

Metagenomic Classification: a Deep Learning Approach

Daniele Bellani

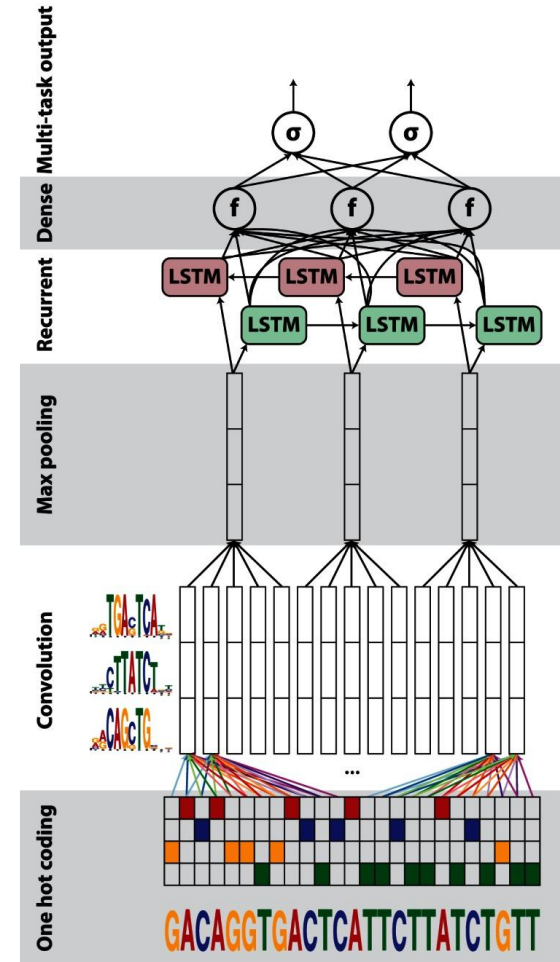
?? • ?? • ????

Goal

Assign a **taxon** to raw genomic slices (**reads**)

Proposed methodology

- Use of Deep Learning classifiers:
 - Convolutional Neural Networks
 - Recurrent Neural Networks (LSTM, GRU)
 - Combinations of the two
- Run it on GPUs for faster execution



Motivations

- Address complex memorization problem with powerful state-of-the-art memorization machines for the first time
- Provide an alignment-free alternative to the most used taxonomy assignment tools.
- Expected advantages:
 - High prediction speed
 - Local variations (mutations) tolerance
 - Independence from reconstructed genomes
 - Can run on GPUs (faster, cheaper and greener than computer clusters) [4]

Related works

Main approaches found in literature:

- Alignment based approaches, i.e. MEGAN [1]
- K-mer based approaches, i.e. Kraken [2]
- Machine learning approaches, i.e. NBC [3]

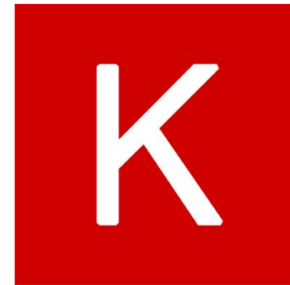
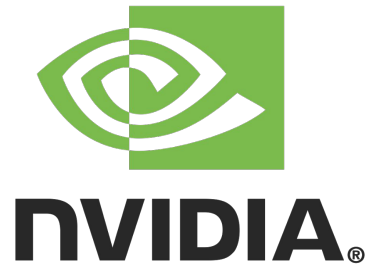
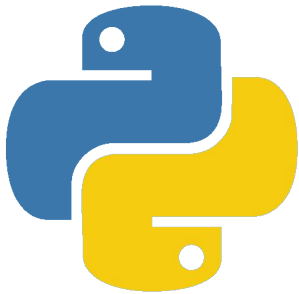
Popular datasets:

- The Human Microbiome Project dataset [5]
- Sargasso Sea dataset [6]
- simBA-5 (Kraken) simulated dataset [7]

Tasks

- Evaluate effectiveness
 - Accuracy
 - Precision
 - Recall
 - ...
- Evaluate efficiency
 - Learning phase time
 - Prediction phase time
 - Memory
- Compare the performances of different DL architectures
- Comparison with current state-of-the-art classification tools

Tools



Bibliography

- [1] Huson, D. H., Auch, A. F., Qi, J., & Schuster, S. C. (2007). MEGAN analysis of metagenomic data. *Genome research*, 17(3), 377-386.
- [2] Wood, D. E., & Salzberg, S. L. (2014). Kraken: ultrafast metagenomic sequence classification using exact alignments. *Genome biology*, 15(3), R46.
- [3] Rosen, G., Garbarine, E., Caseiro, D., Polikar, R., & Sokhansanj, B. (2008). Metagenome Fragment Classification Using *N*-Mer Frequency Profiles. *Advances in bioinformatics*, 2008.
- [4] Huang, S., Xiao, S., & Feng, W. C. (2009, May). On the energy efficiency of graphics processing units for scientific computing. In *Parallel & Distributed Processing, 2009. IPDPS 2009. IEEE International Symposium on* (pp. 1-8). IEEE.
- [5] <https://hmpdacc.org/hmp/overview/data-model.php>
- [6] <ftp://ftp.ncbi.nih.gov/pub/taxonomy/>
- [7] <http://ccb.jhu.edu/software/kraken/>