知能と複雑系 110－6
(1998. 1. 22)

認識の学習を行う認知行動システム

上野 敦志　武田 英明　西田 豊明
奈良先端科学技術大学院大学 情報科学研究科
〒630-01 奈良県生駒市高山町8916-5
TEL: 0743-72-5263
E-mail: ueno@is.aist-nara.ac.jp

あらまし　本稿では，状況認識と行動規則を同時に学習する認知行動システムを提案する．実世界中の認知行動システムでは，環境中の莫大な情報からどのようにして有用な情報を抽出するのかが大きな問題となっている．これは“フレーム問題”と呼ばれる問題である．我々は，フレーム問題を現実的に解決するためには，認識の学習が不可欠であると考える．本稿で提案するシステムは，環境からの報酬に基づいて，連続的な知覚入力空間上で“状況”を抽出し，それを動的に維持する．この状況は，経験的に得られた記号とみなすことができる．このようにして，本システムは，動的な環境中で，知覚入力空間の動的な分節を行うことができる．本システムの有効性をコンピュータシミュレーションによって示す．

キーワード　記号化，分節，フレーム問題，強化学習，状況認識

Learning of the Way of Abstraction in Cognitive Agents

Atsushi Ueno, Hideaki Takeda, Toyoaki Nishida
Graduate School of Information Science,
Nara Institute of Science and Technology
8916-5, Takayama, Ikoma, Nara 630-01, JAPAN
TEL: 0743-72-5263, E-mail: ueno@is.aist-nara.ac.jp

Abstract This paper describes a method for a cognitive agent to learn the way of abstraction and the policy of behavior selection simultaneously. The main problem of agents in the real environment is how to abstract useful information from a large amount of information in the environment. This is called "the frame problem". We consider that learning of the way of abstraction is a key function for solving the frame problem practically. Our developed system extracts "situations" and maintains them dynamically in the continuous state space on the basis of rewards from the environment. This situation can be regarded as empirically obtained symbol. In this way, the system learns the way of abstraction in a dynamic environment. The results of computer simulations are given.

key words symbolization, articulation, frame problem, reinforcement learning, situation classification

1 はじめに

従来、実世界中の人工的な認知行動システム（ロボット）の多くは、工場などのそのために管理された環境中で働いてきた。実世界は、非常に大きな情報量を持っており、さらに動的に変化する環境である。そこで実世界中の認知行動システムには、大量の情報を処理する機能と素早く反応する機能が求められる。従来のロボットは、あらかじめ設計者によって定められた必要な情報のみを環境中から取り出し、処理することによって、この二つの機能を両立させている。管理された環境中では、この方法でロボットをうまく働かすことができた。

近年になって、自然環境や人間の生活環境などの、ロボットのために管理されていない環境中で、ロボットが必要とされるようになってきている。このような環境では、特にタスクが複雑な場合には、ロボットが直面するであろうすべての場面をあらかじめ予測することが非常に難しく、ロボットは、設計者が予測しなかったような場面にしばしば遭遇する。このような不慣れな場面では、ロボットは大量の情報のうちのどの部分に注目したら良いのかが分からず、適切な行動を決定することができない。この問題はフレーム問題と呼ばれる問題で、情報処理の時間の爆発、記述の量の爆発、内部モデルと現実との乖離などの様々な弊害をもたらす。

2 フレーム問題

認知行動システムは、外界の膨大な情報から現在のタスクに必要な情報だけを（枠で囲って）抽出して処理することによって、限定された情報処理能力でタスクを遂行することを可能にしている。しかし、この枠を適切に規定することは、非常に困難な問題である。なぜなら、実世界には非常に多くの事象が複雑に絡み合って存在しており、必要ないと切り捨てたことから、カオティックに重大な影響が生じることがあるからである。これがフレーム問題と呼ばれる問題である。

工場で働くロボットや自動販売機などの単純な作業を繰り返す認知行動システムでは、外界の非常に限られた情報だけを認識するように設計されている。工場の生産ラインではいつも同じ物が流れてくるので、乏しいセンシング能力でも決められた作業をこなすことは可能であるし、自動販売機はコインに似た外形の物しか入らないように設計されているので、コインの外形と重量などを簡単に調べるだけで、販売作業を行うことができる。しかし、工場のロボットは、設計時に予想していなかったような物が生産ライン上を流れてくると、正常に生産を続けることが困難であるし、自動販売機は、外国のコインや変造コインを入れられると、誤って商品を渡してしまうこともある。これはフレーム問題の現れの一つである。

