Review of the litterature

Guillermo Martinez (matricule x)
Dereck Piché (matricule 20177385)
Jonas Gabirot (matricule 20185863)

February 24, 2023

Aptamers

Description of aptamers

insert

Motivation behind the study of Aptamers

insert

Machine learning algorithms

Multilayer Perceptrons

insert

Reccurent Neural Networks

The broad definition of a recurrent neural network is that there are some cycles in the layers. Typical recurrent network adapted for sequences record previous outputs by storing information in the so called hidden state of the network. Since you have to process each token one at a time, their is no parallelization. One advantage of recurrent neural network is that they grow linearly with respect to the amount of tokens as input, which means that their really isn't a practical limit to the number of tokens that it can process with reasonable time complexity.

Transformers

According to our assumptions, the transformer architecture [1] is by far the most appropriate for our task. Transformers use a multi-headed attention mechanism and self-attention. Let be $t_1, \ldots, t_n$ be a sequence of input tokens. Then a

single head will create for each input token t_i an output y_i which is a linear combination of the other tokens given as input. It's complexity is $O(kn^2)$, where k are factors independant of input size. This process is repeated for multiple heads with different parameters (learned). Their outputs are concatened and fed into a fully connected network that combines their ouputs. The architecture is complex and impractical to spell out in greater detail here.

G-flow nets

insert

Currenctly used classical algorithms (State of The Art)

There is currently little research and writing on learning learning with deep learning algorithms. Instead, biology-specific algorithms biology-specific algorithms are favoured, as well as clustering algorithms. clustering algorithms. For example, this article from January 2023 uses an original algorithm that combines clustering methods to find an optimal an optimal aptamer from a selection. https://pubs.acs.org/doi/pdf/10.1021/acssynbio.2c00462. However, some recent papers use deep learning. "Machine learning guided aptamer refinement and discovery" (https://www.nature.com/articles/s41467-021-22555-9) uses a standard MLP neural network to find the most compatible (high affinity) aptamers compatible (high affinity) aptamers with target molecules. The estimation of free energy is a sub-step of the affinity calculation. It performs a truncation step to minimise the length of the aptamer without altering its properties. Another deep learning model with aptamers is AptaNet (https://www.nature.com/articles/s41598-021-85629-0). This model uses an MLP and a CNN to learn the relationship between aptamers and target proteins proteins (Aptamere-protein relations or API). The MLP works best, with a test accuracy of 91.38algorithms such as SVM, KNN and random forests. This model uses a very detailed database containing numerous auxiliary variables measured in the laboratory for each individual, but with only 1000 individuals. No published aptamer model uses transformers or RNNs to predict free energy, so the predict free energy, so our method would be original in this field.

References

[1] Ashish Vaswani et al. Attention is all you need. *CoRR*, abs/1706.03762, 2017.