Extreme learning machines for discovering gene regulatory networks from temporal profiles of expression

M. Rubiolo[1,2], D.H. Milone[2] and G. Stegmayer[2]

[1] CIDISI, UTN-FRSF, Lavaise 610, Santa Fe, Argentina, mrubiolo@santafe-conicet.gov.ar
[2] sinc(i), FICH/UNL-CONICET, Ciudad Universitaria, RN 168 Km. 472.4, Santa Fe, Argentina.

sinc(*i*) Research Center for Signals, Systems and Computational Intelligence (fich.unl.edu.ar/sinc)
M. Rubiolo, D. H. Milone & G. Stegmayer; "Extreme learning machines for discovering gene regulatory networks from temporal profiles of expression"
4th ISCB-LA Bioinformatics Conference, pp. 1-2, 2016.

Background

Reconstructing a gene regulatory network (GRN) from gene expression data is one of the ultimate goals of bioinformatics today. In recent years, several computational methods were proposed to support the discovery of GRN, with different levels of accuracy. Most of them require several input sources to provide an acceptable prediction. Indeed, not only gene expression but also experimental results such as the wild type unperturbed network, steady-state levels of single-gene knockouts and knockdowns are required. Thus, there is an important challenge in reconstructing a GRN only from temporal gene-expression data. We present here a model capable of modeling all possible gene-to-gene regulations just from gene expression data by using a pool of fast and accurate artificial neural networks.

Results

Extreme Learning Machines (ELM) are a new supervised paradigm of neural networks that has gained interest in the last years because of their high learning rate and best performance in terms of generalization capabilities. The essence of ELM is that the learning parameters of hidden nodes are randomly assigned and need not to be tuned, while the output weights can be analytically determined by a simple generalized inverse operation. In our work, a pool of ELMs (Fig 1) is used for modeling each possible regulator-regulated relation just from gene expression time series. The original regulated time series and a time-delayed version of the regulator time series are shown to the ELMs, so the model can learn the temporal dynamics between these profiles. Considering the generalization error of the ELM as a measure of how much modelable is a gene-to-gene relation, it is possible to detect if there is an effective regulation between each pair of analyzed genes.

Table 1 shows the results of the proposed method applied to three real datasets: i) Reverse-engineering and Modeling Assessment (IRMA), a 5 genes network from *Saccharomyces cerevisiae*, whose profiles have 16 time points; ii) *Escherichia coli* SOS pathway, with 8 genes regulating the SOS response of DNA damage and 50 points; and iii) 11 genes of *S. cerevisiae* in a network that controls G1 step of cell cycle with 16 time points. The results obtained (see Fig. 2) show that most connections found are true positives and there are just a few false negatives. The table 1 results are higher than those reported by state-of-the-art methods for the same datasets and under the same experimental conditions.

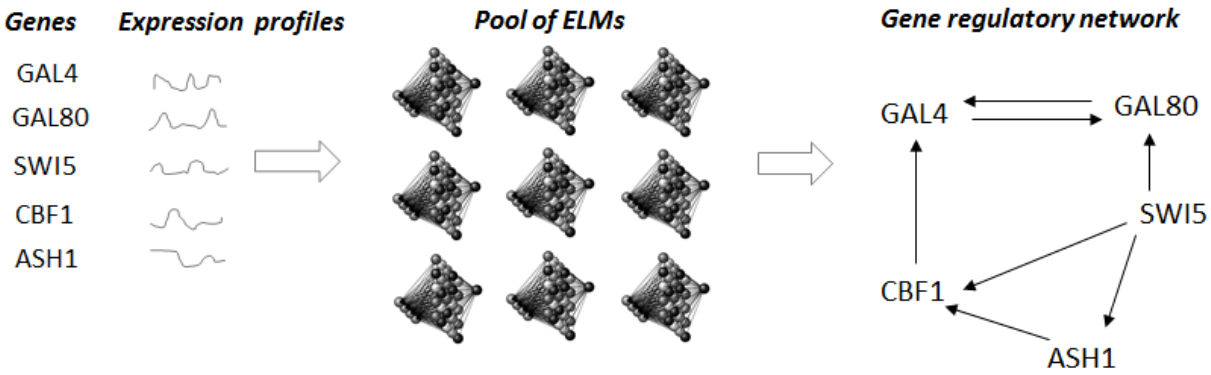

Figure 1. Obtaining the GRN from gene profiles by modeling the regulations with a pool of ELMs.

Table 1. Performance in GRN identification of IRMA, *E.Coli* and *S. Cerevisiae* gene expression data.

	Accuracy [%]	Specificity [%]
IRMA	80.00	83.33
E. Coli	90.63	91.07
S. Cerevisiae	67.77	73.47

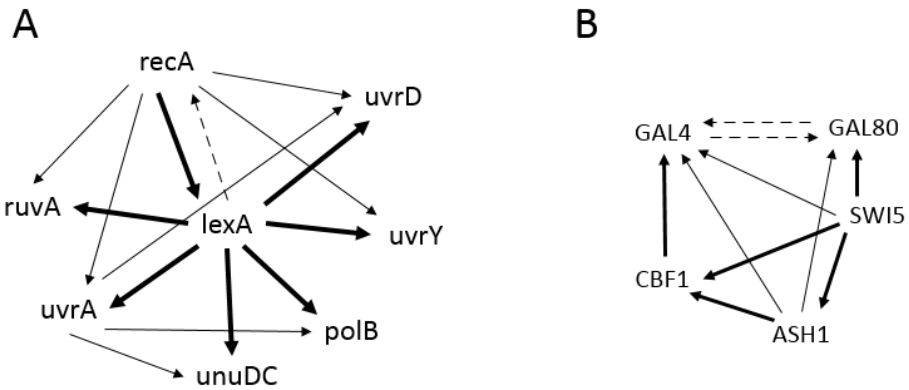

Figure 2.(A) *E. coli* SOS pathway; (B) IRMA network. True positives are shown in thick lines, false positives are shown in thin lines, false negatives are shown in thin-dotted lines.

Conclusions

This work proposed a novel way to infer GRN underlying gene profiles, by which it was possible to determine the correct regulations in three experimental datasets. High levels of accuracy and specificity were obtained for the challenging task of identifying the GRN only considering gene expression data. Our approach has shown to be an accurate and scalable tool to determine the unknown GRN from temporal series of gene expression.

Supported by: CONICET, MINCyT, UNL.

sinc(*i*) Research Center for Signals, Systems and Computational Intelligence (fich.unl.edu.ar/sinc)
M. Rubiolo, D. H. Milone & G. Stegmayer; "Extreme learning machines for discovering gene regulatory networks from temporal profiles of expression"
4th ISCB-LA Bioinformatics Conference, pp. 1-2, 2016.