Implementing Reinforcement Learning in the Chaotic KIV Model using Mobile Robot AIBO

Robert Kozma
Computational NeuroDynamics Lab
Division of Computer Science, University of Memphis
373 Dunn Hall, Memphis, TN 38152, USA
rkozma@memphis.edu

Sangeeta Muthu
Computational NeuroDynamics Lab
Division of Computer Science, University of Memphis
373 Dunn Hall, Memphis, TN 38152, USA
smuthu@memphis.edu

Abstract—We use the biologically inspired dynamic neural network architecture KIV to achieve robust goal-oriented navigation in a physical environment with obstacles. KIV operates on the principle of chaotic neurodynamics, in the style of brains. It performs the task of multi-sensory fusion, recognition, and decision-making in real time. We use the Sony AIBO robot to demonstrate the operation of our algorithm. AIBO’s video camera and infra sensors have been complemented with an external camera for monitoring of the robot’s position. The performance of the autonomous system is evaluated using goal-oriented navigation.

Keywords- Chaos, navigation, KIII and KIV models, AIBO.

I. INTRODUCTION

Intensive research is conducted worldwide towards intelligent techniques for robot navigation. Our approach is an alternative to rule-based systems and it can complement them in the area of sensory information processing and multi-sensory fusion. Various approaches have been used successfully to accomplish robust navigation, like BISMARC [5], ethology inspired hierarchical organizations of behavior [11], behavior-based control algorithm using fuzzy logic [4], subsumption methods [2], and local path planning based on stereo vision [12] to achieve efficient navigation.

In the KIV model described here, intelligence is approached at the level of the salamander brain [6, 8]. KIV is a revolutionary chaotic dynamical memory, which encodes sensory information in the form of aperiodic spatio-temporal oscillations of nonlinear processing elements. An essential part of the KIV model is that it learns from experience. The agent learns continuously as it explores the environment. An example of such learning is observed in babies during their initial development following birth. Initially, the baby has only very basic skills. As she grows up, she learns behaviors and acquires knowledge about the world. Similar approach is used in this model. The agent is unaware of the environment at first. During an exploration phase, however, it keeps on learning until it becomes familiar with it. Then it uses this knowledge for orientation and action selection. At a later stage, it can generalize its knowledge in new environment.

The focus of this paper is to implement a biologically motivated sensory information processing system, the KIV model, which can guide an autonomous agent in finding its way to a goal location. The autonomous agent used herewith is the Sony AIBO robot ERS-220A. In order to use AIBO to our aims, we apply the OPEN-R program provided by SONY. OPEN-R allows us to access sensory signals produced by the robot and implement our own control algorithm, instead of the ones provided by the manufacturer. We call our robot EMMA (Evolving Multi-modular Agent). EMMA has several sensors like a CMOS camera sensor, infrared (IR) distance sensors, acceleration and touch sensors, audio sensors [10]. We augment these with an additional sensor, which is a video camera located above the experimental area. It is used for providing global navigation information for AIBO. It is like the Earth’s magnetic field used by the birds for their long seasonal migration.

This paper first defines the KIV model and the applied navigation algorithm. Then it describes the environment, as well as the results obtained by using the KIV Model in that environment. After this, a performance analysis is conducted to see how well the model performs under various conditions. Finally, we make several conclusions and outline the direction of future work.

II. OVERVIEW OF THE K-BASED CONTROL APPROACH

K sets represent a family of models of increasing complexity. Each level of complexity represents various aspects of operation of vertebrate brains [8, 12]. These models are biologically inspired, and they are built based on the salamander’s central nervous system. They provide a biologically conceivable simulation of chaotic spatio-temporal neural developments at the mesoscopic and macroscopic scale. Second order ordinary differential equations (ODEs) are used to model these K Sets.

The K sets consists of an organized hierarchy of the K0, K1, KII, KIII and the KIV system, in that order, KIV being the highest in the hierarchy. The smallest set, K0, is the basic building block of the K model. It models either an excitatory neuron population or an inhibitory neuron population. It has a weighted input and an asymmetric...
nonlinear sigmoid function. The KI set is built by combining a population of either excitatory K0 sets or inhibitory K0 sets. KII set is formed from KI sets by connecting both the excitatory KI units and the inhibitory KI units by positive and negative feedback connections. The KIII Model has a layered structure with several interconnected KII sets.

