

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

Paolo Mignone^{1,2}, Gianvito Pio^{1,2},, Domenica D'Elia³, Michelangelo Ceci^{1,2,4}

¹Department of Computer Science, University of Bari Aldo Moro, Bari, Italy

²National Interuniversity Consortium for Informatics (CINI), Rome, Italy

³Institute for Biomedical Technologies, National Research Council, Bari, Italy

⁴Department of Knowledge Technologies, Jozef Stefan Institute, Ljubljana, Slovenia

Competing interests: PM none; GP none; DD none; MC none

The reconstruction of gene regulatory networks (GRNs) from gene expression data is pivotal for understanding gene regulatory mechanisms and processes. 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 labelled 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 (Mignone *et al.*, 2020), which can 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 can identify unknown links better. 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 *M. 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 (Zhang *et al.*, 2017; Wang *et al.*, 2017; Long *et al.*, 2014; Huynh-Thu *et al.*, 2010; Aibar *et al.*, 2017; Mignone *et al.*, 2018) and identifies previously unknown functional relationships among the analysed 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 EU Commission's support 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. Aibar S, Bravo Gonzalez-Blas C, Moerman T, Huynh-Thu V, Imrichova H *et al.* (2017). SCENIC: Single-Cell Regulatory Network Inference And Clustering. *Nature Methods* **14**:1083-1086. <http://dx.doi.org/10.1038/nmeth.4463>
2. 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. <http://dx.doi.org/10.1371/journal.pone.0012776>
3. Long M, Wang J, Ding G, Sun J, Yu PS (2014) Transfer Joint Matching for Unsupervised Domain Adaptation. 2014 IEEE Conference on Computer Vision and Pattern Recognition, Columbus, OH, 2014, pp. 1410-1417. <http://dx.doi.org/10.1109/CVPR.2014.183>
4. Mignone P, Pio G. (2018) Positive Unlabeled Link Prediction via Transfer Learning for Gene Network Reconstruction. *ISMIS 2018: The 24th International Symposium on Methodologies for Intelligent Systems*. http://dx.doi.org/10.1007/978-3-030-01851-1_2, 2018.
5. Mignone P, Pio G, D'Elia D, Ceci M (2020) Exploiting Transfer Learning for the Reconstruction of the Human Gene Regulatory Network. *Bioinform.* **36**(5):1553-1561. <http://dx.doi.org/10.1093/bioinformatics/btz781>
6. Wang J, Chen Y, Hao S, Feng W, Shen Z (2017) Balanced distribution adaptation for transfer learning. In *IEEE International Conference on Data Mining (ICDM)*, New Orleans, LA, USA, 2017, pp. 1129-1134. <http://dx.doi.org/10.1109/ICDM.2017.150>
7. Zhang J, Li W, Ogunbona P (2017) Joint geometrical and statistical alignment for visual domain adaptation. In *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, Honolulu, HI, USA, 2017, pp. 5150-5158. <http://dx.doi.org/10.1109/CVPR.2017.547>