Modular Behavior Control for a Cognitive Robot

Palis Ratanaswasd, Will Dodd, Member, IEEE, Kazuhiko Kawamura, Ph.D., Lifetime Fellow, IEEE, and David C. Noelle, Ph.D., Member, IEEE

Abstract—We propose a method for a temporal difference learning-based cognitive robot behavior control in which complex movements from a large library are combined so as to reduce error. This method uses elements of our existing cognitive robot architecture with the addition of a Working Memory System to create a modular control system. This controller is used to drive the motion of the humanoid robot ISAC.

Index Terms—humanoid robots, cognitive robots, cognitive control, motor control, modular control, working memory

I. INTRODUCTION

In the future, robots will be required to exhibit robust performance in a wide range of situations. To move towards this goal, the Center for Intelligent Systems at Vanderbilt University has designed a cognitive robotic control system for the humanoid robot ISAC [1]. A cognitive robot is designed to be equipped with cognition on several levels, from being fluent in routine operations to being capable of adjusting behaviors in the face of unexpected situations. The required flexibility needed to switch between the different levels of cognition arises from self-knowledge and processes that can analyze and modify the mechanisms that support both reactive action and careful deliberation [2]. In order to maintain adaptability in such a complex system, robust control of action must be coupled with the ability to select and focus on important information throughout the task execution. These abilities are known to exist in human as executive functions, and are studied by cognitive psychologists under the rubric of “cognitive control.” Cognitive control observed in humans is thought to be useful for a cognitive robot during action selection process as it guides the robot through the search for component behaviors that might be combined and used to efficiently execute routine tasks as well as to appropriately respond in novel situations.

Biological evidence of cognitive control in humans can be found in the function of the basal ganglia thalamocortical system, the prefrontal cortex (PFC), and the anterior cingulated cortex (ACC). Basal ganglia are involved in the planning and execution of complex motor and cognitive acts [3], while the PFC is involved in guiding these actions by supporting representations of relevant information from interference due to competing information [4]. The ACC is involved in detecting and helping to resolve response conflicts during a task performance [5]. Recent cognitive psychology research supports the idea that working memory plays a very important role in cognitive control and executive function [6]. Being strongly associated with the PFC, working memory can be viewed as a relatively small cache of task relevant information, grouped into “chunks”, that is strategically positioned to efficiently influence behavior [7]. One classical model of working memory is that suggested by Baddely and Hitch in which the control of executive processes is done by a component called the Central Executive. The Central Executive controls two working memory systems, namely the phonological loop and the visuo-spatial sketch pad. These two systems are responsible for both the storage and processing of linguistic and visio-spatial information, respectively [8].

Cognitive control in ISAC’s cognitive robot architecture is enabled by a Central Executive (CE) component that interfaces with a working memory system (WMS) that allows task related information to be actively maintained and analyzed in order to successfully select an action for an assigned task. The ability to select suitable information to maintain is provided by a Temporal Difference (TD) Learning [9] system which is similar in function to the dopamine cells that project to the PFC [10]. Also an important part of a cognitive robot, self reflection allows the robot to reason on past observations to monitor its current performance. It also allows the cognitive robot to derive better performing plans from learned experiences, and to handle situations that appear to have no solution based on the normal behavior of the system [11].

In this paper, we describe the method used by the CE and the WMS for the action execution process, including the

This research has been supported through DARPA/NASA-JSC (Grant #NAG9-1446) and NSF (Grant #EIA0325641). Research was conducted in the Center for Intelligent Systems at Vanderbilt University and at NASA’s Johnson Space Center in Houston.

Palis Ratanaswasd and Will Dodd are graduate students in Electrical Engineering at Vanderbilt University, Nashville, TN 37235 USA (e-mails: palis.ratanaswasd, will.dodd@vanderbilt.edu).

Kazuhiko Kawamura, Ph.D., is a professor of Electrical Engineering and Computer Science Department, and the director of Center of Intelligent Systems at Vanderbilt University, Nashville, TN 37235 USA (e-mail: kaz.kawamura@vanderbilt.edu).