人工知能システムでは、単純な作業の繰り返しではなくて、もう少し自由度が高く複雑なタスクを遂行することを目指している。そのために、多くの人工知能システムでは、認識したことを内部で記号を用いて表現して、その記号表現上での推論によって行動を決定する。そのための記号体系は、あらかじめ設計者によって定義されるが、複雑なタスクにおいて記号を適切に定義することは非常に困難である。通常、システムがタスクの実行に失敗する事態をなるべく避けるために、記号は冗長に定義される。その結果、記述の量が爆発し情報処理の時間が爆発して、考えてばかりでちっとも動かないロボットになってしまう。これもフレーム問題の現れの一つである。

システムの、そしてその設計者の情報処理能力は、実世界と比べるとはるかに限定されたものである。従って、フレーム問題は本質的には解決不能である。人間もまた、実世界の複雑性に比べると乏しい情報処理能力しか持っていないので、フレーム問題を完全に解くことはできない。しかし、フレーム問題の解決不能性にも関わらず、日常生活においては、人間はフレーム問題にほとんど悩まされずに行動しているように思われる。それは、日常的に環境中の膨大な情報のごく一部だけを枠で囲って、その枠の中の情報だけに注目する習慣を身につけているからである。松原らは、これを「フレーム問題の現実的な解決」と名付けている [4]。フレーム問題の現実的な解決の枠を構成しているのは、様々なヒューリスティックスの集まりである。日常の様々な場面でフレーム問題に直面し、それに対処するためのヒューリスティックスを経験的に獲得すること、そして、環境との相互作用を通してそれらのヒューリスティックスを適切に適用することによって、フレーム問題を現実的に解決することが可能になると思われる。

Brooks らによって提唱された行動に基づく知能のアプローチでは、環境との相互作用をなす単純な作業の重ね合わせで複雑なタスクを実現する [1]。これは、たくさんのヒューリスティックスをロボットに持たせることに他ならない。このアプローチでは、環境との相互作用を重視して内部で複雑な推論を行わないので、刺激に対する応答を非常に速くすることができる。しかし、全体として整合性がとれるように個々のヒューリスティックスを設計することは、複雑なタスクの場合には、非常に困難である。また、自動販売機の例で示したように、単純な作業においてもフレーム問題を完全に解決することはできない。従って、このアプローチでは、フレーム問題によって小さな過ちが起こることを前提にして、それを環境中から発見し、大きく広がる前に素早く対処することを目指しているといえる。ヒューリスティックスの獲得は行わないので、同じ場面においては永久に同じ過ちを犯すことになる。

フレーム問題の現実的な解決のためには、環境との相互作用を重視するだけではなく、その相互作用の中でフレーム問題に直面し、適切な情報の枠を経験的に獲得する機能

が必要である。記号処理システムの場合には、現在のタスクに必要な情報だけを内部に記述できるように学習すれば良い。そのために一つには、固定された記号体系の上で、関係ないことを記述しないように学習する手法が考えられる。しかし、固定された記号体系の上で情報処理を行うためには、細かい区別が必要な状況のためにあらかじめ十分に記号を用意しておかなければならないので、記号体系は通常非常に冗長で自由度の大きいものとなる。そのために、関係ないことを無視するために必要な知識だけでも膨大になるし、DennettのロボットR_2D_1のように、無視する作業だけで手一杯になってしまう[2]。これもフレーム問題の現れの一つに他ならない。

記述の量を減らすためには、もう一つには、記号体系の自由度を小さくする手法が考えられる。すなわち、現在のタスクの遂行に必要なことだけを十分コンパクトに書けるような記号を経験的に獲得するアプローチである。これは「抽象化の学習」、または「認識の学習」とみなせる。このアプローチにおいては、情報処理に用いる記号は、ロボット自身がタスクに対する有用性に基づいて環境中から抽象して作り出す。環境の変化などで記号が不適切になった時には、適切になるように修正する。この場合には、現在のタスクに関係ない事象は記述することができないので、記述の量の爆発に悩まされることはない。情報処理の時間に関しては、認識の学習を行う場合には、その分だけシステムの計算量が増える。しかし認識の学習では、各瞬間には情報の枠を少し修正するだけなので、計算量をほぼ一定に少なく抑えることができる。一方、関係ない事象を無視するための推論は、環境によっては際限なく続くことになる。このアプローチでは、無視するための計算が必要ないので、情報処理の時間も抑えられることが期待できる。

3 認識の学習を行う認知行動システム

3.1 認識の学習を行う認知行動システム

人工知能の領域で最も一般的に用いられている抽象化は記号化である。記号表現では、知識をまとまった塊として扱うことができるので、「抽象化された表象に意味付けしやすい」、「情報処理の過程がわかりやすい」、「知識の再利用がしやすい」などの利点が得られる。

