Natural Language

Practical Classes

Luísa Coheur and Rui Henriques

2025

P8

Word embeddings and RNNs

Image generated by ChatGPT

- **Summary**:
  - Training with RNNs
  - Neural embeddings
- **Operational objectives**:
  - Compare FNNs and RNNs in a small dataset scenario
  - Understand some of the different (hyper)parameters
  - Play with neural embeddings
- **This class needs**: a computer
- **Class material**: these guidelines and a jupyter notebook

While no new customer comes knocking on the door...

After what you have learned in your beloved Natural Language classes (you are attending them, right?[1]) you think that you might get back to work and train an RNN on the *spells data* (P8_dataset_cast) to improve results. In this way, maybe Sam[2] will be happy and pay you.

And this is what you do:

1. Take the jupyter notebook (P8_USE+FFNN_RNN.ipynb), and run it in Google Colab (`https://colab.research.google.com`).

2. Run the FFNN, carefully handling the challenges presented in the notebook (do not rush to the solutions, please).

3. Try to improve the FFNN results by exploring different hyperparameterizations.
   The following table may help you organize the results, yet feel free to test other parameters (*loss function, early stopping criteria, regularization*).

Runs	**Sizes of Hidden Layers**	**Num. Epochs**	**Learning Rate**	**Val. Accuracy**	**Test Accuracy**
Setup 1					
Setup 2					
Setup 3					
Setup 4					
Setup 5					

Tabela 1: First Results

4. Do a similar process with the RNN.

5. And you start thinking... when developing machine learning models, small differences in performance metrics (such as accuracy) may arise due to random factors like weight initialization or data shuffling. Observing that one model achieves slightly higher accuracy than another on a single run does not guarantee that this difference is meaningful. Hum... You remember your Professors saying that to rigorously determine whether one model truly outperforms another, it is essential to evaluate them on multiple runs.

   Yes! **Statistical Hypothesis Testing** can come to the rescue by providing a way to *assess whether observed differences in these samples are likely due to chance or reflect a genuine effect*. You start studying this subject, and you should:

   - Define the null hypothesis, in this case:

     $\mathbf{H}_0$: *The two models have the same mean performance.*

   - Run each model multiple times. Cross-validation? Check what else can vary.
   - Choose a statistical test[3]:
     - Choose the **two-sample t-test** (i.e., independent samples t-test) if the runs of each model are independent (e.g., not in the same partitions)
     - If the data are paired (e.g., evaluated on the same splits):
       - Use the **paired t-test** if the differences between pairs are normally distributed[4]

[1] If not, shame on you!

[2] If you do not know who Sam is: shame on you again!

[3] Notice that many more tests exist, depending on the context. This might be important for your master's work, regardless of the scientific area in which you are working.

[4] To check the distribution: a) compute the differences between the paired samples; b) run a Shapiro,ÄìWilk test (for instance).

        - Use the **Wilcoxon signed-rank test** if the differences between pairs are not normally distributed (non-parametric alternative).
    - The score you will attain tells you the direction and magnitude of the difference, but do not forget to look at the **p-value**:
        - If p_value < 0.05 then reject H_0 (meaning that there is a significant difference)
        - If p_value ≥ 0.05 then fail to reject H_0 (i.e., differences are unlikely significant)

6. After replicating the tests in your notebook, you feel confident. Considering the best setups, do RNNs achieve statistically significant improvements over FFNNs?

After class, you find yourself wondering if you'll still have more contact with Inspector Morcela. You feel a certain tension that you can't quite explain. Your intuition tells you that there will be many more missions where you'll have to put your NLP knowledge at the service of humanity. But what kind of missions will they be? And... will you be up to the task? You sigh. You'll need to pay full attention in class.