Team Viserion: Status Update #1

Julie Helmers, Millie Dwyer, Lihan Yao, Shasha Lin

November 16, 2017

1 Affinity Regression

To address our issue with different sequence alignments (as discussed in last week's update), we integrated labeled data with our unlabeled data using Clustal Omega alignment software. This program took both sets of aligned sequences and returned a single set of aligned sequences. Here, labeled data refers to sequences with experimentally confirmed RNA binding profiles. Because we have generated a single alignment for both datasets, we are now able to generate low dimensional representations of labeled data to fit and evaluate affinity regression.

2 Model II: image captioning

As mentioned in our previous report, we have implemented a ResNet encoder + LSTM decoder (an architecture inspired by image captioning task [1]) as our 2nd model, after Seq2Vec. Due to the complexity of ResNet and LSTM, we are still running this model on GPU, and expect to use tSNE plots of learned representation from LSTM hidden vector, and potentially the performance for the downstream affinity regression for quality check.

3 Model III: Cho's Character Level Neural Machine Translation

We implemented [2].

References

[1] "Imagecaptioning." https://github.com/yunjey/pytorch-tutorial/tree/master/tutorials/03-advanced/image_captioning. Accessed: 2017-11-09.

[2] J. Lee, K. Cho, and T. Hofmann, "Fully character-level neural machine translation without explicit segmentation," *CoRR*, vol. abs/1610.03017, 2016.