Chapter 84
Learning Words by Imitating

Thomas Cederborg
INRIA, France

Pierre-Yves Oudeyer
INRIA, France

ABSTRACT

This chapter proposes a single imitation-learning algorithm capable of simultaneously learning linguistic as well as nonlinguistic tasks, without demonstrations being labeled. A human demonstrator responds to an environment that includes the behavior of another human, called the interactant, and the algorithm must learn to imitate this response without being told what the demonstrator was responding to (for example, the position of an object or a speech utterance of the interactant). Since there is no separate symbolic language system, the symbol grounding problem can be avoided/dissolved. The types of linguistic behavior explored are action responses, which includes verb learning but where actions are generalized to include such things as communicative behaviors or internal cognitive operations. Action responses to object positions are learnt in the same way as action responses to speech utterances of an interactant. Three experiments are used to validate the proposed algorithm.

INTRODUCTION

A growing number of experimental results and theories suggest that language is a process that strongly interacts with and is grounded in action and perception (Rizzolatti & Arbib, 1998; Glenberg & Kaschak, 2002; Pulvermuller, Hauk, Shtyrov, Johnsrude, Nikulin, & Ilmoniemi, 2003; Hauk, Johnsrude, & Pulvermuller, 2004). This notion has also been discussed within the robotics community (see for example Cangelosi, et al., 2010; Perani, et al., 2003). That language cannot be separated from action is thus a well-accepted notion. Language and action learning are however still regarded as two different systems, and the problem of how to integrate these two separate systems is sometimes referred to as the symbol grounding problem (see Steels, 2007a, for a description of the problem and solutions). The problem arises when a separate symbol system must be connected to a separate action system. However, if language is learned via a more general system that imitates both actions and language, this difficult problem does not arise (a single imitation learning strategy, learning how to respond to any context, no matter if that context includes speech

or is completely made up of inanimate objects can account for both). This chapter describes one possible such imitation learning system that can learn both non-communicative actions as well as linguistic skills, and tests this system in three experiments. The symbol grounding problem does not arise, simply because there is no separate symbol system that needs to be connected to an action system.

The focus of the present chapter is verb learning–the learning of action concepts and learning that there is a speech utterance or hand sign associated to this action concept. An imitation learner watches two adult humans, one interactant that may speak or make a hand gesture, and one demonstrator that performs an action. After several such interactions, the imitator is confronted with a situation that among other things includes the interactant. The imitator attempts to respond as the demonstrator would have responded. The idea here is that the imitator will treat the interactant (and his/her utterance or hand gesture) as any other part of the context, and if the demonstrator sometimes responds to the interactant, but at other times responds to other elements of the context, the imitator can utilize the same strategy for correctly imitating all of these responses. When viewing a specific demonstrator action, the imitator is not told what this is a response to (either something the interactant did, or something else in the environment). Since the imitator is not told in advance what part of the environment should trigger an action, no bias is displayed for the mode of communication. Thus, in the second experiment, the imitator learns words in speech, as well as words in a sign language, concurrently without problems. The second experiment goes beyond verb learning, as some of the actions that are learnt would look like communicative acts by an outside observer (e.g., responding to speech with a hand sign or describing the environment with a hand sign). In the third experiment, the imitator learns verbs and concurrently learns when to perform operations on an internal cognitive structure, viewing such

internal operations as similar to physical actions. Regardless of whether the actions are responses to a linguistic stimulus or to the properties of an object, the imitator must solve the problem of

1. Identifying which parts of the context and the action are important and
2. Deciding what to do in situations that are similar but not exactly the same.

Tomasello, Carpenter, Call, Behne, and Moll (2005) describe the referential ambiguity of physical actions. They state that, "the exact same physical movement may be seen as giving an object, sharing it, loaning it, moving it, getting rid of it, returning it, trading it, selling it, and on and on depending on the goals and intentions of the actor." This ambiguity is very similar to the type of ambiguity in language acquisition that Quine referred to as the Gavagai problem (Quine, 1960): The problem of how to guess the meaning of a new word when many hypothesis can be formed (out of a pointing gesture), and it is impossible to read the mind of the language teacher.

For example, imagine a demonstrator, that we will call Steve, trying to teach you something new. Steve and you are both looking at a rabbit just in front of the trees of a forest. Steve takes a stone on his right, and throws it towards the rabbit with a parabolic-shaped trajectory. The stone arrives one meter to the left of the rabbit, just below a tree with blue flowers. Now Steve asks you to try to reproduce what he did. In the meantime, the rabbit moves 10 meters away, a cat has arrived next to the tree, and there are no more stones on your right or Steve's right, but a stone on your left and a knife on your right. What should you do? Would you take the stone or the knife? Would you throw it in any direction, trying to reproduce a parabolic trajectory? Or would you throw it towards the left of the rabbit, or onto the rabbit, or to the left of the cat? Or try to throw it just below a tree with blue flowers? Did Steve intend to throw something at the rabbit (but miss it) or to the closest animal?

Related Content

Analysis-by-Synthesis Echo Watermarking
Wen-Chih Wuand Oscal Chen (2008). *Digital Audio Watermarking Techniques and Technologies: Applications and Benchmarks (pp. 152-171).*
www.irma-international.org/chapter/analysis-synthesis-echo-watermarking/8330

Extracting Definitional Contexts in Spanish Through the Identification of Hyponymy-Hyperonymy Relations
Olga Acosta, Gerardo Sierraand César Aguilar (2015). *Modern Computational Models of Semantic Discovery in Natural Language (pp. 48-70).*
www.irma-international.org/chapter/extracting-definitional-contexts-in-spanish-through-the-identification-of-hyponymy-hyperonymy-relations/133875

Cocktail Party Problem: Source Separation Issues and Computational Methods
Tariqullah Janand Wenwu Wang (2011). *Machine Audition: Principles, Algorithms and Systems (pp. 61-79).*
www.irma-international.org/chapter/cocktail-party-problem/45481

Combinatory Categorial Grammar for Computer-Assisted Language Learning
Simon Delamarreand Maryvonne Abraham (2012). *Cross-Disciplinary Advances in Applied Natural Language Processing: Issues and Approaches (pp. 258-272).*
www.irma-international.org/chapter/combinatory-categorial-grammar-computer-assisted/64592

LinkedVis an Information Visualisation Toolkit for RDF Data
Antonio Garroteand María N. Moreno García (2014). *Computational Linguistics: Concepts, Methodologies, Tools, and Applications (pp. 799-815).*
www.irma-international.org/chapter/linkedvis-an-information-visualisation-toolkit-for-rdf-data/108752