CI Seminar A
WS 2017 / 18

How to prepare and hold your talk

Robert Legenstein, robert.legenstein@igi.tugraz.at

Institute for Theoretical Computer Science
Graz University of Technology

Overview

1. How to read your paper
2. How to plan your talk
3. Talk structure
4. Mind the gap
5. Good practice for slide design
6. At the talk

1. How to read your paper (1)

- Understand the paper completely
  - Do not present something you don’t understand

- When you encounter a concept you do not know
  - Check whether a reference is cited. If yes, you may want to consult the cited paper.
  - Search the web for it
  - You can also search for papers using keywords in Google Scholar
  - If you cannot find useful information, send me an email describing your problem.

1. How to read your paper (2)

- Understand the paper completely
  - Do not present something you don’t understand

- When you do not understand some part of the paper
  - Is it because they use a concept you don’t understand? → see above
  - Do they use some special notation that is unfamiliar to you?
  - Send me an email describing your problem.

1. How to read your paper (example)

- When you encounter a concept you do not know

RL^2: FAST REINFORCEMENT LEARNING VIA SLOW REINFORCEMENT LEARNING

Yan Duan[†‡], John Schulman[†‡], Xi Chen[†‡], Peter L. Bartlett[†], Ilya Sutskever[‡], Pieter Abbeel[†‡]
[†] UC Berkeley, Department of Electrical Engineering and Computer Science
[‡] OpenAI

2.3 POLICY REPRESENTATION

We represent the policy as a general recurrent neural network. Each timestep, it receives the tuple (s, a, r, d) as input, which is embedded using a function $\phi(s, a, r, d)$ and provided as input to an RNN. To alleviate the difficulty of training RNNs due to vanishing and exploding gradients (Bengio et al., 1994), we use Gated Recurrent Units (GRUs) (Cho et al., 2014) which have been demonstrated to have good empirical performance (Chung et al., 2014; Józefowicz et al., 2015). The output of the GRU is fed to a fully connected layer followed by a softmax function, which forms the distribution over actions.

2. How to plan your talk (1)

- What are the main messages of the paper?
- Convey these messages and any support for the claims.
- Select the material to present.
  - You do not have to present everything in the paper.
  - Better concentrate on the most important points than to rush superficially over everything.
- Make sure you stay comfortable in time
  - Maybe keep a small margin.
  - This will help to stay calm during the talk.
- Practice your talk and measure the time you need.

2. How to plan your talk (2)

- Concepts / Models not explained in the paper
  - You have to decide whether you want to include it.

  a) If it is just a detail
     - you may ignore it, maybe mention it and give a citation.

  b) If it is important,
     - You should explain it briefly
     - Concentrate on parts important for your paper

2. How to plan your talk (2) (Example 1)

- Concepts / Models not explained in the paper
  - You have to decide whether you want to include it.

RL^2: FAST REINFORCEMENT LEARNING VIA SLOW REINFORCEMENT LEARNING

Yan Duan[†‡], John Schulman[†‡], Xi Chen[†‡], Peter L. Bartlett[†], Ilya Sutskever[‡], Pieter Abbeel[†‡]
[†] UC Berkeley, Department of Electrical Engineering and Computer Science
[‡] OpenAI

Gated Recurrent Units (GRUs):

- Don’t need to be included to understand the main points
- But it should be clear that they are recurrent networks, similar to LSTMs
- However, this may evoke questions: How are they different, etc?
- Therefore it might be good to sketch them briefly.

2. How to plan your talk (2) (Example 2)

- Concepts / Models not explained in the paper
  - You have to decide whether you want to include it.

RL^2: FAST REINFORCEMENT LEARNING VIA SLOW REINFORCEMENT LEARNING

2.4 POLICY OPTIMIZATION

After formulating the task as a reinforcement learning problem, we can readily use standard off-the-shelf RL algorithms to optimize the policy. We use a first-order implementation of Trust Region Policy Optimization (TRPO) (Schulman et al., 2015), because of its excellent empirical performance, and because it does not require excessive hyperparameter tuning. For more details, we refer the reader to the original paper. To reduce variance in the stochastic gradient estimation, we use a

Trust Region Policy Optimization (TPRO)

- Rather complex method
- Mention it
- Describe what it does (not how) and its properties.

