Nonmonotonic Logic and Neural Networks

Reinhard Blutner with Paul David Doherthy, Berlin

1 Introduction

- ✤ Puzzle: Gap between symbolic and subsymbolic (neuron-like) modes of processing
- ✤ Aim: Overcoming the gap by viewing symbolism as a high-level description of the properties of neural networks
- ✤ Method: standard methods of model-theoretic and algebraic semantics. Neural (Re)interpretation of information states as activation states of a neuronal network.
- ✤ Main thesis: Certain activities of connectionist networks can be interpreted as *nonmonotonic inferences*. In particular, there is a strict correspondence between Hopfield networks and weight-annotated Poole systems. Extension of Balkenius & Gaerdenfors (1991).

Intended results

☞ Better understanding of connectionist networks:

Nonmonotonic logic and algebraic semantics as descriptive and analytic tools for analyzing their emerging properties

☞ New methods for performing nonmonotonic inferences:

Connectionist methods (randomised optimisation: simulated annealing) can be adopted for realizing symbolic inferences

☞ Certain logical systems are singled out by giving them a "deeper justification".

Overview

1 Introduction
2 A concise introduction to neural networks
3 Information states as neural activation patterns
4 Asymptotic spreading of activation and nonmonotonic inference
5 Weight-annotated Poole systems
6 The correspondence between symbolic inferences in weight-annotated Poole systems and inferences in connectionist networks (Hopfield nets)

2 A concise introduction to neural networks

General description

A neural network N can be defined as a quadruple <S,F,W,G>:

S Space of all possible states

W Set of possible configurations. $w \in W$ describes for each pair i,j of "neurons" the connection w_{ij} between i and j

F Set of activation functions. For a given configuration $w \in W$ a function $f_w \in F$ describes how the neuron activities spread through that network (fast dynamics)

G Set of learning functions (slow dynamics)

Hopfield networks

Let the interval [-1,+1] be the working range of each neuron

+1: maximal firing rate

0: resting

-1 : minimal firing rate

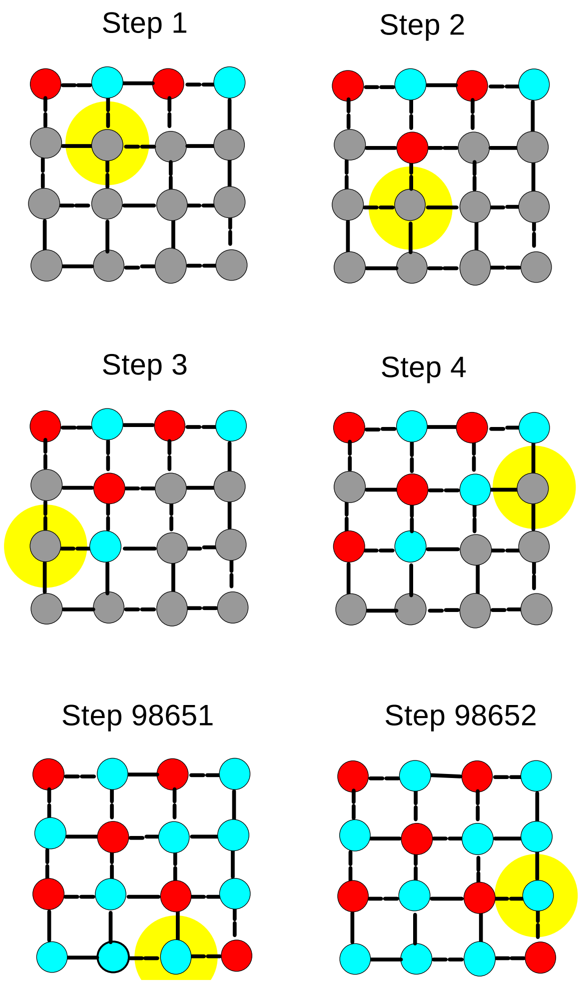
$S = [-1, 1]^n$

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

Aynchronous Updating:

