

「認知科学と人工知能について」

1. 認知科学と人工知能の関係

人工知能から見た認知科学＝人工システムを作る際に応用できる知見を提供

認知科学から見た人工知能＝仮説を検証するのに必要なシミュレーション技法を提供

2. 人工知能の定義と方法論

人間の知的な働きと同じ働きを、機械によって実現することを目的とする科学技術、あるいは、そのような科学技術によって実現される具体的なシステムである。

(人工知能学会編, 人工知能ハンドブック)



知的とは何か？

工学的な立場: 人間の知能を機械で実現しようとするもの(狭義の人工知能)

科学的な立場: 人間の知能のメカニズムを解明しようとするもの(認知科学)



両者の協力関係が必要(広義の人工知能は両方を含んでいる)

記号処理の立場: 記号計算, つまり対象の記号による表現とその形式的な操作によって知的機能を実現する

並列分散処理の立場: ネットワーク構造の上でのパターン変換を情報処理の基本として知的機能を実現する(PDP, コネクショニズム)



記号処理モデルと並列分散処理モデルの統合

3. 人工知能の歴史

1940年代

計算機の誕生

計算理論についての研究

1950年代

人工知能という言葉が誕生

ダートマス会議(1956年)

定理証明, 一般問題解決器, 機械翻訳

1960年代

推論中心の時代

ゲームプログラム

自然言語処理

パーセプトロン(ニューラルネット)

1970年代

知識の時代

エキスパートシステム

知識工学の提唱

フレーム, スクリプトなどの知識表現

1980年代

人工知能の実用化

計算機の高性能化

エキスパートシステムの発展

ニューラルネットワークの復活

不確実性の扱い(ファジィ推論)

遺伝的アルゴリズム

1990年代～

情報環境の変化

データマイニング

エージェント

自律ロボット

4. 人工知能の分野

基礎分野

探索, 知識表現, 学習

応用分野

自然言語処理, 音声理解, ビジョン

エキスパートシステム, ロボット

参考:人工知能学会
のサイトより引用

5. 認知科学の定義

脳と心のはたらきを情報の概念や情報科学の方法論に基づいて明らかにし、もって生物、特に人間の理解を深めようとする知的営み
(岩波講座「認知科学」まえがき)

方法論: 神経科学, 心理学, 情報科学

分野: 知覚, 運動, 記憶, 言語, 思考, 行動, 情動, 注意, 意識

「脳と情報処理」

1. ニューロンについて

脳を構成する細胞:

ニューロン(神経細胞)とグリア細胞

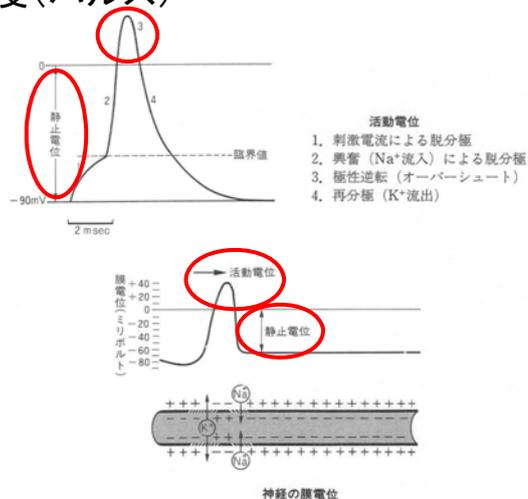
ニューロン: 約140億個, 情報処理を行う

グリア細胞: ニューロンの10倍以上の個数, 栄養供給などニューロンの活動を補助

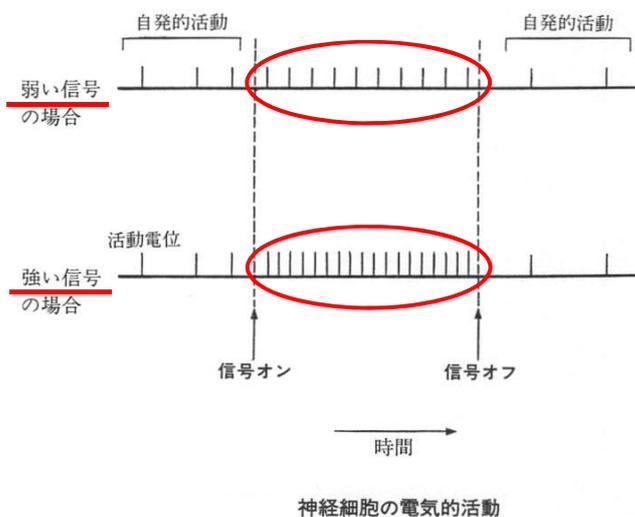
細胞膜の電位

静止電位: 静止状態での膜電位, 約 $-70\text{mV} \sim -90\text{mV}$

活動電位: 興奮により発生する膜電位, 約 $+40\text{mV}$, 数 msec程度(パルス)



