社団法人 情報処理学会 研究報告
IPSJ SIG Technical Report

2007－ICS－147 (5)
2007／3／14

非常時における避難政策の学習による獲得

太田 正幸　山下 倫央　車谷 浩一

産業技術総合研究所
科学技術振興機構 CREST
〒135-0064 東京都江東区青海 2-41-6
E-mail: m-ohta@aist.go.jp tomohisa.yamashita@aist.go.jp k.kurumatani@aist.go.jp

あらまし 我々は、非常時に個々の被災者がどの非常口から脱出するべきかを、マルチエージェントシミュレータ上での学習により獲得する研究を行なっている。このようなマルチエージェント学習を行なう場合、各エージェントが個別に得た経験だけを使って学習を行なうと、しばしば非常に多くの試行回数を必要とする。これは、災害シミュレーションのように1回の試行に時間がかかるものを対象とする場合には、致命的である。この問題に対し、エージェント間で情報交換を行なうアプローチがよく見られるが、エージェントに選好がある場合には誰と交換してもよいというわけではない。そこで、我々は、このような学習の際の情報交換相手を効用関数に基づいて決定する手法を提案する。

A Method to Acquire Effective Evacuation Policy

Masayuki OHTA, Tomohisa YAMASHITA, and Koichi KURUMATANI

National Institute of Advanced Industrial Science and Technology (AIST)
CREST, Japan Science and Technology Agency (JST)
Aomi 2-41-6 Koto-ku Tokyo 135-0064 Japan
E-mail: m-ohta@aist.go.jp tomohisa.yamashita@aist.go.jp k.kurumatani@aist.go.jp

Abstract We are trying to acquire effective evacuation policies with learning on a multi-agent simulator. In such multi-agent environment, the approach that exchanges information between agents is used to promote learning. However, if the agents have some preferences in their action-selection, the information-exchange may cause some problems. To overcome this problem, we propose a method to decide with which agent to exchange information. The effectiveness of the proposed method is confirmed with simulation of evacuation from underground mall.

1. はじめに

災害時に、被災者に対して適切な避難誘導を行なうことは被害軽減のためにとても重要であり、近年、この避難誘導を支援するシステムの研究が盛んに行なわれている [4] [11] [3]。また、自己位置の同定が可能な小型の携帯端末の開発 [10] や、各種センサのためのミドルウェアの整備 [8]、さらに、それらを用いた避難誘導の研究 [2] なども行なわれており、こうしたシステムの実世界での適用も現実味を帯びて来ている。しかし、最終的に避難誘導システムを実現するためには、これらのシステムに加え、効果的な避難経路を導き出す手法が必要となる。このような効果的な戦略を自動獲得する方法として良く用いられているのが、シミュレータ上で複数のエージェントが様々な戦略を試し、効率的な政策を学習するマルチエージェントシミュレーションによる手法 [12] [6] である。このアプローチは、系全体としての性質が分かっていなくても、ミクロな視点での性格が分かっていれば全体としての変化を予測することが可能であり、且つ、比較的容易に実装できるという利点がある。しかし、このよう

な環境において、全エージェントが同時に学習を行なう場合、あるエージェントが政策を変更すると、別のエージェントの政策にも影響が出てしまうことがあり、学習が収束するまでには非常に多くの試行回数が必要になる可能性があるという問題がある。特に、災害シミュレーションのような1回の試行に比較的時間のかかる対象を扱う場合には、この問題は致命的である。この問題に対し、エージェント間で情報交換を行なうことで、少ない試行回数でも学習を可能な限り促進するという解決策がとられることも多い [9] [1] [5] が、どのエージェントと、どの情報を交換すればよいのかということに対する指針のようなものは存在しない。特に、エージェント毎に選好が異なる場合には、他のエージェントからの情報を全てそのまま取り入れることに問題があることは容易に想像がつく。例えば、ある建物からの避難政策を学習する場合、ある地点にいる人がある出口から脱出するのに5分かかったからといって、別の地点にいる人もその出口から脱出するのにかかる時間を5分と見積もることには問題がある。本稿では、我々はこのような環境においてもエージェント間で適切に情報交換を行ない、学習速度を向上させるための手法を提案する。そして、マルチエージェントシステムにより構築された避難シミュレータ上で我々が行なった避難政策の学習において、この方法を適用した実験結果を示す。

以降、第2節において本稿で扱う避難シミュレーションの概要を示し、第3節において学習速度を向上させるための手法を提案し、第4節において、提案手法の効果を確認するための実験を示す。最後に、第5節において、この手法に対する考察を行ない、第6節でまとめを行なう。

