

記号を接地しうる表現として的大脑新皮質カラム構造の検討

Column Structures on the Neocortex as a Basis of Symbolic Representation

山川 宏^{*1}

Hiroshi Yamakawa

^{*1} (株)富士通研究所

FUJITSU LABORATOIRES LTD.

To solve the symbol grounding problem, pre-symbol concepts which can be indicated by symbols should be constructed from the external world input. Because the pre-symbol generation technologies are not achieved until now, comparing with neural mechanisms of human cognitive system will be fruitful. Pre-symbols should be static specific cluster in state space, because they are indicatable by symbols. In the state of the current technologies, some kind of information representation will promote such nature of specificity. One is a representation constrained to low dimensional space by non-linear transformation like the SOM. Another choice is the sparse representation. The hypercolumns which mainly exist on sensory area in human neocortex can associate with the sparse representation, because each hypercolumn has competitive neural mechanism inside. In future, static nature should be introduced for sparse representation such as a invariance search technology, but it is just sprout stage.

1. はじめに

記号接地問題は AI 分野の基本課題であり、主にシステムが外界の対象を、記号として記述することの困難さに係る。

言語学においては、図 1 に示すように、記号が意味する概念としてのシニフィエと、記号がとる表現形としてのシニフィアンがあるとされ、その間の関係は基本的には恣意的である[丸山 81]。よって記号接地問題は、外界の様々な概念を取り出すことでシニフィエを生成/獲得する部分問題と、そのシニフィエにシニフィアンを対応付ける部分問題からなる[山川 95]。

本稿では一つ目の部分問題を工学的な技術として解決する道筋を考える。AI の分野ではシニフィアンは記号と呼ばれるので、以後そう呼ぶことにする。そしてシニフィエを、記号に直接関係付けられる記号前概念とそれを支える多様な表現に分けて考える。そして(レフェランに対応する)外界データを加工/変換する認識処理を通じて多様な概念を生み出し、その一部が記号に接地しうる記号前概念になっていると考える。こうして記号前概念を区別すると、それが記号から直接指し示す事できるという制約から、その生成/獲得に関わる議論を進められる。

AI が人のように多様な対象に対して記号を付与できる汎用性を備えるためには、記号前概念は外界からの経験を利用して教師なし学習で自律的に獲得できる必要がある。最近はこうした技術として深層学習(Deep learning)に期待が寄せられている。

そこで本稿では、まず記号前概念が表現として満たすべき条件とその性質について検討し、それが人の脳においてどのように実現されるかを考察する。ただし記号は述語にも付与されるが、今回は、名詞的な記号を支える記号前概念に議論を絞る。

2. 記号前概念の条件

2.1 記号から指示可能であること

記号前概念を記述する表現として多次元の変数集合からなる状態空間を想定し、そこで扱うデータ中には多数の状態が存在するものとする。ここでは人が記号として理解できる状態空間としては、例えば、画像データを処理して得られる地図空間中の場所であったり、物体の形状を想定している。

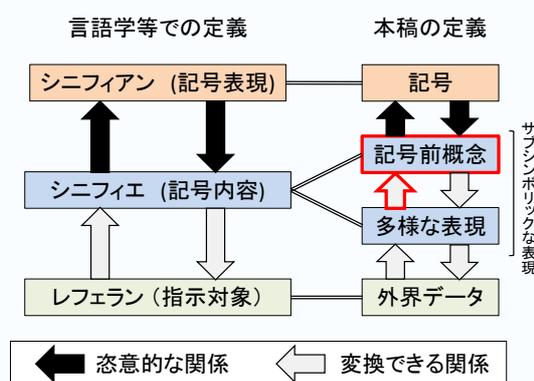


図 1: 記号前概念の位置づけ

本稿では記号前概念とその生成/獲得(赤枠)について検討。

記号が成しうる機能は状態空間内で固定された状態やそのクラスタ(以下、状態クラスタと呼ぶ)を指示することである。しかし一般に全ての個別状態に記号を割り当てると、その数は膨大となり現実的でない。そこで、指示可能な記号前概念が満たす条件は以下であると考えた。

