Text-to-speech and neural networks
Metric spaces and neural networks
Nonmetric spaces and neural networks

Course Trajectory

<table>
<thead>
<tr>
<th>When</th>
<th>What</th>
</tr>
</thead>
<tbody>
<tr>
<td>Weeks 1-3</td>
<td>Basics of three architectures (DAM, perceptron, MLP)</td>
</tr>
<tr>
<td>Weeks 4-6</td>
<td>Cognitive science of DAMs and perceptrons</td>
</tr>
<tr>
<td>Week 7</td>
<td>Connectionism and Cognitive Psychology</td>
</tr>
<tr>
<td>Weeks 8-10</td>
<td>Interpreting MLPs</td>
</tr>
<tr>
<td>Weeks 11-13</td>
<td>Case studies (interpretations, applications, architectures)</td>
</tr>
</tbody>
</table>

Discussion?

- Questions, comments or issues?

Case Study 1: Text To Speech

- There is a long history of developing technologies for converting text to speech
- Klatt (1987) published an influential review that covered several decades of research in this domain

Text To Speech Challenges

- Converting text to speech is a nontrivial task
- There are many irregular relationships between graphemes and phonemes - Ghti could be pronounced fish, from graphemes in enough and nation
- In general, text must be converted to various abstract linguistic representations in order to generate proper, realistic speech
- And the generation of realistic sounding speech is nontrivial too!

Klattalk

- Klatt’s pioneering work at MIT produced a system that was first known as MITalk and later as Klattalk
- Klatt’s interest was in exploring a diversity of representations of text, linguistic structures, and speech variables
- The synthesizer component of Klattalk alone required specifying the values of 19 different control parameters.
**DECtalk**

- DEC developed Klattalk into a commercial system called DECtalk.
- It was a computer peripheral that sold for about $4000 in the 1980s.
- DECtalk looked text up in a dictionary (of irregular words) that converted it to sets of phonemes.
- Otherwise, DECtalk used a set of grapheme-to-phoneme rules.
- DECtalk’s dictionary held 15,000 irregular words and it used more than 1,500 rules.

**Replacing DECtalk**

- Terry Sejnowski and Charles Rosenberg were interested in replacing DECtalk with a neural network.
- Can one replace the thousands of rules and irregular examples of DECtalk with a small, distributed network that can handle regular and irregular pronunciations?

**NETtalk**

- NETtalk was a neural network trained to generate the same I/O behavior as DECtalk.
- 7 groups of 29 input units per group represent letters.
- 80 hidden units.
- Task: generate phoneme for middle input group.
- Trained with generalized delta rule on corpus of 1024 words, informal text.

**NETtalk Learning**

- NETtalk achieved 90% performance after being trained on only 5000 stimuli.
- Klatt was not completely impressed: “In some sense, this is a surprisingly good result in that so much knowledge could be embedded in a moderate number of about 25,000 weights, but the performance is not nearly as accurate as that of a good set of letter-to-sound rules” (Klatt, 1987, p. 770).

**NETtalk Representations**

- Because of its size and complexity, the internal structure of NETtalk was not investigated in detail.
- Sejnowski and Rosenberg explored the hidden layer with Hinton diagrams.
- They concluded that most of the representations were “distributed.”
During the early stages of learning in NETtalk, the sounds produced by the network are uncannily similar to early speech sounds of children. Examples of NETtalk

The phonological mappings produced by NETtalk are efficient encodings for a parallel network and may be comparable to those used by humans.

Descendants of NETtalk have been central in the debate about the kinds of model required to account for reading, as well as symptoms of dyslexia.

Dyslexia is a disorder in reading of words, and can be related to brain injury. 

- Phonological dyslexia is a disorder in which nonwords cannot be read, but the reading of words is unaffected.
- Surface dyslexia is a selective disorder in which there is severe difficulty in reading aloud irregular words, usually revealed in terms of generalization errors; nonwords can be read.
- Deep dyslexia involves semantic errors in reading aloud, visual errors, and an inability to read nonwords.