$s_i(t+1) = \Theta(\Sigma_j w_{ij} \times s_j(t),$ if $i = \text{random}(1,n)$

$s_i(t+1) = s_i(t)$, otherwise



3 Information states in Hopfield networks

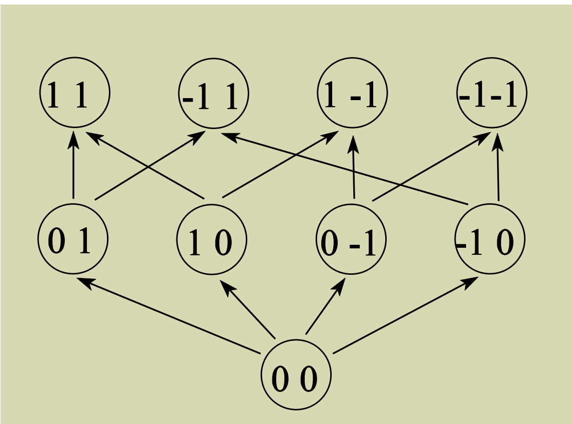
Activation states can be partially ordered in accordance with their informational content

+1: maximal firing rate } indicating maximal
- 1: minimal firing rate } specification
0: resting indicating underspecification

Poset of activation states:

$S = \{-1,0,+1\}^n$

$s \geq t$ iff $s_i \geq t_i \geq 0$ or $s_i \leq t_i \leq 0$, for all $1 \leq i \leq n$.



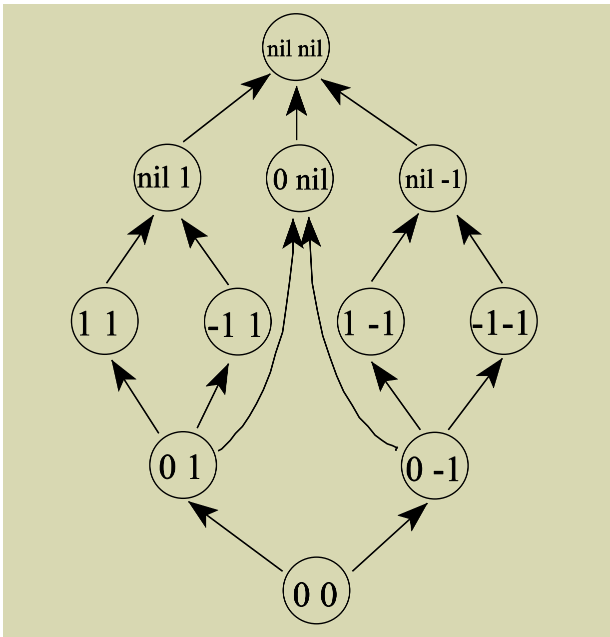
This poset doesn't form a lattice

Extended poset of activation states

$S = \{-1,0,+1,nil\}^n$

nil = "impossible activation"

$s \geq t$ iff s_i = nil or $s_i \geq t_i \geq 0$ or $s_i \leq t_i \leq 0$, for all $1 \leq i \leq n$.



DeMorgan lattice

CONJUNCTION ○: simultaneous realization of two states

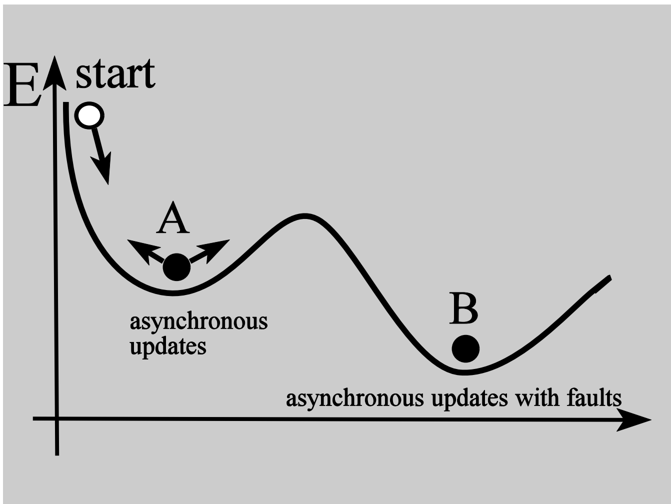
DISJUNCTION ⊕: some kind of generalization.

This fact enables us to interpret activation states as propositional objects (*information states*).

