

## システム情報工学研究科修士論文概要

年 度	平成 26 年度	学位名		修士( 工学 )
専 攻	知能機能システム	専攻	著者氏名	田丸 順基
指導教員氏名 安信 誠二				
論文題目				
予測行動安全度により事故の発生回数を抑制する Profit Sharing				
論文概要				
<p>本論文の目的は、危険を伴う未知環境下において事故の発生回数を抑制する安全な学習アルゴリズムの実現である。危険を伴う未知環境下において、ロボットは行動則を安全に自律獲得することが重要である。なぜなら、事故により故障するとタスクを遂行できなくなるだけでなく、修復のための莫大な出費やその後の作業の妨害になるためである。未知環境下では行動則の事前設計は不可能なため強化学習が有効である。なぜなら、強化学習は学習の際に環境のダイナミクスが不要であるためである。しかし、危険を伴う未知環境下では、多くの強化学習アルゴリズムは学習中に様々な事故を経験することが問題である。これは多くの強化学習アルゴリズムは行動結果の評価を基に行動則を更新するためである。その結果、学習中の事故が故障を引き起こしてしまう。よって、学習中の事故の発生回数を抑制する学習アルゴリズムが必要である。本論文では、学習アルゴリズムは強化学習の一つである Profit Sharing に着目する。これは合理的な行動則を学習しやすいという特徴を持つ。また、事故の発生回数を抑制するために、ロボット設計者が、ZMP など事故に関するセンサ情報(安全情報)を用いて事故を予測して考慮した行動則を設計していることに着目する。ロボット設計者はセンサ情報を構成する空間の2点間の距離が測定可能かつ安全領域と危険領域で構成される性質を利用している。センサの特性から行動による安全領域内のある安全情報の遷移ベクトルが危険領域に到達するまでの遷移回数(行動安全度)を予測することが出来る。つまり、ロボット設計者の予測知識を Profit Sharing に取り入れることで、経験する前に行動安全度を考慮した安全な学習が可能である。そこで、本論文では安全情報から得る予測行動安全度を考慮し、遷移先の状態と最後に出力した行動の行動評価値を毎ステップ予測更新する Profit Sharing を提案する。結果、シミュレーション実験により、従来の Profit Sharing よりも事故発生回数を大幅に抑制した学習が可能であり、従来の Profit Sharing とでは学習の収束が困難な環境でも学習の収束が可能であることを示した。提案手法により危険を伴う未知環境下において安全な行動則の自律獲得に有効であり、惑星や災害地探査ロボットなどの行動則の安全な自律獲得に応用可能であると考えられる。</p>				
審査日	平成 27 年 1 月 28 日			
審査員	(大学名 職名)	(学位)	(氏名)	
主査	筑波大学 教授	工学博士	安信 誠二	
副査	筑波大学 助教	博士(工学)	澁谷 長史	
副査	筑波大学 准教授	博士(工学)	相山 康道	