Seminar

# ARTIST: Artificial Intelligence for Spectroscopy

**Patrick Rinke***

*Aalto University*

**Time: 4: 00 pm, May. 27, 2019 (Monday)**

**2019 5 27 4:00**

**Venue: Room W563, Physics building, Peking University**

**563**

## Abstract

For the study of molecules and materials, conventional theoretical and experimental spectroscopies are well established in the natural sciences, but they are slow and ex- pensive. Our objective is to launch a new era of artificial intelligence (AI) enhanced spectroscopy that learns from the plethora of already available experimental and theoretical spectroscopy data. Once trained, the AI can make predictions of spec- tra instantly and at no further cost. In this new paradigm, AI spectroscopy would complement conventional theoretical and experimental spectroscopy to greatly accel- erate the spectroscopic analysis of materials, make predictions for novel and hitherto uncharacterized materials, and discover entirely new materials.

In this presentation, I will introduce the two AI approaches we have used to learn spectroscopic properties: kernel ridge regression (KRR) and deep neural networks (NN). The models are trained and validated on data generated by density-functional theory calculations for three different molecular data sets: QM9 [1], a well-established small molecule benchmark datasets, a set of amino acid and dipeptide conformers [2] and a set of crystal-forming, optically-active molecules [3]. The molecules are rep- resented by simple, easily attainable numerical descriptors based on nuclear charges and cartesian coordinates [4,5]. The complexity of the molecular descriptor and the diversity of the data sets turn out to be crucial for the learning success, as I will demonstrate for KRR [6]. I will then show, how we can learn spectra (i.e. continuous target quantities) with NNs. We design and test three different NN architectures: multilayer perceptron (MLP), convolutional neural network (CNN) and deep tensor neural network (DTNN). Already the MLP is able to learn spectra, but the learn- ing quality improves significantly for the CNN and reaches its best performance for the DTNN [7]. Both CNN and DTNN capture even small nuances in the spectral shape. In a showcase application of this method, the structures of 10k previously unseen organic molecules are scanned and instant spectra predictions are obtained to identify molecules for potential applications [7].

∗ This work was performed in collaboration with A. Stuke, K. Ghosh, L. Himanen, M. Rupp, C. Kunkel, P. B. Jørgensen, M. N. Schmidt, M. Todorović, and A. Vehtari

[1] R. Ramakrishnan, P. O. Dral, M. Rupp, and A. von Lilienfeld, Sci. Data 1, 140022 (2014)
[2] M. Ropo, M. Schneider, C. Baldauf, and V. Blum, Sci. Data 3 (2016)
[3] C. Schober, K. Reuter, and H. Oberhofer, J. Phys. Chem. Lett. 7, 3973 (2016)
[4] M. Rupp et al. Phys. Rev. Lett. 108, 058301 (2012)
[5] H. Huo and M. Rupp, arXiv:1704.06439
[6] A. Stuke, M. Todorović, C. Kunkel, M. Rupp, K. Ghosh, L. Himanen, and P. Rinke, arXiv:1812.08576
[7] K. Ghosh, A. Stuke, M.