

協調移動フレーム手法による

逆強化学習を用いた移動体の行動予測の実現

梁宇昕[‡]間瀬正啓[‡]

1. はじめに

日本をはじめとする先進国では、少子高齢化に伴う労働人口の減少が社会問題となっている[1]。特に海運では労働人口の減少に伴う人件費の増加が問題となっており、船舶を自律操縦させる自律航行の検討が活発になってきている[2]。自律航行の要素技術として、周辺船舶との衝突を避けながら安全に航行するために周辺船舶の進路方向の変更等の行動を予測する技術が求められている [3]。

船舶のような移動体の行動予測には、対象がある状況で如何に行動するかをルールとして記述し、セルオートマトンのようにエージェントシミュレーションする手法が考えられる[4]。この時、現実では複合的な要因により行動が決定され、全てのルールを記述すると複雑になりすぎるとい問題がある。そのため、逆強化学習[5]を利用し、移動体の軌跡データからルールを自動抽出することを目指す。

逆強化学習では、周囲の環境が同一な条件において対象が行動した十分な数の軌跡データが教師データとして必要となる。Sharifzadeh らは自動車のレーン変更を例にシミュレータ上で教師データを生成し、逆強化学習を適用することで Deep Q-Network として行動ルールを抽出している[6]。しかし、実世界に対し適用しようとする、対象の周辺状況が時々刻々と変化するため、そのままでは適用できない。

そこで、軌跡データから操縦者が注目する部分を順次切り出す協調移動フレームを提案する。時系列に変化する環境のうちの一部を切り出すことで動的な環境を準定常的な環境として扱い、逆強化学習を適用可能とする。

2. 強化学習と逆強化学習

環境を動き回る対象の行動を統計的に学習する手法として、強化学習と逆強化学習が挙げられる。これらの手法は、図 1 に示すマルコフ決定過程(Markov Decision Process (MDP))上でモデル化される。この MDP において、強化学習は報酬関数 $R(s,a)$ と状態遷移確率 $P(s'|s,a)$ が与えられた時に状態価値関数 $V(s)$ 、行動価値関数 $Q(s,a)$ を最大化する方策 $\pi^*(a|s)$ を環境における試行錯誤で探索的に求める手法である。その逆問題である逆強化学習は状態遷移確率 $P(s'|s,a)$ と最適な方策 $\pi^*(a|s)$ が与えられた、すなわち専門家の行動を写像した状態と行動の時系列的な変化の軌跡データが与えられた時に、その要因となった報酬関数 $R(s,a)$ を求める手法である。

本論文では、船舶の行動が最適な方策に基づくものとして、船舶の状態と行動の時系列的な変化の軌跡データから報酬関数を導出する逆強化学習を適用することを考える。適用可能な手法として、MDP の状態を航行する船舶の周囲の環境情報により定義する手法が考えられる。

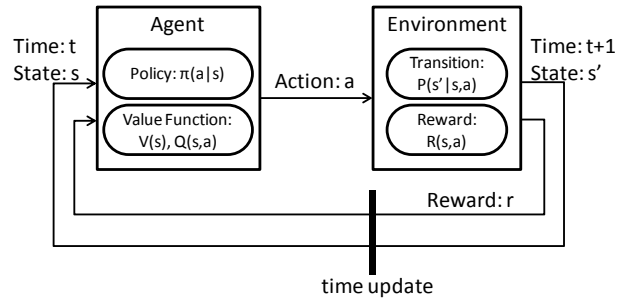


図 1 マルコフ決定過程(MDP)

この場合、周囲の環境情報は学習データ間で同一である事が求められるが、障害物として環境情報に含まれる、動的に航行する他の船舶の存在により環境が同一でなくなってしまう、それぞれを別の学習データとして扱うしかない。その上、周囲の船舶の行動傾向は観察のたびに変わるため、環境が同一な学習データを取得できないことが問題となる。

