# Simulating processes of language emergence, communication and agent modeling

Timo Honkela, Ville Könönen, Tiina Lindh-Knuutila and Mari-Sanna Paukkeri
Adaptive Informatics Research Centre
Helsinki University of Technology, Finland
firstname.lastname@tkk.fi

**Abstract**

We discuss two different approaches for modeling other agents in multiagent systems. One approach is based on language between agents and modeling their cognitive processes. Another approach utilizes game theory and is based on modeling utilities of other agents. In both cases, we discuss how different machine learning paradigms can be utilized for acquiring experience from the environment and other agents.

## 1 Introduction

In this paper, we discuss how the emergence of subjective models of the world can be simulated using different approaches in learning and what is the role of communication and language. We consider, in particular, the role of unsupervised learning in the formation of agents' conceptual models, the original subjectivity of these models, the communication and learning processes that lead into intersubjective sharing of concepts, and a game theoretical approach to multiagent systems including reinforcement learning processes.

An intelligent agent usually has a purpose or a goal, probably set by the designer of the agent, and the agent tries to act rationally for satisfying the goal. For making rational decisions, the agent has to model its environment, perhaps including other agents. In the following, we consider some specific issues related to the modeling.

### 1.1 Complex and non-stationary environments

In the easiest case, the environment in which the agent is located is static, i.e. all properties of the environment remain constant for the whole life-time of the agent. However, often the situation is not so simple. The properties of the environment may vary with time, i.e. the environment is called *non-stationary*. The non-stationarity may be due to the environment itself or the limited resources of the agent. For example, in many real problem instances, the learning agent is not capable to sense the real state of the environment and thus some relevant properties of the environment remain hidden.

Another example of non-stationarity are multiagent systems. In these systems, properties of the environment for each individual agent depend on the actions of all agents located in the real environment. Thus for acting rationally, agents should model also other agents in the system.

Pfeifer and Scheier (1999) stress that the behavior of an agent is always the result of system-environment interaction. It cannot be explained on the basis of internal mechanisms only. They illustrate that the complexity that we as observers attribute to a particular behavior does not always indicate accurately the complexity of the underlying mechanisms. Experiments with very simple robots that merely react to stimuli in their environment have shown that rather complex behavior can emerge.

### 1.2 Subjective versus intersubjective

Moore and Carling (1988) state that "[l]anguages are in some respect like maps. If each of us sees the world from our particular perspective, then an individual's language is, in a sense, like a map of their world. Trying to understand another person is like trying to read a map, their map, a map of the world from their perspective." In many computational approaches to semantics and conceptual modeling, an objective point of view has been used: It is assumed that all the agents have a shared understanding and representation of the domain of discourse. However, Moore and Carling's statement emphasizes the need for explicit modeling of the other's point of view.

This requires modeling of the subjective use of language based on examples, and, furthermore, to model intersubjectivity, i.e. to have a model of the contents of other subjective models.

As an example related to the vocabulary problem, two persons may have different conceptual or terminological "density" of the topic under consideration. A layman, for instance, is likely to describe a phenomenon in general terms whereas an expert uses more specific terms.

### 1.3 Learning agents

Different machine learning techniques can be utilized for adding adaptivity to agent-based systems. There are basically three major learning paradigms in machine learning: *supervised learning*, *unsupervised learning* and *reinforcement learning*. In supervised learning, there exists a teacher having knowledge of the environment, in the form of input-output pairs, and the learning system for which the environment is unknown. The teacher provides samples from the environment by giving correct outputs to inputs and the goal of the learning system is to learn to emulate the teacher and to generalize the samples to unseen data. In unsupervised learning, contrary to supervised learning, there exists no external teacher and therefore no correct outputs are provided. Reinforcement learning lies between supervised and unsupervised learning: Correct answers are not provided directly to the learning system but the features of the environment are learned by continuously interacting with it. The learning system takes actions in the environment and receives reward signals from the environment corresponding to these action selections.

