Ling 5701: Lecture Notes 3
A Model of Neural Activation

Contents

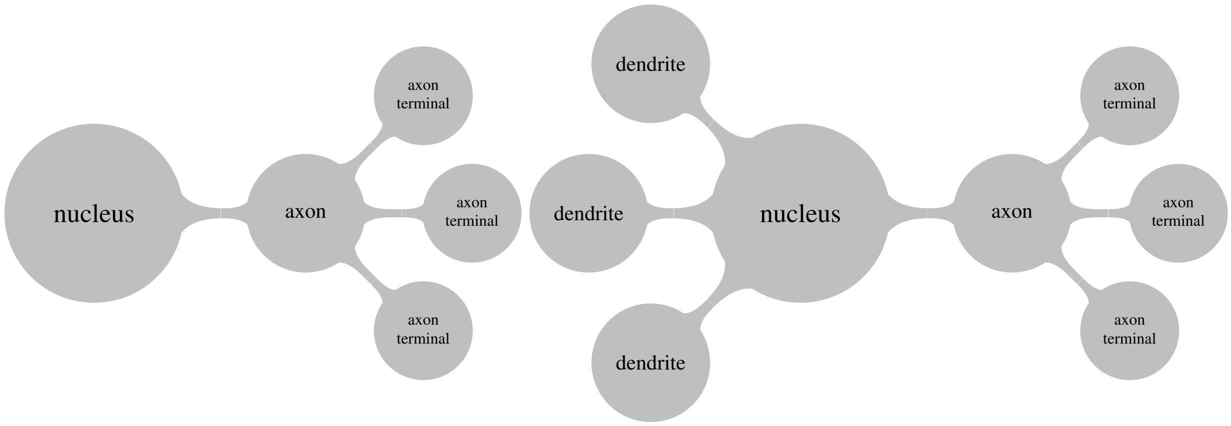
Many psycholinguistic models are defined in terms of neural networks:

- language happens in the brain (brain damage → language damage)
- the brain is composed of neurons
- activation among neurons is associated with linguistic behavior (ERP, FMRI)

3.1 Biology of neural activation

Neurons look like trees, with roots, trunks, and branches. A neuron has:

- **dendrites:** ‘roots’ near other neurons to receive chemical signals
- an **axon:** a ‘trunk’ along which the neuron propagates electric potential
- **axon terminals:** ‘branches’ near other neurons to send chemical signals
- **synapses:** gaps betw. terminals and dendrites that permit thresholding
- **neurotransmitters:** chemicals that carry signals across synapses
- **vesicles:** bubbles in axon terminals that contain neurotransmitters
- **receptors:** attachment sites for neurotransmitters on dendrites



Neurons transmit signals or 'fire' by suddenly changing electric potential:

1. start with more K^+ but much fewer Na^+ ions than outside, creating **membrane potential**;
2. receptors receive neurotransmitters, open **ligand-gated** channels;
3. ligand-gated channels let Ca^{++}/Cl^- in or K^+ out, changing potential
   (this is a **linear** function on the sum of pos/neg ions in the neuron);
4. if potential changes enough, **voltage-gated** channels come open;
5. voltage-gated channels let in many Na^+/Ca^{++} ions; neuron **depolarizes**
   (this is a non-linear **threshold** function on the sum of positive/negative ions in the neuron);
6. depolarization allows **vesicles** to meet surface, release neurotransmitters;
7. depolarization makes voltage-gated channels let out K^+, **repolarize** cell;
8. ion pumps on surface put back Ca^{++},Cl^-,Na^+,K^+, neurotransmitters.

Synaptic connections may be **positive or negative**, e.g.:

1. **pyramidal neurons** may emit neurotransmitters that gate positive ions
2. **interneurons** may emit neurotransmitters that gate negative ions

Synaptic connections also have **weights**:

1. repeated firing removes Mg blockers, so the 'rest state' depolarizes a bit
2. fewer Mg blockers increases phosphate, makes receptors more efficient
3. fewer Mg blockers triggers construction of more receptors (to let in more ions)

