Improving Fragment-Based Deep Molecular Generative Models

Panukorn Taleongpong[1], Brooks Paige[1]

[1]Department of Computer Science, University College London

Overview

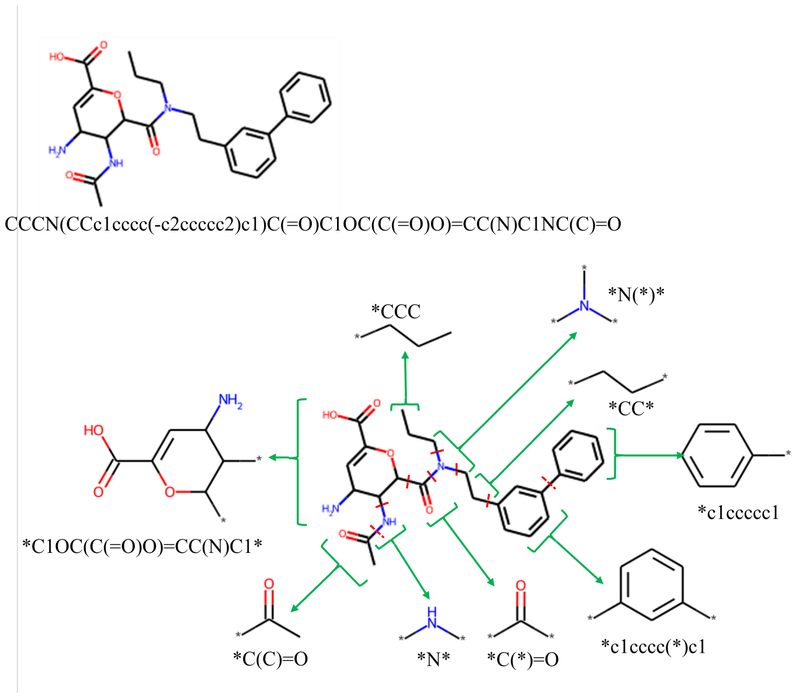


Figure 1: Example of fragments derived from a molecule using the BRICS algorithm. The asterisks represent dummy atoms, and the red dotted lines denote the 'cuts' in the molecule.

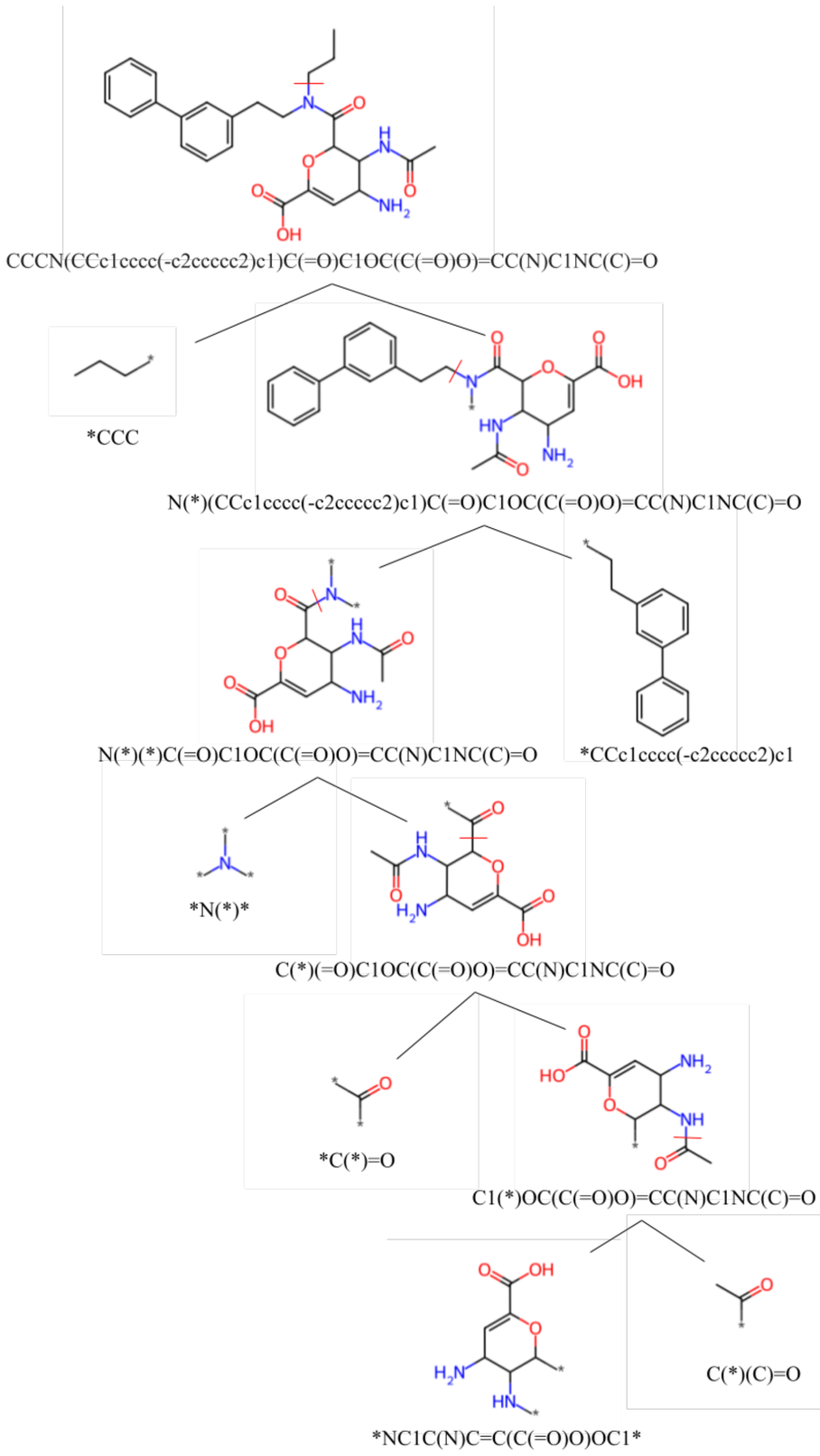
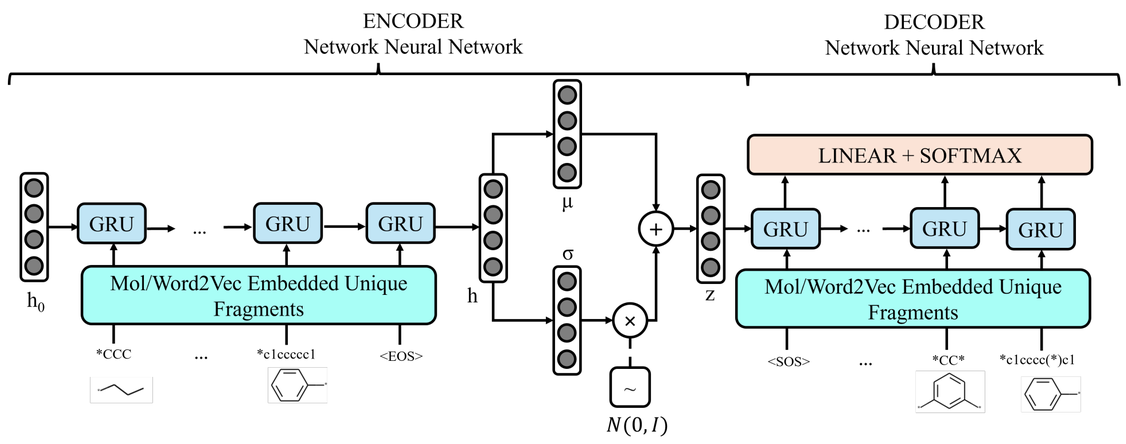


