Neural Monkey:  
The Current State and Beyond

Jindřich Helcl, Jindřich Libovický, Tom Kocmi,  
Tomáš Musil, Ondřej Cífka, Dušan Variš, Ondřej Bojar

Institute of Formal and Applied Linguistics  
Faculty of Mathematics and Physics  
Charles University

March 19, 2018
Neural Monkey

- Open-source toolkit for sequential learning
- Suited for research and education
- Three (overlapping) user groups considered:
  - Students
  - Researchers
  - Newcomers to deep learning
Goals

1. Code readability
Goals

1. Code readability
2. Modularity along research concepts
Goals

1. Code readability
2. Modularity along research concepts
3. Up-to-date building blocks
Goals

1. Code readability
2. Modularity along research concepts
3. Up-to-date building blocks
4. Fast prototyping
Development

- Implemented in Python 3.5 using TensorFlow
- thanks to TensorFlow GPU support using CUDA, cuDNN
- Actively developed using GitHub as the main communication platform

Source code here:

https://github.com/ufal/neuralmonkey
• Multimodal translation  
  (Charles University, ACL 2017)
• Bandit learning  
  (Heidelberg University, ACL 2017)
• Graph Convolutional Encoders  
  (University of Amsterdam, EMNLP 2017)

1
2
3
Simple MT Configuration Example

[main]
output=\"output_dir\"
batch_size=64
epochs=20
train_dataset=<train_data>
val_dataset=<val_data>
trainer=<my_trainer>
runners=[<my_runner>]
evaluation=[
  (\"target\", evaluators.BLEU)]
logging_period=500
validation_period=5000

[en_vocabulary]
class=vocabulary.from_wordlist
path=\"en_vocab.tsv\"

[de_vocabulary]
class=vocabulary.from_wordlist
path=\"de_vocab.tsv\"

[train_data]
class=dataset.from_files
s_source=\"data/train.en\"
s_target=\"data/train.de\"

[val_data]
class=dataset.from_files
s_source=\"data/val.en\"
s_target=\"data/val.de\"

[my_encoder]
class=encoders.SentenceEncoder
rnn_size=500
embedding_size=600
data_id=\"source\"

[my_attention]
class=attention.Attention
encoder=<my_encoder>
state_size=500

[my_decoder]
class=decoders.Decoder
encoders=[<my_encoder>]
attentions=[<my_attention>]
rnn_size=1000
embedding_size=600
data_id=\"target\"

[my_trainer]
class=trainers.CrossEntropyTrainer
decoders=[<my_decoder>]
clip_norm=1.0

[my_runner]
class=runners.GreedyRunner
decoder=<my_decoder>
Define how to load images:

```python
[imagenet_reader]
class=readers.image_reader.imagenet_reader
prefix="/lnet/troja/projects/wmt17-multimodal/data/flickr30k-images"
target_width=229
target_height=229
zero_one_normalization=True
```

And replace encoder with ImageNet network:

```python
[imagenet_resnet]
class=encoders.ImageNet
name="imagenet"
data_id="images"
network_type="resnet_v2_50"
spatial_layer="resnet_v2_50/block4/unit_3/bottleneck_v2/conv3"
slim_models_path="tensorflow-models/research/slim"
```
Sentence Classification

Keep the encoder and replace the decoder (and update the rest)

```python
[my_classifier]
class=decoders.Classifier
data_id="labels"
encoders=[<my_encoder>]
vocabulary=<label_vocabulary>
layers=[200]

[my_runner]
class=runners.PlainRunner
decoder=<my_classifier>

[my_trainer]
class=trainers.CrossEntropyTrainer
decoders=[<my_classifier>]
clip_norm=1.0
```

Pre-trained model parts
Parameters of model parts can be fixed using the gradient blocking module
Supported Features

- Recurrent encoder and decoder with attention
- Beam search decoding with model ensembling
- Deep convolutional encoder
- Self-attentive encoder and decoder
- Wrappers for ImageNet networks
- Custom CNNs for image processing

