

1. はじめに

汎用人工知能の基本能力について考察していきたく思います。人工知能は自律的に自己組織化して、外界に適切に作用していくことが求められます。外界からのデータ獲得はセンサーによってであり、それは外界の持つ構造と外界の動きを人工知能が知るための基本的なデータ収集能力です。また外界への働きかけを行うためのアクチュエータが必要になります。アクチュエータは身体の構造を把握し、且つ、外界の動きの予測を的確に行った上で、獲得した知識を基本として、外界に働きかけることが求められます。

本論では、センサーとアクチュエータの構造と、外界の構造の知識を獲得し、適切な働きかけを行うことに必要な基盤プロセスについて考えていくことにします。

2. センサーシステムの種類

汎用人工知能は、身体と空間・時間を把握するものと、ものの属性や重力を知るものがが必要です。

センサーとして、ハードウェアそのものが値を返すもの（ハードウェアセンサー）と、汎用人工知能が利用するまでにプロセスが介在し、値を返すもの（論理センサー）があります。

そうして、センサーの種類は次のものからなります。

(1) 身体上に分布するもの

皮膚上に分布するセンサーで、温度とか圧力の値とかを返すもので、皮膚上の位置が把握できるものです。皮膚上の位置とは、空間内の身体の配置（姿勢）を反映した座標をもって把握されるものです。

(2) 姿勢を検出するもの

骨格の空間配置と骨格の中の2つの骨のなす角度群を返すものです。また、身体のイメージ（胴周りとか骨の長さとか）を返すものです。空間内の身体の配置（姿勢）を反映した座標をもって把握されるものです。

(3) オブジェクトやオブジェクトの運動を検出するもの

移動センサー、距離センサー、塊センサー、速度センサー、曲率センサーがあります。聴覚センサー、視覚センサー、皮膚感覚センサーの各センサーに、更に詳しく

い情報をもたらす論理センサーとして実現されます。

(4) 身体の周りの時間、空間の広がり、オブジェクトの配置を知るもの

3次元空間と運動状況の把握を行うセンサーです。オブジェクトとオブジェクトの運動を把握するセンサーから、さらにオブジェクト間の関係（空間的關係）を得る、高度の論理センサーです。

(5) 属性を知るもの

色とか、速度とか、重さとか温度とかの物事や運動の属性を知るセンサーです。

(6) オブジェクト間の関係（格）を知るもの

空間關係以外の物事の間關係や運動の状況を把握するためのセンサーです。

(7) 重力を検出するもの

重心の位置と大きさ、ZMPを知るセンサーです。皮膚への圧迫やすべり感覚、骨格に掛る重力圧力といったものです。

(8) 注意の共鳴を実現するもの

他者が自分と同じものに注意を向けていることを保証するセンサーです。視線とか、ハイライトがあたっている状況から注意を向けている事象を特定する能力です。

動いているもの、木々よりも動物を、動物より人間を、人間でも話をしている人に注意を向けることになります。また、目立つ（コントラストが強いもの）ものに注意が向けられます。注意機構にはプロトコルがあるのです。

3. アクチュエータの種類

汎用人工知能は、外界に働きかけるための、骨格システムとそれを動かすアクチュエータが必要になります。

ハードウェアアクチュエータと論理アクチュエータがあります。人間の筋肉では、筋繊維一つ一つの収縮と弛緩を制御するのがハードウェアアクチュエータで、筋肉という筋繊維の集まりをまとめて制御するのが論理アクチュエータです。骨格を管理するアクチュエータは更に高度のアクチュエータになります。

そうして、もっと抽象化しますと、思考過程も、基本的にコマンドによって実現される過程ですから、論理的なアクチュエータを構成すると言って良いでしょう。手をどう動かすかといった最下位の思考要素を、数式を書くといった思考要素が被い、数学の一つの問題を解くといった、大きな思考要素が更に上から覆うことになります。

4. コマンド（記号化）と基本データの構造

汎用人工知能は、意志によって自律的に動く必要があります、意志の下、各種のプロセスを実行していくことになります。そのプロセスを体現するためにコマンドというイメージを記号化して参照できるようにする仕組みが必要になります。

コマンドは、デジタルツインの構造を反映して、次の形式の記号化したデータです。
(動詞[、格、名詞/属性詞]・・・)

