

## 1. はじめに

汎用人工知能を巡る世界は、事象、空間、時間で記述できる世界です。そうして、汎用人工知能の世界とのインターフェースはセンサーとアクチュエータであり、それらから得られるデータを解析し、蓄積し、相互に共起関係をもって認識し、それらデータを利用して世界に働きかけていく。データの蓄積は世界の構造を反映した構造を持ったものとなり、知識ベースとよばれる汎用人工知能の中核要素を構成します。本論では、この知識ベースの構造と知識ベースの利用の機構を記したいと思います。

## 2. 知識ベースの構造

汎用人工知能の基盤は知識と知識の獲得と利用です。その知識には次のものがあるでしょう。

- (1) 事象の記録とそのセット
- (2) 事象に構造を持たせ、情報処理をし易くする仕組み
- (3) 時間と空間を管理する仕組み

空間には世界空間と身体空間と事象空間という階層があります。全てのデータは世界空間に対してどんな配置になるかという座標記述できるものです。そして身体空間は身体の中心的な姿勢の軸に対して固定的に構築され、世界空間に対して持つ方向とか大きさとかで、複数のものが作られます。事象空間は、視覚では、各視野になります。SLAMにより、先ず身体空間に配置され、さらに身体空間から世界空間に配置されます。

事象は、極小のものは画素でありパーティクルです。この画素が集まってオブジェクト画像を構成し、パーティクルはオブジェクトを成します。オブジェクトは他のオブジェクトの部品として位置づけられるものでもあり、これは入れ子構造を構成します。オブジェクトは形状という属性を持ちます。部品オブジェクトはそれが属す枠のオブジェクト内に配置されますから、配置情報を属性として持ちます。

事象は具体的な物である場合と、視覚では見えない概念のものがあります。概念については、テーマと空間、時間によって区切られるステージと登場人物や物によって区切られるシーンに分割され構成され、更に、各オブジェクトの配置や動きのまとまりであるカットに分割構成されます。

事象はセットとグラフとマップと (key,value) とコマンドにより表現されます。これが情報処理の基本になります。これらは、汎用人工知能システムによって振られる固有の名前によって管理されます。マップのデータ (イメージなど) やグラフのデータ (オートマトンなど) などはそれぞれ異なる名前によって管理されますが、マップとグラフが同じ事象の異なる形式のデータだとすると纏めて統一した名前にして管理する必要があります。それがコンセプトです。コンセプトは、その他にフォーカスをもって表現されます。コンセプトは入れ子構造をしますから、その中のどの断面でコンセプトを抱いているかということフォーカスで表現するという事です。

なお、名前の与え方ですが、マップやグラフ、セット、オートマトンといった基板データエリアの創成 (Factory モジュールで一括管理) では、名前は人工知能システムに一意になるように与えられます。その上位のコンセプトは目的に応じて処理システムが固有な名前を付加して管理するようにします。言語が生まれれば、言葉 (単語) によってコンセプトは名付けられることとなります。

知識ベースはコンセプトを管理する機構です。何があっても不変なバックボーンとしての世界を記述するマップ群と、状況により部分的に作られていくマップを基本とする事象空間群からなります。

ただ、空間と時間の管理は海馬システムという知識ベースの一部の機構に於いてのみ存在する物です。メモリを沢山必要とするからです。海馬システム以外の知識ベース部分では、2次元イメージとかコマンドのみが保持されます。

マップは画像ですとイメージになります。断片的な2次元イメージ群とそのイメージ群を3次元の空間の一断面として位置づけた統一を行います。それは世界の認識である、知識ベースになっていきます。3次元世界空間は知識ベースの骨格であり、その3次元マップのノード群がコンセプトであるということになります。

知識ベースは、幾つもの階層からなります。事象、空間、時間の要素は枠と部品群という包含関係の対が入れ子の状態になっているからです。部品としてのマップがあって、それを含む枠としてのマップがあるという関係が、詳細なマップから、最高位のコンセプトに至るまで階層を成すのです。ステージ、シーン、カットというコンセプトの構成もこの枠と部品群という階層の一つであります。

少し飛躍しますが、空間認識で、繰り返しパターンの発見というものが重要ですので、それに触れたいと思います。

まずは、1次元のパターンについて繰り返しを発見することを考えて行って見ます。2次元については、適当な1次元のパスを追って行って、1次元の問題にして解くと良いでしょう。