A KIV model is built from a selection of various KI, KII, and KIII sets. A KIV set is formed by the interaction of 3 KIII sets. It is used to model the interactions of the primordial vertebrate forebrain in the genesis of simple forms of intentional behavior [3, 9]. Interacting aperiodic/chaotic oscillators in the KIV model have the potential of producing macroscopic phase transitions, which provide the mechanism for fast and robust information processing in KIV, in the style of brains. The KIV model has been designed to simulate animal behavior at the level of the amphibian or reptile: simple search, seize or destroy recognizable objects in a relatively limited environment. The reptilian primary sensory paleocortices and components we now model with the KIV set. Paleocortex is the primitive 3-layered precursor of 6-layered neocortex. Details of the differential equations used and the architecture of the complete K sets are given in [1,6,7].

The complete KIV model consists of four major components, out of which three are KIII sets. Namely, one models the hippocampus, one the cortical region and the third describes the midline forebrain. The fourth major component is the amygdala, which is a KII set. Amygdala integrates influences from all parts of the hemisphere, and it provides link to the brain stem and further external parts of the limbic system. Figure 1 shows outline of the KIV model.

Figure 1: KIV model with 3 types of inputs. Exteroceptors (vision, audio, somatosensory) through cortex; Orientation (landmarks, etc) through hippocampus, Interoceptors (emotions, motivation) through the midline forebrain. The Amygdala integrates emotional and sensory information and link to the brain stem and further parts of the motor system [6].

Abbreviations in Fig. 1 are: DG, dentate gyrus; CA1-CA3, Cornu Ammonis (hippocampal sections); PG, periglomerular; OB, olfactory bulb; AON, anterior olfactory nucleus; PC, prepyriform cortex, as given in Parts of the midline forebrain are: diagonal band DB, basal ganglia BG, and hypothalamus HT.

The Hippocampal KIII (HP) deals with the spatial orientation ("Where?" information). The cortical KIII (CR) is responsible for sensory inputs like touch, vision, audition, etc ("What?" information). The Midline forebrain deals with the internal motivation and emotions of the agent ("Why?" information). The amygdala combines these outputs to make decision about the next step to be taken by the agent. Detailed explanation of the complete KIV model is given in [6].

In a simplified KIV model, we have the HP, the CR units, and the amygdala, whereas the midline forebrain is omitted. It is substituted in a simplified form with a reinforcement signal, as it is described in the next section. Fig.2 illustrates the simplified KIV model. The interacting KIII units maintain a high level of autonomy, and have only a small mutually dependent portion in their operation. This small mutually dependent portion, however, is crucial in making the final decision and action.

Figure 2: Schematic view of the simplified KIV model with two major components: the hippocampal and the cortical KIII units. Rectangular boxes model various brain parts [13].

The KIV set is used to process the sensory data and to generate meaningful actions for the robot. There are three steps to this task:

1. Capturing the necessary sensory data;
2. Processing the data using the KIV model and;
3. Making a meaningful decision concerning actions performed by the autonomous agent.

The above steps follow each other without interruption. In fact, once the KIV-based sensory-motor process has been started in a real robot, it just goes on continuously as the life of a real animal after birth. KIV is a dynamical system, in which previous experience intimately interlinked with present state and future action.
III. REAL ENVIRONMENT FOR TESTING AUTONOMOUS AGENT EMMA

A. Sensors and Sensory Processing

The environment used in this paper is of rectangular shape of dimensions approx. 10ft x 14ft, having a single obstacle as shown in Fig. 3. The figure shows five colored balls. Four of the colored balls are landmarks, namely, green, blue, purple and yellow. The red ball is the goal location. These landmarks are used by the hippocampus to decide where the agent is located at any instant of time. In the present experiments, the agent uses only two of its sensors from the available sensors. These are (1) the camera to capture goal and landmark information; and (2) the infrared (IR) distance sensors, to capture distance from any obstacle in its path. The camera can handle long-range information across the entire experimental area. The IR sensors are, on the other hand, are localized: EMMA can sense an obstacle from a distance of max. 0.9m.

B. Data acquisition and learning by EMMA

There are three phases in the working of the KIV model:
1. Learning phase;
2. Labeling phase; and
3. Testing phase.

