

Tutorial at IJCNN'15

## **Learning in indefinite proximity spaces: Mathematical foundations, representations, and models**

### **Tutorial abstract**

Efficient learning of a data analysis task strongly depends on the data representation. Most methods rely on (symmetric) similarity or dissimilarity representations by means of *metric* inner products or distances, providing easy access to powerful mathematical formalisms like kernel or branch-and-bound approaches. Similarities and dissimilarities are however often naturally obtained by non-metric proximity measures which can not easily be handled by classical learning algorithms. In the last years major efforts have been undertaken to provide approaches which can either directly be used for such data or to make standard methods available for these type of data.

The tutorial provides a comprehensive overview for the field of learning with non-metric proximities. First we introduce the formalism used in non-metric spaces and motivate specific treatments for non-metric proximity data. Secondly we provide a systematization of the various approaches. For each category of approaches a comparative discussion is provided addressing complexity issues and generalization properties. We also address the problem of large scale proximity learning which is often overlooked in this context and of major importance to make the method relevant in practice. The discussed algorithms and concepts are in general applicable for proximity based clustering, one-class classification, classification, regression or embedding tasks. Various applications show the relevance of the discussed approaches, which provide a generic framework for multiple input formats. The goal of the tutorial is to give an overview about recent developments in this domain, covering in particular principled approaches as concerns learning in indefinite spaces and its mathematical foundations and extensions to large scale problems.

### **Covered material**

The following items represent the topics of the tutorial:

- Indefinite kernels and pseudo-Euclidean spaces: basic definition; algorithms for correcting indefinite kernels and dissimilarities (eigenspectrum and proxy approaches)
- Models for supervised learning in pseudo-Euclidean spaces and by general similarity functions
- Application examples from the biomedical domain and image processing
- Approximation strategies improving runtime, memory complexity and out-of-sample extension

## URL

Tutorial material will be available at the url:  
[http://www.cs.bham.ac.uk/~schleify/ijcnn15\\_tutorial.html](http://www.cs.bham.ac.uk/~schleify/ijcnn15_tutorial.html)

## Organizers

Peter Tino, University of Birmingham, UK,  
[p.tino@cs.bham.ac.uk](mailto:p.tino@cs.bham.ac.uk)  
<http://www.cs.bham.ac.uk/~ptx>

Frank-Michael Schleif, University of Birmingham, UK,  
[schleify@cs.bham.ac.uk](mailto:schleify@cs.bham.ac.uk)  
<http://www.cs.bham.ac.uk/~schleify>

*Peter Tino* is Professor of Complex and Adaptive Systems at the School of Computer Science, University of Birmingham, UK. He held a Fulbright Fellowship (at NEC Research Institute, Princeton, USA) and a UK-Hong Kong Fellowship for Excellence. Peter is a recipient of three IEEE Computational Intelligence Society Outstanding Paper of the Year awards in IEEE Transactions on Neural Networks (1998, 2011) and IEEE Transactions on Evolutionary Computation (2010). He serves on Editorial Boards of several journals (IEEE Transactions on Neural Networks and Learning Systems (IEEE CIS, since 2013), Scientific Reports (Nature Publishing, since 2011) and Neural Processing Letters (Springer, since 2007)). He has (co-)chaired programme committee of 4 international conferences, while being member on programme committees of more than 90 international conferences. Peter's scientific interests include machine learning, dynamical systems, evolutionary computation, complex systems, probabilistic modelling and statistical pattern recognition.

*Frank-Michael Schleif* received his Ph.D. in Computer Science from the University of Clausthal, Germany, in 2006 and his *venia legendi* in Applied Computer Science in 2013 from the University of Bielefeld, Germany. From 2004-2006 he was working for the R&D department at Bruker Biosciences. From 2006 to 2009 he was a research assistant in the research group of computational intelligence at the University of Leipzig working on multiple bioinformatic projects. In 2010 he joined the Chair of Theoretical Computer Science at the University of Bielefeld and did research in multiple projects in machine learning and bioinformatics. Since 2014 he is a member of the University of Birmingham, UK as a Marie Curie Fellow and PI of the project Probabilistic Models in Pseudo-Euclidean Spaces. His areas of expertise include machine learning, signal processing, data analysis and bioinformatics. Several long term research stays have taken him to UK, the USA, the Netherlands and Japan. He is co-editor of the Machine Learning Reports and reviewer for multiple journals and conferences in the field of machine learning and computational intelligence. He is a founding member of the Institute of Computational Intelligence and Intelligent Data Analysis (CIID) e.V. (Mittweida, Germany), a member of the IEEE-CIS, IEEE-SPS, the GI, the DAGM and secretary of the German chapter of the ENNS (GNNS). He is coauthor of more than 100 papers in international journals and conferences on different aspects of Computational Intelligence, most of which can be retrieved from <http://www.cs.bham.ac.uk/~schleify/>