4 Asymptotic updates and nonmonotonic inference

The *fast dynamics* describes how neuron activities spread through that network. Hopfield networks (and other so-called *resonance systems*) exhibit a desirable property: when given an input state s the system stabilizes in a certain state.

Fact 1 (Hopfield 1982)
The function $E(s) = -\Sigma_{i>j} w_{ij} \cdot s_i \cdot s_j$ is a Ljapunov-function of the system in the case of an asynchronous update function f. I.e., when the activation state of the network changes, E either decreases or remains the same. The output states $\lim_{n\to\infty} (f^n(s))$ can be characterized as *the local minima* of the Ljapunov-function.



Fact 2 (Hopfield 1982)
The output states $\lim_{n\to\infty} (f^n(s))$ can be characterized as *the global minima* of the Ljapunov-function if certain stochastic update functions f are considered ("simulated annealing").

Definition 1 (asymptotic updates)

$ASUP_w(s) =_{def} \{t: t = \lim_{n\to\infty} f^n(s)\}$ [f asynchr. updates with clamping]

Definition 2 (E-minimal specifications of s)

$min_E(s) =_{def} \{t: t \geq s$ and there is no $t' \geq s$ such that $E(t') < E(t)\}$

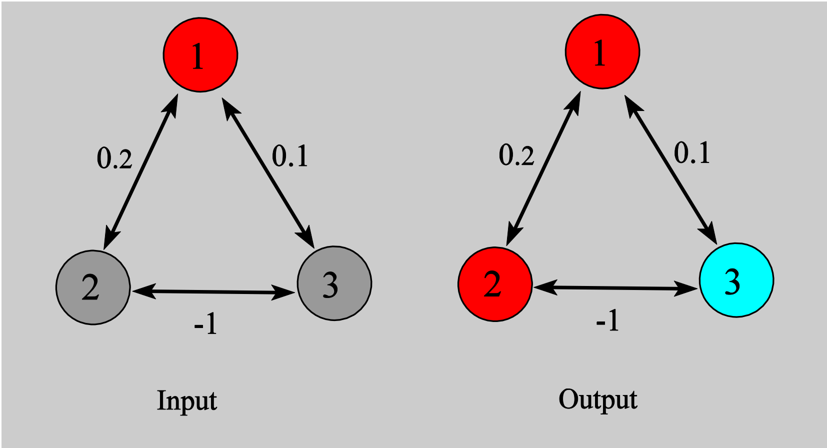
Consequence of fact 2

$ASUP_w(s) =_{def} min_E(s)$, where $E(s) = -\Sigma_{i>j}\, w_{ij} \cdot s_i \cdot s_j$

(energy function)

Example

$$w = \begin{pmatrix} 0 & 0.2 & 0.1 \\ 0.2 & 0 & -1 \\ 0.1 & -1 & 0 \end{pmatrix}$$



			E	
$\langle 1\ 0\ 0\rangle$	$\leq$	$\langle 1\ 0\ 0\rangle$	0	
		$\langle 1\ 0\ 1\rangle$	-0.1	
		$\langle 1\ 1\ 0\rangle$	-0.2	
		$\langle 1\ 1\ 1\rangle$	0.7	
		$\langle 1\ 1\ {-1}\rangle$	-1.1	☜

$$ASUP_w(\langle 1\ 0\ 0\rangle) = \min_E(s) = \langle 1\ 1\ {-1}\rangle$$

Definition 3 (Nonmonotonic inference relation)

$s \mathrel{|\!\sim}_w t$ iff $s' \geq t$ for each $s' \in ASUP_w(s)$

In our example $\langle 1\ 0\ 0\rangle \mathrel{|\!\sim}_w \langle 1\ 1\ {-1}\rangle$

$\langle 1\ 0\ 0\rangle \mathrel{|\!\sim}_w \langle 0\ 1\ 0\rangle$

Fact 3

(i) if $s \geq t$, then $s \mathrel{|\!\sim}_w t$ (SUPRACLASSICALITY)

(ii) $s \mathrel{|\!\sim}_w s$ (REFLEXIVITY)

(iii) if $s \mathrel{|\!\sim}_w t$ and $s \circ t \mathrel{|\!\sim}_w u$, then $s \mathrel{|\!\sim}_w u$ (CUT)

(iv) if $s \mathrel{|\!\sim}_w t$ and $s \mathrel{|\!\sim}_w u$, then $s \circ t \mathrel{|\!\sim}_w u$ (CAUTIOUS MONOTONIC.)