David C. Noelle, Ph.D., is an assistant professor of Electrical Engineering and Computer Science Department at Vanderbilt University, Nashville, TN 37235 USA (e-mail: david.noelle@vanderbilt.edu).
learning of appropriate behavior combinations that allow the robot to adapt its skills in a manner similar to that used by cognitive control processes in humans. The system learns to select behaviors from a large behavior library based on success during past performance and to combine appropriate behaviors in an error-driven manner. This reduces the computational complexity of movement, and allows the system to learn which behaviors are best to combine from experience, without presenting the robot with every possible combination. Without past experience, the system must analyze the task context and slowly deliberate and plan in order to choose behaviors that best fit the task. Once a set of behaviors is chosen successfully, the system will start to select behaviors based on past experience, which will both speed up the action selection process and increase success rate.

II. MULTI-AGENT BASED COGNITIVE ROBOT ARCHITECTURE

ISAC’s cognitive robot architecture is a multi-agent system that incorporates several components inspired by different cognitive processes in humans. Information processing in ISAC is embedded within a multiagent-based software architecture called the Intelligent Machine Architecture (IMA) [1] [12]. ISAC’s cognitive abilities are implemented as collections of IMA agents and memory structures, including two important compound agents: the Human Agent and the Self Agent.

The Human Agent (HA) models the internal state of the humans in ISAC’s environment and serves as a user interface by providing social interaction with the humans. The Self Agent (SA) represents the robot’s “self” by monitoring the robot’s own internal state as well as the progress of task execution. The internal representation of the robot’s self-model is continually updated and enhanced. This allows the system to reason and act based on its status and the context of assigned tasks [13]. This coincides with the fact that the recognition and realization of the existence of one’s self is an important factor of how animals, especially humans, reflect on how they believe and act [14]. The SA also responds to commands given by humans through the HA and is responsible for controlling task execution. Similar to human cognitive abilities, the SA’s task execution takes into consideration the internal state of the robot as well as past experience, emotion, and feelings, all of which are crucial elements of a human’s cognitive behavior [15]. Emotional salience serves as both a reward signal and a filter for incoming information from the enormously complex world in which we live [16]. Through emotions and feelings, as well as cognitive direction, the conscious mind concentrates on the features that have proved to be important in the past.

Memory structures are utilized to help maintain the information necessary for immediate tasks and to store experiences that can be used during decision making processes. The short-term memory (STM) stores sensory information on a structure called the Sensory Ego-Sphere (SES) [17]. The stored information decays over time. The long-term memory (LTM) stores information such as learned skills and semantic knowledge for retrieval in the future. The working memory system (WMS) is modeled after the working memory in humans, which holds a limited number of “chunks” of information needed to perform a task, such as a phone number during a phone dialing task. It allows the robot to focus attention on the most relevant features of the current task and is closely tied to the learning and execution of tasks [6]. In our implementation, the WMS consists of two functional parts: a set of working memory slots that can be accessed by other agents, and a device to select chunks of memory to place in those slots based on the current task context. The WMS works with a component within the Self Agent called the Central Executive (CE) to provide cognitive control similar to that seen in humans. This working memory model is based on models from computational neuroscience as discussed in [6]. In the future, we will incorporate the Emotion System to further help provide feedback concerning the behavior of the system. Figure 1 shows the current implementation of ISAC’s multi-agent cognitive robot architecture.

III. BEHAVIOR CONTROL AND GENERATION USING MODULAR CONTROL

When a problem is given, most of the time it is easier to solve it by decomposing it into smaller ones, allowing each to be solved individually. The method of solving problems in such a modular fashion is common among various fields, including control and robotics. Wolpert and Kawato use modularity in designing adaptive low level controllers, allowing a small set of controller modules to be chosen independently to control the system based on the current situation [18]. Likewise, behavior based robotics benefits from the fact that behaviors can be developed individually and each one yields a different action. Simple behaviors can be grouped to work together, creating a new behavior that can yield more complex deliberation and action [19].
attempt to construct a high level action decision process for a cognitive robot, we have developed a modular control system that uses knowledge of its individual behaviors to predict the outcome. It analyzes its knowledge and chooses the most suitable set of behaviors to generate a complex action. The advantages of implementing modular control are:

--Only a small set of system parameters are required for each behavior. Therefore, changes in system configuration affect only a small number of behaviors.

--Adjustment can be made to an individual behavior without affecting the rest.

--Combinations of behaviors allow novel actions to be generated based on the task context without programming them in advance.