そうして、基本データは、次の、数学的な基本構造を持ったものです。

(1) セット

集合論を体現するデータ表現です。

(2) マップ

画像イメージやものの配置を体現するデータ表現です。

(3) グラフ

ノード列を中核にして、長さ・距離関係やものの分岐を体現するデータ表現です。

5. 身体と姿勢と身体を含む空間の管理

汎用人工知能は自身の身体と自身の姿勢と身体を含む空間のデジタルツインをイメージとして内面（記憶素子上に）持ち、様々な情報処理をしていく必要があります。デジタルツインが全ての処理の基盤になるのです。

デジタルツインは、センサーの各機能をもって、空間上にオブジェクトを把握し、記憶装置上に空間と相似なイメージを配置していくことになります。また、デジタルツインでは、オブジェクトの間の関係や運動に関するオブジェクト間の関係を表現するコマンド（記号）を意味の把握としてイメージに添付することになります。

身体は、腕の長さとか、腕の太さといった3次元の体座標系をデジタルツインとして把握されます。それは、基準点と身体各部に埋め込まれた特殊なセンサーの間の信号の伝わる時間感覚として求められます。

6. 基盤プロセスの種類

汎用人工知能は、汎用であるために幾つかの基盤プロセスを持ちます。

- (1) データの重ね合わせと重なり具合によるクラスタリングにより、パターンとなるデータのチャンクを獲得するプロセス
- (2) 交差法によるデータの間接関係を記号で得るプロセスを持ちます。記号はコマンドという特定の表記をもつデータとなります。
- (3) プロダクションシステム構造を持ちます。
- (4) 生成・変形処理を行い、特定のシーケンスデータを生成するプロセスを持ちます。
- (5) ポピュレーション符号化を行い、汎用人工知能の処理過程に複数の可能性を常に持つようにします。
- (6) 無数の評価項目を持ち、汎用人工知能の処理結果を評価していくことが行われます。
- (7) 思考の実現を図り、柔軟な外界への対応を行います。
- (8) 外界に働きかける行動制御が行われます。

(1) と (2) は、外界のデジタルツインを作るという学習過程であり、(6) は汎用人工知能の意思を発現する過程であり、(7) と (8) は、デジタルツインを基盤として、意志の下に外界に働きかける過程です。(3) と (4) と (5) は、汎用人工知能の情報処理の基盤機構ということになります。

7. 重ね合わせとクラスタリング（多段階解像度解析、枠と部品群、ステージ・シーン・カット）による学習

視覚などのセンサー値からセンサー値が表す空間のデジタルツインを作るために画像を解析し、オブジェクトを学習していく過程で、提示されたセンサー値の為すパターンを整理していく必要があります。重ね合わせを行い、相同と相異の部分を切り出し、相同部分をチャンクとしてクラスタリングしていく技術が必須となります。チャンクが、以降の情報処理で、パターンマッチングして、相同部分を切り出しオブジェクトを認識する手がかかりとなります。

チャンクを作る場合、画像イメージですと解像度の低い枠組みで先ず解析し、重ね合わせして、次に解像度の高い枠組みで解析し、重ね合わせしていきます。そうして、次々に解像度を上げていくのです。これを多段階解像度法と呼ぶことにします。

重ね合わせは、閉曲線を最外枠として切り取り、内部の図形を部品群として行います。外枠がない場合には、仮想的な閉曲線で部品群の枠として設定します。

生活の中で体験する事象群は、時間と空間の枠組みの中で行われるものを管理する単位であるステージというものがあります。次に、そのステージの枠組みの中で、登場人物（注目する対象群）が変わらず続くという一連の事象の流れという枠組みがあります。これをシーンと呼ぶことにしましょう。そうして、各登場人物の所作が生成されます。これをカットと呼びましょう。こうして、生活はステージ、シーン、カットの連続としてチャンクを作ることとして管理できるものになります。

8. 交差法と記号化（変数、コマンド）

交差法はセンサー値群（デジタルツイン）からチャンクを得て、コマンドという記号化表現を得ていくための技術です。

ボールを投げるというシーンを捉えてみます。手でボールを握るというカットに続き、手が前に動き、ボールが手から離れるカットが続きます。ボールは目標の方向に飛んでいくカットが次に続きます。それぞれのカットにコマンドが割り当てられます。これが記号化です。手でボールを握るというカットは、人も点、手も点で、ボールも点で、これらの点が重ね合わされたイメージが一つのコマンドとして把握されます。そうして、「握る」に相当する動詞と、この動詞の周りの「人」と「ボール」が主格とオブジェクト格という関係で結ばれ、コマンドに記されます。

