AIとゲノムとビッグデータ シンポジューム

人工知能と社会

甘利俊一

理化学研究所 栄誉研究員
東京大学 名誉教授

人工知能の衝撃

人間の知的能力を超えるのか？
第4次産業革命？

人工知能の歴史：
記号と論理（知能をプログラムする）
並列分散（ニューラルネットで学習する）

動力技術
機械技術　　技術は止まらない！
材料技術

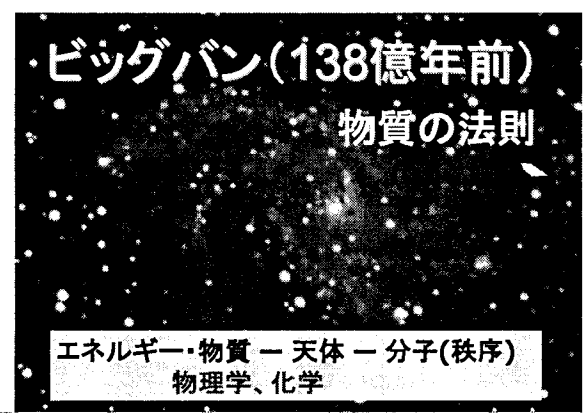
生命技術ー遺伝子編集
情報技術ー知能・心（人のやること）

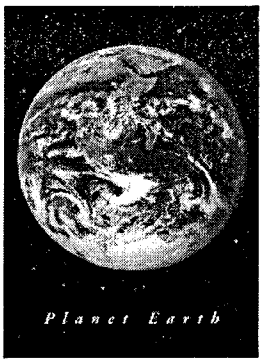
脳ー人間とは何か

思考・言語・意識
人間・社会・文明

宇宙誌と脳　ー 脳ができるまで

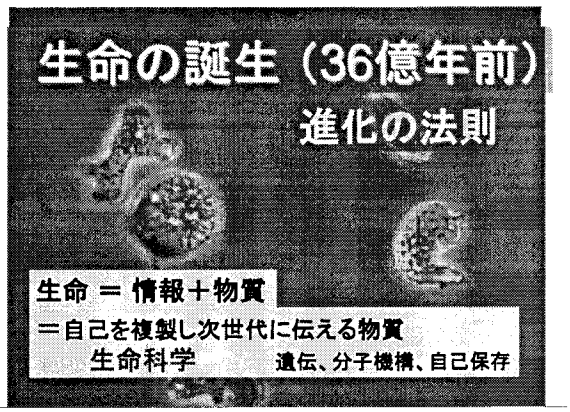
ビッグバン	(138億年前)	物理学・化学
生命	(36億年前)	生命科学
脳・神経系	(5億年前)	神経科学・情報科学
文明・社会	(20万年前？)	脳科学・情報科学・人間科学

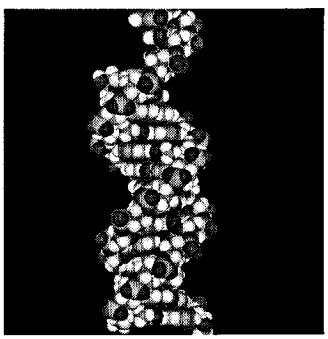




46億年前:
地球の誕生

火の玉、全球凍結





DNA

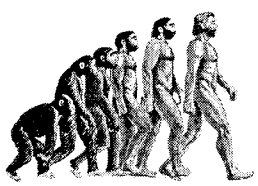
多細胞生物　　環境の情報を利用、
記憶・学習、判断・行動

脳＝情報

脳・神経系（5億年前）：神経科学

類人猿、そして人間
社会に生きる生命；文化と社会

人類の登場（700万年前）



心、意識
文明

猿人、原人
旧人（ネアンデルタール、50万年-3万年前）
新人（ホモ・サピエンス、20万年前-現在）

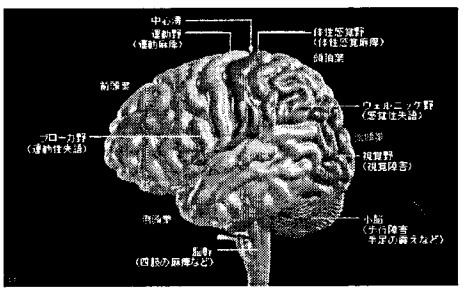
物質の法則:　宇宙

生命の法則:　情報＋物質:　進化

文明の法則:　こころ＋情報＋物質:
社会･文化

脳:大脳、海馬、小脳、脳幹

脳科学:ミクロ－マクロ、理論:神経回路網



人工知能と脳のモデル:

―歴史の要約

第一次ブーム

1956~　AI　　脳モデル
Dartmouth 会議　Perceptron
記号と論理　学習する普遍計算機構
知的推論、ゲーム　線形分離可能
実用的でない!!