In the first phase, the agent explores the environment collecting sensory data. The data that it collects in this experiment is the color intensities of each of the five color balls and the distance of any obstacle from the agent in any of the five forward-looking directions; e.g., E, NE, N, NW and W. The HP processes the color intensities while the CR processes the IR distance values. The five color values and the five distance values are not the only data on which the two KIII units learn. There is a memory for previously taken steps in both HC and CR. These previous steps can range from one memory step to a maximum of five memory steps. The choice of the number of memory step is a crucial factor in deciding the performance of the KIV model, as it will be shown in the next section.

In the second phase, learning takes place only if certain conditions are met. For example, the hippocampal KIII learns if the goal is spotted and that the goal color intensity of the current sensory data is greater than what it was in the previous step. The learning used is reinforcement learning. It is implemented by using Hebbian Correlation rule in CA1 and PC layers of the two KIII systems. Let $X_i$ and $X_j$ denote the activations of the i-th and j-th neurons, respectively; the weight between nodes i and j is represented as $W_{ij}$. Finally, LR represents the learning rate. Using these notations, the weight update is given as:

$$W_{ij}^{\text{(new)}} = W_{ij}^{\text{(old)}} + LR \cdot (X_i X_j) \quad (1)$$

Eq.1 is the basic form of the Hebb learning rule. We use a modification of this rule where the weight update is given as follows:

$$W_{ij}^{\text{(new)}} = W_{ij}^{\text{(old)}} + LR \cdot (X_i - \langle X \rangle) \cdot (X_j - \langle X \rangle) \quad (2)$$

Here $\langle X \rangle$ is the mean activation given by:

$$\langle X \rangle = \frac{\sum_{i=1}^{N} X_i}{N} \quad (3)$$

Positive reinforcement is triggered in hippocampus, if the agent approaches the goal. In the Cortical KIII, a negative reinforcement is triggered whenever the agent approaches an obstacle. The cortical reinforcement is based on the IR sensory readings. The hippocampal reinforcement at first has been based on the evaluation of the camera view of EMMA [10]. It is clear that positive reinforcement requires that EMMA facing the goal otherwise she can’t see it by her internal camera. The external camera described in this paper solved this problem, as it will be described in the next section. The learning phase continues until the agent explores most part of the environment. Examples of the oscillating activation signals collected from the CA1 of the hippocampal and PC layer of the cortex is shown in Fig. 4 and Fig.5, respectively.

At the labeling phase, the activations from the hippocampal and cortical KIIIs are collected as reference patterns. These patterns can be activations representing motion in any one of the directions: forward, back, right or left. These activations are then used for finding the right direction of movement for the agent in testing phase.
During testing, the agent is placed at a randomly selected position location in the environment and is freely left to move on its own. At each step the agent captures the sensory data and passes on to the KIV model. The activations generated from this test input are matched to the stored reference activations. The best match determines the next move selected by the agent. We use k nearest neighbor k-NN method with Euclidean metric to determine the best match with the actual input. steps, which corresponds to 250 ms (the sampling time is 0.5 ms). This sampling time value has been chosen based on previous computational studies [6-7]. The occasional spikes at regular intervals indicate the time segments in Figs. 4 and 5 indicate this theta cycle, when input signals entered the KIII units.

IV. OPTIMIZATION OF PARAMETERS OF THE INDEPENDENT HIPPOCAMPAL & CORTICAL KIII UNITS

There are a couple of factors that influence the performance of the model, i.e. how quickly can EMMA reach the goal. One is the learning rate denoted by (LR); another parameter is the number of past steps EMMA remembers, denoted by (α). Alpha > 1 indicates the presence of a short-term (stack) memory, while Hebbian learning in CA1 and PC represents a long-term associative memory. By varying these factors, the performance of the model can be greatly altered. A typical set of parameters with very good navigation performance are:

- hippocampal LR = 0.8, cortical LR = -0.4;
- hippocampal \( \alpha = 3 \) and cortical \( \alpha = 2 \).

The learning rate and the number of steps chosen as memory, for each of the KIII systems. Before combining the two KIII systems in to a KIV set, it is necessary to find the most favorable parameters at which the two systems will function their best, i.e., give optimal output, so that we can be sure that the KIV model is tuned with the finest parameters. To find out this optimal parameter set, extensive experimentation has been performed for each of the KIII system, with various parameter combinations of the leaning rate (LR) and memory step (α).

