駒澤大学心理学特講 III シラバス
—ディープラーニングの心理学的解釈—

浅川 伸一 *

What I cannot create, I do not understand.

—*Richard Feynman*

1　科目の基本情報

- 履修コード/科目名称:　074401 / 心理学特講Ⅲ A
- ディープラーニングの心理学的解釈
- 開講年度・期:　2019 年 前期
- 開講曜日・時限: 金曜日 1 限
- 単位数　2
- 担当者: 浅川伸一 asakawa@ieee.org
- オフィスアワー:

2　概要

本授業では人工知能に用いられる技術の詳細を検討しながら，その心理学的意味を考えることにします。最終的な目標としては人間と機械の２つ知性はどうすれば構成可能であるかを議論するための素地を提供することを目指します。自動運転が可能となり，囲碁の世界チャンピオンを破り，自動翻訳の精度が向上し，スマートスピーカーが普及するなど AI 技術は毎日のように報道されています。これらの技術はニューラルネットワークモデルに基づいています。とりわけディープラーニング (深層学習) 技術は現在の人工知能の根幹をなしています。現在は第 3 次ニューラルネットワークブームと呼ばれますが 3 度のブームとも心理学者が火付け役でした。2014 年から始まった現在のブームも心理系出身の研究者が先導しました。加えてディープマインドの共同創設者デミス・ハサビスは認知科学出です。このように人工知能と心理学とは同じことを別の側面から理解しようとしているとさえ言えます。このような背景を考慮すれば知的活動の解明を目標とする諸分野において心理学学徒の貢献が期待されていると考えます。

3　到達目標 (ねらい)

深層学習 (ディープラーニング) についての基礎的事項を理解し，心理学との関連を考えるための資料，状況を理解することを目標とします。

4　スケジュール (案)

- Apr.12 イントロダクション，生理学的背景，ニューラルネットワーク，情報理論，サイバネティックス
- Apr.19 人工知能の歴史，計算論的神経科学 vs. 認知科学 vs. 人工知能
- Apr.26 機械学習概論
- May.10 休講

* asakawa@ieee.org

- May.17 パターン認識，パーセプトロン，パンデモニアム，文字認識，画像認識
- May.24 一般画像認識，顔認識，動画認識，意味的画像分節化，畳込みニューラルネットワーク
- May.31 誤差逆伝播法，多層化の工夫，内部表象，表現学習，次元圧縮
- Jun.07 休講
- Jun.14 休講
- Jun.21 脳のモデル，作動記憶，手続き記憶
- Jun.28 リカレントニューラルネットワーク，自然言語処理，系列予測，自動翻訳，文章要約
- Jul.05 単語，文章埋め込みモデルによる意味論
- Jul.12 強化学習, ゲーム AI，経済学，予測報酬誤差
- Jul.19 画像認識と自然言語処理との融合，質疑応答生成，転移学習，マルチタスク学習
- 補講: メタ認知，メタ学習，ハイパーパラメータの自動調整
- 補講: 世界知識，メンタルモデル，メンタルシミュレーション
- 補講: 精神医学 (統合失調症，強迫神経症，依存症，幻覚幻聴)，神経心理学 (意味痴呆，相貌失認，失語，失行)

5 準備学習

可能な限り，事前に必要な情報を提示する予定です。質問は担当者へのメールまたは SNS や掲示板などを介して行ってください。

6 履修上の留意点等

履修制限は設けません。どなたでも履修できます。授業中に検索したり資料を閲覧するために，可能な限り PC を持参してください。スマートフォンでは代用が難しい場合があります。