[指示可能条件]

記号前概念は、記号から指示可能であるために、状態空間上において固定的に特定化された状態クラスタとして表現されていなければならない。

ここで、“特定化された状態クラスタ”とは、ある記号の指し示す状態の範囲である。

こうした状態クラスタを取り出しうる処理としてはクラスタリングや学習ベクトル量子化(LVQ)などがある。LVQは状態集合を有限個のセントロイド(代表的なパターン)に置き換える処理で、そのセントロイドに記号を付与してもよい。LVQはNeural Gasモデルとほぼ同じである。その学習は基本的にk-最近傍クラスタリングのように振る舞う。

2.2 有限集合を表現できること

少なくとも、否定的概念と存在論のある性質を支えるため、記号前概念は有限集合を表現できる必要がある。

まず、2つ以上の要素を持つ有限集合を表現できることにより、ある性質や状態が存在していないことを示す否定的概念の基盤を与えられるだろう。例えば、さまざまな野菜が有限集合として概念化されているからこそ、「緑でない野菜は？」と聞かれたらトマト等(補集合)を答えられ、「透明の野菜は？」という問いには無い(空集合)と答えられる。

さらに集合的表現は、存在論の基盤でもある。存在論は哲学では「存在に関する体系的な理論」、情報科学分野では「概念化(対象とする世界に存在すると考える概念とそれらの間の関係)の明示的な規約」などと捉えられ、しばしば存在論木により表現される。記号前概念が存在論の基盤となるために備えるべき性質の一つとして「整列可能な差異」という性質に着目する。

構造整列説に立脚して、事物の類似性についての評価を調べた心理実験によれば、存在論木における二つの名詞間の階層上での距離を変化させると、“共通性”と“整列可能な差異”の数は、提示された二つのアイテムの階層上での距離が開くにつれて減少する[Markman 00]。

複数の概念が整列可能であることは、互いに対置して比較すべき概念の集合が存在することを示している。整列可能な差異はその集合中の異なる要素を示すことに対応するだろう¹。

以上より記号前概念は次の条件を満たす必要がある。

[集合表現条件]

記号前概念は、有限集合を表現できる必要がある。

状態クラスタが集合の要素(元)となるためには、互いに排他的でなければならない。もしも先の 2.1 節で述べたクラスタリングや LVQ を用いれば、排他的な要素を得られるので、得られた状態クラスタを要素とする集合として扱える。

3. 特定化される表現と類似性

記号前概念が指示可能条件を満たすために、特定化された状態クラスタの生成を促進する条件について検討する。

ここでは、状態を分類するための明示的な教師が存在しない状況を想定する。すると状態を特定化するには、2.1 節で述べた学習ベクトル量子化(LVQ)のような方法を用いた状態間の近接的な類似性に基づいて状態クラスタの領域を特定する(つまり他から分離するための境界を決定する)必要がある。

当然ながら、もし状態間に類似性が存在しなければ、領域は定義できず状態クラスタは作れない。逆に隣接する空間が広すぎると、境界を作ることが難しい。つまり特定化を促進するには状態空間内での近接的な類似性を適切に制御する必要がある。

以下では複数の変数により構成される代表的な表現の性質を[O'Reilly 2012]の”Principle 9”を参考としつつ検討した。

3.1 一般的な尺度の集合からなる多次元表現

一般的な変数に用いられる尺度として、名義・順序・連続がある。特に制約を与えないこれらの尺度変数の集合により張られる多次元の状態空間を考える。

- **連続尺度変数:** 低次元の連続尺度空間は特定化に向いている。例えば2次元座標の地図上で境界を設定することは容易である。しかしその次元数に応じて近接す

る空間が指数関数的に増大するため、高次元においては多くの状態間で類似性が高まる。すると境界の設定が困難となり²、状態クラスタを作ることが難しくなる。

- **順序尺度変数:** 連続尺度と同様の理由で、高次元になると状態クラスタを作ることが難しい。
- **名義尺度変数:** 複数の状態集合から一つを指定する。この変数では、状態間の類似性は定義されない。よってこの情報から状態クラスタを作ることができない³。

