンケートを行うことでシステムの使用感について調査し、システムを評価する。情報の推薦については、推薦により情報を取得した割合や推薦により取得した情報の順位の変化、ユーザーへのアンケートなどによりその妥当性について評価する。

また、システムの改良についても行っていきたい。現状では、カテゴリは情報の発信時のみ付与するようになっているが、取得時にも各ユーザー独自のカテゴリを付与する。カテゴリをフォルダとみなし階層構造を許すようにすると、自分の持っている情報をフォルダごとに自由に整理することも可能となる。この際、情報の推薦については、階層構造内で何親等になるかという距離を定義することでカテゴリ間の類似度が計算できるので、それを他のカテゴリ内の情報への評価として活用すれば、カテゴリに分類したことによるサンプル数の減少を補うことができる [7]。

参考文献

- [1] Sproull, L. and Kiesler, S.: コネクションズ: 電子ネットワークで変わる社会, アスキー出版, 1993.
- [2] Lea, M.: Contexts of Computer-Mediated Communication, Harvester Wheatsheaf, pp.30-65, 1992.
- [3] 野島 久雄: 電子メディア社会の心理学, 情報処理学会論文誌, Vol.40, No.1, pp.66-70, 1999.
- [4] 溝口文雄, 大和田勇人. 帰納学習に基づく情報フィルタリング. 人工知能学会全国大会 (第 10 回) 論文集, 1996.
- [5] Pattie Maes: Agents that reduce work and information, CACM, Vol.37, No.7, pp.30-40, 1994.
- [6] Resnick, P., Iacovou, N., Suchak, M., Bergstrom, P., Riedl, J.: GroupLens: An Open Architecture for Collaborative Filtering of Netnews. CSCW '94 Proceedings. pp.175-186, 1994.
- [7] 沼尾 正行, 横山 甲: 階層化された知識の継承による情報フィルタリング, 情報処理学会「知能と複雑系」研究会, 1999.
- [8] 西田 豊明, 畦地 真太郎, 藤原 伸彦, 角 薫, 福原 知宏, 矢野 博之, 平田 高志, 久保田 秀和: パブリック・オビニオン・チャンネル, 第 2 回 CMCC 研究会シンポジウム, http://www.brl.ntt.co.jp/people/nojima/CmCC/Resources/CmCC_Symposium_Papers/cmcc0206-nishida.pdf, 1999.
- [9] 藤田 邦彦, 亀井 剛次, Eva Jettmar, 吉田 仙, 桑原 和宏: ネットワークコミュニティの可能性-Community Organizer 評価実験結果報告-, 第 3 回 CMCC 研究会シンポジウム, http://www.brl.ntt.co.jp/people/nojima/CmCC/Resources/CmCC_Symposium_Papers/cmcc03-fujita.pdf, 2000.
- [10] 高橋 正道, 北山 聡, 金子 郁容: ネットワーク・コミュニティにおける組織アウェアネスの計量と可視化, 情報処理学会論文誌, Vol.40, No.11, p.3988-3999, 1999.
- [11] 館村 純一: 協調型情報探索を支援する仮想評者とその視覚化, WISS'99, 1999.
- [12] 岡堂 哲雄: 現代のエスプリ別冊 社会心理用語事典 p.91, 至文堂.
- [13] 中村 功: 現代のエスプリ別冊 特集「流行… ファッション」流行と口コミ電話, pp199-209, <http://cc.matsuyama-u.ac.jp/~nakamura/espri.htm>, 2000.
- [14] 安田 雪: ネットワーク分析, 新曜社, 1999.
- [15] Bristol, J.: Enhanced explanations of word of mouth communications; the power of relations, Research in Consumer Behavior, Vol.4, pp51-83, 1990.
- [16] Granobetter, M.: Getting A Job, 1974 (渡辺深訳「転職」ミネルバ書房, 1998).
- [17] 柴内 康文: 電子メディア社会における情報伝播, 第 2 回 CMCC 研究会シンポジウム, http://www.brl.ntt.co.jp/people/nojima/CmCC/Resources/CmCC_Symposium_Papers/cmcc0205-shibanai.pdf, 1999.
- [18] 大谷 武, 南 俊朗: 口コミによる情報資源探索, MACC'1997 一般セッション (情報検索支援), <http://www.kecl.ntt.co.jp/csl/msrg/events/macc97/ohtani.html>, 1997.
- [19] KrackPlot home page, <http://www.contrib.andrew.cmu.edu/krack/>.
- [20] 後藤 滋樹, 野島 久雄: 人間社会の情報流通における三段構造の分析, 人工知能学会誌, Vol.8, No.3, pp.348-356, 1993.
- [21] 伊藤 雄介, 吉田 匡志, 沼尾 正行: 口コミ支援システムの実験, 情報処理学会「知能と複雑系」研究発表会, 2001.
- [22] 伊藤 雄介, 吉田 匡志, 沼尾 正行: 多くの人の評価を経て情報が吟味される口コミ支援システム, 第 15 回人工知能学会全国大会, 2001.

知識の階層構造を用いた協調的情報推薦ネットワークの提案

濱崎 雅弘 武田 英明 河野 恭之 木戸出 正継

1 はじめに

情報過多にある現状を改善するために多くの情報収集支援技術が開発されている。その手法として、情報そのものをテキスト解析等の技術で機械処理し、整理することで支援するもの他に、情報に関する人の知識を利用することで支援するものがある。後者は機械処理では困難な情報の価値的な部分まで扱える利点があるが、人の協力が必要であり、システムが利用者へ多くの負荷をかけるなどの理由で協力が得られなかった場合には、うまく機能しなくなるという問題がある。そこで本研究ではより利用者が協しやすい情報共有の形態と考えられる、ネットワーク型の情報共有システムを提案する。

2 情報共有の負荷

人の知識を用いた情報収集支援技術には、参加者の協力が欠かせない。例えば、協調フィルタリングではある程度の量の情報が集まらなければそもそも機能しない。また、メーリングリスト等のオンラインコミュニティにおいても、参加者が活発に議論する姿勢を維

持し続けなければ情報共有の場として機能しない。

では、参加者の協力が得られない状況はどのようにして起こるのだろうか。それは情報共有に対する負荷が、得られるメリットを上回ったときと考えられる。では、情報共有における負荷とはどのようなものがあるのか。今回、以下のような二つの負荷を考えてみた。

- 情報作成の負荷 情報を提供するだけで、その情報が直接には自分にとって価値の無い物である場合、その情報を作ることに負荷を感じる。
- 情報提供の負荷 苦労して自分で作った情報を、見知らぬ他人にまで公開することへの抵抗感が負荷と感じる。

多くの協調フィルタリングシステムでは、情報に対する 5 段階評価といった作成容易なものを共有情報として用いることで負荷軽減を行っている。だが、それは情報作成そのものの価値を減らすことになる。これは、すでに十分に情報が蓄積され、共有によるメリットを情報提供者が得られるような状態ならば問題ないが、それに至らない場合、例えば蓄積された情報量が不十分な初期段階などにおいては、大きな問題となる。

オンラインコミュニティでは、情報作成の負荷は大きくとも、参加意識によって情報提供の心理的負荷を軽減することで、全体として負荷を小さくしている。だが、コミュニティが肥大化して参加意識が希薄になってきた場合、この構図は崩れてしまう。

Proposal of The Collavolative Filltering Networks using Shared Topic Networks between Persons.

Masahiro HAMASAKI, 奈良先端科学技術大学院大学, Nara Institute of Science and Technology (NAIST).

Hideaki TAKEDA, 国立情報学研究所, National Institute of Informatics.

Yasuyuki KONO, 奈良先端科学技術大学院大学, Nara Institute of Science and Technology (NAIST).

Masatsugu KIDODE, 奈良先端科学技術大学院大学, Nara Institute of Science and Technology (NAIST).

3 協調的情報推薦ネットワーク

3.1 人のネットワーク

以上をふまえ、本研究では、共有を前提としたものではなく自分にとって有用な情報を、友人や知人といった人と人の関係で作られる範囲で共有する、強調的情報推薦ネットワークを提案する。共有情報が自分にとって有用な情報ならば作成の負荷は少なく、また、共有範囲が友人や知人などの見える範囲ならば提供の負荷も少ないと考えられる。だが同時に、共有を前提としてない情報をどのような形で共有し、友人関係から作られるネットワークをどのように広げていくか、という問題がある。

そこで、我々は個人の情報を連結する方法として共通話題ネットワークを、ネットワークを広げる方法として仲介者モデルを提案する。前者は話題を用いることで共有時の知識のギャップを埋め、後者は仲介者を用いることで心理的な負荷を感じさせずに共有範囲を拡充する手法である。

3.2 共通話題ネットワーク

共通話題ネットワークとは、人がそれぞれ持つ話題を、共通性を元につなげたネットワークである。利用者は共通性を持った話題（共通話題）を介して他の人と情報共有を行うことができる。今回は共有情報として Web ブックマークを用い、そして Web ブックマークを用いた共通話題ネットワーク作成には *kMedia* [Takeda00] を用いる。

Web ブックマークのフォルダは Web ブラウザのブックマークに登録する Web ページを階層構造で整理するためのものである。ユーザは任意にフォルダを作り、その中に記録しておきたい Web ページを登録できる。*kMedia* では、その Web ページが登録されたフォルダを一つの話題とみなし、フォルダに登録している Web ページはそのフォルダ（話題）を示す内容であるとする。

ユーザ同士の話題の共通性は、その話題に含まれる情報がどれだけ類似しているかによって決める。*kMedia* ではブックマークフォルダに含まれる情報、すなわちフォルダ内に登録された Web ページの類似性

から話題の共通性を求める。我々は、この手法が人と人をつなぐ技術として利用可能かどうかを被験者を用いて検証した [Hamasaki01]。

3.3 仲介者モデル

仲介者モデルとは、互いに見知らぬユーザ A と B の共通の知人 C が仲介者となり、お互いを紹介する仕組みである。だが、仲介者によって新しい人を紹介してもらっても、紹介される両者はまだ未知の関係であり、また、新たに関係を結ぶかどうかも未定である。そのため、両者には関係を結ぶかどうか判断するのに十分な情報が与えられる必要があると同時に、まだ見知らぬ相手に対して自分を特定できるような情報は与えてはならないという制約がある。

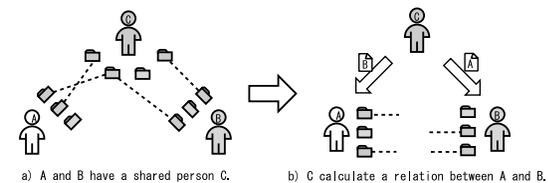


図 1 仲介者モデル

そこで、分類の類似度と、発見された共通話題のみを紹介情報として伝える。この際、相手が持つ話題の名前（=フォルダ名）は伝えられず、自分の持つ話題で相手と共通性のあるものが提示される。つまり友人を介して知り得た相手からは、関係を結ばない限りは、全く相手を特定できない情報しか得られない。

4 今後の課題

本論文では、ネットワーク型の情報共有モデルを提案した。その有効性を検証するためにも、今後は提案に基づく実装を作成し、被験者を用いて評価実験を行いたい。

参考文献

- [Takeda00] H. Takeda, T. Matsuzuka, Y. Taniguchi: Discovery of shared topics networks among people – a simple approach to find community knowledge from www bookmarks, *Proceedings of the PRICAI00*, 2000.
- [Hamasaki01] M. Hamasaki, H. Takeda: Experimental Results for a Method to Discover Human Relationship Based on WWW Bookmarks, *Proceedings of the KES2001*, 2001.

RoboCupRescue における 最少通話による協調の実現

森本 武資 河野 健二 竹内 郁雄

1 はじめに

マルチエージェントシステムの研究では、対象が非常に限定されたトイ的な問題をテストベッドとする場合が多い。トイ的な問題における研究は、問題の本質をとらえやすいが、成果を実際的な問題に適用していくという欠点を併せ持つ。マルチエージェントシステムに関する実際的な問題を研究するためには、様々な事象を対象とするオープンな問題を扱うことが重要である。筆者らは、大規模災害シミュレーションを扱う RoboCupRescue シミュレーション・プロジェクトに参加し、多種類のエージェントが多数活動するマルチエージェントシステムの開発に取り組んでいる。現在、シミュレーションシステムのプロトタイプが開発され、この上で救助活動を行うエージェントの研究が可能になった。しかし、同エージェントの研究が進むにつれ、一見複雑な同システムが実はトイ的な問題に帰着できることが分かり、オープンな問題が本質的に抱える、問題設定と実装の難しさを改めて示す結果となった。本稿では、同システムの概要を示し、筆者らのエージェントの協調を述べ、筆者らのこれまでの成果を紹介する。

2 RoboCupRescue プロジェクト

RoboCupRescue プロジェクトは、大規模災害に対してロバストな社会を創ることを最終目標とするプ

ロジェクトである [1]。現在、シミュレータ上の仮想的な災害空間で活動するエージェントを研究するシミュレーション・プロジェクトと、災害現場で実際に活動するロボットなどを研究するロボティクス & インフラストラクチャ・プロジェクトが進められている。

シミュレーション・プロジェクトが開発している大規模災害シミュレーションシステムは、マルチエージェントシステムのテストベッドとして利用可能な段階に入った。2001 年 4 月に福岡で開催された RoboCupRescue JapanOpen を皮切りに、2001 年 7 月の RoboFesta 関西、8 月の RoboCupRescue2001 in USA において、同システム上で活動するエージェントの救助能力を評価するための競技会が開催された。シミュレーション・プロジェクトは、同システムを発展させることにより、防災情報インフラと結合させ、実際の災害情報とシミュレーションを融合したシステムにし、より有効な意思決定支援に役立てることを目標としている。

3 エージェントの活動目的と災害空間

大規模災害シミュレーションシステムで活動するエージェントには、市民、救急隊、消防隊、道路の寸断を修復する道路啓開隊、通信を行う救急、消防、啓開各隊の司令部の 7 種類がある。エージェントの活動目的は死者数を最小にすることであり、同異種のエージェントが協調することで負傷者を救助する。

同システムは災害発生直後の市街地をシミュレートする。この市街地を災害空間と呼ぶ。死傷の主な原因には埋没と火災が挙げられる。震災直後には家屋

Takeshi Morimoto, 電気通信大学 情報工学科

Kenji Kono, 電気通信大学 情報工学科

Ikuro Takeuchi, 電気通信大学 情報工学科

が倒壊し多数の市民が閉じ込められ負傷する。更に、家屋から逃げられず火に巻き込まれれば、致命的なダメージを被る。救急隊は一刻も早く負傷者を見つけ出し救助しなければならないが、道路は瓦礫などで寸断されており、負傷者の所まで移動できない。負傷者の救出には、まず道路啓開隊が啓開活動により救急隊と消防隊の移動経路を確保し、消防隊が消火活動により負傷者が火に巻き込まれるのを防ぐ必要がある。

