Mining gene regulatory networks
by neural modeling of expression time-series

Mariano Rubiolo[1,2], Diego Milone[2], Georgina Stegmayer[2]

[1]CIDISI, UTN-FRSF, Lavaise 610.
[2]sinc(*i*), FICH-UNL/CONICET, C. Universitaria - Pje El Pozo, R. N. 168
(3000) Santa Fe - Argentina
`mrubiolo@santafe-conicet.gov.ar`

Several machine learning techniques have been developed for discovering interesting and unknown relations between variables from data, even more when these techniques can assist in understanding the behaviour of a complex system. This behaviour can be represented by the interactions between its variables, for instance as a directed graph. A gene regulatory network (GRN) is an abstract mapping of gene regulations in living organisms that can help to predict the system behavior. During last years, many approaches have been proposed to unravel the complexity of gene regulation. Genes interact with one another and these interactions can be measured over a number of time steps, producing temporal gene expression profiles. A hot topic on gene expression data analysis nowadays is the reconstruction of a GRN from such data, revealing the underlying network of gene-to-gene interactions. In other words, the goal is to determine the pattern of activations and inhibitions among genes that make up the underlying GRN.

Given the expression levels of a set of interacting genes measured at different time points, formal methods have been traditionally developed to model gene interactions. As a more recent alternative, the discovery of GRNs by data-driven methodologies has been under study in more recent years. In particular, artificial neural networks (ANN) can model pairs of genes activity over a number of time steps in order to infer genetic networks. Thus, all the possible combinations between genes must be analyzed in order to discover their relations. Using neural networks for this task requires training them to predict a target gene regulation from candidate regulating profiles. By adjusting their synaptic weights, NNs alter their configuration to model each gene connection, which results in a minimum error in predicting a target profile.

This work [1] proposes a novel approach to discover a GRN from temporal genes expression profiles by using a pool of ANNs with temporal delays at the input, named *GRNNminer* (Fig. 1). Each NN is designed to discover, for a target gene profile at the output, the potential regulator of that gene at the input, by modeling their gene-to-gene interaction during a time period. The ability of each trained NN
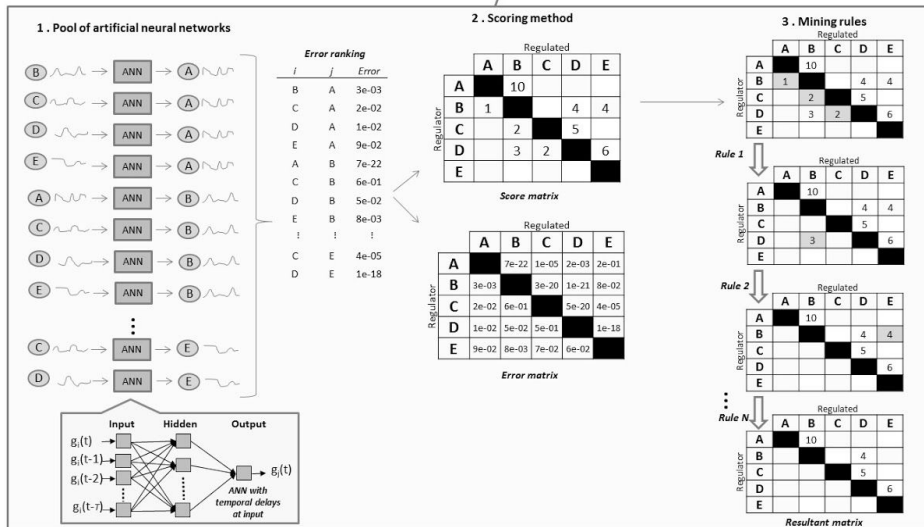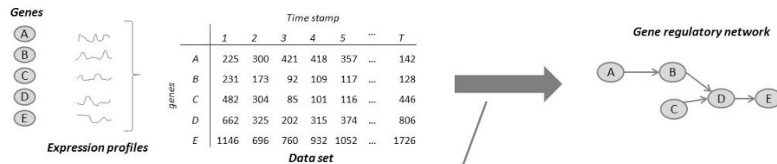
Figure 1. GRNNminer: a three steps approach for discovering relations between variables (genes) from time series (expression profile).

to model each possible relation is judged by the evaluation of the generalization error during the training process. For each possible relationship found, a score matrix is computed to detect the most likely regulations. Then, some rules for mining the unknown GRN from the score matrix are used, which allows to discover each correct gene-to-gene relation from all the possible ones.

Table 1. Comparison of GRNNminer against literature methods.

Methods	Accuracy [%]	Sensitivity [%]
Pearson correlation [2]	92.69	100.00
Rank correlation [2]	83.92	100.00
Mutual Information [3]	65.79	100.00
Knott et al. [4]	99.71	88.89
Smith et al. [5]	98.54	88.89
GRNNminer [1]	**99.71**	**100.00**

Several experiments were done over a known GRN present in an artificial dataset, in comparison against traditional and recent methods in literature. By calculating sensitivity and accuracy measures, experimental results shown at Table 1 demonstrate that the proposal is capable of discovering all the gene-to-gene relations (Sensitivity of 100%) and reconstruct the subjacent GRN , with the least amount of false positives (Accuracy of 99.71%). Moreover, applying the same method over a real dataset, it was possible to obtain the interactions between real genes.

The capability of our mining approach to identify the potential existing relations between variables could be very useful because it allows to focus the attention over a set of variables. GRNNminer could be used as a starting point from where researchers can reconstruct or build a network of interactions, later to be tested or validated through experiments, not only in Bioinformatics but also in any field of knowledge in which time series are the input data.

References

1. Rubiolo, M., Milone, D. H., Stegmayer, G.: Mining Gene Regulatory Networks by Neural Modeling of Expression Time-Series. IEEE/ACM Transactions on Computational Biology and Bioinformatics, vol.12, no. 6, (2015) 1365-1373.
2. Gibbons, J., Chakraborti, S.: Nonparametric Statistical Inference. Fourth Edition: Revised and Expanded. Taylor & Francis, (2014).
3. Cover, T., Thomas, J.: Elements of information theory. Wiley (1991).
4. Knott, S., Mostafavi, S., Mousavi, P.:A neural network based modeling and validation approach for identifying gene regulatory networks. Neurocomputing, vol. 73, no. 13-15, (2010) 2419–2429.
5. Smith, V. A., Jarvis, E. D., Hartemink, A. J.: Evaluating functional network inference using simulations of complex biological systems. Bioinformatics, vol. 18, no. 1, (2002) S216–S224.