伝達された信号には関係なく, 一定
ただし, 信号が強い場合や持続する場合は, 発生頻度
が高くなる



ニューロン:

細胞体

軸索(神経線維):

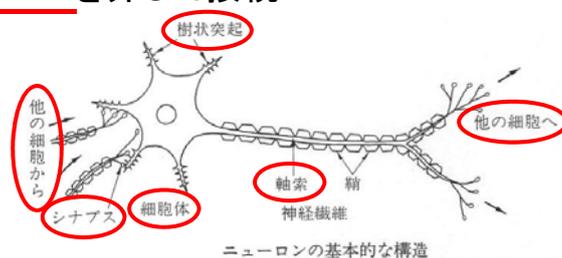
他のニューロンへパルスを送る伝送路

樹状突起:

他のニューロンからの信号を受信

軸索の終末部は樹状突起か細胞体に

シナプスを介して接続

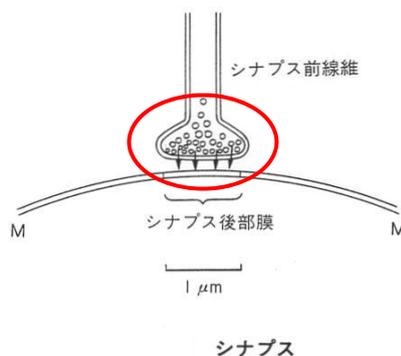


シナプス:ニューロン間の接続部位

シナプスの終末部はシナプス小頭となり, 標的のニューロンに密着

しかし, 直接は接触せず, 電氣的に絶縁

神経伝達物質:活動電位の伝達により, シナプス小胞から放出(例:アドレナリン)



シナプス後電位: 神経伝達物質により, 膜電位が変化
伝達物質の種類によって興奮性と抑制性のものがある

信号の伝達: 活動電位の場所の移動
ニューロンが持つ多数のシナプスの興奮性
および抑制性の電位が時間的, 空間的に
加重されたものが, 閾値を越えると,
活動電位が発生, 他に伝達される
神経細胞内では, 信号は変化しない

以上をまとめると…

入力: 他の多くのニューロンからシナプス結合を介して
情報を受け取る

処理: 入力の総和が, ある値を越えるとパルスが発生

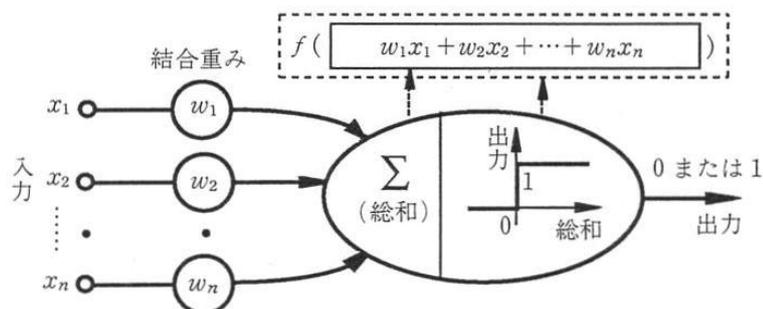
出力: シナプス結合を介して他のニューロンに情報
を送る

シナプス結合には興奮性と抑制性のものがある

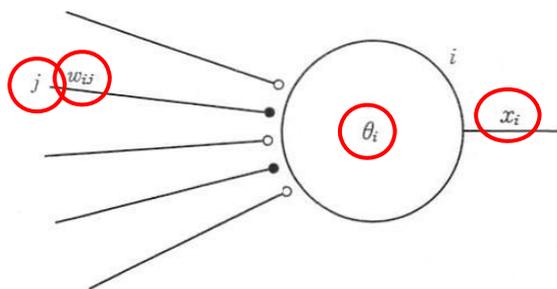
2. ニューロンのモデル

形式ニューロンモデル: (マカローとピッツ)