- ConvNets for sequence classification
- Self-attentive embeddings for sentence classification
- Hierarchical attention over multiple source sequences
- Generic sequence labeler
- Connectionist temporal classification
Console Logging during Training

source: it started the French League season with two wins and two draws, and on Sunday it defeated St. Etienne 5:0.
source: it started the French League season with two wins and two draws, and on Sunday it defeated St. Etienne 5:0.
target: ten do sezo ny francouzské ligy ovoce statého druhého místa s druhého místa zami, v neděli rozstřílel St. Etienne 5:0.
target: ten do sezony francouzské ligy odozval dvěma výhrami a dvěma remizami, v neděli rozstřílel St. Etienne 5:0.

source: it is putting the most pressure on AS Monaco.
source: it is putting the most pressure on AS Monaco.
target: relativitě nejvíce tlaci bota klubu AS Monaco.
target: <unk> <unk> tlak na <unk> AS.
target: relativitě nejvíce tlaci bota klubu AS Monaco.

source: It so far has a balance of 1-1-2 in the league, and over the weekend it parted with Lille with a 1:1 tie.
source: It so far has a balance of 1-0-1-2 in the league, and over the weekend it parted with Lille 1-1 tie.
target: ten má v lize závazník bilanci 1-0-1-2 a o víkendu se rozěsál s Lille 1-1.
target: v <unk> <unk> <unk> a <unk> se <unk> s <unk> <unk> <unk> <unk>.
target: ten má v lize závazník bilanci 1-1 a o víkendu se rozěsál s Lille 1-1.

source: today news appeared on tablet English-language websites stating that the Queens Park Rangers may want Czech.
source: today news appeared on tablet English-language websites stating that the Queens Park Rangers may want Czech.
target: na bulvárech mění anglické kříže nachází výhradně české zprávy.
target: v <unk> <unk> <unk> <unk>.
target: na bulvárech anglických webech dnes vyskočily zprávy.

source: however, according to English bookmakers, Čech and pos; his departure from Chelsea at the last minute is not very likely.
source: however, according to English bookmakers, his departure from Chelsea at the last minute is not very likely.
target: nicméně podle anglických křížů výhod je mohutné zpravodajské mění v pravděpodobnosti.
target: podle anglických křížů výhod je mohutné zpravodajské mění v pravděpodobnosti.

target: target BLEU: 9.055

2018-03-14 12:58:13: Validation (epoch 1, batch number 214968):
2018-03-14 12:58:46: Variable file saved in experiments/czeng_rn_rnm/variables.data
2018-03-14 12:58:46: Best scores saved so far: [9.854902006582597]
2018-03-14 12:59:06: Variables of 'encoder' saved to 'experiments/czeng_rn_rnm/encoderckpt'
2018-03-14 12:59:36: Variables of 'encoder_input' saved to 'experiments/czeng_rn_rnm/encoderinputckpt'
2018-03-14 12:59:26: Best target BLEU 4 on validation: 9.055 (in epoch 1, after batch number 214968)
2018-03-14 12:59:26: Epoch 1/10 Instances 10748805 target/runtime_xent: 6.395 target/train_xent: 2.626 target/bleu-4: 9.055
2018-03-14 12:59:26: Validation time: 11713.46s, Inter-validation: 7901.83s, per-instance (train): 0.025s, per-instance (val): 4.41s
2018-03-14 13:00:07: Validation period setting is inefficient.
Scalar Values in TensorBoard

Losses, evaluation metrics, parameter norms, histograms of gradients
Attention in TensorBoard

TensorBoard | SCALARS | IMAGES | GRAPHS | DISTRIBUTIONS | HISTOGRAMS | PROJECTOR | INACTIVE

- Show actual image size

Runs

Write a regex to filter runs

- san_ctc/
- rnn rnn/
- san_ctc_proj/

Tags matching /.* (all tags)

decoder/attention/image/0
step 10,748,090  Wed Mar 14 2018 12:59:26 GMT+0100 (CET)
Conclusions

Neural Monkey is:

- Actively developing open-source GitHub project
- Suited for researchers, students, and other DL enthusiasts
- Large collection of interesting features from across the NLP sub-topics
- Simple tool to use because of clear and readable configuration files
- Highly modular, therefore also relatively easy to debug
Neural Monkey is:

- Actively developing open-source GitHub project
- Suited for researchers, students, and other DL enthusiasts
- Large collection of interesting features from across the NLP sub-topics
- Simple tool to use because of clear and readable configuration files
- Highly modular, therefore also relatively easy to debug

Thank you for your attention!