Deep learning for spectroscopic data analysis: an evaluation.

G.J. Postma¹, J. Acquarelli², T. van Laarhoven², E. Marchiori², L.M.C. Buydens¹, J.J. Jansen¹
1: Radboud University Nijmegen, Institute for Molecules and Materials
2: Radboud University Nijmegen, Institute for Computing and Information Science
g.j.postma@science.ru.nl

Over the last years, deep learning has gained popularity and relevance because of its effectiveness in classifying images and other multimedia data. Deep learning often uses Convolutional Neural Networks (CNN), composed of a large number of convolutional layers. These layers are the true novelty of CNNs; they facilitate the use of CNN for spectroscopic data sets which frequently are of limited size (‘fat’ data). Moreover, convolution with kernels of a fixed size allows exploitation of the spectral feature locality of spectroscopic data. One important drawback about ANNs from a chemometrics perspective is the limited interpretability of the predictive features in the data. We therefore have developed an embedded feature selection method to provide interpretation of the trained network. We also show that the purely data-driven analysis by CNNs may benefit from analytical chemical domain knowledge, through chemometric data preprocessing.

We have investigated the applicability of CNNs for a wide range of chemometrics case studies, such as the classification of beers, tablets and wines based on Raman or FTIR spectra, the prediction of brain tumours and Alzheimer disease based on Magnetic Resonance Spectroscopic Imaging data, and the classification of hyperspectral images. In case of the latter dependencies on the size of the training sets and methods to enhance the size of the training set were also investigated.

We compared performances of the CNNs with those of the more classical chemometric methods like PLS-DA. In case of hyper spectral images also with a true deep convolutional neural network. We show how CNNs often outperform such more conventional methods, such that embedding CNNs in the knowledge-supported and interpretation-driven approach of chemometrics may be very beneficial.

References: