

## 論文の内容の要旨

### 論文題目

任務環境の不確定性に対応可能な非ホロノミック自律型海中ロボットの  
強化学習による動作計画法

氏名 川野 洋

現在、その開発が実用化の段階に入っている自律型海中ロボットは、そのほとんどが非ホロノミックなロボットであり、Under Actuated Mechanism を持つ。非ホロノミック自律型海中ロボット (NHAUV: Non-Holonomic Autonomous Underwater Vehicle) は横方向推進用の推進器を持たないため横方向への直接的な移動が出来ず、その場回頭が不可能なことが多い。そのような NHAUV は、一般には時不変フィードバック制御の手法では位置制御を行うことが不可能であることが知られており、NHAUV が障害物の密集する環境内を安全に航行するには、NHAUV の限られた運動特性を考慮した動作計画アルゴリズムが必要である。本研究では、そのような NHAUV の動作計画アルゴリズムについて研究する。

NHAUV は、陸上の 4 輪自動車型非ホロノミックロボットなどと比べて、動力学特性が複雑であり、4 輪自動車型非ホロノミックロボットの動作計画で使用されているような A\* などの決定的な経路計画手法を直接適用することが出来ない。具体的には、NHAUV が旋回動作をする際の横滑り現象や、潮流による旋回時の軌道の変形などがその理由である。そこで本研究では、強化学習の手法の一つである準マルコフ決定過程における Q 学習 (SMDP 法) を応用した確率的な動作計画の手法によりこの問題を解決する。

Q 学習は、離散的な状態変数によって記述される環境において適用可能であり、環境の動特性がマルコフ決定過程に従うという仮定の元で適用可能な手法である。そのような Q 学習のサブセットの一つである SMDP 法は、NHAUV の動作計画において、航行所要時間の評価を行うのに有効である。Q 学習を NHAUV に適用する上での問題点としては、一つ目には複雑な NHAUV の動力学特性に起因する非マルコフ性の問題、二つ目には NHAUV の運動を記述する状態変数の多さに起因する状態空間の爆発の問題、三つ目には Q 学習が不安定な学習手法であり学習の収束を補償するのが難しい問題、四つ目には Q 学習によって獲得された行動アルゴリズムを学習を行ったときとは異なる環境で適用するのが難しいという問題などである。特に最後の問題については、任務行動中の想定外のアクチュエー

タ故障の発生や、海中環境における未知な潮流分布や不確定要素を含んだ海底地形の存在を考えると重要な問題であると言える。

本研究では、一つ目と二つ目の問題に対して、SMDP 法と最急降下法を併用した階層的学習アルゴリズムを提案する。三つ目の問題に対しては、教示の手法によって学習の効率化を図り、学習プロセスを安定化することを考える。四つ目の問題の解決法として、Bayesian Network を応用した運動モデルの導入によって、学習中に NHAUV の運動特性を蓄積し、蓄積された知識を再利用することによって、学習の時とは異なる環境に対応可能な動作計画アルゴリズムの獲得が可能であるようにする。さらに、運動モデルの導入は、学習プロセスの効率化にも役に立つ。

本研究で提案する階層的学習アルゴリズムは、上流と下流の 2 層に分かれたモジュールが直列に結合した構造をしている (図 1)。上流モジュールを行動選択部と呼び、下流モジュールをアクチュエータ制御部と呼ぶことにする。行動選択部では、SMDP 法によって計算された  $Q$  値を利用した動作計画が行われ、アクチュエータ制御部では動作計画に従って NHAUV のアクチュエータを制御する。アクチュエータ制御部の学習は最急降下法によって行われる。NHAUV の動作計画を行うためには、少なくとも NHAUV の位置と方位角、方位角速度を参照しなければならないが、そのような多くの状態変数を使って  $Q$  学習のための状態空間を構成すると状態空間の爆発を引き起こすので、本研究ではこれらの状態変数を上流と下流のモジュールに振り分けることによって状態空間の爆発を回避する。すなわち、上流の行動選択部では、NHAUV の位置と方位角、下流のアクチュエータ制御部では方位角と方位角速度を参照する。行動選択部が方位角速度を参照しないことによって起因する非マルコフ性の悪影響をアクチュエータ制御部で補償する。学習中の NHAUV の行動による全ての状態遷移は、運動モデルに蓄積される。行動選択部の学習は、運動モデルのデータを用いて行われる。そのため、一度運動モデルが適切に獲得されれば、それを利用して学習のときとは異なるさまざまな条件に適用可能な動作計画アルゴリズムを導出可能である。例えば、アクチュエータ故障時のための動作計画アルゴリズムや、異なる配置の障害物が存在する海域における動作計画アルゴリズムなどが運動モデルによって導出可能である。