7 成績評価の方法

- 試験 60% (80%)
- レポート 20% (40%)
- 小テスト 20% (40%)
- 平常点 0%

8 評価基準

- 96-100: A
- 91-95: A-
- 86-90: B+
- 81-85: B
- 76-80: B-
- 71-75: C+
- 66-70: C
- 61-65: C-
- <60: F

9 教科書/テキスト

Web 上で公開予定です。各自でダウンロードするなどしてください。

10　参考書

- 『ディープラーニング、ビッグデータ、機械学習 あるいはその心理学』(新曜社, 2014)
- 『Python で体験する深層学習』(コロナ社, 2016)
- 『人工知能学大事典』(人工知能学会編、共立出版 2017)
- 『深層学習教科書 ディープラーニング G 検定 (ジェネラリスト) 公式テキスト』(監修：日本ディープラーニング協会, 共著, 翔泳社, 2018)

11　事前知識

人間や機械の知性に興味があることです。事前知識は必要としません。心理学で使われている統計的推論の概要を知っていると良いとは思われますが，必要な知識ではありません。

12　持参した方が良いもの

PC やタブレットがあるとデモを閲覧するために便利でしょう。また，不明な点はその場検索して調べることができるように授業中の PC，タブレット，スマートフォンの使用は歓迎します。加えて，授業中の質問を SNS などを経由して受け付けることも考えていますので，twitter や Facebook のアカウントがあると良いでしょう。ただしどちらも必ず必要というわけではありません。

講義資料はスマホで閲覧可能だと思われますが，制限がある場合もあります。

13　数学的知識

数式は出てきます。ただし式の意味が不明でも，どのような意味であるのかをわかりやすく解説するようにします。完全に理解するためには，次のような分野の知識が必要です。

- 線形代数
- 解析学
- 確率論
- 統計学
- 情報理論

おおよそ理工系大学 1, 2 年生の履修範囲だと考えてください。詳しく学びたい方のために参考文献を挙げるとすれば，学習院大学田崎晴明先生の数学：物理を学び楽しむためにをお勧めします (田崎, 2018)。

その他に文系向け数学入門書としては，古い本ですが林周二先生の https://www.amazon.co.jp/dp/4121001397 をお勧めします (林, 1967, 1968)。

14　デモサイト

- Google Neural Networks Playground
- 芸術家のための機械学習 Machine Learning for Artists
- 上の YouTube
- 博士号なしのテンソルフローとディープラーニング Learn TensorFlow and deep learning, without a Ph.D.
- プログラマーのための実践ディープラーニング Practical Deep Learning For Coders, Part 1

15　取り上げる話題

1. 生理学的背景，ニューラルネットワーク，情報理論，サイバネティックス Interpreting Deep Neural Networks using Cognitive Psychology, Cognitive Psychology for Deep Neural Networks: A Shape Bias Case Study
2. 人工知能の歴史，計算論的神経科学 vs. 認知科学 vs. 人工知能
3. 機械学習概論
4. パターン認識，パーセプトロン，パンデモニアム，文字認識，画像認識
5. 一般画像認識，顔認識，動画認識，意味的画像分節化，畳込みニューラルネットワーク
6. 誤差逆伝播法，多層化の工夫，内部表象
7. 表現学習，次元圧縮
8. 脳のモデル，作動記憶，手続き記憶
9. リカレントニューラルネットワーク，自然言語処理，系列予測，自動翻訳，文章要約
10. 単語，文章埋め込みモデルによる意味論
11. 強化学習, ゲーム AI，経済学，予測報酬誤差
12. 画像認識と自然言語処理との融合，質疑応答生成，
13. 転移学習，メタ認知，メタ学習，マルチタスク学習，ハイパーパラメータの自動調整
14. 敵対生成学習，生成モデル，ベイズニューラルネットワーク，ベイズ推論，変分推論
15. 世界知識，メンタルモデル，メンタルシミュレーション
16. 意味記憶，作業記憶，反応時間
17. 認知発達，言語獲得，推論
18. 精神医学，統合失調症，強迫神経症，依存症，パレイドリア (幻覚幻聴)
19. 神経心理学，意味痴呆，相貌失認，失語，失行，
20. 社会，倫理，法律

