Do neural models scale up to a human brain?

Roman V. Belavkin (r.belavkin@mdx.ac.uk)

School of Computing Science,
Middlesex University, London NW4 4BT, UK

August 27, 2007

MOTIVATION

- Understanding the main function of an object can give us understanding about its organisation

  Example Engine of a car — one function, different constraints, many implementations.

- What is the main function of the brain?
- Cognitive architectures (ACT–R, SOAR) operate at a high (macro) level. Neural models operate at a low (micro) level.
- Can these models explain or predict macroscopic data about the brain? (e.g. why 10^{11} neurons in the human brain?)
- Are our neural models **sufficient**? (many are **necessary**)

ORGANISATION of HUMAN NERVOUS SYSTEM

Central (CNS)	**Peripheral** (PNS)
Brain $(10^{11}$ neurons) • Forebrain $(2 \cdot 10^{10}$ neocortex) • Midbrain • Hindbrain Spinal cord (10^9)	Somatic voluntary control Autonomic (ANS) • Sympathetic (fight or flight) • Parasympathetic (rest and digest) • Enteric (10^9)

$$\text{PNS} \longrightarrow (\text{inputs})^m\, (\text{CNS})^S\, (\text{outputs})^n \longrightarrow \text{PNS}$$

PNS connects CNS to the outside world through 12 pairs of *cranial* and 31 pairs of *spinal* nerves.

CRANIAL NERVES (12 pairs)

Nerve:	Afferent (IN)	Efferent (OUT)	Fibres
olfactory	smell		$1.2 \cdot 10^7$
optic	vision		$1.2 \cdot 10^7$
vestibulocochlear	hearing, balance		$3.1 \cdot 10^4$
oculomotor		eye, pupil size	$3 \cdot 10^4$
trochlear		eye	$3 \cdot 10^3$
abducens		eye	$3.7 \cdot 10^3$
hypoglossal		tongue	$7 \cdot 10^3$
spinal-accessory		throat, neck	?
trigeminal	face	chewing	$8.1 \cdot 10^3$
facial	2/3 taste	face	10^4
glossopharyngeal	1/3 taste, blood pressure	throat, soliva secreation	?
vagus	pain	heart, lungs, abdominal, throat	?

(Bear, Connors, & Paradiso, 2007; Poritsky, 1992)

$$m_c \approx 4.81 \cdot 10^7\,, \qquad n_c \approx 1.45 \cdot 10^5$$

SPINAL NERVES (31 pairs)

Nerves:	Number
cervical	8
thoracic	12
lumbar	5
sacral	5
coccyx	1

- Spinal nerves are both sensory and motor
- There are 10^9 neurons in spinal cord

$$m_s = n_s \approx 2 \cdot 31 \cdot 4.5 \cdot 10^3 = 2.8 \cdot 10^5$$

$1,1 \cdot 10^6$ fibres in pyramidal decussation (motor fibres which pass from the brain to medulla)

MY ESTIMATES

$$\begin{aligned} m = m_c + m_s &\approx 4.84 \cdot 10^7 \qquad (3 \cdot 10^8) \\ n = n_c + n_s &\approx 4.26 \cdot 10^5 \qquad (9,8 \cdot 10^5) \end{aligned}$$

$2,5 \cdot 10^8$ fibres in corpus callosum (connects the left and right cerebral hemispheres).

Important:

- $m \gg n$
- $S \gg m,\ n$, where $S \approx 10^{11}$ (n. of neurons in the brain)
- $k \ll m$, where $k \in [10^3, 10^4]$ (n. of synapses)

HYPOTHESES

What could be the main function of neurons and the CNS?

- Optimal estimation and control
- Optimal abstract model
- Optimal information coding

OPTIMAL ESTIMATION and CONTROL

Let $x \in X$ be unobserved state of the world with preferences induced by $c : X \to \mathbf{R}$ (cost function).