暗黒期 (1965後半~1970's)　stochastic descent learning (1967) for MLP

第2次ブーム

1970~　AI
エキスパートシステム
(MYCIN, DENDRAL)

1980~　BT (神経回路)
MLP (backprop)
連想記憶モデル、ダイナミックス
一兆円産業か?

沈静化　確率Bayes推論
chess (1997)

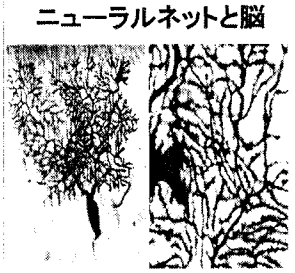
第3次ブーム　2010~　脳型の人工知能(融合)

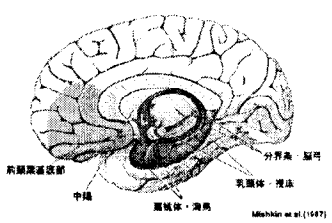
深層学習 Deep learning
(畳み込み多層回路(福島)＋確率勾配降下:日本でなぜ実現しなかったか)
確率推論 (graphical model; Bayesian; WATSON)

深層学習の勝利 ――人間以上の識別能力
パターン認識: vision, auditory, sentence analysis
囲碁:強化学習
時系列とダイナミックス、動的パターン:言語処理

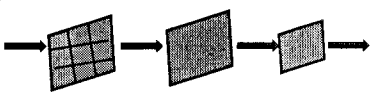
記号と論理　vs　パターンとダイナミックス、学習ー融合

ニューラルネットと脳



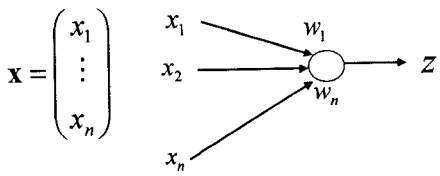


多層パーセプトロン
表現 ダイナミックス 万能性

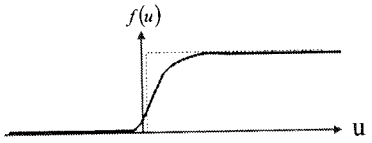


ニューロンの数理モデル

単純モデル



出力関数

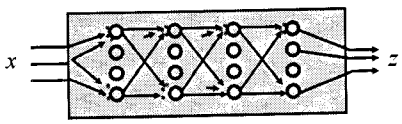


$$z = f(\mathbf{w}\cdot\mathbf{x} - h) \qquad \mathbf{w}\cdot\mathbf{x} = w_1x_1 + \cdots + w_nx_n$$

