Towards Adaptive Human—Robot Interaction

Ing. Ján Magyar

Supervisor:

Prof. Ing. Peter Sinčák, CSc.

Problem specification

- social robots are being used in various fields
- human—robot interaction has been shown to have a positive effect in a number of use cases
- most solutions offer a general robot behavior
- personalization might lead to better human performance

Contents

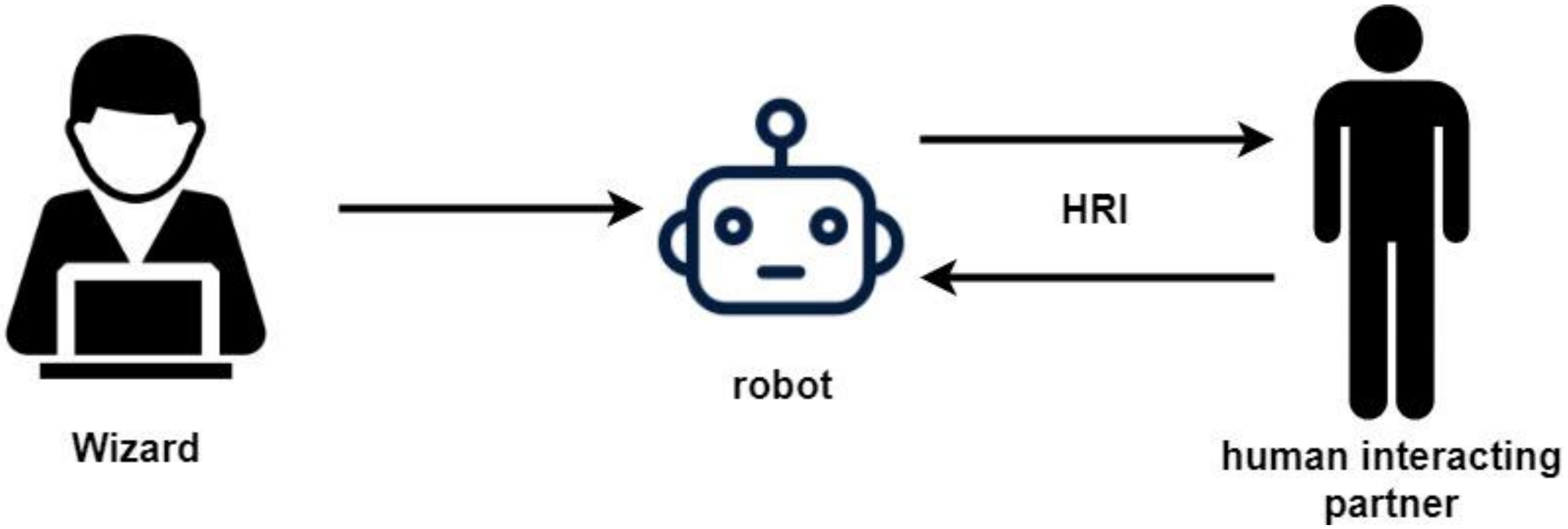
- Human—robot interaction
- Reinforcement learning
- Cloud computing
- Skill development
- Intelligent tutoring systems
- Goals of the thesis
- Testing scenario

Human—robot interaction

- the primary goal is to engage a human
- the interaction should be natural to humans
- using social robots in education
    - higher learning gains
    - increased motivation
    - more engagement

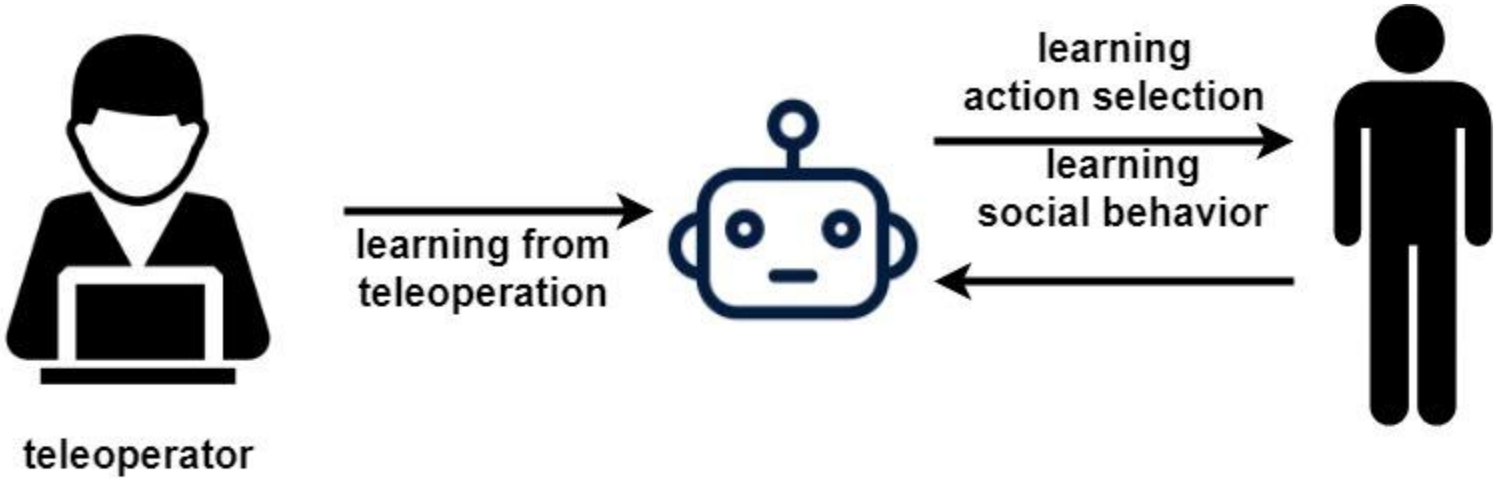
Wizard of Oz methodology

- the robot only appears to be autonomous
- robot is teleoperated by a human
- multiple levels of teleoperation are possible