5 Weight-annotated Poole systems

Knowledge base in

(a) connectionist systems:
- connection matrix
- energy function

(b) symbol systems
- strong and weak (default-) rules

At least for Hopfield systems there is a strict relationship between connectionist and symbolic knowledge bases.

☺ Symbolic systems can be used to understand connectionist systems.
☺ Connectionist systems can be used to perform inferences.

Let us consider the language L_{At} of propositional logic (refering to the alphabet At of atomic symbols)

Definition

A triple <At, Δ, g> is called a weight-annotated Poole system iff

(i) At is a nonempty set (of atomic symbols)

(ii) Δ is a set of consistent sentences built on the basis of At (the possible hypotheses)

(iii) $g: \Delta \rightarrowtail [0,1]$ (the weight function)

Definition

Let T = <At, Δ, g> be a weight-annotated Poole system, and let α be a consistent formula.

(A) *A scenario of α in T* is a subset Δ' of Δ such that $\Delta' \cup \{\alpha\}$ is consistent.

(B) *The weight of a scenario* Δ' is

$$G(\Delta') = \Sigma_{\delta \in \Delta'}\, g(\delta) - \Sigma_{\delta \in (\Delta - \Delta')}\, g(\delta)$$

(C) *A maximal scenario of α in T* is a scenario the weight of which is not exceeded by any other scenario (of α in T).

Definition

$\alpha \mathrel{|\!\sim}_T \beta$ iff β is an ordinary conseq. of each maximal scenario of α in T.

An elementary example

$At = \{p_1, p_2, p_3\}$

$\Delta = \{p_1 \leftrightarrow_{0.2} p_2,\ p_1 \leftrightarrow_{0.1} p_3,\ p_2 \leftrightarrow_{1.0} {\sim}p_3\}$

some (relevant) scenarios of p_1:	G	
$\{\}$	-1.3	
$\{p_1 \leftrightarrow p_2\}$	-0.9	
$\{p_1 \leftrightarrow p_2,\ p_1 \leftrightarrow p_3\}$	-0.7	
$\{p_1 \leftrightarrow p_2,\ p_2 \leftrightarrow {\sim}p_3\}$	1.1	☜
$\{p_1 \leftrightarrow p_3,\ p_2 \leftrightarrow {\sim}p_3\}$	0.9	

Consequently, $p_1 \mathrel{|\!\sim}_T p_2$, $p_1 \mathrel{|\!\sim}_T \neg p_3$

The semantics of weight-annotated Poole systems

Let T = <At, Δ, g> be a weight-annotated Poole system, with At = $\{p_1, ..., p_n\}$. Furthermore, let ν denote a (total) interpretation function for the propositional language L_{At} (ν: At ↣ {-1,1}). The usual clauses apply for the evaluation of the formulas of L_{At} relative to ν:

$[\![\alpha \wedge \beta]\!]_\nu = \min([\![\alpha]\!]_\nu, [\![\beta]\!]_\nu)$
$[\![\alpha \vee \beta]\!]_\nu = \max([\![\alpha]\!]_\nu, [\![\beta]\!]_\nu)$
$[\![\sim\alpha]\!]_\nu = -[\![\alpha]\!]_\nu$.

The following defines a function which indicates how strong a given interpretation ν conflicts with the space of hypotheses Δ:

Definition
$\mathscr{E}(\nu) = -\Sigma_{\delta \in \Delta}\, g(\delta) \cdot [\![\delta]\!]_\nu$ (the energy of the interpretation)

Next, the notions of *model* and *preferred model* can be defined:

Definition
(A) An interpretation ν is called a *model* of α just in case $[\![\alpha]\!]_\nu = 1$.
(B) An interpretation ν is called a *preferred model* of α just in case it is a model of α with minimal energy (w.r.t. the other models of α).