$$u = \mathbf{w}\cdot\mathbf{x} - h$$

層状学習回路網

multilayer perceptron



パーセプトロン Perceptron
バックプロパゲーション Backpropagation

$$L(x,W) = |y - g(x,W)|^2$$

$$w \to w + \Delta w, \quad \Delta w = -c\frac{\delta L(x,W)}{\delta W}$$

深層学習

大量のデータ、計算力

入力を基に正解：内挿・外挿に過ぎない

実験式

現象の予測
日蝕の予測
ケプラーの法則、ニュートン力学

原理の創出・理解ー人間

なぜうまく行くのか：原理がわからない

結果の説明ができない

敵対的例題

囲碁の成功：他のゲームでも（汎用）

深層学習＋強化学習

グーグルの野望

数理脳科学は脳の基本原理を探求する

単純な基本モデル用いる：数理的探索（現実とは違う）

→ 計算論的神経科学
（脳はいかにこの原理を実現したか）

→ AI：技術による原理の実現 （脳とは違う）

脳は基本原理をどう実現したか

進化によるランダムサーチ
使える材料の制約
歴史的な制約

ごたごたの設計の中で精妙な実現：超複雑

人工知能は何をどう実現するか？

人工知能は脳に何を学ぶのか： 心
意識と無意識のダイナミックス

記号 --- 興奮パターン
論理的推論 --- 並列ダイナミックス

AI NN

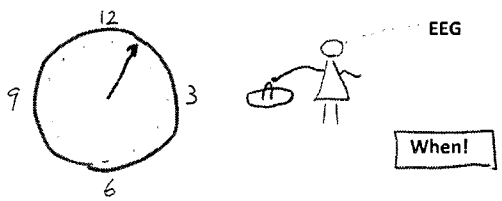
意識の発生

共同作業、自分の意図を自分で知る

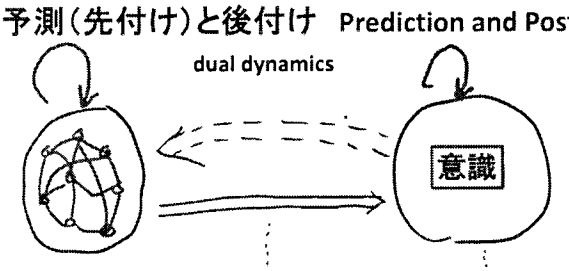
言語： 論理的思考、数学

自己言及

Libet の実験：自由意志



予測（先付け）と後付け Prediction and Postdiction



ダイナミックス 意思決定と行動 反省、正当化、論理

意識と心の役割

心を知る・美、情熱

進化

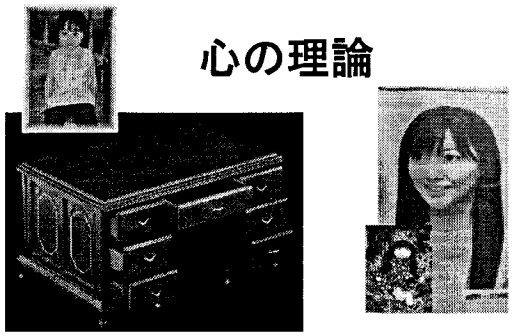
人工知能が脳に学ぶべきこと: 数理的理解

認知;決定;制御;記憶;学習

意識と心の役割; 後付け

連想式記憶システム:知識体系

心の理論



葛藤する心

究極のゲーム
（ultimatum game）

10万円

A: 配分を決定する---- 7万円–3万円
B: 同意または拒否

この時の脳の働き、各領野の確執
利益、公正、その後のこと、評判
二人か社会ゲームか

心を持ったロボットがつくれるか？

人の心の動きを理解する

ロボットが心を持つように見える
（感情移入）

ロボットは合理的（心を持たない方が良い）

ロボットが心を持てるのか？

人間の心:進化の産物

意識、意図、論理、感情

種の生存と個の不合理、不条理 :使命感
喜び、悲しみ、芸術、愛、苦悩、恋愛、倫理観、使命感、宗教、芸術
ただ一度の、かけがいのない人生

ロボットは合理的

人工知能と倫理

人工知能の安全性、制御可能性
暴走: 人間の暴走を範として

金もうけと権力支配
人工知能と戦争;人工知能の金融支配;

支配の道具、格差の拡大

社会への影響: 技術は止まらない:制御できるか

失業問題: AIは仕事を奪うか? 人口減
より高度な仕事を創出

格差の拡大:
ベーシックインカムと人類の家畜化:働く喜びを

人工知能と技術的特異点 2045

人工知能が人間を超えるとき
人工知能が研究し、技術を進める

人間は素晴らしいが、愚かである。

人間はどんな知能システムを作るのか? 正義?
社会の進化と支配

人工知能と未来社会の設計

深層学習を超えてAIは進む
科学研究、技術開発一止められない

社会、文明 その脆弱性・崩壊

我々は何をなすべきか?

地球環境
資源、エネルギー
人工
AIは解決できるのか

社会制度

民主主義——自由と平等
選挙制度の問題か
教育、教養、豊かさ

資本主義vs共産主義
修正資本主義:格差の是正(税制)、社会主義

技術進歩: 種としての人類: 地球のあふれて収奪する

日本の AI の進むべき道：政府の戦略
ブームは終わる

超大国 ⟺ 文化国家

研究開発：物量作戦はだめ
理論とアイデア
中小企業を含む現場との交流；産業の情報化