Predicting Human Brain Activity Associated with the Meanings of Nouns

Tom M. Mitchell, Svetlana V. Shinkareva, Andrew Carlson, Kai-Min Chang, Vicente L. Malave, Robert A. Mason, Marcel Adam Just

Presented by Robert Forgues
November 19th, 2010
Outline

- Functional Magnetic Resonance Imaging
- The Model
  - Motivation
  - Construction
  - Approach
  - Results
  - Discussion
Functional MRI (fMRI)

- Measures blood flow changes in the brain
- 3D images are generated by thinking about words
- Spatial resolution ~1mm
- Temporal resolution ~1 image/sec
Getting fMRI Data

- Images shown to participant as white lines on dark background
- Participant thinks about properties of the item
- Event-related paradigm
  - 3 seconds of stimulus
  - 7 seconds of X
fMRI Example: Bottle

- Red areas denote high activation
- Mean activation is over 60 different stimuli
- The difference images are used during analysis
Motivation for the Model
How Does The Brain Represent Conceptual Knowledge?

- Studied by many fields
  - Neuroscience
  - Linguistics
  - Computational Linguistics
  - Psychology
- Competing theories
- Can we predict it?
Problem

- How to predict fMRI activation?
  - Statistics to the rescue
- Theory behind the model
  - The neural basis of the semantic representation of concrete nouns is related to the distributional properties of those words in a broadly based corpus of the language
The Model
Building a Predictive Model

Predictive model

stimulus word "celery"

Intermediate semantic features extracted from trillion-word text corpus

Mapping learned from fMRI training data

predicted activity for "celery"
Intermediate Semantic Features

- 25 semantic features defined by their co-occurrence with 25 verbs:
  - Sensory: fear, hear, listen, see, smell, taste, touch
  - Motor: eat, lift, manipulate, move, push, rub, run, say
  - Abstract: approach, break, clean, drive, enter, fill, near, open, ride, wear
Text Corpus

- Provided by Google
- Available from Linguistic Data Consortium
- A trillion-word corpus with (1-5)-grams
- Consists of public English web pages
fMRI Data Collection / Processing

- Test Subjects
  - 9 people (5 female, age 18 – 32, right-handed)
  - Each participant generated set of properties for each item prior to session
  - Tasked with thinking of properties of exemplar
  - Consistency across participants unenforced
fMRI Data Collection / Processing

- **The Machine**
  - Siemens Allegra 3.0T scanner
  - Gradient echo EPI pulse sequence
    - TR = 1000 ms, TE = 30 ms, 60° flip angle
  - Seventeen 5-mm oblique-axial slices imaged
    - 1mm gap between slices
  - 64 x 64 acquisition matrix
    - 3.125-mm x 3.125-mm x 5-mm voxels
fMRI Data Collection / Processing

- Stimuli
  - Line drawings of 60 concrete objects from 12 semantic categories
  - Presented 6 times each, randomly permuted for each presentation
  - 3s exposure, 7s rest period
  - 12 extra rest periods of 31s, scattered across session to provide baseline measure
### fMRI Data Collection / Processing

<table>
<thead>
<tr>
<th>Category</th>
<th>Exemplar 1</th>
<th>Exemplar 2</th>
<th>Exemplar 3</th>
<th>Exemplar 4</th>
<th>Exemplar 5</th>
</tr>
</thead>
<tbody>
<tr>
<td>animals</td>
<td>bear</td>
<td>cat</td>
<td>cow</td>
<td>dog</td>
<td>horse</td>
</tr>
<tr>
<td>body parts</td>
<td>arm</td>
<td>eye</td>
<td>foot</td>
<td>hand</td>
<td>leg</td>
</tr>
<tr>
<td>buildings</td>
<td>apartment</td>
<td>barn</td>
<td>church</td>
<td>house</td>
<td>igloo</td>
</tr>
<tr>
<td>building parts</td>
<td>arch</td>
<td>chimney</td>
<td>closet</td>
<td>door</td>
<td>window</td>
</tr>
<tr>
<td>clothing</td>
<td>coat</td>
<td>dress</td>
<td>pants</td>
<td>shirt</td>
<td>skirt</td>
</tr>
<tr>
<td>furniture</td>
<td>bed</td>
<td>chair</td>
<td>desk</td>
<td>dresser</td>
<td>table</td>
</tr>
<tr>
<td>insects</td>
<td>ant</td>
<td>bee</td>
<td>beetle</td>
<td>butterfly</td>
<td>fly</td>
</tr>
<tr>
<td>kitchen utensils</td>
<td>bottle</td>
<td>cup</td>
<td>glass</td>
<td>knife</td>
<td>spoon</td>
</tr>
<tr>
<td>man made objects</td>
<td>bell</td>
<td>key</td>
<td>refrigerator</td>
<td>telephone</td>
<td>watch</td>
</tr>
<tr>
<td>tools</td>
<td>chisel</td>
<td>hammer</td>
<td>pliers</td>
<td>saw</td>
<td>screwdriver</td>
</tr>
<tr>
<td>vegetables</td>
<td>carrot</td>
<td>celery</td>
<td>corn</td>
<td>lettuce</td>
<td>tomato</td>
</tr>
<tr>
<td>vehicles</td>
<td>airplane</td>
<td>bicycle</td>
<td>car</td>
<td>train</td>
<td>truck</td>
</tr>
</tbody>
</table>
fMRI Data Collection / Processing