1. 生理学的背景とニューラルネットワーク

   ■キーワード:　ANN, BNN, 計算論的アプローチと制約

   ■概要

   ■biblio　(Kriegeskorte & Douglas, 2018), (Dayan & Abbott, 2001), (Poggio, Torre, & Koch, 1985), (Hubel & Wiesel, 1959, 1962, 1968; Hubel, 1963; Livingstone & Hubel, 1988), (Hartline & Ratliff, 1954, 1957, 1958),(Poggio et al., 1985),(Edelman, 1997)
2. 人工知能の歴史

   ■キーワード:　GOFAI, symbolic AI, Enbodiment, Computational approach, neuveu AI, ata science, web ontology

   ■概要　記号処理的 AI, 計算論的モデル

   (a)　認知科学 対 計算論的神経科学 対 人工知能

   Cognitive science

   Computational neuroscience

   Artificial intelligence

   ■biblio　(Brooks, 1990, 1991; Russell & Norvig, 2003)
3. 機械学習概論

   ■キーワード:

   ■概要

■biblio

(a) 教師あり学習 (分類，回帰)，教師なし学習，次元削減，クラスタリング

(b) SVM, linear regression, logistic regression.

(c) Random forest, k-means, Nearest Neighbor, PCA, tSNE, SOM, MF(NMF),

(d) Bias/Variance trade-off, Boosting and ensemble mehthod, XGboost

(e) 尤度，entropy, information theory

(f) 最適化手法: 勾配降下法，ニュートン法，SGD momentum(Bottou & Bousquet, 2007), Nesterov momentum(Nesterov, 1983), BGFS(Broyden–Fletcher–Goldfarb–Shanno algorithm)

(g) PCA(Pearson, 1901), SOM(Kohonen, 1985), tSNE(Maaten & Hinton, 2008)

(h) Eigen face, Fisher facePeter N. Belhumeur & Kriegman (1997), Subspace models(Watanabe, 1969), (Fisher, 1936)

4. パーセプトロンとパンデモニアム (Rosenblatt, 1958; Selfridge, 1958)，認識モデル, パターン認識

■キーワード: Peceptrion, Pandemonium, Epistemology, pattern recognition, マッカロックとピッツの形式ニューロン (McCulloch & Pitts, 1943)

■概要 パーセプトロン (Rosenblatt, 1958),

5. 文字認識，画像認識

■キーワード: CNN, dropout, batch normalization, data augmentation, 活性化関数 : sigmoid, tanh, ReLU, softmax

■概要

■biblio (Watanabe, 1969)

6. 物体認識，一般画像認識，畳込みニューラルネットワーク

■キーワード: ANN, BNN, CNN, DNN, Neocognition(Fukushima, 1980; Fukushima & Miyake, 1982), R-CNN, ResNet, Capsule Net

■概要

(a) dropout, data augmentation, batch normalization,

(b) LeNet, Alex Net, GoogLeNet, VGG, Resnet

(c) R-CNN, ResNet, highway net, YOLO, SSD

(d) Semantic Segmentation

(e) One pixel attack

■biblio (Krizhevsky, Sutskever, & Hinton, 2012; LeCun, Bottou, Bengio, & Haffner, 1998)

7. 誤差逆伝播法

■キーワード: BP, BPTT(Bodén, 2002), batch normalization

■概要

■biblio バックプロパゲーション (Rumelhart, Hinton, & williams, 1986)

8. 内部表象，表現学習, 次元圧縮，PCA, tSNE，NMF, MF, SOM

■キーワード: Represenation learning

■概要

9. 脳のモデルとオライリーモデル，二重記憶モデル

■**キーワード**: ANN, BNN, 計算論的アプローチ

■**概要**

■biblio (O'Reilly, Munakata, Frank, Hazy, & Contributors, 2012; Randall C., 2006)

10. 強化学習 (Sutton & Barto, 1998), ゲーム AI，経済学，予測報酬誤差

■**キーワード**: 方策学習, Q 学習, モデルフリー学習

■**概要**

(a) エージェントと環境。状態，MDP
(b) 方策，価値，モデル，Q 学習
(c) DQN, Bellman 方程式，Game AI
(d) self play, A3C, Rainbow, World model

