Neural Monkey
March 1, 2017

Jindřich Libovický
What is Neural Monkey

- tool for sequence-to-sequence learning developed at ÚFAL
- used for neural machine translation, MT postediting, image captioning, multi-modal MT, OCR
- written in Python using TensorFlow, hosted on GitHub

...all are technologies worth knowing
Neural Monkey Overview
Why Contribute?

▶ neural MT has state-of-the-art results
▶ contribution to living internationally used project
▶ separated experiment design and code, highly modular – contribution can focus only to small parts
▶ you can always climb to the 4th floor, knock the door and ask the authors if something is unclear
Factored Translation

- factored input (forms + POS tags / lemmas)
  - what POS granularity is the best (just POS, full positional tags?)
  - rare input words are often split into subword units – how to encode the tags (repeat them? BIO encoding?)

- predict POS on input as regularization – another objective that should bring more information into the encoder

- factored output
  - should we just predict the POS tags on the output or feed them recursively to the decoder?
  - the same question with encoding as with factored input
Visualization

...visualization is the best way of get intuition about the models

- **Debugging web app for trained models**: use Neural Monkey server to visualize attention, fertility, entropy of attention distributions
- visualization to understand training
  - generate video of attention distribution during training
  - interactive visualization of space of word embeddings
- more “scientific” visualization (e.g., use GRU gates to visualize long-distance relations in target sentences)
Improve Neural Monkey LogBook

- visualize higher-level organization of the model (probably using GraphViz)
- bootstrap resampling for estimating confidence intervals and plot comparison graphs with the intervals
- interactive configuration syntax-highlighting with links to GitHub
Reinforcement Learning

- MIXER algorithm
- Self-Critical training
- REINFORCE as used in Google paper
- your own version of REINFORCE algorithm

Warning: challenging and advanced
Technical Contribution to Neural Monkey

- efficient beam search implementation
- saving and loading parts of models (better than now)
- asynchronous validation
- unit-testing network functionality
- and many many more …
If you are interested in neural MT, attend **NPFL116**
*Compendium of Neural Machine Translation*

room S6, Wednesdays, 14:00