1次元のデータとしては、音素とか画素とかの量子化をされていることを利用します。量子を逐一スタックして辞書を作るのです。1量子目を先頭に、次々と量子列を2量子列、3量子列というふうにスタックしていく。また、2量子目を先頭に同じ事をする。3量子目を先頭に同じ事をする。とうとう。そうして、実時間で、そのスタックされた量子列毎に比較して・・・相関係数を求めて・・・一致度が高いものを繰り返しパターンとして認定します。量子化とは、円とか三角とか纏まって認知されるものも含まれる概念です。

### 3. 知識ベースの構築の仕組み

知識は身体の活動を通して獲得されていく物です。もちろん、生得的な知識もあるでしょう。センサーとアクチュエータの持つデータ (KEY,VALUE) の集まりが知識獲得の原点です。

視覚については、次の2つの系統があります。

#### (1) 上丘系

大まかな、事象の配置とパーティクルの間隔や動きを認識します。

#### (2) 外側膝状体系

形状や運動を認識し、事象を認識します。

3次元事象空間の認識は、海馬系の構造で実現します。

センサーデータは世界の中の事象についての情報をもたらしますが、受け取った汎用人工知能の世界には曖昧な意味しか一時には与えてくれません。また、一つのセンサーが全ての情報を解析できるわけでもありません。そこで、各センサー情報を解析・認識するシステムをエージェントとして造り込むとして、センサーやアクチュエータの (KEY,VALUE) の認識処理はマルチエージェントシステムとして実現します。このマルチエージェントの共有する場があって、調停場と呼ぶことにします。汎用人工知能では、この調停場が目的に応じて多数存在することになります。

上丘系はオブジェクトがなんであるかという認知は行わず、高速に画像処理を行い、画面のおおざっぱな状況を把握するという任務を担います。だから、基本的に画像処理であり、学習機構は用いません。

外側膝状体系はオブジェクトの形状と配置から何であるかというコンセプトを得るので、複雑な学習過程をとります。基本的には、次の技術が適用されます。

#### (1) クラスタリング (強化学習)

#### (2) ディープラーニング

#### (3) 交差法

海馬系は事象の3次元の空間配置と1次元の時間配置を管理します。そうして、SLAMにより、各画面の状況を世界空間にマッピングしていく働きをします。

クラスタリングは、事象の重なりを計測することで実現します。あるときには事象はマップの特定の画素群を占めますが、あるときには、別の画素群を占めます。そうすると重なった部分が多いほど、事象を弁別する重要な画素群であると言えます。その弁別がクラ

スタリングです。その重なり解析は、画素の大きさの粗さを粗いときの弁別から精密な画素の大きさの場合の弁別まで、「Coarse to fine」法によって実現します。

また、2つの事象が異なるか相同かということクラスタリングで弁別するときには、重なり度合いを解析する必要がありますから、「t検定」法によると良いでしょう。

視覚については上記のようですが、聴覚や身体覚も同じように、空間配置の認識系と事象の認識系の2つのパスからなります。

音は、音源の位置の認識と、音の認識（周波数分布（音色）、強さ、周期、拍など）があります。身体覚も同様に、身体の配置の認識と、身体に接触しているものがなんであるかの認識を行います。

学習の仕組みですが、次のものがあります。

- (1) ディープラーニング
- (2) クラスタリング
- (3) 強化学習
- (4) 交差法

ディープラーニングと強化学習は他の書物を参照して頂くとして、ここでは、クラスタリングと交差法について少しアルゴリズムに触れたいと思います。

クラスタリングについては、マップ形式に分布する点（量子化要素）の分布密度を解析して塊を切り出す処理で、強化学習に似ています。解析は「coarse to fine 法」で行います。粗い解像度でマップを見て、局所的に最も密度の濃いマス目を塊として分離します。段々と精密な解像度でマップを見ていくことで、クラスタリングが上手く切り出されていることを確認します。なお、1次元の場合には、2つの分布が有意に分離できるかどうかをt検定で確認すると良いでしょう。2次元の場合には、視線という1次元のパスを張って、1次元化してt検定をすると良いと思います。

交差法については、結果をオートマトンプールというグラフ形式の表現で、全体データを分離し、分離したデータの塊を順次結びつけていくことをします。交差法とは例えば次のようなものです。

リング畑で、木に成っているリングを1つ握って、「リング」と言われたとします。真新しい人工知能であれば、リング畑の風景とリングを握った手とリングが、全体として「リング」と判断するでしょう。今度は、スーパーへ行って、「リング」とリング一つを握って提示されたら、背景が交差法によって取り除かれ、リングとリングを握った手を合わせて、「リング」と判断するようになるでしょう。こんどは、テレビに映ったリングを「リング」とい