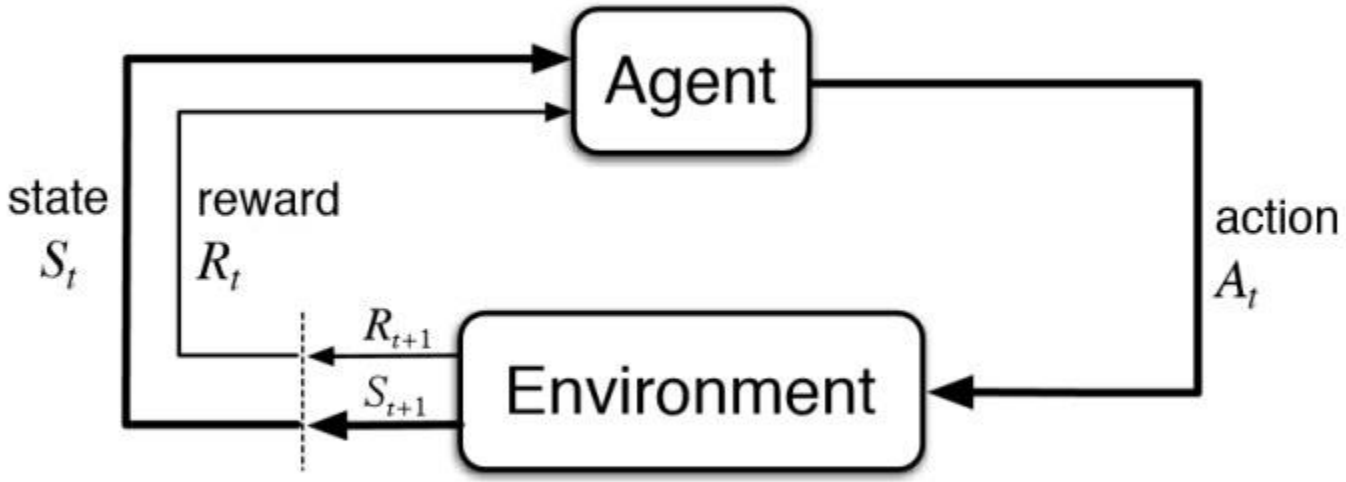
Adaptive human—robot interaction

- learning from teleoperation
- learning social behavior
- learning action selection



Reinforcement learning

- most appropriate for adapting human—robot interaction
- good support for on-line learning
- the robot is able to learn autonomously



Formal definition

- learning takes place through an agent interacting with an environment
- state space of the environment $S = \{S_1, S_2, \ldots, S_n\}$
- action space of the agent $A = \{A_1, A_2, \ldots, A_n\}$
- reward function $R = r(s_t, s_{t+1}, a_t)$
- the goal is to find an optimal policy $\pi^*\colon S \rightarrow A$

Considerations

- exploration vs. exploitation (ε)
- learning rate (α)
- discount factor (γ)
- reward function
- feature selection

Q-learning

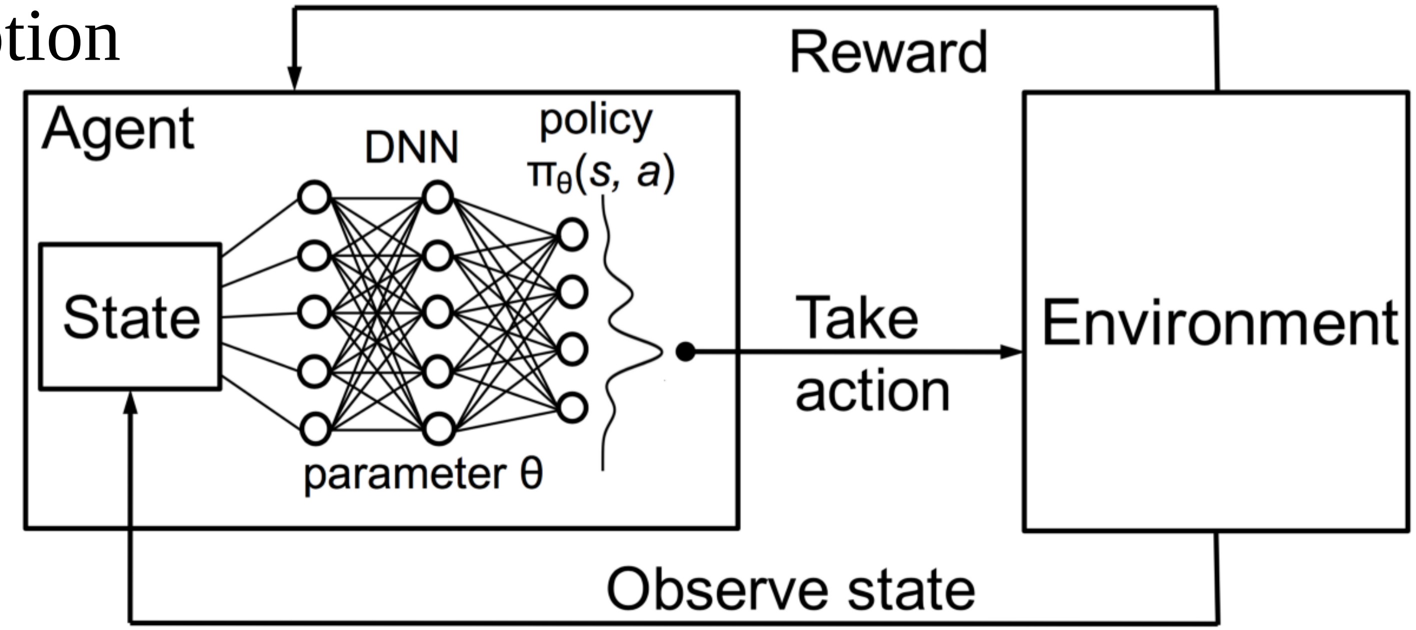
- model-free algorithm
- finds an optimal policy using the Q-function