記号処理に基づく認知行動システムでは、専ら、ルールの学習や論理に基づく学習などの固定された記号体系の上での記号間の関係の学習が行われてきた。記号の学習（すなわち認識の学習）は記号間の関係の学習よりもはるかにゆっくりと進むものなので、近似的に記号体系を固定することができる。その方が効率の面では優れている。しかし、2節で述べたように、実世界中の複雑なタスクに対して記号体系を前もって適切に定めることは、一種のフレーム問題であり、非常に困難である。認知行動システムにおいて認識の学習を行うことによって、不慣れな環境に徐々に慣れてフレーム問題に悩まされなくなることが期待できる。

ここでいう認識とは、通常行われているパターン認識などのように人間の持つ認識をコンピュータ上で再現しようとするものではなくて、システム自身の必要性に基づいて行われる認識である。従来、人工知能の分野で研究されてきた認識（分類）の学習では、分類の結果を利用する人間によって、カテゴリに意味が与えられていた。従って、システム内部だけでは認識系が完結せず、カテゴリに意味を与える人間を含めて認識系が構成されていると考えられる。一方、認知行動システムで行う認識の学習では、システム全体が環境中でうまく行動できるという基準で、カテゴリに意味を持たせることができる。この場合には、うまく行動できるようなカテゴリを構成するための指標をシステムに与えてやる必要があるが、人間が直接カテゴリに意味を与えるのと比べると、はるかに自律的なシステムとなる。

3.2 認識の学習と認識された表象の上での学習の並列実行

認知行動システム中で、認識の学習と認識された表象の上での学習を並列的に実行することを考える。具体的には、図1に示すようなシステムを考える。環境とのフィードバックループとしては、環境からの知覚入力を知覚入力空間上で分節して記号空間にマッピングし、記号空間上で推論を行って行動を出力する。学習としては、行動出力とそれによる環境の変化に基づいて、知覚入力空間上での認識の学習と記号空間上での記号間の関係の学習を行う。認識の学習としては、タスク遂行の必要性に応じて、記号に対応する知覚入力空間中の領域を絶えずゆっくりと調整し続けながら、さらに大きな変更が必要な時には、記号の生成や削除を行う。

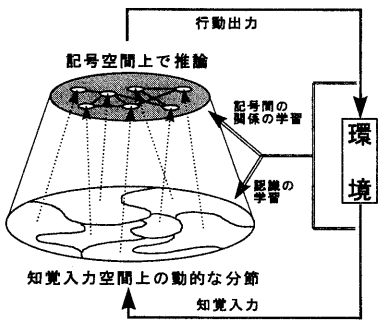


図 1: 認識の学習と記号間の関係の学習の並列実行

このように、認識の学習と認識された表象の上での学習を、オンラインで（すなわち、タスクを遂行しながら）並列して行うことによって、環境とタスクにとって必要十分な記号体系を獲得することが期待できる。また、未知の環境や環境の変化に柔軟に適応することが可能になる。これは、2節で述べたように、フレーム問題を現実的に解決するため

に重要な機能である。一方、認識の学習を常に行い続けることは、効率の面では短所となる。また、タスクをうまく遂行できるような分節を行うための指標を定めなければならないという新たな問題が生じる。これらは、フレーム問題を現実的に解決するためには、積極的に取り組まなければならない問題であると考えることができる。

本稿では、フレーム問題の現実的な解決を目指して、認識の学習と認識された表象の上での学習を並列して行う認知行動システムを提案する。

3.3 状態認識の学習と行動の強化学習の並列実行

認識の学習と認識された表象の上での学習をともに行う認知行動システムは、機械学習の一種、強化学習の領域でわずかに研究が行われている。強化学習とは、「学習者が環境中で行動を起こし、それに対して環境から与えられる「報酬」に基づいて、徐々に環境に適した行動パターンを獲得していく学習」のことであり、通常は離散化された状態表現の上での行動規則の学習を意味する。従って、記号間の関係の学習に相当する。そして、強化学習の前処理としての知覚入力の離散化の学習が、認識の学習に相当する。知覚入力は状態表現に離散化されるので、この学習は「状態認識の学習」である。

強化学習では、記号として「状態」を表す記号しか用いないので、認識の学習と組み合わせるのが容易である。また、タスクを「報酬」という形で端的に表現していて、この報酬を認識の学習のための指標として用いることができるという点でも都合が良い。そのため、本稿で提案するシステムでも、記号間の関係の学習として、強化学習を用いる[1]。

