Seminar: Advanced topics in reinforcement learning and planning

Hector Geffner

Computer Science 6 (Machine Learning and Reasoning)
RWTH Aachen University, Summer 2023

https://www-i6.informatik.rwth-aachen.de/∼hector.geffner

Slides revised from Thomas Noll's (i2)

Outline

- About me
- AI and ML
- Aims of the seminar
- Evaluation
- Important Dates
- Themes; papers
- Final Hints

About me

- I'm new at RWTH; started in January 2023
- Chair of Machine Learning and Reasoning (i6)
- Alexander von Humboldt Professor
- Originally from Argentina; worked in Spain, USA, Venezuela
- I did my PhD at UCLA on AI (Diss. on logic, probab, causality)
- Moved then to **planning**, **generalized planning**, **(Deep) RL**
- ERC Adv. Grant (2020-2025): Representing learning for acting and planning
- Teaching also *Actions and Planning in AI: Learning, Models, Algorithms*

State of AI Research: Learners

$$\textit{Input } x \Longrightarrow \boxed{\textsc{Function } f} \Longrightarrow \textit{Output } f(x)$$

- In **deep learning (DL)** and **deep reinforcement learning (DRL)**, training results in function f_θ
- f_θ given by structure of **neural network** and adjustable parameters θ
  - ▷ In DL, **input** x may be an image and **output** $f_\theta(x)$ a classification label
  - ▷ In DRL, **input** x may be state of game, and **output** $f_\theta(x)$, value of state
- Parameters θ learned by **minimizing error function** by stoch. gradient descent
  - ▷ In DL, error depends on inputs and target outputs in training set
  - ▷ In DRL, error depends on value of states and successor states
- A true revolution in AI still **unfolding**
- **Limitations:** transparency, amounts of data, OOD generalization, understanding

State of AI Research: Solvers (Reasoners)

$$\textit{Input } x \Longrightarrow \boxed{\textsc{Function } f} \Longrightarrow \textit{Output } f(x)$$

- **Solvers** derive output $f(x)$ for **given input** x from **model**:
  - ▷ **SAT:** x is a formula in CNF, $f(x) = 1$ if x satisfiable, else $f(x) = 0$
  - ▷ **Classical planner:** x is a planning problem P, and $f(x)$ is plan that solves P
  - ▷ **Bayesian net:** x is a query over Bayes Net and $f(x)$ is the answer
  - ▷ **Constraint satisfaction, Markov decision processes, POMDPs, . . .**

- **Generality:** Solvers not tailored to particular examples
- **Expressivity:** Some models very expressive; e.g., POMDPs
- **Learners are solvers too:** $\text{argmin}_w \sum_{x \in D} L(x, f_w(x))$ (Diff. programming)
- **Challenge:** Scalability; computation of $f(x)$ is NP-hard
- **Limitation:** models must be known

Learners and Solvers: System 1 and System 2?

Dual process accounts of the human mind assume two processes (D. Kahneman: Thinking, Fast and Slow, 2011; K. Stanovich: The Robot's Rebellion, 2005)

System 1 (Intuitive Mind)	**System 2** (Analytical Mind)
fast	slow
associative	deliberative
unconscious	conscious
effortless	effortful
parallel	serial
specialized	general
...	...
Learners?	Solvers?

Key Challenge in AI

- General **two-way integration** of System 1 and System 2 inference in AI systems
  - ▷ *Learn representations that support reasoning and are general/reusable*

- **Yoshua Bengio**'s challenges reflected in title of his IJCAI 2021 talk:
  - ▷ *System 2 Deep Learning: Higher-level cognition, agency, out-of-distribution generalization and causality*

- **Yann LeCun**'s three challenges, AAAI 2020:
  - ▷ AI must learn to represent the world
  - ▷ AI must think and plan in ways compatible with gradient-based learning
  - ▷ AI must learn hierarchical representation of action plans

Aims of the seminar

- **Critical and independent understanding** of a scientific topic
  - ▷ Acquiring, reading, and understanding scientific literature
  - ▷ May involve **reading** other papers, and carry own **experiments**
  - ▷ Technical content is key; we don't want generalities or blah blah
- **Writing** of your own report
  - ▷ far more than "rewording" one or more papers;
  - ▷ **"deep learning"** is good; but **deep understanding** is better
- **Oral presentation** of your understanding, results
  - ▷ be a teacher; teach us, enlighten us (!)

