Artificial Intelligence and Human Brain Imaging

Tom M. Mitchell

Center for Automated Learning and Discovery
Carnegie Mellon University, USA

Abstract. For many years AI researchers have sought to understand the nature of intelligence primarily by creating artificially intelligent computer systems. Studies of human intelligence have had less influence on AI, partly because of the great difficulty in directly observing human brain activity. In recent years, new methods for observing brain activity have become available, notably functional Magnetic Resonance Imaging (fMRI) which allows us to safely, non-invasively capture images of activity across the brain once per second, at millimeter spatial resolution. The advent of fMRI has already produced dramatic new insights into human brain activity, and how it varies with cognitive task. This breakthrough in instrumentation (and others as well) shifts the balance of utility between building artificial intelligent systems and studying natural intelligence. As a result, we should expect a growing synergy in the future between studies of artificial and natural intelligence.

One intriguing open question regarding fMRI is whether it is possible to decode instantaneous cognitive states of human subjects based on their observed fMRI activity. If this were feasible, it would open the possibility of directly observing the sequence of hidden cognitive states a person passes through while performing cognitive tasks such as language comprehension, problem solving, etc.

We present initial results showing that it is indeed possible to distinguish among a variety of cognitive states of human subjects based on their observed fMRI data. In particular, we have developed machine learning algorithms that can be trained to discriminate among a variety of cognitive states based on the observed fMRI data of the subject at a particular time or time interval. These machine learning algorithms, including Bayesian classifiers, support vector machines, logistic regression, and other methods, use the training data to discover the spatial-temporal patterns of fMRI activity associated with different cognitive states. They can then classify new fMRI observations to distinguish among these states.

This talk will describe results in which our machine learning methods were able to successfully discriminate between states such as "the subject is reading a sentence" versus "the subject is viewing a picture"; "the sentence is ambiguous" versus "the sentence is unambiguous"; and "the word is a noun" versus "the word is a verb." These classifiers are typically trained separately for each human subject, but in one case we were able to train a classifier that applies to new human subjects outside the training set. We will describe these results, the machine learning methods used to achieve them, and a number of directions for future research.