# シミュレーションの基本となる考え方

浅川伸一 <asakawa@twcu.ac.jp>

2002 年 5 月 1 日

## 1　ニューラルネットワークの特徴

ニューラルネットワーク—あるいはコネクショニストモデル、並列分散処理 (PDP) モデルとも呼ばれる— とは脳の神経細胞 (ニューロン) の動作を抽象化した表現であるユニットの集合を用いて人間の情報処理のモデル化をめざす研究分野です。脳内では莫大な数のニューロンが互いに密接に結合されており、ニューロンのネットワークを構成しています。人間の行なうさまざまな行為はすべてニューロンの活動とニューロン間の結合の強度として表現可能であると考えるのがニューラルネットワークです。ニューラルネットワークにおける特徴を挙げるとすれば次の 3 点に要約できます (Farah, 1994; Plaut, 2001)。

分散表現:　ニューラルネットワークにおいては、知識はそれぞれのユニット集団の活性化パターンとして表現される。例えば、ある単語の意味は別の単語の意味とは異なる活性化パターンとして表現されており、類似した概念は互いに類似した活性化パターンとして表現される。

統計的構造の漸進的学習:　長期的な知識はユニット間の関係、すなわちユニット間の結合強度としてネットワーク内に埋め込まれている。ユニット間の結合強度は学習によって徐々に変化する。すなわち、学習により外界から提供される情報 (単語間の類似度や相互関係など) の統計的性質が徐々に獲得される。

相互作用:　ユニットは密接に連結されており、相互に影響しあう。すなわちユニット間の結合強度に応じて、互いに活性化パターンを強め合ったり、弱め合ったり、振動したりというような複雑な相互作用をする。

ニューラルネットワークを破壊することで言語を含むさまざまな認知機能の障害をコンピュータ上に再現できる。人間の認知機能とニューラルネットワークプログラムとを同一視し、かつ、ニューラルネットワークを部分的に破壊することと脳損傷を同一視することとによって、近年の認知神経心理学は大きく変化してきました。

## 2　ニューラルネットワークの損傷の解釈について

ニューラルネットワークを用いた脳損傷のシミュレーション研究では特定の認知機能を遂行するためのニューラルネットワークモデルをコンピュータプログラムとして実現し、構築されたニューラルネットワークの一部を破壊することによって対応する部位が損傷を受けたときに生ずる症状をプログラムの出力として表現することをめざしています。こうした研究はいわば人工脳損傷とも言える研究です。人間の言語活動を理解する上でも、あるいは実際の脳損傷患者の症状を理解するためにも、コンピュータを用いた人工脳損傷研究は重要だと思われます。倫理上の制約から実際の人間の脳を破壊して実験を行なうことは不可能だからです。ニューラルネットワークによる人工脳損傷研究は、実際の脳損傷患者を扱う神経心理学に対して強力な道具を提供していると言えると思います。

ニューラルネットワークの破壊実検、切除実験は重要な情報を与えてくれますが、その解釈には注意を要します。「ラジオから抵抗を取り外したらピーッといったからといって、その抵抗がピーの抑制中枢であるとはいえません。ほとんどの人間は噂話が好きですが、グループはある人がそのグループを離れると、その人の噂話をします。そのときその人をそのグループの噂話抑制中枢とするようなものです。

ニューラルネットワークは相互作用をする複雑なシステムとしての行動と、そのシステムが損傷したときの効果との複雑な関係を推論する手段を提供しています。そのような推論が明示的で機能論的な検証、すなわちシミュレーションによる検証が可能になったことが重要なのです。特定の認知機能の検査結果とその認知システムの障害の部位を特定することの関連は、かつて信じられていた程明らかな関係にはありません。

## 3　システムの等価性

下図のニューロン $n_1$ とニューロン $n_2$ はいずれも閾値が $1$ で単に遅延回路として働くものとし、入力がそのまま単位時間後の出力となるとしましょう。ニューロン $n_3$ は少なくとも $x_3, x_4, x_5$ のいずれかから入力を受けたときに単位時間後に発火します。ニューロン $n_4$ は二つ以上の入力があったときに発火します。したがって $x_3, x_4, x_5$ のいずれにも入力がなければ、この回路全体の出力は $x_1$ と $x_2$ とが同時に入力されたときだけ発火します。他方、もし $x_3, x_4, x_5$ のいずれかに入力があれば、この回路の出力は、$x_1$ または $x_2$ からの入力があれば発火します。すなわち $x_3, x_4, x_5$ は、このネットワークが $x_1$ と $x_2$ が AND 回路として動作するか OR 回路として動作するかを決定する制御ニューロンと見なすことができます。このシステムを実験者が外から見るとき、$x_1$ と $x_2$ はシステムへの入力ニューロンであることが

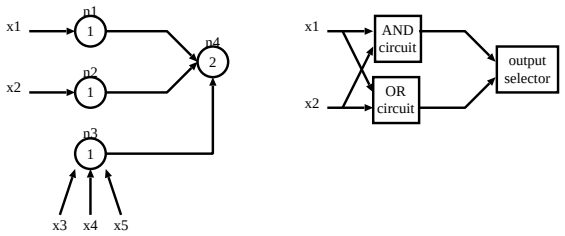


Figure 1: 同じ機能でも内部構造が異なる 2 つのシステム。入出力は両者とも同じなので、損傷実験をするか内部変数をモニターしないかぎり分離できない

分かりますが、どこか別のサブシステムから来ている AND と OR を計算する別々のサブシステムがあり、そのどちらかをシステムの出力として選択する出力セレクターがあると思って、右のようなブロック回路を描くことも可能なのです。

## 4　システムの構造と解釈

一つのシステムの行動が完全にわかったとしても、そのことから構造が一義的に決まるわけではありません。システム $S$ を分析して、それが行動的に二つのシステム $S_1$ と $S_2$ の結合とみなすことができ、またそのように分解するとシステム $S$ がよく説明できるにしても、それを持ってシステム $S$ の構造を $S_1$ と $S_2$ の二つのシステムに機械的に分解することはできません。もし心を幾つかの心的過程に分解できたとしても、その心的過程を脳の別々の部位に帰することはできませんし、その逆も成り立ちません。観察された行動のもとにある内部交互作用の詳細を説明する前に、内部構造と状態についての解剖学と生理学に留意しなければならないのです。

ある機能がないからと言って、その脳部位が無活動であるとは言えません。何かの異常によってニューロン $x_3$ が発火し続けるとニューロン $n_4$ はいつも OR ゲートになり、AND にはなりません。したがってある機能が働かないということは、その神経回路が働かないということではなくて、ある仕方で働かないということなのです。すべてのニューロンが活動し、大筋においてすべての神経路が適切に結合していても、小さな結合や閾値に異常があると異常行動となります。ラジオの「ピー音抑制」の例を考えればわかりやすいと思います。脳構造の意味づけは慎重に行なう必要があるのです。

## 文献

Farah, M. J. (1994). Neuropsychological inference with an interactive brain: A critique of the locality assumption. *Behavioral and Brain Sciences*, *17*, 43–104.

Plaut, D. C. (2001). A connectionist approach to word reading and acquired dyslexia: Extension to sequential processing. In M. H. Chirstiansen & N. Charter (Eds.), *Connectionist Psycholinguistics* chapter 8, (pp. 244–278). Westport, CT: Ablex Publishing.