本研究では、2 種類の教示の手法を導入する。一つ目は詳細教示と呼ぶもので、学習中の NHAUV の行動選択の仕方を教示者の手で直接制御する方法である。詳細教示を受けた NHAUV は、教示を受けた通りに行動し、その行動結果を運動モデルに蓄積していく。教示者によって与えられた経験は、運動モデルを使った SMDP 法による学習の際に評価され、教示されたのが良い行動の場合は高い報酬が与えられるが、悪い行動であれば低い報酬が与えられる。そのため、詳細教示の手法においては、教示者のミスを含んだ教示内容によって NHAUV が間違った行動を学習することを防止できる。二つ目は包括教示と呼ぶもので、学習中の NHAUV が危険な状況に陥るのを常に防止する機能を持つ Behavior Based アルゴリズムのハードコーディングによって行われる。NHAUV は、未経験の状態にある

ときや、任務空間の外部に位置する時には、包括教示によって与えられた Behavior Based アルゴリズムによって行動選択を行う。この手法により、例えば潮流中で学習を行う NHAUV が任務環境を遠く離れた場所まで潮流によって流されてしまうのを防止することができ、学習プロセスを安全に続行させることが出来る。

本研究では、未知の潮流分布に対応するために、異なる一様流中条件での航行を想定した複数の動作計画アルゴリズムを、運動モデルを利用した学習で導出し、それらを併用する手法を採用する。実際の航行においては、NHAUV の位置における潮流条件に合った動作計画アルゴリズムを随時選択して航行制御に使用することとする。また、その形状に不確定要素を含む海底地形に対応するために、Behavior Based アルゴリズムによる縦方向の高度制御アルゴリズムと、本研究で獲得された 2 次元平面内航行用の動作計画アルゴリズムを併用する手法をとる。すなわち運動モデルを用いた動作計画アルゴリズムの導出の際に、NHAUV の上下動の運動性能の限界を考慮することによって、NHAUV の運動性能では回避不可能な険しい斜面を避けつつも目的地まで NHAUV を誘導可能な動作計画アルゴリズムを導出する。

本研究で提案するアルゴリズムはシミュレーションによってされる。想定する NHAUV モデルとして、東京大学生産技術研究所によって建造された R-One Robot を想定する (図 2)。図 3 に、本研究で提案する学習アルゴリズムによって獲得された運動制御アルゴリズムによる垂直推進器故障時の NHAUV の潮流中での垂直上昇動作を示す。NHAUV は垂直推進器の代わりに昇降舵を利用して垂直上昇を行っている。この動作は、教示者によって与えられたものではなく、NHAUV が学習中の経験から自動的に導き出したものである。図 4 に本研究で提案する学習アルゴリズムによって獲得された動作計画アルゴリズムによって、NHAUV が水平面内の 2 次元円柱近傍を航行する様子を示す。円柱周りの潮流分布は NHAUV にとっては未知なものであり、このような未知な潮流分布中でも NHAUV の運動性能を考慮した動作計画によって適切に航行制御が行われているのがわかる。図 5 に、誤差を含む海底地形モデルを利用して導出された、定高度航行用の動作計画アルゴリズムによる NHAUV の航行の様子を示す。NHAUV は定高度航行により、NHAUV の運動性能では回避不可能な険しい山を回避しつつ目標地点への到達に成功しているのがわかる。

以上のシミュレーションの結果より、本研究で提案する学習アルゴリズムによって、さまざまな不確定要素を含む海中環境にて適用可能な動作計画アルゴリズムを NHAUV に獲得させることが可能であることが示された。