以上より、独立な変数の集合としての状態空間は、低次元の連続/順序変数空間による多次元空間の場合のみが特定化に適している。しかし、一般に知能システムが扱う情報は高次元であり、記号前概念はそれを扱う必要がある。

3.2 空間に規定された表現

前記のように、単に変数の集合を扱うのではなく、空間 x を導入し、その各点における分布 $f(x)$ を表す表現がある。

典型例として、物理分野における様々な場(電場・磁場等)や地理的空間上の温度分布や、画像などの多次元センサ入力がある。ここで f はスカラー量、ベクトル量など様々な量が想定され、場を規定する空間 x は、物理的な空間座標や時間座標である。

場の表現においては、例えば画像処理における領域分割のように、稠密な空間 x を f の値に基づいて分割することで状態クラスタを得る処理がよく行われる。こうした処理は空間 x が低次元であれば、互いに近接する空間は大きくならず実用的である。

こうして、場に基づく表現では、空間 x を事前に設計することで、特定化された状態クラスタを作り、その集合を得ることが促進されるが、反面それに縛られることにもなる。つまり新たな空間をつくり出す汎用性の面では弱みがある。

ところで、先に述べた順序尺度変数が値 x を保つ場合に、一つの x のみ $f(x)=1$ で、他の x については $f(x)=0$ という場の表現に変換することはダミー変数化と呼ばれる。また連続尺度変数についても、ビンで分割することなどでダミー変数化できる。そして空間 x 上で入力データ中の出現数を累積すればヒストグラムとなり、おなじ空間 x 上で、確率分布が記述される確率空間として利用できる。

3.3 低次元空間への射影(多変量解析の技術)

高次元状態空間の全体を、低次元の表現に射影もしくは圧縮する多変量解析手法は、可視化技術としても有用である。

代表的手法である主成分分析は線形変換により値の分布全体の中で寄与度が大きい上位数個の成分のみに着目する。これはそれ以外の成分では差が大きい二つの状態を同一視する欠点や、高次元空間で曲がった分布にフィットできない弱点があり、状態クラスタの特定においては問題がある。

この問題を解決するのが、位相情報を保持しながらの非線形変換により、入力データを低次元に射影する多様体学習である。例えば、自己組織化マップ(SOM)、Isomap などがある。仮に、高次元空間中の状態分布を二次元平面にフィットした場合、類似性はその平面上に限定して定義されるので状態クラスタを作り易くなる。

例えば状態クラスタを抽出する Neural Gas モデルを SOM に組み合わせた長谷川らによる SOINN (Self-Organizing Incremental Neural Network)では画像・音声中の概念を記号に結びつけることに成功している[He 07]。

² この現象は次元の呪いの一面であると思われる。

³ ただし状態毎には出現頻度の類似性を考えることはできる。

¹ 概念が集合になっていなくても、その共通性は見つけられる。

3.4 スパース表現

脳の神経回路は高次元中で少数の神経細胞が活動するスパース表現を持つことからのヒントもあり、近年は入力データをスパースな内部表現に変換して情報処理を行う技術が注目されている。たとえば深層学習においては畳み込み層で得られる表現にスパース性を仮定することが重要となる。

スパース表現は、多数の変数の活動を競合させて、個々の状態を記述する変数を少数個(k個)程度に制限(正則化)して得られる情報表現である。一方で非負値行列因子分解のように明示的な正則化に依らずに同様の表現が得られる手法もある。

こうして表現全体の次元数は通常は入力空間より大きくなるが、各状態を記述する次元数は小さく保たれる。よって入力空間における状態間に豊かな類似性が存在していれば、写像されたスパース表現においては、多数の低い類似性を持つ状態間関係が無視され、個々の状態において近接する他状態の数が抑制される。こうしてスパース表現上では、境界を作ることによる特定化は進めやすいだろう。