• Data Processing
  - Statistical Parametric Mapping software SPM2
  - Corrected for slice timing, motion, linear trend
  - Temporally filtered using a 190s cutoff
  - Normalized into MNI and resampled
  - A single fMRI mean image was created for each presentation of an item by taking mean of item
The Model: A Two-Step Approach

- Step 1: lookup stimulus word, generate normalized semantic feature vector

Semantic feature values: “celery”
- 0.8368, eat
- 0.3461, taste
- 0.3153, fill
- 0.2430, see
- 0.1145, clean
- 0.0600, open
- 0.0586, smell
- 0.0286, touch
- ...
- ...
- 0.0000, drive
- 0.0000, wear
- 0.0000, lift
- 0.0000, break
- 0.0000, ride
The Model: A Two-Step Approach

- Step 2: train the model so that it can predict neural activity

\[ y_v = \sum_{i=1}^{n} c_{vi} f_i(w) \]
Training

- Alternative models trained based on different sets of intermediate semantic features
- Trained and evaluated using cross validation
  - Trained repeatedly with 58 / 60 stimuli
  - Tested with 2 stimuli left out
  - On each iteration, model was given stimuli and corresponding fMRI, required to match
  - Performed 1770 times
The Model: Theoretical Details

- Two key theoretical assumptions:
  1. semantic features are reflected in statistics
  2. some thoughts can be represented as a linear sum

- The training data determines which locations are modulated by which aspects of word meanings (all voxels are considered)
  - 500,000 parameters, for each 25 verbs, there were 20,000 voxels that coefficients were learned on
Results
Celery and Airplane

Predicted and observed fMRI images for celery and airplane after training on 58 other words
A Closer Look

- The model trained on 58 / 60 words and has 1770 test pairs in leave-2-out
- How well can it predict fMRI images for the other two words?
  - Chance is 0.50
  - Accuracy above 0.61 is significant ($p < 0.05$)
  - The accuracy for subjects P1 through P9 was 0.83, 0.76, 0.78, 0.72, 0.78, 0.85, 0.73, 0.68, 0.82
  - Mean accuracy over the 9 subjects was 0.77
Accuracy

- The model is differentially accurate in different parts of the brain

Rendering of the correlation between predicted and actual voxel activations for words outside the training set. Clusters contain at least 10 contiguous voxels.
Can We Extrapolate Beyond Training Data?

• Predicting in a new semantic category
  - Retrain, but drop all examples belonging to the same semantic category as the two held-out words
  - Mean = 0.70

• Predicting when the two held-out words are in the same category
  - Harder to differentiate between words
  - Mean = 0.62
Testing Even Greater Diversity

- The model was given 1000 high frequency words
- Tested with leave-one-out, using 59 / 60 nouns
- Given the fMRI for the hold out, and 1001 candidate words, can it predict the image?
  - Mean over 9 participants = 0.72
Evaluating The Model Beyond Qualitative Measures

- Examining the learned basis set of fMRI images for the 25 verbs
  - The learned signatures cause the model to predict neural activity representing a noun in brain areas that perceive actions described by a verb to the degree that the noun co-occurs with said verb
  - Example: noun $n$ will exhibit neural activity in the *gustatory cortex* to the degree it co-occurs with *eat*
Surprising Result

The model generated all of these images from the 20,000 coefficients learned from the verbs. It was not given any information with respect to physical location voxels. This emerged from training.
How About Other Semantic Features?
Discussion

- There is a direct and predictive relationship between statistics of word co-occurrence in text and neural activation associated with word meanings.

- Neural representations of concrete nouns are in part grounded in sensory-motor features, but involve other regions as well.
Questions?