

学 位 論 文 題 名

ベイジアンネットによるエージェントの  
確率的知識表現に関する研究

学位論文内容の要旨

機械学習の一つとして知られる強化学習は、報酬という外界からの入力を手がかりに方策を最適化することで、対象となる環境に適応することを目的とする学習の枠組みである。環境との相互作用を通して強化学習エージェントが報酬を得る過程において、状態と行動という組のデータが生成されるが、経験強化型の強化学習アプローチの一つである利益共有法では、報酬を利用したデータ系列の重み値の更新により方策の学習を行う。

その一方で、マルチエージェント系のような不確実性を有する環境でエージェントが行動選択する場合、確率的推論システムによる行動の選択も有効となる。確率的推論を行うモデルの形式として、近年、グラフィカルモデルの一つとして知られるベイジアンネットに関する研究が盛んに行われている。ベイジアンネットは、対象から得られたデータの背景の同時確率構造を、非循環性有向グラフ的に表現する知識表現系モデルとして知られ、不確実性を有する領域における確率的な知識表現として広く利用されている。

ベイジアンネットを他のシステムに組み込んで利用する場合、そのネットワーク構造は事前情報から設計されることが多い。ここで、強化学習エージェントが試行錯誤の過程で得たデータ系列や報酬のように、ネットワークに対応するデータが得られる場合、実用的なデータマイニングへの実装例も数多く報告されているMDLやAIC等の情報理論的モデル選択を利用できる。情報理論的モデル選択によってサンプルデータから構築されたベイジアンネットは、強化学習エージェントが得たデータ系列と報酬をもとに表現された、エージェントの確率的知識であるとみなすことができ、これを利用することによって以下のような有効性を期待できる。

1. ネットワークの構造を利用して、強化学習エージェントの得た方策を改善可能
2. ネットワーク構造の特徴を解析することによって、ネットワークに対応する環境の情報を抽出可能

これらの考えをもとに本論文では、強化学習エージェントのデータ系列と報酬から、情報理論的モデル選択手法によって構築したベイジアンネットシステムを確率的知識として利用した方策改善法を提案する。提案する手法では、母体となる方策学習機構に利益共有法を適用し、ベイジアンネットの構造から正の報酬が得られる確率(事後確率)が最大となるエージェントの入力と行動の対を取り出して、これを

もとに重み値更新を行うことによって方策改善を図っている。

本論文では、計算機実験によって提案手法の有効性を検証し、提案手法において構築されるベイジアンネットによるエージェントの環境情報表現について議論している。

第1章では、序論として研究の背景および本研究の目的について述べている。

第2章では次章以降の準備として、強化学習とベイジアンネットについて概説している。はじめに、帰納学習や類推学習等いくつかの機械学習法について簡単に紹介し、それらの応用としてのエキスパートシステムについて概説した後、近年エージェントの学習の枠組として注目されつつあるリアクティブプランニングについて紹介している。引き続いて、提案システムにおける方策学習機構として利用する強化学習法について、環境同定型および経験強化型という強化学習の二つのアプローチにおける理論と現在の研究状況について述べている。次に、提案システムにおける方策改善機構の基盤となるベイジアンネットについての理論的背景や主な適用分野について説明し、本論文における同時確率分布のモデル化について示した後、ネットワークの構造決定に用いられる手法として、K-2 アルゴリズムと MDL 基準を用いた情報理論的モデル選択手法について紹介している。

第3章では、冒頭においてベイジアンネットと強化学習に関する従来研究について概説している。引き続いて、本論文で提案する、確率的知識表現を利用した強化学習エージェントの方策改善システムの枠組について、典型的なマルチエージェント問題であるエージェント追跡問題を例に説明した後、システムの各種設定について述べている。まず、ベイジアンネットにおけるノード、およびネットワークの構造学習に利用するサンプルデータについて定義し、次に、具体的な方策改善の手順を示している。また、同時確率分布の計算が不可能であるようなサンプルデータの取り扱いとして、ノードの追加的学習法を紹介した後、提案システムの特性についての議論を行っている。

第4章では、提案システムの特徴および有効性の検証のために行う計算機実験として取り上げる、エージェント追跡問題について紹介している。実験環境として、エージェントの数や環境の規模が異なる2次元平面を用意し、実験で使用するエージェントや環境に関する各種パラメータ、および構築するベイジアンネットの設定について紹介した後、具体的な実験方法について示している。

