

強化学習型マルチエージェントによる交通信号制御に関する研究

メタデータ	言語: jpn 出版者: 公開日: 2017-10-05 キーワード (Ja): キーワード (En): 作成者: メールアドレス: 所属:
URL	http://hdl.handle.net/2297/16678

氏名	参 沢 匡 将
生 年 月 日	
本 籍	長野県
学 位 の 種 類	博士 (工学)
学 位 記 番 号	博甲第 669 号
学 位 授 与 の 日 付	平成 16 年 3 月 25 日
学 位 授 与 の 要 件	課程博士 (学位規則第 4 条第 1 項)
学 位 授 与 の 題 目	強化学習型マルチエージェントによる交通信号制御に関する研究
論文審査委員(主査)	木村 春彦 (工学部・教授)
論文審査委員(副査)	中山 謙二 (工学部・教授) 村本健一郎 (工学部・教授) 山根 智 (工学部・教授) 南保 英孝 (工学部・講師)

学 位 論 文 要 旨

Abstract

A multi-agent system is effective as a technique of building a large-scale and complex system. However, in case of forming an organization by learning, in order to take into consideration several agents' condition, the number of recognized states becomes huge and much time is required for learning. It is the problem which should be solved when applying to real environment.

In this paper, we propose self reinforcement learning and vicarious reinforcement learning in order to shorten learning time. Self reinforcement learning is the technique of creating a reinforcement (self reinforcement) by considering the success or failure of reinforcement using the knowledge which is known beforehand. Therefore, it is possible to learn to several rules of the same state with a executed rule. Self reinforcement learning is effective when the same recognized state was again, however it is thought that an effect is small when recognizing various states such as the first stage of learning. On the other hand, vicarious reinforcement learning is the technique of learning the rule in the similar states of a executed rule. Therefore, it is effective in the first stage of learning. We simulate this algorithm using traffic signal control problem which needs a large-scale and complicated system and show the usefulness.

1. 序論

大規模で複雑なシステムを構築する手法としてマルチエージェントシステムが挙げられる。また、学習手法として強化学習が用いられることが多い。強化学習は事前知識を必要とせず、試行錯誤によって学習することができることが特徴である。しかし、試行錯誤を繰り返さなければならぬため、学習に多くの時間が必要である。特に、マルチエージェントシステムのように学習すべき状態が膨大である場合、顕著である。これは、マルチエージェントシステムを実環境に適用する場合に大きな問題点となる。

本研究では、学習時間の短縮手法として自己強化学習、代理強化学習を提案し、大規模で複雑なシステムである交通信号制御に適用し、その有効性を検証する。

2. 交通信号制御

交通信号制御の目的は複雑な交通現象となる交差点に信号機を設置し、円滑な交通流を確保することである。各信号機はサイクル（信号の表示が青、黄、赤と一巡するのに要する時間）、スプリット（各方向に通行権が与えられている時間）、オフセット（隣接信号機間の青信号開始時間のずれ）によって制御される。また、制御地域は、単独制御（1つの交差点を単独に制御する）、系統制御（隣接信号機間を関連付けて制御する）があり、制御方式は、予め決められたパラメータで制御する固定制御と、学習などによって適応的に制御する可変制御がある。本研究では、学習によって適応するため可変制御であり、自己強化学習を用いる場合は単独制御、代理強化学習を用いる場合は系統制御である。

3. 自己強化学習による交通信号制御

従来の強化学習は環境からの強化信号によって学習するため、環境にコントロールされている。しかし、人間の場合、強化信号自体を自らコントロールすることがある。これは、強化信号を自ら操作し、自らに与える強化信号（自己強化信号）によって学習手法であり、自己強化学習と呼ばれる。強化学習に多くの時間が必要となるのは、実行したルールのみ学習しているためである。そこで、自己強化学習によって複数のルールを学習することができれば、学習時間短縮が可能であると考えられる。自己強化学習を行うためには事前知識が必要であるが、これは環境特性などの予め既知な知識や関連研究による知識を用いることが可能である。つまり、本研究で提案する自己強化学習は事前知識を用いて、環境から得られる強化信号に対してその成否を考察し、実行していないその他のルールにも強化を与えることによって学習する手法である。特に競合集合は条件部が同一であるため、自己強化学習が作成しやすいと考えられる。