Requirements on Report

- Independent writing of a report of 12–15 pages
  - ▷ Be crisp, clear, and to the point; depth and substance
  - ▷ Make justice to the technical content of the papers
  - ▷ Convey this in written and oral communication
- Complete set of references to consulted literature
- Correct citation of important, relevant work
- **Plagiarism**: taking text blocks without sources: exclusion from seminar
- Font size 12pt with standard LaTeX layout
- Language: English; correct usage of spelling and grammar expected

Requirements on Talk

- Talk of 30 minutes
- Available: projector, presenter, [laptop]
- Focus your talk on the audience, teach us
- Descriptive slides:
  - a few lines lines of text (≤ 10)
  - tell a story
  - convey substance (technical content)
  - structure is important
  - be crisp and clear
- Language: English
- Avoid spelling mistakes; finish in time
- Ask and answer questions

RWTH AACHEN UNIVERSITY

Seminar
–Evaluation Sheet–

General Information

This document provides various criteria used to assess the performance of seminar participants. There are four milestones, namely

- detailed outline,
- full report,
- presentation slides, and
- seminar talk.

Each milestone has its own set of requirements which are evaluated during the semester—allowing students to get intermediate feedback on their performance. A certain amount of points is granted for each fulfilled requirement. If all requirements are fulfilled (i.e. not **failed**), the final grade for the course is given by the sum of the collected points using the following table:

Points:	≥ 30	≥ 33	≥ 36	≥ 39	≥ 42	≥ 45	≥ 48	≥ 51	≥ 54	≥ 57
Grade:	4.0	3.7	3.3	3.0	2.7	2.3	2.0	1.7	1.3	1.0

In exceptional cases, we reserve the right to deviate from this grading scheme (e.g. bonus points for going above and beyond). Moreover, **meeting the deadlines is a must!** If the deadline is missed by no more than one day, 3 points are deducted. Missing a deadline by more than one day (i.e. more than 24 hours) will lead to the seminar to be considered "not passed". In very exceptional cases (such as being ill for a longer period), the deadline may be extended, please contact us proactively!

Hint 1: Look at our website for additional material on how to set up good reports/presentations.

Hint 2: Communicate potential issues with your supervisor as soon as possible.

Student Name: _____

Detailed Outline — 10 Points

The goal for this milestone is to provide an overview of the overall structure (section headers, main definitions/theorems/examples) as well as an initial part of one of the main sections of the report. When the detailed outline is due, students are expected to be familiar with their topic and the associated literature.

Deadline: Please refer to our website.

Overview of Structure (4 Points) Score: _____

The future contents of the individual (sub-)sections should become clear to the supervisor. If necessary, short bullet points can be added for clarification.

Initial Content (3 Points) Score: _____

There should be at least one page of actual content for one of the main sections.

Language (3 Points) Score: _____

The report can be in German or English. We expect proper scientific writing and the correct usage of spelling and grammar.

Student Name: _____

Full Report — 20 Points

By the deadline for this milestone, the full report should be finished and proofread multiple times.

Deadline for the version we give feedback on: Please refer to our website.

Deadline for the final version: Please refer to our website.

Mathematical Notation (2 Points) Score: _____

Mathematical notation should be used wherever necessary to provide clear, well-defined descriptions. It should be consistent both within the report and with other literature.

Document Layout / Typesetting (2 Points) Score: _____

Definitions and equations must be readable, using suitable LaTeX environments. Bullet points must be appropriately used, cross-references to other parts of the report should be correct and used in a reasonable manner. Figures should not distract from other content.

Understandability (4 Points) Score: _____

The content of the source material should be explained in the report in such a way that no prior knowledge of the source material is necessary to understand the report. Non-trivial prerequisites that are not explained in the report should be clearly stated in the report and communicated to the supervisor in advance.

Examples (2 Points) Score: _____

Examples should be used to demonstrate the general theories. If possible, a single running example is better than many individual ones and a custom example is better than one copied from the source material.

Amount of Content (2 Points) Score: _____

The report should have 12–15 pages. Significant deviations (e.g. an additional appendix with figures) should be discussed with the supervisor.

Language (3 Points) Score: _____

The report can be in German or English. We expect proper scientific writing and the correct usage of spelling and grammar.

Student Name: _____

Presentation Slides — 10 Points

In this milestone the presentation slides are due. You may use a different language (German or English) than the report. It is also fine to e.g. present in German with English slides.

Deadline: Please refer to our website.

Structure (2 Points) Score: _____

There should be a clear thread running through the presentation. The different sections should build upon each other, leading to a clear conclusion. Relevant definitions are given before their usage, but are introduced as late as possible to avoid mental overload.

Layout of Slides (4 Points) Score: _____

Slides should not be used to convey all the information, but to support the oral presentation. They are not a script, so they should be simpler than what is used during lectures.