11. 言語モデル，リカレントニューラルネットワーク自然言語処理

■**キーワード**: 系列予測，自動翻訳，RNN,
(a) vanishing/exploding gradient
(b) Grad clip, grad check, BiRNN, QRNN, lasabore net, ESN
(c) BPTT, BiRNN, QRNN, LSTM, GRU
(d) Language models, NER, BLEU, NER
(e) Kalman filter (particle filter), AR(ARMA, ARIMA).
(f) Semantics or meaning(word2vec, GolVe)
(g) TDIDF, LSI, LSA, topic model
(h) 注意 (seq2seq, transformer),
(i) 機械翻訳, sentence vector(ELMo, BERT, Tough-to-beat)
(j) sacre BLUE, negative log likelihood, perprexity
(k) NIC, MS-COCO, VQA
(l) NTM
(m) 単語埋め込みベクトルによる意味表現
(n) 文章埋め込み表現によるマルチタスク学習
(o) ELMo, BERT

■**概要**

12. 意味論

■**キーワード**: word2vec, SD 法，Topic Model, LSA(LSI), MF, NMF

■**概要**

13. 転移学習，メタ学習

■**キーワード**: Transfer learning, Meta-learning, One-shot learing, 自叙伝的記憶，エピソード記憶,

14. 画像認識と自然言語処理との融合

■**キーワード**: NIC，skip-thought, embedding, representation learning

■**概要**

15. 敵対生成学習，生成モデル

■**キーワード**: GAN, WGAN, beta GAN, VAE, EM, ELBO, KL divergence

■**概要**

16. 世界知識，メンタルモデル，メンタルシミュレーション
17. 発達

■**キーワード**: クワインのパラドクス，形状バイアス，語彙爆発，一撃学習，エピソード記憶，一撃学習に拠る顔認知

■**概要**

18. 精神医学，統合失調症，強迫神経症，依存症，パレイドリア (幻覚幻聴)

■**キーワード**: ニューロモデュレーション，神経伝達物質，フリーエナジー原理

■**概要**

■biblio q Friston, Stephan, Montague, & Dolan (2014), Friston, Kilner, & Harrison (2006),Friston (2007)

19. 神経心理学

■**キーワード**: ，意味痴呆，相貌失認，失語，失行，

■**概要**

■biblio 岩田 (1996); Shallice (1988); 山鳥 (1985) Warrington (1975); Warrington & McCarthy (1987)

20. disentagled，autoML

■**キーワード**: Disentagled，autoML，auto-sklearn，変分推論 VAE，EM，Variational EM，KL divergence，ELBO(Blei, Kucukelbir, & McAuliffe, 2017)

■**概要**

21. 社会，倫理，法律

■**キーワード**:

■**概要**

引用文献

浅川 伸一. (2001a). ニューラルネットワークの数理的基礎. 守 一雄・都築 誉史・楠見 孝（編）『コネクショニストモデルと心理学』 (pp. 166–203). 京都: 北大路書房.

浅川 伸一. (2001b). 脳損傷とニューラルネットワークモデル—神経心理学への適用例—. 守 一雄・都築 誉史・楠見 孝（編）『コネクショニストモデルと心理学』 (pp. 51–66). 北大路書房.

浅川 伸一. (2002). 文字知覚のための２段階神経回路網モデルとその破壊実験による離断仮説と視覚性障害仮説の検討. 『神経心理学』, *18*, 92–100.

Asakawa, S. (2014). Semantics with or without categorization. In A. Costa & E. Villalba (Eds.), *Horizons in neuroscience research* (Vol. 16, pp. 139–178). New York, NY: NOVA science publishers.

浅川 伸一. (2015a). 『ディープラーニング、ビッグデータ、機械学習あるいはその心理学』. 東京: 新曜社.

浅川 伸一. (2015b). ニューラルネットワーク. 榊原 洋一・米田 英嗣（編）『発達科学ハンドブック』 (Vol. 8, pp. 94–104). 東京: 新曜社.

