

Natural Language Processing

Seq2seq for Text Generation



BITTIGER

The Lifelong Learning Platform of Silicon Valley

Copyright Policy

All content included on the Site or third-party platforms as part of the class, such as text, graphics, logos, button icons, images, audio clips, video clips, live streams, digital downloads, data compilations, and software, is the property of BitTiger or its content suppliers and protected by copyright laws.

Any attempt to redistribute or resell will result in the appropriate legal action being taken.



We thank you in advance for respecting our copyrighted content.

Outline



1. Multiclass
2. Seq2seq
3. Task: Generate poems

Multiclass



BITTIGER

The Lifelong Learning Platform of Silicon Valley

Classifier Layer



- Final layer of a DNN: can be seen as a classifier
- Binary classification:

```
model.add(Dense(1, activation=tf.nn.sigmoid))
```

- Multiclass: # class = num_class

```
model.add(Dense(num_class, activation=tf.nn.softmax))
```

More Preprocessing



- Label must be “one-hot”

Original labels: [“1”, “2”, ... “n”], e.g., n=5

One-hot label: [[1,0,0,0,0], [0,1,0,0,0], ... , [0,0,0,0,1]]

- Keras has a helper:

```
tf.keras.utils.to_categorical(labels, num_classes=5)
```

- For Keras, labels must start from 0!
- Keep original labels for later

Redefine Loss



- Binary classification loss: `'binary_crossentropy'`
- Multiclass classification loss: `'categorical_crossentropy'`

[Source: Mikolov et al, NIPS'13]

Metrics?



- You may remember F-score
- However, F-score is NOT included in Keras
- Use `sklearn`, but remember `sklearn` requires **original labels!**

```
from sklearn.metrics import f1_score  
  
y_true = [0, 1, 2, 3, 4, 2]  
y_pred = [0, 2, 1, 4, 3, 1]  
  
f1_score(y_true, y_pred, average=None)
```


Seq2seq



BITTIGER

The Lifelong Learning Platform of Silicon Valley

Sequence to sequence

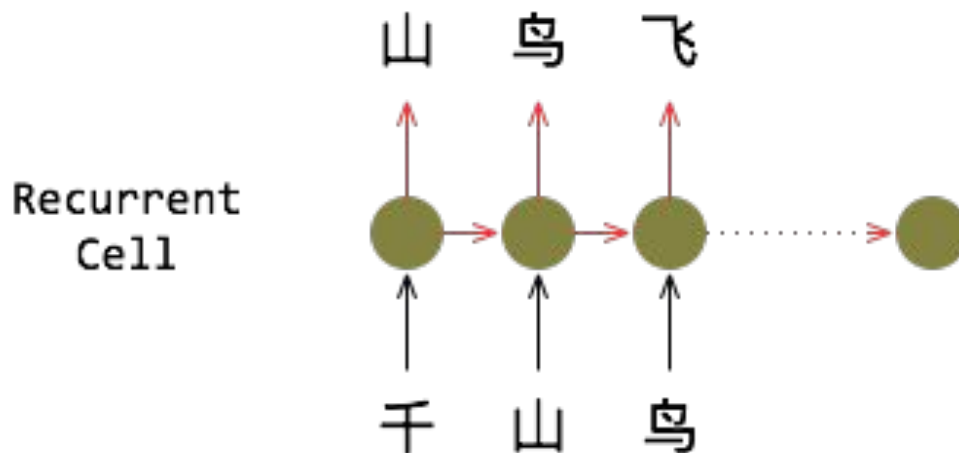


- Input and output are both sequences
- We can use this structure to do e.g., language modeling
 - Given previous words, predict the next word
 - Example task: generate Chinese poems
- A natural choice of neural network is RNN (more specifically, LSTM)
 - Output depends on previous input(s)
- However, as we mentioned, the sequence length cannot be too long

RNN for Language Modeling



- Output is input shift by 1 time step



Dataset



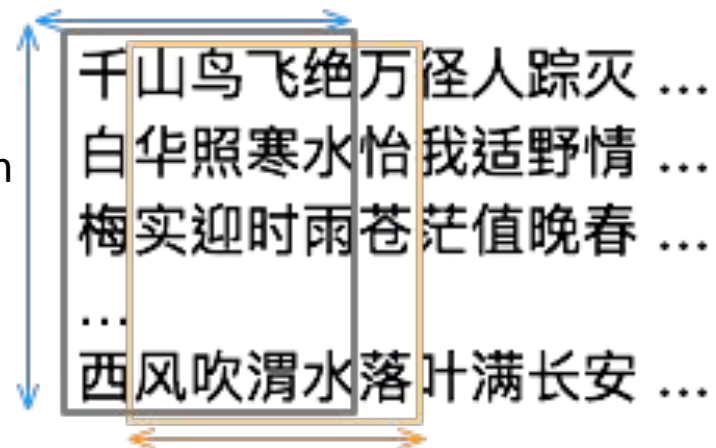
- You can download 全唐诗 from the web
- ~43,000 poems
- Need preprocessing like we mentioned before
- You can also add more custom data!

Preprocessing for LM training



- In practice:
 - Concatenate all poems into 1 vector
 - Reshape it to form a matrix
 - Take a part of the matrix as one batch

1 batch of input



1 batch of output

Loss



- Average loss of the entire sequence
 - `tf.contrib.seq2seq.sequence_loss`
 - Weighted cross-entropy loss for a sequence
 - Can set mask for padding symbols (we won't be using this now)

Tensorflow Basics



BITTIGER

The Lifelong Learning Platform of Silicon Valley

Major steps



- Define hyperparameter
- Input generator
- Build model structure
- Training operations
- Training controller
- Testing operations

Major Object Types



- Operation (op): add, reshape, etc.
 - No gradients
- Variable: weights, embeddings
 - Have gradients (can be turned off)
- Placeholders: a way for you to assign values into the model
 - No gradients
- Session: like a thread that actually send and receive data to/from model

Main Concept



- Build “computation graph”: define elements in the model and how they should be calculated/connected. **Not actually running!**
 - Caveat: It is very difficult to find bugs because the shape of the tensors are flexible. `print` **does not work!**
 - Includes calculating loss and gradients.
- Training controller creates data flow and records output from graph

Training Techniques



- During training, sometimes it's helpful to decay the learning rate over time
- Today, we will use the decay method after training for some epochs, or so-called “warmup” epoch.
- After warmup, we can decrease learning rate exponentially.

Code: Generate Chinese Poems



BITTIGER

The Lifelong Learning Platform of Silicon Valley