Dyslexia's symptoms are difficult to explain using simple boxologies:

1. Semantic errors (e.g., BLOWING “wind”, VIEW “scene”, NIGHT “sleep”, GONE “lost”).
2. Visual errors (e.g., WHILE “white”, SCANDAL “sandals”, POLITIC “politics”, BADGE “bandage”).
3. Function-word substitutions (e.g., WAS “and”, ME “my”, OFF “from”, THEY “the”).
4. Derivational errors (e.g., CLASSIFY “class”, FACT “facts”, MARRIAGE “married”, BUY “bought”).
5. Non-lexical derivation of phonology from print is impossible (e.g., pronouncing nonwords, judging if two nonwords rhyme).
6. Lexical derivation of phonology from print is impaired (e.g., judging if two words rhyme).
7. Words with low imageability/concreteness (e.g., JUSTICE) are harder to read than words with high imageability/concreteness (e.g., TABLE).
8. Verbs are harder than adjectives which are harder than nouns in reading aloud.
9. Functions words are more difficult than content words in reading aloud.
10. Writing is impaired (spontaneous or to dictation).
12. Whether a word can be read at all depends on its sentence context (e.g., FLY as a noun is easier than FLY as a verb).

The success of NETtalk paved the way for other researchers to explore networks that converted text into something else. Geoffrey Hinton and Tim Shallice, for instance, began to study networks that were models of reading. These networks mapped, for example, graphemes to phonemes— but included intermediate semantic representations too. Issue was whether such models could provide an alternative to classical, multiple route models, like Coltheart’s DRC.

Hinton and Shallice began with a small set of primitive features— letters, words, and semantic features, and defined mappings between them.
Hinton and Shallice explored a number of different network architectures to map one kind of feature into another, all motivated as models of reading. Key architecture mapped graphemes through semantics to phonemes.

A Variety Of Architectures

Connectionist Neuroscience

Hinton and Shallice explored the effects of a variety of lesions of their networks after training was completed. They produced errors associated with deep dyslexia.

A Variety Of Architectures

Hinton and Shallice explored a number of different network architectures to map one kind of feature into another, all motivated as models of reading. Key architecture mapped graphemes through semantics to phonemes.

Connectionist Neuroscience

Hinton and Shallice explored the effects of a variety of lesions of their networks after training was completed. They produced errors associated with deep dyslexia.

A Variety Of Architectures

Hinton and Shallice explored a number of different network architectures to map one kind of feature into another, all motivated as models of reading. Key architecture mapped graphemes through semantics to phonemes.

Connectionist Neuroscience

Hinton and Shallice explored the effects of a variety of lesions of their networks after training was completed. They produced errors associated with deep dyslexia.

A Variety Of Architectures

Hinton and Shallice explored a number of different network architectures to map one kind of feature into another, all motivated as models of reading. Key architecture mapped graphemes through semantics to phonemes.

Connectionist Neuroscience

Hinton and Shallice explored the effects of a variety of lesions of their networks after training was completed. They produced errors associated with deep dyslexia.
A Converged Network

- With 169 training patterns, the network converged in 5078 sweeps
- It had internalized a metric space
- The weights at the end of training were highly systematic
- However, the weights did not appear to represent distances!
- For instance, weights could not be used to predict distances on the map

Hidden Units In Space

- Dawson et al. (2005) were able to treat hidden units as if they occupied particular locations on the map of Alberta
- Key property: Hidden units defined a plane with a particular direction of view
- Direction was how distance was being measured!
- Imagine being on a flat plane with a small number of distinct landmarks (e.g., trees) in view, and having the task of measuring the angular displacement between pairs of landmarks. If the only tool available was a sextant, 
so that the only measurement that could be taken was the angular separation between pairs of landmarks, one would have to make a rough estimate of the distance between locations by taking a sextant reading between pairs of landmarks. Nearly meaning would be that if a sextant reading was high, then the two landmarks were far apart, and if it was near zero, then the two landmarks were near one another” (Dawson et al., 2005, p. 44)