$$Q(s_t, a_t) \leftarrow (1-\alpha) \cdot Q(s_t, a_t) + \alpha \cdot (r_t + \gamma \cdot \max_a Q(s_{t+1}, a))$$

- constructs a Q-table for Q-values
- multiple variations: SARSA, double Q-learning, etc.

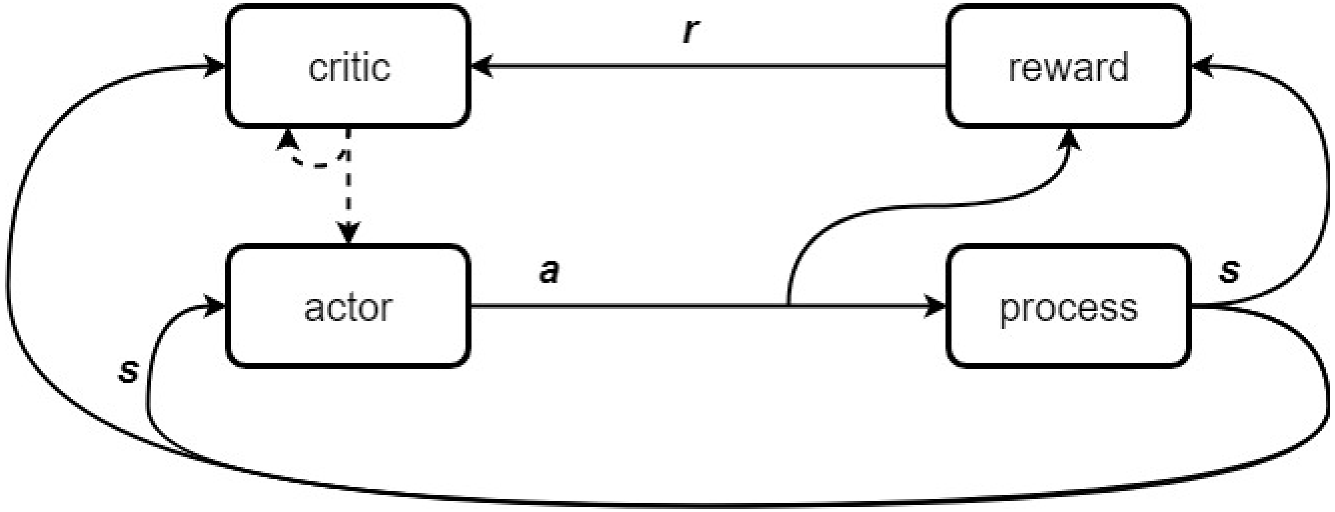
Deep Q-learning

- addresses the problem of Q-learning for huge state spaces
- represents the agent as a neural network
- input is the state description
- output is Q-values
- experience replay for balanced learning



Actor-critic methods

- combination of policy-based and value-based methods
- the actor produces actions through a policy function $\pi(s,a,\theta)$
- the critic evaluates the current policy through a value function $q(s,a,w)$

