

## 1. はじめに

汎用人工知能の中心課題は、言語と思考システムの実現です。本コンテンツは、この課題に挑戦し、課題を克服するシステムの構造の詳細を設計することです。

言語の獲得と思考の実現は、汎用人工知能が生活する中で得る体験を、秩序だって整理し、意味という物を構築し、その意味体系が作る知識を利用できるようにしていくことで可能となります。体験は、汎用人工知能が自律的に生成する行動と環境からのセンサー入力による情報の結合です。その時に、価値ある・・・目的を持った行動ができるように、優位な行動と身体に障害を及ぼす行動との弁別です。基本的な評価機能が汎用人工知能に備わっていないければなりません。

センサー入力の主な物は、視覚と聴覚です。それらのセンサーデータは時間と空間を形作り、サイバー空間として汎用人工知能内部に保存し、利用されます。言語は、体験の記号化によって獲得される能力です。汎用人工知能はセンサーデータが持つ構造に着目した特徴パターンに記号を割り振り、外界に有効な行動として作用していく手段を構築していきます。

音声は、音素が最小の単位となり、画像は部分画像が単位となり全体が認識されます。意味のある音素列や部分画像はチャンクとして纏められ、全体の情報はそのチャンクの連なりのグラフ構造として表現されます。ただし、一次元配列である音声のグラフは音素列をチャンクとする時系列のグラフですが、画像は2次元でありますから2次元空間の図形をチャンクとし、さらにチャンクの上位の2次元画像の内部でのその図形が占める位置と大きさをチャンクとするノードを図形チャンクの前部に持つグラフとなります。

ここで注意すべきことは、音素といえども音波のある時点の特徴を表現するものですから、全く確定的に決まる物で無く、ある種の曖昧性をもって、典型的なパターンを持つ物が音素として獲得されると考えるべきでしょう。そこで、音声処理については、音素の獲得から単語の獲得、そして文章の生成・認識へと設計を進めていきます。

## 2. 画像知識の獲得

2次元の静止画像を基盤として、時間変化を組み入れた動画像と3次元に空間を拡張した世界空間とが展開されます。そこで先ず、静止画像の解析を設計していきます。

視野とフォーカスが基本機構になります。視野の中にオブジェクトとオブジェクトの属性とオブジェクトの配置が捉えられます。フォーカスは解析が行われる場を決定し、このフォーカスを移動していくとで、視野・・・視野を越えて、世界の情報を構築していくことになります。

画像解析では、2つの機能モジュールからなります。

- (1) オブジェクトがなんで有るかを判別するモジュール（外側膝状態系）
- (2) オブジェクト間の距離とか配置、属性を判別するモジュール（上丘系）

先ず、オブジェクトがなんで有るかを判別するモジュールの設計から始めようと思いません。

オブジェクト画像は枠と部品群という構造を持っています。部品は更に枠と部品という風な構造を持ち、画像は入れ子構造をしています。

枠は、閉曲線であります。また、色々な画像要素が集まっていれば、それらを包含する閉曲線が想定できます。想定される閉曲線も枠であります。

枠が同一のオブジェクトのものであるとはどう判断すべきでしょうか。赤ちゃんが、母親を生活の中で繰り返し見ます。その中で、母親の画像はこんなものだと記録することになる分けですが、不変である母というオブジェクトの一つの表現である画像であるとはどう判断するかということです。父親の画像かも知れませんし、母親の愛猫の画像を見るかも知れません。

声の周波数分布の違いとか、顔つきの似ているというようなことが手がかりになるでしょう。特徴の不変性を捉えてオブジェクトの同一性を推定していくのだと考えてこれからは論じていきます。顔の変化はガウス雑音の範囲の変化でしょう。そういう相同性の解析ということが行われるはずで、その為にも、画像や音声の特徴というものを解析し、記憶し、利用していくことは重要な人工知能技術要素であると言えます。特徴とはオブジェクトとその属性の塊です。そうして、オブジェクトとその属性は、汎用人工知能が生活していく中で、獲得していく記号になります。直線はどんなものか、相同と相異という2つの評価項目から、定義できるように学習していく能力によって獲得された概念（記号）です。これらは、基本的にオントロジーとなります。

このように枠と部品群という入れ子と相同性の弁別から、視野をフォーカスが動くとき、オ

オブジェクト候補の画像要素をチャンク（ノード）とするグラフが作られます。

チャンクは、画像要素で、大きさと回転は正規化した2次元配列として保存され、画像要素が属する枠画像での位置と大きさと回転の値を属性として添付し、枠画像に接続（グラフとして）されます。正規化した2次元配列は名前付けられ、あとの解析で利用できるようにします。