2. 避難シミュレーション

本稿では、ある地下街からの避難に関するマルチエージェントシミュレーションを対象とする。シミュレーションの動作イメージを図1に示す。このシミュレーションでは、地図はノードとリンクで表現され、エージェントはその上を移動することのみが可能である。地図上にはいくつかの非常口 $e_j \in E$ が存在し、エージェントの目的は、可能な限り短時間で、いずれかの非常口から外に脱出することである。各エージェント a_i の初期位置は決まっており、そこから外に脱出するまでにかかった時間を評価値とし、その効用の期待値を評価関数 $F_i(e_j)$ として個別に保持している。エージェントはシミュレーションの最初にそれぞれ自分が向かう非常口を基本的にはグリーディーな選択に従いつつ、ある一定確率 ϵ $(0 \leqq \epsilon < 1)$ でランダムに非常口を選択し、そこに向かって最短経路で移動する。エージェントの移動速度 v は、混雑の度合に応じた変化を再現するため、以下の式に従うものとした。

$$v = v_0 \times \left(1.0 - \frac{0.8}{1.0 + \exp(c - p)}\right)$$

(ただし、v_0 は自分しかいない場合の速度、c は現在地 (ノードまたはリンク) の基準容量、p は現在地の人口を表わす) シミュレーションは、全エージェントが脱出を終えるか、あらかじめ設定したステップ数の上限に達するまで継続し、終了した段階のステップ数を全体の政策の評価とした。

毎回シミュレーション終了後に、各エージェント a_i は、全体の脱出終了時間ではなく、個別の脱出時間に応じて次の式で効用関数 $F_i(e_j)$ を更新する。

$$F(e_j) = (1 - \alpha)F(e_j) + \alpha F(e_j)'$$

(ただし、α $(0 < \alpha < 1)$ は学習率を表わす定数、$F(e_j)'$ はシミュレーション結果として得られた、自分または他のエージェントが e_j に到着するのに要したステップ数) 他のエージェントから得た情報による更新も、上記と同じ式を用いて行なわれる。

この学習方法は、各エージェントが利己的に自分の効用を追求するものであり、得られる結果は全体として最適な解では無いかも知れない。しかし、その代わり、誰かの犠牲の元に全体の効用を最適化するというアプローチは取られていない。非常時のような場合には、全員が指示に従うとは限らないため、このようにして得られた結果は逆に説得力を持つものと考えられる。

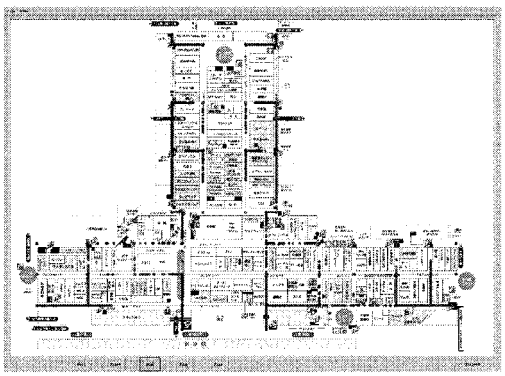

図1 シミュレーションの動作イメージ
(背景画像は八重洲地下街の Web Page より)

3. 効用関数に基づく情報交換相手の決定

前述の通り、マルチエージェント環境において、個々の経験だけをもとに学習を行なう方法は効率が悪い。しかし、個々のエージェントが選好を持つ場合にはどのエージェントからの情報でもそのまま信じて用いて良いわけではない。この問題に対し、我々は、自分と政策の近い

エージェントは選好も近いという仮定に基づき、情報交換可能なエージェントを決定する、次のような手法を提案する。まず、各エージェント $a_1, a_2, \cdots, a_n$ にはそれぞれ選択肢 $e_1, e_2, \cdots, e_m$ があり、エージェント a_i の e_j に対する効用関数を $F_i(r_j)$ と表現するものとする。このとき、政策ベクトル $V_i = \{F_i(e_1), \cdots, F_i(e_m)\}$ を考え、この政策ベクトル空間における a_i と $a_{i'}$ の距離 $D(i, i')$ を以下の式により定義する。

$$D(i,i') = \sqrt{\sum_{j=1}^{m} (F_i(e_j) - F_{i'}(e_j))^2}$$

そして、この距離が近いエージェントとだけ情報交換を行ない、それ以外のエージェントとは情報交換を行なわないという規則に基づき学習を進める。この政策ベクトルは学習過程のものを用いるため、特に学習初期においては、選好に関係なく、どのエージェントとも情報を交換する可能性がある。そういった問題に関する調整が必要になる可能性は充分にあるが、以下に示す実験では、そのような調整は行なっていない。

