Deep Learning and Temporal Data Processing

LSTM in TensorFlow

Andrea Palazzi

July 17th, 2017

University of Modena and Reggio Emilia

Agenda

Synthetic Sequence Dataset

Learning to Count

References

Synthetic Sequence Dataset

Synthetic Sequence Dataset

For this practice I prepared a **synthetic dataset** consisting in $\mathbf{2^{20}}$ **binary sequences**.

For each input sequence, the target is the number of ones in the sequence.

From an implementation standpoint, the target is encoded as one-hot vector. Thus, examples (x, y) from the dataset looks like the following:

input	target
00110011111000111101	00000000000100000000
01000010100001010000	00000100000000000000
11101110010111011110	00000000000001000000

Synthetic Sequence Dataset

The dataset can be found in `synthetic_dataset.py`.

Loading the data is as simple as:

```
from synthetic_dataset import SyntheticSequenceDataset
synthetic_dataset    = SyntheticSequenceDataset()
```

Synthetic data are automatically either generated or loaded from cache (if existent) the first time that dataset property `data` is accessed.

Learning to Count

Goal

Our task is to **count the number of ones in the binary sequences**.

The goal of this practice is to implement and train a **LSTM** [1] network to do so.

Useful Functions

To this purpose, you may find useful the following functions:

- `tf.contrib.rnn.LSTMCell`
- `tf.nn.dynamic_rnn`
- `tf.transpose`
- `tf.gather`
- `tf.layers.dense`

Please refer to the docs to know the exact API.

Good Luck!

References

References i

[1] S. Hochreiter and J. Schmidhuber.
Long short-term memory.
Neural computation, 9(8):1735–1780, 1997.