



Contribution ID: 251

Type: **Poster (Chemistry SIG)**

Unsupervised learning used for separation and sub-clustering of cancer samples

Deep learning techniques show significant promise in improving the results of unsupervised learning algorithms such as the Self Organising Map (SOM) which allow for clustering and visualisation of datasets. One such approach is the use of autoencoders[1] which aim to encode a dataset in a way which provides an improved representation of the data. The Growing Hierarchical Self-Organising Representation Map (GHSORM) is a model fusing the denoising autoencoder and the Growing Hierarchical Self-Organising Map[2]. The use of the GHSOM proves to be helpful in sub-grouping clusters, which aren't fully separated by a single SOM, in a compact form. There are a number of potential applications for this model however the focus of this project is making use of genetic datasets to cluster cancer samples. It is hoped that the results achieved can be used as a proof of concept that, using a limited gene set, gene expression data can be used to classify and sub-classify cancer samples. The combined effect of the computational expense of training the denoising autoencoder as a pre-processor and the use of gene expression resulting in a large feature space necessitates high performance computing. The embarrassingly parallel nature of the denoising autoencoder algorithm provides an opportunity for significant speed-up.

Presenter Biography

Primary author: COULSON, Matthew (UCT SCRUB Lab)

Co-authors: WINBERG, Simon; NAIDOO, Kevin J. (University of Cape Town)

Presenter: COULSON, Matthew (UCT SCRUB Lab)

Session Classification: Chemistry and Material Science SIG Seminar

Track Classification: Chemistry and Material Science SIG Seminar