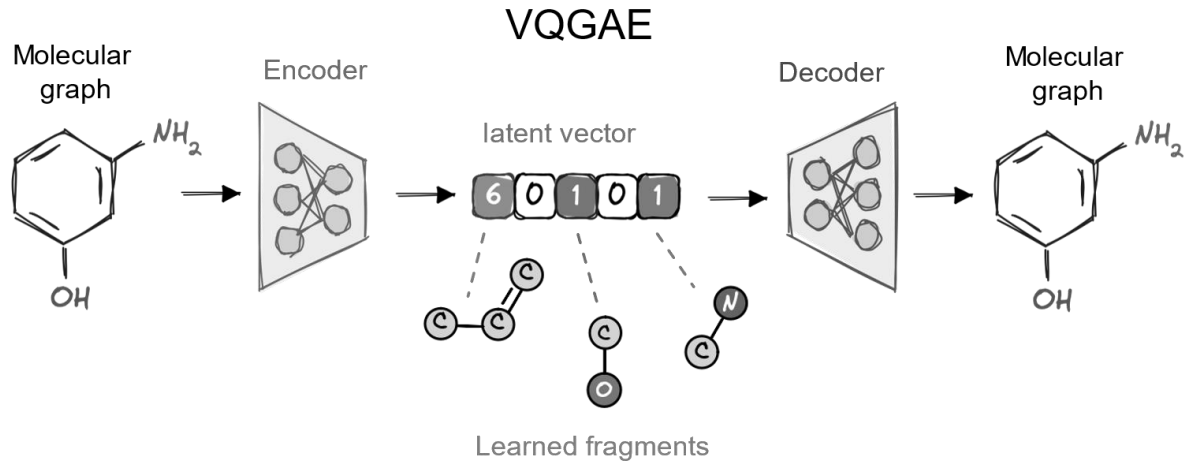
Construction of order-independent molecular fragments space with vector quantised graph autoencoder

Tagir Akhmetshin[1], Arkadii Lin[1], Timur Madzhidov[2], Alexandre Varnek[1]

[1]*Laboratory of Chemoinformatics, UMR 7140, CNRS University of Strasbourg, 4 rue Blaise Pascal, 67000 Strasbourg, France*

[2]*Chemistry Solutions, Elsevier Ltd, Oxford, United Kingdom*

In de novo molecular design, the Vector Quantised Graph AutoEncoder (VQGAE) [1] is a new generative architecture that addresses whether autoencoders can autonomously learn and use molecular fragments during training. By combining a vector quantization procedure with a graph convolution network (GCN), VQGAE generates fragment vectors that replace atom vectors, ensuring permutation invariance and improving similarity ranking performance. Trained on the ChEMBL v.27 dataset, VQGAE achieved a reconstruction rate of 80% and demonstrated that its latent vectors are unbiased by atom order and effective in predicting molecular properties. QSAR benchmarking against conventional fragment descriptors and other autoencoders confirmed VQGAE's ability to extract meaningful structural features. Furthermore, its generative capability was validated in the inverse QSAR design of tubulin inhibitors, producing structures with favourable synthetic accessibility and drug-likeness scores, demonstrating VQGAE's potential in molecular optimisation.



Bibliography :

[1] T. Akhmetshin, A. Lin, T. Madzhidov, A. Varnek. ChemRxiv. (2023). doi:10.26434/chemrxiv-2023-5zmvw.