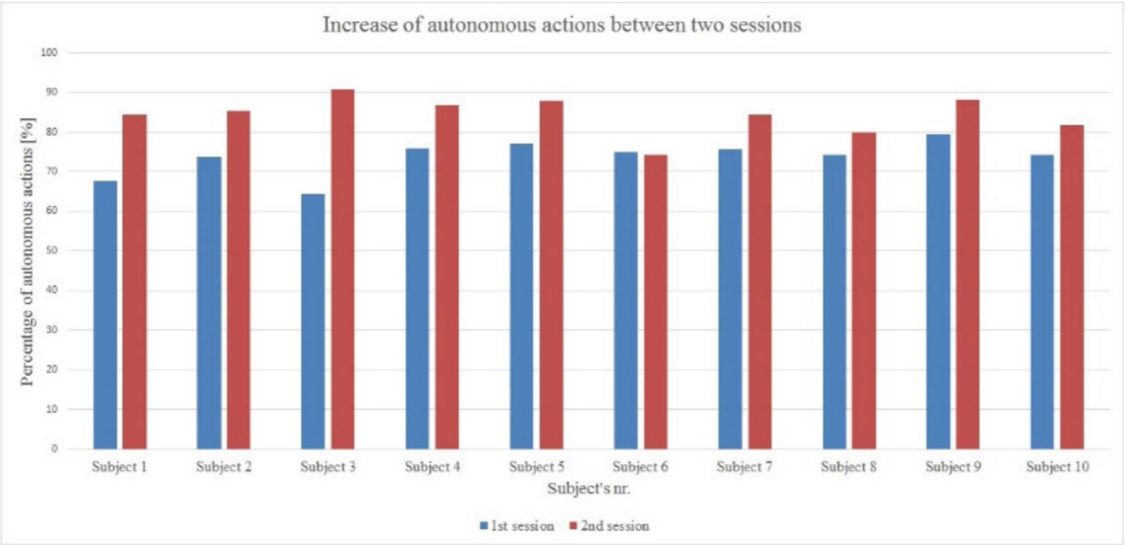


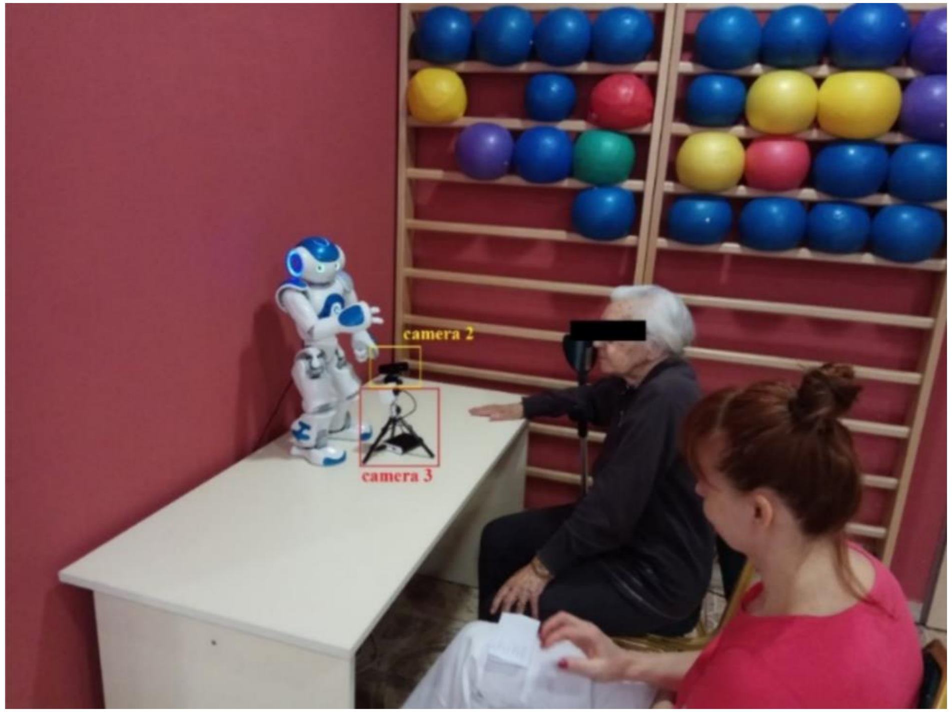
Learning from teleoperation

- the Wizard's behavior is considered to be optimal
- the robot learns to imitate the Wizard's action selection
- usually on-line learning
- human supervision still necessary
- tested and evaluated in cognitive stimulation therapy

Learning from teleoperation

- robot became more autonomous in structured interaction
- personalization was successful



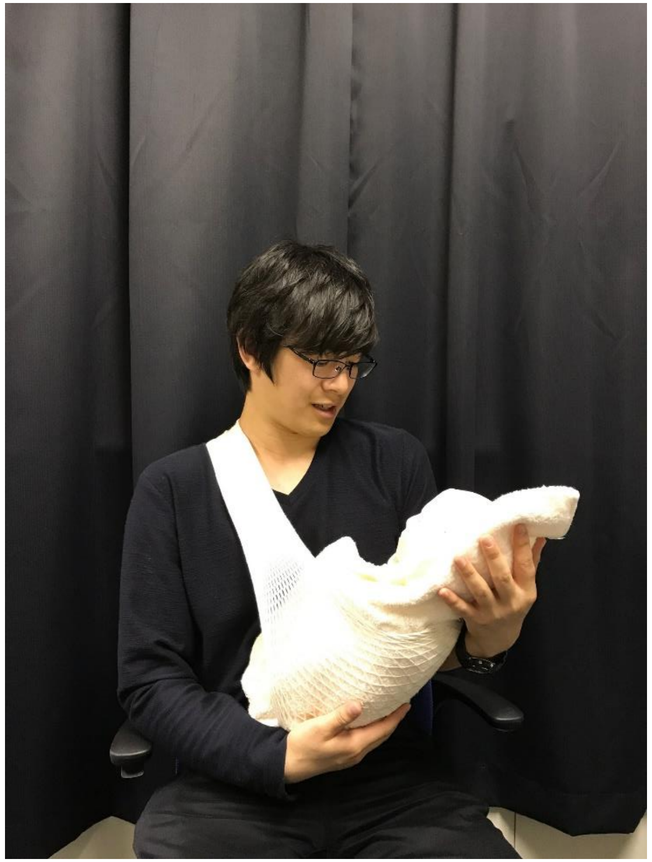


Learning social behavior

- targets social aspects of human—robot interaction
- the robot learns how to behave not what to do
- tested and evaluated in elderly care (with ATR)
    - fully teleoperated
    - semi-autonomous
    - fully autonomous