正規化して作られる2次元配列は重ね合わされてガウス誤差の範囲ないで、一致するか、分離するかを各画素毎に評価されます。それによって、配列の分類を行い、分類によって名前を付けられ、知識要素としてサイバー空間（イメージを保存）と知識ベース（記号情報を保存）に分散して保存されます。

次に、オブジェクト間の関係を得るモジュールについて設計していきます。

オブジェクト間の関係の基本的なものは距離と方向です。接触しているか離れているかは2つのオブジェクト間の距離で表現できます。あるオブジェクトの周りを巡るというのは、距離があって、方向が順次変わって、最後に元に戻ることで表現できます。To とか from も離れている状態から接触状態に変わる事とか、その反対に動く事として表現できます。オブジェクト間の関係はコマンドで、そのなかの格で表現されます。

動作がコマンドの要のデータになります。外界の認識はオブジェクトの動作を記号化することにあります。その動作に現れるオブジェクト群とそのオブジェクトが互いに影響し合う関係にあることを動作を要に格として表現します。これがコマンドの全てです。動画画像イメージがあって、それらを重ね合わせを主体にして分析する交差法で、オブジェクトとその格と動作が精密な記号で表現されていくのが学習です。

触覚は身体を起点に外界の構造を認識する手段になります。それは骨格の構造を自分で知っている汎用人工知能で可能なことです。動画画像も触覚も、3次元のイメージを創り出し、そこから交差法で、コマンドに獲得していくことで、世界を内面に獲得していくことになります。

そのようなコマンドの中で、特徴的で基盤的なものは、音声認識で基本となる音素のようにコマンド体系の核となるものを形成します。汎用人工知能が日常生活の中で様々な作業コマンドを生成するときには、この核となるコマンド群が基盤となって作業コマンドを生成することにより高速な人工知能となります。特に、動画画像を解析してコマンドを作る場合が重要です。

上記の通り、画像イメージを沢山学習していくと、入力のイメージが既存のどのイメージに近いかということ学習済みの全パターンについてマッチングして調べねば成りません。

超平行処理が特に必須なのは、パターンマッチングですね。入力データが、記録にある無数のパターンとマッチングし、その記録の中から最もマッチングするものの ID を得るということ。それとマッチングには ID とそのマッチング度合いである確信度が返されると思うのですが、それらを入力データの全体との関連・・・データは構造をもっていますから、その全体の中での占める位置との関係から全体を推定したり、全体から部分を推定したりします・・・で統合的に集計しないといけません。これも超平行処理が必須。パターンマッチングは神経回路には得意な処理なわけですが、IT は不得意とするところです。IT 的にやるとすると VSAM 型のパターン群の管理とか、ハッシュというインデクス技術を使うかということになると思います。脳の回路網であることを IT でシミュレーションするとなると、脳回路網の有る部分の纏まった処理を CPU のある時間スパン内で実行するプロセスとして表現することになります。それが、脳全体の処理時間内で IT システムが実行するためのスレッドが幾つあれば良いかと計算すれば、汎用人工知能で脳をシミュレーションできるということではないでしょうか。・・・・ま、いろいろ考えています。いずれは、AI チップでと。

パターンマッチング処理を、近傍検索という手法に置くと、高速なパターンマッチングが実現できます。それは、動画像処理に位相を埋め込むという処方です。

例えば、線分の向きがどれだけであるかというのを、(向き、 $3\pi/4$ ) というように幾何イメージなくオントロジーを key にする 1 次元配列のなかに値を (識別子) を設定した辞書を用意します。Key は幾何構造を表わすオントロジーの数だけありますから、こうした辞書が key の数だけあります。Key に対応する 1 次元配列中の同一の識別子は一つのパターンを表現することになります。この配列群を検索してある画像要素の近傍の画像を検索できるようになり、近い程、パターンは一致するということが直ぐに認識できるようになります。

更に、領野を分けて、動画像の種類 (カテゴリ) 毎に異なる領野に保存するようにすると高速な処理が実現します。例えば、顔画像は特徴的な属性 (key カテゴリ) に分けて領野を設定します。人間の領野とか、花の領野とか有るわけです。

### 3. 音声知識の獲得

音声は、実波動というセンサー値の時系列と周波数で見る時系列の2種を考慮して実現してできるセンサーデータです。可聴周波数は、10から2万 Hz ですから、サンプリング周波数は4万 Hz で、窓は100ミリ秒で実波動をフーリエ変換して周波数空間で解析していけば宜しいということです。