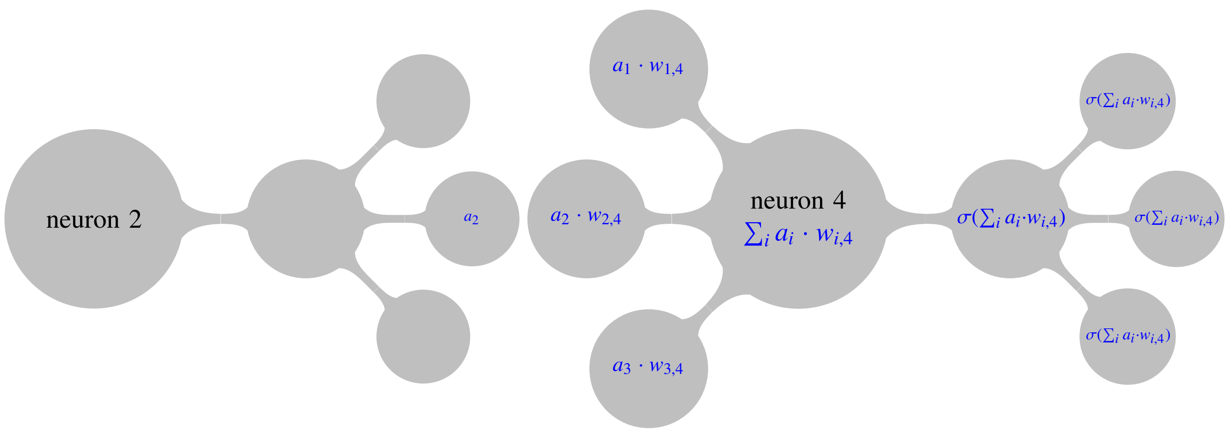
3.2 A model of neural activation

The **linear function** and **threshold function** can modeled mathematically:

1. a_i, a_j: real-valued activation of artificial neural units i and j
2. $w_{i,j}$: real-valued weight (pos/neg) of connection from unit i to unit j
3. $\sum_i a_i \cdot w_{i,j}$: connection-weighted (**linear**) sum of impinging neural units
4. σ: sigmoid (S-shaped) **threshold** function, e.g. logistic: $\sigma(x) = \frac{1}{1+e^{-x}}$

$$a_j = \sigma\left(\sum_i a_i \cdot w_{i,j}\right)$$

For example, if neuron 2 impinges on neuron 4 (neurons 1 and 3 not shown):



But neurons don't have real-valued activation; they fire all-or-none if they reach threshold.

Neural models may therefore be more similar to **clusters** of neurons.

Neurons in the cortex seem to be organized into columnar clusters:

- neurons in the same cluster seem to fire together
- clusters may have real-valued (or at least graded) activation

3.3 Mental states for concepts/referents as distributed activation

Do individual clusters correspond to concepts/referents ('localist' model)? Inconsistent w. plasticity.

Activation for concepts/referents may be instead **distributed** over clusters [Horton and Adams, 2005]:

- **mental states** for concepts or referents are characterized by patterns of activation, e.g.:

'airplane'	'celery'	'my house'	
.0	.0	.58	← neuron/cluster closest to center of motor cortex
.58	.50	.0	← neuron/cluster second closest to center of motor cortex
.0	.50	.58	← neuron/cluster third closest to center of motor cortex
.58	.0	.0	⋮
.0	.50	.0	← neuron/cluster closest to center of auditory cortex
.58	.0	.58	← neuron/cluster second closest to center of auditory cortex
.0	.50	.0	⋮

- maybe 20,000 clusters in human cortex: 20,000-dimensional space; room for many ideas!
  (in contrast, physical space has only 3 dimensions: $L\times W\times H$, color space has 3: $R\times G\times B$)
- mental states for concepts are locations/regions/coordinates in this space ('vector-space')
- there's no actual limit on the number of states/concepts/referents, just potential for confusability
- if sparsely encoded (many units inactive), we can have mixture states of several referents at once!

3.4 Models of activation over time

Over time (e.g. during sentence processing), the activation of neurons/clusters changes.

These changes can be modeled with **Recurrent Neural Networks** [Elman, 1991]:

- the model is defined in terms of a 'context' vector of neural units, as shown above;
- activation of the context vector defines a mental state, as noted above;
- the context vector is connected to sensory units (observations);
- the context vector is also connected to *itself* at previous time step, forming a **circuit**;
- the model learns to **transition** between states by associating each previous and current state
  (these associations are determined by synaptic weights, as we'll see later);
- the learned transitions define a sequence of mental states for any sequence of observations.

We will assume this kind of transition model, with transitions defined by synaptic weights.

Experiments with these models have shown learning of syntax:

- word order predictions
- number agreement

3.5 Mental states for concepts are composed of features

Mental states for concepts are distributed over the cortex in different brain areas:

- visual cortex (posterior)
- auditory cortex (medial, bilateral)
- motor cortex (medial, dorsal)

Mental states therefore have various **features**: visual, auditory, proprioceptive, ...

- features may be encoded by several neuron or cluster units (boxes in the vectors)

Working memory may be modeled with **temporal features** [Howard and Kahana, 2002]:

- temporal feature values change over time
- recurrent learning builds associations between present and past contexts
- recent past events easily cued from current temporal features (STM)
- distant past events cued not so easily, need other features (LTM)

References

[Elman, 1991] Elman, J. L. (1991). Distributed representations, simple recurrent networks, and grammatical structure. *Machine Learning*, 7:195–225.

[Horton and Adams, 2005] Horton, J. C. and Adams, D. L. (2005). The cortical column: a structure without a function. *Philosophical Transactions of the Royal Society of London - Series B: Biological Sciences*, 360(1456):837–862.

[Howard and Kahana, 2002] Howard, M. W. and Kahana, M. J. (2002). A distributed representation of temporal context. *Journal of Mathematical Psychology*, 45:269–299.