3. 提案手法

人間の行動の指針に照らし合わせると、その人間が行動を行う時に考慮する情報は、その人間が観察している情報に限定される。このような情報として、大域的な情報からその観察している情報のみを局所的に抽出し、環境として扱う空間を局所化すれば、位置が変化する船舶の数を減らして状態の数を削減できる。この手法は、予測精度を落とさずに単一の軌跡データから複数の学習データを抽出し、逆強化学習に用いる学習データのレコード数を増加させることが期待できる。

そこで、行動データから運転者が注目する局所的な環境情報を局所化・相対化して順次抽出する協調移動フレーム((Co-Moving Frame(CMF)))を提案する。船舶に対して CMF を適用した場合の処理イメージを図 2 に示す。まず、軌跡データを第一人称の視点に基づき相対化する。例として、前方から 2 隻の船舶が行動分析対象船に接近する状況を 2 枚の局所環境(CMF)データとして切り出す①。この 2 枚の CMF では、対象の船舶の周辺の 2 隻の船舶の行動パターンはほぼ同一である。したがって、このようなデータを束ねることで、類似した環境における様々な軌跡データを含むデータセットを作成できる②。

第一人称視点による協調移動フレームを図 3 のように実装する。行動分析対象の運転者が観測する情報として、他の船舶・土地・目的地との相対距離および、目的地に向けて航行するために、目的地の方向と距離を追加する。

以上より、軌跡データのデータセットに対し CMF を適用することで、定常的な環境における様々な軌跡データを含むデータセットを作成できる。これにより、逆強化学習を適用し、行動ルールを抽出できるようになる。

‡ 日立製作所研究開発グループ

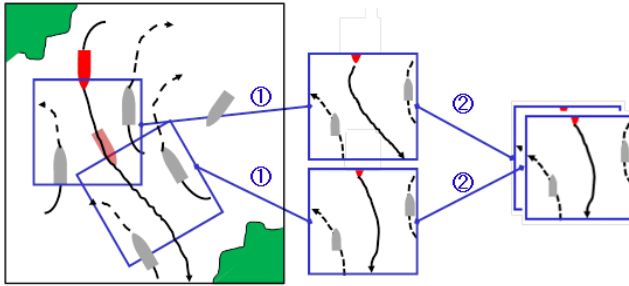


図2 CMFを用いるデータ処理のイメージ

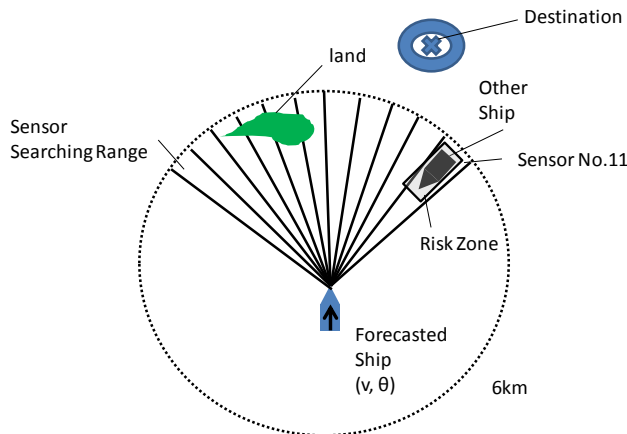


図3 CMFの実装概念

4. 評価

抽出した行動ルールの精度について評価を行うため、東洋信号通信社が販売する2015年9月1日～30日の東京湾内のAISデータ[7]を用いて、混雑した海域にある船舶の進路変更を予測する実験を行った。浦賀水道と横浜港に囲まれた海域(東経139.618~140.1度, 北緯35.342~35.545度)を航行する船舶を対象として評価を行った。この海域は、東京湾内の複数の港に入出港するために必ず通過する必要があり、船群が6方向からこの海域に入出入りする。そのため、船舶が互いの行動を予測する必要があるが、人間でも正確に予測する事は困難で事故が頻発しており[8]、行動予測による衝突回避に対するニーズが高い海域である。

衝突事故を緊急回避するためには、少なくとも3分以上前に衝突リスクを知る必要があるといわれている[9]。そこで、対象船の5分先の進路方向の変化について、左折45°、同30°、同15°、直進、右折45°、同30°、同15°のいずれかに該当するかを予測する。利用するデータのうち、9月1日から20日までのデータを用いて学習を行い、9月21日から9月30日のデータに対して予測精度の評価を行った。本評価では、学習データ、評価データ共に35分以上連続して座標が観測されたレコードを軌跡データとして用いた。

