Eric Bach

Curriculum Vitae

Muusantori 5A 18, 00350 Helsinki, Finland
+358 40 5893344
eric.bach@aalto.fi

About me

I am a computer scientist and machine learning expert with strong knowledge in computational metabolomics.

I work quality- and solution-oriented with the attitude to provide the right computational tools for challenging problems in metabolomics.

I am a team player with effective communication skills and ready tell my opinion.

Keywords Knowledgable, Curios, Self-critical, Foresightful, Motivated, Friendly

Education

8/2016 – present **Doctoral Candidate**, *Aalto University*, Espoo, Finland.
In KEPACO bioinformatics research group under supervision of Prof. Juho Rousu
Topic: Machine Learning for Computational Metabolomics

4/2013 – 8/2016 **M.Sc. Computer Science**, *Friedrich-Schiller-Universität (FSU)*, Jena, Germany.
Major: Digital Image Processing, Minor: Machine Learning & Data-Mining

10/2015 – 7/2016 **Master's thesis**, *Aalto University*, Espoo.
Metabolite Identification using Magnitude-preserving Input Output Kernel Regression

10/2009 – 4/2013 **B.Sc. Applied Computer Science**, *FSU*, Jena.
Minor (40 ECTS): Computational Neuroscience, Studies encompassed 210 ECTS
Thesis: Adaption of a Semantic Segmentation Algorithm for Remote Sensing Data

Doctoral studies

Abstract I am developing machine learning (ML) methods for the automatized annotation of high-throughput untargeted metabolomics data arising in liquid chromatography (LC) tandem mass spectrometry (MS^2) experiments. My focus lays in the integration of orthogonal information, e.g. LC retention times and MS^2 spectra, to improve the molecule identification performance.

Methods On the machine learning side of my research I mainly utilise kernel methods, e.g. support vector machines. I develop novel approaches to integrate the orthogonal information sources in LC-MS^2 experiments. My tasks range from the mathematical description of my models till their practical implementation mainly using Python. On the metabolomics side, I aim to deeply understand the LC-MS^2 technology and measurement processes. I have a comprehensive knowledge of the data generation and the required processing, which allows me to map problems in metabolomics to ML frameworks.

Programming Projects related to Cheminformatics (selected)

`ROSVM` (`github.com/bachi55/rosvm`) Python package implementing a ranking support vector machine for **molecule retention order prediction** optimized using conditional gradient descent. The object orientated implementation is compatible with `sklearn` to allowing seamless integration of `ROSVM` to ML pipelines. Feature extraction for molecules is implemented using `rdkit`.

`msms_scorer` (`github.com/aalto-ics-kepaco/msms_rt_score_integration`) Library for the integration of MS^2 spectra and LC retention order scores aiding the **identification of small molecules** in metabolomics by molecular structure candidate ranking. The scores are integrated using a markov random field and rankings are provided using approximated marginal inference on random spanning tree ensembles. The library allows for efficient parallel inference.

`massbank2db` (`github.com/bachi55/massbank2db`) Python package to **extract data and meta-information from the online mass spectral library** MassBank. It implements a parser to extract information from MassBank entries as well as routines to organize the entries into sub-datasets with homogeneous meta-information. The parsed information is stored an SQLite database and enables ML practitioners to easily access MassBank for their research.

Relevant technical expertise

Machine Learning (selected)

Classical ML (>6 years) Kernel methods and structured prediction (metabolomics, mass spectrometry & molecule data), Random Forests (remote sensing and automotive applications), Gaussian Mixture Models (MRI image segmentation)

Deep Learning (>1 years) Convolutional Neural Networks (CNN) (visual object recognition), Graph-convolutional NN (molecule property prediction & representation learning)

Programming Languages (selected)

Python (>4 years) Machine learning (`sklearn`), Database administration (`pandas, sqlite3`), Molecule processing (`rdkit`), Scientific Computing (`numpy, scipy, numba`)

R (>6 years) Data visualization & organization (`ggplot, data.frame`), Processing molecular structures (`rcdk`), Statistical analyses

Matlab (>6 years) Machine Learning (kernel methods, graphical models), Prototyping

Other (selected)

SQLite (>3 years) Organising experimental results, unification of heterogeneous data source

Cheminformatics (>4 years) Molecule structure processing and visualization using `rdkit` & `CDK`, Feature extraction for ML purposes

Publications

Eric Bach, Simon Rogers, John Williamson, and Juho Rousu (2020). "Probabilistic framework for integration of mass spectrum and retention time information in small molecule identification". In: *Bioinformatics* 37.12, pp. 1724–1731

Sandor Szedmak and **Eric Bach** (2020). "On the generalization of Tanimoto-type kernels to real valued functions". In: *arXiv preprint arXiv:2007.05943*

Eric Bach, Sandor Szedmak, Céline Brouard, Sebastian Böcker, and Juho Rousu (2018). "Liquid-chromatography retention order prediction for metabolite identification". In: *Bioinformatics* 34.17, pp. i875–i883

Céline Brouard, **Eric Bach**, Sebastian Böcker, and Juho Rousu (2017). "Magnitude-Preserving Ranking for Structured Outputs". In: *Asian Conference on Machine Learning*, pp. 407–422

Björn Fröhlich, **Eric Bach**, Irene Walde, Sören Hese, Christiane Schmullius, and Joachim Denzler (2013). "Land cover classification of satellite images using contextual information". In: *ISPRS Annals of the Photogrammetry, Remote Sensing and Spatial Information Sciences* 3