エージェントは同異種間で協調する必要があるが、個々のエージェントの視覚能力と通信能力は非常に制限されており、エージェントは限られた情報だけで協調を実現しなければならない。加えて、災害空間はダイナミックに変化する。負傷者は徐々に消耗し、火災は急激に広がり、他のエージェントは動きまわる。つまり、行動の対象自体が時々刻々と変化する。エージェントは、行動対象の変化に応じて、活動と協調のプランを柔軟に変更できなければならない。

4 大規模災害シミュレーションシステム

4.1 構成と処理

シミュレーション・プロジェクトが開発している大規模災害シミュレーションシステムは、複数のモジュールから構成される(図1)。モジュールには、個々の災害をシミュレートする Sub-simulator、災害空間の地理情報を与える GIS (Geographical Information System)、災害空間の状況を視覚化する Viewer、これらを統合する Kernel がある。個々のエージェントも、システムを構成するモジュールの一つとして位置づけられる。

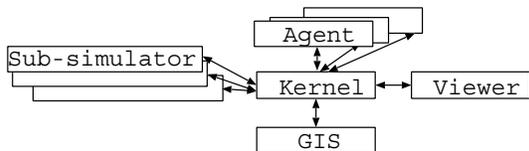


図 1 シミュレーションシステムの構成

災害空間のシミュレーションは 1 分刻みで進行する。災害空間上の 1 分を 1 ターンと呼ぶ。エージェントは 1 ターン毎に、周囲の状況を知覚し、行動を決

定し、それをシステムに伝える。行動は次のターンに反映される。具体的には、以下の繰り返しによって、シミュレーションが進行する。

1. Kernel が個々のエージェントに視覚情報を送る
2. エージェントが Kernel に行動内容を伝える
3. Kernel が行動内容を Sub-simulator に送る
4. Sub-simulator が更新結果を Kernel に伝える
5. Kernel が統合結果を Sub-simulator に送る
6. 災害空間の時間を 1 分進める

エージェントは視覚情報を受けてから 1 ~ 2 秒以内に行動を決定しなければならない。

4.2 災害空間のモデルと問題点

災害空間は建物や道路、人間等のオブジェクトの集合として表現され、個々のオブジェクトは複数のプロパティを持つ。災害空間は 500 m 四方の市街地であり、約 2500 個のオブジェクトによって構成される。

オブジェクトには、一般家屋、避難所、各隊の司令部といった建物オブジェクト、道路とその端点を表すオブジェクト、川とその端点を表すオブジェクト、市民、救急隊、消防隊、道路啓開隊、車といった移動物体オブジェクト、災害空間の場所や風向き等を表す世界オブジェクトがある。しかし、全 15 種類のうち、世界、川、川の端点、車の 4 種類は実質的には使用されない。災害空間上に存在するのは、建物と道路、人間だけである。

建物には 13 個のプロパティがあり、道路には 15 個、人間には 8 個あるが、システムにより実装され、かつエージェントの問題解決に必要なプロパティはわずかである(表1)。

表 1 問題解決に必要なプロパティ

オブジェクト	プロパティ
建物	位置, 燃焼度, 面積
道路	位置, 幅, 閉塞幅, 車線数
人間	現在地, 生命力, 負傷度, 埋没度

災害空間の仕様は複雑だが、その複雑さには問題設定を豊かにするという点において有意味とはいえない部分がある。加えて、複雑な仕様が開発が追いつ

表 2 エージェントの能力

種類	能力
市民	Sense, Hear, Move(Route), Say(String)
救急隊	Sense, Hear, Move(Route), Say(String), Tell(String), Rescue(Human), Load(Human), Unload
消防隊	Sense, Hear, Move(Route), Say(String), Tell(String), Extinguish(Building)
道路啓開隊	Sense, Hear, Move(Route), Say(String), Tell(String), Clear(Road)
救急隊司令部	Hear, Say(String), Tell(String)
消防隊司令部	Hear, Say(String), Tell(String)
道路啓開隊司令部	Hear, Say(String), Tell(String)

かず、現在開発中のプロトタイプ・シミュレーションシステムの実装は不安定である。不安定な実装の上では、協調以前に、単体で活動できるエージェントの実現すら難しく、実際、筆者らがエージェント開発に要した労力の大半は、単体行動の実現のために費された。

このような問題設定の難しさや、それに付随するシステムの安定性の問題は、オープンな事象を扱うマルチエージェントシステムが本質的に抱える問題である。RoboCup サッカー [2] でも、我々はこれと同じ問題に直面したが、その際もやはり、協調の前に、まず単体の行動能力の向上が必要であるという認識に立ち、世界トップクラスのサッカー・エージェントを開発した [3]。複雑な問題設定の中では、マルチエージェントの協調について、トイ的な問題をテストベッドとするのとはまったく別の苦労がある。

4.3 エージェントのモデル

個々のエージェントは、数人の人間の集まりを表している。市民は家族を、救急、消防、道路啓開隊は小隊を、司令部はそこで働く人達の集まりである。災害空間上では、72 家族の市民と、5 隊の救急隊、10 隊の消防隊、10 隊の道路啓開隊、そして司令部が各 1 棟活動する。

エージェントは、行為とその対象をシステムに伝えることにより行動を起こす。ある経路 *route* に沿って移動するには、*Move(route)* をシステムに伝える。経路は道路や建物のリストで表現する。同様に、人間 *human* の救助は *Rescue(human)*、救急車への乗せ降ろしは *Load(human)*、*Unload*、火災 *building* への放水は *Extinguish(building)*、道路 *road* の啓開は *Clear(road)* をシステムに伝える (表 2)。

エージェントが情報を獲得する手段には、視覚と通信がある。個々のエージェントは、火災と半径 10m 以内の建物や道路、他のエージェントなどを視覚でき、肉声と電気通信による通信ができる。肉声は話者から半径 30m 以内にいるエージェントに聞こえ、小隊による電気通信は同種の小隊と司令部に伝わる。司令部による電気通信では、他種の司令部にもメッセージが伝達される (図 2)。個々のエージェントの発話

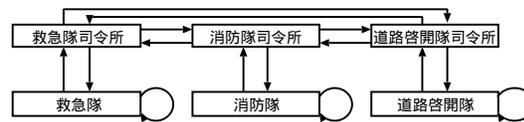


図 2 電気通信の流れ

と聞き取りはそれぞれ 1 ターンに 4 メッセージに限られている。聞き手は、届いたメッセージの中から 4 つを話者によって選別し、内容を聞き取ることができる。1 つのメッセージの情報量は、1 センテンス程度に制限されている。小隊と司令部の通信回数の制限が同じであるのは一見すると不自然だが、実際に東京消防庁に置かれている通報用の電話の台数が 10~20 台程度であることを考えれば、妥当な問題設定である。

しかしながら、エージェントの通信能力の制約は非常に厳しい。メッセージの伝達には 1 ターンの遅延がある上に、肉声は伝えられる対象が限られており、電気通信はメッセージの到達率が低いからである。

遅延は同種の小隊間ならば許容範囲だが、異種小隊間の電気通信の伝達経路は、話者 同種司令部 異種司令部 異種小隊と長く、3 ターンもの遅延が生じる。

肉声ならば、異種小隊間でも遅延を 1 ターンに抑

異種端末間の協調による高齢者・障害者の移動支援, Robotic Communication Terminals

矢入(江口) 郁子 猪木 誠二

1 はじめに

人間にとって「移動」とは、自立的かつ快適に生活するための手段として極めて基本的、かつ必要不可欠な行動である。しかし現在の社会環境下では、視覚・聴覚・下肢駆動機能に障害を持つ高齢者・障害者の場合、移動に不可欠な認知・駆動・情報入手の3つの要素行動に問題が生じるため、自立的な移動が困難となっている。そこで筆者らは、幅広い高齢者・障害者を対象に、障害の種類、レベル、障害歴の違い、重複障害などの多様性を考慮に入れ、認知・駆動・情報入手の3つの要素行動を補助する移動支援システム, Robotic Communication Terminals(以後, RCT と略記)の実現を目指し、研究を行っている[矢入 01]。これまで高齢者・障害者の移動支援の研究では、携帯電話や自律走行ロボットを使った視覚障害者のナビゲーションシステム[東海 99][Kemmerling 98][小谷 97]や、肢体障害者のための知能化車椅子[足達 99]など、高齢者・障害者のうちの一部のユーザを対象に、移動の問題の一部の解決を試みる研究は数多くなされてきた。一方、健常者を主対象にした研究では、認知・情報の入手の補助を行う歩行者用 ITS の実現が国の政策として位置付けられ[国土交通省 01]、産業界では携帯電話や環境に埋め込むビーコンを用いた早期実

現可能なシステムの開発が活発化している。RCT ではこれらのシステムの先を見据え、環境やユーザのモニタリング、異種端末間の協調などの技術を用いた、より広い観点での認知・駆動・情報の入手の補助方法を提案し、歩行者用 ITS の高度化に貢献することを目指す。

本稿では、RCT の概要、研究目標を示し、マルチエージェント研究の観点からみた RCT の位置付けを述べる。そして最後に、筆者らのこれまでの成果および今後の展開を紹介する。

2 Robotic Communication Terminals

2.1 RCT による移動支援

RCT は、以下の3つのタイプの端末からなる。
<環境端末> アンテナのような役割を持ち、道路や駅などに設置される端末である。それぞれの環境端末は設置された場所の周辺を監視し、障害物や人・車の有無などの環境に関する情報を検出する。環境端末が検出した情報は、周辺を移動中のユーザに伝えられるだけでなく、遠隔地のユーザにも移動経路を決めるための材料として提供される。

<ユーザ携帯型移動端末> 携帯電話や PDA を進化させた端末である。この端末を持ち歩くユーザはインターネットを通じて目的地への経路案内や交通機関の事故などの情報を入手する。そしてさらに環境端末との通信によって、周辺や遠隔地の障害物や人・車の有無などの情報を入手することも可能である。これらの情報は、ユーザの身体の状態に合わせて加工され、ユーザの身体の状態に合わせて設計された画像・

Robotic Communication Terminals Supporting Mobility of the Elderly and Disabled People by Communications among Heterogeneous Systems

Ikuko Eguchi YAIRI, Seiji IGI, 独立行政法人通信総合研究所, Communications Research Laboratory, Independent Administrative Institution

音声・触覚を利用したインタフェースを介して提示される。

<ユーザ搭乗型移動端末> 高齢者・障害者のための乗り物を進化させた端末である。ユーザ携帯型移動端末の機能に加え、ユーザの操縦通りに運転できる通常の乗り物としての機能、さらにセンサー及び環境端末からの情報を利用した障害物の自動回避などの運転支援機能も装備される。また、この端末にはユーザの身体の状態に合わせて設計されたハンドルやジョイスティックなどの操縦系、電動のスクータや車椅子のボディが用いられる。

これらの役割の異なった端末同士が通信し互いに協力しあうことで、認知・駆動・補助の 3 つの要素行動を補助し、ユーザの市街地での移動を支援する。図 1 に RCT による移動支援のイメージを示す。例えば移動端末は、移動中のユーザから離れた場所をモニタする環境端末の「こっちからは何も来ないよ」というメッセージを近傍の環境端末を通して入手し、ユーザに「前からは何も来ません」という情報を伝えることができる。また、自転車を検出した環境端末は、移動方向や位置から危険と判断した近傍の移動端末に、「自転車きたよ、ぶつかる可能性があるよ。」というメッセージを送る。それを受けた移動端末は、例えばユーザ搭乗型であれば自動停止してやり過ごすなどの補助を行ったり、ユーザに危険情報を伝えることができる。

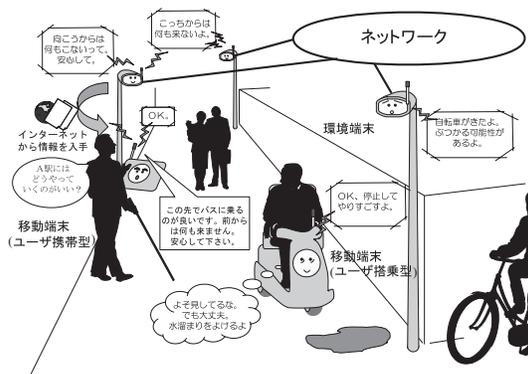


図 1 RCT による移動支援のイメージ

2.2 RCT の目標

RCT では、5 年以内に以下の目標 1~3 の要素技術を実現し、システムの実用化への基礎を築くことを目指している。

<目標 1> ユーザの多様性の適切な分類方法と、その分類に基づいた支援タスクおよび移動端末のハードウェアデザイン、移動端末のソフトウェアのユーザへの適応方法

RCT では高齢者・障害者の多様性を考慮した支援タスクを、個々のユーザに適した方法で実行する移動支援を目指す。具体的には (1) 移動に不可欠な認知・駆動・情報入手の 3 つの要素行動と身体の状態との関係に注目し、ユーザを 20~30 個程度のグループに分類、(2) そのグループ毎に支援タスク、移動端末のハードウェアのデザインの指針を設定、(3) 移動端末のソフトウェアが、入力された身体の状態や嗜好などに合わせて動作したり、使い込むうちにユーザの癖や習慣に合わせて適応する仕組みを実現する。

<目標 2> 環境端末、移動端末およびユーザ間の協調によって、移動支援タスクを場面、状況、ユーザの状態に合わせて実行するためのシステムアーキテクチャ

環境端末、ユーザ搭乗型・ユーザ携帯型移動端末の 3 つのタイプに分類される多様な端末からなる分散協調システムとして全体システムを構成する。また、RCT のサブシステムとしての環境端末および移動端末を、環境やユーザとインタラクションし適応するロボット的なシステム、すなわちロボティックシステムとして構成する。そして、移動端末とユーザとの密なインタラクションによって、RCT の全体システムとユーザ間の協調を実現する。これらのシステムアーキテクチャをプロトタイプシステムの製作を通して示す。

<目標 3> 実世界、ネットワーク世界からの情報の取り込み・加工・蓄積、およびその情報の伝達方式

RCT では、認知・駆動・情報の入手の移動の要素行動の補助のうち、情報の入手の補助に最も注目している。「実世界」「インターネット」の 2 つをユーザに提供する情報のリソースとし、「専用の情報提供用サーバの構築と利用」「既存のインターネット上の多

数のサーバ上の情報の有効利用」の 2 つの方法による情報の入手の補助の具体的なアプリケーションを提案し、プロトタイプを作成する。そして、2 つの情報リソースからユーザに情報を提供するまでの、取り込み・加工・蓄積・伝達の方式を示す。

2.3 マルチエージェント研究としての RCT

RCT は以下の 3 つの種類のエージェントから構成されるマルチエージェントシステムとして捉えることができる。

- (1) 端末内部エージェント: 異なる機能を持ち、群として環境端末、移動端末の単体端末の内部を構成する。
- (2) 端末間通信エージェント: 個々の環境端末、移動端末を代表し、端末間の通信を担い、協調動作を実現する。
- (3) 外部通信エージェント: インターネットなどのネットワーク上に構築された他のシステムとの通信を担い、協調動作を実現する。

認知・駆動・情報の入手の補助タスクがどの種類のエージェントによって実行されるかを表 1 に示す。

表 1 RCT の 3 つの種類のエージェントと補助タスクの
実行 (はユーザ搭乗型移動端末のみ)