4. 実　　験

前述の避難シミュレーションにより、本稿で提案する政策空間での距離に基づく手法の有効性を確認する。実験では、図 1 の地下街に 1785 のエージェントを配置し、自分の経験だけに基づいて学習した場合と、それに加えて政策空間上で最も近いエージェントから入手した情報も利用して学習した場合とを比較した。この実験で用いた提案手法では、政策空間において最も近い 1 つのエージェントだけから情報を受け取る方法を取った。このとき、自分の経験に対する学習率は $\alpha = 0.05$、他のエージェントから得た情報に対する学習率は $\alpha = 0.025$、探索確率は $\epsilon = 0.1$ (いずれも固定) として行なった。

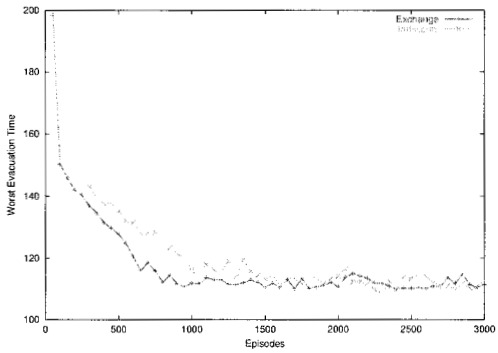

図 2　結　　果

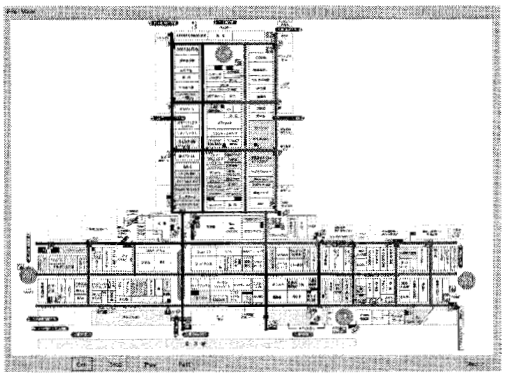

図 3　結　　果

学習過程において、全エージェントが脱出し終わるまでにかかった時間を 50 エピソード毎に平均してプロットしたものを図 2 に示す。このグラフから、どちらの手法を用いた場合でも、最終的には同程度に効果的な政策が得られる結果となっていることが見て取れる。また、学習した政策に基づいて避難の様子を再現すると、どちらの方法を用いた場合でも同様に、最後に残ったエージェントが全ての非常口から同時に脱出するような政策が得られたことを確認することができる。しかし、そのようなほぼ同様の政策を獲得するまでに要した時間は提案手法を用いることにより、大幅に短縮することに成功した。提案手法は、政策空間での距離を計算する必要があるため、1 回の試行にかかる時間は多少長くなるという欠点があり、実際に学習が収束するまでにかかる時間を短縮できるかどうかは問題に依存する。しかし、災害シミュレーションなど、1 回の試行に比較的多くの時間を必要とする問題については充分な効果を期待することが可能である。この実験において、提案手法を用いた場合の方が少ない試行回数で収束しているが、提案手法では効用関数の更新回数そのものが多い。従って、この実験だけでは、常に試行回数を減らすことが可能だという結論は得ることはできない。しかし、この方法によって選んだエージェントから受け取った情報を用いて学習を行なっても、間違った方法への学習はなされておらず、少なくとも利用すべき情報の選択には成功していると言うことができる。このことから、政策空間での距離に基づいたこの方法には充分に利用価値があるといえる。最終的に提案手法により獲得した政策を図 3 に示す。この図は、各地点からどの非常口を目指すべきかが色分けされている。この結果も通常の手法で学習を行なった場合との大きな差は見受けられなかった。

5. 考　　察

本稿で用いられた実験では、実際には、初期位置が物理的に近いエージェント間で情報交換を行なえば、問題の性格上、提案手法と同等かそれ以上の効果が期待できるものと思われる。しかし、特定の問題に対し、必ずしも適切な情報交換相手が直感的に分かるとは限らないため、どのような問題に対してでも適用可能な本提案手法には利用価値があると言える。また、実験では最も政策の近いエージェントだけから情報提供を受けたが、情報をもらう相手は複数であっても構わない。どこまで遠い相手から情報提供を受けて良いのかといった政策空間上での範囲を動的に設定する方法などが将来課題として考えられる。