Let $y \in Y^m$ be observed, $u \in U^n$ the estimate or control. The optimal

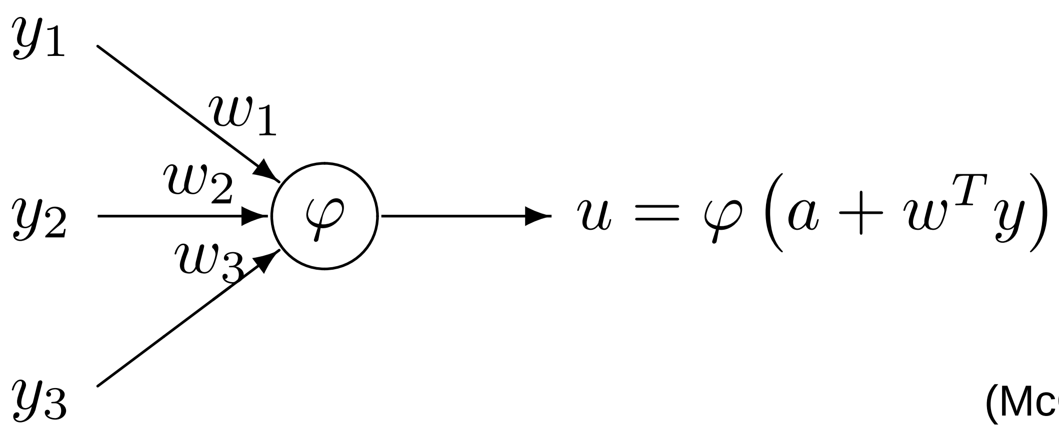
$$\begin{aligned} u^*(y) &= \arg\min_{u(y)} E\{c(x, u(y)) \mid y\} \\ &= \arg\min_{u(y)} \int c(x, u(y)) P(\mathrm{d}x \mid y) \end{aligned}$$

For quadratic cost (e.g. $c = |x - u(y)|^2$) the optimal is

$$u^*(y) = E\{x \mid y\} \approx E\{x\} + B^T(y - E\{y\})$$

For Gaussian x, linear is optimal (Stratonovich, 1959; Kalman & Bucy, 1961)

NEURON as a LINEAR MODEL



(McCulloch & Pitts, 1943)

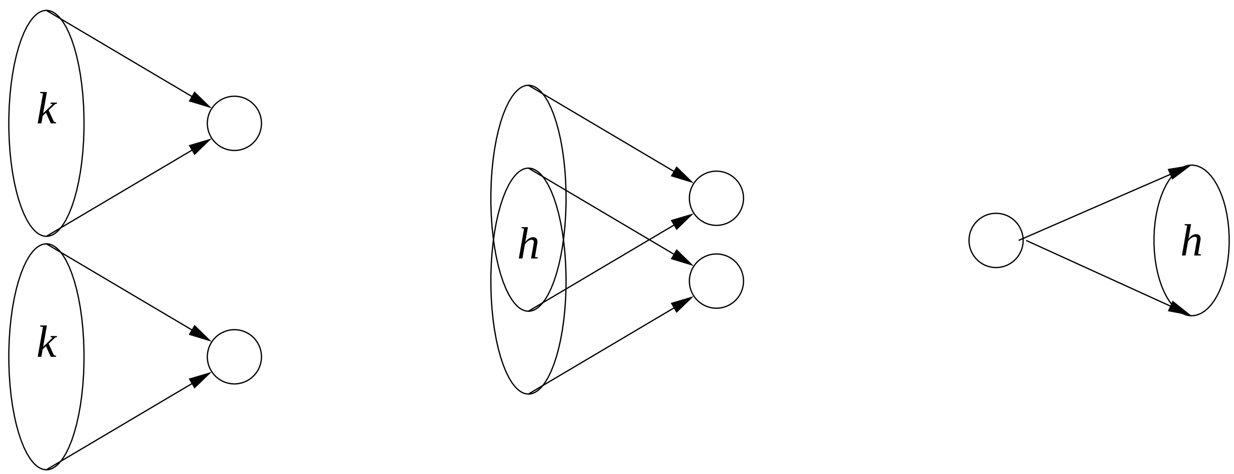
- If we let $a = E\{x\} - B^T E\{y\}$, and $W = B$, then NN implements optimal linear transformation (estimation or control).
- Hebbian learning $w_i \approx \beta_i(y, u)$ (Hebb, 1955; Sejnowski, 1977)
- Principal or independent components analysis using NN (Oja, 1982; Hyvärinen & Oja, 1998), self–organising maps (Kohonen, 1982)
- It is possible to do linear $u : Y^m \to U^n$ with a single layer ($S = 0$) of n neurons with $k = m$ (but $k \ll m$)

