A Neural Network Metaphor for Organizations

Reimer Kühn

Department of Mathematics, King’s College London

Heidelberg, Jan 2010
Overview

- Neurons; what individual neurons do
- Neurons and neural networks: dynamics and attractors
- Information processing and learning
- Learning in Neural Nets – Version Space
- Learning Systems in General
- Main Implications & Limitations
- Summary
Neurons

diversity of neurons

schematic neuron

formal neuron

Human brain: $\sim 10^{11}$ neurons, each connected to $\sim 10^4$ (input & output), $\lesssim 5 – 6$ handshakes between any pair
What Individual Neurons Do: Linear Separation

- Neurons: two state threshold elements, firing or non-firing
  \[ S_i \in \{1, 0\} \]

- Input output relation
  \[ h_i = \sum_j J_{ij} S_j \equiv J_i \cdot S \]
  \[ S_i = \Theta(h_i - \vartheta_i) \]

- Synaptic couplings:
  \[ J_{ij} \begin{cases} > 0 & \text{; excitatory} \\ < 0 & \text{; inhibitory} \end{cases} \]

Classification by linear separation
Neurons and Neural Networks

- Neurons interacting in networks: output of a neuron is used as input by others at later times (important when feedback loops exist).

- Dynamics (can be deterministic or probabilistic)

\[
S_i(t + \Delta t) = \Theta \left( \sum_j J_{ij} S_j(t) - \vartheta_i \right)
\]

\[
\text{Prob}\{S_i(t + \Delta t) = 1\} = \Phi \left( \sum_j J_{ij} S_j(t) - \vartheta_i \right)
\]

- Deterministic dynamics: two types of global states: transient or persistent; the latter can be stable attractors (fixed points, limit cycles). ⇒ associative memory, motion control, . . . .

- Probabilistic dynamics: fluctuations about attractors of deterministic dynamics (long-lived sets of states).
Recursive Architecture — Attractors

Recursive architecture

Attractors depend on \( \{J_{ij}, \theta_i\} \): created by ‘learning’ or ‘adaptation’
Information Processing & Learning

- neural firing states $\iff$ information
- neural dynamics $\iff$ information processing
- synaptic couplings $\iff$ information processing capabilities
- processing capabilities (synapses) evolve through learning/training
- Correspondence (high level)
  - neural firing states $\simeq$ brain states $\iff$ cognitive states
  - neural dynamics: generates representations of world and acts on them; interprets, generates actions/reactions
  - learning $\iff$ adapt representations to improve adequacy, performance, success rate, survival probability, rewards, ...
Learning in Neural Nets — Version Space

- Adapt $\{J_{ij}\}$ to stabilize a desired set of (sequences of) neural activity patterns ($\simeq$ ‘procedures’) $\{S^\mu\}$, $\mu = 1, 2, \ldots, p$.

$$S^\mu_i(t + \Delta t) = \Theta(\sum_j J_{ij}S^\mu_j(t) - \vartheta_i) \quad \forall i, \mu \quad (\ast)$$

- Could be fixed points, limit cycles, input-output pairs (classifications of sensory data)

- Every new pattern puts new constraints on $\{J_{ij}\}$. Version space:

$$V_p = \{J_{ij}; \{J_{ij}\} \text{ satisfy (\ast)}, \mu = 1, 2, \ldots, p\}$$

- Version space shrinks

$$V_0 \supseteq V_1 \supseteq V_2 \supseteq V_3 \supseteq \ldots$$

- For given architecture/structure, there may be $p_{\text{max}}$ s.t. $V_p = \emptyset$ for $p > p_{\text{max}}$, independently of learning algorithm/strategy.
Learning in Neural Nets — Version Space

- Value of $p_{\text{max}}$ depends on architecture and (statistical) properties of $\{S^{\mu}\}$. Typically of the order of the number $N$ of input channels (adaptable parameters) per neuron.
  — For random $\{S^{\mu}\}$, $p_{\text{max}} = 2N$.

- Architectures/structures have restrictions on the problems they can represent/solve. E.g.
  $$S_0 = \text{XOR}(S_1, S_2) \neq \Theta(J_{01}S_1 + J_{02}S_2 - \vartheta_0)$$
  for any $J_{01}, J_{02}, \vartheta_0$.

- As $p \rightarrow p_{\text{max}}$ for given problem class, the version space can become fragmented (disconnected). Finding solutions that accommodate new patterns can become difficult (even impossible without violating some constraints ‘on the way’ to the new solution).
Learning Systems in General

- Any complex system able to generate and adapt representations of the world can be said to learn.

- Need not be an individual agent, but could be a team, an organization, or a society of complex agents.

- Information exchange via language, actions, gestures, attitudes. . .

- \(\Leftrightarrow\) collective information processing, if exchanged information influences/constrains (re-)action on receiving side.

- The collection of agents learns, if interactions (evaluations of exchanged information) are adapted, e.g. in order to improve performance, utility functions, . . .
Main Implications

• Collective information processing in organizations will be guided by (dynamic) attractors
  – subsets of system states or sequences of system state generated and reinforced by dynamics at the system level
  – $\simeq$ spontaneous symmetry breaking & pattern formation
  – multiplicity/diversity of attractors in systems with given interactions

• Limits on representability or solvability of tasks, and problems due to fragmented version spaces as systems approach their capacity limits will exist for collections of complex agents as well.
Limitations

- In what follows, explore whether the hypothesis about existence of fundamental limitations of representability etc in organizations survives scrutiny.

Do this via series of approximations.