2. How to plan your talk (3)

- Math. derivations / proofs
  - Decide whether you want to present it
    - Is it important? Does it convey some insight or is it rather standard?
  - Proofs / derivations should be explained on an intuitive level (what is the basic idea of the proof?)
  - It is not a good idea to only write down the formulas.

3. Talk structure (1)

- The talk should be well structure, so that the audience knows
  - where they are currently,
  - how the current content is related to the material as a whole, and
  - where the journey is heading.
- A good talk tells a story while conveying scientific content.

3. Talk structure (2)

These are just suggestions, you can be creative.

- Usually you start with an introduction:
  - What is the problem that is tackled?
  - What are the preliminaries, prior approaches?
  - Basic idea behind the solution or the basic message.
- The structure of the main part depends more on the content.
- You end with conclusions.
  - Repeat the main messages and
  - bring them in context to the contemporary research in the field.

4. Mind the gap

- The listeners often have a hard time in talks. Be aware that:
  - You know the content very well, the listeners may hear it for the first time.
  - Think about your talk from the viewpoint of the audience!
  - The ability to absorb information and bring it into context is limited → speakers usually overestimate
  - Try to explain the content clearly and intuitively.
- Note that the audience always interprets what you are saying.
  - Some interpretations may be wrong without anyone noticing.
  - Later, such listeners may get lost when they realize that their interpretation is inconsistent.

5. Good practice for slide design (1)

- Do not pack too much content on a slide.
- Formulas: define the symbols on the slide.
- Citations:
  - do not use numbers such as [1], the audience cannot browse to the bibliography
  - put a brief citation directly on the slide, such as [Legenstein et al., J Nsci 2011]

5. Good practice for slide design (2)

- Formulas
  - Explain them in an understandable / intuitive manner
  - Sometimes, meaning can be assigned to parts of a formula.
    - Indicate that on the slide with sub brackets / color.
  - Avoid unnecessary formulas
  - Don’t put unexplained formulas on the slides
  - If you introduce some formula / definition on one slide, and need it on a later slide,
    - it is often a good idea to repeat it (maybe somewhat smaller) as a reminder.

5. Good practice for slide design (Example)

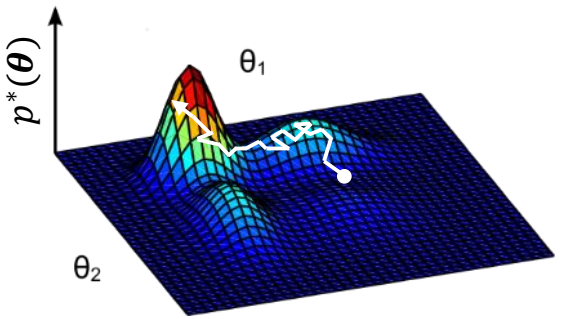
- Formulas

Consider a network with parameters $\boldsymbol{\theta}$ and the following parameter dynamics:

$$d\theta_i = b\left(\frac{\partial}{\partial \theta_i}\log p^*(\boldsymbol{\theta})\right)dt + \sqrt{2bT}\, d\mathcal{W}_i$$

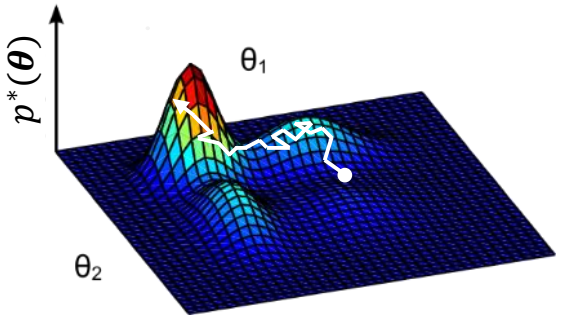
Noise

b……...sampling speed (learning rate)
$p^*(\boldsymbol{\theta})$…target parameter distribution
$d\mathcal{W}_i$…..increments of a standard Wiener process
T……..”temperature” controls the amount of noise



5. Good practice for slide design (Example)

- Formulas



Hence, the parameter dynamics are:

$$d\theta_i = \begin{cases} \sqrt{2bT}\, d\mathcal{W}_i & \text{for an inactive connection } (\theta_i \leq 0) \\ b\left(-\alpha - \dfrac{\partial}{\partial \theta_i} E_{\boldsymbol{\theta}}(\boldsymbol{Y}^* | \boldsymbol{X})\right) dt + \sqrt{2bT}\, d\mathcal{W}_i & \text{for an active connection } (\theta_i > 0) \end{cases}$$

