INTERNATIONAL ATOMIC ENERGY AGENCY
UNITED NATIONS EDUCATIONAL, SCIENTIFIC AND CULTURAL ORGANIZATION

INTERNATIONAL CENTRE FOR THEORETICAL PHYSICS
34100 TRIESTE (ITALY) - P.O.B. 586 - MIRAMARE - STRADA COSTIERA 11 - TELEPHONE: 2240-1
CABLE: CENTRATOM - TELEX 460392 - 1

SMR/302-35

COLLEGE ON NEUROPHYSICS:
"DEVELOPMENT AND ORGANIZATION OF THE BRAIN"
7 November - 2 December 1988

"A Review of Neural Nets"

John G. TAYLOR
King's College
Department of Mathematics
London, U.K.

Please note: These are preliminary notes intended for internal distribution only.

A REVIEW OF NEURAL NETS.

PURPOSE:

(a) 'DUPLICATE' BRAINS ⟶ Si BRAINS

(b) '>IMITATE' BRAINS ⟶ SLAVE BRAINS

TAKING OVER FROM EXPERT SYSTEMS (?) ⟶

N.N's COMPOSED OF:

(a) 'neurons'
(b) connection 'weights'
(c) learning rules

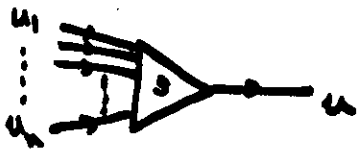
(a)–(b): static nets
(a)–(c): learning nets

HISTORY: 1943 McCulloch & Pitts BDN-logic ⟶ serial computers
1957 Rosenblatt ⟶ perceptron
1969 Minsky & Papert ⟶ dead perceptron
1961 Caianiello ⟶ BDN nets ⟶ 'thoughts' + learning
1982 Hopfield ⟶ energy E for BDN nets ⟶ CAMs
>1980 McClelland Rumelhart Winston, Grossberg ⟶ PDP connectionism ⟵ brain res.

I STATIC NETS.

1.1. Deterministic

1. BDN: $u_i = 0$: off
$= 1$: on



$u = 0$ if $\sum_i a_i u_i < s$
$= 1$ $\geq s$

weights: +ve excitat.
−ve inhibit.

$u = \theta\left(\sum a_i u_i - s\right)$

Logic functions:

X ∨ Y

X ∧ Y

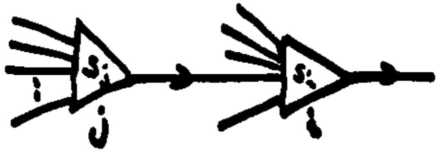
etc.

A Net of BDN's can perform any logical function

(McC+P
1943

2. BDN Nets

(Caianiello '61



Common time delay τ

$$u_i(t+\tau) = \theta\Big(\sum_{j=1}^{N} \underbrace{a_{ij}}_{\text{weights}} u_j(t) - s_i\Big)$$

φ inputs φ outputs

$\boxed{\text{Net activity to cycles}}$ (reverberations)

state space $2^N = \Big(\overbrace{\begin{smallmatrix}0\\1\end{smallmatrix}\ \begin{smallmatrix}0\\1\end{smallmatrix}\ \cdots\ \begin{smallmatrix}0\\1\end{smallmatrix}}^{N}\Big)$
points

(must always return)

Caianiello's program : cycles ~ 'thoughts'
+ learning (slow) for Δa_{ij}

But
- many cycles $(\# > (10^{10^9})^{10^{10^9}}$ if $N \sim 10^{10})$
- which connectivities?
- 2^N cycles for fixed data : too many

(J. Clark
Phys Rep
Feb '88

3. Hopfield Nets

(Pr. Nat. Acad. Sci
79, 2554, '82
(Science 233, 625, '86
(with Tank)

$a_{ij} = a_{ji}$, $a_{ii} = 0$

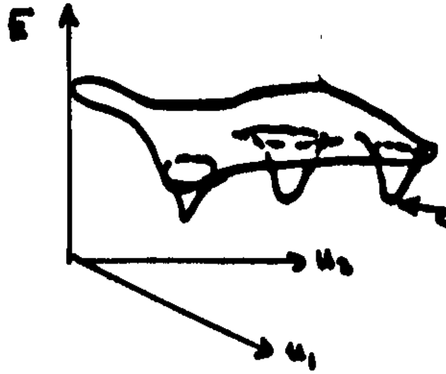
$\exists$ energy function

$$E = -\sum_{i<j} a_{ij} u_i u_j + \sum_{j=1}^{N} u_j s_j$$

$$E(u_i = +1) - E(u_i = -1) = s_i - \sum_j a_{ij} u_j$$

$\leq 0 \quad \Leftrightarrow \begin{cases} \text{switch } u_i \text{ on (or leave on) if } \sum a_{ij}u_j > s_i \\ \quad \text{"} \quad u_i \text{ off (" " off) " } \sum a_{ij}u_j < s_i \end{cases}$

$\sim \Big\{$ no increase in E if update neurons singly as BDNs.



descent to minima of E.

~ Content Addressable Memory.

Speed up by noise :

(false min)

- basins of attraction

(intertwined?)

T.S.P. →

What info. stored? $$a_{ij} = \sum_{s}(2u_i^{(s)}-1)(2u_j^{(s)}-1)$$

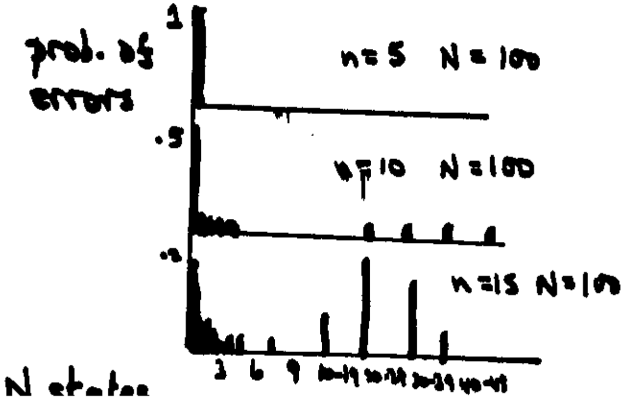
→ (±1 ~ spin glasses)

stored activities

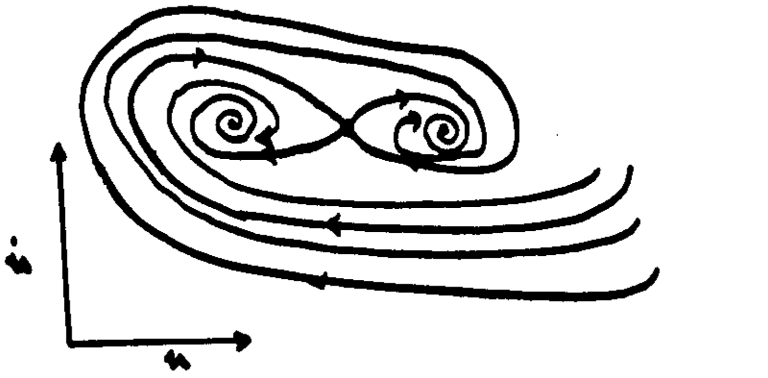
+ find $\sum_j a_{ij}u_j^{(s)} \approx u_i^{(s)}$

(on average) reproducing

Too large N gives errors:



⇒ $\lesssim .14N$ states



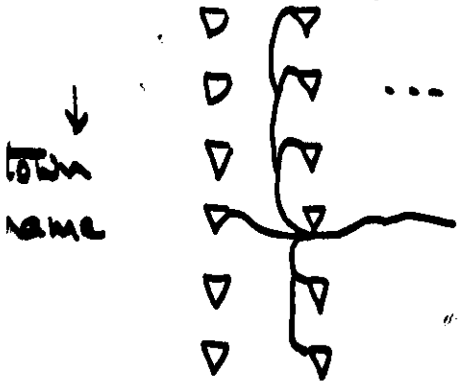
intertwined basins of attraction

or chaos

chaotic attractor

velling Salesman problem:

→ order of visiting



inhibit same order (column)

" town (row)

$$E \sim \sum_{x,y} d_{xy} u_{xi} u_{y\,i+1} + \sum_{x,i,j} u_{xi} u_{xj} + \sum_{x,y,i} u_{xi} u_{yi}$$

30 towns

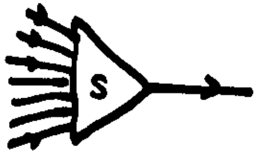
10^{30} possible routes

1 of 10^7 solutions in one iteration

1.2. Pulse Processing

Pulses with freq. f

prob of firing $p = f \cdot \Delta t$



firing prob $f\Delta t = \sum_{f_{r_1}+\dots+f_{r_l}>s} (f_{r_1}\Delta t_{r_1}) \cdot \dots \cdot (f_{r_l}\Delta t_{r_l})$

Becomes more non-linear as s increases:

$s=0$: $f\Delta t = \sum f_r \Delta t_r + \sum f_r f_s \Delta t_r \Delta t_s + \dots$

$= \text{linear} + \dots$

$s = \sum f_r$: $f\Delta t = \prod f_r \Delta t_r$

- pulse proc. with decay
- refractory period

13 Linear Nets.

(Hartline et al – inv ret.
J.G.T. – vert. ret.
(Cooper et al

$$\underbrace{y_i}_{\text{outputs}} = \sum_{j=1}^{N} A_{ij} \underbrace{x_j}_{\text{inputs}}$$

