

平成21年6月1日現在

研究種目：若手研究（B）  
 研究期間：2007～2008  
 課題番号：19700138  
 研究課題名（和文）ベイジアンネット混合モデルを用いた環境適応型オンライン学習システム

研究課題名（英文）  
 An On-line Learning System Adapting to Environmental Changes with the Mixture of Bayesian Networks

研究代表者  
 北越 大輔（KITAKOSHI DAISUKE）  
 東京工業高等専門学校・情報工学科・講師  
 研究者番号：50378238

研究成果の概要：本研究では、機械学習の一つである強化学習と、確率モデルとして知られるベイジアンネットの混合モデルを用いた環境適応型オンライン学習システムを提案し、これを実装した移動エージェント（ロボット）の複雑・動的な環境での適応性能評価を目的とした。提案手法はノイズを含む環境や、環境の様々な変化に対して柔軟に適応可能であること、および指数型混合モデルに固有の特徴として、広範な環境情報を少数のベイジアンネットによって表現可能であることが実験的に示された。

## 交付額

(金額単位：円)

	直接経費	間接経費	合計
2007年度	2,200,000	0	2,200,000
2008年度	1,100,000	330,000	1,430,000
年度			
年度			
年度			
総計	3,300,000	330,000	3,630,000

研究分野：総合領域

科研費の分科・細目：情報学・知能情報学

キーワード：機械学習，強化学習，混合モデル，環境適応型学習システム，Khepera II

## 1. 研究開始当初の背景

現在までに、行為者（移動エージェント・ロボット）の行動選択指針（以降方策と呼ぶ）を最適化することを目的とした手法が数多く提案されてきている。これらの手法は様々な分野に適用され、一定の成果を得ているが、最適な方策の学習が困難な状況も少なからず存在する。

特に近年、実世界において一般的に想定される、

- ・入出力にノイズ等の不確実性が含まれる
- ・環境が動的に変化する

といった特徴を有する状況に対して柔軟に適応可能なシステムに関する研究が盛んになりつつあり、このような要求を十分に満たすシステムの開発が求められている。

ロボットやエージェント等の行為者を上述のような環境へ効率的に適応させることを目的とする学習手法としては強化学習が挙げられる。強化学習は、環境から与えられる報酬を手掛かりに試行錯誤的に最適な方策を獲得する機械学習アプローチの一つである。強化学習はその実装の容易さ等から、これまで多様な手法が提案され、様々な場面への適用が試みられてきたが、その適応性能、および適応に要する計算資源（処理時間）の双方の面で利用者を十分に満足させる成果は報告されていない。一方、不確実性を有する問題に対して確率的な意思決定を行うことは有効であることが知られているため、ベイジアンネットワーク等の確率モデルを用いた推論を利用することで、環境への適応を目指す研究も盛んになりつつある。しかしながら、準備した確率モデルをロボットの行動決定や複雑な環境への適応へ効率的に利用する方法についての研究は、いまだ不十分である。加えて、ロボットやエージェント、もしくは環境に対するどの要素を対象とすれば、上記目的を達成するための効果的なネットワークモデルを構築可能となるか、という課題についても十分な議論がなされているとは言い難い。

## 2. 研究の目的

1. で述べたような背景のもとで申請者は、不確実かつ動的な状況への適応に対して一定の成果が得られている強化学習とベイジアンネットワークを相補的に活用する枠組みを提案する。

この着想は、我々人間がある種の問題を解く際のプロセスから得られている。例えば我々人間が何らかの問題（例：パズルや迷路等）を初めて解こうとする場合、試行錯誤的な過程を経て解を得ることが少なくない。また、一旦解決した問題についてはより効果的な解法を獲得する可能性も高まる。ここでの人間の振舞は強化学習的であると言える。一方、いくつかの問題を経験した上で新たな問題に挑戦するような場合、過去の経験と現在の問題との類似性を考慮した上で、新たな問題の解決のために経験を活用することも可能である。

申請者は、上述した人間の経験的知識利用による環境変化への適応能力をエージェント（ロボット）へ付加するため、環境適応型オンライン学習システムを提案する。オンライン学習システムを実装したエージェントは強化学習によって現在の環境に対する適切な方策の獲得を目指しながら、過去にエージェントが得た知識を活用して、未知環境を含む多様な環境変化への柔軟な適応を目指す。ここで、エージェントが過去に獲得した個々の環境に関する知識はベイジアンネッ

