Neural Networks 72 (2015) 109–122

Contents lists available at ScienceDirect

Neural Networks

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

2015 Special Issue

Development of compositional and contextual communicable congruence in robots by using dynamic neural network models

Gibeom Park, Jun Tani *

Department of Electrical Engineering, KAIST, Yuseong-gu, Daejeon, Republic of Korea

ARTICLE INFO

Article history:
Available online 30 September 2015

Keywords:
Dynamic neural network model
Socially intelligent robot
Self-organization
Compositional semantics
Functional hierarchy

ABSTRACT

The current study presents neurorobotics experiments on acquisition of skills for "communicable congruence" with human via learning. A dynamic neural network model which is characterized by its multiple timescale dynamics property was utilized as a neuromorphic model for controlling a humanoid robot. In the experimental task, the humanoid robot was trained to generate specific sequential movement patterns as responding to various sequences of imperative gesture patterns demonstrated by the human subjects by following predefined compositional semantic rules. The experimental results showed that (1) the adopted MTRNN can achieve generalization by learning in the lower feature perception level by using a limited set of tutoring patterns, (2) the MTRNN can learn to extract compositional semantic rules with generalization in its higher level characterized by slow timescale dynamics, (3) the MTRNN can develop another type of cognitive capability for controlling the internal contextual processes as situated to on-going task sequences without being provided with cues for explicitly indicating task segmentation points. The analysis on the dynamic property developed in the MTRNN via learning indicated that the aforementioned cognitive mechanisms were achieved by self-organization of adequate functional hierarchy by utilizing the constraint of the multiple timescale property and the topological connectivity imposed on the network configuration. These results of the current research could contribute to developments of socially intelligent robots endowed with cognitive communicative competency similar to that of human.

1. Introduction

Recently, studies on socially intelligent robots (Breazeal, 2004; Dautenhahn, 2007) have attracted much attention in the research field of intelligent/cognitive robotics. The main motivation of these studies is to investigate theories and methods for building robots that can perform human-like interactions with other agents, including human as well as other robots, autonomously (Breazeal, 2004). Studies on socially intelligent robots inherit some of their design philosophy, as discussed in behavior-based robotics, sourced from Rodney Brooks (Brooks, 1986) in the late 1980s. Although conventional studies on intelligent robots attempted to add behavioral components at a later time, having the basic components on "intelligence" for thinking and cognition built first, the researchers in behavior-based robotics fields did otherwise. By following the thoughts of embodied mind (Rosch, Thompson, & Varela, 1992), they considered that these two processes of thinking and acting should be organized inseparably. Similarly, the researchers in socially intelligent robotics consider the processes of thinking, acting, and communicating as one inseparable process (Billard & Dautenhahn, 1998; Breazeal, 2004; Dautenhahn, 2007).

Recently, not only academia, but also commercial industries, have made great efforts in the development of socially intelligent robots for possible use as home-robots or pet-robots. Such examples can be easily found, including the dog-like robot, AIBO, developed by Sony (Fujita & Kageyama, 1997) and a human-interacting humanoid robot, Pepper, by Aldebaran (2014). They proposed new types of home entertainment for family members through interaction with these robots.

Some other researchers, especially in the research field referred to as developmental robotics (Asada et al., 2009; Cangelosi et al., 2010; Lungarella, Metta, Pfeifer, & Sandini, 2003), have tried to apply various psychological aspects evidenced in human infant development, in building cognitive models of socially intelligent robots. At the same time, they attempted to contribute to the understanding of the underlying mechanism or principles for the development of social cognitive functions via their reconstruction in

* Correspondence to: Room 516, N1 Building, 291 Daehak-ro(373-1 Guseong-dong), Yuseong-gu, Daejeon 305-701, Republic of Korea. Tel.: +82 42 350 7428.
E-mail address: tani1216jp@gmail.com (J. Tani).

http://dx.doi.org/10.1016/j.neunet.2015.09.004
0893-6080/

robotics experiments, by utilizing psychologically and neurobiologically plausible models. These social cognitive functions include learning to imitate action demonstrated by others (Billard, 2002; Demiris & Hayes, 2002; Gaussier, Moga, Quoy, & Banquet, 1998; Ito & Tani, 2004; Schaal, 1999), emergence of turn taking skills such as switching between following and followed among simulated agents (Iizuka & Ikegami, 2004) as well as between human and robots (Nadel, Revel, Andry, & Gaussier, 2004), or joint attention for achieving coordinated behaviors between human and robots (Nagai, Hosoda, Morita, & Asada, 2003; Triesch, Teuscher, Deak, & Carlson, 2006). However, such reconstructions of social cognitive functions through learning have not well addressed the problem of how cognitive competency of systematicity (Cummins, 1996; Fodor & Pylyshyn, 1988) can be developed and also how the contextual flow in particular social cognitive tasks can be captured via iterative learning of the social experiences in the adopted tasks.

