AAAI Technical Report SS-12-02
Designing Intelligent Robots: Reintegrating AI

Can a Robot Learn Language as a Child Does?*

Stephen E. Levinson, Logan Niehaus, Lydia Majure, Aaron Silver, and **Luke Wendt**
Department of Electrical and Computer Engineering
University of Illinois at Urbana-Champaign, Urbana, IL 61801

Abstract

This paper gives a brief retrospective of a research project begun in 1987 and continuing to the present on the topic of language acquisition by an autonomous humanoid robot. We recount the motivations for, theoretical bases of and experimental results on this subject. Important results include novel models and algorithms resulting in interesting linguistic function of our robots.

The Strong Theory of AI is clearly expressed in Turing's seminal 1950 paper in which he proposes the infamous "Turing Test" for intelligent behavior. The intuition behind his elaborate argument is that the Universal Turing Machine is capable of performing virtually any symbol manipulation process and is therefore sufficient for creating a mental model of the world. The preferred realization of this idea is a synthetic model, complete in every detail, that computes a symbolic representation of meaning from natural language text. This process is to be based on predetermined operations on predefined symbols. The symbols and operations are to be determined by some combination of scientific observation, introspection and divine inspiration.

In the penultimate paragraph of this paper, Turing offers an astounding and often overlooked alternative suggesting that the symbols and relations amongst them could be learned from real-world sensory data. In fact, he urges that both approaches be tried. It may be argued that both approaches conform to the Strong Theory of AI but in significantly different ways. The direct synthesis is predicated on a discrete symbolic model perfectly isomorphic to reality and unaffected by any uncertainty present in the physical world. This approach assumes that the sensorimotor periphery may safely be ignored. The alternative theory uses the computational power of the Turing Machine to analyze the physical processes from which distributed symbols and structures derive. Thus, cognitive function emerges from physical measurement and mathematical description of the experience of and participation in reality.

A unification of these two complementary interpretations of the Strong Theory leads to the following hypothesis about brain, mind and language. The disembodied mind is a fantasy. Thought is almost exclusively the product of the vast associative memory called the brain. The memory is able to capture spatiotemporal order and represent it episodically. Thus there can be no isolated perceptual or cognitive functions. Memory is built up from instincts by the reinforcement of successful behavior in the real world at large. As a cognitive model of reality is acquired, a linguistic image of it is formed primarily in response to semantic information. Other levels of linguistic structure exist to make semantics robust to ambiguity. When the language is fully acquired, most mental processes are mediated linguistically and we appear to think in our native language which we hear as our mind's voice.

The argument outlined above is discussed in detail in chapters 9 and 10 of (Levinson 2005) and serves as the basis for our ongoing research on the role of sensorimotor function, associative memory and reinforcement learning in automatic acquisition of spoken language by an autonomous robot. Over the years we have made some progress toward this most ambitious goal. We have integrated visual navigation and object manipulation under voice command (Squire and Levinson 2007) and then augmented it with syntax learning (McClain 2007). In these experiments we spoke naturally to the robot while pushing it around its "playpen". This resulted in the acquisition of lexical semantics for physical objects such as "green ball" and "red can". The visual navigation ability was learned by means of a Markov Decision Process whose parameters were estimated by a variant of Q-learning. The perceptual and linguistic abilities were simultaneously learned by means of a nested collection of HMMs whose parameters were estimated by an incremental stochastic gradient algorithm. Later we added a simple grammatical inference algorithm so that the robot was able to learn the production rules of a small context-free grammar and the lexical semantics of action verbs and the compositional semantics of short sentences. In this experiment we also provided speech synthesis so that the robot could perform an action and then use the grammar generatively to explain its actions.

As members of the EUCog project, we recently acquired an iCub humanoid robot. This has enabled us to study anthropomorphic fine motor control (Silver, Wendt, and Levinson 2010) and language acquisition (Majure et al. 2010).

*This work was supported in part by the Office of Naval Research under grant N00173-10-1-G-017.

The motor control experiments are based on the design of a PID controller to allow the robot to balance a ping-pong ball at any designated location on a flat plate held in the robot's hand based on visual perception. We are now studying the process of reinforcement learning of the controller by means of a Q-learning algorithm where the Q-function computes the reward of a state/action pair. These experiments are a prelude to getting the robot to learn to stand upright and walk.

At present we are studying cortical models for associative memory (Duda and Levinson 2010). These are ab initio simulations beginning at the cellular level in which we use well-known cellular electrodynamics to produce spiking neurons. These are clustered into small networks whose collective properties such as firing density and phase synchrony are taken as information bearing features. The small networks are interconnected by fixed neural pathways into larger networks whose non-linear dynamics display multiple metastable equilibria that can be used as memories.

Another ongoing study is that of the relationships among motor control, spatial reasoning, and the semantics of action words. Here we address the question of what kind of stochastic model can learn a symbolic representation of spatial reasoning derived from sensorimotor function and how can that representation be used to support the deep semantics of language. We are conducting experiments to get the robot to learn to reproduce simple hand and arm gestures from direct demonstration and/or imitation and the spoken descriptions of the gestures (Niehaus 2011).

References

Duda, A. M., and Levinson, S. E. 2010. Nonlinear dynamical multiscale model of associative memory. In *Proc. IEEE 9th Int. Conf. on Machine Learning and Applications (ICMLA 2010)*, 867 872.

Levinson, S. 2005. *Mathematical Models for Speech Technology*. New York, NY: John Wiley and Sons Ltd.

Majure, L.; Niehaus, L.; Duda, A.; Silver, A.; Wendt, L.; and Levinson, S. 2010. Integrating language and motor function on a humanoid robot. In *Proc. Int. Conf. on Intelligent Robots and Systems (IROS 2010)*.

McClain, M. 2007. Semanticbased learning of syntax in an autonomous robot. *Int. J. Humanoid Robotics* 4(2):321 346.

Niehaus, L. 2011. Autonomous learning of action-word semantics in a humanoid robot. Master's thesis, University of Illinois at Urbana-Champaign, Urbana, IL.

Silver, A.; Wendt, L.; and Levinson, S. E. 2010. isionbased balancing tasks for the icub platform: A case study for learning external dynamics. In *10th IEEERAS Int. Conf. on Humanoid Robots*.

Squire, K., and Levinson, S. 2007. HMM-based semantic learning for a mobile robot. *IEEE Trans. on Evolutionary Computation* 11:199–212.