--The system considers only a small set of appropriate behaviors based on a specific task, which saves system resources.

--The system learns to adapt behaviors it already knows to fit given task requirements.

--The system chooses the best behavior based on task context as well as past experience.

Within the framework of ISAC’s cognitive robot architecture, modular control is used to generate actions in response to an assigned task. This process takes place within the WMS, which contains behaviors that are selected based on past experience. That is, the CE selects a small number of behaviors that are believed to be the best for the current task, and loads them into the WMS based on past experience as measured within a Reinforcement Learning (RL) module. The CE then activates the modular controller with the selected modules in the WMS. The CE determines the amount each selected behavior will contribute to a particular action based on its relevancy to the task context, and the CE produces appropriate weights for the behaviors within the WMS. The combination of control signals is then sent to the robot for action execution. Figure 2 shows the functions of the CE during an action selection process, based on skill knowledge and past experience.

An advantage of a modular control approach as compared to other methods is that it allows for a gross reduction of the state-space while retaining the benefits of error-based behavioral control. It prevents the system from having to interpolate and execute a task using several hundred LTM units at the same time, and assists the system with learning to achieve better performance over time.

Figure 3 shows the interaction of components in the cognitive robot architecture with relation to modular control.

**Fig. 3. Modular control in cognitive robot architecture and related components**

### A. Behavior Selection

ISAC’s behaviors are currently derived during an initial behavior learning process, during which the robot is teleoperated through sequences of actions. The motion streams are recorded and parsed by Spatio-Temporal ISOMAP (ST-ISOMAP) [20], and stored in the Procedural Memory section (PM) in the Long Term Memory (LTM) [21]. At this time, all ISAC’s behaviors involve arm movements. In the future, we will include behaviors that control the movements of other types of actuators, including the head and hands.

Behaviors in the LTM are loaded into the WMS upon the reception of a task assignment. When the system has little experience in the task, it must rely on an analysis of task context based on goal information and the physical configuration of the robot. The execution result is monitored. The RL module in the CE decides which behaviors should be loaded into the WMS based on the performance of prior executions of the particular task. Performance evaluation is performed using the result of a task execution, either success or failure, and stored a measure of expected reward for reinforcement learning.

The number of behaviors loaded into the WMS at one time is currently limited to three, which is similar to the capacity of human working memory [22]. Restricting to a small number of behaviors helps to limit the size of the search space during action selection, allows only a small amount of system resources to be used, and results in fast response times.
**B. Prediction of Behavior Outcome: State Estimation**

Within the WMS, each behavior is assigned a state estimator. A state estimator uses the trajectory information associated with its behavior to estimate the execution output of the next sampling time step in the control loop. The estimation is found using an interpolation technique called Verbs/Adverbs – a technique originally used to produce believable motion for computer animation [23]. The Verbs/Adverbs technique uses radial basis functions to interpolate for a set of trajectories that brings the robot closest to the goal state. The estimation is stored in the form of state variables that are sent to the CE for “relevancy” computation.

**C. Task Relevancies**

In order to select the most appropriate behaviors to perform a task, the CE calculates the degree of “Relevancy” of a behavior to the given task and compares these quantities across behaviors. The behaviors that have high relevancy are biased by the behavior selector for current action execution.

The relevancy value used in the current modular control system determines how a behavior can bring the state of the system toward the targeted goal state while assuring that the system state at next time step is achievable. To ensure the achievability of the next state, the CE takes the state estimations and subtracts from the actual current state. The resulting difference, called the estimation error, indicates the amount of movement the robot has to make in order to reach the estimated configuration in the next time step. Smaller estimation error is preferred because the movement is likely to be attainable. In addition, to determine the relevancy to goal information the CE determines the distance between the goal state and the state estimation. The resulting error, called the goal state error, indicates how close to the goal the behavior will take the system. The calculation of both the estimation error and goal state error are done individually for each behavior.