補助タスク	(1) 端末内部 エージェント	(2) 端末間通信 エージェント	(3) 外部通信 エージェント
認知			—
駆動		—	—
情報の入手			

認知の補助タスクに関しては、端末内部エージェントが端末の周囲の環境の認識を行い、端末間通信エージェントが認識結果を近傍の端末に送り届けるという役割分担がなされる。また情報の入手の補助タスクに関しては、端末内部エージェントがユーザの要求の解釈および情報の加工・提供を行い、端末間エージェントが個々の端末に蓄積された実世界データを探索・獲得し、外部通信エージェントがインターネットなどのネットワーク上に構築された他のシステムに蓄積されたデータを探索・獲得するという役割分担がなされる。

端末内部エージェントのように、機能の異なったエージェント群で一つの知的システムを構築するアプローチは、古くから AI およびロボットの分野で研究がなされてきた。また端末間通信エージェントのように、複数の端末間の協調によってタスクを実行するアプローチは、群ロボットによるタスク実行の研究や [Balch 94]、端末間を移動するエージェントによる通信の研究 [佐藤 00] などがなされている。そして外部通信エージェントのように、インターネット上のサーバのデータから有効な情報を探索・獲得し提供するアプローチは、サーチアルゴリズムの研究や [Levy 96]、データマイニングの研究 [Cooley 97] などの多数の研究がなされてきた。しかし、これらのアプローチの全てを含む統合的なマルチエージェントシステム研究は例を見ない。RCT は人間の実世界での移動を支援するこれまでにない複雑なマルチエージェントシステムの枠組を示す一つのテストベッドであると言える。

3 筆者らの研究の現状

研究開始から現在までの約 2 年間の成果と今後の計画を述べ、研究の現状を紹介する。

3.1 ユーザの調査

2.2 の目標 1 のユーザの多様性の分類の課題への取り組みとして、以下の (a) ~ (c) に示す調査を行ってきた。

(a) 高齢者・障害者の身体の状態と移動の問題との関係の予備調査

身体の状態と移動問題の関係を具体化するため、高齢者および視覚・聴覚・下肢駆動機能に障害をもつ障害者、計 13 名を対象にヒアリングを実施した [Yairi 00]。その結果、視覚障害者は未知の環境・人のいない環境・変化が多い環境での移動に、聴覚障害者は視野外から接近する動物体の認識、アナウンス情報の入手に、そして下肢駆動機能障害者は転倒防止のための歩行面の詳細な状態認識、パリアフリー情報の現場での探索による入手に、それぞれ問題が集中することがわかった。また、障害が重く障害者となった年齢が高いほど、問題の危険度や頻度が増すこと、および使用する補助器具や障害の種類に固有の特徴

的な問題の存在も明らかになった。高齢者の場合は、問題点を絞り込むことは困難であったが、前述した障害者の抱える問題のいずれかに分類されることが多かった。また、加齢により情報の入手を面倒と感じるようになり外出の意欲が低下してきた、という訴えが特徴的であった。

(b) 身体の状態と移動の問題との関係からのユーザ分類の可能性の予備調査

(a) の結果をもとに身体の詳細な状態の質問と、移動の問題を 5 段階の数値で回答する質問からなる調査表を設計し、アンケート調査を実施した。調査表は首都圏の高齢者・障害者団体を通じて配布され、計 698 名の回答を得た。障害の種類、レベル、使用する補助器具、性別、障害の先天性/後天性に関する身体の状態の分類ごとに、移動の問題に関する質問の回答の平均を比較したところ、分類項目と回答の間に強い相関関係の認められる質問や、ある分類項目を境に回答に有意差が認められる質問の存在が明らかになった。このことにより、障害の種類、レベル、使用する補助器具、性別、障害の先天性/後天性をパラメータとして、抱える移動の問題の違いからユーザをグループに分類可能であること、およびパラメータからユーザの抱える移動の問題を推定可能であることが確かめられた。

(c) ユーザ分類のための第 1 次調査

(b) の結果を踏まえ、ユーザ全体を大まかに分類するために全員に同一の調査表を用いてアンケートを実施する第 1 次調査と、さらに詳細なグループにユーザを分類するためにグループ毎に異なった調査表を用いてアンケートを実施する第 2 次調査を計画した。現在は第 1 次調査を実施中である。(b) では、身体障害者の約半数が 65 歳以上であるという現状から、障害者の中年層、若年層の回答数が不十分となり、年齢や障害歴による分類の可能性を確認することはできなかった。そこで今回の調査では中年層、若年層も含めた十分な数の回答を得るため、東京、大阪、京都の 3 大都市の高齢者・障害者団体を通して調査表を配布し、2001 年 10 月現在で約 3500 名の回答を得た。現在は、移動の問題の生じるメカニズムを表現するモデルを決定するための共分散構造分析や、グループに分類するための因子分析などを行っている。ユーザ分類

のためのデータ解析はクラスタリング問題の一種とみなすこともでき、従来手法 [Cheeseman 88] [Fisher 87] を適用してみる予定である。

3.2 研究インフラの構築

2.2 の目標 2 および 3 に取り組むための研究インフラとして、以下の (a) - (c) に示すプロトタイプシステムを作成した。

(a) 画像による道路監視システム

環境端末の道路監視機能のプロトタイプとして、画像による道路監視システムを製作した。このシステムは、図 2 に示す屋外に設置するカメラ、デュアル CPU の PC、キャプチャボードと画像処理ライブラリから構成される。そして、動物体の形状、位置、移動速度を検出し、歩行者や自転車、自動車などの動物体の種類や、動物体 静止する障害物 動物体という状態変化の判別を行う [吉水 01]。さらに移動端末の位置情報を通信で受け、検出した動物体との位置関係から、動物体を危険物体と判断し、移動端末へ通知することが出来る。道路監視システムの動作の様子を図 3 に示す。下部右側の GUI では監視すべき道路の部分として自動的に切り出された領域の画像が、下部左側の GUI では道路上の動物体として検出された部分の画像が表示されている。そして上部の GUI では検出された動物体が四角形で囲まれ、人、自動車、自転車などの種類の判別結果が文字で示されている。本システムの現在の処理速度は、1 フレーム中の動物体の数が 10 個以下の状態で毎秒 5 フレームである。また性能は、晴天の日中かつ無風状態であれば、移動端末へ



図 2 道路監視システムの屋外設置カメラ 2 基の外観

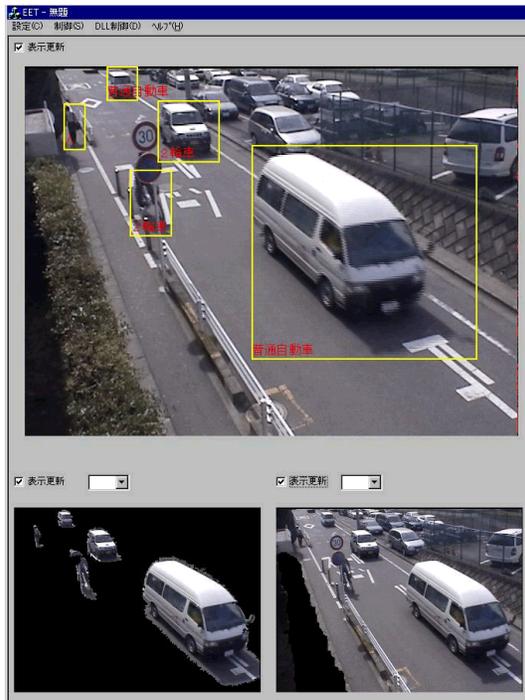


図 3 道路監視システムの動作の様子

の安定した危険物体の通知が可能である [香山 01] .

今後は、処理の高速化と、天候の条件の緩和、夜間への対応などの工夫を重ねながら、画像以外のセンサの導入による環境端末のアーキテクチャの提案、実世界情報の蓄積・ユーザへの伝達方式の実現などの課題に取り組む予定である .

(b) 高齢者向けユーザ搭乗型移動端末

移動端末のプロトタイプの一つとして、足腰が少々弱った高齢者向けユーザ搭乗型移動端末、Intelligent City Walker (以後、ICW と表記) のハードウェアを製作した . 図 4 に全体像を示す . ICW には、障害物を検出するための超音波・赤外線センサ、障害物回避や操縦補助のための制御装置、他の端末との通信やインターネット接続のための無線 LAN、ユーザと端末との対話のための装置としてのタッチパネルが搭載されている [南山 01] . ICW は他の端末との通信機能、ユーザとの対話機能、ユーザの運転補助機能を持つセミオートノマススクータである . 現在は運転補助機能を中心に全機能のインプリメントを行っている . 図 5 に目の前に飛び出してきた障害物 (段ポ-

ル箱) を回避する屋外での実験の様子を示す . 搭乗者は回避動作をより効果的にデモするためにハンドルから手を離している . また、見通しの悪い交差点で環境端末から自動車との衝突の可能性の通知を受け、交差点手前で自動停止し、通知された自動車をやりすごす実験も行なった . さらにこれらの実験に加えて、高齢者の試乗による評価実験 [田中 01] を行い、山口きらら博などの展示会での一般人の反応とあわせて、ユーザインタフェースの設計に活用している .

今後はセンサやユーザとの対話用装置の改良を重ねながら、運転補助機能の性能の向上、対話インタフェースのデザインの向上を行い、移動端末のシステムアーキテクチャの提案の課題に取り組む予定である .

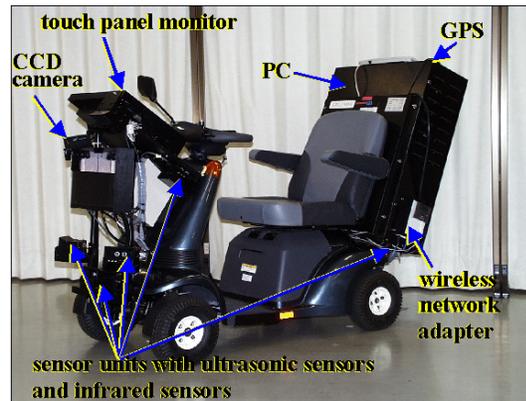


図 4 高齢者向けユーザ搭乗型移動端末 ICW



図 5 屋外での障害物回避実験の様子

(c) バリア・バリアフリー情報地図データベース

実世界の情報を専用のサーバを用いてインターネット経由で発信し、ユーザの情報の入手を補助するための具体的アプリケーションとして、歩行者用道路に関するバリア・バリアフリー情報地図データベースのプロトタイプを製作している。このデータベースは、東京都小金井市(国分寺市の一部を含む)の500分の1の高精度の地図に、これまでに実現されていなかった歩行者用道路ネットワークを付加し、段差などのバリア情報、盲人用信号などのバリアフリー情報を道路ネットワークにリンクさせて埋め込んだものである[桑原 01][吉岡 01]。図6に道路ネットワークの概要を示す。道路ネットワークは歩行者用道路を表すライン、ラインの始点・終点としてのノード、ラインの近傍に存在する停留所などの施設を表しラインにリンクされるポイント、の3種類のオブジェクトから構成される。図7に埋め込む情報の種類、項目をライン、ノード、ポイントの階層構造から示す。これらの埋め込む情報の種類、項目は、視覚・聴覚・下肢駆動機能に障害をもつ障害者・高齢者、および障害者・高齢者支援団体の指導者の計14人に、詳細なヒアリングを実施して決定された。また地図データベースに蓄積した情報は、実際に現地を踏破して収集された。

現在は、2002年中にインターネット上で一般に公開することを目指し、表示やルート検索用インタフェースを製作中である。今後は身体の状態に適したルートの検索やナビゲーション文を自動生成するソフトウェアの実現を試みながら、道路監視システムとのドッキングによって、データベースに実世界情報を蓄積する課題にも取り組む予定である。

4 おわりに

本稿では、異種端末の協調によって高齢者・障害者の自立的移動を支援する Robotic Communication Terminals を紹介した。そして、マルチエージェント研究としての位置付け、筆者らの取り組みの現状や今後の展開を述べた。

RCTによる移動支援は、健常者を対象とした支援にも容易に適用可能であると期待される。それに関わらず高齢者・障害者の支援に限定して研究対象と

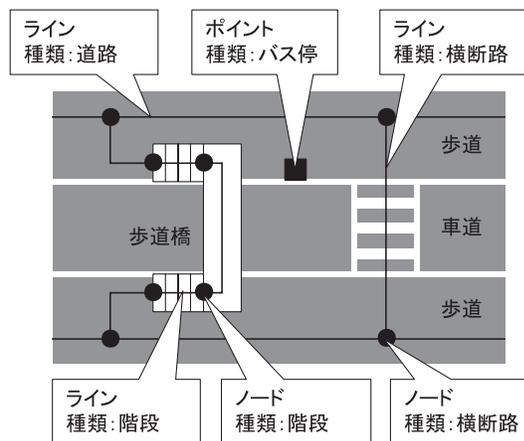


図 6 歩行者用道路ネットワーク概要

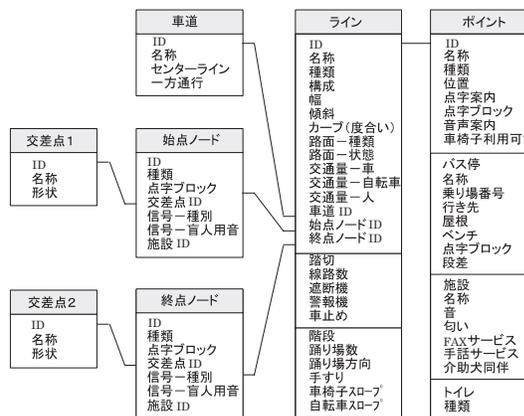


図 7 バリア・バリアフリー情報 DB 概要

する理由には、高齢者・障害者を健常者とは異なるグループとして切捨てて支援システムを実現し、後付けでバリアフリーの実現を目指す従来の手法と異なり、まず高齢者・障害者の立場から網羅的な支援方法を考え、高齢者・障害者の延長線上に健常者が位置すると捉えることこそが、真の意味でのバリアフリーのつながるといふ筆者らの信条が挙げられる。

また RCT は、このような社会的貢献の面だけでなく、画像・音声認識、プランニング、モバイルエージェント、情報検索、データマイニングなど、マルチエージェント以外にも多数の AI 研究のトピックを含む学術的研究対象であるという一面も持つ。筆者らの取り組みは、これらのトピックの研究を行うための一揃いの研究インフラの構築が完了した段階にあると

