

ルール生成機構をもつ

1H-8

強化学習型マルチエージェントシステム

上田 茂雄 阿部 倫之 服部 進実

金沢工業大学情報工学科

1 はじめに

ネットワークに存在するソフトウェアエージェントは、環境変化に適した行為を適応的に選択していくための自律性や学習能力を必要としている [3]。ここで、各サーバの性能や提供すべきサービスを意識してエージェントの振舞いを定義する場合、環境変化を事前に予測して必要な知識ベースを構成することは難しい。従って、環境状態とそれに対応する適切な行為が事前に決定できない場合、環境状態と行為の関係（行為決定ルール）を経験的に学習していくメカニズムが必要である。本稿ではニューラルネットワークの環境適応能力に着目し、行為決定ルールを自律的に学習していく推論メカニズムについて述べる。また、このメカニズムを用いたマルチエージェントシステムの構成について述べる。

2 強化学習型エージェント

2.1 知識構成

これまでのエージェントの知識構造は主に環境から得られる条件部と環境に働きかける行為部の組み合わせによって構成されてきた。しかし、一般的に得られる環境情報は独立に存在することはなく、それぞれに関連性を持つことが多い。そこで、この条件部と行為部を、ニューラルネットワークを模倣したネットワーク構造で構成することにより、再構成可能な知識構造の実現を図った。この知識ネットワークは環境情報を入力として扱い、出力を行為（メソッド）とすることで入力情報の関連性を崩すことなく非決定的知識を表現することができる。この知識を強化学習によって洗練することにより、予測が難しい環境変化に対して柔軟に追従できる知識構造を実現する。この知識ネットワークは複数の layer で構成される。

Multi-Agent System

based on Reinforcement Learning and Rule Generator

Shigeo Ueda, Noriyuki Abe, Shimmi Hattori

Kanazawa Institute of Technology

7-1 Ohgigaoka, Nonoichi, Ishikawa 921-8501, Japan

これらの layer は、

- 環境情報を取り入れる input layer
- 行為を処理する output layer

に分かれている。output layer は多段になることもある。各 layer 間は重み（ウエイト）を持ったリンクによって結合されている。最終 output layer 以外の output layer を hide layer と呼ぶ。

この知識ネットワークの構成を図1に示す。

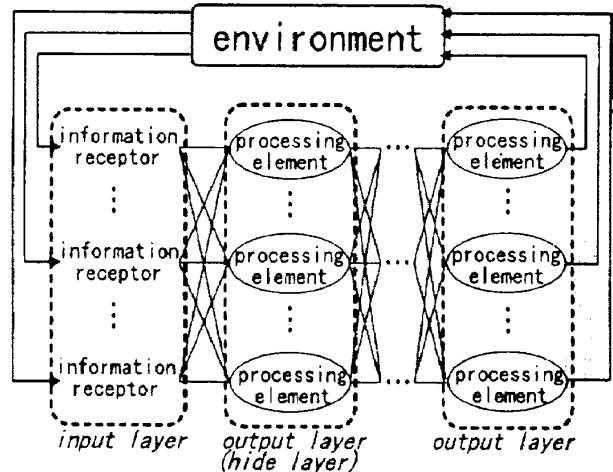


図 1: 知識ネットワークの構成

2.2 強化学習

強化学習は、報酬を契機にして行為の選択過程を思考錯誤的に洗練していく教師なし機械学習方法であり、環境に対する自律的適応能力の実現が期待できる [3]。

ここで、サーバを管理するエージェントでは、時間とともに変化する環境の中で自己の負荷を制御し、最大限にシステムに対して貢献することが望まれる。そのため、本マルチエージェントシステムでは、行為の選択過程に活性度を設け、行為の実行によって発生した環境変化の状況に基づいて報酬値を算出し、それを各リンクに分配することで強化学習を実現している。

学習には、行為の選択直後に行われる活性化モードと環境変動時に行われる報酬分配モードがある。

活性化モードにおける学習では、output layer の中で 行為を実行したエレメントに活性値を与え、この活性値を input layer 方向に向かって分配する。活性値は行為の貢献度を表わしており、リンクの重みに比例した分配を行う。活性化モードでは行為の評価は行われない。報酬分配モードでは、環境変動に同期したタイミングで報酬値が与えられると、各エレメントとリンクに対して、活性値に比例した配分率で報酬値が分配させる。