The current research aims for reconstruction of cognitive mechanisms with systematicity and context dependency in a neurobiologically plausible manner in a macroscopic sense, which enables robots to perform communicably congruent tasks with human subjects via learning from own sensory-motor experiences. Here, systematicity especially in language processing refers to the cognitive capability of a human to infer the meaning of unknown sentences from known sentences, by extracting compositional semantic rules from them. In the adopted communicably congruent tasks in the current study, we utilize human gestures characterized by systematicity as communicative modality. (It has been shown that human natural gesture recognition capability is also endowed by systematicity (Cassell, Kopp, Tepper, Ferriman, & Striegnitz, 2007; Streeck, 1993).) More specifically, a humanoid robot is tutored to generate specific compositional motor primitive patterns as corresponding to imperative gestures demonstrated by the human subject. An importance here is that the imperative gestures demonstrated by the human subject consists of various combinatorial sequences of movement patterns by following a predefined compositional semantic rules. For example, the robot is requested to respond by generating particular sequences of movement primitives either in order or in reverse order as well as either slowly, normally or quickly as specified by imperative gesture demonstrated.

A technical challenge is to achieve systematicity whereby the robot should become able to infer the underlying meaning or intention for newly demonstrated gesture patterns by means of generalization via learning of prior experienced ones. A difficulty arises because the adopted task uses continuous dynamic patterns of human gesture as well as robot motor response for the communication channel instead of language of discrete symbol sequences. Those dynamic patterns are not always repeatable with variance in their profiles and features. Additionally, in the adopted communicable congruence task, there are no cues that explicitly indicate the structures of the adopted tasks to the robot. First, there are no explicit cues that indicate types of movement patterns performed by the human subject. The demonstrated movement patterns could be either movement primitives to be memorized for regeneration or commands in specifying order of regular or reverse, or with a speed of slow, normal or fast. The underlying type structures should be learned from scratch out of iterative experience of continuous perceptual patterns in demonstrated gesture. Also, there are no explicit cues to segment on-going task flow between the human demonstration phase and the robot response phase in the course of continuous alternation of these two phases. The turn taking between these two phases should be developed autonomously in the course of learning of examples.

Furthermore, the robot should be able to keep the context of the task flow during each session—observing the imperative gesture first, and then generating corresponding behaviors, and meanwhile, the accumulated context in the previous session should be reset in the beginning of new session. Such control of the contextual flow, as well as the control of turn taking, are considered to belong to another human specific cognitive capability, which should be developed gradually via iterative learning of task examples. Although the current cognitive task using a particular set of imperative gestures may not replicate our everyday social cognitive behaviors exactly, human, certainly, use gestures to achieve communication characterized by systematicity and context sensitivity (Arbib, 2012; Bowie, 2008; Kendon, 2004). It can be said that the current robotics experiment attempts to model such social cognitive competency of human in an abstract manner.

Here, the technical challenges focused in the current research are summarized as: (1) the adopted MTRNN can achieve generalization by learning in the lower feature perception level by using a limited set of tutoring patterns, (2) acquisition of compositional semantics with generalization for achieving communicable congruence tasks characterized by systematicity, (3) development of cognitive mechanism for controlling the turn-taking process as well as controlling the contextual flow as situated to on-going task processes. The aforementioned technical challenges of targeting development of the cognitive competency characterized by systematicity and context dependency out of lower level perceptual experiences are considered to be novel as compared to those prior-existing studies aiming to reconstruct social cognitive functions via learning, as previously mentioned. Therefore, these technical challenges could contribute significantly to the realization of socially intelligent robots.

For the purpose of accomplishing the aforementioned challenges, the current study takes on an approach based on the paradigm of dynamical systems and self-organization in modeling the development of target cognitive-behavioral processes, because the research outcomes accumulated for two decades have shown that this approach is one of the best to account for the essence of the embodied cognition (Beer, 2000; Clark, 1999; Kelso, 1997; Lewkowicz & Lickliter, 1995; Port & Van Gelder, 1995; Tani, 1996; Thelen, 1994). The current research especially follows the results from the study conducted by Yamashita and Tani (2008), which conjectured on how actional compositionality may be developed by self-organization of particular dynamic neural network models via iterative learning of sensory-motor experiences. They showed that functional hierarchy for generating complex behaviors can be developed through iterative learning of sensory-motor experiences by utilizing the reported multiple timescales recurrent neural network (MTRNN) model. In their study, it was shown that a set of behavior primitives can be learned in the fast timescale dynamics of the lower level subnetwork, whereas sequential combinations of these behavior primitives are learned in the slow timescale dynamics of the higher level subnetwork. The current study is related also to a robotics study on the associative learning between proto-language and behaviors conducted by Sugita and Tani (2005) by utilizing a version of recurrent neural network (RNN), so-called the recurrent neural network with parametric bias (RNNPB). This study investigated how the compositional semantic rules can be extracted with generalization from the iterated tutoring experience by learning.

The current research attempts to apply these frameworks to realization of a communicably congruent response of robots to humans i.e., robots generating corresponding or expected actions to gestures demonstrated by a human, which is characterized by systematicity and context sensitivity. Detailed analysis on the results of the robotics experiment will clarify how the cognitive competency necessary for achieving the aforementioned communicative skills can be developed in the course of self-organizing adequate dynamic structures in the adopted dynamic neural network model. Next section will describe general idea of the adopted human–robot communicably congruent task which is followed by the descriptions of the employed model and a set of experiments performed.