ったら、手が削除され、「リンゴ」の意味が最終的に決定されるでしょう。

交差法の考え方は明確ですが、処理は重たい物になります。データを制限する必要があります。例えば、音素10個までの記号を単語獲得の対象とするというようなことで有り、1文をSVOC文型を十分内包できるだけの単語数に制限するというようにです。

交差法で、原語を獲得するのは、基本的なものを対象とし、ある時期が来たら、対象に制限を設けないアルゴリズムへ移行します。単語合成、複合語合成とかですね、それによって、文法の能力を豊かにして行くのです。

#### 4. 知識ベースの利用の仕組み

知識ベースを構成したり、利用したりするには、自律的なプログラム・・・エージェントとして実現するシステムが有効になります。それは、汎用人工知能という巨大システムを作るために、モジュール分解を高度に独立したものとしておくことが現実的な実現方法だからです。そうして、エージェントは自律的に動作していくようにします。

エージェントは、複数のものが独立に動作するというマルチエージェント環境を構成します。その構成では呼び出されるエージェントと呼び出すエージェントという上下関係の中にあるものと、必要に応じて会話する同僚関係になるものがあるでしょう。いずれにしても、あるエージェントを埋め込んだときには、他のエージェントと特別なプロトコル（特定の両者が納得している引数渡しとか）を決めるわけにはいきません。汎用的なプロトコルであるべきです。そのため、通信のためのオントロジーが必要になります。そのオントロジーは知識ベースの基盤として整備されている物であります。

以下、知識ベースの利用について記したいと思います。

思考と行動が知識の利用の中心的な事柄です。思考は、次の段階からなります。

- (1) 気づき
- (2) 問題発見
- (3) 課題設定
- (4) 課題解決
- (5) 評価

相同と相異という共起の基本機構を手がかりに、いままでのデータと今のデータの間を把握して、「気づき」で事象を切り出します。「問題発見」も「気づき」で漠然とした違い、相同を把握しているものを吟味して、どこが問題かを発見することです。このように「新しい発見」というものに結びつくのですが、その要素は既存の知識です。プリンタの故障という発見があったとしても、コンピュータシステムがプリンタが故障すると止まるという知識があって始めて、認知できることです。プリンタの故障もインク切れとかジャムとかの経験があって・・・設計時に想定される物もある・・・の認知活動です。そのように、思考過程は連想のネットワークなのです。ここで、知識ベースがオートマトンプールを基盤になっていることが想定されます。とくに、オートマトンのステートの枝分かれが大きな所は、プロダクションシステムとして構成し、高速化と大規模化を実現します。

行動は、行動発現の根本に「意志」があります。意志は手足・指といった身体の動きで大きな行動表現のコマンド（意志の表現）を接地し、更に、細かな1つ1つの筋肉の動きにまで落とし込んだコマンドまで入れ子構造で規定します。そうして、感覚が入力になって、

各筋肉を動かす (key,value) の出力群が発現します。そしてそれは、各身体の状態 (姿勢と動き) によって決まりますから、オートマトンプールとして表現されるシステムと成っていきます。学習は、このオートマトンプールを構成・発展させるものです。

思考と行動が交差する場が、目標管理思考で、無意識下の自動系である直観に対して、意識下の意志を持った思考プロセスです。そうして、その思考プロセスの利用するデータ範囲が調停場というものになります。

なお、行動制御では、思考の直観に対応する物は、反射系になります。重力補償システムや、手足のバランスを取ったり、運動の基本動作を発現したり、行動の誤差補償などは反射系を構成して、意識下の行動制御は、その反射系の動きを加減していくように設計されるべきです。一つの反射系は一つのオートマトンを構成します。

反射系は幾重にも重なって合成されていき、それぞれの反射系は、また、「意志」によって制御されるものでもあるという風に、多重の制御系を成します。例えば、重力を利用した歩行の各身体部位のリズムは、腱反射などの反射系をまとめて実現する反射系であり、また、「意志」の下に歩行から、走るとか、足を蹴り上げるなどへと制御されていくものです。



## 5. オートマトンプール

言語や思考、行動の基盤はオートマトンですので、少し、オートマトンについて詳しく記述していきます。

ステートマシンという意味で、「オートマトン」というものを捉えて来ていますが、言語や思考などを考えていきますと、ステートマシンのステートにプロセスを配置すべきことが言えます。それは、入力と出力というのが、設計時に静的に決定づけられないということが言えるからです。例えば言語表出について考えて見ます。