The following notion is the semantic counterpart to the syntactic consequence relation $\alpha \mathrel{\supset\!\!\sim}_T \beta$:

Definition
$\alpha \mathrel{\supset\!\!\approx}_T \beta$ iff each preferent model of α is a model of β.

Theorem
For all formulas α and β of L_{At}: $\alpha \mathrel{\supset\!\!\sim}_T \beta$ iff $\alpha \mathrel{\supset\!\!\approx}_T \beta$.

6 Integrating Poole systems and Hopfield networks

Bringing about the correspondence between connectionist and symbolic knowledge bases, we have first to look for a symbolic representation of information states.

Symbolic representation of information states

Let $\langle S\cup\bot, \geq\rangle$ be the extended poset of activation states for a neural network with n elements.

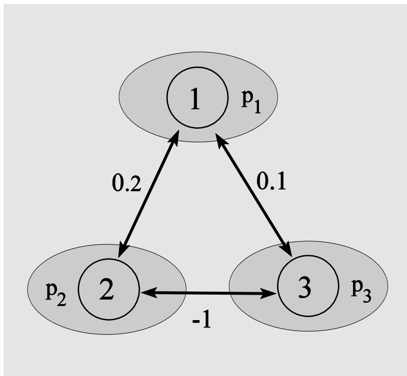
Definition

The triple $\langle S\cup\bot, \geq, \upharpoonleft\ \downharpoonright\rangle$ is called a *Hopfield model* (for L_{At}) iff $\upharpoonleft\ \downharpoonright$ is a function assigning some element of $S\cup\bot$ to each atomic symbol and obtaining the following conditions:

$$\upharpoonleft \alpha\wedge\beta \downharpoonright = \upharpoonleft \alpha \downharpoonright \circ \upharpoonleft \beta \downharpoonright, \quad \upharpoonleft {\sim}\alpha \downharpoonright = -\upharpoonleft \alpha \downharpoonright.$$

A Hopfield model is called *local* (for L_{At}) iff it realizes the following assignments:

$$\upharpoonleft p_1 \downharpoonright = \langle 1\ 0\ \ldots\ 0\rangle$$
$$\upharpoonleft p_2 \downharpoonright = \langle 0\ 1\ \ldots\ 0\rangle$$
$$\ldots$$
$$\upharpoonleft p_n \downharpoonright = \langle 0\ 0\ \ldots\ 1\rangle$$



With regard to local Hopfield models each state can be represented by a conjunction of literals (atoms or their inner negation);
e.g. $\langle 1\ 1\ 0\rangle = \upharpoonleft p_1\wedge p_2 \downharpoonright$, $\langle 1\ 1\ {-1}\rangle = \upharpoonleft p_1\wedge p_2\wedge{\sim}p_3 \downharpoonright$.

Translating Hopfield networks into weight-annotated Poole systems

Consider a Hopfield system (n neurons) with connection matrix w, and let At = $\{p_1, ..., p_n\}$ be a set of atomic symbols. Take the following formulae of L_{At}:

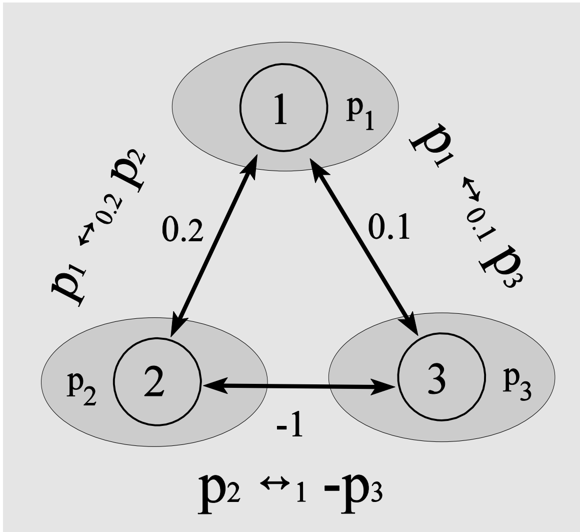
$$\alpha_{ij} = (p_i \leftrightarrow \mathrm{sign}(w_{ij})\, p_j), \text{ for } 1 \leq i < j \leq n$$