強化学習システムと組み合わせる認識の学習としては、フレーム問題を現実的に解決するという観点から考えると、連続入力から、オンラインで、報酬の類似度に基づいて学習を行うことが望ましい。連続入力が可能であるという性質は、実世界からの知覚入力のためには欠かせない。オンラインで学習を行うという性質は、現在のタスクへの特化と環境の変化に対する適応性のために欠かせない。また、強化学習においてはタスクは報酬の中に表現されているので、知覚入力空間の分節のための指標として報酬を用いることは、タスクにとって適切で抽象度の高い抽象化のために望ましいといえる。強化学習のための状態認識の学習法は数多く存在するが、連続入力から、オンラインで、離散的な状態表現（すなわち記号表現）を、報酬に基づいて学習する手法は、筆者らの知る限りでは、石黒らのEOP（empirically obtained perceiver）を用いる手法[3]だけである。しかし、この手法では、タスクを遂行しながら状態を分割することは想定しているが、一度分割した境界線は固定してしまうので、柔軟性にはやや欠けるといえる。本稿で提案するシステムでは、タスクを遂行しながら状態を切り出し、切り出した後も状態の形状を調整し続けるような認識の学習を行う。この柔軟性は、不慣れな環境に徐々に慣れてフレーム問題に悩まされないようになるためには非常に重要である。

[1] しかし、強化学習システムは、記号として「状態」しか用いることができないので、記号処理システムとしては非常に単純なものである。将来的には、さらに複雑な記号処理システムと認識の学習を組み合わせる方向に拡張していく必要がある。

4 システムの概要

この節では、認識の学習と記号間の関係の学習を共に行う認知行動システムとして、状況遷移ネットワークシステム（Situation Transition Network System、STNS）を提案し、その概要を説明する。STNSは、図2に示すように、状況認識器と、状況遷移ネットワークと、複数の行動モジュールからなる。一回の動作では、センサから与えられる連続的な知覚入力から状況認識器を用いて現在の状況を認識し、状況遷移ネットワーク上で部分的プランニングを行って、一つの行動モジュールを作動させる。

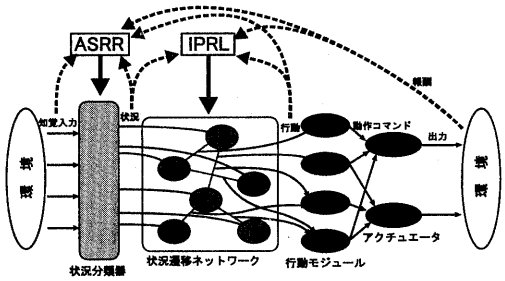


図 2: システムの構成

STNSは、環境中で何か評価し得る出来事が起こった時に、環境から報酬を得る。報酬というのは、一般に強化学習で用いられている報酬と同様に、数値で与えられる評価のことである。STNSは、この報酬に基づいて、状況認識の学習と行動規則の学習の両方を行う。状況認識の学習のために「Adaptive Situation Recognition based on Rewards（ASRR）」を、行動規則の強化学習のために「Interleave Planning-based Reinforcement Learning（IPRL）」を提案する。

ASRRでは、知覚入力、報酬、および行動の類似度に基づいて同一とみなし得る領域、すなわち「状況」[2]を連続的な知覚入力空間（または単に「入力空間」と呼ぶ）から切り出し、その後の経験に基づいて形状を調整しながら維持する。この状況表現では、固有の意味を持った抽象度の高い領域を表すことができる。

IPRLでは、状況表現の上で行動規則の強化学習を行う。具体的には、状況間の遷移確率と各遷移で得られる報酬の期待値を最尤推定してワールドモデルを構成し、その上で

[2] 一般的な強化学習における「状態」に相当する。

前向きの部分的プランニングを行うことによって行動を決定する。学習としては、ワールドモデルの最尤推定しか行わないので、状況表現が変化しても、行動規則を素早く収束させることができる。

STNS では、 ASRR と IPRL を組み合わせて、状況認識と行動規則を、タスクを実行しながら同時に学習する。そのため、状況の形状が適切であれば行動はタスクを成功させるし、行動がタスクを成功させ続ければ状況の形状が適切に維持される。この状況と行動の相互依存関係によって、タスクに特化しながら環境やタスクの変化に柔軟に適応できる認知行動システムが実現できる。また、このオンライン学習によって、入力空間中の、タスクを実行する上で頻繁に経験する領域について、特に詳しく学習することができるので、学習が効率的に進む。

ASRR と IPRL における学習は、図 2に示すように、システムがタスクを実行しながら観測した知覚入力、認識された状況、選択された行動、得られた報酬をデータとして行われる。このデータは、一定期間、履歴データベースに蓄えられて、その後、消去される。そのため、環境やシステムの内部表現が変化した場合にも、古過ぎるデータに惑わされることなく柔軟に対応することが期待できる。

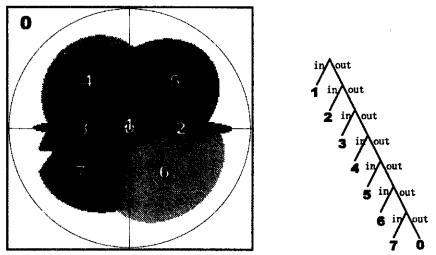
続いて、 ASRR について説明する。頁数の関係上、本稿では、 IPRL に関する説明は割愛する。詳細な説明は文献 [6]、簡単な説明は文献 [5] を参照されたい。

