Learning Virtual Sensors for Cognitive States
(10-702 Lecture 1/13/03)

Tom M. Mitchell
Rebecca Hutchinson, Marcel Just,
Sharlene Newman, Francisco Pereira,
Radu Niculescu, Xuerui Wang

Paper draft available:
www.cs.cmu.edu/~tom/nips02.ps
Can we train classifiers to decode instantaneous cognitive state?

- → virtual sensors for cognitive states
- much analysis of *average* fMRI response
- little consideration of this question!
Classifiers for Cognitive States

Difficult!

– Data very noisy
– High dimensional
– Sparse training data
Approach

• Learn \( fMRI(t, \ldots, t+k) \rightarrow \text{CognitiveState} \)

• Classifiers:
  – Gaussian Naïve Bayes, SVM, kNN

• Feature selection/abstraction
  – Select subset of voxels (by signal, by anatomy)
  – Select subinterval of time
  – Average activities over space, time
  – Normalize voxel activities
  – …
Study 1: Word Categories

[Francisco Pereira]

- Family members
- Occupations
- Tools
- Kitchen items
- Dwellings
- Building parts

- 4 legged animals
- Fish
- Trees
- Flowers
- Fruits
- Vegetables
Word Categories Study

- Ten neurologically normal subjects
- Stimulus:
  - 12 blocks of words:
    - Category name (2 sec)
    - Word (400 msec), Blank screen (1200 msec); answer
    - Word (400 msec), Blank screen (1200 msec); answer
    - ...
  - Subject answers whether each word in category
  - 32 words per block, nearly all in category
  - Category blocks interspersed with 5 fixation blocks
Training Classifier for Word Categories

Learn $fMRI(t) \rightarrow \text{word-category}(t)$
- $fMRI(t) = 8470$ to $11,136$ voxels, depending on subject

Feature selection: Select $n$ voxels
- Best single-voxel classifiers
- Strongest contrast between fixation and some word category
- Strongest contrast, spread equally over ROI’s
- Randomly

Training method:
- train ten single-subject classifiers
- Gaussian Naïve Bayes $\rightarrow P(fMRI(t) \mid \text{word-category})$
Results

Classifier outputs ranked list of classes
Evaluate by the fraction of classes ranked ahead of true class
– 0=perfect, 0.5=random, 1.0 unbelievably poor

<table>
<thead>
<tr>
<th>Experiment</th>
<th>A</th>
<th>B</th>
<th>C</th>
<th>D</th>
<th>E</th>
<th>F</th>
<th>G</th>
<th>H</th>
<th>I</th>
<th>J</th>
</tr>
</thead>
<tbody>
<tr>
<td>12 classes</td>
<td>0.13</td>
<td>0.17</td>
<td>0.04</td>
<td>0.12</td>
<td>0.06</td>
<td>0.069</td>
<td>0.2</td>
<td>0.04</td>
<td>0.14</td>
<td>0.05</td>
</tr>
<tr>
<td>6 classes</td>
<td>0.45</td>
<td>0.52</td>
<td>0.4</td>
<td>0.5</td>
<td>0.42</td>
<td>0.38</td>
<td>0.52</td>
<td>0.35</td>
<td>0.50</td>
<td>0.29</td>
</tr>
</tbody>
</table>

Try abstracting 12 categories to 6 categories
e.g., combine “Family Members” with “Occupations”
Impact of Feature Selection

<table>
<thead>
<tr>
<th>Feature Selection Method</th>
<th>A</th>
<th>B</th>
<th>C</th>
<th>D</th>
<th>E</th>
<th>F</th>
<th>G</th>
<th>H</th>
<th>I</th>
<th>J</th>
</tr>
</thead>
<tbody>
<tr>
<td>Feature Average Rank</td>
<td>0.11</td>
<td>0.18</td>
<td>0.04</td>
<td>0.12</td>
<td>0.065</td>
<td>0.085</td>
<td>0.20</td>
<td>0.047</td>
<td>0.13</td>
<td>0.054</td>
</tr>
<tr>
<td>Activity p-value</td>
<td>0.049</td>
<td>0.042</td>
<td>0.024</td>
<td>0.07</td>
<td>0.029</td>
<td>0.032</td>
<td>0.049</td>
<td>0.02</td>
<td>0.078</td>
<td>0.013</td>
</tr>
<tr>
<td>Activity p-value per ROI</td>
<td>0.13</td>
<td>0.15</td>
<td>0.051</td>
<td>0.12</td>
<td>0.058</td>
<td>0.068</td>
<td>0.13</td>
<td>0.039</td>
<td>0.15</td>
<td>0.042</td>
</tr>
<tr>
<td>Random</td>
<td>0.14</td>
<td>0.18</td>
<td>0.065</td>
<td>0.13</td>
<td>0.072</td>
<td>0.082</td>
<td>0.22</td>
<td>0.052</td>
<td>0.15</td>
<td>0.059</td>
</tr>
<tr>
<td>Use all voxels</td>
<td>0.13</td>
<td>0.17</td>
<td>0.044</td>
<td>0.12</td>
<td>0.058</td>
<td>0.069</td>
<td>0.2</td>
<td>0.043</td>
<td>0.14</td>
<td>0.05</td>
</tr>
</tbody>
</table>

(a) Voxel in 1200 most active only
(b) Voxel in 1200 best discriminators only
(c) Voxel in both groups
(a) Distribution of discriminating voxels for subject 1 (the more red the better)

(b) Overlap (red) of the active (yellow) and discriminating (blue) 1200 voxel subsets for sub. 1
(c) Distribution of discriminating voxels for subject J (the more red the better)

(d) Overlap (red) of the active (yellow) and discriminating (blue) 1200 voxel subsets for sub. J
(c) Distribution of discriminating voxels for subject J (the more red the better)

(a) Distribution of discriminating voxels for subject I (the more red the better)
[Haxby et al., 2001]

(c) Distribution of discriminating voxels for subject J (the more red the better)
Study 1: Summary

- Able to classify single fMRI image by word category block
- Feature selection important
- Is classifier learning word category or something else related to time?
  - Accurate across ten subjects
  - Relevant voxels in similar locations across subjs
  - Locations compatible with earlier studies
Cognitive model:

Observed image sequence:

Hypothesized intermediate states, representations, processes:

What We’d Like
Challenge: virtual sensors to track sequence of cognitive states

\[ a = 6, \ldots \quad 3x + a = 2 \]

- recall correct
- recall error
- transform correct
- transform error
- answer
- read problem

\[ \text{start} \]

\[ \text{time} \rightarrow \]