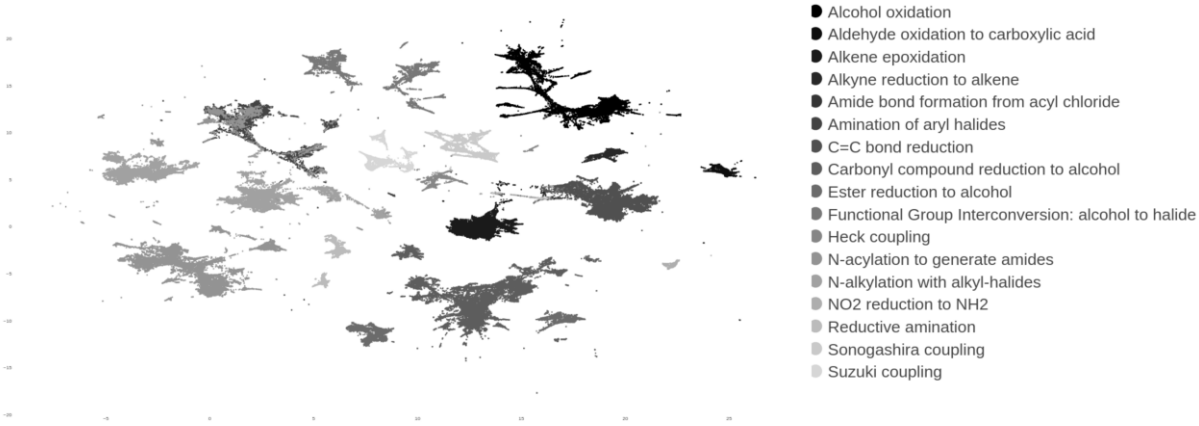
A Multi-Scale Chemical Reaction Database Visualisation

Tarik Boumaza[1,2], Thomas Galeandro-Diamant[1], Hamamache Kheddouci[2]

[1]*DeepMatter Group ; Leeds (United Kingdom).*

[2]*LIRIS Laboratory ; Villeurbanne (France).*

Data is at the core of machine learning applications. Visualisation is a powerful way to understand where gaps and outliers lie within reaction databases [1] and to identify ways of improving or complementing these databases. This poster introduces an application tailored for chemical reaction database visualisation. Using ICClassify [2], we aggregate reactions that share the same reaction centres. We then convert the reaction, given by a SMILES, to a numerical vector using the BERT fingerprint [3], which focuses on the reaction centres. These fingerprints are then processed for dimensionality reduction via UMAP [4], leading to a 2D interactive visualisation that groups reactions by class. This application allows users to explore a whole database (currently SPRESI and Pistachio), or explore any reaction class of interest (such as the Suzuki reaction), examining multiple parameters, including the structure of reactants or products, and the reaction conditions (solvent, catalyst, temperature). While this aspect of the project is still under development, preliminary results show promising potential in enhancing our understanding of reaction databases. The second significant feature is the ability to compare reaction databases through a heatmap, visualising the density of points in a 2D space, representing the user selected reaction classes. Packaged as a web application, this methodology offers a user-friendly platform for chemists and researchers to visualise and analyse chemical reaction databases.



Bibliography :

[1] M.Andronov; M.V.Fedorov; S.Sosnin. Exploring chemical reaction space with reaction difference fingerprints and parametric t-SNE. ACS Omega (2021) 6(45):30743-30751.
[2] H.Kraut; J.Eiblmaier; G.Grethe; P.Löw; H.Matuszczyk; H.Saller. Algorithm for reaction classification. Journal of Chemical Information and Modeling (2016) 53(11):2884–2895.
[3] P.Schwaller; D.Probst; A.Vaucher; V.Nair; D.Kreutter; T.Laino; J-L.Reymond. Mapping the space of chemical reactions using attention-based neural networks. Nature Machine Intelligence (2021) 3:1–9.
[4] L.McInnes; J.Healy. UMAP: Uniform Manifold Approximation and Projection for dimension reduction. 02 2018.