FUSING INFORMATION SOURCES THROUGH CONVOLUTIONAL NEURAL NETWORKS FOR GENE REGULATORY NETWORK INFERENCE

Mariano Rubiolo, Matías Gerard, Leandro Vignolo

Research Institute for Signals, Systems and Computational Intelligence, sinc(i), FICH/UNL-CONICET,

Ciudad Universitaria, 3000 Santa Fe, Argentina

Background:

The inference of gene regulatory networks underlying a gene expression dataset has received special attention by scientists for understanding the role of each gene and their relationships in the biological processes at molecular levels. Several inference methods based on machine learning models have been proposed, considering biological information in some cases, with different performance levels. This work proposes to fuse the inferred networks by applying both Extreme Learning Machine and Genetic Algorithms models with the biological information available in Gene Ontology for the involved genes. Inspired by how a Convolutional Neural Network identifies a color image analyzing the typical three colors channel, this approach considers the reconstructed networks after running the models and its biological information as three information channels. To test this proposal, synthetic datasets were built using the *E. coli* and *Yeast* models provided by the GeneNetWeaver tool of DREAM challenge, and were used for obtaining the best hyperparameters configuration for the convolutional neural network. Several simulations of the same gene regulatory network were presented to the convolutional model in order to obtain a more accurate regulatory network between those genes.

Results:

The performance of the baseline ELM model (F1 = 0.636), is improved with application of the convolutional model over the obtained raw gene regulatory network, with an F1 of 0.682.

If the resultant network from the previous model is fused with the network inferred by the genetic algorithm model, the convolutional neural network improves their performance, scaling the F1 up to 0.741. And, if the biological information layer is added to the input, the convoluted network performs better than the previous, up to 0.846 for the F1 measure.

Conclusions:

Our experiments show that complementing machine learning models predictions with biological information of the genes involved in gene regulatory networks as the input of a convolutional model has demonstrated an improvement on the final identification of the networks underlying the gene expression data. Future work involves a deeper convolutional model hyperparameters evaluation, regarding not only model layers but also biological information at the input.

Poster Session: http://bit.ly/cab2c-2021-posters
Poster Rubiolo et al: https://youtu.be/e9_a_1tmQKE