トにより表現されるものとし、現在の環境に対する適切な確率モデルを表現するため、本研究ではこれら複数のベイジアンネットワークを構成要素とする混合モデルを利用する。申請者らはベイジアンネットワーク混合モデルとして、一般的に広く用いられる線形混合モデルとともに、新たに指数型混合モデルを導入した。提案するシステムの枠組を図1に示す。

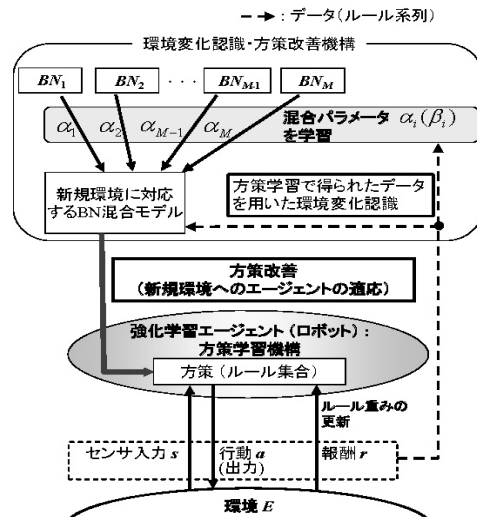


図1: 提案する学習システムの枠組

強化学習エージェントは現在の環境において適切な方策を獲得するため、入力（例：センサ情報）に対する適切な出力（行動）を、環境から与えられる報酬を手掛かりに学習していく。一方、学習の過程で得られるエージェントの入出力（ルール）と報酬に関するデータは、エージェントが現在存在している環境が、過去に経験したどの環境にどの程度類似しているかを判断するために用いられる（図中の混合パラメータ $\alpha_i$ は、現在エージェントの存在する環境を表現するため、各ベイジアンネットワーク（図中のBN）がどのように（どの程度）利用されているかの尺度を表す）。また、現在の環境が他の環境へと変化した際、エージェントは混合モデルを用いて環境が変化したことを認識する。変化の認識後、新規環境に対応する混合パラメータの調整が行われ、調整後の混合モデルで表わされる確率的知識をもとにエージェントの方策が更新される。確率的知識による方策の更新（方策改善）を行うことで、提案システムを実装したエージェントは環境変化への迅速かつ柔軟な適応を可能とする。

本研究の目的は、提案した上記のシステムに関して、次に示す各項目を明らかにすることである。

- (1) これまでの研究成果をもとに、提案する環境適応型オンライン学習システムの複雑、動的な環境に対する適応性能を評価すること。

- (2) 提案システムの仮想環境下での性能を評価するとともに、実世界の環境における性能、特徴についても移動ロボット (Khepera II) を利用して調査すること。
- (3) 提案システムの適応性能と、システムに導入した混合モデルとの関係について考察すること。
- (4) 混合モデルの環境表現能力とモデルの構成要素 (ベイジアンネット) との関連性について調査すること。

### 3. 研究の方法

- (1) 平成 19 年度の初頭より、移動エージェントの入出力に対するノイズの影響等を考慮した場合における提案システムの性能について評価するため、エージェントナビゲーション問題を適用例として計算機実験を実施した。調査項目の詳細は以下の通り。

- ① 提案する環境適応型オンライン学習システムのノイズに対する頑健性、および動的環境に対する適応能力を評価する。何種類かの正規ノイズを移動エージェントの入力へ加えた状態で、提案システムを実装したエージェントのタスク成功率や、環境切り替え後の新規環境に対する学習能力等について検討する。具体的には、障害物配置を変更した複数の環境を用意して一定時間ごとに環境を切り替え、変化後の環境への適応速度等をもとに考察を実施する。提案システムの比較対象として、提案システムで実装している強化学習法 (利益共有法) および利益共有法の改良版を用い、同様の設定で実験を行う。
- ② 2 種類のベイジアンネット混合モデル (線形混合, 指数型混合) の特徴、およびそれらを実装したシステムの適応性能を比較する。混合モデルを構成する個々のベイジアンネットの各環境に対する混合パラメータ比較を通して、二つの混合モデルの特徴、およびどちらの混合モデルが環境変化への適応に効果的であるか等について考察する。
- ③ 混合モデルの構成要素として適当なベイジアンネットについて検討する。計算機上で仮想的に構築した、ノイズを含まない環境で収集したデータから構築したネットワークと、一定のノイズが含まれる状況におけるデータをもとに構築したネットワークを準備し、それぞれを混合モデルの構成要素として実装したシステムの性能について比較す