第5章では実験結果として、各実験設定において構築されたベイジアンネットを示し、ノード間の結合数、およびタスク達成に要したエージェントの行動数等を紹介した後、これらの結果をもとにベイジアンネットによる環境情報の表現能力、および方策改善システムの特徴や有効性について考察している。実験結果と考察から、本論文で提案した方策改善システムを利用することによって、強化学習エージェントの方策が改善されること、および、構築されたベイジアンネットが他のエージェントを含む環境についての確率的知識表現となっていることを示している。また、方策改善法の適用が困難であるような環境についての議論も行っている。

第6章では、結論として、本論文で得られた結果についてまとめ、今後の研究の方向性と課題を示している。

# 学位論文審査の要旨

主 査 教 授 栗 原 正 仁  
副 査 教 授 宮 腰 政 明  
副 査 教 授 工 藤 峰 一  
副 査 教 授 佐 藤 義 治

## 学 位 論 文 題 名

### ベイジアンネットによるエージェントの 確率的知識表現に関する研究

ベイジアンネットは、真偽の二値をとる命題を区間  $[0,1]$  の実数をとる確率変数に一般化し、それらの間の因果関係を条件付き確率によりモデル化してネットワーク表現したものであり、不確実性を有する環境における知識表現法の一つとして研究が進展してきている。特に近年、信念伝播アルゴリズムと呼ばれる高速な確率推論アルゴリズムが発見されてからはその実用性が一気に加速され、種々の応用への展開が図られつつある段階になってきた。しかし、個々の応用領域に応じて、いかなる確率変数や因果関係を抽出して知識を表現し、さらにその知識をどのように効果的に利用するかという点については、まだ未知な点が多く、さまざまな応用事例を通じてその知見を蓄積していく必要がある。

本論文は、そのような研究の流れの中で、応用領域を利益共有法による強化学習に基づくエージェントの行動決定の問題にしぼり、環境から得られる報酬、およびエージェントの入力-行動対の系列からなるサンプルデータによって構築されるベイジアンネットの確率推論と利益共有強化学習法の協調に基づいてエージェントの方策を改善するシステムアーキテクチャを開発し、標準的な例題によって実験的にその特徴と有効性を明らかにしている。その成果は次の2点に要約できる。

#### 1. ベイジアンネットによるエージェントの確率的知識表現法の開発

著者の開発したアーキテクチャは、エージェントから渡されるサンプルデータから学習アルゴリズムによってベイジアンネットを構築する部分と、そのベイジアンネットから確率推論によって得られる知識を教師信号的にエージェントに与える部分から成っている。前者においては、報酬および入力-行動対の系列から確率変数の値を重複なく抽出することによって確率モデルを構成し、サンプルデータの全体か

らその同時確率分布を求め、最小記述長 (MDL) 基準を用いたモデル選択と未観測データを補うノード追加法によってベイジアンネットを構築する手法を開発している。後者においては、ベイジアンネットの報酬ノードと確率的依存関係の強いノードに着目し、正の報酬が得られる確率が最大となる入力-行動対の重み更新に基づく方策改善法を導入している。独立したこの2つの部分から構成される類似のシステムは今のところ事例がなく、その独創性は評価できる。

## 2. 焼きなまし法による非線形最適化のヒューリスティックス

前項のシステムで最も計算量を要する部分はMDL基準の最小化である。これは、離散的でかつ非線形な最適化問題であり、一般にはNP困難な問題となっているが、著者は、その解法として焼きなまし法を採用して、その効果的な適用法を明らかにしている。焼きなまし法は熱力学的なモデルをシミュレートすることによって最適解を求めようとするアルゴリズムのパターンであり、特定の問題領域にそれを適用する際には、領域に応じたヒューリスティックな問題表現やパラメータ設定が必要となる。著者は、標準的な例題を通じて、そのヒューリスティックスを論じている。すなわち、データが表す分布のランダム性の指標である同時エントロピーやベイジアンネットの構造に依存して、いかに温度減少率を設定すれば、効率よく質の良い解が求められるかを実験的に明らかにしている。

これを要するに、著者はベイジアンネットによるエージェントの確率的知識表現法を考案し、それを利用して利益共有強化学習エージェントの独創的な方策改善システムアーキテクチャを開発したものであり、知能情報工学ならびにシステム情報工学の発展に対して貢献するところ大なるものがある。よって、著者は、北海道大学博士 (工学) の学位を授与される資格あるものと認める。