Big Data analytics for knowledge transfer among organisms while reconstructing Gene Regulatory Networks

Paolo Mignone[1,2], Gianvito Pio[1,2], Domenica D'Elia[3] and Michelangelo Ceci[1,2,4]

1 Department of Computer Science, University of Bari Aldo Moro, Bari 70125, Italy
2 National Interuniversity Consortium for Informatics (CINI), Roma 00185, Italy
3 CNR - Institute for Biomedical Technologies, Bari 70126, Italy
4 Department of Knowledge Technologies, Jozef Stefan Institute, Ljubljana 1000, Slovenia

Abstract

The reconstruction of gene regulatory networks (GRNs) from gene expression data is pivotal for the understanding of gene regulatory mechanisms and processes. Because of the importance of the elucidation of gene complex interactions and functions for the identification of genes involved in diseases, the development of new and more effective methods and tools for the GNR has been receiving more attention in the latest years. In this context, machine learning and big data analytics tools can be considered fundamental. However, most existing methods *(i)* produce poor results when the amount of labeled examples is limited or when no negative example is available and *(ii)* they are not able to exploit information extracted from GRNs of other (better studied) related organisms.

We overcome these limitations by proposing an innovative *transfer learning* method, called BioSfer [1], which is able to exploit the knowledge about the GRN of a source organism for the reconstruction of the GRN of the target organism. In the first stages, we identify two predictive models to discover unknown links for both the considered GRNs. In the final stage, we build a new geometrically-combined model, which is able to better identify unknown links. Moreover, the proposed method is natively able to work in the positive-unlabeled setting, where no negative example is available, by fruitfully exploiting a set of unlabeled examples. In our experiments, we reconstructed the human GRN, by exploiting the knowledge of the GRN of Mus musculus. The qualitative analysis showed that the proposed method is able to identify biologically plausible gene regulations that are not identified by other tools. Results showed that the proposed method outperforms state-of-the-art approaches [2-7] and identifies previously unknown functional relationships among the analyzed genes.

Availability of data and materials

The system, the adopted datasets and all the results are available at:
http://www.di.uniba.it/~mignone/systems/biosfer/index.html

Acknowledgements

We acknowledge the support of the EU Commission through the project MAESTRA - Learning from Massive, Incompletely annotated, and Structured Data (Grant number ICT-2013-612944) and of the National Research Council (CNR) Flagship Project InterOmics.

References

[1] Mignone, P., Pio, G., D'Elia, D., Ceci, M., Exploiting Transfer Learning for the Reconstruction of the Human Gene Regulatory Network, Bioinform. 36(5): 1553-1561 (2020)

[2] Zhang, J., Li, W. & Ogunbona, P. Joint geometrical and statistical alignment for visual domain adaptation. In2017 IEEEConference on Computer Vision and Pattern Recognition, CVPR 2017, Honolulu, HI, USA, July 21-26, 2017, 5150–5158(2017).
[3] Wang, J., Chen, Y., Hao, S., Feng, W. & Shen, Z. Balanced distribution adaptation for transfer learning. In2017 IEEEInternational Conference on Data Mining, ICDM 2017, New Orleans, LA, USA, November 18-21, 2017, 1129–1134 (2017).
[4] M. Long, J. Wang, G. Ding, J. Sun and P. S. Yu, "Transfer Joint Matching for Unsupervised Domain Adaptation," 2014 IEEE Conference on Computer Vision and Pattern Recognition, Columbus, OH, 2014, pp. 1410-1417, doi: 10.1109/CVPR.2014.183.
[5] Huynh-Thu V, Irrthum A, Wehenkel L, Geurts P (2010). “Inferring regulatory networks from expression data using tree-based methods.” PLoS ONE, 5(9), e12776. doi: 10.1371/journal.pone.0012776.
[6] Aibar S, Bravo Gonzalez-Blas C, Moerman T, Huynh-Thu V, Imrichova H, Hulselmans G, Rambow F, Marine J, Geurts P, Aerts J, van den Oord J, Kalender Atak Z, Wouters J, Aerts S (2017). “SCENIC: Single-Cell Regulatory Network Inference And Clustering.” Nature Methods, 14, 1083-1086. doi: 10.1038/nmeth.4463.
[7] Mignone, P., Pio, G., Positive Unlabeled Link Prediction via Transfer Learning for Gene Network Reconstruction. ISMIS 2018 : the 24th International Symposium on Methodologies for Intelligent Systems, doi: 10.1007/978-3-030-01851-1_2, 2018