L1 norm regularization

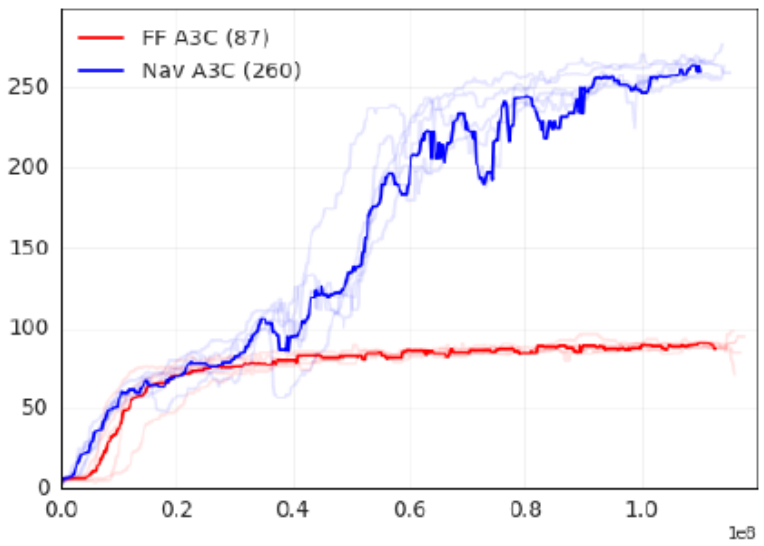
error backpropagation

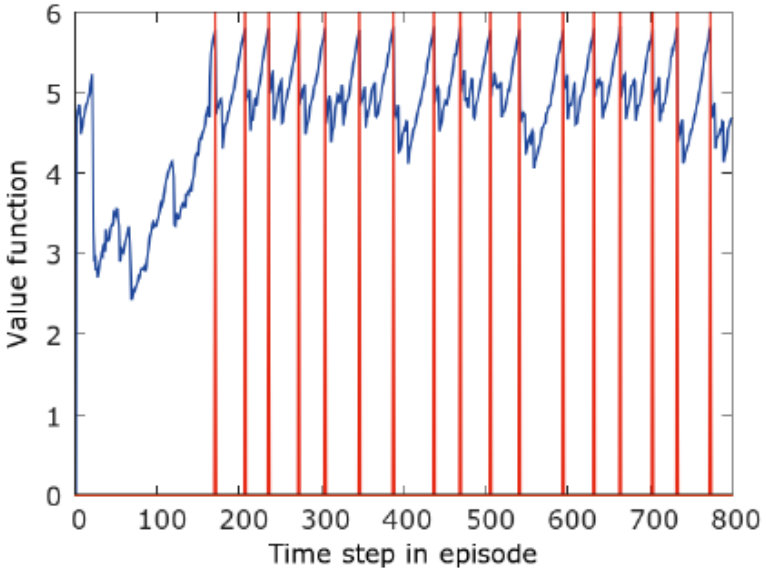
gradient noise

5. Good practice for slide design (3)

- Figures
  - Help to explain the content or show results.
  - Make sure that the quality is OK and everything (axis labels etc.) is well legible.
  - Make sure that it is clear what is shown and what the axes mean
    - I.e., what value is plotted on the y axis in dependence of what on the x axis.
    - Note: this sounds trivial, but it is surprisingly often unclear.

5. Good practice for slide design (Example)





Bad figure

- no axis labels
- text on bottom right can hardly be read.
- Legend on top left is cryptic

OK figure

- But it should be clear what “Value function” means here –
- and what the red vertical lines are.

6. At the talk

- Usually, you stand during the talk.
  - You can also walk slowly (makes it more dynamic).
- If you don’t know what to do with your hands,
  - it is often a good idea to hold something like a laser pointer.
- A laser pointer is very useful to point at the currently relevant parts of the slide.
- Do not read your sentences from paper or the slides.
  - The slides can guide you
    - so you do not get stuck.
  - But don’t stare at your slides all the time.
  - It is optimal to generate your sentences online.
    - Then the talk gives a very natural impression.