まずは、音素を獲得することを考えて行かねばなりません。実波動と周波数パターンからなる特徴をもった音のチャンクを切り出して記録していくことになります。

交差法で、音波のチャンクをノードとするグラフを作っていくことで、音素群をチャンクとして抽出していきます。そのときの手続きは次の通りです。

- (1) チャンクの音波の先頭(100ミリ秒の窓)の特徴を **key** とし、その **key** を先頭に持つチャンクの名前をリストとして辞書を構成し、パターンマッチングを高速に行えるようにします。
- (2) チャンクに名前を付け **key** とし、チャンク本体の配列を **value** とする辞書を構成します。
- (3) 音波の配列を既存のチャンク群とマッチングさせて、マッチングすればそのまま、マッチングしないところは、新しいチャンクとしてグラフを成長させます。
- (4) チャンクで音波要素が近いものを一つの音素として記号化します。音素が100ミリ秒を越えるほど長いものであっても、解析の途中では断片化しますが、常に同じ音素列が生成されることから、長音素も言語として有意義な単位として一つのチャンクとなって機能していきます。

次に、単語と文法の獲得を設計していきます。

単語の獲得は、音素列をベースとする交差法で、音素獲得と同じ方法で行われます。文法は、単語の生成順序をグラフで表わし、それを単語に与えられるオントロジー素性によって抽象化していくことで、獲得が行われます。と同時に、コマンドという深層表現と言語という表層表現の対応をとります。対応はグラフとなります。

次に、文章表出の制御を設計します。

文章表出は文法の獲得と同じ方法で獲得できます。原因・結果・説明・例題というような言語表現の項目は生活の中で抽象化されて獲得し、それを文章の流れを示すグラフとして表現し、言語表出を制御します。言語表出時に、知識ベースのこの構造(コマンド群)は原因であるとか、結果であるとか評価しながらこのグラフに埋め込んで行くことで実現する

機能です。

次に音声の抑揚制御の獲得を設計します。

感情をオントロジーとして、コマンド群の生起する感情を評価し、また、抑揚の体験を沢山獲得することで、抑揚のチャンクを作り、チャンクのグラフを作り、これに当てはまるように評価結果を言語表出時に重ね合わせ、抑揚の制御を実現します。

中心周波数によらず（男性は低音、女性は高音というようなこととか、歌の音階などで、音素は不変であるという事実があります）に音素を定める為には、周波数空間での包絡線の形が同じものは同じ音素になると提示しておく必要があります。

#### 4. 言語の獲得

言語は、汎用人工知能の意志と強く関係しています。意志は行動を発し、行動をプランニングし、結果を評価して、意味のある目的を持ったものにしていきます。そのときに使うのは、イメージデータの集まりであるサイバー空間と記号で意味や行動を表現しているコマンドの集まりである知識ベースです。

静止画像からは、オブジェクトの切り出しとオブジェクトの配置や属性が定められ、その時間変化から動作が定められます。この動作を中心として、オブジェクトとその属性を引数とし、その引数の関係を格として冠を成し、その冠でオブジェクトと属性を纏めた物がコマンドです。このコマンドが意志の発現になります。

言語は、サイバー空間のイメージとそれに関連するコマンド群を意味の深層表現として表層表現を対応させるグラフを獲得することが重要な出発点となります。それは、コマンドと表層表現が対応付けられて言語獲得機構である交差法モジュールで提示されることで実現します。

更に、言語では、知識ベースの内容から、相手が求めてくれるであろう知識情報を表出していくという重要な機能があります。知識ベースの特定の部署にフォーカスを与えるのが、語ろうとしている知識です。そうしてその語りの順序を決めるのがシナリオであり、グラフ構造した知識です。

このグラフのノードは、特定のコマンドであり、サイバー空間のイメージです。そのコマンドやイメージとマッチングするフォーカスが当たった知識を取り出し、コマンドとして纏めて表層表現生成モジュールに渡します。シナリオグラフのノードは交差法によって精密化していきます。