言える。今後は、製作中のプロトタイプシステムの完成度を高めながら、ソフトウェアの実装を中心に AI 研究のトピックに深く切り込み、学術的な貢献を目指したい。

参考文献

- [足達 99] 足達, 中西, 他: 顔の視覚情報処理を用いた知的車椅子. 日本ロボット学会誌, vol.17, no.3, pp.113-121 (1999).
- [Balch 94] Balch, T. and Arkin, R. C., :Communication in Reactive Multiagent Robotic Systems. Autonomous Robots, 1, pp.1-25(1994).
- [Cheeseman 88] Cheeseman, P., et al. : Autoclass : A bayesian classification system. In Proceedings of the Fifth International Workshop on Machine Learning, pp. 54-64 (1988).
- [Cooley 97] Cooley, R., et. al. : Web Mining: Information and Pattern Discovery on the World Wide Web, Proceedings of the 9th IEEE International Conference on Tools with Artificial Intelligence (IC-TAI'97), November 1997.
- [Fisher 87] Fisher, D., H. : Knowledge Acquisition via Incremental Conceptual Clustering. Machine Learning, 2, pp.139-172(1987).
- [香山 01] 香山, 吉水, 他: 高齢者・障害者の移動を支援するロボティック通信端末のための一般道路監視システム. 第 19 回日本ロボット学会学術講演会, 1M23 (2001).
- [Kemmerling 98] Kemmerling, M., et al. : An Orientation and Information System for Blind People based on RF-Speech-Beacons. TIDE third congress, <http://www.dinf.org/tide98/tide98.htm>, Helsinki, July (1998).
- [国土交通省 01] 国土交通省: ITS <http://www.its.go.jp/ITS/j-htm/index/indexHBook.html> (2000).
- [小谷 97] 小谷, 清弘, 森: 視覚障害者のための歩行ガイドロボットの開発. 映像情報メディア学会誌, Vol.51, No.6, pp.878-885 (1997).
- [桑原 01] 桑原, 矢入, 他: 高齢者・障害者向け移動支援 GIS. AI 人工知能学会全国大会, 3A1-05 (2001).
- [Levy 96] Levy, A. Y., et. al. : Querying Heterogeneous Information Sources Using Source Descriptions. Proceedings of the Twenty-second International Conference on Very Large Databases, Bombay, India, pp.251-262(1996).
- [南山 01] 南山, 永合, 他: 高齢者移動支援のためのユーザー搭乗型端末 (IWC: Intelligent City Walker). AI 人工知能学会全国大会, 3A1-04 (2001).
- [佐藤 00] 佐藤, :モバイルエージェント. コンピュータソフトウェア, Vol.17, No.2, pp45-54(2000).
- [田中 01] 田中, 矢入, 他: 高齢者移動支援のためのユーザー搭乗型移動端末 (ICW: Intelligent City Walker) - 運転補助機能実装のためのユーザー評価実験. ヒューマンインタフェースシンポジウム 2001, pp.481-484 (2001).
- [東海 99] 地域振興のための電波利用に関する調査研究会: 視覚障害者を支援する情報通信システムに関する調査研究報告書, 東海電気通信監理局電波監理部企画課, March (1999).
- [Yairi 00] Yairi, I. E. and Igi, S.: A Self-sustained Moving Support System for Aged and Disabled Persons. In Proc. The 6th International Conference on Intelligent Autonomous Systems, pp.692-697 (2000).
- [矢入 01] 矢入, 猪木: 高齢者・障害者の自立的移動を支援する Robotic Communication Terminals(RCT). 人工知能学会誌, vol.16, no.1, pp.139-142 (2001).
- [吉水 01] 吉水, 矢入, 猪木: 歩行者支援のための外乱を考慮した道路監視システム. AI 人工知能学会全国大会, 3A1-02 (2001).
- [吉岡 01] 吉岡, 矢入, 他: バリア・バリアフリー情報を蓄積・利用した移動支援 GIS. ヒューマンインタフェースシンポジウム 2001, pp.479-480 (2001).

えられるが、半径 10m の視野では他のエージェントの位置の把握が難しい上に、30m の肉声伝達範囲では声が届く前に移動してしまっている場合が大半である。肉声による通信は、埋没した市民、や司令所のような動かない聞き手に対してのみ有効であるといえる。

電気通信には、話すメッセージが多い程、聞き手がメッセージを聞く確率が低くなるという、通信量と到達率のジレンマがある。例えば、消防隊 10 隊と消防隊司令所 1 棟がそれぞれ 4 回電気通信した場合、次のターンには、各消防隊に 40 ものメッセージが届く。消防隊が内容を聞き取れるのは 4 メッセージだけであるから、届いたメッセージの 10% しか聞き取れないことになる。したがって、重要なメッセージを聞き逃さないためには、各エージェントが、最小限しか話さないようにする必要がある。

5 実現したエージェント

5.1 単体の行動能力の向上

複数のエージェントが協調的活動を行うためには、個々のエージェントが基礎的な行動を確実に遂行できる能力を備える必要がある。RoboCupRescue における基礎能力とは、迅速な移動能力である。筆者らは、通過可能性に基づいた経路探索により到達確率の高い経路を構成することで、迅速な移動を実現した。これには災害空間の以下の性質を利用した。

- 一度通過した経路は渋滞を除き確実に通過できる
- 閉塞は道路毎に判定されるため、未確認道路数が少い程、閉塞遭遇率が低くなる
- 車両は高速に移動できるため、移動距離を考慮しなくてよい

経路構成の方針は、まず通過可能を確認した道路のみで構成を試み、次に未確認道路数最少の経路の構成を目指す。それでも目的地までの経路が見つからなければ、閉塞も含む経路を構成する。このアルゴリズムは、最良優先探索に、以下のコスト関数(式 1)を与えることで実現できる。ここで、 M は最大経路長より大きな値とする。

$$\text{costFunc}(\text{Road}) = \begin{cases} 1 & \text{通過可能} \\ M & \text{未確認} \\ M^2 & \text{閉塞} \end{cases} \quad (1)$$

こうして得られた経路上の移動を繰り返すことで、エージェントは閉塞状況をリアルタイムに学習しながら、その時点で知り得る最良の経路を移動する。

5.2 救急隊

死傷者に直接関わる救急隊の活動は非常に重要である。多数の市民が倒壊家屋に埋もれているのに対し、救急隊は 5 隊しかいない。救急隊は、救助活動の効果を最大限に引き出すために、負傷の程度によって負傷者を選別し、生命の危機に瀕している負傷者を最優先に救助する“トリアージ”[4]に基づいて活動する。具体的には、負傷者を以下の 4 段階に分けて救助する。

1. 即救助 ... 救助しないと助からない
2. 救助 ... 自力で避難はできないが当面生き残れる
3. 救助不要 ... 自力で避難できる
4. 救助断念 ... 救助しても助からない

救急隊は「即救助」「救助」の負傷者を発見した時のみ救助を始め、「救助不要」「救助断念」の負傷者を発見しても捜索を続ける。同じ場所に複数の負傷者がいる場合には、まず「即救助」の救助を優先する。優先度が同じ場合には、确实のため、埋没度の少い負傷者の救助を優先する。

負傷者は時々刻々と消耗し傷も深くなるため、迅速な救出活動が必要である。救出にかかる時間を短縮するためには、救助能力を集中させなければならない。救急隊は、「即救助」「救助」の優先度をもつ負傷者を発見した時に、電気通信によって他の救急隊に報告し、報告を受けた救急隊はその場所に駆けつける。複数の報告が同時に届いた場合には、迅速に到達できる場所にいる負傷者の方に駆けつけ、別の負傷者の救助中に報告を受けた場合は、移動せず救助を続ける。この協調により、負傷者を救出できる確率が向上する。

5.3 消防隊

火災は状態が急激に変化するため、特に素早い対応が必要な災害である。ごく初期の火災は 1 隊の消防

隊でも消火可能だが、それ以降は多数の消防隊による集中放水が必要となる。消防隊には、目的の火災に即座に移動できる能力が強く求められる。また、災害空間には同時に複数の火災が発生するため、火災の状態を見極め、放水を適切に振り分けなければならない。

通信は遅延を伴うため、急変する火災への放水の分散を、通信で実現することは難しい。各自の判断で迷わず消火に取り組む方が効果的だと考えられる。火災は視界に関係なく見られるため、全ての消防隊は、火災自体の状態とその周囲の火災状況の認識を共有している。個々の消防隊は、これらを基に打ち立てた消火戦略を軸に、通信を使わない暗黙の協調を実現する。

個々の消防隊は、火災の状態、火災周囲の状況、火災と自分との距離から、火災の放水優先度を見積り、優先度が高い火災に放水/移動する。図 3 の数字は優先度を表す。1 が最優先であり、これは放水範囲内であって、消火価値が高く、ごく初期段階の火災を表す。もしこの条件を満たす火災があれば、消防隊はそれに放水する。

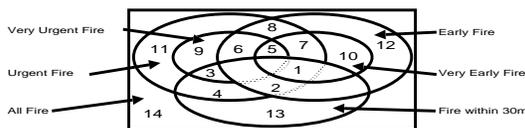


図 3 火災の放水優先度

火災の消火価値は、その周囲 18m の火災状況から決める。周囲に未着火の家屋が沢山ある場合、消火価値は高い。逆に火災が沢山ある場合や鎮火済みの家屋しかない場合、その火災は焼け野原の中にあり消火価値は低い。ただし、未着火と鎮火済みの両方がある場合には、鎮火家屋は延焼防止線の役割を果たすため、その火災を消火すれば延焼防止線を結べるため、消火価値が高い。消火価値は以下のプログラムから求める。

```
消火価値 (Fire) = 未着火数-火災数 × 減少率
if 未着火数 > 0
then 未着火数 += 鎮火数
```

協調機構はトリビアルである。消防隊は移動を極力避け、放水に専念したい。上述の戦略に基づいた行動選択過程において、遠くの火災に移動する際に、移動率(式 2)の確率で移動するかしないかを判断することにより、他の消防隊が該当火災に放水している可能性が高いならば、移動せずにすぐに放水できる火災を優先する。

$$\text{移動する確率} = \frac{\text{目的の火災数}}{\text{消防隊数}} \quad (2)$$

5.4 道路啓開隊

救急隊と消防隊にとって、移動は救助・放水のための手段であり、移動に時間をとられたくない。道路啓開隊は彼らの移動回数を減らすことを目的とする。啓開隊は閉塞に遭遇しても自力で道を切り開けるため、未確認道路の再探索に時間を費すよりも、最少道路数の経路を突き進む方が効率が良い場合が多い。また、最少道路数経路の確保は、他のエージェントの経路構成方針(式 1)とも合致しており、暗黙の協調 (RoboCup サッカーにおける locker-room agreement [5]) を実現しているといえる。最良優先探索に与えるコスト関数は以下ようになる。

$$\text{costFunc}(\text{Road}) = \begin{cases} 1 & \text{通過可能} \\ M & \text{未確認 or 閉塞} \end{cases} \quad (3)$$

さらに、道路啓開隊は、他のエージェントからの啓開要請に対応する。救急隊と消防隊は、通りたのに通れない道路と、通りたかったのに通れなかった道路を道路啓開隊に報告し、道路啓開隊は報告された閉塞を優先的に切り開く。道路啓開隊は、道路を啓開するたびにそれを救急隊と消防隊に報告することで、他のエージェントの道路状況に関するワールドモデルの精度を高める。

6 実験

上述の 3 つの協調行動、救助能力の集中、閉塞と啓開の報告、放水能力の分散について、その有効性を調べるために実験を行った。実験を行った災害空間では、震災直後に 4 棟が出火し 33%の道路が閉塞する。

表 3 協調の有無と活動結果

集中救助	閉塞報告	放水分散	死者数	火災数	鎮火時間	救出人数
しない	しない	しない	25	141	203	32
する	する	しない	25	53	167	34
する	しない	する	26	32	178	32
しない	する	する	24	21	81	29
する	する	する	23	10	41	37

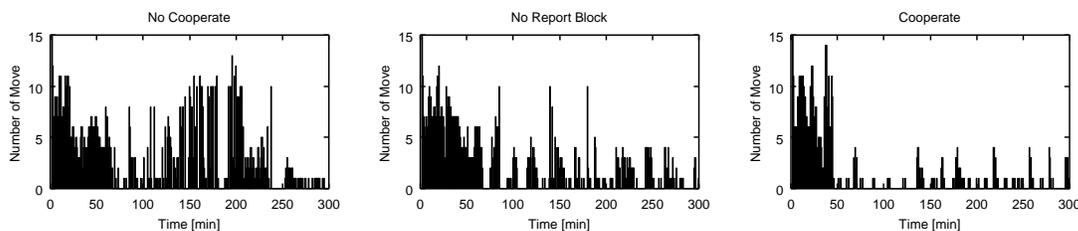


図 4 協調の有無と移動回数

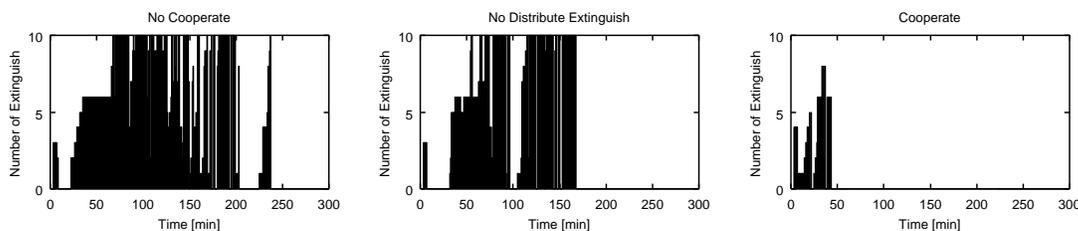


図 5 協調の有無と放水回数

実験結果を図 3 に示す。

救助能力を集中した方が、しない場合よりも多くの負傷者を救出できている。閉塞と啓開を報告した場合と、放水能力を分散した場合も、しない場合より速く鎮火が完了している。また、3つの協調行動を全て行うことで、効果的な救助活動、消火活動を実現している。

閉塞報告と放水分散の動きを詳しく見るために、協調の有無による時間毎の移動回数(図4)と放水回数(図5)を示す。

図4は、左から順に、協調しない場合、啓開報告をしない(他は行う)場合、3つの協調を行う場合の、救急隊と消防隊の、時間毎の移動回数を示している。協調しない場合は鎮火が完了するまで、多数のエージェントがほぼ絶え間なく移動し続けている。移動の多さは、目的地に到達するまでに時間がかかっている

ことを示している。啓開報告をしない場合(中央の図)は、放水分散によって無駄な移動を抑えているため、移動するエージェントが少ない。全ての協調を行う場合、消防隊は初期段階で鎮火を完了する。それ以後の移動は、救急隊が次の負傷者の場所に向かうための周期的なものがほとんどである。

図5は、協調しない場合、放水分散しない場合、全ての協調を行う場合の時間毎の放水回数を示している。協調しない場合は、初期段階に放水できず、火災が燃え広がった頃になってようやく放水できるようになっている。放水分散しない場合も初期消火に失敗しているが、閉塞報告によって火災に近けるため100ターンあたりから全ての消防隊が放水できている。全ての協調を行う場合は、消防隊は初期消火に成功し、わずかな放水で鎮火を完了している。これは放水分散によって無駄な移動を抑えられているためである。

火災に移動する際も、他に移動するエージェントが少いため、啓開隊によるサポートを受けやすくなる。

以上により、実装した協調の有効性が確かめられた。

7 競技会での評価

RoboCupRescue プロジェクトは、大規模災害シミュレーションシステム上で活動するエージェントの救助能力を評価するために、競技会を開催している。筆者らが開発したエージェント YabAI は、これまでに開催された全ての大会、RoboCup JapanOpen (2001 年 4 月)、RoboFesta (2001 年 7 月)、RoboCup2001 in USA (2001 年 8 月) で優勝した。米国で開かれた第一回世界大会 (4ヶ国 7 チームが参加) ではイランのチーム (Arian) と接戦を繰り広げたが、それを除けば他を圧倒していたといえる。