5 状況の認識とその学習

STNS では、 ASRR を用いて、知覚入力空間中の特定の領域、すなわち「状況」に相当する記号を学習する。 ASRR において、各状況の意味は行動の結果の類似性によって定義される。類似性の基準は環境から与えられる報酬とする。すなわち、一つには、ある特定の行動によって大きな報酬が得られる領域を状況として切り出す。また、もう一つには、ある特定の行動によってそのような報酬の見込みの高い既存の状況に遷移する領域を状況として切り出す。切り出した後も、状況と行動の相互依存関係によって、状況の形状を常に調整し続ける。そのため、状況を、データが不十分な早期の段階で切り出し、データが増えるにつれてタスクに適応させることができる。その結果、学習が速く進むことが期待できる。また、状況を環境の変化に適応させることができる。このように行動の結果の類似性に注目することによって、エージェント自身の環境中での経験のみに基づいた記号の定義が可能になる。このような手法によって、抽象度が高い状況を切り出すことができる。強化学習の状態表現の学習法として状態の形状の調整を行う手法は他に見当たらず、 ASRR が斬新な手法であるといえる。

5.1 状況の認識

ASRR では、入力空間は図 3a のように分割される。状況の認識には、図 3b に示す判別木を用いる。判別木の各ノードにおいては、現在の知覚入力が、入力空間中の各ノードが対応する領域に入るか、入らないかで判別を行う。各状況は重なりあっていて、上にある状況から順に判別が行われる。状況として切り出されていない余白部分の領域は、状況 0 と名付ける。



a. 入力空間（2 次元入力の場合）　b. 判別木

図 3: 入力空間と判別木（0 ～ 7 は状況を示す）

ASRR においては、状況 0 以外の各状況が固有の意味を持っている。それは、「その状況から特定の行動（条件行動と呼ぶ）をすると、特定の結果が得られる」という意味である。特定の結果には、「大きな正の報酬が得られる」という結果と、「他の特定の状況に遷移する」という結果の二種類がある。前者の結果を意味として持つ状況を R 状況（Reward の R）、後者の結果を意味として持つ状況を T 状況（Transition の T) と名付ける。

状況 0 以外の状況の形状は、その状況の意味に対して定められる正事例と負事例によって決定される。正事例とは、履歴データベース中の全データのうち、条件行動をして特定の結果が得られたデータのことである。負事例とは、その状況に分類されたデータのうち、条件行動をして特定の結果が得られなかったデータのことである。

状況の認識は、図 4に示すように、巨視的な認識と微視的な認識を組み合わせて行う。巨視的な認識では、入力空間中の正事例集団からのマハラノビス距離がある値以下の領域、すなわち超楕円体の内部を状況とみなす。微視的な認識では NN 識別法（nearest neighbor classification）を用いる。ただし、 NN 識別法で通常用いられるユークリッド距離の代わりに、正事例集団の標準偏差で標準化した距離を用いる。

状況に入るかどうか判別する際には、まず現在の知覚入力が超楕円体の内部に入るかどうかを調べ、内部に入る場合は NN 識別法によりその状況に入るかどうかを調べる。超楕円体は簡単なパラメータだけで表現されるので、巨視的な認識法では認識や学習のための計算が非常に速い。しかし、境界面が整い過ぎているので、図 4a に示すように、

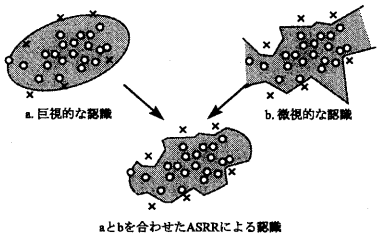


図 4: ASRR による認識 (○は正事例、×は負事例)

誤分類を避けることができない。そこで、微視的な認識法を組み合わせる必要がある。

NN 識別法による微視的な認識では、非常に精密に判別することができる。また、学習は、事例の位置を記憶するだけなので、非常に速い。一方、記憶容量と認識のための計算時間がかなり大きいということが欠点である。ASRR では、NN 識別法を用いる前に巨視的な認識で足切りを行うことによって、計算時間を節約している。また、負事例が不足すると、図 4b に示すように、NN 識別法だけではあちこちに破れが生じる。巨視的な認識を組み合わせることによって、図 4c に示すように、この破れを補うことができる。

ASRR では、多くの事例（特に正事例）によって状況の形状を維持している。これらの事例は、システムが報酬を集める過程で頻繁に経験されたものである。そのため、ASRR における状況は、報酬に到達するために通過する経路としての役割を持つ。

