
This is an electronic reprint of the original article.
This reprint may differ from the original in pagination and typographic detail.

van Dijk, Aalt D J; Lähdesmäki, Harri; de Ridder, Dick; Rousu, Juho
Selected proceedings of Machine Learning in Systems Biology

Published in:
BMC Bioinformatics

DOI:
[10.1186/s12859-016-1305-1](https://doi.org/10.1186/s12859-016-1305-1)

Published: 13/12/2016

Document Version
Publisher's PDF, also known as Version of record

Published under the following license:
CC BY

Please cite the original version:
van Dijk, A. D. J., Lähdesmäki, H., de Ridder, D., & Rousu, J. (2016). Selected proceedings of Machine Learning in Systems Biology: MLSB 2016. *BMC Bioinformatics*, 17(Supplement 16). <https://doi.org/10.1186/s12859-016-1305-1>

This material is protected by copyright and other intellectual property rights, and duplication or sale of all or part of any of the repository collections is not permitted, except that material may be duplicated by you for your research use or educational purposes in electronic or print form. You must obtain permission for any other use. Electronic or print copies may not be offered, whether for sale or otherwise to anyone who is not an authorised user.

INTRODUCTION

Open Access



Selected proceedings of Machine Learning in Systems Biology: MLSB 2016

Aalt D. J. van Dijk^{1,2,3}, Harri Lähdesmäki⁴, Dick de Ridder³ and Juho Rousu^{4*}

From The 10th International Workshop on Machine Learning in Systems Biology (MLSB)
Den Haag, The Netherlands. 3-4 September 2016

Introduction

Biology is rapidly turning into an information science, thanks to enormous advances in the ability to observe the molecular properties of cells, organs and individuals. This wealth of data allows us to model molecular systems at an unprecedented level of detail and to start to understand the underlying biological mechanisms. The burgeoning field of systems biology creates a huge need for methods from machine learning, which find statistical dependencies and patterns in these large-scale datasets and use these to establish models of complex molecular systems. MLSB is a successful series of workshops that aims to provide a scientific forum for the exchange between researchers from Systems Biology and Machine Learning, to promote the exchange of ideas, interactions and collaborations between these communities.

MLSB started in 2007 and since 2008 has been co-located with major conferences in computational and systems biology (ECCB 2012, 2014; ISMB/ECCB 2011, 2013; ICSB 2010) or machine learning (ECML 2008–9, NIPS 2015), in order to engage the relevant wider communities. The workshop has constantly attracted around 80 participants or more, 2016 not being an exception: the workshop was fully booked, participant number only limited by the room capacity.

MLSB2016 took place as a two-day pre-conference workshop of the European Conference on Computational Biology, in the Hague, The Netherlands. The focus of the contributions to MLSB ranged from more methodological to more applied, and clearly demonstrated the use of machine learning to address biological questions. Selected submissions were invited based on the papers presented in the workshop. This supplement contains a reviewed selection of six full papers that cover a large panel of topics in Machine Learning devoted to Systems Biology.

Two of the manuscripts [1, 2] deal with the analysis of epigenomic marks. Lukauskas et al. [1] present an approach to cluster and visualize these marks. Their approach adaptively rescales genomic distances in order to enable clustering regions of interest with similar shapes. Park et al. [2] apply association rule mining in order to find differential combinatorial chromatin modification patterns.

Two additional papers describe aspects of (unsupervised) network reconstruction [3, 4]. Affeldt et al. [3] present a consensus method based on spectral decomposition. The basic idea here is to first identify related variables, and then in a second step perform multiple parallel local network reconstructions from which a global network is inferred. The second contribution related to network construction, Heinävaara et al. [4], describes aspects of L1-penalised sparse precision matrix estimation. L1-regularisation is often applied in network reconstruction, and this manuscript demonstrates that it is important to check whether the conditions of consistency are likely to be met by the dataset and the problem at hand. In addition to these two papers focussing on network reconstruction, a third contribution, Veríssimo et al. [5] also deals with networks, using network-based features for regularization in survival analysis. They propose a method that applies network centrality measures to constrain models where the outcome is patient survival and the features are genes. Finally, Gönen [6] presents a Bayesian multiple kernel learning algorithm, which trains a binary classifier with a sparse set of active gene sets using a sparsity-inducing prior. This method is subsequently generalized to a multi-task learning setting to model multiple related datasets conjointly.

