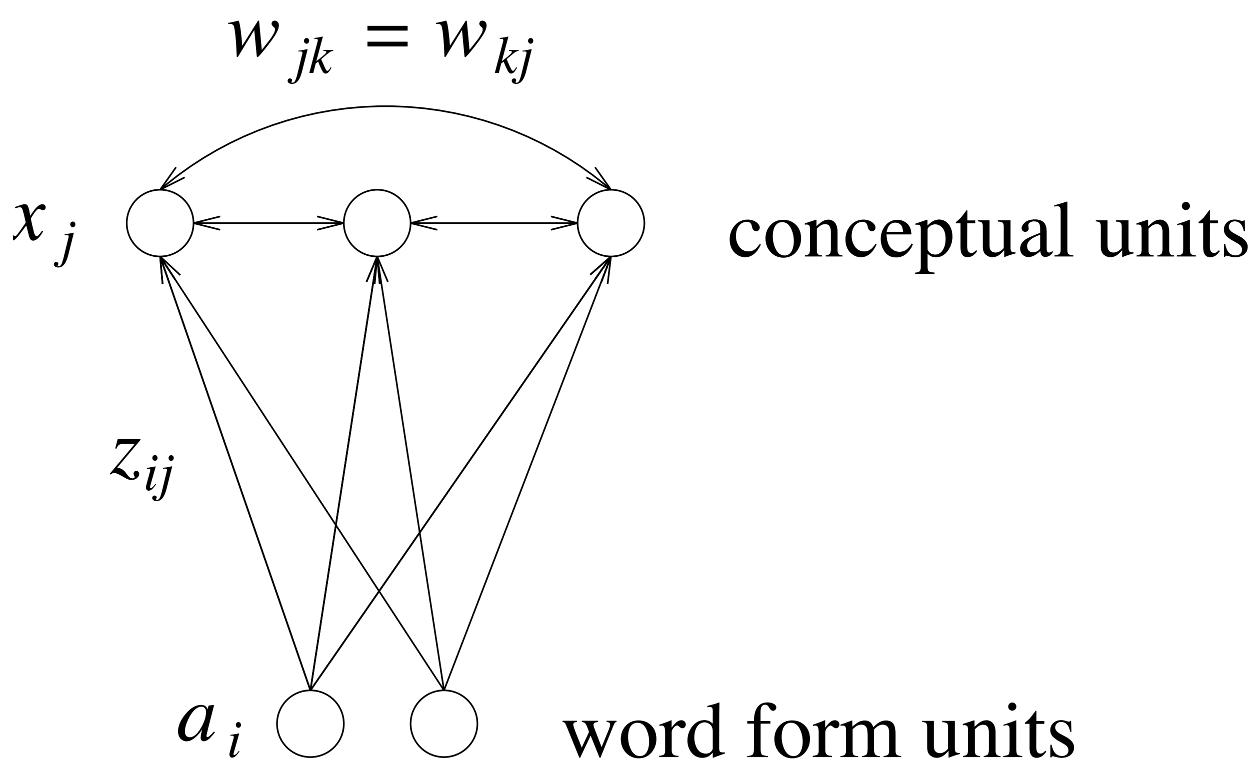
Cogsci

Virginia de Sa

Application of Hopfield Memories to Understanding Human Conceptual Memory
(From a paper with Ken McRae and Mark Seidenberg) Link to paper

Modeling the Computation of meaning from orthographic representation (word)



Input representation : 379 units representing triples of letters (including leading and trailing blanks)

e.g. CANARY is represented by activity on the units _CA CAN ANA ARY RY_

(distributed coding that includes some letter order information)

Output representation: 646 units representing semantic properties collected by Ken

e.g. *has feathers*, *is dangerous*, *has leaves*, ... A conceptual unit is given activation 1 if more than 5 of 30 subjects listed the property for the concept

Connectivity Details

The conceptual units are fully interconnected as a Hopfield network.