5.2 状況認識の学習

ASRR では状況の抽出と動的な維持によって、状況認識の学習を行う。ASRR における状況抽出の条件は、次の二つである。

1. **R 状況** 全入力空間中で、ある特定の行動によって特定の大きな（r^R_{min} 以上）報酬が得られたデータが、N^R_{init} 個以上存在する。

   T 状況 状況 0 の中で、ある特定の行動によって状況 0 以外の特定の状況に遷移し、大きな（r^R_{min} 以上）報酬が得られなかったデータが、N^T_{init} 個以上存在する。

2. 既存の状況の中に、同じ意味を持つ状況が存在しない。

上の二つの条件を満たした領域は状況として抽出される。抽出する際に、抽出条件中の特定の行動を条件行動にして、特定の行動の結果と共に、新たな状況の固有の意味として定められる。そして、その意味にしたがって、履歴データベース中に含まれる正事例となり得るデータを、すべてその状況の正事例集団に集める。負事例は、抽出時には存在しない。

ASRR では、入力空間全体が状況 0 であるところから学習を始めて、ある行動で報酬に到達する R 状況 A、ある行動で R 状況 A に到達する T 状況 B …というように、次々に状況を抽出し、入力空間を分割しながら学習が進む。

ASRR では、状況を動的に維持するために、次のような操作を行っている[3]。

1. 新たなデータが得られる度に、正事例集団、負事例集団を更新する。
2. 正事例集団から各正事例までのマハラノビス距離を定期的に調べて、最も遠い正事例までの距離を超楕円体の境界とする。
3. 報酬に近い状況が上になるように、定期的に状況の重なり方を更新する。
4. 状況の意味が不適切である場合に、意味を切り替える。
5. 維持できない状況、不要な状況を削除する。

6 実験

STNS の性能を調べるために、コンピュータシミュレーションで、ナビゲーションの実験を行った。頁数の関係上、本稿では、STNS が獲得した状況認識の様子を簡単に示すにとどめる。学習の過程やシステムの性能などの詳細は、文献 [6] を参照されたい。

6.1 2 次元入力のナビゲーション

まずはじめに、簡単な 2 次元入力のナビゲーション問題の結果を示す。問題の設定を図 5に示す。

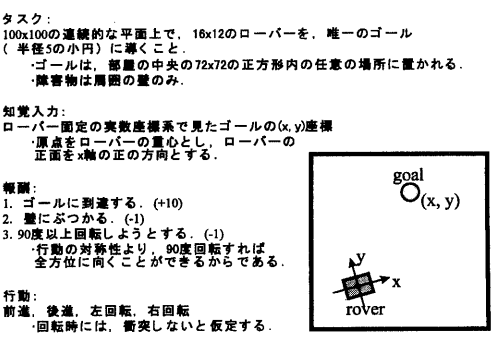


図 5: 2 次元入力のナビゲーション問題

このナビゲーション問題の政策を学習する STNS をつくり、シミュレーションで実験を行った。ゴールに到達した時は、任意の位置にゴールを、ゴール以外の位置にローバーを置き直して学習を続けた。学習は、平均的には、累積行動数 1200 回程度で収束した。典型的な収束後の入力空間を図 6a に、理想的な入力空間を図 6b に示す。真ん中の小円がゴールに到達した領域である。この図から、ほぼ理想的な状況認識と行動規則が獲得できているといえる。

[3]詳しくは、文献 [6] を参照のこと。

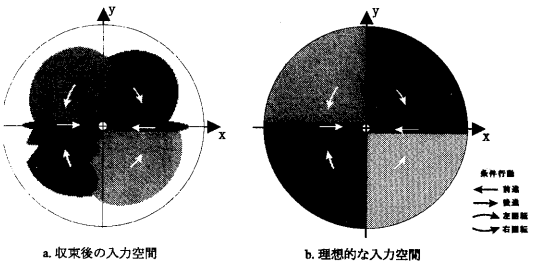


図 6: 収束後の入力空間と理想的な入力空間

次に、環境の変化に対する適応性を調べた二つの実験の結果を示す。上の実験で学習が収束したSTNS（図6aに示したもの）を用いて、環境を変化させて再学習する実験を行った。環境の変化としては、センサの取り付け角度がずれてしまう場合と、左車輪の回転速度が落ちてしまう場合を想定した。図7に、センサ取り付け角度が15°ずれた後に再学習された入力空間の例を、図8に、左車輪の回転速度がもとから15%低下した後に再学習された入力空間の例を示す。この図から、環境の変化に適応して、ほぼ理想的な状況認識と行動規則が獲得できているといえる。

