Providing Cognitive Functions for Interactive Learning with Speech and Multimodal Processing

Hartwig Holzapfel, Alex Waibel
interACT Research, Universität Karlsruhe (TH)
hartwig@ira.uka.de

Extended Abstract

Cognitive systems can roughly be categorized into systems that try to imitate human cognition and systems that try to imitate cognition on a functional basis, to compare which functionality human cognition can achieve. Both approaches are driven forward with usually different research approaches. In this paper we address influences to our work by both approaches and especially look for hints that can help us further in designing a cognitive system, while focussing on the development of an interactive learning component that uses speech and multimodal information.
An important part of cognition is the ability to acquire new knowledge though learning mechanisms, which enhances an artificial system with the ability to develop or adapt to a new environment. In contrast to most learning algorithms applied in machine learning today, which mainly work with offline learning on training samples, learning in a cognitive system needs to be performed autonomously and through interaction with the environment or with other agents/humans.
We present ongoing work in this area of interactive learning with first experiments on having the robot learn words, objects and their semantic interpretation, and argue for a modality independent approach to knowledge acquisition through dialog.

To equip the robot with the ability to learn new information we apply an approach which first detects information that is inconsistent with the existing knowledge base, estimate the need of acquiring more information about it, and then applying most promising actions to incorporate newly acquired knowledge, and storing the results in the knowledge base.
This approach was firstly evaluated on a task where the robot learns new words for previously unknown objects and secondly applied to interactively learning to now persons using speech and vision.
Our approach to interactive semantic learning and knowledge acquisition first uses out-of-vocabulary (OOV) detection to find unknown words and at a second stage, a dialog acquires semantic information about the object from a human user.