この表現では、個々の状態は高々k個程度の変数によって記述されるという事前知識(制約)を導入しているが、これは SOM における制約よりもかなり緩く、高い類似性を持つ状態を引き離す可能性はかなり減少するだろう。

4. 一般多値属性としての新皮質カラム構造

記号接地可能な AI の実現に向けて、それを実現している人の脳における表現との対応付けは有用だろう。

4.1 記号前概念を蓄積する大脳新皮質

記号前概念が脳の何れの器官に蓄えられるかを考える。人は 2.2 節で述べた存在論を含む様々な概念を想起し、これに対して記号を付与できる。一方、脳において多様な表象を保持しうる器官は大脳新皮質であろう、特に人間においては視聴覚に関わる領野上での情報処理を通じて外界の多様な概念を認識している。またそれ以外に該当する脳内の器官は見当たらない、そこで、人で発達していると思われる記号前概念は大脳新皮質上のこうした領野に保持されると想定するのが妥当であろう。

4.2 汎用的な表現単位としての一般多値属性

個々の記号は対象の特定の側面を記述するので、それを支える記号前概念は多様な表現単位の集合体となるだろう。一方、3 章の議論においてスパース表現は記号前としての条件を満たしていることを述べた。そこで、コンパクトなスパース表現を記号前概念の表現単位とし、これを一般多値属性と呼ぶことにする。

こう名付けたのは、まず 2.2 節に述べたように3つ以上の集合要素を表現する機能を反映して多値と冠した。次に大脳新皮質上の表現単位には主要な尺度を統一的に表現しうる汎用性が必要と考え、一般と冠することにした。なぜなら大脳新皮質上の一般多値属性は、記号前概念としての条件を満たすと同時に、入力空間における元データを記号前概念に変換(写像)する様々な認識処理においても利用されるためである。

そこで一般多値属性を構成するコンパクトなスパース表現が、名義尺度、順序尺度、連続尺度の何れをも表現できることを以下で述べる。

スパース表現では多くの変数集合の中から k 個程度の変数が競合的に活動することで、ある状態が表現される分散表現である。二つの状態 x_1, x_2 を考えた時に x_1 で活動する変数と、 x_2 で活動する変数に重なりがあればそこに近接した類似性を定義できる。するとよく知られた多値変数の尺度は隣接関係(類似性)が一次的となる特殊ケースとみなせる。よって、

- ◆ 隣接関係を無視できれば「名義尺度」となり個物を指定できる。
- ◆ 隣接関係を一次元にした場合に、「連続尺度」や「順序尺度」になる
- ◆ 上記において、近接類似性が滑らかならば「連続尺度」に、特定化が進み相互に離散的な状態クラスタが生じていれば「順序尺度」になる。

以上より、大脳新皮質上の記号前概念の構成単位として一般多値属性を想定することが可能である。

4.3 一般多値属性を表現するカラム構造

大脳新皮質の神経回路は複雑で、未だ完全には解明されていない。しかし以下で議論するように、大脳新皮質の感覚野を構成するハイパーカラムが一般多値属性を表現すると考えた。

神経科学では、人の大脳新皮質の視覚野や聴覚野は多数のハイパーカラムにより構成されると考えられている。哺乳類は新皮質上に水平分化した 6 層構造を持つが、その 6 層構造と直交した柱状の繰り返し構造がハイパーカラムである。単純化すれば、各ハイパーカラムは図 2 に示すように、100 個程度のミニカラムの集合体である。ミニカラムは 100 個程度の興奮性細胞からなり、(ミニカラム内の局所的な抑制細胞の制御により)概ね二値を表現する¹。

各ハイパーカラムにおいては、その中のミニカラム全体から入力で活性化される抑制性のバスケット細胞が、全てのミニカラムに対して抑制的に作用する。これによりミニカラム同志の活動が競合するとことで少数のミニカラムによるスパースな興奮状態が属性値を表現する(k-Winner Take All)[Johansson 07]。つまり、ハイパーカラムは一般多値属性を表現しうる。