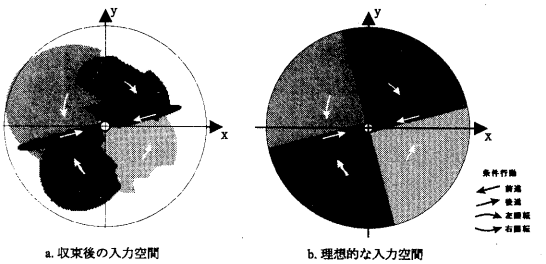


図 7: センサ取り付け角度が15°ずれた場合

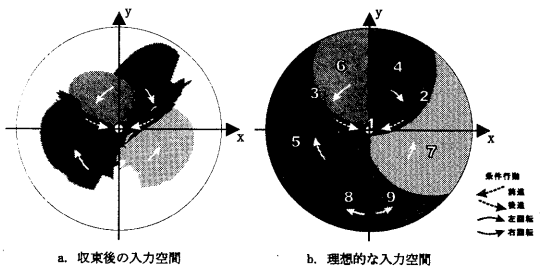
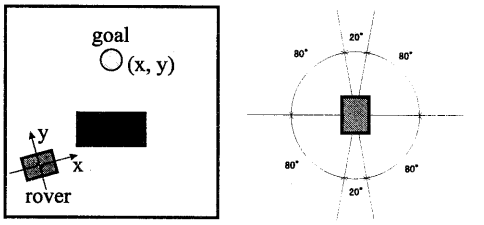


図 8: 左車輪の回転速度が15%低下した場合

いずれの故障の場合も、環境の小さな変化（センサ取り付け角度のずれが10°，15°の場合や、回転速度の低下が5%の場合）に対しては、状況の形状を変化させるだけで素早く対応できた。学習中の性能も、過去の知識を利用することによって、著しく悪化することなく維持したまま適応できた。また、環境の大きな変化（センサ取り付け角度のずれが20°，25°，90°，180°の場合や、回転速度の低下が10%，15%の場合）に対しては、報酬から離れていて歪みの大きい状況が消滅して過去の大きく誤った知識の影響を断ち切った上で、生き残った状況と新たに生成した状況で自律的に状況表現を再構成した。このような環境中の小さな変化に対して記号に対応する領域の形状を調整するシステムも、環境中の大きな変化に対して既存の記号を消去して新たな意味の記号を定義し直すシステムも、筆者らの知る限りでは、STNS以外には存在しない。環境の変化が大きくて状況表現を再構成した場合にも、ゼロから学習した場合に比べて、遅くとも累積行動数約1000回[4]遅れで同等の性能に達していて、安定性はあるといえる。

6.2 8次元入力のナビゲーション

最後に、複雑なアプリケーションとして扱った8次元入力のナビゲーション問題の結果を示す。図9aに示すように、6.1節の問題で用いた作業空間中に障害物を一つ置いた。ゴールは図の位置に固定した。ローバーには、ゴールセンサからの知覚入力に加えて、図9bに示す障害物センサからの知覚入力を与える。障害物センサは図の範囲内で最も近い障害物（周囲の壁を含む）までの距離を出力する。その他のタスク、報酬、行動の設定は、すべて6.1節の問題と全く同じものを使用する。



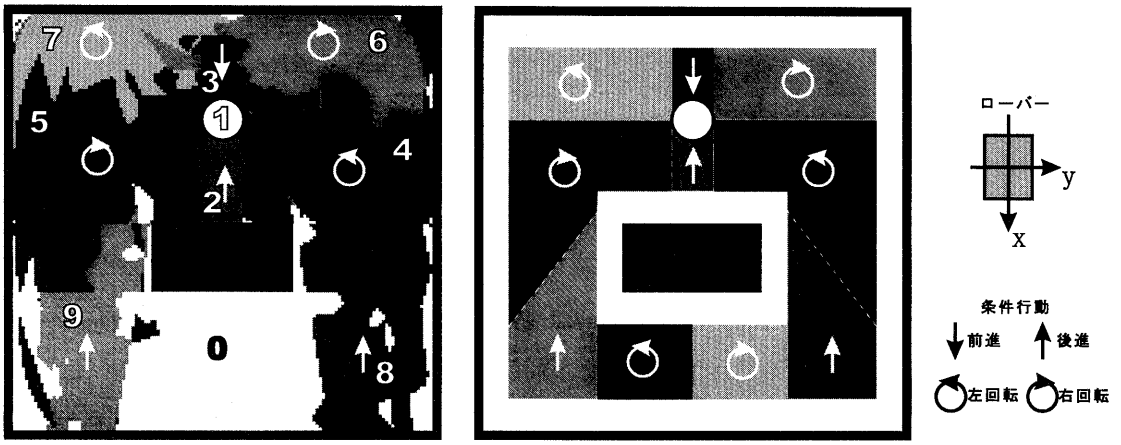
a. 作業空間　　b. 障害物センサ

図 9: 問題設定

このナビゲーション問題の政策を学習するSTNSをつくり、シミュレーションで実験を行った。学習が十分進んだ時（累積行動数28000回）の作業空間分割の例を図10aに、理想的な作業空間分割を図10bに示す。これらの図に示すように、ゴールに近い領域では、理想的な状況と対応が取れる程度の状況認識が獲得されている。しかし、ゴールから離れた領域では、2次元入力の場合と比べてはるかに長く学習しているにも関わらず、適切な状況認識が獲得できなかった。その原因は、主に、状況の形状を学習する能力の低さであると思われる。