Level of Detail (4 Points) Score: _____

Explanations can be example-driven. The level of detail should be adapted to the audience (i.e. other participating students).

Student Name: _____

Seminar Talk — 20 Points

The seminar talk is the last milestone in which the acquired knowledge on the topic should be presented to the audience.

Timing (2 Points) Score: _____
The time limit should not be exceeded. Students should manage their time by themselves.

Preparation (2 Points) Score: _____
The speaker should be prepared for giving the presentation. They should know the content and order of their slides.

Spoken Language / Style of Presentation (2 Points) Score: _____
The audience should understand what the speaker is saying on a language level. The speaker should speak freely and convey a certain confidence in the presentation.

Understandability (6 Points) Score: _____
It should be possible for the audience to follow the provided explanations without prior knowledge of the report or source material.

Questions and Answers (8 Points) Score: _____
Questions may be asked by other students and/or by attending supervisors during or after the talk (if they are asked during the presentation, the time limit is extended accordingly). The questions should be answered appropriately and in an understandable manner.
Hint 1: Some questions might be outside the scope of the topic. Sometimes an appropriate answer to those questions is a simple "I don't know"—followed by an educated guess.
Hint 2: Questions are usually asked in good faith out of interest in your topic, i.e. they should not be confused with an oral examination. In particular, students should not hesitate from asking questions to other students. The affect of student questions on the overall grade is negligible.

Important Dates: Deadlines

- April 14: Topic/paper preferences due
- May 15 Detailed outline due (2 pages, pdf)
- June 16: Full report due
- July 12: Presentation slides due
- July 18–19: Seminar talks

Missing a deadline causes exclusion from the seminar

Send **pdf**s by mail to me with corresponding **Subject** line

Selecting Your Topic: Procedure

- Indicate your preferences (first, second, third) in email to me by **April 14th**
- I'll do my best to find an adequate topic-student assignment.
- Assignment will be published on web site before **April 21st**
- Language for report, oral presentation, and communication is **English**

Withdrawal:

- You have up to three weeks to refrain from participating in the seminar.
- Later cancellation (by you or by us) causes a not passed

The topics

- Reinforcement learning/Planning
- Model-based RL/General RL
- Deep learning: Transformers
- Large-Language Models (RL, Planning)
- Other

Tentative, partial list of themes/papers (1/2)

- **RL/Planning**
  - *Playing Atari with Deep Reinforcement Learning*. V. Mnih, K. Kavukcuoglu, D. Silver, et al. 2013
  - *Mastering Chess and Shogi by Self-Play with a General Reinforcement Learning Algorithm*. D. Silver, T. Hubert, J. Schrittwieser, et al. 2017
  - *Mastering Atari, Go, Chess and Shogi by Planning with a Learned Model*. J. Schrittwieser, I. Antonoglou, T. Hubert, et al. 2020
  - *A Generalist Agent* S. Reed, K. Zolna, E. Parisotto, et al. 11/2022
  - *Transformers are Sample-Efficient World Models* V. Micheli, E. Alonso, F. Fleuret. 3/2023

Tentative, partial list of themes/papers (2/2)

- **Deep Learning: Transformers**
  - *End-to-end symbolic regression with transformers*. P. Kamienny, S. d'Ascoli, G. Lample, F. Charton. 2022
  - *Linear algebra with transformers*. F. Charton. 2021
  - *What Can Transformers Learn In-Context? A Case Study of Simple Function Classes*. S. Garg, D. Tsipras, P. Liang, G. Valiant. 1/2023
- **LLMs (RL, Planning)**
  - *Training language models to follow instructions with human feedback (ChatGPT)* L. Ouyang, J. Wu, X. Jiang, D. Almeida, et al. 3/2022
  - *ReAct: Synergizing Reasoning and Acting in Language Models*. S. Yao, J. Zhao, D. Yu, et al. 3/2023
  - *Guiding Pretraining in Reinforcement Learning with Large Language Models*. Yuqing Du, Olivia Watkins, Zihan Wang, Cédric Colas, Trevor Darrell, Pieter Abbeel, Abhishek Gupta, Jacob Andreas. 2/2023
  - *On the Planning Abilities of Large Language Models (A Critical Investigation with a Proposed Benchmark)* K. Valmeekam, S. Sreedharan, M. Marquez, A. Olmo, S. Kambhampati. 2/2023

Final Advice

- Take your time to process, read, understand the literature, try experiments, etc.
- It's your work but: discuss with your fellow students, talk to me, etc.
- Be crisp and clear in your understanding, your report, and your presen
- Be proactive and don't let the course evaluation spoil you the fun of learning
- Excellent opportunity for learning on your own, with some help from me

Look forward to an enjoyable and productive seminar!