論 文 の 内 容 の 要 旨

論文題目　　部分観測環境におけるエージェントの自律的行動獲得

氏 名　　井 上 康 介

身体性を有するエージェントが実環境で動作する際には，エージェントによる現在おかれた状況の認識およびそれに基づく動作決定の方法は，身体・環境・タスク間の実際の相互作用の様態に依存するものとなる．このとき，エージェントはその状況認識および動作決定のための内的機構を実際の環境・タスクとの相互作用に立脚して自律的に規定する必要があり，このためには状況認識機構・動作決定機構の学習による獲得が必要となる．エージェントがこれらを獲得する上では，エージェントの身体に属するセンサ・アクチュエータに基づく情報が利用されるが，このとき身体性から帰結される以下の問題が解決される必要がある．

（1） エージェントは自らの持つセンサ系が与える局所的な情報を用いて状況認識を行わなければならない．このとき，観測入力の持つ局所性に起因して，エージェントの観測はMarkov性を失うため，環境は部分観測Markov決定過程（POMDP）としてモデル化され，このモデルに基づく状況認識を行う必要がある．このとき，エージェントは環境の部分観測性に対処するため，状況認識において自らの持つ短期記憶を利用することによって観測におけるMarkov性の破れを除去しなければならない．

（2） 特定の状況においてエージェントが受け取る観測入力は，実際の環境との相互作用を行う時点で初めて明らかになるものである．従って，エージェントが観測入力をいかに解釈し，それに基づいて状況を規定するかの写像関係を予め設計者がエージェントに与えることは困難であり，観測入力の解釈様式は実際の環境との相互作用に立脚して規定されなければならない．

（3） エージェントが複雑なタスクを実現する場合，エージェントはタスク全体を構成する個々のサブ・タスクを実行するための動作原理を有していなければならない．ところが，エージェントが身体性に起因する上記2点の拘束を受けている場合，エージェントは現在達成すべきタスクの認識及び個別のタスクにおける現在の状況の認識の双方を，身体・環境間の相互作用に立脚した形式で実現せねばならない．

本研究では，上記３点の問題を踏まえて，状況・動作に対する評価として，外部から各時点における動作の評価が即時報酬として与えられるという前提の下で，この即時報酬に基づいて身体・環境・タスク間の相互作用に立脚した状況認識機構および動作決定機構をエージェントが獲得する手法を提案する．このように外部から与えられる即時報酬に基づいて行動獲得を行うことから，提案手法は一種の教示システムと見なすことができる．教示システムとしての提案手法の優位点は，エージェントの状況認識方法を予め設計する必要がなく，またエージェントの各時点における動作の評価という少ない教示情報に基づいて教示が可能であるという点である．

第３章では，単一のタスクに対して状況認識機構・動作決定機構を獲得する学習手法を提案した．ここでは，環境の部分観測性に対処するために，エージェントの観測・動作の短期記憶情報に基づく状況識別を行うことにより Markov 性を回復し，観測情報に基づく状況識別の方法を身体・環境・タスク間の相互作用に立脚して適応的に獲得する．具体的には，短期記憶に基づく状況識別を表現する決定木構造の内的状態表現を利用し，この決定木に対して観測情報に基づく状況識別を表現する分岐を適応的に追加するという方法をとる．図１には提案手法における内的状態表現を示す．ただし，図中 o_τ は時刻 τ における観測に基づく識別に対応するレイヤ，a_τ は時刻 τ における動作に基づく識別に対応するレイヤであり，図中 t は現在時刻を示している．各状況に対応する内的状態は，葉ノードに対応する．

提案手法では，この内的状態表現をタスク試行を繰り返す過程で逐次的に獲得する．具体的には，開始時には内的状態表現は単一の状態からなり，これに対して逐次的に状態分割すなわち決定木における分岐を加えてゆく．状態構成の目的は，各内的状態において同一の動作に対応して得られる報酬が同一となることである．したがって，同一の内的状態を過去訪れた際に行われた同一の動作に対する報酬の履歴群において，その報酬がばらつきをもつとき，その状態は分割される．ここで，報酬のばらつきには以下の２つの原因が考えられ，それぞれに対応する対処が必