交通信号制御に適用した例を示す。自己強化信号として以下の値を与える。

$$Effect = GreenTime - In \times ThroughTime$$

ここで、*GreenTime* は実行した青時間、*In* はセンサによって観測される交差点への流入台数、*ThroughTime* は車両1台当たりの交差点通過時間である。本研究では *ThroughTime* を事前知識として用いた。この *Effect* は青時間の有効度を示すものであり、*Effect* が小さければ十分な青時間が与えられていなかったことを示し、大きければ余分な青時間を与えていたことになる。そこで、*Effect* を E_{short} （青時間が短い）、 E_{good} （青時間が適切）、 E_{long} （青時間が長い）の3種類に区分する。それぞれの状況に応じて、例えば、 E_{short} であれば、実行したルールより長い青時間のルールの重みを大きくし、その他のルールを小さくするように学習する。よって、1回のルール実行において同一状態の複数のルールを学習することが可能となり、学習時間短縮が期待できる。実験結果の一例を図1に示す。本シミュレーションでは平均待ち時間によって評価を行っている。図1から、従来研究では5~10時間の学習時間が必要であったが、提案手法であるCoop Signalは学習時間短縮が行えていることがわかる。

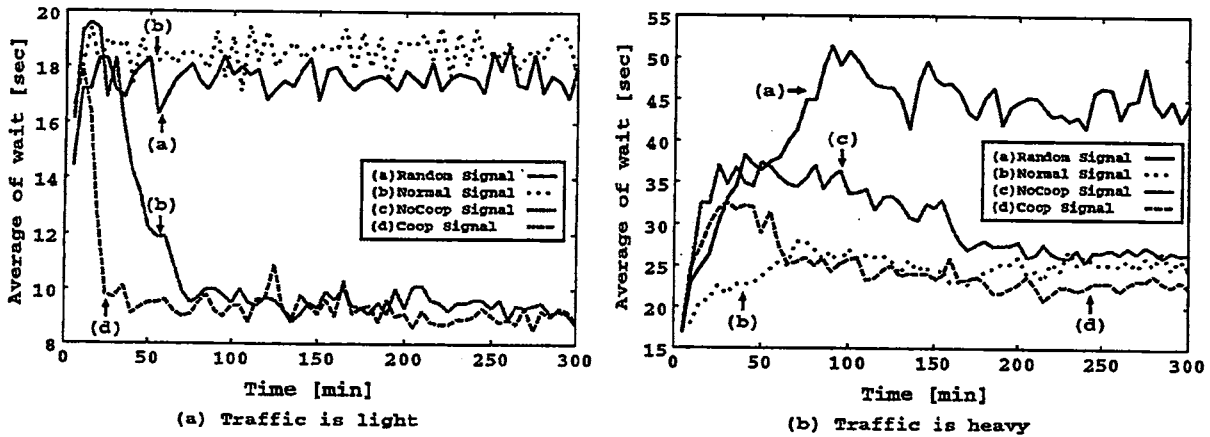


図 1: シミュレーション結果 (自己強化学習)

4. 代理強化学習による交通信号制御

自己強化学習は同一状態の複数のルールを学習することによって学習時間の短縮を行っているが、この学習では再び同一状態を認識した場合に効果があると考えられる。しかし、学習初期段階のような様々な状態を認識する場合、同一状態を認識することが少ない。つまり、状態数が膨大な場合には自己強化学習は効果が小さいと考えられる。そこで、本研究では代理強化学習を提案する。代理強化学習とはモデルとなる能力や置かれている環境が類似するエージェントが受ける強化を見聞し、自分の立場に置き換えて（代理強化を用いて）学習する手法である。つまり、代理強化学習を行うためには条件として類似する能力と環境が必要である。マルチエージェントシステムを構築する場合、環境内のエージェントの能力を同一とすることは可能であるが、環境に関しては、環境特性は学習初期段階では明白ではないなどの問題点がある。そのため、前述の定義にしたがった手法は困難であると考えられる。そこで、エージェント内の各状態を主体と考え、実行したルールの状態をモデルとし、その状態に与えられた強化を参考に、類似する状態にも強化（代理強化）を与える手法を提案する。つまり、1つのルール実行において、複数の状態のルールを学習することが可能となり、自己強化学習に比べ、学習初期段階において学習されている状態が多く存在するため、行動がランダムに選択される確率が低下し、学習時間の短縮が期待できる。

交通信号制御に適用した例を示す。代理強化学習を行う類似した状態として待ち行列が異なる場合と通行権が異なる場合の2種類に対して行うものとする。本研究では、停止回数を削減を目的とするために車両の状態を推測しているため、その仮定により代理強化学習が可能である。シミュレーション結果を図2に示す。図2では(a)自己強化学習、(b)待ち行列に対する代理強化学習、(c)通行権に対する代理強化学習、(d)待ち行列、通行権に対する代理強化学習である。図2より多くの状態に対して学習を行うことによって学習時間短縮が可能であることがわかる。また、ランダム選択率も同様の傾向があることがわかる。