浅川 伸一. (2016a). 『Python で体験する深層学習』. 東京: コロナ社.

浅川 伸一. (2016b). 深層学習をめぐる最近の熱狂. 『基礎心理学研究』, *35*, 149–162.

浅川 伸一. (2012). 脳損傷患者の症例から見た読字過程. 川崎 惠里子（編）『認知心理学の新展開—言語と記憶』. 京都: ナカニシヤ出版.

Blei, D. M., Kucukelbir, A., & McAuliffe, J. D. (2017). Variational inference: A review for statisticians. *arXiv*.

Bodén, M. (2002). A guide to recurrent neural networks and backpropagation. *the Dallas project*.

Bottou, L., & Bousquet, O. (2007). The tradeoffs of large scale learning. In *Advances in Neural Information Processing Systems* (Vol. 20). Cambridge, MA, USA: MIT Press.

Brooks, R. A. (1990). Elephants don't play chess. *Robotics and Autonomous Systems*, *6*, 3-15.

Brooks, R. A. (1991). Intelligence without representation. *Artificial Intelligence*, *47*, 139-159.

Dayan, P., & Abbott, L. F. (2001). *Theoretical neuroscience*. Cambridge, MA: MIT press.

Edelman, S. (1997). Computational theories of object recognition. *Trends in Cognitive Sciences*, *1*, 296–304.

Fisher, R. A. (1936). The use of multiple measures in taxonomic problems. *Annual Eugenics*, *7*, 179–188.

Friston, K. (2007). The free-energy principle: a rough guide to the brain? *Trends in Cognitive Sciences*, *13*, 293–301.

Friston, K., Kilner, J., & Harrison, L. (2006). A free energy principle for the brain. *Journal of Physiology*, *100*, 70–87.

Friston, K. J., Stephan, K. E., Montague, R., & Dolan, R. J. (2014). Computational psychiatry: the brain as a phantastic organ. *The Lancet Psychiatry*, *1*, 148-158.

Fukushima, K. (1980). Neocognitron: A self-organizing neural network model for a mechanism of pattern recognition unaffected by shift in position. *Biological Cybernetics*, *36*, 193–202.

Fukushima, K., & Miyake, S. (1982). Neocognitron: A self-organizing neural network model for a mechanism of visual pattern recognition: Competition and cooperation in neural nets. In S. ichi Amari & M. A. Arbib (Eds.), *Lecture notes in biomathematics* (Vol. 45, pp. 267–285). Berlin, Heidelberg, New York: Springer-Verlag.

Hartline, H. K., & Ratliff, F. (1954). Spatial summation of inhibitory influences in the eye of limulus. *Science*, *120*, 781.

Hartline, H. K., & Ratliff, F. (1957). Inhibitory interaction of receptor units in the eye of limulus. *Journal of General Physiology*, *40*, 357–376.

Hartline, H. K., & Ratliff, F. (1958). Spatial summation of inhibitory influences in the eye of limulus, and the mutual interaction of receptor units. *Journal of General Physiology*, *41*, 1049–1066.

Hubel, D., & Wiesel, T. N. (1959). Receptive fields of single neurones in the cat's striate cortex. *Journal of Physiology*, *148*, 574–591.

Hubel, D., & Wiesel, T. N. (1962). Receptive fields, binocular interaction and functional architecture in the cat's visual cortex. *Journal of Physiology*, *160*, 106–154.

Hubel, D., & Wiesel, T. N. (1968). Receptive fields and functional architecture of monkey striate cortex. *Journal of Physiology*, *195*, 215–243.

Hubel, D. H. (1963). The visual cortex of the brain. *Scientific American*, *209*, 55–63.

岩田 誠. (1996). 『脳とことば — 言語の神経機構』. 東京: 共立出版.

Kohonen, T. (1985). *Self-organizing maps*. Springer-Verlag.