る。

- (2) 提案システムの実環境での適応能力について評価するための準備として、実験用高性能計算機と移動ロボット (Khepera II) およびその周辺機器を購入。
- (3) 実環境での移動ロボットを用いたロボットナビゲーション問題を適用例とした実験を実施するため、環境の検討と整備を実施する。環境としては、ロボットに対する障害物および外壁としてアクリル材を、ロボットの目指すべきゴールとしてライトを用意する (ロボットは赤外線センサにより光を探知し、一定量以上の変化があった場合にゴール到達を認識すると設定)。
- (4) 計算機実験の結果をもとに、実環境への適用へ向けた提案システムの改良, 拡張に着手する。具体的には、以下のような修正を追加。

- ① 移動エージェントの入出力部をロボットの入力 (センサ情報) と出力 (モータへの指令) へ変更。
- ② 指数型混合モデルのモデル構成にかかる計算量抑制を目的としてプログラムを修正。
- (5) 平成 20 年度の初頭より、実世界における環境下での提案システムの性能評価を行うため、ロボットナビゲーション問題を取り上げて実験を実施する。調査項目の詳細は以下の通り。
- ① 提案システムの実環境下でのノイズ頑健性, 適応能力について評価する。障害物配置を変更した複数の環境を用意し、類似した環境間での変化, および類似性の少ない環境間での変化が生じる場合の双方について調査を実施する。計算機実験の場合と同様, 利益共有法とその改良版を用意して同一設定での実験を行い, 提案システムの結果と比較する (計算機実験の結果, 指数型混合モデルを実装した提案システムの方が線形混合を用いたシステムより優れた性能を示すことが確認されたため, 実世界における実験での提案システムでは指数型混合モデルのみを実装し, 従来手法との比較を実施する)。
- ② 指数型混合モデルを提案システムに実装する際に問題となり得る, 個々のベイジアンネットの混合処理にかかる計算量 (計算時間) の問題について調査し, 問題に対する解決策について議論する。
- ③ より実用的, かつ一般的な問題への適用へ向けた, 提案システムのさらなる改良について考察する。

#### 4. 研究成果

##### (1) 主な研究成果

- ① 計算機実験の結果、ノイズの含まれない“理想的な環境下”で収集されたデータから構築されたベイジアンネットの混合モデルを実装した環境適応型オンライン学習システムは、システムを実装したエージェントの存在する環境におけるノイズの有無に関わらず、多様な環境において効率的な振舞いを示すこと、動的な環境変化に対して柔軟な適応能力を示すことを確認した（表1：ゴール到達回数は8000試行中の到達回数を示し、行動選択回数はゴール到達した試行における行動選択回数の平均値である）。表1において、PS および PS' は通常の利益共有法とその改善法を表し、LM および EM は線形および指数型混合モデルを実装した提案システムの結果を表す。実験では4つの環境が1000試行ごとに切り替わり、それを2度繰り返した結果（計8000試行）について評価している。提案システムの性能は従来の利益共有法およびその改良版よりも優れていることが明らかとなった。加えて、提案システムに実装する二種類の混合モデルのうち、指数型混合モデル（EM）の方が線形混合モデル（LM）と比較してより広範な環境情報を表現可能であるため、環境変化への適応性能の面でも線形混合モデルよりも優れていることが示された。

表1：計算機実験の結果

	PS	PS'	LM	EM
ゴール到達総数	4166.3	7206.4	7402.4	7614.3
行動選択回数	195.1	143.4	136.2	130.8

- ② 混合モデルの構成要素として適切なベイジアンネットの特徴について検討した結果、個々のネットワークの特徴が際立っている（=ネットワーク間の特徴に差がある）場合、提案システムの適応性能がより高くなることを確認した。ここで、各ネットワーク間の特徴の差異を測るための基準は、確率モデルのそれぞれが表す同時確率間の Hellinger Distance と呼ばれる距離関数によって計測した。
- ③ 実世界における実験においても、上に示した計算機実験での結果と同様の傾向を示すことが確認された（表2：結果は6つの環境を10試行ごとに切り替えた、計60試行における値となっている）。環境を表現するベイジアンネット間の特徴の違いを上記①で使用した Hellinger Distance によ

表2：実環境下での実験結果

	CB1			CB2		
	EM	PS'	PS	EM	PS'	PS
ゴール到達総数	36.7	31.3	29.0	34.0	27.3	26.3
行動選択回数	239.8	309.4	322.7	269.8	331.6	332.6

て測定し、その平均値を環境変化の激しさで見なした。実験では、表2で示されるCB1における平均値の値よりもCB2における値が大きくなるような環境設定を用いた（すなわち、CB2の方がCB1と比較して、環境変化前後の環境の差異が大きい）。CB1、CB2の双方において提案システム（EM）が最大のゴール到達回数、最小の行動選択回数を示している。

- ④ 指数型混合モデルは線形混合モデルと比較して構造が複雑であるため、モデルの構成、および再構成に多くの時間を要する。一方で、線形混合モデルよりも少数の構成要素（ベイジアンネット）で、より広範な環境情報を表現可能である。混合モデルの再構成は、現在の構成要素数では対応しきれない環境が現れ、それに対応するための新たなネットワークを導入する際に必要となる。しかしながら、指数型混合モデルではモデル自体が持つ上記の特長のため、再構成の頻度そのものが非常に少なく抑えられる。従って、モデル再構成に伴う計算量の問題は（実験開始時のモデル構成にかかる計算を除き）実用上ほぼ起こり得ず、計算機実験、実世界における実験の双方においても同様の問題は確認されなかった。

##### (2) 国内外における位置づけ、インパクト等

- ① これまで、国内外においてロボットやエージェントの環境変化への適応を目指す研究は数多く行われてきている。これらの研究の多くは一定の成果を得ているものの、適用範囲が緩やかな環境変化に限定される、新たに適応すべき環境が現れるたびに、適応のための新たな情報を導入する必要があるため、多くの計算資源や計算時間を要する、等といった課題を抱えていた。提案する環境適応型オンライン学習システムでは、ベイジアンネット混合モデル（特に指数型混合モデル）を学習者の経験的知識として導入した結果、少数の構成要素のみを利用して、それらの混合パラメータを調整しながら非常に多様な環境への適応を可能とした。環境変化への適応能力と、適応に必要な計算資源・計算時間の抑制を両立可能であるという意味で、提

案システムは非常に独創的かつ実用的であるといえる。

- ② 提案する学習システムは、これまでに述べた通り、ノイズへの頑健性、環境変化への追従性を有する上、他の問題領域へも容易に適用可能であるため、従来の学習システムでは適用が困難な、ノイズや不確実性を含む広く一般的な環境においても良好な振舞を学習可能となることが期待される。加えて、システムで用いる混合モデルの構成要素であるベイジアンネット自体は、ネットワーク内の個々のノード間の依存関係を有向グラフ表現したグラフィカルモデルの一種であるため、利用者に対して視覚的にも理解しやすい情報を提供する。また、方策の学習機構として利用する強化学習も、直観的に理解しやすいシンプルなアルゴリズムであると考えられているため、提案システムの枠組み自体は広く一般に受け入れられ、普及へ向けた技術的障害も比較的少ないと予想される。

## 5. 主な発表論文等

〔学会発表〕（計1件）

Daisuke Kitakoshi, Hiroyuki Shioya, and Ryohei Nakano, “Noise Tolerance of an on-line Adaptive Learning System using a Mixture of Bayesian Networks”, Joint 4th International Conference on Soft Computing and Intelligent Systems & 9th International Symposium on Advanced Intelligent Systems, 2008年9月20日発表.

## 6. 研究組織

### (1) 研究代表者

北越大輔 (KITAKOSHI DAISUKE)

東京工業高等専門学校・情報工学科・講師

研究者番号：50378238