なおハイパーカラムにおいて、SOM のように低次元空間に状態を配置する制約は想定しづらい。よって現状ではスパース表現と看做するのが妥当であろう。

以上述べたように、ハイパーカラムは一般多値属性として機能しうる神経回路を持つので、人の大脳新皮質が、記号前概念を支えるという仮説が補強される。

4.4 記号を支えるカラム構造の進化

大脳新皮質が特に哺乳動物で進化した点から見ても、記号を支える一般多値属性の所在は大脳新皮質のカラム構造と考

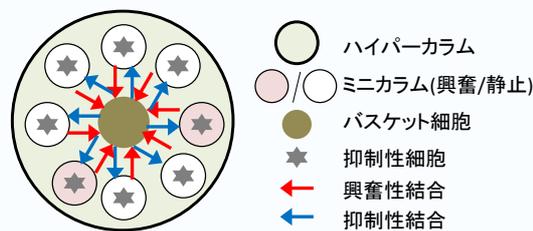


図 2: 大脳新皮質上の表現単位であるハイパーカラム

本図では、競合を勝ち抜いた 2 個のミニカラム(ほぼ二値表現)の活動により一般化多値変数としての状態を表現。
*この回路への入出力は存在するが、省略されている。

¹カラム構造と一般多値属性との対応関係についての理解を促すため、極端な $k=1$ の場合(Winner Take All)を考える。するとミニカラムと属性値が一意に対応する。この場合は名義尺度となり状態間の類似性は表現できない。これは神経科学における「お婆さん細胞」に相当する、現在はその存在可能性は低いと見られている。

えることに蓋然性があることを述べる。

個々の興奮性神経細胞の出力はスパイクと呼ばれるパルス状の電気信号である。そしてサブミリ秒区間内のスパイク頻度は出力の強度を示せると解釈できる。よって個々の興奮性神経細胞が表現できるのはさほど解像度の高くない連続値であり、一般多値属性の表現としては単純に過ぎる。

次に、哺乳動物が外界からの情報を概念化することで認識を行う主要な脳器官は、海馬もしくは大脳新皮質と考えられる。しかし海馬の興奮性神経細胞は構造化されておらず、一般多値属性を実現しているとは考え難い。次に大脳新皮質に目を移すと、ラットなどの齧歯類ではカラム等の何らかの構造は発達しておらず、ネコ等のある程度高次の哺乳類から存在する¹。

こうしてみると、一般多値属性を神経回路として実装しうるカラム構造は、記号とは無関係の何らかの理由²で大脳新皮質の特定の領野において進化的に獲得され、人において記号前概念として利用されるようになったと推察できる。

5. 不変な表現を得るために

5.1 不変性が状態クラスタを固定化する

ここまで議論しなかったが、記号前概念に必要な指示可能条件のために、状態クラスタは固定されている必要がある。

状態クラスタが潜在的に特定化しうるものであっても、それが入力空間中で移動や変形してしまえば記号により指し示すことはできない。例えば視覚情報処理においては、おなじ物体からのイメージも常にアフィン変換(並進/回転/拡大縮小)をうける。

そうした状態クラスタに対しては、異なる変数集合から得られる類似の分布を等価なものとして扱う不変性に基づき、クラスタを固定化する前処理が必要がある。

5.2 生成された空間における不変性

変数同志の互いの位置づけ(隣接関係など)を規定する空間 x が存在すれば(3.2 節参照)、その性質を活かした不変性を利用できる。例えば画像においては二次元空間を基底とすることによりアフィン変換を定義できる。

しかし SOM で得られる新たな低次元ユークリッド空間や、オートエンコーダ・ニューラルネットワークで得られるスパース表現においては、入力空間が畳み込まれている。このため新しい空間においては仮定できる不変性を特定できない。例えば SOM における隣接する任意の二つの代表点に対して何らかの共通する不変性を見出す手段がない。