Hidden Unit As Sextant

- Each hidden unit could be seen as a sextant, delivering angles or bearings towards pairs of cities
- Connection weights were strongly correlated with this model
- But this means that each hidden unit delivers an inaccurate distance measure

Coarse Coding

- Individual inaccuracy is dealt with by having multiple views (multiple hidden units with different bearings)
- Pooling these inaccurate, but varied, responses together generates accurate distance readings
- A committee of sextants!
- This kind of coding should be able to cope with violations of metric properties!

Case Study 3: Nonmetric Space

- Judgments of similarity are not symmetric
  - The judged similarity of North Korea to China exceeds the judged similarity of China to North Korea
- Judgements of similarity violate the triangle inequality
  - Jamaica is similar to Cuba
  - Cuba is similar to Russia
  - but Jamaica is not similar to Russia at all

Antisymmetric Space

- Direction is a spatial relation that is a radical violation of symmetry – it is antisymmetric
- This is because if \( d(x,y) \) delivers direction, then \( d(x,y) = -d(y,x) \)
- Dawson & Boechler (2007) explored a multilayered network of value units that was trained to deliver directional judgments
- Given two cities, deliver the compass rose direction from one to the other
Asymmetric Training

- Again, a network with 7 hidden units, trained on 169 patterns, converged after 7645 sweeps of training.
- Hidden unit behavior reflected the asymmetry of the task.
- Hidden units in a 13 x 13 city matrix had large asymmetries of both net inputs and of activities.

<table>
<thead>
<tr>
<th>Hidden Unit</th>
<th>Preferred Attraction</th>
<th>preferred Attraction Off</th>
<th>Correlation Between</th>
<th>&quot;W&quot;</th>
<th>&quot;W&quot;</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>-0.2</td>
<td>-0.13</td>
<td>0.72</td>
<td>0.52</td>
<td>0.32</td>
</tr>
<tr>
<td>2</td>
<td>-0.3</td>
<td>-0.15</td>
<td>0.71</td>
<td>0.50</td>
<td>0.31</td>
</tr>
<tr>
<td>3</td>
<td>-0.4</td>
<td>-0.17</td>
<td>0.69</td>
<td>0.48</td>
<td>0.29</td>
</tr>
<tr>
<td>4</td>
<td>-0.5</td>
<td>-0.19</td>
<td>0.67</td>
<td>0.46</td>
<td>0.27</td>
</tr>
<tr>
<td>5</td>
<td>-0.6</td>
<td>-0.21</td>
<td>0.65</td>
<td>0.44</td>
<td>0.26</td>
</tr>
<tr>
<td>6</td>
<td>-0.7</td>
<td>-0.23</td>
<td>0.63</td>
<td>0.42</td>
<td>0.24</td>
</tr>
<tr>
<td>7</td>
<td>-0.8</td>
<td>-0.25</td>
<td>0.60</td>
<td>0.40</td>
<td>0.23</td>
</tr>
</tbody>
</table>

More Coarse Coding

- When hidden unit activity was plotted in terms of two cities in the context of a preferred direction, it was clear that the hidden unit system coarse coded output direction.

Coarse Coding Again

- Coarse coding is further revealed by looking at patterns of activity when city pairs are presented that all map onto the same output direction.
- Interaction of three hidden units is required; simple local feature detection is not evident!

Head Direction Cells

- The hidden units in the network are analogous to head direction cells.
- These cells are also coarsely tuned.
- Current theories combine these cells into a system of overlapping sensitivities, as coarse coding would require.
- The bottom figure shows networks of cells; the greyer the cell the higher the activity.
- Head direction is mediated by parallel processing.