強化学習では、最も一般的には、入力空間を均等なグリッドで分割して離散化する。この8次元入力の問題において、入力空間を距離は3段階、角度は10°刻みで36段階に離散化したとすると、状態の数は78,732個になる。この状態表

[4]履歴データベース中の全データの数に等しい。



a. 学習が十分に進んだ時の作業空間分割　　b. 理想的な作業空間分割

図 10: 学習後の作業空間分割と理想的な作業空間分割

現では、分割が非常に粗く、しかもそれにもかかわらず状態の数がかなり大きいので、適切な行動政策の学習はかなり困難であると思われる。それと比較すると、ASRR による状況表現は、タスク遂行の必要性に応じて緻密で、かつ抽象度が高いという点で優れているといえる。

7 結論

本稿で提案した STNS では、認識の学習と認識された表象の上での強化学習を並列して実行することによって、従来の強化学習システムよりも柔軟な状態表現を実現している。その結果、与えられたタスクに特化しつつ、不慣れな環境に柔軟に適応できる認知行動システムを構成することができた。STNS は、記号に対応する知覚入力空間中の領域を、筆者らの知る中で最も柔軟に調整することができる認知行動システムである。この柔軟性は、フレーム問題の現実的な解決のためには、非常に重要である。

しかし、このシステムは、記号の学習としては知覚入力空間からの「状況」の抽出、記号を用いた情報処理としては強化学習しか行わないので認知行動システムとしては非常に初歩的なものである。従来の人工知能システム中で用いられているようなもっと強力な記号体系、もっと複雑な情報処理をこの種の認知行動システム中に組み込むことは、今後の大きな課題である。

また、この種の経験に基づいて学習を行うシステムの共通の欠点として学習が遅いという点があげられる。その原因の一つは、学習を全くの白紙状態から行うことであろう。生物にも莫大な年月をかけて培われた遺伝的学習の成果が、先天的に備わっている。この種の認知行動システムにも、何らかのアプリオリな知識を設計時に埋め込んでやり、実際の経験に基づいて、新たな知識を獲得したり、従来の知識に修正を加えたりするのが現実的なアプローチであろう。

本稿で提案した認識の学習を行う認知行動システムのアプローチは、行動に基づく知能のアプローチと対立するものではなく並立するものである。STNS では、環境とのフィードバックループは一重であり、システムと環境との相互作用の点では不十分である。複雑なタスクを頑強に遂行するためには、行動に基づく知能のアプローチのように、独立して環境と相互作用する複数のモジュールでシステムを構成し、その各々のモジュールが認識の学習を行うような仕組みが望ましいと思われる。

認識の学習と認識された表象の上での学習を同時に行う認知行動システムは、まだほとんど扱われていない研究領域であり、フレーム問題の現実的な解決を目指してこの種の研究を発展させていくことは、実世界中の知能ロボットを実現する上で非常に有用であると思われる。

参考文献

[1] Rodney A. Brooks. Intelligence without representation. *Artificial Intelligence*, Vol. 47, pp. 139–159, Jan. 1991.

[2] Daniel Dennett. コグニティブ・ホイール: 人工知能におけるフレーム問題 (原題: Cognitive wheels: the frame problem of ai) . 現代思想, Vol. 15, No. 5, pp. 128–150, 1990. 信原幸弘 訳.

[3] Hiroshi Ishiguro, Ritsuko Sato, and Toru Ishida. Robot oriented state space construction. In *Proceedings of the 1996 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS 96)*, volume 3, pp. 1496–1501, Nov. 1996.

[4] 松原仁. 人工知能におけるロボットの役割. 日本ロボット学会誌, Vol. 14, No. 4, pp. 478–481, May 1996.

[5] 上野敦志, 堀浩一, 中須賀真一. 自律エージェントのための状況認識と行動規則の同時学習. 第 30 回人工知能基礎論研究会, pp. 19–24, Sep. 1997. 人工知能学会研究会資料 SIG-FAI-9702.

[6] 上野敦志. 自律システムのための状況認識と行動規則の同時学習. PhD thesis, 東京大学大学院 工学系研究科 航空宇宙工学専攻, 1997.