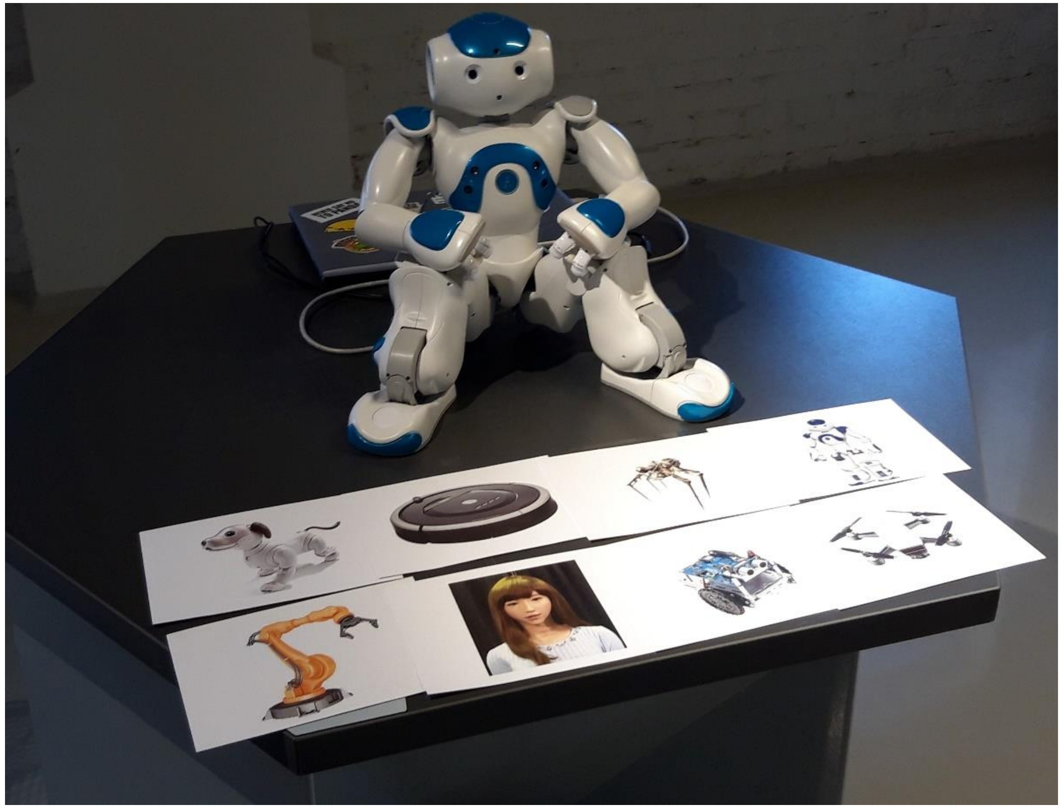
Learning social behavior

- robot was simultaneously learning from teleoperation and from the interaction
- robot did not learn the teleoperator’s behavior, but action selection was not random
- increased interaction time
- elderlies talked more
- more eye contact



Learning action selection

- the robot learns from the interaction autonomously
- robotic tutor in a museum
  - simple game
  - learning question selection
  - optimal game experience: two correct answers and an incorrect one



Learning action selection

- used information from emotion recognition to get data about players – inaccurate
- support that ideal gameplay can be accomplished through careful question selection
- learning module implemented and deployed, currently under testing and evaluation

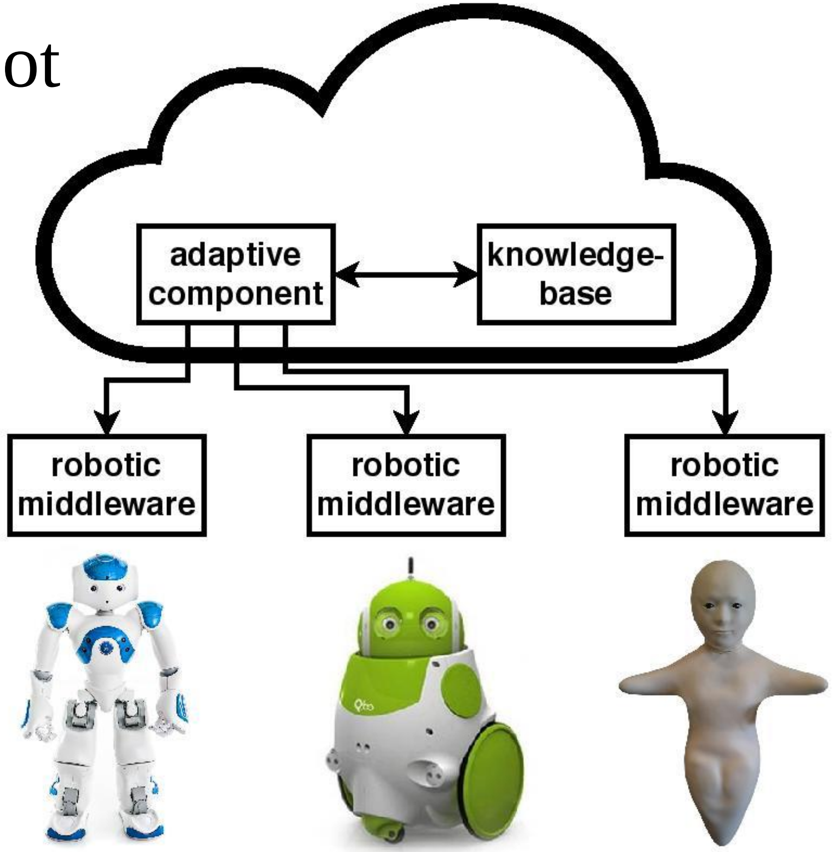
category	**correct**	**incorrect**	**no answer**	**total**
mobile robot	52	54	28	134
industrial robot	94	18	21	133
robotic dog	106	9	26	141
humanoid	68	39	18	125
android	52	34	30	116
nanobot	71	34	16	121
domestic robot	94	27	21	142
drone	135	5	11	151

Cloud computing

- ubiquitous, convenient, on-demand network access to a shared pool of configurable computing resources
- cloud robotics – functionality implemented on the cloud
- advantages for
  - offloading computation
  - collective learning
  - knowledge-base on the cloud
  - public solutions

Cloud robotics for personalization

- methods of artificial intelligence might require dedicated hardware
- adaptation mostly independent of the robot
- common knowledge-base
- experience sharing