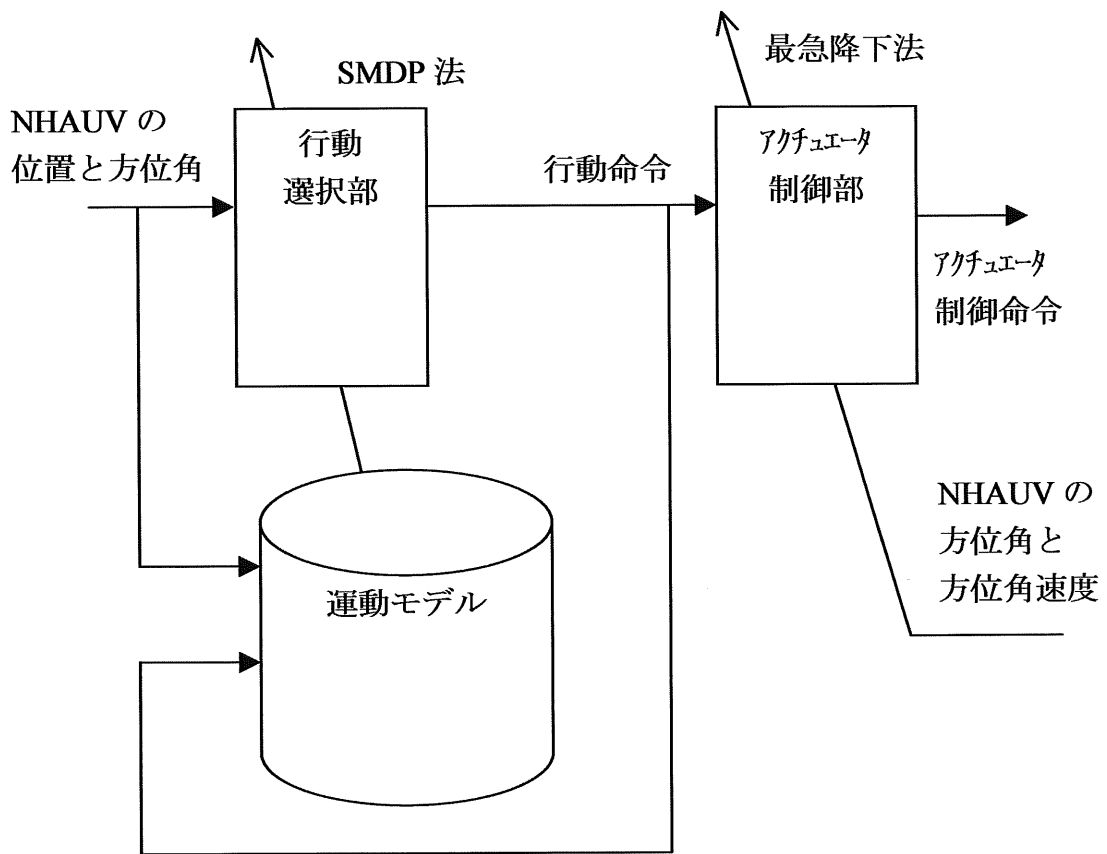


図1 階層的学習アルゴリズムの全体構成

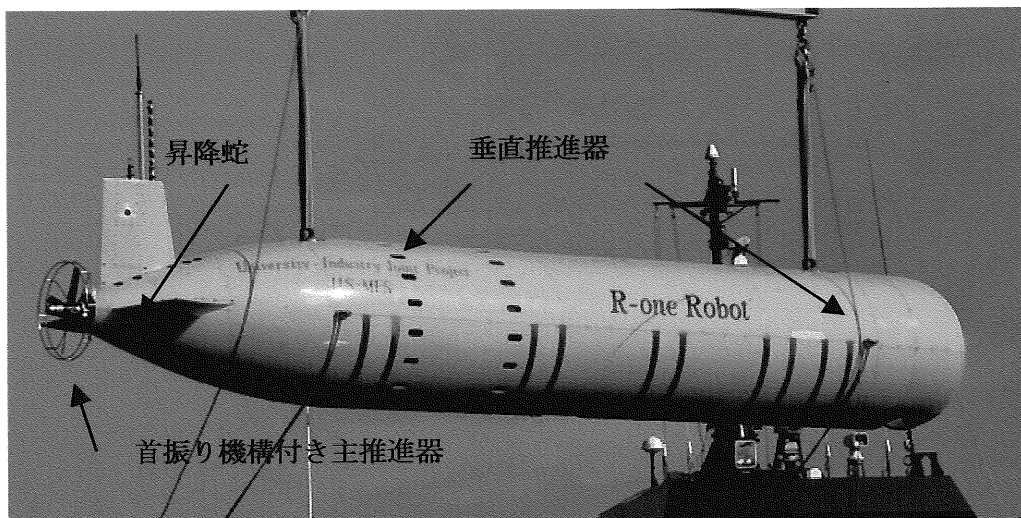


図2 本研究で非ホロノミック AUV モデルとして想定する R-One Robot