--Let the \( i^{th} \) behavior that is loaded into the WMS be \( B_i \). Computed at current time \( t \), the state estimation of \( B_i \) of next time step \( t+1 \) is \( \hat{x}_{t+1}^i \). The delayed estimation \( \hat{x}_{t}^i \), is evaluated by the CE to find two types of errors:

- The current state error, \( e_{c_i}(t) = x_i - \hat{x}_{t}^i \)
- The goal state error, \( e_{g_i}(t) = x^* - \hat{x}_{t}^i \)

where
- \( x_i \) is the actual state of the robot at time \( t \), and
- \( x^* \) is the goal state

--Task relevancy, \( \lambda_i \), for behavior \( B_i \) at time \( t \) is then

\[
\lambda_i = \alpha_c e^{-\|e_{c_i}(t)\|^2/\sigma_c^2} + \alpha_g e^{-\|e_{g_i}(t)\|^2/\sigma_g^2}.
\]

The scaling factors, \( \sigma_c \) and \( \sigma_g \), are chosen such that values of the two terms are not too different, and resulting exponentials are not too small to cause computation underflow. The relevancy weights, \( \alpha_c \) and \( \alpha_g \), are assigned according to the type of task being executed.

**D. Behavior Selector**

From the relevancies vector \( \lambda(t) \), the behavior selector within the CE calculates a weights vector \( w(t) \) that will be used to control the amount that each behavior will be involved during an action. That is, a weight \( w_i(t) \) is calculated from a thresholded task relevancy multiplied by a normalization scale factor as shown in Figure 4.

![Weight calculation of the Behavior Selector](image)

In Figure 4, \( \int \) represents a threshold function

\[
Thr(x) = \begin{cases} 
    x, & x \geq T \\
    0, & x < T 
\end{cases}
\]

The threshold, \( T \), is a constant that is tuned empirically, based on experimental results.

During action execution, each selected behavior module in the working memory independently produces a control signal that controls the robot along its trajectory. Magnitudes of all the control signals are weighted and summed before finally being sent to the robot controller. As control signals are combined with time varying weights, the resulting action does not resemble only one behavior but adjusts itself based on the task context. The combination of the same set of behaviors will produce a different action when a different task is given.

**E. Reinforcement Learning**

The role of the RL module is to select the behaviors that are appropriate for the current task. A reward is calculated by the CE after attempting the task with the given set of behaviors. This reward is then fed into the RL module so it can learn which behaviors to select the next time. The reward is equal to the average relevancies of each behavior being loaded during one action execution, discounted by the task completion time. Average relevancies, in this case, indicate how much a set of selected behaviors is suitable for that particular task based on goal information and prediction error. Thus, the most suitable set of behaviors that completes the task most quickly is likely to be selected in the future. ISAC learns which set of behaviors to select for each task. New, similar tasks can be built on the learned values.

Currently this is implemented by a simple look-up table that attempts to estimate the utility of each behavior and is used to select the top three behaviors for the task. This selection system is also configured to conduct a random exploration for a small percentage of task attempts in order to prevent initially explored behavior from “hogging” activation, and to guarantee that the system will converge on a usable,
reasonably explored state. In the future, we will integrate a more powerful “Working-Memory” temporal difference learning toolkit into the system, which will judge the behaviors as groups through the use of neural networks. This toolkit will also allow for the system to generalize over different contexts and tasks and to learn to group similar behaviors. [7]

IV. EXPERIMENT

We have conducted an initial set of experiments for studying robot behavior learning based on a working memory model and modular control. The first of these experiments is intended to test the following abilities in the context of generating behaviors for a single task:

--The ability to choose appropriate behaviors out of a number of behaviors to perform a task based on past experience
--The ability to switch the priority of behaviors during the course of an action based on the current state of the robot
--The flexibility to combine behaviors to achieve a new behavior when the stored behaviors are not adequate to perform the task

In the future, the capability of the working memory to filter through large amounts of information and determine what is important for the task will be tested. As ISAC’s range of movement is very limited, only a few behaviors can be taught and performed. However, we expect modular control to be much more effective in robots with higher ranges of movement such as NASA’s Robonaut. This approach is also not limited to the humanoid robot platform but would work on a mobile platform as well.

A. Experimental Design

ISAC is endowed with a set of initial knowledge. This consists of behaviors learned in the past, such as hand waving, reaching, and arm sweeping. These behaviors are currently explicitly taught, though in the future they will be automatically learned by recording common movements as they are generated by the modular controller. This set of behaviors is taught to ISAC prior to the task assignment.

For the task, ISAC is asked to reach to a point on the table from its resting configuration. This point is chosen so that any one of the stored behaviors alone cannot be interpolated to perform this reaching successfully. Figure 5 illustrates the point that ISAC has to reach, as well as interpolated motions of stored behaviors.