- Start with society of two-state (yes/no) agents:

\[ h_i^\mu(t) = \sum_j J_{ij} S_j^\mu(t) , \quad S_i^\mu(t + \Delta t) = \Theta(h_i^\mu(t) - \vartheta_i) \quad \forall i, \mu \]

Standard theory for McCulloch-Pitts neurons applies. Typically \( p_{\text{max}} = \mathcal{O}(N) \). Precise values are known and depend on statistics (\( p_{\text{max}} = 2N \) for random patterns).
First critique: states of agents are not binary.

Response: assume $S_i \in \mathbb{R}$ and graded response dynamics

$$h_i(t) = \sum_j J_{ij} S_j(t) \ , \quad S_i(t + \Delta t) = g_i(h_i(t))$$

for some (monotone) response functions $g_i$.

Embedding patterns (with tolerance $\varepsilon$) then requires:

$$S^\mu_j(t) \in I^\varepsilon,\mu_{j,\text{out}}(t) \implies h^\mu_i(t) = \sum_j J_{ij} S^\mu_j(t) \in I^\delta,\mu_{i,\text{in}}$$

so that $S^\mu_i(t + \Delta t) \in I^\varepsilon,\mu_{i,\text{out}}(t + \Delta t)$

Modification of standard theory for graded response neurons applies. Typically $p_{\text{max}} = \mathcal{O}(N)$. Precise values are known and depend on statistics, input/output tolerance $\varepsilon$ and the shape of the transfer-function $g$. 
Second critique: state variables communicated between agents are not simple scalars (language, gestures, actions, ...).

Response: enlarge dimension. $h_i = (h_i^a) \in \mathbb{R}^K$.

$$h_i^a(t) = \sum_{jb} J_{ij}^{ab} S_j^b(t), \quad S_i^a(t + \Delta t) = g_i^a(h_i(t))$$

Embedding patterns (with tolerance $\varepsilon$) then requires multi-dimensional generalization of previous argument.

Want (dropping temporal aspects in I/O tolerance values):

$$S_{ij}^{b\mu} \in I_{ij,\text{out}} \implies h_i^{a\mu} = \sum_{jb} J_{ij}^{ab} S_j^{b\mu} \in I_{i,\text{in}}^{b\mu}$$

so that $S_i^{a\mu} \in I_{i,\text{out}}^{a\mu}$

Same conclusions, although computations have not been done. Expect $p_{\text{max}} = \mathcal{O}(KN)$
Third critique: Argument ignores that agents are complex and have internal adaptable structure to ‘compute’ outputs.

Response: could take internal adaptable structure into account

\[ g_i^a(h_i(t)) = g_{w_i}^a(h_i(t)) \]

in which \( w_i \) stands for the collection of adaptable parameter within a function-class representable by agent \( i \).

Conclusion about existence of fundamental limitations are not altered, if internal parameters are taken into account. Get further enlargement of expected \( p_{\text{max}} \).

Only partial results known. If \( g_i^a \) represents the computation of a feed-forward neural network with \( M \) hidden nodes, get an extra factor \( O(\ln M) \) for random binary patterns.
Fourth critique: Individuals are not machines (they act autonomously & unpredictably ...)

Response: A difficult one. A tentative answer could be that as far as the influence of this feature on the question of existence of fundamental limitations is concerned, one could model the unpredictability and autonomy by replacing the $g_i^a$ by random functions which produce a range of outputs with a statistics constrained by the inputs.

The requirements of stability of collective information processing patterns would then have to be replaced by conditions of the form that with sufficiently high likelihood, an agent would have to produce $S_i^{a\mu} \in I_{i,\text{out}}^{a\mu}$ in context $\mu$.

If this is not guaranteed, there would be too many malfunctioning agents jeopardizing the reliable execution of the pattern (the ‘collective procedure’) $\mu$. This would be perceived as costly by the organization, creating pressures to improve.
What should I make of all this?

- Note that distributed information processing and learning is ubiquitous; it doesn't begin and it doesn't stop at the individual level.

- Note hierarchy of levels: neurons → cortical columns → brain organization (visual, somatosensory, auditory, olfactory .. cortices, cerebellum ...) → brain → society of brains.

- Suspect corresponding hierarchy of levels of meaning (for most of which, both at the microscopic end and at the large collective end, we have as yet no established concepts).

- Information exchange and processing between levels, both ways: 'up and down'

- ➞ Ask the question which is the theme of this workshop: (Where are the) Human boundaries?
Specifically: the notion of attractors in dynamical systems with feedback could rationalize long-lived patterns in societies and organizations (dominant paradigms, fashions & trends in art/science/economy)
• Specifically: the discussion of existence and implication of fundamental limits
  – could rationalize some of the major transformations in history as finding ‘extra dimensions’ to accommodate solutions to problems previously unsolvable, given the structure (eg. hunter/gatherer to agricultural transition, invention of writing to create reliable long term memory, recently the world wide web to allow fast exchange of annotated data, texts, and theories, software engineering concepts to handle large software projects)
  – It might be useful if policy makers were aware of limitations and shrinking of version spaces with added complexity driven by added problem solving requirements: Creation of institutions/rules/laws often double-edged: new institutions may liberate some elements of a society by relieving them of tasks, but they also put constraints in the form of additional coordination requirements. See to it that effects of extra constraints do not outweigh liberating effects.
Summary

- Neural information processing as paradigm for decentralized & collective information processing
- Structure and limits of representability (version spaces)
- Society of brains as information processing systems
- Implications of limits of representability and functionality for such systems
- Interlaced hierarchy of levels of information processing & meanings
- Human Boundaries