In order to demonstrate the effects of these parameters on hippocampal KIII’s performance, we need to isolate the hippocampal system from the KIV model and test its performance. This has been achieved by setting the learning rate of the cortical KIII to zero. In addition, the connection matrix that links the cortical KIII to the hippocampal KIII in the KIV model, is also set to zero. This is to ensure that there is no interaction between the cortex and the hippocampal system. This way the configuration of the Hippocampal system can be fine-tuned to achieve optimum performance. The optimization criterion aims at conditions to reach the goal in minimum number of steps.

The range of LR is between 0 and 1 [10]. We run experiments for different incremental values of hippocampal LR, starting from 0.2. Also, we need to find out how much short-term memory (α) is needed by the system. The short-term memory, in our case,
remembers the agent’s previous steps in addition to the current step. Figure 6 shows the effect of LR and $\alpha$ on the performance of Hippocampal KIII. It is seen that LR = 0.8 and alpha = 3 gives the optimum performance. Similarly we can derive optimum tuning for the cortical KIII as LR = -0.4 and alpha =2. Details of the optimization are given in [10].

V. RESULTS OF NAVIGATION WITH EMMA USING KIV MODEL

In this section, results of navigation using the integrated KIV system are introduced. In order to be able to provide a continuous and autonomously operated global navigation signal to the amygdala, we use the camera positioned above the experimental area. It is to be pointed out that the external camera is part of EMMA’s the autonomous sensory system. In order the aid the exploration of the environment, we placed several obstacles in the environment. We used wall following strategy during the training. It is important to outline, that we did not train EMMA wall following and during the testing the obstacles are to be removed or rearranged to make sure that we really observe a generalization behavior.

Figure 7 show and example of the learning trajectory traveled by EMMA in the wall-following model (left wall following). In a next training session, we have right wall following as well. As a results, we have generated about 8-10 examples of left turns and about the same number of right turns that are reinforced in KIV. Clearly, there are much more correct forward moves in such a training session. To some extend animals have bias toward forward movement, which has advantages. In order to prevent overly dominating forward bias during learning, EMMA has learnt forward movements, if is happened after a turn (left or write). Based on such strategy, in the given environment, we had about 20 forward examples to be learnt by EMMA.

![Figure 6: Plot of memory steps ($\alpha$) versus the average # of steps to goal for each LR; case of the hippocampal KIII system.](image)

![Figure 7A: The dots mark EMMA’s locations detected by the external camera using wall-following training mode in an environment with obstacles.](image)

![Figure 7B: Snapshot of the environment with obstacles.](image)

During testing, we have removed the obstacles. We have also covered the goal (red ball) so it can’t see the goal. This way we wanted to test if she can move toward the goal based on the relative location of the learned landmarks (the other colored balls). Figure 8 shows an example of the trajectory obtained by EMMA in the testing mode. Clearly, it is not a perfect trajectory, but clearly, EMMA moving toward the goal, step-by-step. She tests the result of her move and if she made a mistake, she turns and corrects it.

V. CONCLUSION

In this paper, we have introduced the simplified KIV model. We have discussed KIV can be applied to solve realistic problems such as real-time navigation with mobile robot EMMA. We have demonstrated that
with this methodology, EMMA can work autonomously to achieve the specified goal using KIV-based control. This is an essential proof-of-concept, which will be followed by detailed quantitative performance studies.

Following the initial success of the navigation in the experiments, we plan to run a detailed series of experiments with various environmental configurations and various locations for EMMA. It is clear that the method should be further improved and fine-tuned, concerning the parameters of the reinforcement learning and other memory effects. The implementation of the amygdala is very simplified in the present work. It will be elaborated in the future based on our studies in other application domains [13].

Since, real life parameters are involved, the degree of difficulty is very high. This is because the sensory data is very noisy in real environment. For example, shadow on objects in environment, extent of agents visibility, non-uniform light, etc. Our K-based algorithm is based on very general principles, not limited to the simple environment like the one described in this paper. It can show it’s competitive edge for complex environments like the Martian terrain [11]. Applications to that domain are in progress.

REFERENCES