CONSTRAINTS on CONNECTIVITY

Assume a layered structure with r_i nodes in 'layer' i

k — maximum number of inputs (synapses) from $i+1$

h — maximum number of output connections (axon branches) to i



$\max r_{i+1} = k r_i$ $\qquad\qquad$ $\max r_i = h r_{i+1}$

We are talking about the same vertices connecting i and $i+1$

$$r_{i+1} = r_i \frac{k}{h}$$

PARTIALLY CONNECTED FORWARD NETWORKS

Using boundary conditions $r_0 = n$, $r_i = n\left(\frac{k}{h}\right)^i$, $r_{l+1} = m$. Thus

$$m\left(\frac{h}{k}\right)^{l+1} = n$$

The number of layers (the *order of connectivity*)

$$l = \frac{\ln m - \ln n}{\ln k - \ln h} - 1 \tag{1}$$

Total number of hidden nodes

$$S = \sum_{i=1}^{l} r_i = n\sum_{i=1}^{l}\left(\frac{k}{h}\right)^i = m\sum_{i=1}^{l}\left(\frac{h}{k}\right)^i \tag{2}$$

ESTIMATING l and S

- Set $m = 4.84 \cdot 10^7$, $n = 4.26 \cdot 10^5$
- For $\frac{h}{k} = .9995$, using (1) and (2) we get

$$l = 9461\,, \qquad S = 0.96 \cdot 10^{11}$$

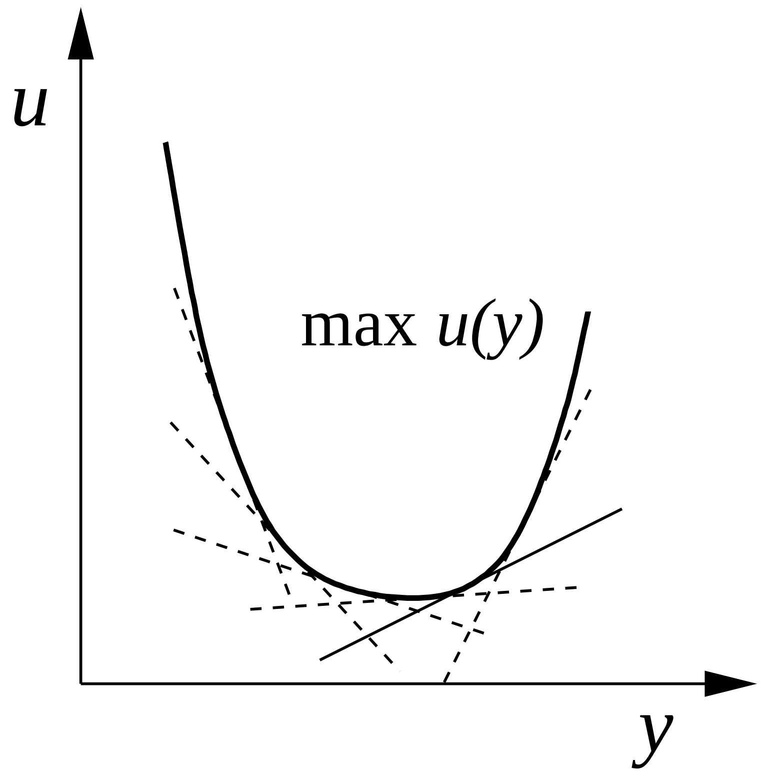
- Note that $h, k \in \mathbf{N}$, and if k is minimised, $\frac{h}{k}$ maximised, then $h = k - 1$.
- For $\frac{h}{k} = .9995$, we have