## 2 Unsupervised learning of conceptual systems

In the following, we consider the unsupervised learning of conceptual systems. We first study the modeling an individual agent that learns to create a conceptual space of its own and learns to associate words and expressions with conceptual space. The notion of *conceptual space* is taken from Gärdenfors who also presents the basic motivation and framework for dealing with conceptual spaces (Gärdenfors, 2000). After considering one individual agent, we consider a multiagent system in which a shared conceptual system is formed in a self-organized manner.

### 2.1 One agent

The *Self-Organizing Map (SOM)* (Kohonen, 1982, 2001) is an unsupervised learning model which is often considered as an artificial neural network model, especially of the experimentally found ordered "maps" in the cortex. The SOM can be used as a central component of a simulation model in which an agent learns a conceptual space, e.g. based on data in which words are "experienced" in their due contexts (Ritter and Kohonen, 1989). This approach has been used, for instance, to analyze the collection of Grimm fairy tales. In the resulting map, there are areas of categories such as verbs and nouns. Within the area of nouns a distinction between animate and inanimate nouns emerges (Honkela et al., 1995). The basic idea is that an agent is able to form autonomously a conceptual mapping of the input based on the input itself.

### 2.2 Community of agents

The basic approach how autonomous agents could learn to communicate and form an internal model of the environment applying self-organizing map algorithm was introduced, in a simple form, in (Honkela, 1993). Later we developed further the framework that would enable modeling the degree of conceptual autonomy of natural and artificial agents (Honkela et al., 2003). The basic claim was that the aspects related to learning and communication necessitate adaptive agents that are partially autonomous. We demonstrated how the partial conceptual autonomy can be obtained through a self-organization process. The input for the agents consists of perceptions of the environment, expressions communicated by other agents as well as the recognized identities of other agents (Honkela et al., 2003). A preliminary implementation of a simulated community of communicating agents based on these ideas did not succeed to fully demonstrate the emergence of a shared language (Honkela and Winter, 2003).

When language games (Wittgenstein, 1953) was included in the simulation model, it resulted in a simple language emerging in a population of communicating autonomous agents (Lindh-Knuutila, 2005). In this population, each agent was able to create their own associations between the conceptual level and the emerged words, although each agent had a slightly different conceptual representation of the world. The learning paradigm for the conceptual learning was fundamentally unsupervised, but the language learning tested has been so far supervised, i.e. the communicating agents are provided feedback of the outcome of the game as well as the "right answer". The reinforcement learning and the unsupervised learning models for language games remain to be implemented.

# 3 Game theoretical approach for multiagent systems

As discussed in Section 1, an intelligent agent usually has a goal. For studying agent based systems formally, e.g. for developing learning algorithms for intelligent agents, it is useful that the goal can be expressed mathematically. Traditionally approach is to define an *utility function* for the agent, i.e. there is a scalar value connected to each possible action measuring the fitness of the action choice for satisfying the goal of the agent. The utility function is often initially unknown and must be learned by interacting with the environment. Machine learning techniques can be utilized in this learning process, consider e.g. (Könönen, 2004b; Könönen and Oja, 2004; Könönen, 2004a).

## 3.1 One agent

In single-agent systems, achieving the rational behavior is simple: the agent always selects an action with the highest utility value. There exists a vast number of learning methods that utilize, in one way or another, the rational decision making in agent-based systems.

## 3.2 Community of agents

The branch of science studying decision making in single-agent systems is called *decision theory*. However, when there exists multiple active decision makers (agents) in the same environment, decision theory is not suitable for achieving rational behavior any more. *Game theory* is an extension of decision theory to multiagent systems. In game theory, agents explicitly model the dependency of their utility functions on the actions of all agents in the system. The goal of the agents is to find *equilibrium actions*, i.e. the actions that maximize their utility values assuming the all agents will use the same equilibrium. Jäger (2006) applies game theory to examine the shape formation in the conceptual spaces of the agents (Gärdenfors, 2000).