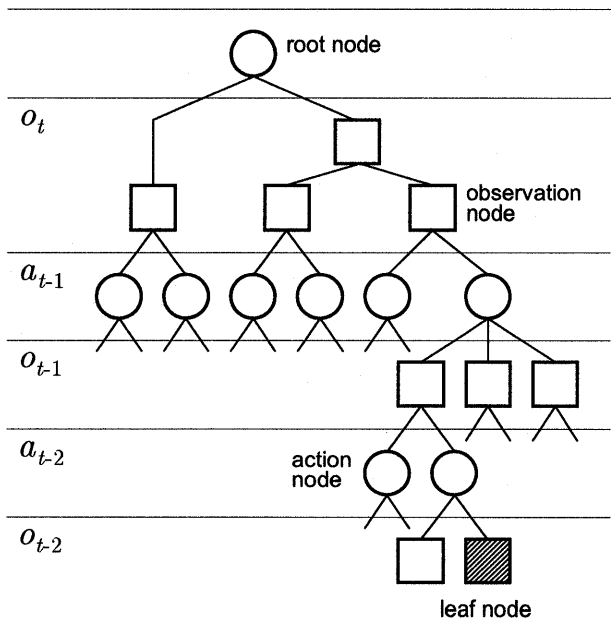


図１　内的状態表現

要となる：（1）本来観測情報に基づいて識別可能な状態を混同している：この場合はより詳細に観測値の識別を行うために観測情報に基づく分岐を追加する，（2）環境の部分観測性に起因する知覚騙し問題：観測値による識別は不可能であるため，より過去の短期記憶を参照して状況識別を行うために，より下層のレイヤを追加する．提案手法ではこれら2つの原因を，注目する内的状態における特定の動作に対して過去に得られた報酬の観測空間上での分布に対する統計的指標を用いて判別する．

このようにして獲得された内的状態表現内の各内的状態には行動方策が内包され，この方策を得られた報酬に基づいて学習することにより行動の獲得を実現する．

知覚騙し問題を顕著に含む単純な通路状グリッド環境におけるナビゲーション問題のシミュレーションに提案手法を適用し，適切な内的状態表現の獲得に基づく行動の獲得を確認した．

第4章では，上で提案した単一タスクに対する行動獲得手法を複数タスクに拡張する．身体性を有するエージェントが自律的に状況認識機構を獲得する場合，その状況認識の様式は身体・環境・タスクに依存するものとなるため，個別のタスクの識別と個々のタスク上での状況の識別との双方を身体性に即して実現されなければならない．提案手法では，この問題を解決するために，個々のタスクに対する状況認識・動作決定機構の獲得を行った上で，現在扱っているタスクの識別を実現するタスク識別機構を，身体性に即して，即ち自らの観測・動作の履歴情報だけに基づいて実現するという階層的構造を獲得するための学習スケジューリング手法を採用する．提案手法では，個別のタスクに対する行動を獲得したエージェントを未知のタスクに対して適用し，それぞれのタスクを行っていることの蓋然性を内的に表現する確信度と呼ばれるパラメータを利用することでタスク識別を実現するが，ここでは以下の2点の問題を解決しなければならない：（1）個別のタスクに対する学習過程において得られた経験データをタスク識別における判断基準として利用するが，このデータが既に十分得られているという保証がない，（2）タスクを識別するためには一定の動作系列を実行する必要があるが，このタスク識別行動を行った時点から，対応するタスクを遂行するまでの行動が，対応するタスクに対する学習過程で獲得されているという保証がない．提案手法ではこれらの問題を，経験データおよび行動の獲得を行う追加学習によって解決する．

単純な通路状グリッド環境上における，複数のスタートからのゴールへのナビゲーション問題を複数タスクと位置づけたシミュレーションにより，提案手法によって複数タスクの識別・遂行を実現する行動の獲得を確認した．

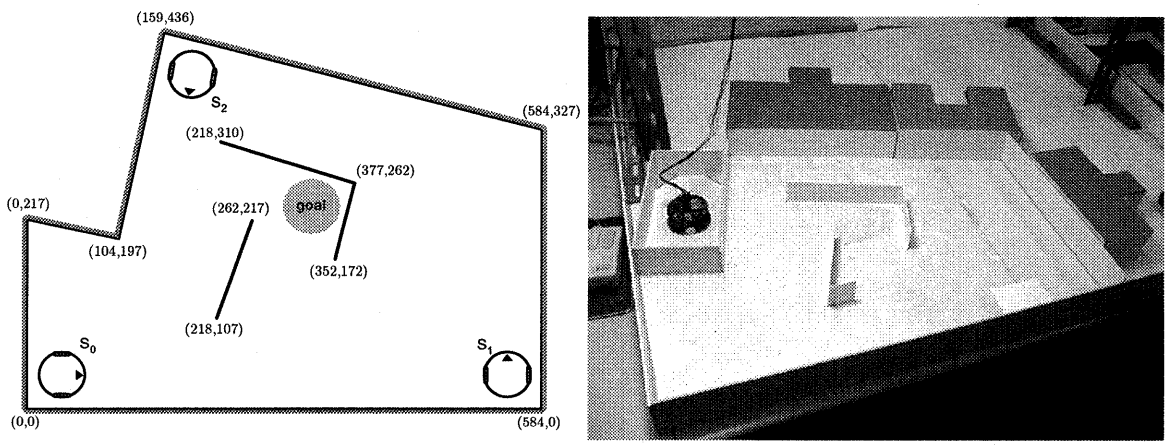
第5章では，以上の提案手法を実移動ロボット Khepera およびそれを想定した実際的なシミュレーションにより検証した．検証1として，実移動ロボットにおける行動の獲得の確認を行い，検証2として，即時報酬の付与原理がエージェントの身体性に立脚しない形で実装されている場合に対するシミュレーションにより，設計・教示労力の小さい教示システムとしての提案手法の動作を確認した．