次に、輪投げを考えてみます。この場合、ボール投げと同じコマンドシーケンスが生成されます。違うのは、「ボール」が「輪」であることです。ここで、交差法で、「ボール」も「輪」も交換可能なコマンド、（投げる、主格、人、オブジェクト格、「変数」）というコマンドが抽象化されます。この交差してコマンドの引数とオブジェクトや動作を抽象化する過程を交差法とよぶことにします。

このように、視覚での運動については、画像イメージをそのまま記録すれば、デジタルツインが構成できますが、論理的な、目に見えないものについては、セットとマップとグラフという基本データを中間において、抽象化したイメージで、デジタルツインを作る必要があります。

例えば、「必要とする」という概念のデジタルツインは、2つ以上のセットに所定の事象が共通オブジェクトして存在することです。

例えば、「長い」という概念のデジタルツインは、2つのグラフがあつて、一方のグラフのノードの数が多というイメージを作って把握します。

9. プロダクションシステム

パターンマッチング用コマンドに処理プロセスとしてのコマンドを連想させたものがプロダクションシステムです。このプロダクションシステムの寄って立つ作業記憶域は、デジタルツインになります。デジタルツインの中のイメージを操作して評価や思考、行動を生成していく過程で、プロダクションシステムは基本プロセスを構成します。

パターンマッチングは、多段階解像度法によって、大体こんな感じというマッチングから、精密なパターンマッチングまで、同時に行うことが肝要です。また、マッチングの結果はポピュレーション符号化によって、いくつかの可能性を考慮することになります。

10. 生成・変形処理

言語の生成や行動のアクチュエータシーケンスを生成するときに、処理を手続きだった、頑健なものとするための技術です。

デジタルツインの基本的な構造から、言語とか行動アクチュエータを基本的な形で生成します。その後、詳細な付加情報とか生成されたものの間の関連情報から基本的な形の生成物を変形させ、実際の出力とします。

言語生成の場合の生成・変形処理の例を次にあげます。

デジタルツインで、女性がスーパーで買い物しているイメージがあったとします。それは、私が頼んでやってもらっているとします。このシーンを表す言葉を生成しようというわけです。まず、「彼女は買い物する」とデジタルツインのシーンをコマンド形式の情報を使って、言語で生成します。これに、「私はそのような状況を作る」というコマンドからの情報から、「彼女は」を「彼女に」と、「買い物する」を「買い物してもらおう」と変形し、主語「私は」を追加し、「私は、彼女に買い物してもらおう」という完成形の言語表現を得ます。

次に、行動制御における、生成・変形処理を考えます。

意志の働きで、姿勢の変化（目標の姿勢を提示する）を計画し、行動要素（チャンク）のシーケンスを生成します。脳では、補足運動野などの高次の行動生成領野の働きです。これで、基本的なアクチュエータの時間変化のパターンが生成されます。これを、一連の姿勢変化パターンを制御データとして、目的中心の滑らかな最適軌道を起こすように修正する作業を行い、実際のアクチュエータの値シーケンスとしていきます。この修正作業は、脳では小脳の仕事と考えられます。

修正用シーケンスの姿勢変化パターンとの対になった学習は、パーセプトロン（ディープラーニング）で行います。

1 1. ポピュレーション符号化と曖昧性処理

一つの判断を、確実に一つ得ることはまれです。曖昧性のある、いくつもの可能性のある判断からなるのが普通です。また、広い立場から、再度評価し、もっとも可能性のある判断を導き出すために、ポピュレーション符号化のテーブルを利用することも有効です。

パターン認識について、ポピュレーション符号化を考えてみます。例えば、丸い円の中に横に二つの小さな円が並んでいたら、顔だという可能性は60%あるだろうと判断し、テーブルに記します。その他に、皿にミカンが置かれていると判断もあり得ます。40%としましょう。あるいは、ただの模様という状況もありえます。50%かなとして、テーブルに記します。あとで、周りの状況から、マスクをした顔だと判断するのにこのテーブルを利用します。

