A Computational Cognitive Model of Mirroring Processes: A Position Statement

Mor Vered and Gal A. Kaminka
Bar Ilan University, Ramat Gan, Israel

Abstract

In order to fully utilize robots for our benefit and design better agents that can collaborate smoothly and naturally with humans we need to understand how humans think. My goal is to understand the mirroring process and use that knowledge to build a computational cognitive model to enable a robot/agent to infer intentions and therefore collaborate more naturally in a human environment.

Robotics as a field has evolved tremendously over the past few years and yet there still remains a gap in the ability to utilize robots for our benefit. For the most part robots do not behave correctly in social settings (involving either humans or other robots), because they generally do not possess the computational mechanisms required for social intelligence.

In this particular field where robots fail, humans succeed. I posit that when building socially-capable robots one should rely on general social intelligence building blocks, built into the brains of robots, rather than grafted on per mission. These building blocks would then cause the robot to behave socially, regardless of its specific task. I propose to investigate human intention recognition, the ability to infer the unobservable intentions, states and plans of other humans from their observable behavior.

In investigating the intention recognition building block, I intend to draw from what is known within the fields of neuroscience, psychology and cognitive science. Humans have the ability to quickly infer others intentions and therefore seamlessly adapt to new social situations. This ability has been hypothesized to come from the newly discovered mirror neuron system for matching the observation and execution of actions within the adult human brain. The mirror neuron system gives humans the ability to infer the intentions/mental states leading to an observed action using their own internal mechanism.

Mirror neurons have first been discovered to exist in macaque monkeys in the early 90’s (Pellegrino et al. 1992). These neurons for manipulation were seen to fire both when the monkey manipulated an object in a certain way and also when it saw another animal manipulate an object in a similar fashion. Recent neuroimaging data indicate that the adult human brain is also endowed with a mirror neuron system where it is attributed to high level cognitive functions such as imitation, action understanding, intention attribution and language evolution. Recently many researchers have linked the operation of a mirror system in the human brain to the motor system since most of the mirror neuron systems have been found in motor areas(Wolpert, Ghahramani, and Flanagan 2001).

Following this principle comes the hypothesis that the mirror system implements an internal model for control, much the same as the motor system. It is based on this internal model hypothesis that we will try to construct an efficient mirroring model.

In order to do so we will explore two specific, very different types of internal models(Oztop, Kawato, and Arbib 2006):

• An inverse model transforms a desired sensory consequence, a state, into the motor commands that would achieve that certain state (Wolpert and Kawato 1998). When dealing with mirror systems it is supposedly used to map the observed state into a rough motor representation which is now available for imitation, answering the question ”which action brought me to this state?”.

• A forward model is a relatively recent discovery in the field of motor learning and has been linked to mirror systems only recently (Carr et al. 2003). A forward model represents the opposite direction from an inverse model. It indicates the causal direction, mapping motor command into their predicted sensory consequences, states, answering the question ”If I perform this action, in which state will I be?”. When dealing with mirror systems it is supposedly used in order to anticipate the outcome of possible actions and therefore be able to choose the action that best leads me to the desired goal state.

There have been several neuroimaging studies that showed that areas associated with motor actions have also been shown to be active during imitation and observation, actions attributed to the mirroring system(Grezes et al. 2001)(Fadiga et al. 1995). This gave way to a recent proposal that, like the motor system, the mirror system consists not only of an inverse model but a forward model as well and that the two systems may not be fundamentally different(Wolpert, Ghahramani, and Flanagan 2001). In our model we intend
to explore this proposal by building both a forward model implementation and an inverse model implementation.

The difference between most plan recognition models and the mirror neuron system is in the format of the plan library. While in plan recognition models the plan is built for recognition and the usage of the plan library is meant only in order to match observations, in a mirroring model all knowledge of plans is supplied in a form suitable for both execution and recognition. In order to recognize plans, the agent uses its own internal execution library. This latter format enables us to execute learning easily and deal with problems arising from the incompleteness of a given plan library.

I posit that an agent that uses the same cognitive process as a human, i.e. mirroring, will be better equipped to infer intentions and consequently exhibit better behavior within socially challenging settings. There are two immediate advantages to mirroring as a component to AI plan recognition. The first is the reduction in storage space derived from the usage of the same execution library both for execution and recognition. The second substantial advantage is the ability to implement learning/imitation, derived from the fact that there is now a direct link between executable plans and recognized plans. This is due to the fact that recognized plans are now also represented in their executable form. Arising from this is the ability to execute anything I can recognize and recognize anything I can execute.

Therefore, as part of the quest of gaining the ability to utilize robots for our benefit my high level goal is to understand how humans think, as an aid in designing better agents that can collaborate smoothly and naturally with humans. My specific goal, is to understand the mirroring process and use that knowledge to build a computational cognitive model to enable a robot/agent to infer intentions and therefore collaborate naturally with humans and robots.

References