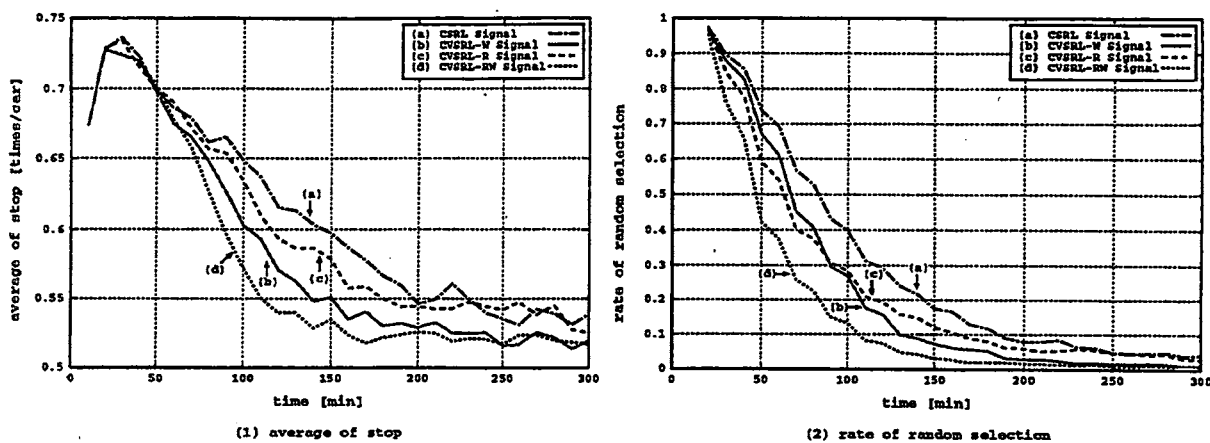


図 2: シミュレーション結果 (代理強化学習)

5. シミュレーション実験による比較評価

自己強化学習, 代理強化学習の特性を検証するために比較実験を行った. 結果の一例を図 3, 4 に示す. 評価対象は平均停止回数である. 図 3 は直進率 50% と直進率 90% の場合のシミュレーション結果である. 図 3 から直進率 90% の場合は代理強化学習の効果が大きい, 直進率 50% の場合は逆効果となっていることがわかる. これは, 直進率 90% の環境は通行権毎に交通流が明白になるため, その特徴を捉えやすいが, 直進率 50% の環境では通行権毎に交通流が明白ではないため, 特徴を捉えにくいためであると考えられる. 図 4 は交通量を変化させた場合と流入率を変化させた場合のシミュレーション結果である. 図 4 から交通流を変化させた場合, 代理強化学習が最も良く適応できていることがわかる. これは, 環境は変化するが, 制御方式は大きく変える必要がなく, 多くの状態に対して学習を行っている効果であると考えられる. しかし, 流入率を変化させた場合, 代理強化学習が環境変化時に逆効果となっていることがわかる. これは, エージェントが認識していない環境が変化している場合であり, 代理強化学習が間違った学習を行っているためであると考えられる.

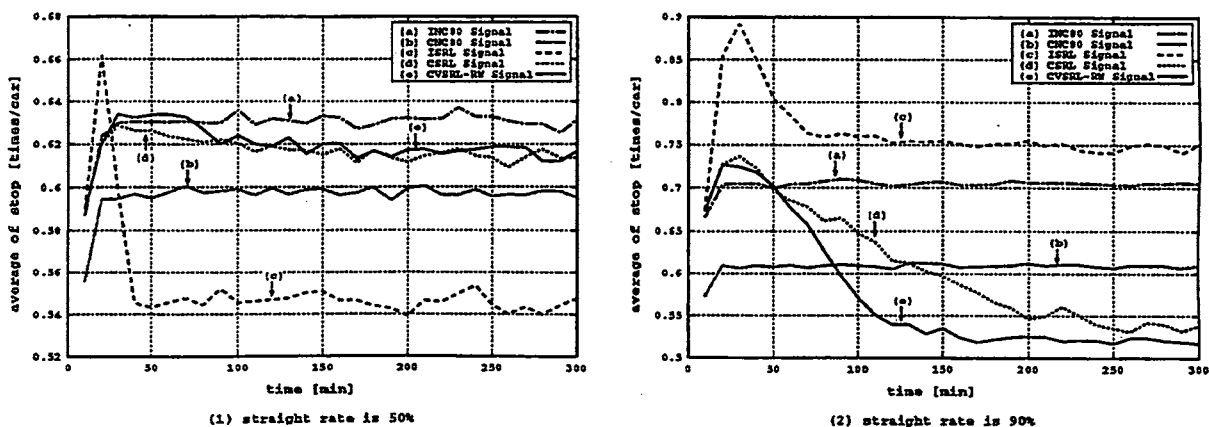


図 3: シミュレーション結果 (直進率)