Skill development

- knowledge acquisition vs. skill development
- stages of competence

Types of practice

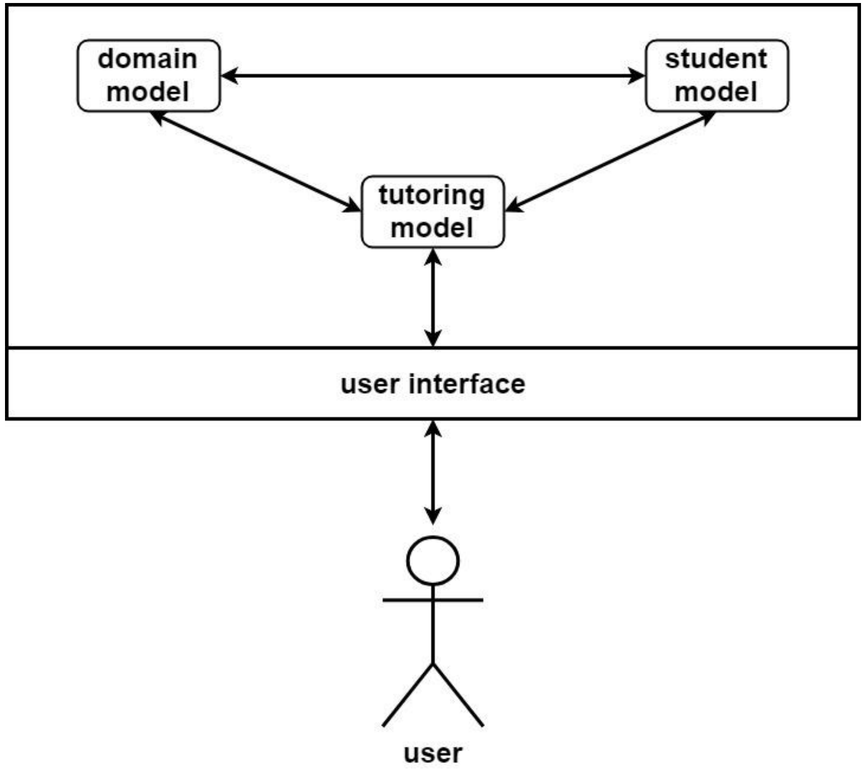
- naïve practice
  - sufficient to reach a level of competence
  - rote repetition
  - gradual degradation of the skill
- purposeful practice
  - well-defined and specific goals for each session
  - involves feedback
  - the learner performs just beyond his comfort zone

Deliberate practice

- a special type of purposeful practice
- more structured
- relies on years of gathered experience
- targets the learner’s mental representation

Intelligent tutoring systems

- emulate a learning strategy of a real learning environment
- can successfully provide personalized learning experience
- four basic components
  1. domain model
  2. student model
  3. tutoring model
  4. user interface