提案手法に基づき、政策の近いエージェント同士が同じ情報を元に学習を続けて行くと、全員の効用関数が非常に近い集団ができ、その集団内では効用が大幅に低下する危険性がある。すなわち、これはエージェントが利己的な場合に大域情報がエージェント集団の学習に悪影響を及ぼす状況[7]と同様である。この問題を避けるため、本稿で示した実験では、自分が直接得た情報と、他のエージェントから間接的に得た情報に対し、異なる学習率を適用した。本稿で示した実験では、これらの学習率は経験的に決めた値を使用したが、将来的には、この学習率に対し、何らかの理論的根拠が示す必要がある。また、これ以外にも、情報交換の頻度，学習率を後半にかけて、減衰させる必要性など，まだ数多くの議論の余地が残されており、これらは全て将来の課題である。

6. おわりに

マルチエージェント学習を行なう際、エージェント間で情報交換を行なうことにより学習の試行回数を軽減するため、政策空間における距離に基づいて情報交換相手を決定する方法を提案した。本稿では、非常時における避難政策を学習する例にこの手法を適用し、その有用性を確認した。この方法の欠点は政策空間上で近いエージェントを探すのに多くの計算量を必要とすることだが、その分、学習の試行回数が少なくて済むため、災害シミュレーションなど、1回の試行に時間がかかる問題に対しては特に有効である。また、この方法は環境についての情報が一切無くても利用できることから、様々な問題に幅広く適用可能であるという利点も持っている。実際に適用する際の学習率の制御方法など、将来課題が数多く残されてはいるが、この手法により、少なくとも利用すべき情報の選択には成功しているおり、将来的にも様々な応用が期待できるものと思われる。

文　　献

[1] A. Garland and R. Alterman. Learning procedural knowledge to better coordinate. In *Proceedings of the Seventeenth International Joint Conference on Artificial Intelligence*, pages 1073–1079, 2001.

[2] Y. Inoue, A. Sashima, and K. Kurumatani. Indoor navigation system for emergency evacuation in ubiquitous environment. In *UbiComp2006 Adjunct Proceedings, CD-ROM*, page Posters22, 2006.

[3] Y. Nakajima, H. Shiina, S. Yamane, H. Yamaki, and T. Ishida. Disaster evacuation guide using a massively multiagent server and gps mobile phones. In *IEEE/IPSJ Symposium on Applications and the Internet (SAINT-07)*, 2007.

[4] H. Nakanishi, S. Koizumi, T. Ishida, and H. Ito. Transcendent communication: Location-based guidance for large-scale public spaces. In *Proceedings of the 2004 Conference on Human Factors in Computing Systems*, pages 655–662, 2004.

[5] L. Nunes and E. Oliveira. Learning from multiple sources. In *Proceedings of the Third International Joint Conference on Autonomous Agents and Multiagent Systems (AAMAS 2004)*, volume Vol.3, pages 1106–1113, 2004.

[6] M. Ohta, T. Koto, I. Takeuchi, T. Takahashi, and H. Kitano. Design and implementation of the kernel and agents for the robocup-rescue. In *Proceedings of The Fourth International Conference on MultiAgent Systems*, pages 423–424, 2000.

[7] M. Ohta and I. Noda. Reduction of adverse effect of global-information on selfish agents. In *Proceedings of Seventh International Workshop on Multi-Agent-Based Simulation*, pages 7–14, 2006.

[8] A. Sashima, N. Izumi, and K. Kurumatani. Location-aware middle agents in pervasive computing. In *Proceedings of the 2004 International Conference on Pervasive Computing and Communications*, pages 820–826, 2004.

[9] M. Tan. Multi-agent reinforcement learning: Independent vs. cooperative agents. In *Proceedings of the Tenth International Conference on Machine Learning*, pages 330–337, 1993.

[10] 車谷浩一, 山下倫央, 和泉憲明, 幸島明男, and 和泉潔. 愛・地球博グローバル・ハウス統合情報支援システム - CONSORTS アーキテクチャによる 情報提供・会場運営支援システム. **情報処理**, 47(2):105–108, 2006.

[11] 中西英之, 小泉智史, 石黒浩, and 石田亨. 市民参加による避難シミュレーションに向けて. **人工知能学会誌**, 18(6):643–648, 2003.

[12] 田所諭, 高橋友一, 高橋宏直, 畑山満則, 松野文俊, 太田正幸, 小藤哲彦, 竹内郁雄, 松井武史, 桑田喜隆, 兼田敏之, 渥美政保, 野邊潤, and 北野宏明. ロボカップレスキュープロジェクト (解説). **人工知能学会誌**, 15(5):798–806, 2000.