YabAI の最大の勝因は、個々のエージェントの高い基礎行動能力にあり、本題となるべき個々のエージェントの自律や複数エージェント間の協調というテーマにあまり触れていない。この結果は、現在の大規模災害シミュレーションシステムが、エージェントの分散協調の研究以前の段階にあることを示している。

8 まとめ

様々な事象を対象とするオープンな問題を扱うために、RoboCupRescue シミュレーションプロジェクトが開発中の大規模災害シミュレーションシステム上で活動する防災救助エージェントについて研究を行った。しかし、一見複雑な同システムが実はトイ的な問題に帰着できることが分かり、オープンな問題が本質

的に抱える、問題設定と実装の難しさを改めて示す結果となった。協調について若干の知見が得られたが、今後は、問題自体、Sub-simulator 群のさらなる改善が必要である。しかしその分、問題自身がオープンであるといえる。このような問題はマルチエージェントの協調について、トイ的な問題をテストベッドとするのとはまったく別の苦勞があることを報告した。RoboCup サッカーでもこのようなオープン性はあったが RoboCupRescue ではオープン性が桁違いに大きい。しかし、より現実に即した問題であり、このようなある種、黎明期の混乱の中、このような複雑な (怪奇と付け加えてもよいくらいの) 問題設定の中でも、エージェントの協調について、我々の RoboCup サッカーの研究と同様、問題が複雑であれば協調の前にまず個々のエージェントのスキルの向上が必要であるという一般的な知見が得られたということは結論できるであろう。

参考文献

- [1] 田所, 北野編: ロボカップレスキュー: 大規模災害救助への挑戦, 共立出版 (2000)
- [2] P. Stone, T. Balch, G Kraetzschmar (Eds.): RoboCup 2000: Robot Soccer World Cup IV, lecture notes in Computer Science. Springer Verlag, Berlin, 2000.
- [3] 鈴木隆志, 栗田英明, 竹内郁雄: RoboCup シミュレーションエージェントの能力とワールドモデル, 情報処理学会研究報告, Vol. 2000 No. 27((GI-2) pp. 25-32, (2000)
- [4] 高橋友一他: 大規模災害におけるマルチエージェントシステム, MACC2000 (2000)
- [5] Peter Stone and Mnuela Veloso: Task Decomposition, Dynamic Role Assignment, and Low-Bandwidth Communication for Real-Time Strategic Teamwork, Artificial Intelligence (AIJ), 1999.

未来状態の予測を利用したマルチエージェント強化学習法

長行 康男 柴田 直樹 伊藤 実

1 まえがき

マルチエージェント系における適応行動の実現は、工学及び認知科学の観点から興味深い課題である。その中でも、学習による適応行動の自律的獲得に関する研究が、強化学習 (RL) [1] の発展を契機として近年注目を集めている。

マルチエージェント系において、個々のエージェントが実行する行動の善し悪しは、他エージェントが実行する行動に依存する。この点に注目したマルチエージェント強化学習 (MARL) 法 (マルチエージェント系における強化学習法) として、他エージェントの政策 (行動決定関数) を推定しながら学習を進行する手法が提案されている [2][3][4]。これら 3 つの MARL 法では、推定した他エージェントの政策を基に他エージェントの未来の行動を予測し、その予測行動を考慮に入れながら学習を進行する。これらの MARL 法は、シングルエージェント系を対象として提案された RL 法である Q 学習 [5] を基盤にしている。そして、学習時に他エージェントの政策 (を基に実行される行動) を考慮に入れるため、学習関数である Q 関数を、状態 s と行動 a の関数 $Q(s, a)$ から、状態 s とすべてのエージェントの行動 a^1, \dots, a^n の関数 $Q(s, a^1, \dots, a^n)$ に拡張している。しかしながら、このように Q 関数を拡張した場合、学習関数はすべてのエージェントの行動に依存するため、学習空間サ

イズはエージェント数の増加に対して指数関数的に増加する。強化学習において、学習空間サイズの指数関数的増加は『次元の呪い』と呼ばれ、学習を大幅に遅らせる原因となる。

本稿では、他エージェントの政策を推定しながら学習を進行する MARL 法で、エージェントの行動に依存しない学習関数を利用した新たな MARL 法を提案する。提案する MARL 法は、TD 学習 [6] を基盤にしたもので、学習関数として V 関数 $V(s)$ を用いる。提案する MARL 法では、他エージェントの政策推定に加えて、状態遷移確率関数の推定を行う。そして、推定した状態遷移確率関数と推定した他エージェントの政策を基に未来の状態を予測し、その予測状態における V 関数値を利用しながら強化学習を進行する。

本研究では、マルチエージェント系のモデルとして、先行研究 [2][3][4] と同様、マルコフゲームの枠組みを採用する。

本稿では、提案する MARL 法と拡張 Q 関数 $Q(s, a^1, \dots, a^n)$ を利用した MARL 法 [4] を、マルコフゲームの枠組みでモデル化した 2 体エージェント問題と 3 体エージェント問題に適用する。そして、それらの実験結果を比較することにより、提案手法を評価する。

2 マルコフゲーム

マルコフゲームは、マルチエージェント環境における行動決定問題のモデルで、組 $\langle n, S, A^1, \dots, A^n, T, R^1, \dots, R^n \rangle$ で定義される。ここで、 n は環境内に存在するエージェントの数、 S は環境状態の有限集合、 A^k ($k = 1, \dots, n$) はエージェント k

A multi-agent reinforcement learning method with the estimation of future states.

Yasuo NAGAYUKI, Naoki SHIBATA, Minoru ITO,
奈良先端科学技術大学院大学, Nara Institute of Science
and Technology

の行動の有限集合, T は環境状態の遷移確率関数, R^k はエージェント k の報酬関数である.

各離散時間ステップ $t = 0, 1, 2, \dots$ において, エージェント k ($k = 1, \dots, n$) は, 現在の状態 $s_t \in S$ を観測し, 行動 $a_t^k \in A^k$ を実行する. そして, 状態は $s_{t+1} \in S$ に遷移し, エージェント k は環境から直接報酬 r_{t+1}^k を受け取る. 状態の遷移は状態遷移確率関数 T に従う. この関数は時不変な状態遷移確率

$$T(s' | s, a^1, \dots, a^n) =$$

$$\Pr(s_{t+1} = s' | s_t = s, a_t^1 = a^1, \dots, a_t^n = a^n) \quad (1)$$

の集合で表される. ここで, $\Pr(s' | s, a^1, \dots, a^n)$ は, 状態 s でそれぞれのエージェントが行動 a^1, \dots, a^n を実行したときに状態が s' へ遷移する確率を表す. エージェント k が環境から受け取る直接報酬 r_{t+1}^k も確率的で, その期待値は報酬関数

$$R^k(s, a^1, \dots, a^n) =$$

$$E\{r_{t+1}^k = r^k | s_t = s, a_t^1 = a^1, \dots, a_t^n = a^n\} \quad (2)$$

で表される. ここで, $E\{r^k | s, a^1, \dots, a^n\}$ は, 状態 s でそれぞれのエージェントが行動 a^1, \dots, a^n を実行したときにエージェント k が受け取る直接報酬 r^k の期待値を表す.

マルコフゲームにおいて, エージェント k の目的は, 式 (3) で表される関数を最大にするような自分 (エージェント k) の政策 π^k を見つけることである.

$$V^{\pi^k}(s) = E\left\{\sum_{n=0}^{\infty} \gamma^n r_{t+n+1}^k | s_t = s, \pi^1, \dots, \pi^n\right\} \quad (3)$$

ここで, 政策 π^k は各状態において各行動を選択する確率への写像 ($\pi^k : S \times A^k \rightarrow [0, 1]$), $\gamma \in [0, 1]$ は割引率と呼ばれるパラメータである. π^1, \dots, π^n は, それぞれエージェント $1, \dots, n$ の政策を表す. 式 (3) の右辺は, それぞれのエージェントが時刻 t 以降の行動を政策 π^1, \dots, π^n に従って選択したときに, エージェント k が受け取る割引報酬の和の期待値を表す. V^{π^k} は (エージェント k の) V 関数と呼ばれる. 以下, V^{π^k} を V^k と書く.

3 マルチエージェント強化学習

本稿では, マルコフゲームにおける新たなマルチエージェント強化学習 (MARL) 法を提案する. ここで, 本

研究で取り扱うマルコフゲームは, 不完全情報ゲーム [7] を仮定し, 以下の 3 つの条件を満たすものとする.

- 両エージェントが同期して行動を実行する.
- 両エージェントがお互いの行動を観測できる.
- エージェント間のコミュニケーションは存在しない.

マルコフゲームにおいて, エージェント k の V 関数はすべてのエージェントの政策に依存している. これは, エージェント k の目的 (V 関数の最大化) が, 他エージェント (エージェント k 以外のエージェント) の政策に依存していることを意味している. この点に注目した MARL 法として, 他エージェントの政策を推定しながら学習を進行する手法がいくつか提案されている [2][3][4]. これら 3 つの MARL 法では, 推定した他エージェントの政策を基に他エージェントが未来に実行する行動を予測し, その予測行動を考慮に入れながら学習を進行する. これらの MARL 法は, シングルエージェント環境を対象として提案された RL 法である Q 学習 [5] を基盤にしている. そして, 学習時に他エージェントの政策 (を基に実行される行動) を考慮に入れるため, 学習関数である Q 関数を, 状態 s と行動 a の関数 $Q(s, a)$ から, 状態 s とすべてのエージェントの行動 a^1, \dots, a^n の関数 $Q(s, a^1, \dots, a^n)$ に拡張している. この拡張した Q 関数を用いた場合の学習空間数は,

$$|S| \times |A^1| \times \dots \times |A^n| \quad (4)$$

である. ここで, $|S|, |A^1|, \dots, |A^n|$ は, それぞれ集合 S, A^1, \dots, A^n の要素数を表す. 式 (4) は, Q 関数を $Q(s, a^1, \dots, a^n)$ と拡張することにより, 学習空間がエージェント数の増加に対して指数関数的に増加することを示している. 強化学習において, 学習空間の指数関数的増加は『次元の呪い』と呼ばれ, 学習を大幅に遅らせる原因となる.

本稿では, 他エージェントの政策推定を利用した MARL 法で, エージェントの行動に依存しない学習関数を利用した新たな MARL 法を提案する. 本稿で提案する MARL 法は, TD 学習 [6] を基盤にしたもので, 学習関数として V 関数 $V(s)$ を用いる. ここで, 学習空間の数は, エージェント数に関わらず, $|S|$ である. 提案する MARL 法では, 他エージェントの政策推定に加えて, 状態遷移確率関数 T の推定を行う. そして, 推定した状態遷移確率関数と, 推定した他エージェントの政

策を基に、自分（エージェント k ）がどの行動を実行すれば、未来にどの状態に遷移するかを式 (5) で予測する。

$$\hat{P}^k(s'|s, a^k) = \sum_{a^{o_1}} \cdots \sum_{a^{o_{n-1}}} \hat{\pi}_k^{o_1}(a^{o_1}|s) \cdots \hat{\pi}_k^{o_{n-1}}(a^{o_{n-1}}|s) \hat{T}^k(s'|s, a^1, \dots, a^n) \quad (5)$$

ここで、 $\hat{P}^k(s'|s, a^k)$ は、エージェント k の観点から、自分が状態 s で行動 a^k を実行したときに状態が s' へ遷移すると予想される確率を表す。 $\hat{T}^k(s'|s, a^1, \dots, a^n)$ は、エージェント k が推定した状態遷移確率関数で、状態 s でそれぞれのエージェントが行動 a^1, \dots, a^n を実行したときに状態が s' へ遷移すると予想される確率を表す。 $\hat{\pi}_k^{o_1}, \dots, \hat{\pi}_k^{o_{n-1}}$ は、エージェント k が推定した他エージェントの政策を表す。ここで、 o_1, \dots, o_{n-1} は、 k 以外のエージェント（他エージェント）のいずれか 1 体に対応するものとする。 $\hat{\pi}^o(a^o|s)$ という表記は、他エージェント o が状態 s で行動 a^o を実行すると予想される確率を表す。本稿で提案する MARL 法では、式 (5) の \hat{P}^k で予測した未来の環境状態 s' に対する V 関数値 $V^k(s')$ を考慮しながら行動選択を行い、強化学習を進行する。

以下で、本研究で採用する他エージェントの政策の推定法、状態遷移確率関数の推定法、そして、提案する MARL 法を示す。

3.1 状態遷移確率関数の推定法

状態遷移確率関数 T の推定には、式 (6) を利用する。

$$\hat{T}^k(s^*|s, a^1, \dots, a^n) = \frac{C(s, a^1, \dots, a^n, s^*)}{N(s, a^1, \dots, a^n)} \quad (6)$$

ここで、 $N(s, a^1, \dots, a^n)$ は、状態 s でそれぞれのエージェントが行動 a^1, \dots, a^n を実行した回数、 $C(s, a^1, \dots, a^n, s^*)$ は、状態 s でそれぞれのエージェントが行動 a^1, \dots, a^n を実行したときに状態が s^* へ遷移した回数を表す。

3.2 他エージェントの政策の推定法

他エージェントの政策の推定には、我々が以前提案した手法 [4] を採用する。以下に、その手法を示す。

時刻 t において、他エージェント o が状態 s_t で行動 a_t^o を実行したとする。そのとき、状態 $s = s_t$ で実行可能な、すべての行動 $a^o \in A^o$ に対して、式 (7) に従っ

て $\hat{\pi}_k^o$ を更新する。

$$\hat{\pi}_k^o(a^o|s) \leftarrow (1 - \theta) \hat{\pi}_k^o(a^o|s) + \begin{cases} \theta & (a^o = a_t^o) \\ 0 & (\text{otherwise}) \end{cases} \quad (7)$$

ここで、 $\theta \in [0, 1]$ は観測した行動を将来の行動予測時にどれくらい考慮するかを決定するパラメータである。式 (7) の更新則によって $\sum_{a^o \in A^o} \hat{\pi}_k^o(a^o|s) = 1$ が保たれることに注意する。本研究では、それぞれのエージェントが他エージェントの行動集合 A^o について予め知っていることを仮定する。式 (7) の推定法が強化学習エージェントの政策の推定に適していることが実験的に示されている [4]。

3.3 未来状態の予測を利用したマルチエージェント強化学習法

以下に、本稿で提案する未来状態の予測を利用した MARL 法の学習の流れを示す。

1. 現在（時刻 t とする）の状態 $s_t \in S$ において、エージェント k は、式 (8) の soft-max 関数で与えられる政策 π^k に従って行動 a^k を選択する。

$$\pi^k(a^k|s_t) = \frac{e^{J(s_t, a^k)/\tau}}{\sum_{b \in A^k} e^{J(s_t, b)/\tau}} \quad (8)$$