空間的加算と閾値処理だけに単純化したもの
しかし, 原理的な計算能力は汎用計算機と同等



モデル化されたニューロン (他のニューロンからの
重みづけ和が, 0 以上ならば 1 を出力し, 負ならば何も
出力しない)

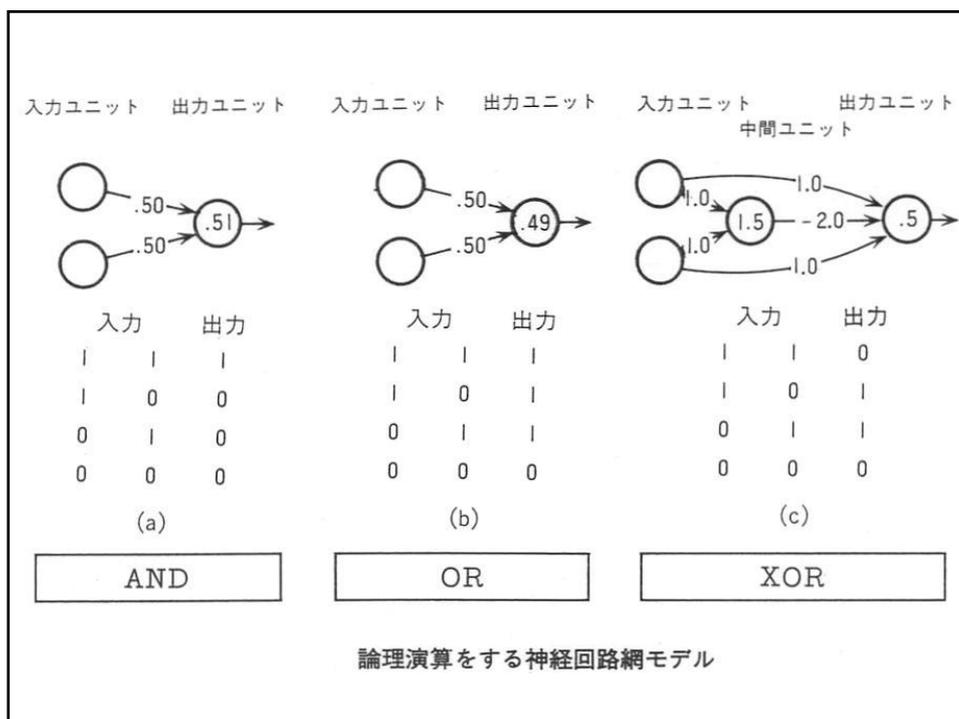


McCulloch と Pitts の論理ニューロンモデル

$$x_i(t + \Delta t) = I \left[\sum_j w_{ij} x_j(t) - \theta_i \right]$$

ただし

$$I[y] = \begin{cases} 1 & (y \geq 0) \\ 0 & (y < 0) \end{cases}$$



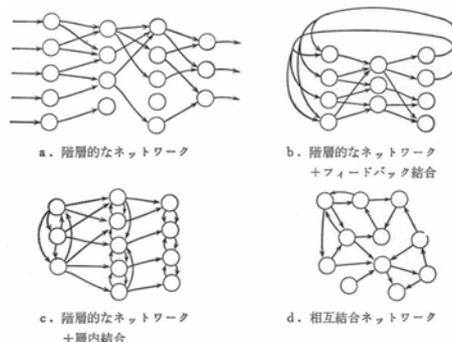
3. 神経回路網(ニューラルネットワーク)のモデル

ニューロンに対応した多数のユニットを結合させ、ネットワークを構成

階層的であるか、相互結合的であるか、

さらには、フィードバック結合を含むか含まないか、

などにより、分類される



典型的なネットワークの形

2つの観点

ニューロン間をどのように結合するか

(アーキテクチャ, 自己組織化)

ニューロン間の結合の強さをどのように決定するか

(学習)

実験的神経科学

神経生理学, 神経解剖学, 神経生化学, 神経学などの
基礎医学や神経心理学, 臨床神経学などによるアプ
ローチ

機能局在の研究

全体論と対立

研究方法

脳疾患の臨床観察

破壊実験

非観血的活動観察(PET, MRI, SQUIDなど)

