

BITS :: Call for Abstracts 2019 - Oral communication

Type	Oral communication
Session	Machine Learning in Bioinformatics
Title	Transfer Learning for Gene Network Reconstruction
All Authors	Mignone P(1,3), Ceci M(1,3), Pio G(1,3), Malerba D(1,3), D'Elia D(2)
Affiliation	(1) Department of Computer Science, University of Bari Aldo Moro, Bari (2) Institute for Biomedical Technologies, CNR, Bari (3) Infolife Lab - National Interuniversity Consortium for Informatics (CINI), Rome

Motivation

The regulation of gene expression is a complex process that involves different actors (e.g., modification enzymes, regulatory proteins or non-coding RNAs) and that can be modulated at different levels and by different mechanisms. The extent to which a gene can be expressed depends by endogenous or exogenous factors that vary by the type of tissue or organ in which the cell is located [1]. When some control mechanisms are compromised, cells undergo a series of modifications that can bring to their transformation in cancerous cells [2].

Studying and reconstructing Gene Regulatory Networks (GRNs) is therefore fundamental to understand gene expression regulation within cells. The computational reconstruction of GRNs from gene expression data is receiving increasing attention in recent years, due to the valuable insights that its results provide for the understanding of complex diseases. Most of existing methods reconstruct the links of the network through machine learning approaches, by analyzing known examples of interactions among genes. However, existing methods often suffer when they are adopted in real biological applications, because of the limited number of labeled examples (i.e., validated interactions) and the absence of negative examples (i.e., confirmed absent interactions). To this aim, transfer learning techniques can be employed to leverage additional knowledge from an external source GRN to better reconstruct a target GRN.

Methods

We propose a transfer learning method which can combine the model for the reconstruction of the GRN of a target organism with that for the reconstruction of another, possibly related, source organism. Methodologically, our approach consists of four stages. The first two stages aim to solve the problem of positive-unlabelled learning. In particular, in the first stage, we apply a clustering algorithm on the positive links, which aims at identifying prototype links, i.e., centroids, that summarize the different kinds of positive links. In the second stage, we assign a weight for each unlabelled link, according to its similarity with the identified centroids. In the third stage, we exploit both the labeled links and the set of weighted unlabelled links to learn a classification model, through a machine learning method which considers possibly different weights assigned to training examples. Finally, in the fourth stage, we combine the models obtained for the source and the target organisms, to obtain a new model that better describes the target network, also according to the knowledge acquired from the (reconstructed) source network.

Results

In our experiments, we reconstructed the human GRNs, by exploiting the knowledge of the mouse GRNs. Existing gene interactions (ground truth) were taken from BioGRID [3], while all the interactions were described with expression data from Gene Expression Omnibus (GEO) [4].

The evaluation was performed by analyzing the correctly classified existing interactions, in terms of the recall@k measure, which has been effectively adopted in [5], since it avoids imposing any assumption on the unlabelled examples. We compared the quality of the reconstructed network with that obtained without the exploitation of the source network, and we found that the knowledge acquired from the source network was helpful. Moreover, a comparative analysis with different state-of-the-art transfer learning approaches showed that our method was able to outperform them, both in terms of accuracy of the reconstruction and in terms of computational efficiency.

Finally, a qualitative analysis highlighted that our method was able to return interesting, previously unknown, interactions among genes, possibly involved in neurological diseases as well as in other human disorders. Among them, our method predicted an interaction between LINC00657 and RPL39, between NBPF8P and ND4, and between EEF1A1P5 and UBEN. The functional relationship of these genes, that was not detected by other tools, appears to be plausible, according to information retrieved from the literature. Therefore, our method can be considered a very useful tool for the accurate prediction of (possibly previously unknown) interactions among genes: researchers may use our tool as a reliable source for experimental lab-validations of unknown function and role of many genes.

References

- 1) Klemm SL, Shipony Z, Greenleaf WJ (2019) Chromatin accessibility and the regulatory epigenome. *Nature Reviews Genetics*, 20:207–220
- 2) Zhang L et al. (1997) Gene Expression Profiles in Normal and Cancer Cells. *Science*, 276, (5316):1268-1272
- 3) Stark C. et al. (2006) Biogrid: a general repository for interaction datasets. *Nucleic Acids Res.*, 34(Database-Issue), 535–539
- 4) Barrett T et al. (2013) NCBI GEO: archive for functional genomics data sets--update. *Nucleic Acids Res.* 41(Database issue): D991-5
- 5) Pio G et al. (2015) ComiRNet: a Web-based System for the Analysis of miRNA-gene Regulatory Networks, *BMC Bioinformatics* 16 (Suppl 9), S7

Info

-

Figure

-

Availability

-

Corresponding Author

<i>Name, Surname</i>	Paolo, Mignone
<i>Email</i>	paolo.mignone@uniba.it
<i>Submitted on</i>	29.04.2019

Società Italiana di Bioinformatica

C.F. / P.IVA 97319460586

E-mail bits@bioinformatics.it

Sede legale Viale G. Mazzini, 114/B - 00195 Roma

Website bioinformatics.it

message generated by sciencedev.com for bioinformatics.it 10:37:43 29.04.2019