Goals of the thesis

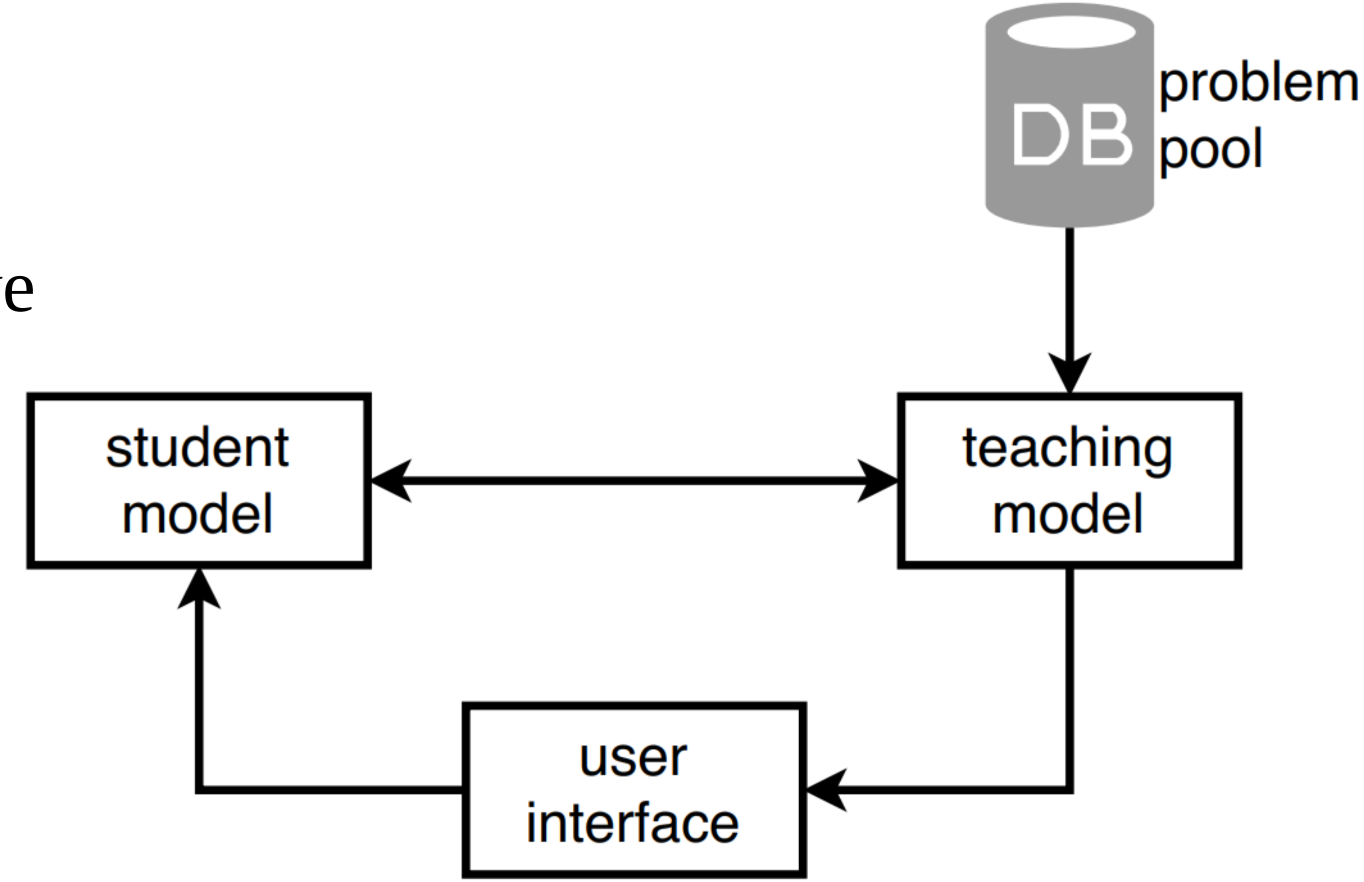
Scientific goal 1

Design an on-line reinforcement learning method tailored for the adaptation of robotic tutors for higher learning gains.

- **Measure of success**: analyze the learning gains of students using an adaptive robotic tutoring system. In an adaptive tutoring system, learning should take place at a near-constant rate without the student reaching a plateau.
- **Expected scientific contribution**: novel student modeling approach; reinforcement learning application that can be generalized for personalization of human—robot interaction

Scientific goal 1

1. student model - adaptive
2. teaching model - adaptive
3. problem pool
4. user interface



Scientific goal 2

Test and evaluate a robotic tutor that uses reinforcement learning to personalize the human learning process

- **Measure of success**: compare learning gains of students using a tutoring system with adaptive question selection with learning gains of students using a tutoring system with randomized question selection
- **Expected scientific contribution**: novel method for assessing level and success of personalization in human—robot interaction

Scientific goal 3

Conduct a long-term test of robotic tutor

- Measure of success: compare learning gains of students interacting with a robotic tutor with the learning gains of students learning through a non-embodied tutoring system. Evaluate the novelty effect of the robotic tutor in a long-term interaction.
- **Expected scientific contribution**: detailed analysis of the novelty effect and embodiment effect in human—robot interaction

Technological goal 1

Design and implement a cloud service providing reinforcement learning models

- **Measure of success**: test and evaluate the cloud service in a real-life use case.
- **Expected scientific contribution**: cloud-based implementation of reinforcement learning methods available to researchers

Technological goal 2

Design and implement a cloud-based application for personalization of human–robot interaction

- **Measure of success**: test and evaluate the web application in a real-life use case.
- **Expected scientific contribution**: create a framework for implementing virtual robots on the cloud defining a physical robot's functionality

Testing scenario

- during tests we will explore three main questions:
  - whether adaptive question selection yields higher learning gains
  - whether using a robotic tutor instead of a simple user interface has a positive effect on the learning experience
  - whether a robotic tutor’s novelty effect wears off with time
- two application scenarios
  - cognitive stimulation therapy
  - skill development

Testing in elderly care

1. records from previous therapy sessions will be used to pre-train the tutoring system
2. all participating elderlies will be informed of the goals of the experiment, their questions about the system will be answered
3. a small number of therapy sessions will be carried out
4. the sessions will be evaluated in two aspects
    - social aspect – acceptance of the robotic tutor
    - therapeutic aspect – whether the therapy was natural and successful

Testing in skill development

1. participating students will be informed of the goals of the experiment, their questions about the system will be answered
2. the students' performance at the skill will be assessed
3. skill development will take place across multiple study sessions
4. the students' skill level will be measured at the end of the learning process
5. the students' skill level will be measured after a longer period of time to measure long-term retention

Thank you for your attention!

Question 1 – Ing. Michal Gregor, PhD.

As a part of scientific goal 1, the author intends to construct a student model, which will “predict whether the student’s answer to a given question will be correct or incorrect” using reinforcement learning. Why is reinforcement learning required here? It seems that supervised learning would suffice.

Question 1a – Ing. Michal Gregor, PhD.

The student modeling problem seems to be clearly non-stationary, since the student keeps learning over time. How is this going to be addressed?

Question 1b – Ing. Michal Gregor, PhD.

Will the learning need to start from scratch for every student? Is there any knowledge that will be transferrable across students?

Question 1c – Ing. Michal Gregor, PhD.

How are the questions going to be represented? Is there going to be any generalization across questions, which would allow related questions to have similar probabilities of correct/incorrect answer?

Question 2 – Ing. Michal Gregor, PhD.

