

## Series 4, March 13th, 2017 (A simple PoS tagger in tensorflow)

**The Model** We want to build a simple feed-forward neural network to perform PoS-tagging. That is, given a word  $w_t$  and  $m$  previous words  $w_{t-m}, \dots, w_{t-1}$  we want to predict the PoS tag  $y_t$ . We choose a simple approach consisting of the following steps:

- Embed words as vectors  $x_i \in \mathbb{R}^d, i \in \{t-m, \dots, t\}$ .
- Concatenate the target word vector and all vectors in the window into a single feature vector  $x_t^{(1)} \in \mathbb{R}^{(m+1)d}$
- Obtain a hidden representation  $x_t^{(2)} \in \mathbb{R}^{d'}$
- Apply a softmax over all PoS tags
- Optimize wrt. the cross entropy loss between our predictions  $p(y)$  and the label  $y^*$

### The Implementation

- Make yourself familiar with tensorflow [1] and install [2] it on your local machine (no GPU required, nor recommended). Please try to find solutions regarding installation on the web – most issues are common ones. Please understand that we cannot answer general tensorflow questions. You can be sure, the web is full of answers.
- Download the skeleton code at [3]. It does some basic steps like loading the data for you and provides a simple code structure. You are welcome to change or ignore it.
- Downloads and unpack the dataset at [4].
- Complete the implementation by following the hints in the code skeleton.

**Experiments** The purpose of this exercise is to get you familiar with tensorflow and deep learning routines. The model is too simple and the data too small to see stellar<sup>1</sup> results, yet you should be able to see effects of different architectures. Start with the small dataset<sup>2</sup>. The following experiments are recommendations and we encourage you to try whatever you feel curious about.

- Run your model for  $d = 50, m = 3, |V| = 20000$  using a single hidden layer of the form

$$x_t^{(2)} = \text{relu}(Wx_t^{(1)} + b)$$

- What accuracy do you get on the validation set? Use tensorboard [7] to watch the curves...
- Increase  $m$  (and optionally  $d$ ) and watch training and validation error. Do you see overfitting?
- Increase  $|V|$  to 50000. How does the accuracy change?
- Add another hidden layer with dimension  $d''$  and vary the dimensions  $d'$  and  $d''$ . What effect do you observe on training and test accuracy?
- Vary the window size and plot the validation accuracy over  $m$ . When is the model saturated?

<sup>1</sup>Feel free to check out [6] and use standard training sets such as WSJ.

<sup>2</sup>Feel free to use the large dataset. Depending on your memory, you might need to change the input processing to hold the corpus in memory only once and create  $x, y$  batches on the fly. You might want to decrease `dev_sample_percentage` as well.

- How much information is in the target word  $w_t$ ? Remove  $x_t$  from the feature vector to find out.
- How much information does the window carry? Find out by providing no window at all.

If you are curious to run longer experiments, you can ssh to euler [5] and run

```
module load new gcc/4.8.2 python/3.6.0
```

to get a running python tensorflow implementation. Before you allocate dozens of cores, use `bjob_connect` and `top` to investigate how many cores tensorflow is actually using. Typically this not more than four. In any case you **must** set `inter_op_parallelism_threads` and `intra_op_parallelism_threads` in tensorflow to match the number of cores that you ordered when submitting jobs. The admins keep an eye on this.

## References

- [1] [https://www.tensorflow.org/get\\_started/get\\_started](https://www.tensorflow.org/get_started/get_started)
- [2] <https://www.tensorflow.org/install/>
- [3] <http://www.da.inf.ethz.ch/teaching/2017/NLP/material/pos-tagger.zip>
- [4] <https://polybox.ethz.ch/index.php/s/pvSDsyyqRlDkaH3>
- [5] [http://brutuswiki.ethz.ch/brutus/Getting\\_started\\_with\\_Euler](http://brutuswiki.ethz.ch/brutus/Getting_started_with_Euler) (no registration required)
- [6] [https://www.aclweb.org/aclwiki/index.php?title=POS\\_Tagging\\_\(State\\_of\\_the\\_art\)](https://www.aclweb.org/aclwiki/index.php?title=POS_Tagging_(State_of_the_art))
- [7] [https://www.tensorflow.org/get\\_started/summaries\\_and\\_tensorboard](https://www.tensorflow.org/get_started/summaries_and_tensorboard)