$$k = 2 \cdot 10^3$$

- Recall that $k \in [10^3, 10^4]$ (n. of synapses of an average neuron)

OPTIMAL NONLINEAR FILTERING

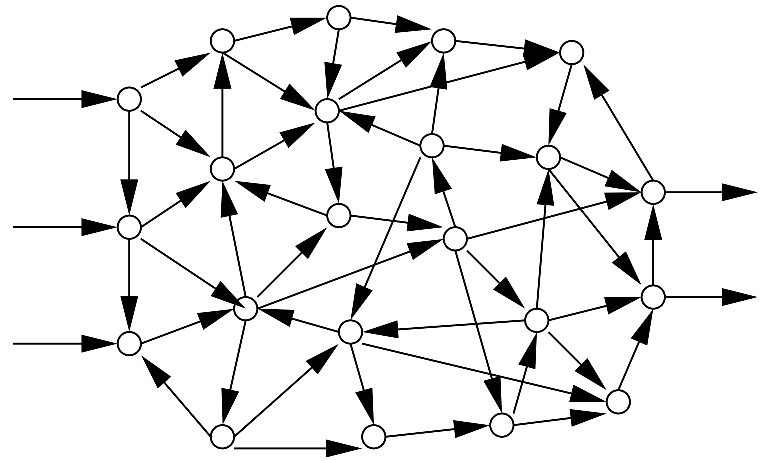
- Several linear units with an extra layer can approximate nonlinear functions.
- Optimal linear algorithms require only the first two moments, but are optimal only for Gaussian P. Similar algorithms are optimal in the sense of $\max P$ and suitable for non–linear problems (Stratonovich, 1959).
- For small k, P on $Y^k \subset Y^m$ the Gaussian approximation can be sufficient.



- Small $k \longrightarrow$ faster convergence.
- The sum of Dirac δ–measures (i.e. Gaussians with $\sigma^2 = 0$) can be used to approximate any P.

RECURRENT NETWORKS with CONSTRAINTS

NN as directed graph $G = (V, E)$



$$\sum_v \deg^+(v) = \sum_v \deg^-(v) = |E|$$

Constraints:

$$\deg^+(v) \leq k\,, \quad \deg^-(v) \leq h$$

Maximising $|E|$ for fixed $|V|$ (or minimising $|V|$ for fixed $|E|$)

$$\max_G |E| = (S + n)k = (S + m)h$$

$$S = \frac{mh - nk}{k - h} \tag{3}$$

For $m = 4.84 \cdot 10^7$, $n = 4.26 \cdot 10^5$, $k = 2 \cdot 10^3$ gives

$$S \approx 0.96 \cdot 10^{11}$$

OPTIMAL ABSTRACT MODEL

- Directed graph, $G = (V, E)$, can represent an abstract model. Each link between two nodes is a binary relation, and a path of l nodes between input and output nodes can be seen as l–operator between y and u.
- Fully connected directed graph represents Cartesian product $Y \times \cdots \times U$ — all possible relations (not interesting).
- The mind can be seen as a subset $G \subset Y \times \cdots \times U$ representing the most important operators.