コマンドの抽象化について設計していきます。学習開始時には、具体的なオブジェクトとか行動とか格の値でもってコマンドは構成されていきます。そのコマンド群が多く集まると交差法によって抽象化が行えるようになります。「母」とか「父」とか「兄」とかから「家族」更には「人間」といったコンセプトの下、コマンドが構成できていきます。ここを明確に制御するには、オントロジー素性というものを導入すると良いでしょう。「母」はオントロジー「人間」とか、オントロジー「家族」というような属性で意味が構成されるとするのです。もちろん、オントロジー素性も学習によって獲得していくもので、その学習は強化学習と交差法が基本手段になっています。知識ベースに「人間」というオントロジー素性のオブジェクトとコマンドがあって、「先生」というオブジェクトがあって、そのオントロジー素性に「人間」があれば、この「先生」というオブジェクトは知識ベースのオントロジー素性「人間」が振られたコマンドが適用できるということです。ここがパターンマッチング技

術の基本になります。

そうして、意志システムの発現である言語の基本はコマンドです（言語の深層表現）。コマンドは、（動詞[、格、名詞・属性詞]・・・）の形式のデータであり、他の下位のコマンドにより支えられ、最下位のコマンドは、イメージによって支えられ物です。

この意志の行動システムに関するコマンドの構築は、身体の動きの中のフォーカスが当たったものに記号を振ることで行われます。身体の部署は主格のオブジェクトとなり、身体が作用する相手のオブジェクトは目的格のオブジェクトになります。そうして、全体の動作のイメージは動詞となります。これが、基本となるコマンドの土台です。石を握ってクルミを破壊したとすると、石はなにかの（例えば **by**）格となって破壊という行動のオブジェクト（石）を結びつけます。初めは何かの記号の格を割り振るのですが、いろいろな身体動作を経験し記録していく中で、交差法によって明確な概念となっていきます。動詞とオブジェクトも精密な記号割り当てへと収束していきます。このようにして意志システムは高度な処理となっていきます。

言語の表層表現は、実際の言葉・・・文章です。身体的作用や視覚や触覚で得る世界のイメージとその時の音声という言葉が同時に汎用人工知能に与えられると、そのイメージと言語という記号が連想関係に置かれます。イメージは上記のようにコマンドを構成しますが、一方で、文章がこのコマンドに対応します。このような体験を多くしていきまると、交差法を体験に適用して、コマンドの各引数に正確に文章の単語が対応していくように言語システムを設計できます。

こうして言語が獲得されていきます。



## 5. 思考基盤知識の獲得

思考は評価機構の配下の次の3要素からなります。

- (1) 直観
- (2) 調停場
- (3) 目標管理思考

サイバー空間のデータも知識ベースのデータも位相が埋め込まれ、情報の近さという関係が構築され、意味的に近いものを直ぐに連想できるようにします。位相は共起性を時間を追って蓄えていくことで設定されます。

この連想を実現するのが直観です。近いデータ要素は多数ありますから、それらを調停場に一時的に置いて、目標とする結果に合致している物か評価します。それらを意識下で実現するのが、目標管理思考です。試行錯誤と評価を繰り返し、目標に近づいて行きます。さらに、思考や行動のプランニングや保留・再開の管理やシミュレーションして予測を評価します。そうして、目標管理思考の中で実現した思考行動のパスは記録され、将来に直観の一部にしたり、思考の高速化に寄与したりします。

なお、連想は静的に決まったデータを発火させるだけで無く、動的に決まった操作（プロダクションシステム）が起動して生成するデータであることは重要です。将棋の駒の配置の最適なものを連想するという課題は正にこの動的操作が介在した連想です。

直観の獲得は、試行錯誤で思考し行動する中で体験する原因と結果と大きな目標とを共起させ、グラフとして定着することから始まります。その定着したグラフを抽象化することによって、プロセスを作り、パターンを定め、プロダクションシステムを組織化していきます。そうして直観は完成します。

調停場は、直観の学習時に使った評価項目に関連するプロセスを連想させていくことで、形成されます。

目標管理思考は、本来、試行錯誤の過程ですが、その思考・行動は常に目標を持つ意志システムのイメージと共に提示されます。その目標に関連して思考過程を組織化（グラフで表現する）していくことで充実していきます。

## 6. 思考の実現

思考は、シミュレーションによる予測を直観という機構もとに行い、実行動としての目標管理思考を実効していくことで、実現されます。

言語の表出と思考は、共に意志システムの下で制御される能力です。シナリオに従うのを基本とします。断片的なシナリオを試行錯誤や位相空間での近いものを検索することで、長い・・・目的までに至る言語表出や思考の完遂を実現します。

言語表出も思考も、シナリオに従いますが、常に目標項目を意志の下にもって（コマンドで表現される）、逐次評価されます。評価の優先順位とか、評価項目、評価値の大きさとかは生活するなかで獲得していくオントロジーになります。そうして、その評価によって言語表出や思考のプランニングや保留・再起の制御が行われます。