ステートとは、文脈のある状態でしょう。起承転結ならば、今書き出しで、舞台説明がいるとか。論述ならば、研究の目的と結論を書くとかします。このように経験によって、発話のステートとステートの大まかな遷移関係が獲得されていきます。そのステートで何を語るかということは、個別の目的によって・・・対象とする知識体系によって・・・異なります。場合によっては、この知識の目的は何かなど考え、知識ベースを解析することになります。結論は何かなどという目的の下、知識ベースを解析して、発話すべきことを決定します。入力も出力もステートが決められたときには定まっていない事象です。ですから、オートマトンは、ステートマシンというだけでなく、プッシュダウンオートマトンとかチューリングマシンのようなものであると言えます。

なお、起承転結の「起」で「舞台説明」とか「結論を書く」という処理をしますが、その処理自体も「起」の下位のオートマトンのステートとなります。このように、単一のコマンド群（「記述する結論を決定する」のような）にまでオートマトンの階層が構成されます。

ステートに貼られるプロセスも学習によって得ます。それが汎用人工知能の根本技術です。コマンドの獲得ということ、連想の獲得ということで、プロセスは作られていきます。意志というものも重要になります。意志はステートの選択に関わってくる。入力がステート遷移を決定しますが、意志もステート遷移を決定していくのです。

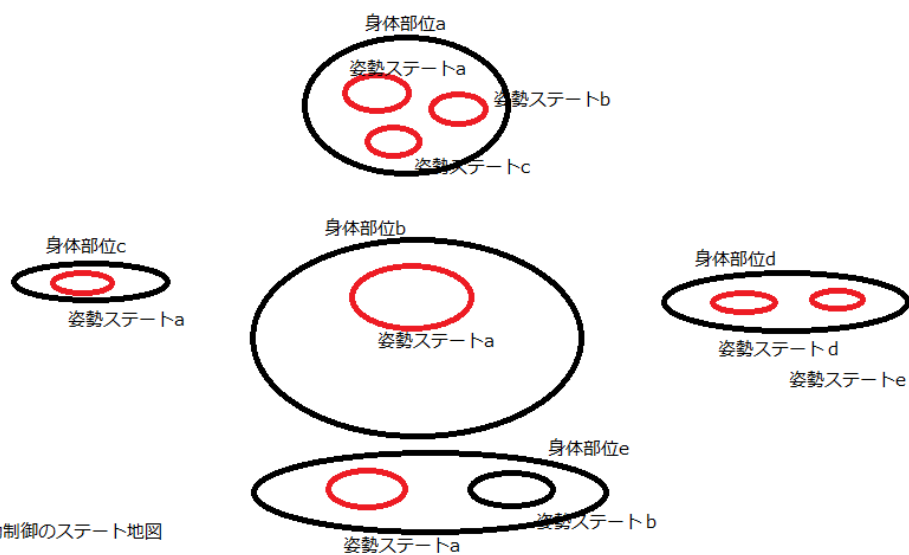
そんな一大体系がオートマトンプールです。

数式処理ですと、式のある部分の変化の前後の様子をもってステートとします。行動制御ですと、姿勢の変化の前後の様子がステートとまります。それは、「変化」がその時の「意志」とか「入力・出力」を規定するからです。

変化はコマンドです（コマンドのオペレーション項である動詞は変化を表わします）。コマンドは、変化前のイメージと変化後のイメージを対応付けます。逆に、イメージの変化からコマンドを推定できるわけで、その変化は知識ベースを解析するときの基盤となるデー

タなのです。

なお、行動制御は身体各センサーとアクチュエータの関係を基盤としていて、その上に姿勢のステートが配置されます。その図を以下に示します。



図で、「姿勢ステート a」は、「身体部位 a」、「身体部位 b」、「身体部位 c」、「身体部位 e」に分散して存在しています。

## 6. おわりに

2016年10月25日の「鉄腕アトムの設計書」からずっと汎用人工知能の技術を語ってきました。なんども考え直し、技術書を読み直し、そんな中で必要な技術は出そろったかどうかと内省してきました。部分的な技術は出そろっているというのが現在の結論です。問題は、それら技術群をどう纏めていくかということです。人工知能の応用といっても、パラメータの設定とかで個々の問題に対してそれなりの工夫が必要です。既存の技術をどう展開していくか、それぞれの汎用人工知能開発者の創意工夫が待たれるところだと感じます。

本論は、汎用人工知能技術のバックボーンはこんな所だろうということを示すことで、目的を達成できたと思います。

おわり