The wordform units are feed-forward connected to the Hopfield network.

The idea is that forward activation from the word form units will get you in the vicinity of a word meaning (and also bias the Hopfield units) and the Hopfield net will converge to the meaning.

Activity Weights

The weights between output units were calculated using the modified Hopfield rule of Tsodyks and Feigelman (1988)

$$w_{ij} = \frac{1}{n_1} \sum_p [(x_{ip} - \mu_p)(x_{jp} - \mu_p)]$$

where n_1 is the number of conceptual units and μ_p is the fraction of activated concepts for pattern p, x_{ip} is 1 if pattern p posseses property i (and 0 otherwise)

Weights from word form to conceptual units are calculated using a similar Hebbian style learning rule.

$$z_{ij} = \frac{1}{n_2} \sum_p [(a_{ip})(x_{jp} - \mu_p)]$$

where n_2 is the number of word-form units and a_{ip} is the normalized activation of the i^{th} word-form unit in pattern p.

Word-form patterns were normalized to remove the dependence of word-length according to

$$a_{ip} = \frac{2u_{ip}}{||u_p||} \forall p$$

where u_{ip} is 1 if pattern p contains the triple of letters represented by word form unit i and 0 otherwise;

Simulating the Network

To start we clamp the wordform units for a word and initialize 60 random conceptual units to activity .25 (and the rest 0) then we propagate activity

$$x_i(t+1) = g(c_1 \sum_j (z_{ji} a_j) + c_2 \sum_j (w_{ji} x_j(t)) - \theta)$$

$$g(x) = .5tanh(c_3 x) + .5$$

we used $c_1 = .85$, $c_2 = 0.33$, $c_3 = 400$ and $\theta = .0105$

Results

The Model learned 80 of the 84 trained patterns (with 646 output units)

Some *errors*

has wings: BUDGIE

is an animal: BUZZARD, CANARY, EAGLE

is large: CANARY, CHICKEN, DUCK

is dangerous CANNON

is edible CARROT, RADISH, ZUCCHINI

has leaves CARROT

eats HAWK

has four legs MOUSE

is loud CANARY, MISSILE

worn by women TROUSERS

Simulating Semantic Priming

Semantic Priming Experiment:

Is it an object?

LAMP (200 msec PRIME)

Simulating Semantic Priming

Semantic Priming Experiment:

Is it an object?

LAMP (200 msec PRIME)

########## (50 msec mask)

CHANDELIER

Semantic Priming Simulations

Semantic priming was simulated by activating the word-form for the units and computing its conceptual representation. With the prime's meaning active the word-form units are then clamped for the target word and convergence latency for the target concept is watched.

We used 3 measures of convergence:

- number of iterations for error to drop below 1
- number of iterations for error to be within .1 of its stable value
- number of iterations for error to be within .01 of its stable value

All gave similar results

Semantic Priming Simulations

Findings from Studies on People: **For biological kinds (but not artifacts), the similarity in terms of shared correlated properties between the prime and the target significantly predict priming time. For artifacts (but not biological kinds), the number of shared properties between the prime and the target significantly predict priming time.**

The model showed a similar result (for all 3 measures of convergence). The reason is that biological kinds have many highly correlated properties and this leads to strong weights between the properties and fast activation of the properties from others. Artifacts however have very few correlated properties and the speed of activation of a property is simply a function of whether it is already activated or not.

Simulating Property Verification

Property Verification Experiment:

> indicate as quickly as possible whether the target property is reasonably true of the entity to which the concept name refers

DEER (400 msec)

Simulating Property Verification

Property Verification Experiment:

indicate as quickly as possible whether the target property is reasonably true of the entity to which the concept name refers

DEER (400 msec)

hunted by people

Property Verification Simulations

Findings from Studies on People: **With normed frequency balanced, people are faster to verify a property if it is strongly correlated with other properties of the concept (Faster to verify DEER-**$hunted\ by\ people$ **than**

DUCK-*hunted by people*)