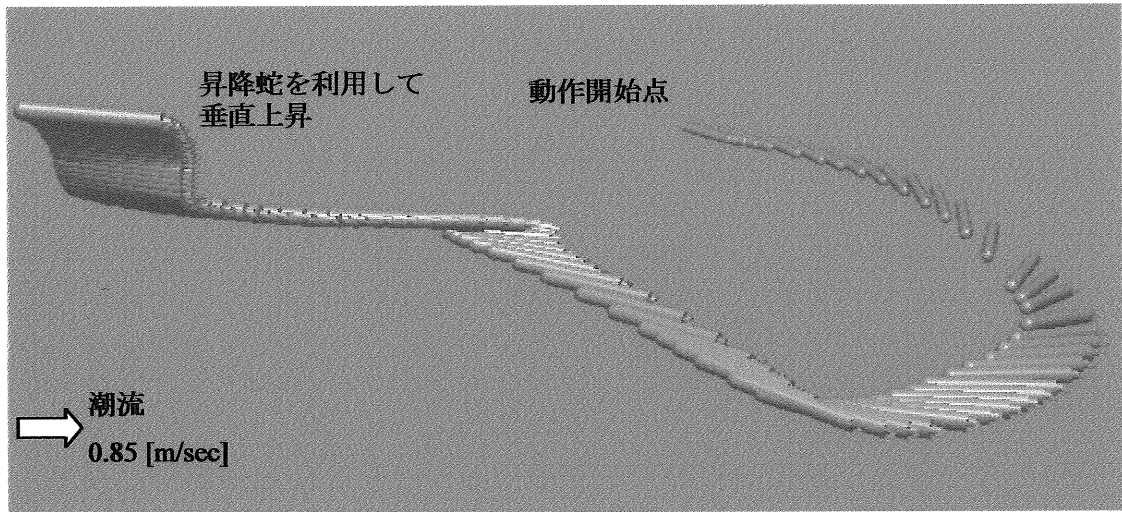


図3 獲得された制御アルゴリズムによるアクチュエータ故障時の垂直上昇動作  
 ( 垂直推進器が故障しており、代わりに昇降舵が使用されている )