All in all, the special issue reflects the depth and diversity of data analysis and modelling challenges that the field faces, and the variety of methods that are used to tackle them.

* Correspondence: juho.rousu@aalto.fi

⁴Department of Computer Science, Aalto University, 00076 Aalto, Finland
Full list of author information is available at the end of the article



Acknowledgements

We wish to thank the MLSB Programme Committee for their support.

Declaration

This article has been published as part of *BMC Bioinformatics* Volume 17 Supplement 16, 2016: Proceedings of the Tenth International Workshop on Machine Learning in Systems Biology (MLSB 2016). The full contents of the supplement are available online at <http://bmcbioinformatics.biomedcentral.com/articles/supplements/volume-17-supplement-16>.

This supplement contains a selected subset of papers presented at the workshop MLSB2016, Machine Learning in Systems Biology, The Hague, The Netherlands, September 3–4 2016.

Funding

The workshop was generously sponsored by contributions from the Dutch Organisation for Scientific Research (NWO), the Helsinki Institute of Information Technology (HIT) and five companies: BaseClear, Bayer, Enza Zaden, Philips and RijkZwaan.

Authors' contributions

All authors read and approved the final manuscript.

Competing interests

The authors declare that they have no competing interests.

Author details

¹Biometris, Wageningen University and Research, Droevendaalsesteeg 1, 6708 PB Wageningen, The Netherlands. ²Applied Bioinformatics, Wageningen University and Research, Droevendaalsesteeg 1, 6708 PB Wageningen, The Netherlands. ³Bioinformatics Group, Wageningen University and Research, Droevendaalsesteeg 1, 6708 PB Wageningen, The Netherlands. ⁴Department of Computer Science, Aalto University, 00076 Aalto, Finland.

Published: 13 December 2016

References

1. S Lukauskas, R Visintainer, G Schweikert, G Sanguinetti. DGW: an exploratory data analysis tool for clustering and visualisation of epigenomic marks. *BMC Bioinformatics*. 2016; 17(Suppl 16). doi:10.1186/s12859-016-1306-0.
2. SH Park, S-M Lee, Y-J Kim, S Kim. ChARM: Discovery of combinatorial chromatin modification patterns in hepatitis B virus X-transformed mouse liver cancer using association rule mining. *BMC Bioinformatics*. 2016; 17(Suppl 16). doi:10.1186/s12859-016-1307-z.
3. S Affeldt, N Sokolovska, E Prifti, J-D Zucker. Spectral Consensus Strategy for Accurate Reconstruction of Large Biological Networks. *BMC Bioinformatics*. 2016; 17(Suppl 16). doi:10.1186/s12859-016-1308-y.
4. O Heinävaara, J Leppä-Aho, J Corander and A Honkela. On the inconsistency of l1-penalised sparse precision matrix estimation. *BMC Bioinformatics*. 2016; 17(Suppl 16). doi:10.1186/s12859-016-1309-x.
5. A Veríssimo, A L. Oliveira, M-F Sagot and S Vinga. DegreeCox: a network-based regularization method for survival analysis. *BMC Bioinformatics*. 2016; 17(Suppl 16). doi:10.1186/s12859-016-1310-4.
6. M Gönen. Integrating gene set analysis and nonlinear predictive modeling of disease phenotypes using a Bayesian multitask formulation. *BMC Bioinformatics*. 2016; 17(Suppl 16). doi:10.1186/s12859-016-1311-3.

Submit your next manuscript to BioMed Central and we will help you at every step:

- We accept pre-submission inquiries
- Our selector tool helps you to find the most relevant journal
- We provide round the clock customer support
- Convenient online submission
- Thorough peer review
- Inclusion in PubMed and all major indexing services
- Maximum visibility for your research

Submit your manuscript at
www.biomedcentral.com/submit