検証1では，図2に示す環境におけるナビゲーションに対して，ゴールへ到達するために必要な最小動作ステップ数の増減に基づく即時報酬に基づいて行動獲得を行った．ただし，図中 S_0，

S_1およびS_2が各タスクに対応するスタート点である．検証では，これらのタスクに対する行動獲得を，実移動ロボットのセンサ・モータ系を再現した実際的なコンピュータシミュレーション上での学習によって実現し，獲得された行動を実移動ロボットによる実験で検証した．この結果，実際のロボットによるナビゲーションタスクに対して提案手法による行動獲得が可能であることを確認した．

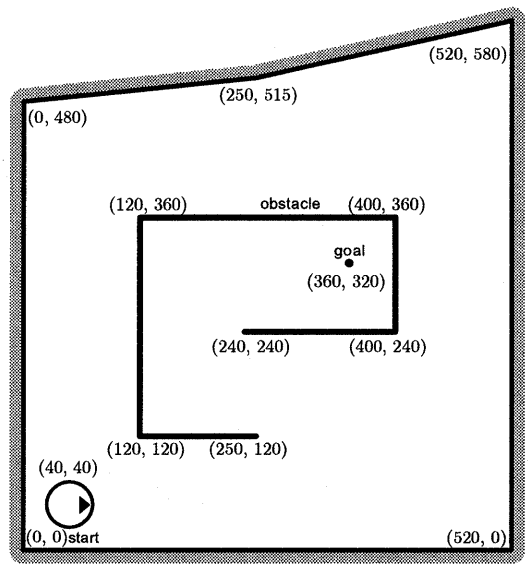
検証2では，図3（A）に示す環境において，図中（B）に示すWavefront法に基づくポテンシャルによって即時報酬を与え，この即時報酬による単一のスタート・ゴールに対するナビゲーション行動の獲得をシミュレーション上で行った．図中（C）は学習過程の各試行において消費された動作ステップ数を示しており，最小のステップ数によるゴール到達行動が獲得されていることが示されている．これにより，即時報酬がエージェントの行動原理に依存していない場合についての提案手法による行動獲得が示された．

第6章では，提案手法に対する考察と評価を行った．まず単一タスクに対する学習手法に関して，計算量・記憶量，環境の形状や学習パラメータに対する性能の依存性，状態表現における観測値の利用の意義，および与える即時報酬の満たすべき条件を議論し，実環境への適用に関しては，誤差の影響を議論した．また，提案手法の適用の可能な範囲とそれを限定している要因，およびそれに対する展望について議論した．

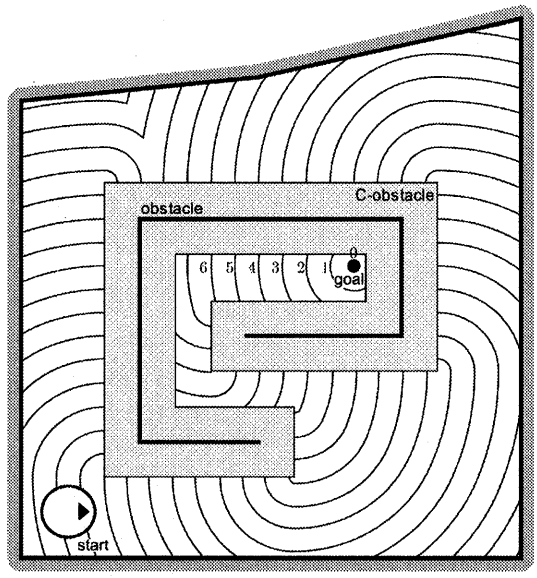


（A）環境の配置　　（B）実験環境

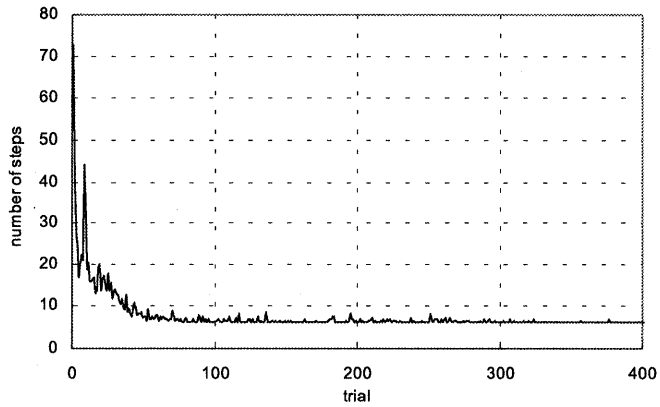
図2　検証1の環境



（A）環境の配置



（B）Wavefront ポテンシャル



（C）学習結果

図３　検証２

以上