
Using Machine Learning and Cognitive Modeling to Understand the fMRI-Measured Brain Activation Underlying the Representations of Words and Sentences
(0423070-FY04)

Tom M. Mitchell
Carnegie Mellon University

Marcel A. Just
Carnegie Mellon University

A number of recent fMRI studies have reported significant and repeatable differences in fMRI brain activation when human subjects perceive pictures or words describing objects from different semantic categories (e.g., pictures or words that describe tools, buildings, or people). It is currently possible to determine with good accuracy which of several semantic categories a person is thinking about, based on their brain activation.

We propose new research that builds on these recent discoveries, and seeks to understand (1) human brain activity associated with different semantic categories of objects and actions (nouns and verbs); (2) whether the brain activity associated with semantic categories can be partitioned into more primitive semantic components (e.g., does the brain activity associated with words about tools factor into one component characterizing the tool's visual appearance and a second component characterizing the motor actions involved in using the tool?); and (3) how brain activity associated with individual words is combined into more complex patterns when reading word pairs or simple phrases and sentences.

This research involves:(1) applying machine learning algorithms to discover cortex-wide brain activation patterns associated with particular semantic domains, (2) developing a computational model of human language processing that instantiates the representational principles discovered and that makes specific, testable predictions, and (3) conducting new fMRI studies to obtain novel data about human semantic category representations.

The intellectual merit of the proposed research is multifaceted. If successful, our research will lead to new scientific insights into how the brain organizes information about meanings of words, objects, and actions. It will also lead to new methods for fMRI data analysis, especially for discovering complex temporal-spatial patterns of fMRI activation that accurately distinguish different mental states. The research will also lead toward a new paradigm for developing computational cognitive models and fitting them to empirical data obtained from fMRI and from behavioral measures.

The broader impacts of the proposed research will be amplified by specific outreach activities to several communities. In addition to publishing our scientific results in the cognitive and computational neuroscience literature, we will also actively engage this community by disseminating our new experimental fMRI data through the NSF-funded fMRI Data Center, and by documenting and publishing our new data analysis algorithms on the internet. We will proactively engage the statistical machine learning community, which has much to contribute to development of new fMRI analysis methods, and will develop and disseminate teaching materials for the undergraduate and graduate educational community,including fMRI data sets.

Finally, our proposed research has potential impact on the medical research community, especially regarding the study of neurological conditions such as Alzheimer's disease, dyslexia and high-functioning autism - three areas entailing a language disturbance in which we already have active research collaborations, providing a direct conduit for transferring new scientific insights that may arise from this research.