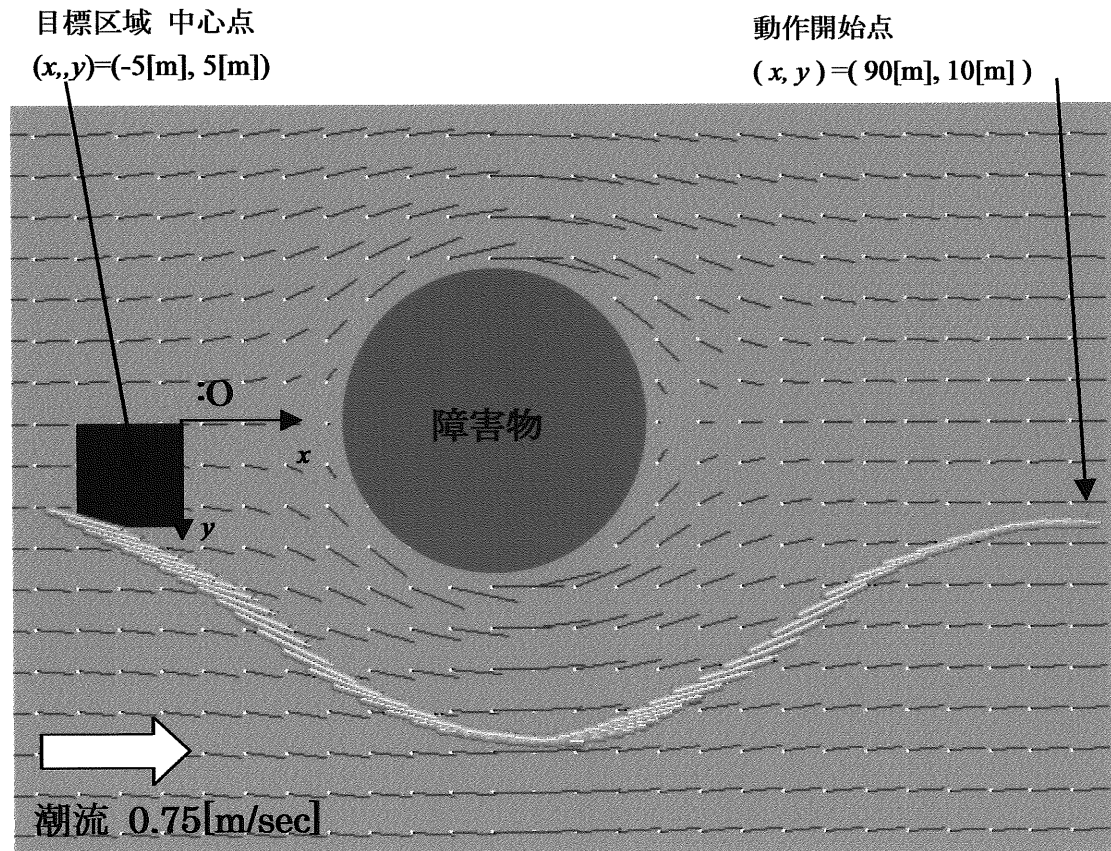


図4 非均一未知潮流分布中における非ホロノミック AUV の動作軌跡

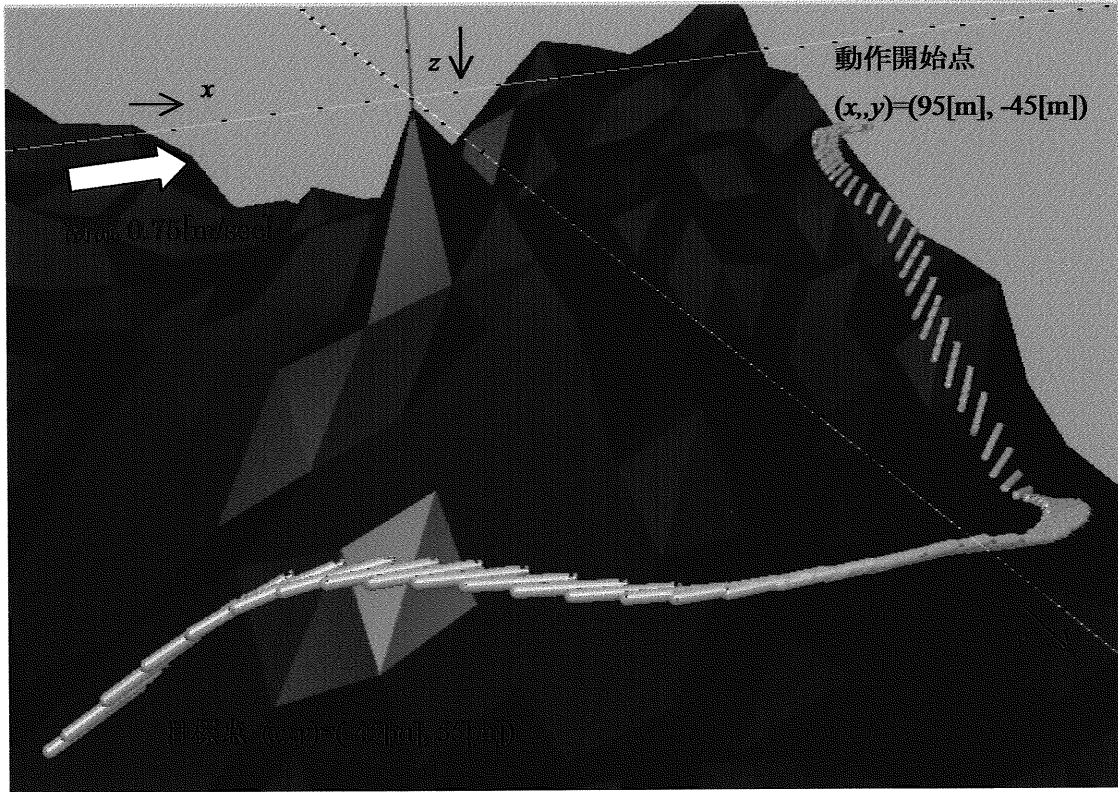


図5 定高度航行を想定した動作計画アルゴリズムによる NHAUV の動作の様子