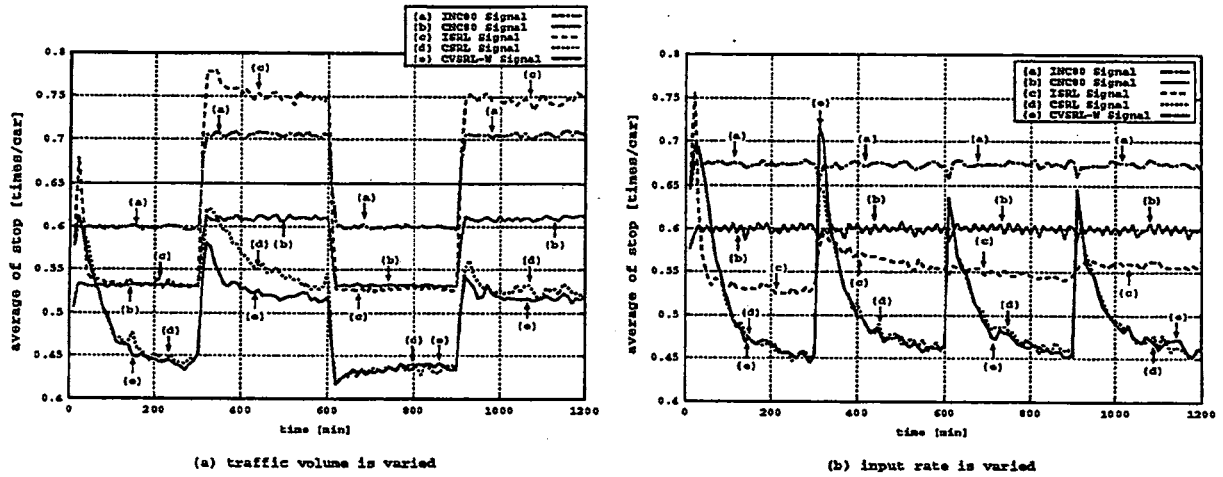


図 4: シミュレーション結果 (動的環境)

6. 結論

本研究では、マルチエージェントシステムにおける強化学習では、学習に多くの時間が必要であるという問題点に対して、自己強化学習と代理強化学習を提案した。これらを交通信号制御に適用してその有効性を検証した。シミュレーションの結果、交差点に特徴がある環境では代理強化学習が有効であるが、交差点に特徴がない環境では逆効果となり、自己強化学習の方が良い結果となることがわかった。また、動的環境に関しては逆効果となることもあるが、全体的に代理強化学習が良い結果となることがわかった。以上のように環境によって適している学習手法が異なるため、環境に応じた学習を行うハイブリッド型の学習手法が必要であると考えられる。また、本研究で提案した自己強化学習、代理強化学習は交通信号制御ばかりでなく、様々な分野に適用可能であると思われる。

学位論文審査結果の要旨

平成16年1月27日に第1回学位論文審査委員会を開催、1月29日に口頭発表、その後第2回審査委員会を開催し、慎重審議の結果以下の通り判定した。なお、口頭発表における質疑を最終試験に代えるものとした。

マルチエージェント環境のような大規模で複雑な環境では、学習すべき状態が膨大となり、学習の収束に多くの時間が必要であった。本論文では、強化学習の学習時間短縮手法として、自己強化学習、代理強化学習を提案し、大規模で複雑なシステムを必要とする交通信号制御に応用し、その有効性を示した。提案した自己強化学習は、予め既知である知識を用いて強化の成否を考察し、自ら強化信号を作成するため、実行したルールと同一状態の条件を持つルールに対して学習することが可能となり、学習時間短縮を可能にした。しかし、自己強化学習による同一状態の複数ルールの学習では、再び同じ状態を認識した場合には効果があるが、学習初期段階のような様々な状態を認識する場合には効果が小さいと考えられる。そこで、エージェント内の各状態を主体と考え、実行したルールの状態をモデルとし、その状態に与えられた強化を参考に類似する状態にも強化（代理強化）を与える手法を提案した。これによって、1回のルール実行において複数状態の複数ルールの学習が可能となり、学習初期段階における効果が期待できる。

以上の研究成果は、自己強化学習、代理強化学習に貢献し、また交通信号制御などの実用分野にも有効であり、本論文は博士（工学）に値するものと判定した。