これにより各エレメントの閾値やリンクの重みが変わり、環境情報(条件部)と適切な行為の関連性が強化され、結果として必要なルールが適応的に学習されていく。

3 マルチエージェントシステム

章2で示した強化学習型エージェントの実現例として、マルチサーバシステムを制御するマルチエージェントシステムの構成を示す。

システムの概念構成を図2に示す。

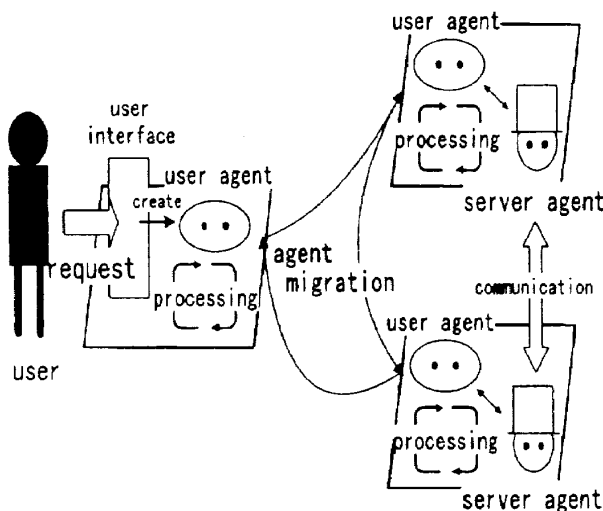


図2: マルチエージェントシステムの概念図

本システムはサーバエージェントとクライアントから送られてくるクライアントエージェントの2つから構成される。サーバエージェントは自己の管理するサーバのメモリや処理能力を環境情報として扱い、クライアントエージェントからリクエストされたサービスをどのように処理するのかを決定する。また、サーバエージェントは、積極的にサーバ群の負荷を分散させている場合に多くの報酬値が与えられるようにしている。なお、クライアントエージェントは定義した処理をモバイルさせることによってサーバに処理を依頼する。このマルチエージェントシステムの構成を図3に示す。

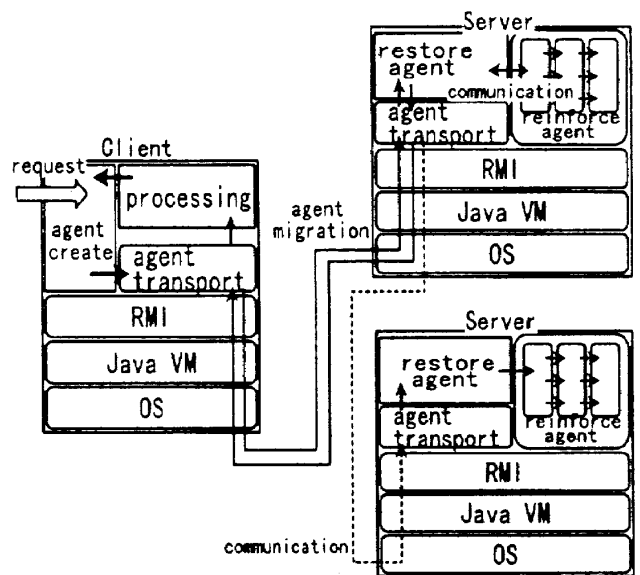


図3: マルチエージェントシステムの構成

クライアントエージェントはクライアントの要求や行き先などを持ってサーバに送られる。その後サーバエージェントとネゴシエーションを行い、サーバエージェントは処理を受理するか否かを決定する。処理が受理されない場合には、他のサーバエージェントに処理が委託され、自律的な負荷分散が行われる。環境の変化に対しては、サーバの負荷を平滑化することにより適した行為が選択されるように、エージェントの行為決定ルールが自律的に洗練されていく(ルール生成)。

4 おわりに

本稿では、ルール生成機構をもつ環境変動に適応する強化学習型エージェントとそれを用いたマルチエージェントシステムについて述べた。現在、Javaを用いてマルチエージェントシステムを作成しており、エージェントの動作やコミュニケーション方法について検討を行っている。今後、学習能力について実験的に評価を行う予定である。

参考文献

- [1] J.H.Holland, K.J.Holyoak, R.E.Nisbett, P.R.Thagard, "Induction: Process of Inference, Learning, and Discovery", the MIT press(1986)
- [2] 畷見達夫, "強化学習", 人工知能学会誌, Vol.9, no.6, pp830-836(1994)
- [3] 石田亨, 山田誠二, 他, "特集「エージェントの基礎と応用」", 人工知能学会誌, Vol.10, no.5, pp662-711(1995)