Reinforcement Learning: Introduction and Overview

Winter 2017/18

Stefan Riezler

Computational Linguistics & IWR
Heidelberg University, Germany
riezler@cl.uni-heidelberg.de

Organization of Class

- Tuesday, 11:15-12:45
- Schedule and reading list will be posted and updated on `http://www.cl.uni-heidelberg.de/courses/ws17/reinforcement/`

Assessment

- Paper presentation
  - Sign up for a session from schedule: 2 preferences, by email to `sekretariat@cl.uni-heidelberg.de`, subject: REINFORCE COURSE, until Nov. 21
  - Presenter: Read papers, send two technical questions to others one week ahead, slide presentation in session
  - All others: Answer two technical questions per email
- Term paper
  - Implementation project and its description
  - or in-depth discussion of theoretical questions

Textbooks

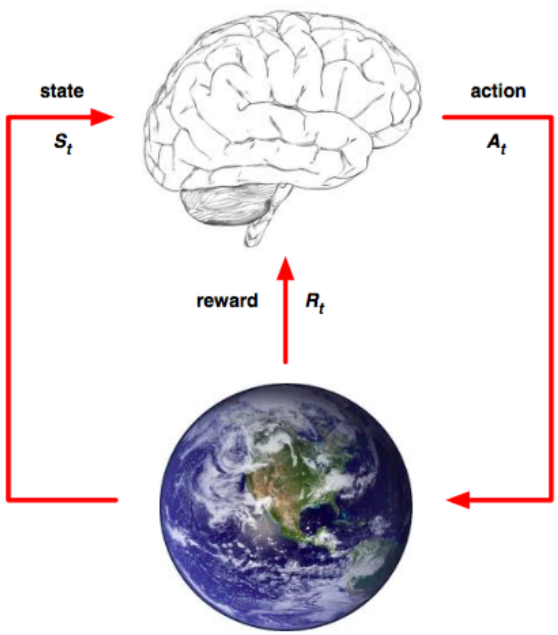
- Richard S. Sutton and Andrew G. Barto (2017, 2nd edition, in progress): Reinforcement Learning: An Introduction. MIT Press.
  - `http://incompleteideas.net/sutton/book/the-book-2nd.html`
- Csaba Szepesvári (2010). Algorithms for Reinforcement Learning. Morgan & Claypool.
  - `https://sites.ualberta.ca/~szepesva/RLBook.html`
- Dimitri Bertsekas and John Tsitsiklis (1996). Neuro-Dynamic Programming. Athena Scientific.
  - = another name for deep reinforcement learning, contains all proofs, analog version can be ordered at `http://www.athenasc.com/ndpbook.html`

Reinforcement Learning (RL) Philosopy

- *Hedoninistic* learning system that *wants* something, and adapts its behavior in order to maximize a special signal or *reward* from its environment.
- *Interactive* learning by trial and error, using consequences of own actions in uncharted territory to learn to maximize expected reward.
- *Weak supervision signal* since no gold standard examples from expert are available.

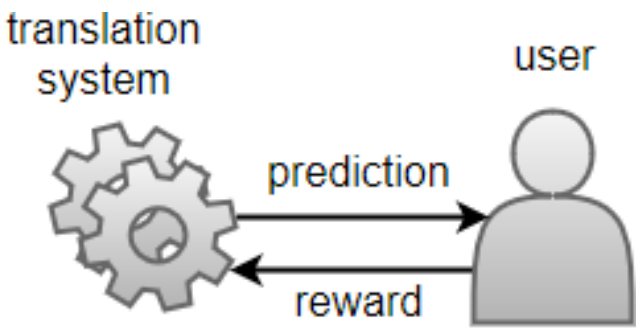
Reinforcement Learning Schema

- RL as Google DeepMind would like to see it (image from David Silver):



Reinforcement Learning Schema

- A real-world example: Interactive Machine Translation



  - action = predicting a target word
  - reward = per-sentence translation quality
  - state = source sentence and target history

Reinforcement Learning Schema

Agent/system and environment/user interact

- at each of a sequence of time steps $t = 0, 1, 2, \ldots$,
- where agent receives a state representation S_t,
- on which basis it selects an action A_t,
- and as a consequence, it receives a reward R_{t+1},
- and finds itself in a new state S_{t+1}.

Goal of RL: Maximize the total amount of reward an agent receives in such interactions in the long run.