Knowledge-Based Systems 57 (2014) 8–27

Contents lists available at ScienceDirect

Knowledge-Based Systems

journal homepage: www.elsevier.com/locate/knosys

Biologically inspired layered learning in humanoid robots

Hamed Shahbazi [a,*], Kamal Jamshidi [b], Amir Hasan Monadjemi [b], Hafez Eslami [b]

[a] Department of Mechatronic Engineering, University of Isfahan, Isfahan, Iran
[b] Department of Computer Engineering, University of Isfahan, Isfahan, Iran

ARTICLE INFO

Article history:
Received 25 December 2012
Received in revised form 17 November 2013
Accepted 2 December 2013
Available online 14 December 2013

Keywords:
Neural networks
Oscillatory neurons
Layered learning
Central pattern generator
Policy gradient algorithm
Humanoid robots

ABSTRACT

A hierarchical paradigm for bipedal walking which consists of 4 layers of learning is introduced in this paper. In the Central Pattern Generator layer some Learner-CPGs are trained which are made of coupled oscillatory neurons in order to generate basic walking trajectories. The dynamical model of each neuron in Learner-CPGs is discussed. Then we explain how we have connected these new neurons with each other and built up a new type of neural network called Learner-CPG neural networks. Training method of these neural networks is the most important contribution of this paper. The proposed two-stage learning algorithm consists of learning the basic frequency of the input trajectory to find a suitable initial point for the second stage. In the next stage a mathematical path to the best unknown parameters of the neural network is designed. Then these neural networks are trained with some basic trajectories enable them to generate new walking patterns based on a policy. A policy of walking is parameterized by some policy parameters controlling the central pattern generator variables. The policy learning can take place in a middle layer called MLR layer. High level commands are originated from a third layer called HLDU layer. In this layer the focus is on training curvilinear walking in NAO humanoid robot. This policy should optimize total payoff of a walking period which is defined as a combination of smoothness, precision and speed.

1. Introduction

The problem of robot locomotion is where neuroscience and robotics converge. This common part is the pattern generators in the spinal cord of vertebrate animals called "Central Pattern Generators" (CPGs). CPGs are neural circuits located in the end parts of the brain and first parts of the spinal cord of a large number of animals and are responsible for generating rhythmic and periodic patterns of locomotion in different parts of their bodies [1]. Although these pattern generators use very simple sensory inputs imported from the sensory systems, they can produce high dimensional and complex patterns for walking, swimming, jumping, turning and other types of locomotion [2]. The idea that human nervous system has a layered mechanism in generating complex locomotion patterns with only simple stimulations is a provocative one which is intended to be modeled in this paper.

Learning in humanoid robots deals with a large number of challenges. For example, the robot should overcome noisy and nondeterministic situations and reduce unwelcome perturbations [4]. The state space is continuous and multidimensional, thus it is impossible to search systematically in that space. The fact that there is no explicit mapping between intentions and actions in a humanoid robot is a big issue that should be solved [5].

In this paper we intend to train to perform a curvilinear walk in a NAO soccer player robot using a hierarchical layered learning paradigm. The proposed method uses a basic CPG based walk controller built of Learner-CPG Neural Networks (LCPGNNs). In this manner, any kind of complex behavior can be trained into a CPG neural network and it can be used in the movement of different types of robots.

In the next section related works in the field of humanoid robot locomotion and learning will be reviewed and the advantages and disadvantages in each method will be discussed. In this section we also introduce NAO platform which is used in this research. Section 3 is dedicated to the proposed model of layered learning in this work. It introduces each layer of our learning platform and explains different correlations between the layers. The CPG layer is explained in Section 4. The role of the arms and coupling of them with other joints is explained in this section. Another important concept is the feedback pathways which are discussed here. The mathematical discussion about the learning algorithm used for Learner-CPG neural networks is presented in Section 5. Here the two-stage learning algorithm which can train each oscillator

* Corresponding author. Tel.: +98 311 7935620.
E-mail addresses: shahbazi@eng.ui.ac.ir (H. Shahbazi), jamshidi@eng.ui.ac.ir (K. Jamshidi), monadjemi@eng.ui.ac.ir (A.H. Monadjemi), eslami@eng.ui.ac.ir (H. Eslami).

0950-7051/$ - see front matter © 2013 Elsevier B.V. All rights reserved.
http://dx.doi.org/10.1016/j.knosys.2013.12.003

neuron and its synaptic connection in a LCPGNN is explained. Section 6 introduces the MLR layer and its learning mechanism. We use reinforcement learning in this layer to find an optimal policy for the CPG layer. Policy parameterization and payoff function is discussed here as well. Section 7 includes the experimental results. Here some of the implementations and results in WebotsTM simulator and simulink of Matlab are presented. In Section 8 the highest layer HLDU, its functions and capabilities are briefly discussed. Section 9 includes the conclusion and future works.