For $A_{ij} = u_i u_j$

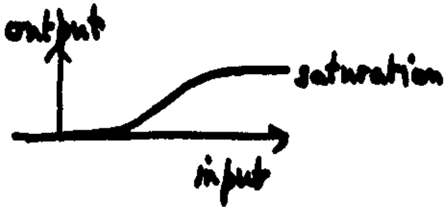
$$Y_i = u_i \sum_j (u_j x_j) = u_i (\underline{u} \cdot \underline{x})$$

projection of $\underline{u}$ on $\underline{x}$

Doesn't notice input orthogonal to $\underline{u}$

Generalise to $A_{ij} = \sum u_i^{(\alpha)} u_j^{(\alpha)}$ (Cooper et al

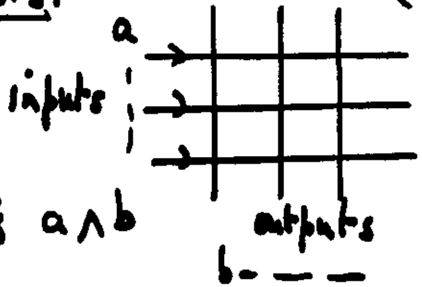
. more general neurons



II LEARNING RULES.

2.1. Associative Nets.

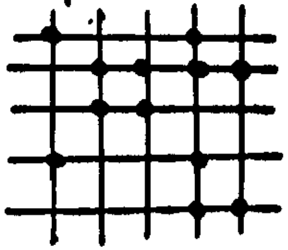
(Willshaw



. turn nodes on if a ∧ b

. then play in a pattern; set threshold on b at total activity on a → read out associated b pattern

e.g.



learnt to associate

01100 → 01100
01001 → 00011
10010 → 10010

Limitations if too many inputs → Cb models (Albus
Marr

2.2. Cost Function (Supervised L.)

e.g. quadratic cost fns

$$E = \frac{1}{2} \sum_{\underbrace{k}_{\text{presentations}}} (\underbrace{y^{(k)}}_{\text{actual}} - \underbrace{d^{(k)}}_{\text{desired}})^2$$

desired } outputs

$$y_i^{(k)} = \sum_j a_{ij} x_j^{(k)}$$

$$\frac{\partial E}{\partial a_{ij}} = \sum_k (y_i^k - d_i^k) x_j^k$$

he $\Delta a_{ij} = \varepsilon . \frac{\partial E}{\partial a_{ij}} = \varepsilon \sum_k (y_i^k - d_i^k) x_j^k$ (gradient descent

P $\Delta a_{ij}^{(n)} = \varepsilon (y_i^n - d_i^n) x_j^n$

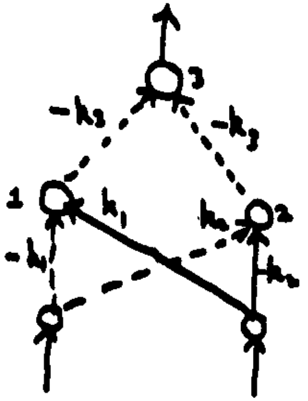
1. Hidden Units

Needed for complex tasks

eg. parity problem

$(00) \to 0$
$(01) \to 1$
$(10) \to 1$
$(11) \to 0$

$a_1 x_1 + a_2 x_2 - s$ 1



$a_1 + a_2 = s$
$0 = s$
$a_1 - s = 1$
$a_2 - s = 1$
} contradiction

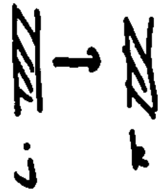
(Gorse

thresh: $1 = 2 = -\varepsilon$, $3 = -k_3 - \varepsilon$

input	to 1	to 2	fr 1	fr 2	to 3	fr 3
00	0	0	1	1	$-2k_3$	0
01	k_1	$-k_2$	1	0	$-k_3$	1
10	$-k_1$	k_2	0	1	$-k_3$	1
11	0	0	1	1	$-2k_3$	0

Need to find $\frac{\partial E}{\partial y_j}$ ← hidden

Do for feed forward net by back error prop.



$$\frac{\partial E}{\partial y_j} = \frac{\partial E}{\partial y_k} . \frac{\partial y_k}{\partial x_k} \frac{\partial x_k}{\partial y_j}$$

✓ ✓ ✓ from forward prop. of activity

Slow.

· Competitive learning (Kohonen

2.4 Unsupervised Learning.

$\Delta a_{ij} \propto u_i u_j$ (Hebb)

eg. application to development of feature detectors in VC (Linsker Pr. Nat. Acad. Sci '86

III APPLICATIONS

Mortgage risk assesment }
Electric motor testing } slaves

Feature detectors }
Small ganglia } brain research

Selverston
(Model Neural Nets
& Behaviour

- vision
- text recognition
- speech → text

IV FUTURE.

(Pappert

1. No clear math. structure ⇉ choice of neuron activity
   " " learning rules
2. Gap : low level processing (vision, ...)
   — cognitive level (attention, concept formation
3. Is CAM used by brain?

   Is any f.p. or cyclic activity used by brain?

   Better neuronal models (+noise)