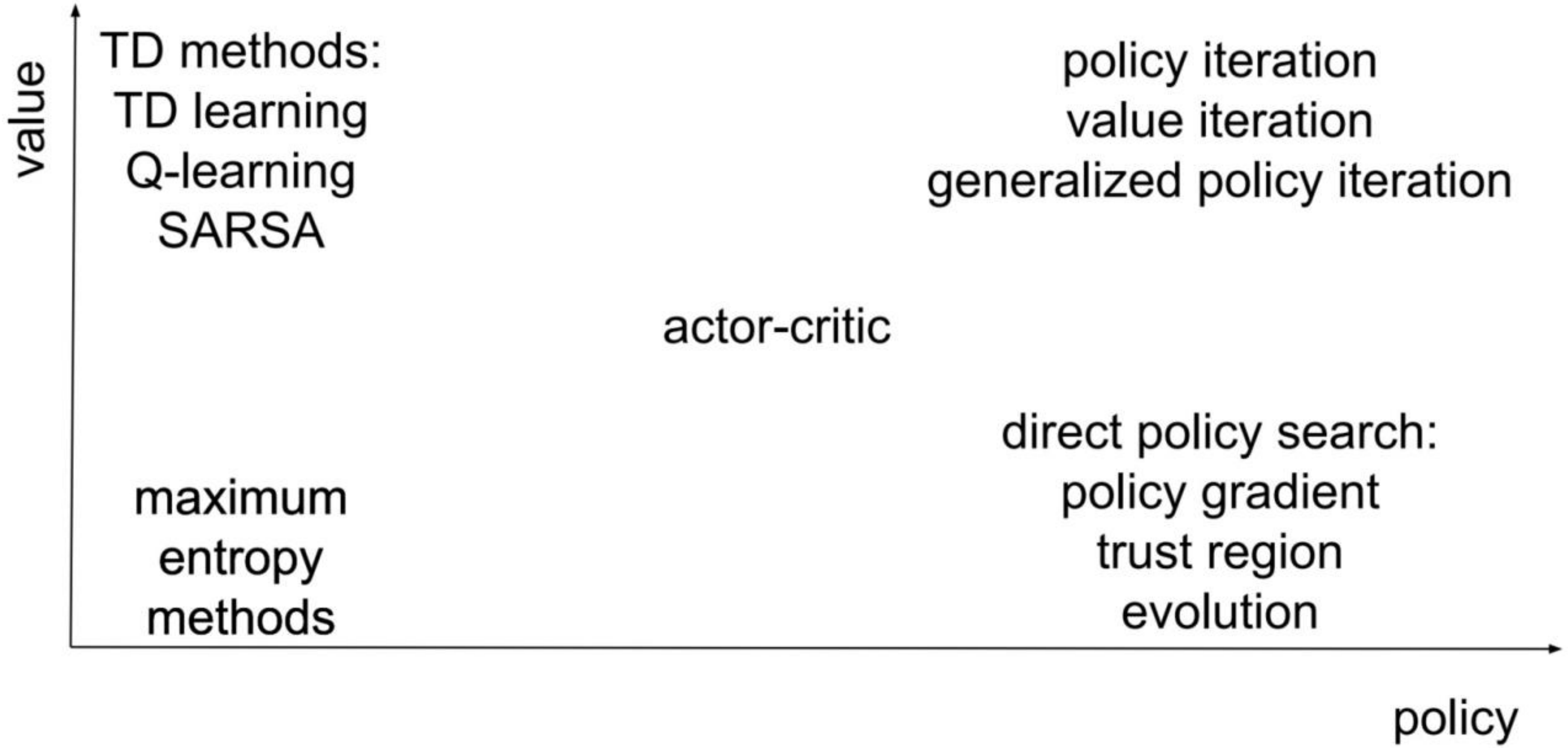
With respect to the “problem pool with corresponding levels of difficulty” (part of scientific goal 1): will the system rely on prior expert knowledge to determine the relative difficulty of questions? Or else, will it be learned solely from interaction with the student?

Question 3 – Ing. Michal Gregor, PhD.

With respect to technological goal 1: how are the benefits of “knowledge sharing and collective learning” to be realized? The description of the system seems to indicate that the experience of the RL agents will largely be user-specific.

Question 4 – Ing. Michal Gregor, PhD.

What source is Figure 3.2 from? It does not seem to come from [74] as indicated. Could the author explain the reasoning behind it? Why are actor-critic algorithms, which work with both: an explicit policy and an explicit value function in the middle of the figure and algorithms such as value iteration are in the top right corner, even though they are clearly value-based and the (implicit) policy is derived from the value function?


value
TD methods:
TD learning
Q-learning
SARSA
policy iteration
value iteration
generalized policy iteration
actor-critic
direct policy search:
policy gradient
trust region
evolution
maximum
entropy
methods
policy

Question 5 – Ing. Michal Gregor, PhD.

The author states on page 32 that “Q-values for terminal states are never updated but set for the reward observable in the given state”. Could he explain what is meant by this? It follows from the standard definition of the state- and the action-value function that the value of any terminal state should be zero, because no further rewards can be expected from that point onwards.

Question 1 – Kaori Yoshida, Dr.

I totally agree with the technical elements presented in Chapter 4. There is a description that "By deploying the adaptive model on the cloud we ensure that all physical robots display the same behavior and rely on the same version of the robotic system." in Section 4.3 (pp. 51). How many types of adaptive models do you currently expect? Also, please let us know the reason for that.

Question 2 – Kaori Yoshida, Dr.

There is a description that “In an adaptive tutoring system, learning should take place at a near-constant rate without the student reaching a plateau.” in Section 5.1.1 (pp. 53). Please let us know why you considered so. Isn’t it “learning should take place at flexibly different rate with the student reaching a plateau” in adaptive tutoring system?

Question 3 – Kaori Yoshida, Dr.

There is a description that “Evaluate the novelty effect of the robotic tutor in a long-term interaction.” in Section 5.1.3 (pp.56). How long do you expect for “long-term interaction”? For example, could you argue in comparison with the period when students get used to tablets or multimedia materials?

Question 4 – Kaori Yoshida, Dr.

Cognitive stimulation therapy in elderly care may be an on-going project, but the relationship with skill development for students is unclear. There is a description that “Although the two use cases are similar in nature, their purpose is different.” in Chapter 6 (pp. 61). The design of methods and scenarios of cognitive stimulation therapy in elderly care can be understood well. However, it’s hard to understand the design of skill development using a web application with a graphical user interface and a robot system with a robotic tutor. For example, could you explain the differences in design between cognitive stimulation therapy and skill development by comparing user and task characteristics?