Theoretically, game theory provides a perfect tool for modeling multiagent systems. In practice, there are many problems with game theory. For example there can exist multiple equilibria and the agents should coordinate which one they will select. For calculating an equilibrium, the agents should not only model their own utility function but also the functions of all other agents in the system. This is often intractable and therefore some other methods, e.g. Bayesian techniques, should be used for creating more coarser models of other agents. By using these "external" models, the game theoretical problem reduces to the simpler problem solvable by using decision theory.

# 4 Discussion

Language does not need to be viewed plainly as a means for labeling the world but as an instrument by which the society and the individuals within it construct a model of the world. The world is continuous and changing. Thus, the language is a medium of abstraction rather than a tool for creation and mediation of an accurate "picture" of the world. The point of view chosen is always subject to some criteria of relevance or usefulness. This is not only true for the individual expressions that are used in communication but also concerns the creation or emergence of conceptual systems. It makes sense to make such distinctions in a language that are useful in one way another.

In this paper, we have discussed the role of unsupervised learning in the formation of agents' conceptual models in a non-stationary environment and a game theoretical approach for multiagent systems. In the future, we will study game theoretical models for language games and how to apply machine learning approaches for solving these games.

# References

Peter Gärdenfors. *Conceptual Spaces*. MIT Press, 2000.

Timo Honkela. Neural nets that discuss: a general model of communication based on self-organizing maps. In S. Gielen and B. Kappen, editors, *Proceedings of ICANN'93, International Conference on Artificial Neural Networks*, pages 408–411, Amsterdam, the Netherlands, September 1993. Springer-Verlag, London.

Timo Honkela and Juha Winter. Simulating language learning in community of agents using self-organizing maps. Technical Report A71, Helsinki University of Technology, Laboratory of Computer and Information Science, Espoo, Finland, 2003.

Timo Honkela, Ville Pulkki, and Teuvo Kohonen. Contextual relations of words in Grimm tales analyzed by self-organizing map. In *Proceedings of ICANN-95, International Conference on Artificial Neural Networks*, volume 2, pages 3–7. EC2 et Cie, 1995.

Timo Honkela, Kevin I. Hynnä, and Tarja Knuuttila. Framework for modeling partial conceptual autonomy of adaptive and communicating agents. In *Proceedings of Cognitive Science Conference*, 2003.

Gerhard Jäger. Convex meanings and evolutionary stability. In *Proceedings of 6th Int. Conf. on the Evolution of Language*, 2006.

Teuvo Kohonen. *Self-Organizing Maps*. Springer, Berlin, Heidelberg, 2001.

Teuvo Kohonen. Self-organizing formation of topologically correct feature maps. *Biological Cybernetics*, 43(1):59–69, 1982.

Ville J. Könönen. Policy gradient method for team Markov games. In *Proceedings of the Fifth International Conference on Intelligent Data Engineering and Automated Learning (IDEAL-2004)*, pages 733–739, Exeter, UK, 2004a.

Ville J. Könönen. Asymmetric multiagent reinforcement learning. *Web Intelligence and Agent Systems: An International Journal (WIAS)*, 2(2):105–121, 2004b.

Ville J. Könönen and Erkki Oja. Asymmetric multiagent reinforcement learning in pricing applications. In *Proceedings of the International Joint Conference on Neural Networks (IJCNN-2004)*, pages 1097–1102, Budapest, Hungary, 2004.

Tiina Lindh-Knuutila. Simulating the emergence of a shared conceptual system in a multi-agent environment. Master's thesis, Helsinki University of Technology, Laboratory of Computer and Information Science, 2005.

Terence Moore and Chris Carling. *The Limitations of Language*. Macmillan Press, Houndmills, 1988.

Rolf Pfeifer and Christian Scheier. *Understanding Intelligence*. MIT Press, Cambridge, MA, 1999.

Helge Ritter and Teuvo Kohonen. Self-organizing semantic maps. *Biological Cybernetics*, 61(4):241–254, 1989.

Ludwig Wittgenstein. *Philosophical Investigations*. Blackwell Publishers, Oxford, 1953.