CS 273B Paper Review: **Attend and Predict: Understanding Gene Regulation by Selective Attention on Chromatin**

Tymor Hamamsy, Tristan McRae, Mamie Wang, Timothy Wu

Summary

In this paper, Singh *et al*. presented an attention-based deep learning method (AttentiveChrome) to predict gene expression from genome-wide chromatin signals. The method focuses on interpreting the importance of distinct epigenetic histone marks and how they interact in gene regulation. AttentiveChrome used a hierarchical attention architecture with multiple Long short-term memory (LSTM) modules to encode the input signal of histone modification (HM) marks and model their interactions. The authors evaluated their model across 56 different cell types in humans using five core HM marks from the Roadmap Epigenome Project (REMC). They presented two major findings. First, they found that their proposed architecture is more accurate compared with previous studies that also predicted gene expression from HM marks using convolutional neural networks (CNNs). Second, they concluded that attention weights trained from AttentiveChrome architecture provided a better interpretation compared with several popular visualization methods.

The proposed architecture used LSTMs to find sequential patterns in input sequences and between HMs and used attention layers to highlight the importance of selective HMs as well as the regions of importance within each HM. Input sequences for each HM were passed through a bidirectional LSTM (BiLSTM) and attention layers to create a representational vector for each HM. The HM vectors were then passed through a second pair of BiLSTM and attention layers to create one final representation. A softmax classifier then outputted a prediction on whether a gene was on or off.

Previous research on gene expression prediction from chromatin modification marks focused on traditional machine learning approaches, such as linear regression, SVM, and random forests. These techniques typically require strong assumptions about feature, and have trouble learning the complex spatial dependencies of chromatin marks. The authors previously published a method called DeepChrome which employed CNN for the same task; while the DeepChrome method outperformed traditional machine learning methods, AttentiveChrome outperforms DeepChrome and the baseline CNN. The authors presented five variations of the AttentiveChrome method and the best performing one, LSTM-α, had improved performance over the baseline CNN model in 50 of the 56 cell types. The LSTM-α,β and LSTM-Attn models were a close second, both outperforming the baseline in 49 of the 56 cell types.

Critiques

The paper presented a deep learning architecture that combined bidirectional LSTMs (BiLSTMs) with attention weights to predict whether a gene will be on or off. Attention layers were able to provide a level of interpretability for spatial dependencies among histone modification mark signals. The ability to model interactions among mark signals along with providing a metric to interpret the predictions made the architecture novel compared with previous machine learning studies.

One critique is the use of an LSTM on HM level data. LSTMs are optimized to find sequential patterns in long input sequences; however, there is no inherent sequence to HMs and at the HM level, the inputs have a relatively short length of 5. A fully connected or convolutional layer would have been more appropriate here than an LSTM.

The study trained/validated the models for each of the 56 cell types separately. For each cell type, the 19,802 gene samples were divided into three separate, equal-sized folds for training (6601 genes), validation (6601 genes), and testing (6600 genes). One critique on training/testing is that rather than pooling different cell type data or training models on one cell type and testing out of sample in other cell types, the models were trained and tested on each of the 56 individual cell types independently. The reported results represent the average performance of all 56 models across the 56 different cell types. It would be more rigorous to show if the model can work independent of cell type, i.e. validating/testing out of sample in another cell type or with pooled data. Additionally, in the results section they averaged across all of the different cell types, more detailed reporting/explanation of performance across the different cell types is necessary.

The model performance was then evaluated using AUC. However, the paper failed to present other classification metrics and potentially overlooked the problem of class imbalance. Class imbalance could cause AUC to be a poor indicator of classifier performance. Precision-recall can provide a better statistic to inspect if class imbalance exists.

The authors also do not show how the other models that they mention besides the baseline CNN (i.e. traditional machine learning) performed on this data, which would be valuable as an additional baseline.

The authors used the α-attention weights to visualize which gene region the AttentiveChrome is focusing its attention on. They used a histone modification mark (Hactive) that was not used during the training as a reference for active gene regions and used the Pearson correlation to Hactive read coverage as a validation to the interpretability of the attention weight. However, the comparison was performed on the averaged profile, which might obscure the gene-wise performance. It is recommended to have an evaluation metric that has higher granularity.