目標管理思考の1つのプロセスが起動されると、その関連項目が直観によって調停場に展開されます。それらを、目標に合致しているか評価し、また、色々な関連事象と合致しているか評価します。もっとも良いと判断した評価項目で・・・有利・不利を判定し・・・次のステップをプランニングします。こうして、目標管理思考の歩を進めていくことで、思考が実現します。

評価システムの基本はセンサーデータです。センサーの中で、価値あるものと指示するセンサーと、身体に・・・意志に・・・ダメージを与えるということを指示するセンサーが重要になります。全ての評価はこの2つのセンサーで評価されます。そうして、その基本的なセンサーの上に、オントロジーとして様々なコンセプト（イメージやコマンド）が作られ、そのコンセプトを満たしているか、真逆かということの評価する1次元の配列によるデータエリアが確保され、評価が実効されます。評価は連想であり、または、ディープラーニングで得られる重み付き投票システムであります。

自律システムである汎用人工知能では、意志システムが常に動作しているように、常になんらかの評価が多重に行われていることで、外界の変化に対応しているのです。

## 7. おわりに

思考と言語は深く関係があります。構造も同じものであるとも言えます。それぞれのデータエリアであるコマンド体系は共通です。言語はコマンドという深層表現を言語表出である表層表現に対応させるものであります。思考は目的を設定して、その解を得ていく一連の行為として実現しています。思考と言語は技術的に同じ物であると言えます。

思考と言語の働きの前提となるのは、音声認識、動画認識、触覚と骨格感覚などの身体性認識であります。認識は、これらのセンサー、アクチュエータ空間内で、繰り返し同じパターンが現れることから生まれる能力です。音声は時間と空間内での繰り返しですし、動画は身体世界の空間内の繰り返し現れるというパターンです。

繰り返すパターンを獲得していくことは、時間によるイメージの重ね合わせを見ることから始まります。あるパターンはなんでも重なり、有るパターンは殆ど繰り返さない。こうしたことから、重なるの多いパターンをチャンクとして、一般的なイメージはグラフで表わされることが分かります。このグラフを作る過程が交差法です。重ね合わせ作業がクラスタリングであり強化学習です。

ここで切り出されたパターンに記号を割り振り、コマンドとして組み立て、思考、行動、言語活動をするのが意志システムです。意志の下、イメージはサイバー空間を作り、記号は知識ベースを作り、利用されていくのです。

最後に、`python` を利用してモジュールを設計していくことについて説明します。

データエリアやプロセスは領野分散処理という構造を持ちます。代表的 (オントロジーとしての) コンセプトによって領野分割は定められます。そのコンセプトに近いデータエリアがそのコンセプトが代表を務める領野に保存されます。

各領野には次の特別な管理データがあります。

- (1) 名前アドレス変換辞書
- (2) 上位から下位への名前参照がサイバー空間や知識ベースのデータエリアの基本ですが、これらを利用して行くには、下位から上位への参照機構もなければ成りません。これがインバース辞書で、領野毎に設けられます。
- (3) ファイル化管理機構

古いデータはファイル化してキャッシュである生きているデータが占めるメモリを有効に使う必要があります。`Pandas` の `Siries` とかデータエリアを使えば、それは `RDB` のテーブルとしてファイル化します。`Numpy` の配列を使う場合には通常のファイルに格納します。

#### (4) 名前の出現カウントテーブル

チャンクの獲得や、チャンクの連鎖の獲得をするのに出現カウントを利用します。また、チャンクの重畳によるクラスタリング分類学習の基盤になります。

この文書で現れた代表的なデータエリアとして、有向グラフとステートマシンとイメージとコマンドがあります。その構造を提示したいと思います。

##### (1) 有向グラフ

ノード：エッジ名リスト

エッジ：遷移先ノード名

これは2種類（2領野からなる）の辞書からなると設計できます。

##### (2) ステートマシン

ステート（ノード）：遷移名リストと内包するステートマシン名と出力記号リスト

遷移（エッジ）：遷移記号リストと遷移先ステート名

ステートマシンはステート名と遷移名リストの辞書、ステート名と内包するステートマシン名の辞書、ステート名と出力リストの辞書、遷移名と遷移記号リストの辞書、遷移名と遷移先ステート名の辞書という5つの基本辞書から構成されます。

##### (3) イメージ

基本的に配列ですが、データフレームを用いて、要素情報リストを配列形式で保存することが望まれることがあります。

##### (4) コマンド

コマンド引数のリストの合体リストとして構成します。

おわり