ここで、政策 $\pi^k(a^k|s)$ は状態 s で行動 a^k を選択する確率を表す。 τ は温度パラメータと呼ばれ、行動選択のランダムさを調整するパラメータである。 $J(s, a^k)$ は、V 関数 V^k の \hat{P}^k (式 (5)) に関する期待値で、式 (9) で与えられる。

$$J(s, a^k) = \sum_{s^*} \hat{P}^k(s^*|s, a^k) V^k(s^*) \quad (9)$$

2. エージェント k は、手続き 1 で選択した行動 a_t^k を実行する（ここで、他エージェントも同期して行動 $a_t^{o_1}, \dots, a_t^{o_{n-1}}$ を実行する。環境の状態は、状態遷移確率関数 T に従って s_t から s_{t+1} へ遷移する）。エージェント k は、他エージェントの行動 $a_t^{o_1}, \dots, a_t^{o_{n-1}}$ を観測する。また、環境から直接報酬 r_{t+1}^k を受け取る。

3. エージェント k は、関数 $\hat{\pi}_k^o$ ($o = o_1, \dots, o_{n-1}$) を式 (7) に従って更新し、関数 N 、関数 C をそれぞれ式 (10)、式 (11) に従って更新する。

$$N(s_t, a_t^1, \dots, a_t^n) \leftarrow N(s_t, a_t^1, \dots, a_t^n) + 1 \quad (10)$$

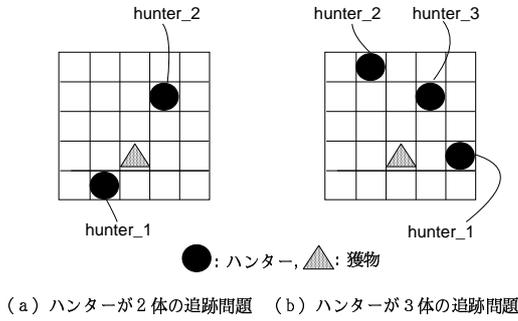


図 1 追跡問題のグリッド空間

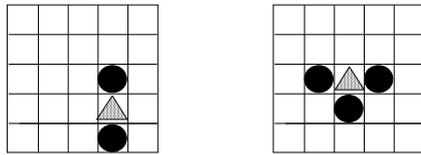


図 2 捕獲状態の例

$$C(s_t, a_t^1, \dots, a_t^n, s_{t+1}) \leftarrow C(s_t, a_t^1, \dots, a_t^n, s_{t+1}) + 1 \quad (11)$$

そして、状態 s_t における V 関数値を式 (12) に従って更新する。

$$V^k(s_t) \leftarrow (1 - \alpha)V^k(s_t) + \alpha(r_{t+1}^k + \gamma V^k(s_{t+1})) \quad (12)$$

ここで、 $\alpha \in (0, 1]$ は学習率と呼ばれるパラメータである (式 (12) の更新式は TD 学習 [6] における更新式と同じものである)。

4. 学習の終了条件を満たしていれば学習終了。そうでなければ t に 1 を加えて、手続き 1 に戻る。

4 実験

4.1 追跡問題

本研究では、実験に使用するタスクとして追跡問題 [8] を取り上げる。追跡問題は、複数のハンターが獲物を追いかけて捕獲する課題である。以下に、本研究における追跡問題の問題設定を示す。

- 2次元 (5×5) のグリッド空間中に、複数のハンターと 1 体の獲物が存在する (図 1)。ここで、グ

リッド空間の上と下、左と右の境界は繋がっているものとする。

- 本研究では、ハンターを『エージェント』と定義する。本研究では、ハンターが 2 体の場合 (図 1 (a)) と、3 体の場合 (図 1 (b)) の 2 つの実験を行う。
- 各時間ステップ毎に、ハンターと獲物は、それぞれ 1 つの行動を同期して実行する。ここで、ハンターが実行可能な行動は、隣接する上、下、左、右のグリッドへ移動する、現在位置に留まる、の 5 通りとする。また、獲物が実行可能な行動は、隣接する上、右のグリッドへ移動する、現在位置に留まる、の 3 通りとする。

- ハンターの目標は獲物を捕獲することである。ここで、捕獲の定義は、

- ハンターが 2 体の実験では、『2 体のハンターが獲物を上下、あるいは左右から挟んだ状態』 (図 2 (a))

- ハンターが 3 体の実験では、『獲物に隣接する 4 近傍のグリッドうち、いずれか 3 つのグリッドを 3 体のハンターが占有している状態』 (図 2 (b))

とする。

- 初期配置から獲物が捕獲されるまでを『1 エピソード』とする。獲物が捕獲されると、ハンターと獲物はグリッド空間中にランダムに初期配置され、新たなエピソードを開始する。
- 環境状態は、それぞれのハンターと獲物の相対位置の組合せ $s = (p^1, \dots, p^n)$ とする。ここで、 p^i ($i = 1, \dots, n$) は hunter- i と獲物の相対位置を表す。例えば、図 1 (a) では $s = ([1, 1], [-1, -2])$ 、図 1 (b) では $s = ([-2, 0], [1, 2], [-1, -2])$ である。
- 獲物は学習を行わず、3 通りの行動の中から 1 つの行動を確率的に選択する。実験では、右へ移動する確率と現在位置に留まる確率をそれぞれ $\frac{2}{5}$ 、上へ移動する確率を $\frac{1}{5}$ としている。この行動選択確率は時不変とする。

以上の問題設定では、報酬関数について言及していないが、報酬関数が式 (2) を満たすように設定された場合、この追跡問題はマルコフゲームの条件を満たす。

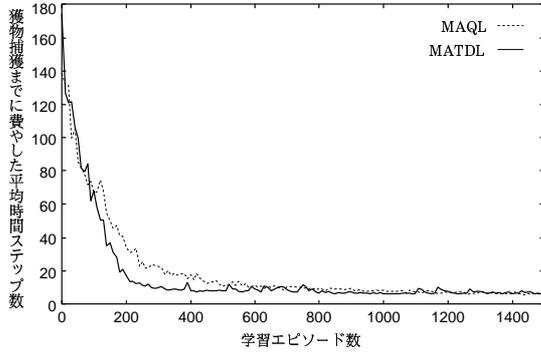


図 3 獲物捕獲までに費やした平均時間ステップ数：
ハンターが 2 体の実験。

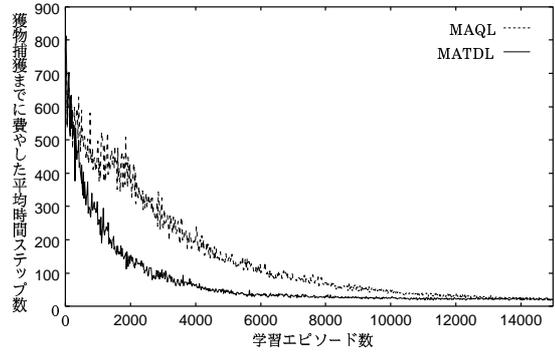


図 4 獲物捕獲までに費やした平均時間ステップ数：
ハンターが 3 体の実験。

4.2 他エージェントの政策推定を利用したマルチエージェント Q 学習法

提案した MARL 法の性能を評価するための比較対象として、本研究では、他エージェントの政策推定を利用したマルチエージェント Q 学習法 [4] (以下 MAQL と書く) を取り上げる。MAQL では、学習関数に Q 関数 $Q(s, a^1, \dots, a^n)$ を用いており、3 章で述べたように、学習空間数 ($|S| \times |A^1| \times \dots \times |A^n|$) は、エージェント数の増加に対して指数関数的に増加する。MAQL の学習の流れを以下に示す。

1. 現在 (時刻 t とする) の状態 $s_t \in S$ において、エージェント k は、式 (13) で与えられる政策 π^k に従って行動 a^k を選択する。

$$\pi^k(s_t, a^k) = \frac{e^{\bar{Q}(s_t, a^k)/\tau}}{\sum_{b \in A^k} e^{\bar{Q}(s_t, b)/\tau}} \quad (13)$$

ここで、 $\bar{Q}(s, a^k)$ は式 (14) で与えられる関数である。

$$\bar{Q}(s, a^k) = \sum_{a^{o_1}} \dots \sum_{a^{o_{n-1}}} \hat{\pi}_k^{o_1}(a^{o_1}|s) \dots \hat{\pi}_k^{o_{n-1}}(a^{o_{n-1}}|s) Q^k(s, a^1, \dots, a^n) \quad (14)$$

2. エージェント k は、手続き 1 で選択した行動 a_t^k を実行する (ここで、他エージェントも同期して行動 $a_t^{o_1}, \dots, a_t^{o_{n-1}}$ を実行する。環境の状態は、状態遷移確率関数 T に従って s_t から s_{t+1} へ遷移する)。エージェント k は、他エージェントの行動 $a_t^{o_1}, \dots, a_t^{o_{n-1}}$ を観測する。また、環境から直接報酬 r_{t+1}^k を受け取る。
3. エージェント k は、他エージェントの政策 $\hat{\pi}_k^o$

($o = o_1, \dots, o_{n-1}$) を式 (7) に従って推定 (更新) する。そして、状態 s_t 、行動 a_t^1, \dots, a_t^n における Q 関数値を式 (15) に従って更新する。

$$Q^k(s_t, a_t^1, \dots, a_t^n) \leftarrow (1 - \alpha) Q^k(s_t, a_t^1, \dots, a_t^n) + \alpha (r_{t+1}^k + \gamma \max_{a^k} \bar{Q}(s_{t+1}, a^k)) \quad (15)$$

4. 学習の終了条件を満たしていれば学習終了。そうでなければ t に 1 を加えて、手続き 1 に戻る。

4.3 実験結果

提案したマルチエージェント強化学習法 (以下、MATDL と書く) と MAQL を追跡問題に適用した。ハンターが 2 体の場合の実験結果を図 3 に、3 体の場合の実験結果を図 4 に示す。図の横軸は学習エピソード数、縦軸は 1 エピソード中で獲物捕獲までに費やした平均時間ステップ数を表す。図の結果は、10 学習エピソード毎に、そのときまでの学習性能を評価するため、初期配置を変えた 100 評価エピソード (このエピソードでは学習を行わない) の実験を行ない、その平均時間ステップ数を示したものである。2 つの学習法 (MATDL と MAQL) で使用した学習パラメータは $\alpha = 0.3 \times \text{decay}^{\text{num_ep}}$ 、 $\gamma = 0.9$ 、 $\tau = 0.1 \times \text{decay}^{\text{num_ep}}$ 、 $\theta = 0.5 \times \text{decay}^{\text{num_ep}}$ である。ここで、 decay は減衰係数、 num_ep は学習エピソード数を表す。減衰係数 decay は、ハンター数が 2 体の実験では $\text{decay} = 0.9977$ 、3 体の実験では $\text{decay} = 0.99977$ としている。減衰係数 0.9977、0.99977 は、それぞれ $0.9977^{1000} \approx 0.1$ 、 $0.99977^{10000} \approx 0.1$ となるように選ばれた値である。エージェントが環境から受け

取る直接報酬 r^k は、獲物捕獲時に $r^k = 1.0$ 、それ以外の時に $r^k = -0.05$ としている。すべての $s \in S$, $s^* \in S$, $a^1 \in A^1, \dots, a^n \in A^n$ に対して、V 関数、Q 関数、関数 N 、関数 C のそれぞれの初期値は、 $V^k(s) = 0.0$, $Q^k(s, a^1, \dots, a^n) = 0.0$, $N(s, a^1, \dots, a^n) = 0$, $C(s, a^1, \dots, a^n, s^*) = 0$ としている。また、すべての $s \in S$, $a^o \in A^o$ ($o = o_1, \dots, o_{n-1}$) に対して、関数 π_k^o の初期値は $\pi_k^o(s, a^o) = 0.2$ としている。

図 3, 図 4 の結果は、ハンター数が 2 体の場合、3 体の場合の両方で、MATDL が MAQL よりも学習が速いことを示している。また、獲物捕獲までに費やした平均時間ステップ数の差 (MATDL-MAQL) と比率 (MATDL:MAQL) は、ハンター数が 3 体の場合の方が、2 体の場合より大きいことを示している。

4.4 考察と今後の課題

2つの学習法 MATDL と MAQL の主な相違点は以下の3つである。

1. 学習関数として、MATDL では V 関数 $V(s)$ 、MAQL では Q 関数 $Q(s, a^1, \dots, a^n)$ を用いている。
2. MAQL の学習空間は、ハンター数が 2 体の場合、MATDL の 25 (5^2) 倍、3 体の場合、MATDL の 125 (5^3) 倍である。
3. MATDL では状態遷移確率関数 T を式 (6) より陽に推定しているのに対して、MAQL では学習 (試行錯誤の経験) を通じて、Q 関数中で陰に推定される。

MATDL が MAQL よりも学習が高速である原因は、これら 3 つの違いが影響していると考えられるが、獲物捕獲までに費やした平均時間ステップ数の比率 (MATDL:MAQL) が、ハンター数が 2 体の場合より 3 体の場合の方が大きいことより、項目 2 の学習空間の増加が大きく寄与していることが予想される。より詳細な調査は今後の課題である。

5 あとがき

本研究では、未来状態の予測を利用したマルチエージェント強化学習法を提案した。提案したマルチエ

ージェント強化学習法では、他エージェントの政策と、環境の状態遷移確率関数を推定し、推定したそれらの関数を利用して、どの行動を実行すればどの環境状態に遷移するかを予測した。そして、その予測した環境状態における V 関数値を基にどの行動を実行すればよいかを決定し、強化学習を進行した。提案したマルチエージェント強化学習法では学習関数として V 関数を採用した。提案したマルチエージェント強化学習法をマルコフゲームの枠組みでモデル化した追跡問題に適用し、実験を行った結果、ハンターが 2 体の場合、3 体の場合の両方の実験で、他エージェントの政策推定を利用したマルチエージェント Q 学習法 [4] より学習が高速になった。また、その学習の高速化は、2 体エージェント問題よりも、3 体エージェント問題の方がより顕著に現れた。学習が高速になった原因の詳細な調査は今後の課題である。

参考文献

- [1] R. S. Sutton, and A. G. Barto, Reinforcement Learning: An Introduction, MIT Press, Cambridge, Massachusetts, 1998.
- [2] M. L. Littman, "Markov games as framework for multi-agent reinforcement learning," Proc. 11th International Conference on Machine Learning, pp.157-163, New Brunswick, New Jersey, USA, July 1994.
- [3] J. Hu, and M. P. Wellman, "Multiagent reinforcement learning: theoretical framework and an algorithm," Proc. 15th International Conference on Machine Learning, pp.242-250, Madison, Wisconsin USA, July 1998.
- [4] Y. Nagayuki, S. Ishii, and K. Doya, "Multi-agent reinforcement learning: An approach based on the other agent's internal model," Proc. 4th International Conference on Multi-Agent Systems, pp.215-221, Boston, Massachusetts, USA, July 2000.
- [5] C. J. C. H. Watkins, and P. Dayan, "Technical Note Q-Learning," Machine Learning, vol.8, no.3, pp.279-292, 1992.
- [6] R. S. Sutton, "Learning to predict by the methods of temporal differences," Machine Learning, vol.3, pp.9-44, 1988.
- [7] G. Owen, Game Theory: Third edition, Academic Press, San Diego, California, 1995.
- [8] M. Benda, V. Jagannathan, and R. Dodihiawalla. "On optimal cooperation of knowledge sources". Technical Report BCS-G2010-28, Boeing AI Center, 1985.