Definition

For each connection matrix w the *associated Poole system* is defined as $T_w = \langle At, \Delta_w, g_w \rangle$ where the following clauses apply:

(i) $\Delta_w = \{\alpha_{ij}: 1 \leq i < j \leq n\}$

(ii) $g_w(\alpha_{ij}) = |w_{ij}|$



Under certain conditions (no isolated nodes) it can be shown that each (partial) information state is completed asymtotically. Consequently, $ASUP_w(s)$ contains only total information states. This fact allows us to prove the following theorem:

Theorem

Assume that the formulae α and β are conjunctions of literals. Assume further that the Poole system T is associated to the connection matrix w. Then

$$\lceil \alpha \rfloor \mathrel{\vdash\!\sim}_w \lceil \beta \rfloor \quad \text{iff} \quad \alpha \mathrel{|\!\sim}_T \beta$$

7 Conclusions

☺ Weight-annotated Poole systems can be used to understand connectionist systems. Nonmonotonic inferences ($\alpha \mid\!\sim_T \beta$) as an analytic tool to understand emerging properties of connectionist networks.

☺ Weight-annotated Poole systems are singled out by giving them a "deeper justification".

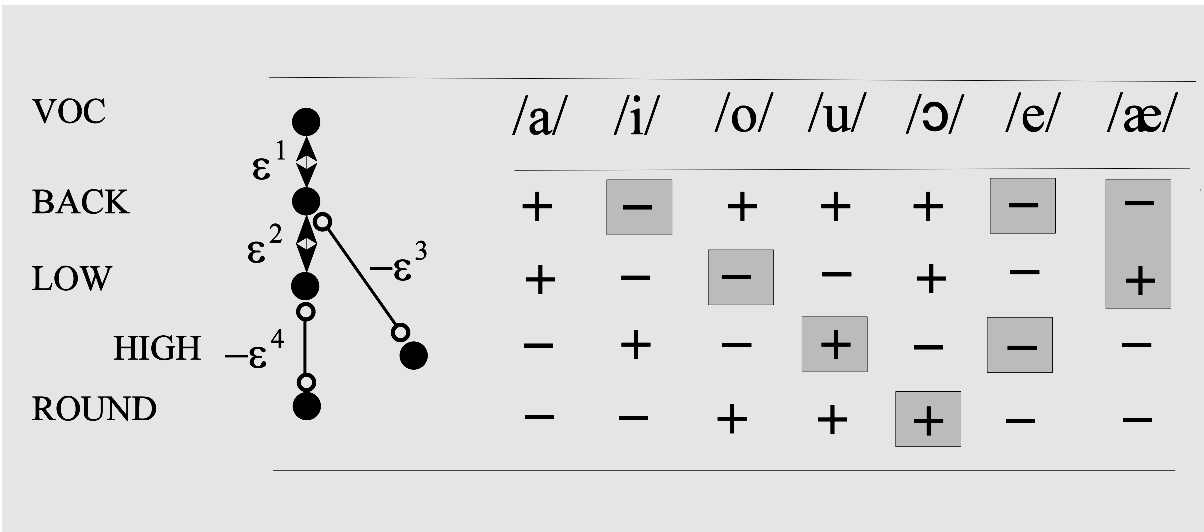
☺ Connectionist systems can be used to perform nonmonotonic inferences. Efficiency?

Appendix: An example from phonology

−back	+back	
/i/	/u/	+high
/e/	/o/	−high/−low
/æ/	/ɔ/	+low
	/a/	

The phonological features may be represented as by the atomic symbols BACK, LOW, HIGH, ROUND. The generic knowledge of the phonological agent concerning this fragment may be represented as a Hopfield network using exponential weights with basis $0 < \varepsilon \leq 0.5$. Furthermore, make use of the following **Strong Constraints**:

LOW → ~HIGH; ROUND → BACK



Assigned Poole-system

VOC $\leftrightarrow_{\varepsilon^1}$ BACK; BACK $\leftrightarrow_{\varepsilon^2}$ LOW
LOW $\leftrightarrow_{\varepsilon^4}$ ~ROUND; BACK $\leftrightarrow_{\varepsilon^3}$ ~HIGH

(These default rules are in strict correspondence to Keane's markedness conventions)