Language Acquisition from Neural and Sensorimotor Systems

Henrik Jacobsson, Geert-Jan M. Kruijff
DFKI, German Research Center for Artificial Intelligence, Saarbrücken (Germany)

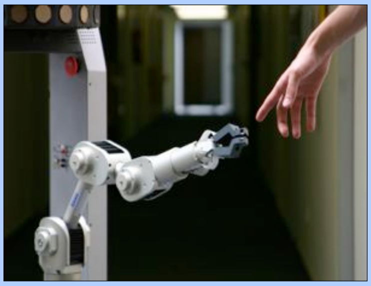
Maria Staudte
Saarland University, Saarbrücken (Germany)

IST-4250
CogSys Cognitive Systems

Overview

Key Ideas:

- Discrete models of situated embodied experience can be acquired through active learning
- Acquired models will be strongly robo-centric
- Mediation between sensorimotoric experience and high-level cognitive processes (such as language and planning) can be established through situated dialogue
- Negotiation of the mapping between robocentric and human terminology



Approach:

- Bottom up active learning of embodied concepts solely on the basis of actions and perceptions of the robot forms the robo-centric ontology
- Method: Rule extraction from recurrent neural networks (RNN-RE) discretization of dynamic systems
- Top-down categorization by the human user
- Method: Situated dialogue
- Integration: the CoSy Architecture Schema Toolkit (CAST)

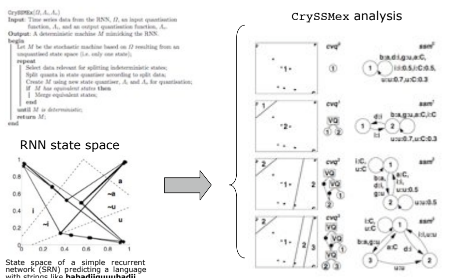
System Implementation

Method

Through the novel algorithm, `CrySSMEx` [1], extraction of rules from Recurrent Neural Networks (RNNs) [2] is possible for deep context-free grammars, for large RNNs (10^3 state nodes), chaotic systems and, for example, from echo state networks predicting natural language sequences [3]. A trained RNN in combination with `CrySSMEx` is a form of language acquisition. `CrySSMEx` analyses the state space of the RNN by learning how inputs affect the state and output of the RNN

- Anytime generation of finite state description (from simple, stochastic machines to more detailed descriptions).
- Hierarchical decomposition of state space.
- Active selection of disambiguating state data
- Parameter free and deterministic.

`CrySSMex` Example



State space of a simple recurrent network (SRN) predicting a language with strings like **babadiguuubadii...**

I, Robot

Hardware:
People Bot, Katana Arm, Pan-tilt-unit & Stereo Vision camera, SICK laser scanner, table sensor (IR), bumping sensors, sonars, haptic sensors
Software:
CoSy Architecture Schema Toolkit (CAST) [4]
Subarchitectures:
Dialogue system with VoIP interface, Vision, Spatial modelling, Planner, SLAM, Ontological reasoning, Manipulation, Crossmodal binding between subarchitectures

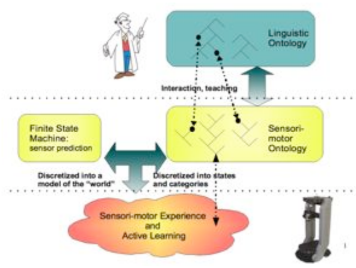
Rule Extraction for Robots

The goal is to find the "grammar" underlying the dynamics in the "dialogue" between the robot and its environment. CrySSMEx supports this by extracting a stochastic finite state machine description of the system. Instead of analysing an RNN, CrySSMex will be applied to the sensorimotor feedback loop. The inputs to the RNN are replaced with actions, and the output with perception.

Accompanying the machine is a hierarchical description of how the continuous state space of the dynamic system is mapped to the extracted discrete states (i.e. grounding the grammar in the system). This hierarchical description can be considered CrySSMEx's ontology of the state space.

We essentially want to view the learning of perceivable consequences of actions, in different situations (states), as a language acquisition problem.

The mapping of the extracted ontology to the user's ontology can be viewed as a socially guided machine learning task.



Goal Scenario

The goal scenario is to let `CrySSMEx` control the robot in a simple setting in which the robot can start from "scratch" to build up a grounded embodied world model. After a while it can hopefully sufficiently predict the consequences of its actions and has established a hierarchical model of world states. At this point, it should be possible to tag states and sets of states with labels through user interaction, e.g. by the user explaining to the robot that "your current action is called *bumping*" the robot could infer that the current state is part of a family of states that are involved in something which can be labelled "bumping", assuming that also closely related states may be labelled "bumping" (something which may be verified through an interactive clarification dialogue). At this point there will essentially be a connection between a concept of the robot's world view, i.e. an instance of an enumeration of perception, and a human concept. Our conjecture is that to learn this concept is a trivial translation once the robot has the capability of predicting "bumping". Once the association is made, the user should be able to give high-level commands that can be translated to planning goals, e.g. "avoid bumping".

References

[1] H. Jacobsson, "The crystallizing substochastic sequential machine extractor - CrySSMEx," *Neural Computation*, vol. 18, no. 9, pp. 2211–2255, 2006.
[2] H. Jacobsson, "Rule extraction from recurrent neural networks: A taxonomy and review," *Neural Computation*, vol. 17, no. 6, pp. 1223–1263, 2005.
[3] H. Jacobsson, S. L. Frank, and D. Federici, "Automated abstraction of dynamic neural systems for natural language processing," in *Proceedings of IJCNN'2007*.
[4] N. Hawes, A. Sloman, J. Wyatt, M. Zillich, H. Jacobsson, G. Kruijff, M. Brenner, G. Berginc, and D. Skocaj. Towards an integrated robot with multiple cognitive functions. In Proceedings of the Twenty-Second Conference on Artificial Intelligence (AAAI-07), 2007.