2. Related works

There are many approaches to solve bipedal skill learning issues [6]. As an alternative to the methods using pre-recorded trajectories [7,8], ZMP-based approaches [11] or methods using heuristic control laws (e.g. Virtual Model Control (VMC) [12]), the CPG based methods are introduced, using some biological perspectives. They encode rhythmic trajectories as limit cycles of nonlinear dynamical systems. Coupled oscillator-based CPG implementations offer miscellaneous features such as the stability properties of the limit-cycle behavior (i.e. the ability to overcome perturbations and compensate their effects), the smooth online modulation of trajectories through changes in the parameters of the dynamical system, and entrainment phenomena when the CPG is coupled with a mechanical system. Examples of CPGs applied to biped locomotion are included in [13,14]. Matsubaraa et al. discussed a CPG-based method for biped walking combined with policy gradient learning [10].

A drawback of the CPG approach is that most of the time these CPGs have to be custom-made for a specific application. There are few techniques to construct a CPG for generating an arbitrary input trajectory. Righetti et al. represented a model for constructions of a generic model of CPG [13]. This method is a Programmable Central Pattern Generator (PCPG) by applying dynamical systems and some differential equations for developing a training algorithm. The learner model is based on the works of [15] a Hebian learning method in dynamical Hopfs oscillators. The programmable central pattern generator was used to generate walking patterns for a Hoap2 robot. This Hoap2 can increase its speed without falling to the ground. Using this type of generic CPGs they trained the PCPGs with sample trajectories of walking patterns of the Hoap-2 robot provided by Fujitsu. Each trajectory is a teaching signal to the corresponding CPG controlling associated joints.

Hackenberger initiated some proceedings [16] on programmable CPG model included in [13] in order to use a nonlinear feedback policy for balancing a humanoid robot during a walking gait. This system consists of two modules: A polar-based PCPG which reproduces a walking trajectory, and a reinforcement learning agent responsible for modifying the walking patterns. This paradigm can use programmable central pattern generators and enables them to incorporate gyro feedbacks into the system definitions generating the walking trajectories. Degallier et al. [17] defined a modular generator of movements called Unit Pattern Generators (UPGs) and combined them to build CPGs for some robots with great degrees of freedoms. He applied his framework to interactive drumming and infant crawling in iCub humanoid robot.

3. Layered learning architecture

The idea of layered learning in multi agent systems was introduced in [19] by Stone. He investigated the use of machine learning within a team of soccer player's 2D agents. Using hierarchical task decomposition, layered learning enables us to learn a specific task at each level of the hierarchy. Here a hierarchical learning framework for walk learning in soccer player humanoid robots is designed. Our model composed of 4 different layers. Designs of these layers are inspired from biological hierarchy of the nervous system in human [22]. These layers are called HLDU, MLR, CPG, LLJ layers. Fig. 1 illustrates this model. In this section we discuss the overall hierarchical model, specific function of each layer and relations between the layers.

- HLDU layer: High Level Decision Unit (HLDU) is a model of Cerebrum part of the brain cortex. Cerebrum controls learned behaviors and memory in human being and makes up about 80% of the brain mass [20]. In this region high level commands for different motor behavior, vision, hearing and speaking are generated. In our model this is the place of decision making which learns to analyze input images, process them and send commands/vision feedbacks to the next layer. The Image captured by the robot cameras which determine the local position of the robot with respect to the desired path and this position generates the immediate speed and precision feedbacks. In the current study, The special commands generated by this layer determine a curvilinear path on the ground.
- MLR Layer: Mesencephalic Locomotor Region (MLR) layer is responsible for making suitable policies for the lower parts of neuronal system. This region is located in the midbrain and has descending pathways to the spinal cord via the reticular formations. Here is the center for decisions related to locomotion. Different decisions from higher parts of the brain are entered into this region and it produces some types of high-level and low-level electrical stimulation in order to modify the behavior of central pattern generators [20]. The level of stimulation can modulate the speed of locomotion or translation of the gaits [3]. This region is modeled as a policy learner which gets parameterized inputs (path commands and vision feedbacks) from HLDU layer and generates a policy vector for the next layer. A policy of walking is a stimulation of CPG layer that is formulated as a policy gradient learning problem on some open parameters of the CPG layer. In our previous works [21,22] we did not consider the effects of feedback pathways in this policy vector. In this paper, however the gyro and foot pressure feedbacks are added to the CPG layer and consequently the policy should consider the effects of these state variables in the learning process. This value can determine the instantaneous smoothness of walking. Instantaneous smoothness, speed and precision are combined to compute total payoff of a walking experiment.
- CPG layer: The third layer is the Central Pattern Generator (CPG) layer that consists of some Learner Central Pattern Generator Neural Networks (LCPGNNs). CPG layer is connected to the LLJ layer and sends motion trajectories generated from a high level decision command to the PIDs. The fundamental building block of the LCPGNNs is oscillatory neurons (*o-neurons*), designed and introduced in this paper. These o-neurons have the property of learning the frequency of a periodic input signal and changing it based on some sensory input. Usually, the frequency of an o-neurons can be controlled by a specific parameter in the state representation. In this paper a learning algorithm is introduced for finding specific parameters of o-neurons and synaptic connection weights in a LCPGNN.
- LLJ layer: Low Level Joint (LLJ) layer is composed of PID controllers located in the robot hardware. This layer is directly connected to the robot and controls each Degree Of Freedom (DOF). Its input is the desired positions of the joints which are generated from CPG layer and it also receives the previous joint values as a feedback. It can calculate error and generate appropriate voltages to produce required torques and speeds.