精神物理学

研究方法

脳を情報処理機械とみなし, その機能を
ブラックボックス的に解析

不良設定問題

網膜でとらえた視覚信号に基づいて
視覚世界を推定する過程には,
計算不能な問題が含まれる
(色認知や立体視など)

情報表現の研究

反応選択性

視覚系の神経細胞は, ある特定の図形に
選択的に反応する特性をもつ

シナプス可塑性

学習により獲得, 神経細胞活動により
シナプス結合が変化

シナプス可塑性→反応選択性→認知

シナプスの可塑性

記憶・学習の本質: 脳の可塑性

シナプス結合の可塑性: 新しいシナプスの形成

シナプス伝達の可塑性: 伝達効率の増加(長期増強)
と減少(長期抑制)

ヘブの学習則:

シナプス前細胞からシナプス後細胞へのシナプスが
活性化され、同時にシナプス後細胞が興奮したとき
のみ、このシナプスが増強される、
つまり、伝達効率(シナプス荷重)は増加する

ニューラルネットワークの学習

教師信号の有無で分類される

教師なし学習の例:

ヘブ型学習アルゴリズム

ヘブの学習則に基づく学習

教師あり学習の例:

誤差逆伝搬法(バックプロパゲーション)

学習アルゴリズム:

教師信号と実際の出力の誤差を用い、

この誤差が0になるように結合荷重を学習する

ニューラルネットワークの研究

理論的神経科学の立場

生理学的・解剖学的に厳密な脳神経系のモデルを構築する

コネクショニズムの立場

厳密な脳神経系のモデルは求めず、心のモデルを構築する

実用化の立場

情報処理分野で応用できればよい

ニューロコンピュータの研究開発

脳神経系のモデルをヒントにした新しいコンピュータを開発する

理論的神経科学

神経回路や脳に関する理論やモデルに基づく研究

脳の計算理論(マー)

脳を理解するための3つのレベル

計算理論のレベル:

知的機能を脳が解いている計算問題としてとらえ、それを明らかにする

表現とアルゴリズムのレベル:

その計算問題がどのような計算手続きで、またどのような表現形式を用いて解かれているかを知る

ハードウェアのレベル:

そのアルゴリズムを実行する脳の演算装置の構造を知る

3つのレベルを統一的に行うこと、つまり、脳の実験的科学で明らかにされた脳のハードウェアの知識を拘束条件とする理論的研究、逆に計算論的予測に立脚した脳の実験的研究が必要である

心の研究の流れ

ヒポクラテス(BC4世紀)

心の座は脳にある

デカルト(17世紀)

心身二元論:心と体のはたらきを完全に分離

心=自由意思, 理性, 創造性→科学の対象ではない

体=機械の原理, 刺激と反応→科学の対象となる

行動主義心理学(20世紀初頭~)

人間や動物の行動を調べるだけでよい

行動=外界の刺激に対する特定の反応

行動について科学的に研究するのに, 心の概念を用いる必要はない

実験室での「条件付け」を観測, 法則化する

ゲシュタルト心理学(20世紀初頭～)

高等動物の行動は刺激と反応では説明できない面が多い

思考作用を対象とする

行動＝外界の刺激に対する合理的な適応

内観(感じたことを言葉で表現させる)を基礎とする

認知心理学(1950年代～)

心を脳のソフトウェアと考え、情報処理的アプローチをとる

モデル化とシミュレーションによる分析

「記憶と学習」

1. 記憶と学習の定義(心理学的立場)

記憶と学習

両者は非常に密接に関連, 厳密に区別することは難しい場合もある

学習

経験がのちの行動に影響するようなかたちで「こころ」の中味に比較的永続的な変化を生じさせること

知識構造に比較的永続的な変化を生じさせること(認知科学的立場)

※学習が成立するためには記憶が必要

記憶

(広義) 経験の効果を時間を越えて存続させるもの

(狭義) 「情報を蓄える」(受動的)

「おぼえる」・「思い出す」(意図や意識を伴う)

3つの側面: 記憶機能, 記憶システム, 記憶情報

2. 記憶

記憶の過程:

記銘(符号化), 保持(貯蔵), 想起(検索)