環境状況の変化に応じて 自己の報酬を操作する学習エージェントの構築

森山 甲一 沼尾 正行

Abstract

The authors aim at constructing an autonomous agent learning appropriate actions in several situations of a Multi-Agent environment. Under this aim, the authors constructed an agent having abilities that can handle the agent's own rewards and distinguish the situations, and we saw that the agent can act well in each of the situations of the environment composed of homogeneous agents. In this paper, on the other hand, we check whether the agent can also act well in the environment in which the situations switch over.

1 はじめに

自律エージェントの学習とは可能な選択肢の選択基準を獲得することであり、通常は何らかの外部評価に基づいてその評価値を最大化する選択肢を選ぶようにする。しかし、例えば複数の評価基準が存在してそれらの並立が難しい環境ではどの評価基準に従うかの選択の問題があり、またエージェントが複数存在するマルチエージェント環境では社会的ジレンマの問題がある。

本研究では、マルチエージェント環境における社会

的ジレンマの問題に着目する。社会的ジレンマの下では、個々が利益を追求すればするほど（合理的に行動するほど）実際に得られる利益が減少するため、個々は非合理的に行動することが求められる。社会科学の分野では[1]、この場合に何らかの権威者（政府など）が法律や税制などを整備することによりジレンマ状況を解決する研究がなされている。しかし、例えば我々人間の場合は状況に応じて非合理的に行動できるため、法律などによる細かな規制がない場合でも社会生活を成立させることが可能である。

そこから、権威者による上位の制御がない場合でもジレンマ状況を解決できる人間のような非合理的エージェントを構築することが考えられる。しかし、一方では従来の評価を最大化する合理的行動の学習に関する研究の蓄積があり、それらを有効利用することが望ましい。そこで本研究では、自分に与えられた外部評価を操作して内部評価を生成し、それを従来の学習手法に適用することにより非合理性を実現するエージェントを構築することを目的とする。

確かに、例えば自己の内部評価が他者の利益を含むように予め設計者が設計すればエージェントはジレンマ状況を克服できるだろう。しかし、我々の人間社会のような現実のマルチエージェント環境は常にジレンマ状況であるとは限らず、むしろ動的に変化する。そのため、時と場合によっては他者への影響が無視できる状況も存在し、その中では上記のように設計されたエージェントは明らかに非効率なものとなる。また、そのようなエージェントは開かれたマルチエージェント環境に起こり得る状況の変化には対応できない。つ

Construction of a Learning Agent Handling Its Rewards According to Changes of Environmental Situations.

Koichi Moriyama, 東京工業大学 情報理工学研究所 計算工学専攻, Department of Computer Science, Graduate School of Information Science and Engineering, Tokyo Institute of Technology. koichi@nm.cs.titech.ac.jp

Masayuki Numao, 同上, ditto. numao@cs.titech.ac.jp

まり、現実の開かれたマルチエージェント環境で適切に行動するエージェントを構築する場合には、予め設計者が設計するのではなく、エージェント自身がその環境中でジレンマ状況を克服することが求められる。

これらの目的の下で筆者らは、外部評価から内部評価を生成する能力に加えて環境の状況を識別する能力をエージェントに付与し、状況に応じて自己の内部評価を生成するエージェントを構築した [4]。そしてそのエージェントを集めた同質なマルチエージェント環境において適切にジレンマ状況が克服できる一方、非ジレンマ状況においても比較的良好な結果が得られることを示した [4]。しかし、そこでは異なる状況それぞれについて実験を行っており、実験途中で状況が変化する場合を扱っていない。そこで本稿では途中で状況が変化するような実験を行ない、文献 [4] で提案した手法について検証する。

本稿は以下のような構成になっている。2 節で本研究で対象とするマルチエージェント環境を分類し、3 節では文献 [4] で提案したエージェントの構築手法について述べる。4 節では共有地の悲劇 [2] [8] をモデル化したゲーム [3] と筆者オリジナルの狭路問題という 2 種類の問題について、途中で状況が変化する実験を行なった結果を述べる。5 節でその実験結果について考察を行ない、6 節でまとめと今後の課題を示す。

2 マルチエージェント環境

本研究では個々のエージェントとエージェント社会全体の関係に着目して、環境を次の 3 つの状況に分類する [4]。

- 非干渉状況: 全てのエージェントが利益を追求する時に社会全体が最良
- 泥沼状況: 全てのエージェントが利益を追求しない時に社会全体が最良
- 競合状況: 一部のエージェントが利益を追求し、残りが利益を追求しない時に社会全体が最良

それぞれの状況で望ましい行動を 2 人ゲームの表の形で表 1 に示す。

非干渉状況は個々が単に利益を追求すれば良いため非ジレンマ状況に当たる。泥沼状況は個々が自分の利

表 1 各状況における望ましい行動組み合わせ

(a) 非干渉状況		
	利益追求	利益不追求
利益追求	○	
利益不追求		

(b) 泥沼状況		
	利益追求	利益不追求
利益追求		
利益不追求		○

(c) 競合状況		
	利益追求	利益不追求
利益追求		○
利益不追求	○	

益をより良くしようと合理的に行動するほど全体が悪化し個々の利益も減少する状況であり、囚人のジレンマ問題はこれに当てはまる。競合状況は多くの個体が同時に合理的に行動してもそれらの要求を満たせないタイプの状況であり、狭い空間などの排他的な資源によって競合が発生する状況に相当する。

我々の人間社会のような現実の開かれたマルチエージェント環境は、これらの状況が複雑に組み合わさっていると考えられる。従って、そのような環境で行動するエージェントには、上記のそれぞれの状況で適切な学習を行なう能力とともに状況を識別する能力が必要となるだろう。

3 学習エージェントの構築

本研究ではエージェントの学習手法として、明確な教師を必要としない強化学習 [6] を用いる。強化学習は報酬というただ 1 つのスカラー強化信号を用いて、それを可能な限り大きくする行動を選択させようとする学習手法である。以下では強化学習のうちの代表的手法である Q-learning [7] を用いる。これは状態 s と行動 a の組み合わせに伴う価値 (報酬の見込み値) $Q(s, a)$ (Q 値) を学習サイクルごとに以下の式により更新し、行動選択の際にこの値を用いるものである。 r_{t+1} は今回の行動提示で得られた報酬を、

α, γ はそれぞれ学習率と割引率を表す^{†1}。

$$\begin{aligned} Q_t(s, a) &= Q_{t-1}(s, a) \quad \text{if } s \neq s_t \text{ or } a \neq a_t, \\ Q_t(s_t, a_t) &= (1 - \alpha)Q_{t-1}(s_t, a_t) \\ &\quad + \alpha(r_{t+1} + \gamma \max_b Q_{t-1}(s_{t+1}, b)). \end{aligned} \quad (1)$$

しかし、これは報酬 r_{t+1} を最大化する合理的行動を選択するように学習する手法である。そのため、マルチエージェント環境、特に泥沼状況や競合状況では各エージェントはどのように学習を行なえば良いだろうか。

その 1 つの解として筆者らは学習の際に報酬値を調整することを提案した [4]。そこで用いた手法は、エージェント A_i の時刻 t の報酬 $r_{i,t+1}$ にパラメータ $\lambda_{i,t+1}$ を加えたものを A_i の学習用報酬 $r'_{i,t+1}$ とし、これを (1) 式の r_{t+1} の代わりに用いるものである。なお、以下では表記の繁雑さを避けるために自分 (A_i) を表す添字 i を省略する。

$$r'_{t+1} = r_{t+1} + \lambda_{t+1}. \quad (2)$$

本研究ではまず泥沼状況に有効な λ_{t+1} として

$$\lambda_{t+1} = \sum_{A_k \in N_i \setminus A_i} r_{k,t+1} \quad (3)$$

を用いる。この式で $N_i \setminus A_i$ はエージェント A_i の近隣のエージェントからなる集合 N_i のうち自分自身を除いたものを表している。(2) 式と (3) 式を合わせると、これは自分及び近隣他者の現在の報酬の和を学習用報酬として用いることを意味し、自分が合理的に行動することにより周囲に悪影響が現れる環境において有効である。

また、競合状況に有効な λ_{t+1} として

$$\lambda_{t+1} = r_{t+1} - r_t \quad (4)$$

を用いる。これは報酬の時間変化を考慮したもので、前回の行動による報酬と今回のその差を強調するものである。

これらの λ_{t+1} はそれぞれ対応する環境については有効であるが、異なる環境ではあまり芳しくない。例えば (3) 式を用いた場合、非干渉状況では通常の強化学習のみを用いた場合よりも結果が悪い。また (4) 式を用いた場合には、泥沼状況では文字通り泥沼に

はまってしまふ。ところが非干渉状況でこの式を用いるとエージェントの合理性をより高める結果となり、良好な結果が得られる^{†2}。これらの点から、現実の開かれたマルチエージェント環境で行動するエージェントは λ_{t+1} を適切に選択しなければならない。すなわち、環境が泥沼状況の際には (3) 式を、非干渉状況・競合状況の際には (4) 式を自動的に選択するような能力をエージェントは保持しなくてはならない。

そこで筆者らはこの両者を状況によって自動的に選択する手法を考案した [4]。それには、個々のエージェントがどのようにして状況を識別するかということが重要になるわけだが、提案手法では以下 2 つの状況識別仮定を置き、少なくとも 1 つが満たされる場合に泥沼状況と認識するようにした。

$$Q_{t-1}(s_t, a) < 0 \quad \text{for all } a. \quad (5)$$

$$r_{t+1} < Q_{t-1}(s_t, a_t) - \gamma \max_b Q_{t-1}(s_{t+1}, b). \quad (6)$$

(5) 式は現在どのような行動を実行しても今後正の報酬が見込めないことを意味している。(6) 式は (1) 式から考案したものであり、左辺は実際に獲得した報酬を表している。一方で通常の Q-learning の場合、Q 値が収束した際にはこの両辺が一致することから、右辺が現在までに学習した報酬の予測値を意味していると考えられる。すなわち、(6) 式は実際に得られた報酬が予測値よりも小さいことを表しており、これは報酬の予測値を学習した時点よりも現在の環境が悪化していることを意味している。

ところが、本研究では (2) 式により調整した学習用報酬 r'_{t+1} を Q 値の学習に用いているため、(6) 式の左辺は学習用報酬 r'_{t+1} でなくてはならない。しかし、これは r'_{t+1} を計算するためにどちらの λ_{t+1} を選択するかを決めるための式であるので、 r'_{t+1} をここで用いることは出来ない。そのため実際の獲得報酬 r_{t+1} で代用していることに注意を要する。

筆者らは、以上の仮定を満たす場合に (3) 式を、満たさない場合に (4) 式を用いて報酬を操作することを提案している [4]。

^{†1} 以下、時刻の表記は Sutton らの流儀に依る [6]。ここでは時刻 t の状態 s_t で行動 a_t を実行すると、状態 s_{t+1} に遷移し報酬 r_{t+1} が与えられるとする。

^{†2} 非干渉状況では合理的行動が望ましいことに注意。

4 実験

本研究では実験環境として、共有地の悲劇 (n 人囚人のジレンマ) [2][8] をモデル化したゲーム [3] と筆者オリジナルの狭路問題を用いる。いずれの実験も同質なエージェント 10 台で行なう。実験に用いるエージェントは、行動選択をランダムに行なうランダムエージェント、通常の強化学習を行なう通常エージェント、(3) 式を用いて報酬を操作する近隣報酬合計エージェント、(4) 式を用いる報酬差分エージェント、更に (3) 式と (4) 式を (5)(6) 式を用いて自動的に選択する自動選択エージェントの 5 種である。学習率 α ・割引率 γ はともに 0.5 とする。行動選択には Q 値から求めた温度 $T = 1$ のボルツマン分布

$$p_t(a) = \frac{\exp(Q_{t-1}(s_t, a)/T)}{\sum_b \exp(Q_{t-1}(s_t, b)/T)} \quad (7)$$

による確率 $p_t(a)$ を用いる [6]。

4.1 共有地の悲劇ゲーム

これは共有地の悲劇 (n 人囚人のジレンマ) [2][8] をモデル化したゲーム [3] である。以下ではこのゲーム環境において泥沼状況と非干渉状況を作り出し、エージェントの集合がどのように振る舞うかを確認する。

各エージェントは利己的・協力的・利他的の 3 種の行動のいずれかを一齐に提示し、エージェント社会全体における行動の組み合わせにより決められる報酬を獲得する。この一連の流れを 1 サイクルと称する。エージェント A_i の行動に対して社会全体に対するコスト c_i が付随する。ここでは利己的・協力的・利他的な行動についてそれぞれ $+\Delta r_c, 0, -\Delta r_c$ の各値をとるものとする。泥沼状況においては $\Delta r_c = 1$ 、非干渉状況においては $\Delta r_c = 0$ とした。一方、エージェント A_i の報酬はそれぞれ $3 - r_c, 1 - r_c, -3 - r_c$ とする。ここで r_c は共有コストと称し、 $r_c \triangleq \sum_i c_i$ と定義する。この場合に、泥沼状況では全てのエージェントが利他的に行動した場合に全体の報酬合計が最大 70 となり、全てが利己的に行動した場合に -70 と最小になる。一方で非干渉状況では $r_c \equiv 0$ となり、各エージェントの獲得する報酬は他のエージェントに依

存しないため、全てのエージェントが利己的に行動した場合に報酬合計は最大 30 となる。この環境で重要な点は、個々のエージェントの立場から見ると、泥沼状況においても他者の行動に関係なく利己的行動が常に最大の報酬をもたらすことである。

ここで一例としてゴミの分別を考えてみよう。焼却の際のダイオキシンの発生を防ぐために、可燃ゴミと不燃ゴミを分別するように求められているとする。ダイオキシンは塩素化合物を含むゴミを約 400°C で燃やすことにより生じ、発ガン性が疑われている [5]。この場合に上記の利己的・協力的・利他的の 3 種の行動はそれぞれ「分別しない」「自分のゴミは分別する」「集積場にある他者のゴミも分別する」となる。もし分別しなかった場合には、自治体はダイオキシンの発生を防ぐために税金を用いて人を雇い、ゴミの分別をさせることになるだろう。この場合に使われる税金が上記の共有コストに相当する。一方では、たとえ利他的に集積場の他者のゴミまで分別するとしても、その人が自治体に支払う税金が減るわけではない。そうするとゴミの分別は面倒なので、分別をしないという利己的な行動が最も好ましいと思う人が生じても不思議ではない。しかし、もし住民全員が利己的な行動を採ったらどうなるだろう。これは増税をもたらすかもしれないため、泥沼状況となるだろう。