評価は9月1日から20日までのデータセットから抽出された学習データのレコード数と、対象の船舶の進路方向の変化を予測した際の正解率について、CMFを適用した提案手法と適用しない従来手法とを比較した。従来手法では軌跡データから、地理的情報(緯度経度情報に紐付く陸地の有無)を用いて、セルオートマトンベースで行動モデルを作成した。

評価結果を表1に示す。従来手法では、評価対象のデータセットから1,486レコードの学習データが抽出されたのに対し、CMFを用いた場合は、約30倍の46,638レコードを学習データとして抽出できた。これにより周囲の船舶を考慮しても十分なデータ量が確保できるようになり、行動予測精度が従来手法の52.0%に対して提案手法では81.2%と29.2%上昇した。この精度は船が相互のGPS座標を連絡した場合に、知見に基づき相手船の方向を算出した精度である82%に匹敵する[10]。

表1 行動予測精度評価結果

Methodology	Number of Available Train Data	Turn Action Forecasting Accuracy
Conventional Method	1,486 [records]	52.0 [%]
Proposed Method	46,638 [records]	81.2 [%]

5. おわりに

先進国では、労働人口の減少に対して、自律操縦による労働代替が求められている。安全な自律操縦実現のためには、周囲の移動体の行動を精度良く予測する技術が求められている。似た状況における多数の行動データがあれば、逆強化学習により行動を統計的に分析できることが知られているが、現実には周辺の状況は変化するため、似た状況は滅多に発生せず学習データ量が不足することが問題となっている。そこで、行動データから操縦者が注目する部分を順次切り出す協調移動フレームを提案した。東京湾のAIS航跡データを用いた評価の結果、学習データ数が約30倍に増えて逆強化学習が適用可能となり、従来手法では52.0%の精度でしか予測できなかった対象船舶の行動を81.2%の精度で予測できた。これより、自律操縦における周囲の移動体の行動予測を実現する見込みが得られた。

参考文献

- [1] 総務省, “平成26年版 情報通信白書 我が国の労働力人口における課題,” (2014).
- [2] Rolls-Royce, “Remote and Autonomous Ships the Next Steps,” AAWA White Paper, (2016).
- [3] 日本海難防止協会, “安全航行のいろは” (2011). <http://www.nikkaibo.or.jp/pdf/iroha.pdf> (2017年4月24日閲覧)
- [4] 大鑄史男, 小野木基裕, “セルオートマトン法による避難流動のシミュレーション,” 日本オペレーションズ・リサーチ学会和文論文誌, 51巻, 94-111, (2008).
- [5] Pieter Abbeel and Andrew Y. NG, “Apprenticeship Learning via Inverse Reinforcement Learning,” ICML2004 (2004).
- [6] S. Sharifzadeh, I. Chiotellis, R. Triebel, and D. Cremers, “Learning to Drive using Inverse Reinforcement Learning and Deep Q-Networks,” NIPS workshop on Deep Learning for Action and Interaction (2016).
- [7] 東洋信号通信社, “AIS航跡履歴データ,” http://www.toyoshingo.co.jp/site/portal/service/shipinfo/service/data_archive.html (2017年4月24日閲覧)
- [8] 南真紀, 庄司るり, “東京湾におけるAISデータを用いた衝突危険性の評価について,” 日本航海学会講演予稿集, 3巻, 1号 (2015).
- [9] 戸野 淳司, 有村 信夫, 丹羽 康之, 沼野 正義, 岡崎 忠胤, 劉 峭, “衝突予防援助装置の機能向上と支援効果の評価に関する研究,” 海上技術安全研究所報告, 8巻, 3号, (2008).
- [10] 山口太郎, 久保田周治, “ZigBee センサネットワークを用いた小型船舶事故防止システムの検討,” 電子情報通信学会東京支部学生会研究発表会, 講演番号88番, (2016)