If ISAC cannot reach the goal position within a certain amount of time, it reports that the task is failed; otherwise it will report it as a success. ISAC will be asked to repeat the task for a number of trials to give ISAC some opportunity to experiment with various movements and weed out the ones that should be used. As the number of trials increases, the rate of failure per number of trials should decrease.

B. Results

When ISAC is presented with the “pointing” task, the CE looks for behaviors in the LTM with high expected rewards for a task with the keyword “point.” The experiment shows that the response time of this process is quite fast when ISAC already has some experience with the task. However, if such experiences do not exist, other information must be used to reason about behavior selection, which results in a slower response. At this time, information used includes the starting physical configuration of the robot in the behavior trajectory and average distance between the ending position of behavior exemplars and the goal position. In the future, the TD-Learning module will be able to generalize to a certain extent about which behaviors might be appropriate in novel situations.

In this experiment, behavior selection was not very efficient at the beginning, as ISAC had no experience in the task. After a few trials on the same task, ISAC started to pick similar sets of behaviors that produced successful results. One must note that the benefit of faster response time was not obvious in this experiment because only a small set of behaviors was presented. However, the efficiency of the behavior selection process could easily be observed as the success rate became higher as the number of trials increased. Figure 6 shows the number of trials of each behavior selected by ISAC and corresponding success rates.

By observing the trajectory data of the resulting motion, we can see that motions are combined together in the form of a weighted summation of control signals. Since each behavior cannot move toward the goal individually, modular control combines them to create a new motion that has a greater likelihood of reaching goal. Weights are distributed to each
behavior differently during the course of execution, showing
dynamic adaptability of modular control to the task context.
Figure 7 illustrates the weight distribution of behaviors during
an action.

This experiment is only a preliminary test of the system, but
it does show some positive points of using modular control in
cognitive robots. Some features that we are expecting to see
in future experiments include the ability to focus on task
context and to maintain only relevant behaviors in the
presence of a secondary task.

V. CONCLUSION

We are developing a cognitive robot control system that
learns to use its skill knowledge, stored in the form of
behaviors within a large library, to generate actions suitable
for assigned tasks. The robot should learn, by itself, to choose
the most appropriate set of behaviors for a particular task over
time. In this paper, we introduced a method called modular
control, which utilizes a biologically inspired working
memory system to hold task related information during task
execution, and uses a central control mechanism called the
Central Executive to select a small set of skills believed to be
most suitable for task execution. The CE loads skills into the
WMS and executes them. Execution results are used to
update selection criteria in the future through reinforcement
learning.

By using modular control along with a working memory
system, we find that the robot is capable of choosing the right
set of behaviors after a small number of trials. The behaviors
are selected correctly once they are learned, allowing rapid
response times in later executions. Motions from executions
reveal that the combination of selected behaviors makes use of
time varying weights.

The reported research is currently limited to handle a single
task of arm control. Future research will include task
switching performance and the controlling of different types
of actuators, including the head and hands. Other types of
bias will be added to widen the dimension of learned
experience, such as an emotion system, so more complex
actions can be learned with more quickly.

REFERENCES

Spratley, and K. A. Hambuchen, “Multitagent-based cognitive robot
architecture and its realization,” Int’l. Jo. of Humanoid Robotics, vol. 1,
Workshop on Robot and Human Interactive Communication (RO-MAN),
D. Cohen, “Anterior cingulate cortex, error detection and the online
computational model of working memory”, Models of working memory:
Mechanisms of active maintenance and executive control, (A. Miyake
biologically inspired adaptive working memory for robots,” AAAI Fall
Symp. Workshop on the Intersection of Cognitive Science and Robotics:
dopamine in regulating prefrontal function and working memory,” In S.
Monsell & J. Driver, eds., Control of Cognitive Processes: Attention and
multi-agent approach to self-reflection for cognitive robots.” 11th Int.
Conf. on Advanced Robotics (ICAR), Coimbra, Portugal, Jun. 30 – Jul.
for integrated service robot development,” Proc. of IEEE Systems, Man
research,” Performance Metrics for Intelligent Systems Workshop
[14] Hauser, M. D., Wild Minds: What Animals Really Think, New York:


