4M－4

構成的帰納学習と強化学習の統合による知的エージェントの学習効率の向上*

宮本 行庸† 上原 邦昭†

神戸大学 工学部 情報知能工学科‡

1. はじめに

学習能力を持った知的エージェントの実現方法として，強化学習 [1] が有力視されているが，従来の強化学習は最適解への収束の遅さが欠点として指摘されている．その要因として，学習に多くの事例が必要なこと，獲得した経験からの知識の抽出が困難なことの二点が挙げられている．学習の効率化のためには，報酬が得られる事例の特徴を新たに生成する必要があることが指摘されている．本研究は，事例から新たに特徴を生成する手法として，構成的帰納学習 [2] を強化学習に導入し，対象領域に適切な特徴を構成して学習効率の向上を図ることを目的としている．

2. 構成的帰納強化学習

本研究で提案する構成的帰納強化学習は，観測事例そのものからでは評価が困難な問題領域に対し，新たな特徴を逐次的に生成・選択して，従来の強化学習における収束の遅さを解消するものである．

事例は，観測状態，選択された行動，環境の変化に伴う報酬，頻度，遷移後の状態より構成される．また，目的とする概念を報酬により五つのクラス (成功，正，無報酬，負，失敗) に大別する．各クラスは，それぞれ特徴階層と呼ばれる DNF 形式で表現された特徴の集合を持ち，クラスに見合った報酬を得られる特徴を明確に表現する集合となっている．本アルゴリズムは，強化学習に基づいて，観測，検索，行動決定，実行・報酬獲得，特徴生成の順に行われる．

検索過程では，観測状態が報酬の高い順に特徴階層に照合され，報酬の予測値が立てられる．報酬の予測値が実際に得られた報酬と一致すれば，予測が正しいとみなし，予測された事例の頻度を増加する．しかしながら，報酬の予測値は実際に得られる報酬と常に一致するとは限らない．一致しない場合は，両者を適切に識別できる他の特徴が存在すると仮定して，両者間で異なる属性を利用して新たな特徴を生成するようにしている．

この手続きにより，従来の強化学習では各事例をそれぞれ一つの特徴とみなしていたのと比較して，報酬別の分類が効率化されることになる．具体的には，予測された遷移後の状態と実際の遷移状態との相違属性を軸に他の属性との連言をとり，これを新たな特徴としている．以下，全ての相違属性に対して同様の操作を行った後，生成された特徴の集合を新たに特徴階層に加えている．

しかしながら，加えられた特徴階層の特徴数は非常に莫大なものとなるため，適切な特徴を選択する必要がある．選択の基準としては，報酬と頻度の大きさの二点である．この条件に基づいて試行を繰り返す毎に適切な特徴を選択し，各クラスの特徴を洗練するようにしている．

3. 実験

本実験は，図 1 のような閉じた平面を想定し，この環境内に放たれた知的エージェントの挙動を観測する [3] ．学習対象とする概念は，報酬の得られる特徴の DNF 表現で，事例数に直すと約 16 万通りにも及ぶ．報酬の得られる事例を全て残しておくには非常に無駄が多く，特徴生成による絞り込みが必要とされる．獲得したい特徴は報酬別に 4 種類× 4 方向の 16 種類あり，その中には，特定の属性に依存しないものや，複数の属性の連言で表される特徴も含まれており，従来

*Scaling up Learning Rate by Integrating Constructive Induction into Reinforcement Learning

†Yukinobu Miyamoto and Kuniaki Uehara

‡Dept. of Computer & Systems Engineering, Faculty of Engineering, Kobe University

1-1 Rokkodai, Nada, Kobe 657, Japan

表 1: 生成された特徴

報酬	行動	頻度	状態	報酬	行動	頻度	状態
1.0	2	477	*1***3**	−0.5	2	4109	*1***2**
1.0	1	481	1***3***	−0.5	2	13068	*1*******
1.0	8	1100	***1***3	−0.5	1	1767	1***2***
0.5	4	887	**2***0*	−0.5	1	13138	1*******
0.5	4	81	**2*****	−0.5	4	7	**1***2*
0.5	2	141	*2***3**	−0.5	4	1	**1*****
0.5	2	1699	*2*******	−1.0	1	10	3***1***
0.5	1	890	2***0***	−1.0	1	292	3*******
0.5	1	729	2*******	−1.0	2	217	*3***2**
0.5	8	507	***2***0	−1.0	2	10	*3*******
0.5	8	893	***2****	−1.0	8	7	***3***2
				−1.0	8	2	***3****

の単純な強化学習のみでの学習は困難な問題となっている．試行中に生成されたクラス別特徴階層を表 1に示す．表 1の状態は，エージェントの正面より時計回りに近距離，遠距離でそれぞれ 4 属性ずつ，計 8 属性で表現されている．表 1より，成功報酬クラスの特徴階層には不要な特徴が含まれていない，単一の属性で表現できるクラスには不要な特徴が含まれ易いという事実が読みとれる．第一の事実より，成功報酬を得られる特徴は検索の優先度が高く，最も頻繁に生成・消去が行われるクラス内にあるので，試行を繰り返すことによって充分な洗練化がなされ，目的概念に必要な特徴を得ることができていることが分かる．第二の事実は，単一属性による特徴で表現可能なクラスには，複数の属性による特徴も含まれることがあり，検索の際には複数の属性による特徴を優先したために起こったと考えられる．この問題の対策としては，予測に失敗した特徴の信頼度を大幅に下げることが挙げられる．生成される特徴は最初に遭遇した事例に依存しているが，報酬獲得に必要な属性の DNF 表現を適切に抽出し，事例の再描写に成功している．

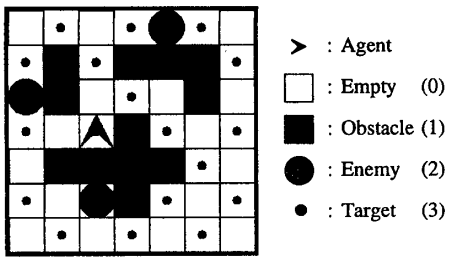

図 1: 学習環境

また，試行中の特徴数の推移 (図 2) より，学習中期以降は新たな特徴が生成されていない，生成された特徴数 23 は獲得すべき特徴数 16 に近いという事実が読みとることができる．第一の事実より，学習は初期段階で完了していることが分かる．初期段階で獲得した特徴には不要なものは含まれておらず，以降では，それまでに生成された特徴についての優先順位の変更のみが行なわれていると考えられる．第二の事実より，観測状態を生成された特徴表現に変換する際に，生成された特徴のみを用いると，概念の再描写に関して極端に冗長な表現にならないことが分かる．以上より，一定試行回数後は新たな特徴が生成されておらず，生成された特徴が後に観測される同じ特徴を有する異った事例をも良く反映しているため，学習に必要な事例数を大幅に削減できていることが分かる．

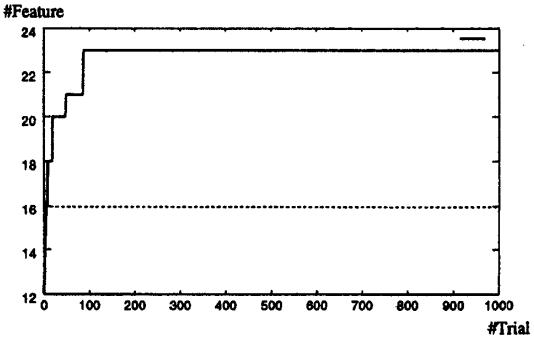

図 2: 特徴数の推移

4. おわりに

本稿では，強化学習に構成的帰納学習を導入し，従来の強化学習の問題の解消を試みた．実験では，学習に必要な事例数が大幅に削減され，学習効率の向上という目的を達成している．また，得られた特徴が少数へと絞り込まれ，特定の報酬が得られる状態を DNF 表現で簡潔に抽出することに成功している．

参考文献

[1] 畝見 達夫, "強化学習," 人工知能学会誌, Vol.9, No.4, pp.830-836 (1994).

[2] 滝 寛和, "構成的帰納学習とバイアス," 人工知能学会誌, Vol.9, No.6, pp.818-822 (1994).

[3] R. Maclin and J. W. Shavlik, "Incorporationg Advice into Agents that Learn from Reinforcements," *Proc. of the 12th National Conference on Artificial Intelligence*, Vol.1, pp.694-699 (1994).