OPTIMAL INFORMATION CODING

- Consider CNS as a function of random variable, $u : Y^m \to U^n$.
- If $u(y)$ is not an isomorphism, then information contained in y is generally destroyed ($|Y|^m \geq |U|^n$).
- For entropically stable y, we only need to encode

$$e^{H_y} \leq |Y|^m \,, \qquad \text{(where } H_y = -E\{\ln P(y)\}\text{)}$$

- The optimal code approaches uniform $P(u)$ such that

$$\max_{P(u)} H_u = |U|^n = e^{H_y} \leq |Y|^m$$

- If $|U| = |Y|$ (e.g. 2), then $n \leq m$, and still encodes all information.

NN for OPTIMAL CODING

- Many ANN algorithms maximise the entropy of the output. For example, ICA can be implemented using

$$u^*(y) = \arg\min_{u(y)} \left(\sum_{i=1}^{n} H_{u_i} - H_u \right)$$

  The above minimum corresponds to maximum H_u (optimal coding).

- Linear ICA can be implemented using single layer network, which does not correspond to $S \approx 10^{11}$ and $k \approx 10^3$.

- The constraints on connectivity lead to ‘multilayered’ network, and therefore the brain may implement nonlinear $u(y)$.

OPTIMAL CODING

- A network of S units has the capacity to communicate $|U|^S$ realisations.
- However, h units receive the same information, and the real capacity is $|U|^{S/h}$.
- Preserving information between input and output (perfect communication) means

$$|Y|^m = |U|^{S/h}, \qquad m = \frac{S}{h} \qquad \text{(e.g. } |Y| = |U| = 2\text{)}$$

- Using our estimates of m and h, we obtain

$$S \approx .97 \cdot 10^{11}$$

CONCLUSIONS and DISCUSSION

- It is possible that the brain implements the optimal (nonlinear) control and optimal coding. Their combination is a familiar variational problem

$$F = \min_{P(\mathrm{d}u|x)} (E\{c(x, u(y))\} + \lambda E\{\ln P(\mathrm{d}u \mid x)\}) = R - TC$$

(Free energy)

- Are neural models **sufficient**? We need to consider:
  - Partially connected, multilayer (nonlinear) networks
  - Achieves maximum connectivity (or minimum number of nodes)
  - Local and bounded connectivity leads to cell–assemblies (Hebb, 1955) (may lead to topology preserving mapping like in SOM).
- A particular organisation of the brain is likely the result of optimisation due to additional constraints: Sensory (m), motor (n), h/k–ratio.

References

Bear, M. F., Connors, B. W., & Paradiso, M. (2007). *Neuroscience: Exploring the brain* (Third ed.). PA: Lippincott Williams & Wilkins.

Hebb, D. O. (1955). Drives and the C.N.S. (conceptual nervous system). *Psychological Review*, *62*, 243–254.

Hyvärinen, A., & Oja, E. (1998). Independent component analysis by general nonlinear Hebbian–like learning rules. *Signal Processing*, *64*, 301–313.

Kalman, R. E., & Bucy, R. S. (1961, December). New results in linear filtering and prediction theory. *Transactions ASME, Series D–J. Basic Engineering.*, *83*, 94–107.

Kohonen, T. (1982). Self–organized formation of topologically correct feature maps. *Biological Cybernetics*, *43*, 59–69.

McCulloch, W., & Pitts, W. (1943). A logical calculus of ideas immanent in nervous activity. *Bulletin of Mathematical Biophysics*, *5*, 115-133.

Oja, E. (1982). A simplified neuron model as a principal component analyzer. *Journal of Mathematical Biology*, *15*, 267–273.

Poritsky, R. (1992). *Neuroanatomy: A functional atlas of parts and pathways*. Hanley & Belfus Inc.

Sejnowski, T. J. (1977). Storing covariance with nonlinearly interacting neurons. *Journal of Mathematical Biology*, *4*, 303–321.

Stratonovich, R. L. (1959). Optimum nonlinear systems which bring about a separation of a signal with constant parameters from noise. *Radiofizika*, *2*(6), 892–901.