Kriegeskorte, N., & Douglas, P. K. (2018). Cognitive computational neuroscience. *21*, 1148–1160.

Krizhevsky, A., Sutskever, I., & Hinton, G. E. (2012). ImageNet classification with deep convolutional neural networks. In F. Pereira, C. Burges, L. Bottou, & K. Weinberger (Eds.), *in Advances in Neural Information Processing Systems 25*. Montréal, Canada.

LeCun, Y., Bottou, L., Bengio, Y., & Haffner, P. (1998). Gradient-based learning applied to document recognition. *Proceedings of the IEEE*, *86*, 2278–2324.

Livingstone, M., & Hubel, D. (1988). Segregation of form, color, movement, and depth: Anatomy, physiology, and perception. *Science*, *240*, 740-749.

Maaten, L. van der, & Hinton, G. (2008). Visualizing data using t-sne. *Journal of Machine Learning Research*, *9*, 2579-2605.

Marr, D. (1982). *Vision*. San Francisco, USA: W. H. Freeman and Company.

McCulloch, W. S., & Pitts, W. (1943). A logical calculus of the ideas immanent in nervous activity. *Bulletin of Mathematical Biophysics*, *5*, 115–133. (in Neurocomputing, Anderson, J. A. and Rosefeld, E., Neurocomputing, MIT press, chapt 2, 1988)

Nesterov, Y. (1983). A method of solving a convex programming problem with convergence rate o(1/k2). *Soviet Mathematics Doklady*, *27*, 372–376.

O'Reilly, R. C., Munakata, Y., Frank, M. J., Hazy, T. E., & Contributors. (2012). *Computational cognitive neuroscience* (1 ed.). Wiki Book.

Randall C. O' Reilly. (2006). Biologically based computational models of high-level cognition. *Science*, 91-94.

Pearson, K. (1901). On lines and planes of closest fit to systems of points in space. *Philosophical Magazine*, *2*, 559–572.

Peter N. Belhumeur, J. a. P. H., & Kriegman, D. J. (1997). Eigenfaces vs. fisherfaces: Recognition using class specific linear projection. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *19*, 711–720.

Poggio, T., Torre, V., & Koch, C. (1985). Computational vision and regularization theory. *Nature*, *317*, 314–319.

Rosenblatt, F. (1958). The perceptron: a probabilistic model for information storage and organization in the brain. *Psychological Review*, *65*, 386–408. (In J.A. Anderson and E. Rosenfeld (Eds.) Neurocomputing (1988), MIT Press)

Rumelhart, D. E., Hinton, G. E., & williams, R. J. (1986). Learning representations by back-propagating errors. *Nature*, *323*, 533–536.

Russell, S. J., & Norvig, P. (2003). *Artificial intelligence a modern approach*. Upper Saddle River, New Jersey 07458, USA: Pearson Education, Inc.

Selfridge, O. G. (1958). Pandemonium: a paradigm for learning. In *Mechanisation of Thought Processes: Proceedings of a Symposium*

Held at the National Physical Laboratory (Vol. 1, pp. 513–526). London, HMSO.
Shallice, T. (1988). *From neuropsychology to mental structure*. New York, NY: Cambridge University Press.
Sutton, R. S., & Barto, A. G. (1998). *Reinforcement learning*. Cambridge, MA: MIT Press.
田崎 晴明. (2018). 『数学 -物理を学び楽しむために-』.
Warrington, E. K. (1975). The selective impairment of semantic memory. *Quarterly Journal of Experimental Psychology*, *27*, 635–657.
Warrington, E. K., & McCarthy, R. A. (1987). Categories of knowledge further fractionations and an attempted integration. *Brain*, *110*, 1273–1296.
Watanabe, S. (1969). *Knowing and guessing*. New York: NY, USA: John Wiley and Sons.
山鳥 重. (1985). 『神経心理学入門』. 東京: 医学書院.
林 周二. (1967). 『数学再入門 I』. 東京: 中央公論社.
林 周二. (1968). 『数学再入門 II』. 東京: 中央公論社.