しかし、ダイオキシンは約 400°C で燃やすことにより生じるため、これよりも高い温度でゴミを焼却するようになれば発生を防ぐことが可能である。また、ダイオキシンを除去する装置も世の中に存在する [5]。従って、高温でゴミを焼却するダイオキシン除去装置付きの焼却炉を自治体が建設した場合には、それ以後ダイオキシンの観点からは住民は可燃ゴミと不燃ゴミを分別する必要がなくなる。そうすると分別をしないことによる共有コストの増加は 0 となり、非干渉状況に移行したと考えられるだろう。

(3) 式に現れる近隣エージェント集合 N_i は

$$N_i \triangleq \{A_k \mid k = (i+j) \bmod 10, j = 0, 1, 2, 3\} \quad (8)$$

で定義する [3]。更に、各エージェントの保持する状態は近隣エージェントの行動の組み合わせとする。

最初は泥沼状況だが 30000 サイクル後に突然非干渉状況へ変化した場合の実験結果を図 1 に示す。こ

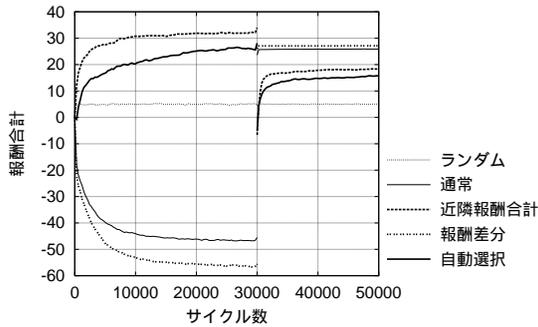


図 1 共有地の悲劇ゲームの実験結果

のグラフは 100 回の実験の平均をベジエ曲線で近似したものであり、状況が変化する 30000 サイクルでグラフを分割している。縦軸が 10 台のエージェントの報酬合計、横軸がサイクル数である。30000 サイクルまでを見てみると「通常」は時間とともに報酬和が減少している。これは個々が合理的行動をするほど社会が悪化して個々の利益も減少する泥沼状況が生じていることを意味する。また、「近隣報酬合計」が最良の結果をもたらす一方で、「報酬差分」は「通常」よりも悪い結果となっている。「自動選択」は「近隣報酬合計」よりは遅れるが上昇している。

また、最初から非干渉状況である場合の実験結果を図 2 に示す。グラフによると「報酬差分」が最良の結果をもたらしており、それに続くのが「通常」となっている。一方では、「近隣報酬合計」が途中で上昇が止まってしまっており、「通常」より悪くなっている。「自動選択」は「近隣報酬合計」と「報酬差分」のほぼ中間の結果となっている。

ここで、図 1 の 30000 サイクル以後と 図 2 を比較すると、状況が途中で変化する場合の結果が初めから非干渉状況であるものよりも悪くなっていることが分かる。これは「ランダム」を除く全てのエージェント集合に言えることであるが、(2) 式により学習用の報酬を調整する 3 種のエージェントのうち「自動選択」の悪化が大きいことが見て取れる。

4.2 狭路問題

以下では筆者オリジナルの狭路問題を用いて競合状況と非干渉状況を作り出し、エージェント集合がど

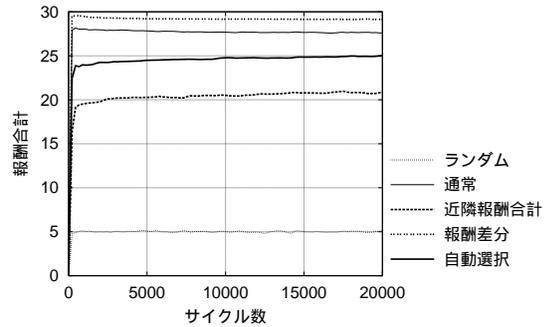


図 2 共有地の悲劇ゲーム (非干渉状況) の実験結果

のように振る舞うかを確かめる。

狭路問題では、路上駐車のため道路に車がすれ違わない狭い部分があり、狭い部分の両端に同時に 2 台の車が現れたと仮定する (図 3 (a))。問題はこの狭い部分を 2 台の車がいかにして早く通過するかというものである。車 (エージェント) の取り得る選択肢として「進む」と「待つ」の 2 つがあり、各エージェントにとって最良の結果は相手を待たせて自分が先に通過することである。しかし両者が「進む」を選択すると、狭い部分ですれ違ってしまうので結局両者とも通過できない。一方で両者が共に「待つ」を選択すると何も状況は変化しない。そのため、この場合にはエージェントの一部が利益を追求 (「進む」を選択) し、残りが利益を追求しない (「待つ」を選択) ことが求められ、競合状況となる。

もし、ある時点でこの路上の車が両方とも移動してしまったとすると (図 3 (b))、もはや 2 台の車は相手の車を意識することなくこの場所を通過すること

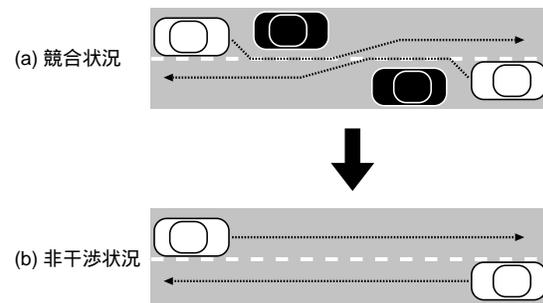


図 3 狭路問題

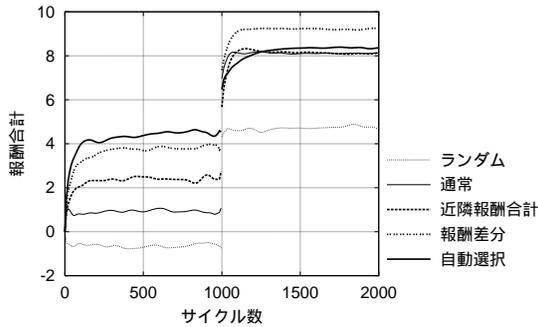


図 4 狭路問題の実験結果

が出来るようになる。従って、この場合には非干渉状況となる。

実験では 10 台のエージェントをランダムに 2 台ずつに分け、各組ごとに両者が行動を提示する。競合状況では各組で 1 台のみが「進む」となった場合にその「進む」を提示したエージェントは通過成功となる。一方で「待つ」を提示したエージェントは(相手なしで)再び行動を提示する。また、両者の行動が一致した場合には両者とも再び行動提示を行なう。毎回の行動提示の際に両者には負の報酬が与えられ、「進む」を提示して通過に成功したエージェントには正の報酬が与えられる。非干渉状況では「進む」を提示したエージェントは無条件で通過に成功し正の報酬が与えられ、「待つ」を提示したエージェントには負の報酬が与えられて再び行動提示を行なうことになる。どちらの状況においても $2^2 = 4$ 回の行動提示が行なわれた後にはエージェントが残っていてもその組は終了とする。その場合には正の報酬は与えられない。このようにして全ての組が終了したならばそれを 1 サイクルと称し、再び最初から繰り返す。

エージェントに与えられる報酬は、「進む」を提示して通過に成功したエージェントに +1、それ以外には行動提示ごとに -0.5 とした。つまり報酬合計の理論上の最大値は競合状況で $(1 + 0.5) \times 5 = 7.5$ 、非干渉状況で $(1 + 1) \times 5 = 10$ となる。各エージェントは行動提示ごとに学習を行なう。

近隣エージェント集合 N_i は組分け後の自分と相手とし、もし相手が先に通過に成功した場合には自分のみとする。各エージェントの状態は、自分の組の残り

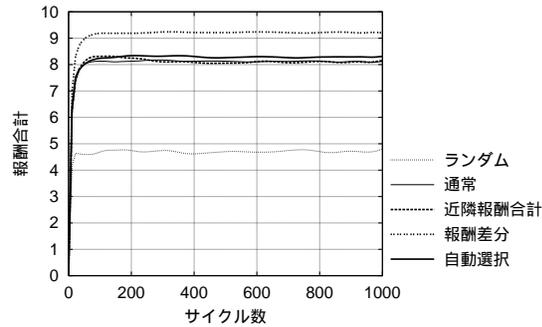


図 5 狭路問題 (非干渉状況) の実験結果

エージェント数と相手 ID の組み合わせとした。

最初は競合状況であったが 1000 サイクル後に非干渉状況となった場合の実験結果を図 4 に示す。描画方法は図 1 と同様であり、状況が変化する 1000 サイクルでグラフを分割している。1000 サイクルまでのグラフによると「自動選択」が最良の結果をもたらしており、それに続くのが「報酬差分」となっている。このように、「自動選択」がその元となっている「報酬差分」よりも良い結果となった原因の究明は今後の課題である。「近隣報酬合計」も「通常」よりは良いのだが「報酬差分」より悪いことが分かる。

また共有地の悲劇ゲームと同様に、初めから路上駐車のない非干渉状況における実験結果を図 5 に示し、これを図 4 の 1000 サイクル以後と比較する。この場合にはどちらも結果に差はなく、「報酬差分」が最良で「ランダム」を除く他のエージェント集合はほぼ同様の結果となっている。

5 考察

本稿では文献 [4] で筆者らが提案した手法について、途中で状況が変化する環境で実験を行なった。この実験結果について以下で考察を行なう。

まず 4.1 節の共有地の悲劇ゲームでは、「ランダム」を除く全てのエージェントについて、環境が泥沼状況から非干渉状況に移行した後(図 1 の 30000 サイクル以後)の結果が初めから非干渉状況であった実験の結果(図 2)より明らかに悪くなっている。これは状況変化前に学習した結果の影響が出ているためと考えられる。

また、報酬を調整した 3 種のうち「自動選択」の結果の悪化が最も大きく、「近隣報酬合計」よりも劣るものとなっている。この原因としては、状況識別仮定のうちの (6) 式の左辺に r'_{t+1} の代わりに r_{t+1} を用いたことが影響していると考えられる。このゲームにおける泥沼状況では獲得報酬 r_{t+1} が最小となる利他的行動が望まれるため、必然的に r'_{t+1} に対する λ_{t+1} の影響が大きくなり、 r'_{t+1} と r_{t+1} が全く異なる値になると思われる。すると学習の結果である Q 値も r_{t+1} から得られる値とは異なるものとなり、結果として (6) 式の判定に誤りが生じるのではないだろうか。またより本質的な問題として、(6) 式では学習が収束した状態を想定しているが、実験開始直後ではそれは望めないという点と、そもそも環境の非マルコフ性により Q -learning では収束性すら保証されていないという点が問題となっているのかもしれない。

一方、4.2 節の狭路問題の場合には、競合状況から非干渉状況に移行した後 (図 4 の 1000 サイクル以後) の結果は初めから非干渉状況であった実験の結果 (図 5) とほぼ同じである。すなわち、この場合には状況変化前の学習の悪影響が見られない。これは (4) 式を用いれば状況変化前後どちらにも対応できるというように競合状況と非干渉状況は性質が似ているためと考えられる。しかし、非干渉状況において「自動選択」が「近隣報酬合計」とほぼ同じ結果になったということは、提案手法による状況の識別がうまく機能していないことを意味している。

もし、エージェントの知覚能力がもっと豊かであったらどうであろうか。共有地の悲劇ゲームの例では、自治体が新焼却炉を建設した場合に通常は住民に対して何らかのアナウンスがあるだろう。また狭路問題の例では、運転者は路上駐車がなくなったことを視覚から認識できるはずである。これらのようにエージェントが豊かな知覚能力により状況変化を認識できるならば、エージェントは状況が変化する 1 つの環境を状況が不変な複数の環境に分割することが可能となる。この場合には、それぞれの環境中では文献 [4] における実験と同等となり、過去の学習の悪影響を受けずに学習することが可能である。あるいは (5)(6) 式を用いて状況を推測しなくても、初めから (3) 式か (4) 式

のどちらを利用すれば良いかが自明となるかもしれない。しかし、この議論は対象とする問題をエージェントの知覚の問題に転嫁しただけであり、状況変化が知覚できない場合については何も述べていないことに注意が必要である^{†3}。本研究はこの点を踏まえて、与えられる外部評価を基に限られた知覚によって学習を行ない、環境に適応するエージェントを構築することを目的としている。

6 まとめ

本稿では筆者らが提案した手法 [4] について新たにを行った実験の結果を示した。文献 [4] ではマルチエージェント環境の 3 種それぞれの状況が不変な場合の実験のみを行っていたが、本稿では途中で状況が変化する場合の実験を行った。その結果、共有地の悲劇ゲームにおける泥沼状況から非干渉状況への移行の際には、提案手法は状況の変化に対応する能力が乏しいことが判明した。従って提案手法は、いろいろな状況が組み合わさっている現実の開かれたマルチエージェント環境に対応するためのエージェントを構築するという観点からは不合格であるといえる。しかし、考察で言及した通り、(6) 式の左辺を r_{t+1} で代用した点を改良することによって、少なくとも報酬を調整する個々の手法 ((3)(4) 式) と同等な適応力を得られるのではないかとと思われる。これは行動選択用の Q 値と状況識別用の Q 値を分けて考えることで可能になると思われるが、単純に考えても空間使用量が倍になる点がデメリットである。一方で狭路問題の場合には、非干渉状況において提案手法による状況の識別がうまく機能しないことが判明した。この原因を究明してより良い状況識別法を構築することも必要となるだろう。

それ以外の今後の課題としては、状況識別の自動化の追求、異質なエージェントからなるマルチエージェント環境における提案手法の有効性の確認、(3) 式における近隣エージェントからの情報の不要化と異なる強化学習法の適用が考えられる [4]。

^{†3} これは、資源の枯渇の問題などの状況変化が徐々に起こる問題について顕著であると思われる。

References

- [1] 出口. 複雑系としての経済学. 社会科学のフロンティア 6. 日科技連, 東京, 2000.
- [2] G. Hardin. The Tragedy of the Commons. *Science*, 162:1243–1248, 1968.
- [3] S. Mikami and Y. Kakazu. Co-operation of Multiple Agents Through Filtering Payoff. In *Proc. 1st European Workshop on Reinforcement Learning, EWRL-1*, pp. 97–107, Brussels, Belgium, 1994.
- [4] 森山, 沼尾. 自己の報酬を操作する学習エージェントの構築. 人工知能学会 第 45 回人工知能基礎論研究会資料 (SIG-FAI-A101), pp. 15–20, 東京都, 2001.
- [5] T. G. Spiro and W. M. Stigliani. *Chemistry of the Environment*. Prentice-Hall, Englewood Cliffs, NJ, 1996. (岩田 元彦, 竹下 英一 訳. 地球環境の化学. 学会出版センター, 東京, 2000).
- [6] R. S. Sutton and A. G. Barto. *Reinforcement Learning: An Introduction*. MIT Press, Cambridge, MA, 1998.
- [7] C. J. C. H. Watkins and P. Dayan. Technical Note: Q-learning. *Machine Learning*, 8:279–292, 1992.
- [8] X. Yao and P. J. Darwen. An Experimental Study of N-Person Iterated Prisoner's Dilemma Games. *Informatica*, 18:435–450, 1994.