1 2. 評価システム

センサーの値をもって評価とします。このようにセンサーと評価システムを捉えると、学習によって評価項目（たとえば、「問題が解ける」というような）を獲得していくことも必要です。「問題が解ける」というような評価項目は基本的に、快適とか達成感とか、メリットとかの基本センサー項目に接地していますが、より高速に評価を行うために設けられるものです。

評価は評価されるデジタルツインの構造をパターンマッチングし、評価項目を結果として出力するというプロダクションシステムになります。

1 3. 思考システム

注視事象の共鳴というものが先ずは重要になります。これによって、他者との情報のやり取りができるようになります。この発展したものが、ミラーニューロンシステムで、この共鳴によって言語システムが実現します。

思考の中核は、シミュレーションによる予測過程です。シミュレーションもシミュレーションの結果の評価過程もプロダクションシステムによって構築されます。

思考システムは、次の3つの構造からなります。

(1) 直観

評価システムと同じ構造をしたプロダクションシステムです。結果は調停場というデジタルツイン空間にポピュレーション符号化で保存します。直感中には、意志は作用できません。

(2) 目標管理思考

コマンドのネットワーク（ステートマシン）で、各コマンド処理毎に評価を意志の下で行い、予測や思考を実現していく過程です。評価によって直観を動かすことも行い、調停場の様子から更に思考を展開していきます。

(3) 調停場

作業用のデジタルツイン空間です。

1 4. 行動制御システム

重力に対応する、安定した姿勢を維持する技術について述べます。

重心の大きさと重力の向きと、身体が地面と接地している点群との間の関係のみが姿勢の安定を決定する考慮点です。手足胴体の組み合わせの様子は関係ありません。簡単な平衡の条件を力学的に解く問題になります。

歩いている場合、動的な安定を考える必要があります。この場合でも、時間を追ったシーンの力学的平衡状況を簡単な計算で求めることができます。

行動制御で発するアクチュエータの値群も、生成・変形処理で生成していくことになります。先ず、骨格の変化の方向と強さから基本的なアクチュエータの値シーケンスを生成します。それは、学習によって得た情報に基づくものです。デジタルツインに根を持つ情報です。そのような、シーケンスが連なって実際の動きになるわけですが、そのシーケンスのつながりによって、基本値のシーケンスを変形し、なめらかな、最適な運動となるようにしていく過程が続くのです。そうして、最終的なアクチュエータ値のシーケンスは生

成されるのです。

行動制御は複数の行動を同時に行ったり、計画を立てて、順に行ったりしていくことを行います。その中では、常に、この行動は目的に合っているかを評価していくことで、統一だった生活ができるように保障していくことが必須になります。その中で、多段階解像度法も有効になります。行動での多段階解像度法とは、次のようなものです。

長野から仙台に行くという大枠の意思から、バスに乗って長野駅まで、長野駅から新幹線で大宮まで、大宮から仙台まで新幹線で行くというふうに計画はブレークダウンし、実際に行動していきます。さらにその行動は、朝ごはん食べてから玄関まで歩くとかにブレークダウンし、計画にはない本箱で列車で読む本を物色するとかの行動も差し挟みます。さらに、右足を出し、左足を出しといった行動の基礎的なレベルまでブレークダウンします。こうゆう風に、多段階解像度法で行動制御は見えていく必要があるのです。

15. おわりに

汎用人工知能のプロセスは、全て、コマンドの実行として実現できます。そのことは、コマンドが常に世界を表現するデジタルツインとつながっていますから、世界を表現するものになっていることが保証されているからです。

更に、多段階解像度法を用いると、大枠なコマンドの実行から、微細な目的別の行動を作り出す、あるいは知識のスタックを作る強力な外界対応力を持つようにすることができます。その上で、本論の技術は思考や行動が実現されることを目指した技術となっています。

汎用人工知能のデータベースは、デジタルツインの他に思考過程を構築するネットワーク・プロダクションシステムと行動制御を構築するネットワークシステムと言語認識・生成ネットワークシステムがあります。これらデータベースは、重ね合わせと公差法とパーセプトロンによって学習していき、生成・変形文法により、外界に働きかけを行うこととなります。

汎用人工知能は、センサーとアクチュエータによって世界とつながり、世界のデジタルツインを構築し、コマンドとコマンドの束であるプロダクションシステムによって全てのプロセスが実現されているシステムであるということが出来ます。

おわり