短期記憶(一次記憶)

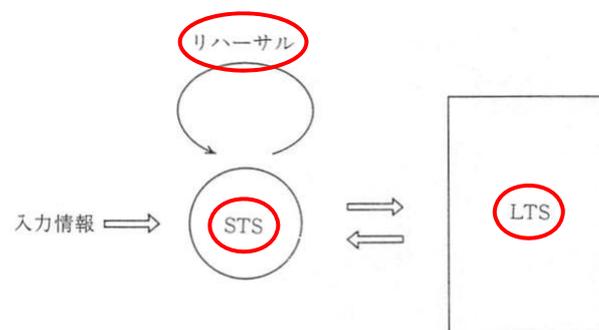
意識として心の中に留めている内容

長期記憶(二次記憶)

必要に応じて意識化されたり, 無意識的な行動として現れたりする過去の膨大な記憶

2貯蔵庫モデル

短期貯蔵庫と長期貯蔵庫



貯蔵庫モデルの基本形

リハーサル

短期記憶の容量は小さく, リハーサルしていないと減衰する

しかし, リハーサルを繰り返した情報は長期記憶となる

維持リハーサル

短期記憶貯蔵にとどめておくリハーサル

精緻化リハーサル

長期記憶貯蔵への転送の原因となるリハーサル
(関連付けやイメージ化)

「マジカルナンバー7±2」

短期記憶の容量は, 項目あたりの情報量によらず, 7項目程度

宣言的記憶

さまざまな「事実」に関する記憶

手続き的記憶

作業を行うときに参照する「やり方」に関する記憶

意味記憶

単語の意味や概念などに関する記憶, 一般的な記憶

エピソード記憶

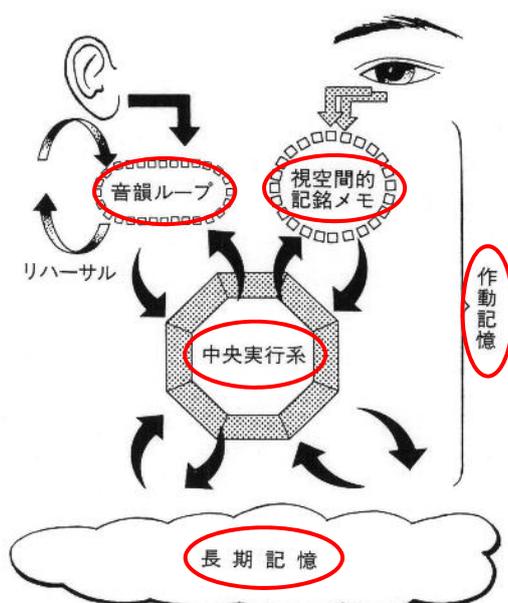
特定の時間や場所などの文脈情報を含む, 個人が過去に経験した出来事に関する記憶

作動記憶(作業記憶)

短期記憶の概念を発展させたもの

認知過程での情報の処理機能を重視する

記憶以外の認知機能との関わりを捉えようとする



バドリーの作動記憶のモデルの概略図 (Baddeley, 1986, 1992; 図は Logie, 1995 に準拠)

3. 学習

例からの学習(帰納的学習)

解説よりも例そのものから学ぶ方が分かりやすいことがある

- ・例そのものを記憶し, 類似事例に利用する場合
- ・例から一般的な法則を抜き出し記憶し, 利用する場合

説明による学習(演繹的学習)

例だけで学習することはまれで, 説明が伴うことが多い

- ・背景となる知識などについての説明を利用する場合
- ・関連する既存知識を利用する場合(自己説明)

4. 記憶と学習のモデル(情報处理的立場)

記憶情報の表現

記号表現

記号処理モデル: 1つの対象を1個の記号に対応させる

思考や言語に関わる記憶・学習のモデルに適している

パターン表現

パターン情報処理モデル: 多くの情報単位からなるパターンの集まりで表現する

局所表現

1個のニューロンによって1つの対象を記憶する
→認識細胞(おばあさん細胞)モデル

分散表現

多数のニューロンに分散して情報が保持される

スパース表現

一部のニューロン群にわたって表現される

神経回路網(ニューラルネットワーク)による記憶と学習
のモデル化

ニューロンの可塑性が脳の記憶機能の根拠