つまり生成された空間では、利用しうる不変性を設定できない問題がある。現状で一般物体認識などに必要な高次特徴量を獲得できない理由は主にここに起因する可能性がある。

5.3 脳に学ぶ汎用的な不変性探索

前節の議論より、学習により新たに生成された空間において不変性を設定する手段が必要である。

このために[山川 13]は、変数集合中から、互いに等価に扱える変数の部分集合を探索する等価性構造抽出技術の研究を進めている。この技術では、”局所時間内での出現パターンが類似する部分変数集合は結合して処理しうる”という比較的一般性の高い事前知識を仮定している³。

5.4 不変性を支える均質な階層性

画像情報におけるアフィン変換のように入力データ内に何らかの不変性が存在する場合を考える。すると対応付けを行う変数集合を得るための変換処理は揃っている必要がある。

大脳新皮質においては、並列してほぼ同じ処理を実行する各領野を階層的に積み上げた構造を持ち、これが均質性を満たすことに貢献していると思われる。

6. おわりに

記号接地問題の克服を目指し、本稿では記号の接地基盤となる記号前概念の生成についての検討を進めた。この種の議論は過去にもあるかもしれないが、近年の機械学習の進展や神経科学の知見蓄積をふまえての再検討には有意義であろう。

本稿では、記号前概念は固定的でなおかつ特定できる状態クラスタであると考えた。そして特定化を促進しうるスパース表現は大脳新皮質のカラム構造に対応付けられる。そうであればスパース表現を持つ深層学習をベースに不変性を探索する技術を組み合わせてゆけば、記号前概念の自動獲得技術を実現できる見込みがあるだろう。つまり、記号接地問題の解決は手の届く範囲に近づきつつあるのではないだろうか。

なお 3 章で述べた表現技術の進歩をみると、次第に制約が緩和されることで一般化が進んでいるようである。そもそも記号前概念においては表現空間中において距離が必須であるかどうかも定かではない(ただし不変性のためには距離は有用そうに思える)。汎用的な記号前概念というものは、位相空間⁴に僅かな制約を加えたような表現になるのかもしれない。

本稿の検討は些か未完であり、今後とも先行研究を見つつ議論と熟慮を重ねたい。そして不足している理論や技術を探り当てつつ記号接地可能な AI の実現に向けて歩を進めたい。諸先生方よりご意見・ご批判などいただければ幸いです。

参考文献

- [丸山 81] 丸山圭三郎, ソシユールの思想, 岩波書店, 1981.
- [山川 95] 山川宏, パターンベースド知能システム - 学習能力から見たシンボルグラウンディング問題の検討 -, 情報統合ワークショップ(IIW-95), pp.167-175. 1995.
- [Markman 00] Markman A.B., Gentner D., Structure mapping in the comparison process, *Am J Psychol.* 113(4), 501-38, 2000.
- [O'Reilly 2012] O'Reilly, R.C. et. al., *The Leabra Cognitive Architecture: How to Play 20 Principles with Nature and Win!*. S. Chipman (Ed) Oxford Handbook of Cognitive Science, Oxford: Oxford University Press, 2012.
- [He 07] He X, Ogura T, Satou A, Hasegawa O. Developmental word acquisition and grammar learning by humanoid robots through a self-organizing incremental neural network, *IEEE Trans Syst Man Cybern B Cybern.* Oct;37(5):1357-72, 2007.
- [Johansson 07] Christopher Johansson, Anders Lansner, Towards cortex sized artificial neural systems, *Neural Networks* 20, 48-61, 2007.
- [山川 13] 山川宏, 局所多次元時系列の関係表現としての性質の実験的検討, *Proc. JSAI2013*, 3H4-OS-05c-2in, 2013.

¹各ハイパーカラムは一万個程度の興奮性細胞を必要とするため、サイズの小さな脳では実現できないことも関係しそうである。

²この理由として、神経細胞数の自乗で増える接続のボトルネック回避、集合的概念の優位性、深い階層処理等が考えうる。

³この技術は、大脳新皮質上の記憶構築に関わる海馬上の情報表

現をヒントとして提案された。

⁴集合に要素どうしの近さや繋がりに関する情報(位相, topology)を付け加えたもの。数学における空間概念。