Knowledge Acquisition in Dialogue with the interACT Receptionist Robot

Hartwig Holzapfel
Universität Karlsruhe
Karlsruhe, Germany
hartwig@ira.uka.de

Abstract

Recently there has been great advance in building humanoid robots and other autonomous systems. Acquiring new knowledge through interactive learning mechanisms is a key ability for such systems in a natural environment. In recent and ongoing work we focus on approaches for natural learning that enables an autonomous system, such as a humanoid robot, to acquire new information through multimodal and natural language dialogues with humans.

1 Introduction

Key aspects of learning approaches with natural language dialogue are that the learning mechanisms are performed autonomously and only through interaction with the environment or with other agents/humans. This means that in contrast to classical learning methods, **dialog based learning (DBL)** doesn't require a human operator to manually annotate data, such as in supervised learning. Rather, knowledge is extracted from information acquired interactively. We have applied such 'learning dialogues' to person ID, social information about persons and object learning in the domain of human-robot interaction. In contrast to supervised learning, learning dialogues are error-prone, as are background processing methods that initialize or extend the knowledge base using information extraction on the World-Wide-Web. To reduce errors and improve knowledge base quality, robust dialogue strategy design, coupling of dialogue and speech recognition, and multimodal processing have been examined. As the human user is an important resource for learning dialogues, evaluation of such a system considers a complexity of dialogue metrics, subjective user feedback, knowledge base quality and learning success.

2 The interACT receptionist as a scenario for dialogue-based learning

In this abstract we briefly introduce our learning scenario, outline the learning method, and summarize key aspects, including conducted and ongoing experiments.

A scenario where the approach of learning dialogues are examined in a social setting has been established with the interACT receptionist robot, which is located in a corridor in the entrance area of our institute. The receptionist is a simple robotic platform with speech capabilities and stereo vision. A stereo camera is mounted on a pan-tilt unit for visual perception and person tracking. The task of the robot is to engage in a dialogue with persons and create a model of people working at the institute. The model includes names, face snapshots, organizational roles, research interests and other information. In the background, social network models represent group structures and researcher collaboration. By modelling such information, the system serves as an autonomous web site administrator for a Who-is-Who page, which is accessible from within the institute, and which reflects the system's knowledge. Its learning task includes learning of new persons, removing persons from the knowledge base and correcting the knowledge base.

The dialogue approaches presented here have been implemented and tested in the dialogue manager Tapas (Hartwig Holzapfel, 2008). It integrates with the Janus speech recognition toolkit and IBIS decoder, and the integrated real-time tracking software Arthur[1]. The following sections describe approaches and experiments with the interACT receptionist scenario. An application to semantic category acquisition and object learning for a humanoid robot is presented in (Holzapfel et al., 2008).

[1] http://isl.ira.uka.de/ nickel/arthur/

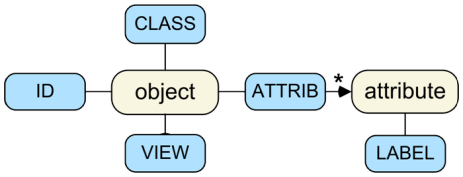

Figure 1: knowledge entity model

3 A dialogue approach to dynamic knowledge acquisition

Real life environments inherently carry dynamically changing information, which the learning system tries to model. Figure 1 shows the model of a basic entity in the knowledge base. In the receptionist task, entities represent persons. The robot needs to interpret visual and acoustic data (VIEWs) and understand the person's name (ATTRIButes). The initial challenge is to identify a person the first time with the correct name. Afterwards, the system associates an ID with the person and can recognize the person, also using face ID.

The interACT receptionist dialogue system comprises several dialogue modules for greeting, identification, questions about person, robot info and goodbye. We, as well as others, have found that a modular dialogue strategy supports the development of a complex system, with separate modules for different dialogue tasks (Holzapfel and Waibel, 2007). Furthermore, it allows to switch single module implementations, without influencing other modules. By using a modular approach, we could also integrate the identification module, which has been trained by a reinforcement learning approach, with other, handwritten modules. In ongoing work we evaluate the system in a long-term study with speech recognition experts, who work at the institute, and different series of naive users, as institute visitors, where the system was running 24 hours a day.

3.1 Integration with speech recognition and dialogue strategies

The receptionist learning scenario requires recognition of person names, project names or unknown terms for research interests that are not covered by the speech recognizer's vocabulary. Thus it is necessary to learn unknown words by OOV-detection and dynamic vocabulary approaches or spelling. A first integration of these components in dialogue has been reported in (Holzapfel et al., 2006). Subsequently we have improved name recognition and word learning, e.g. multiple speech recognition passes and name selection from social network data to improve recognition rates. In further ongoing work we address error-rate sensitive integration in the dialogue strategy, e.g. as in (Holzapfel and Waibel, 2008), where Reinforcement Learning with a multimodal user simulation is applied to optimize person identification dialogues.

3.2 Multimodal processing

Another important aspect of the learning scenario is processing of multimodal data. In contrast to other identification systems we address open set identification, which includes unknown persons, and online learning for both, user ID and names. Besides integration of face ID and voice ID (Ekenel and Jin, 2006), we analyzed confidence measures for multimodal ID, and user ID estimation with belief networks (Holzapfel and Waibel, 2008b), which improve integration with the dialogue strategy that can then dynamically decide which modality to trust.

References

Hazim K. Ekenel, Qin Jin. 2006. ISL Person Identification System in the CLEAR Evaluations. *CLEAR Evaluation Workshop*, Southampton, UK.

Hartwig Holzapfel. 2008. A Dialogue Manager for Multimodal Human-Robot Interaction and Learning of a Humanoid Robot. *Industrial Robots*, 35(6):528–535, Emerald Publishing.

Hartwig Holzapfel, Daniel Neubig and Alex Waibel. 2008. A Dialogue Approach to Learning Object Descriptions and Semantic Categories. *Robotics and Autonomous Systems, Special Issue on Semantic Knowledge in Robotics*, 56(11):1004-1013 Elsevier.

Hartwig Holzapfel, Thomas Schaaf, Hazim Kemal Ekenel, Christoph Schaa and Alex Waibel. 2007. A Robot learns to know people - First Contacts of a Robot. *KI 2006: Advances in Artificial Intelligence, Lecture Notes in Computer Science*, 4314:302–316.

Hartwig Holzapfel and Alex Waibel. 2008. Learning and Verification of Names with Multimodal User ID in Dialog. *International Conference on Cognitive Systems (CogSys 2008)*, Karlsruhe, Germany.

Hartwig Holzapfel and Alex Waibel. 2007. Behavior Models for Learning and Receptionist Dialogs. *Interspeech 2007*, Antwerp, Belgium.

Hartwig Holzapfel and Alex Waibel. 2008. Modeling Multimodal User ID in Dialogue. *SLT Workshop*, Goa, India.