Figure 2: Example of fragments derived from a molecule using Breadth-First Fragmentation algorithm. The asterisks represent dummy atoms, and the red dotted lines denote the 'cuts' in each recursion.

1. Deep molecular generative models are promising in exploring the molecular space, estimated to be 10^{60}.
2. We propose a novel fragmentation and fragment reconstruction algorithm that produces fragments much more efficiently whilst guaranteeing the reconstruction of the original molecule.
3. Our model demonstrates superior performance metrics of the generated molecules by simply applying a factor to the sampling variance rather than using low-frequency masking [1].

Model



Results

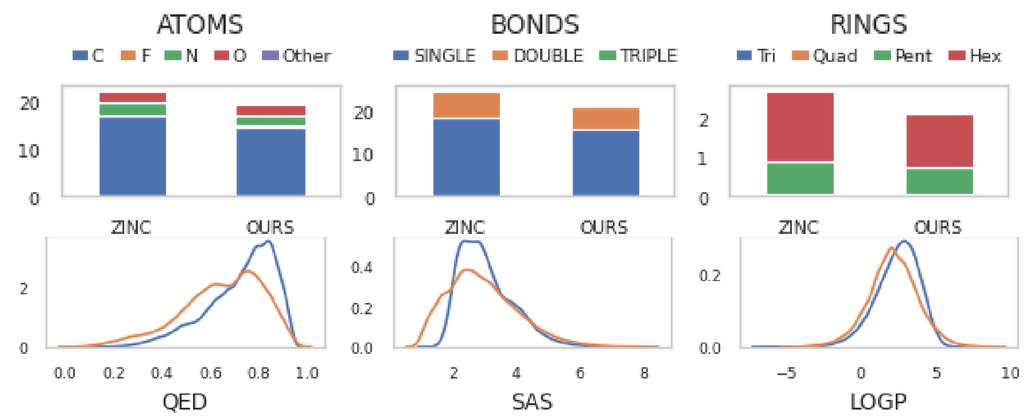


Figure 3: Top: Statistics of sampled molecular characteristics. Bottom: Density of sampled molecular properties where the blue line corresponds to ZINC. Our model: c = 25, Mol2Vec Embedding.

Model	Valid	Novel	Unique	Sample Rate (ms/mol)
Word2Vec Embedding				
Podda	1	0.992	0.460	
Mol2Vec Embedding				
Ours (c = 20)	1	0.997	0.661	123
Ours (c = 25)	**1**	**0.997**	**0.793**	**211**
Low Frequency Mask Sampling				
Podda (LFM)	1	0.995	0.998	

Table 1: Results of 20,000 molecules sampled

References

[1] Marco Podda, Davide Bacciu, and Alessio Micheli. A Deep Generative Model for Fragment-Based Molecule Generation. In *Proceedings of the Twenty Third International Conference on Artificial Intelligence and Statistics*, pages 2240–2250. PMLR, June 2020. ISSN: 2640-3498.