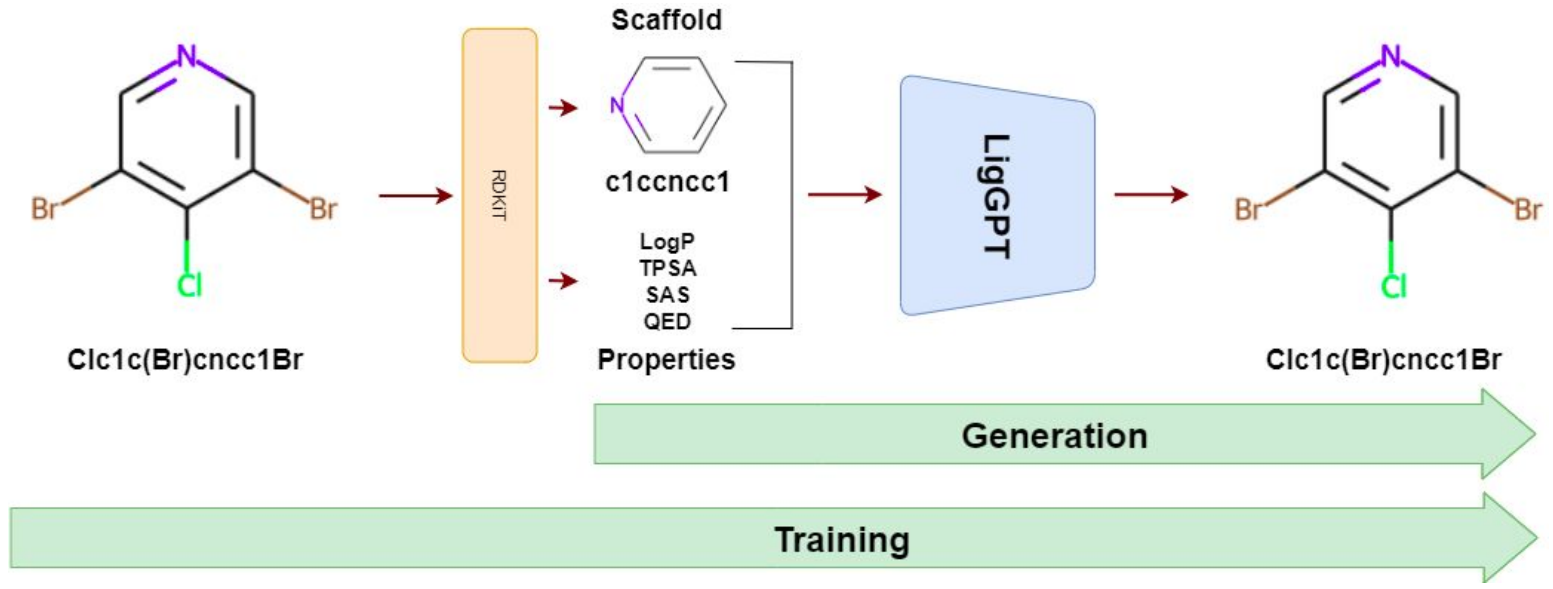
INTERNATIONAL INSTITUTE OF INFORMATION TECHNOLOGY
HYDERABAD

R&D SHOWCASE 2021
Technology, Social Impact

LigGPT: Molecular Generation using a Transformer-Decoder model

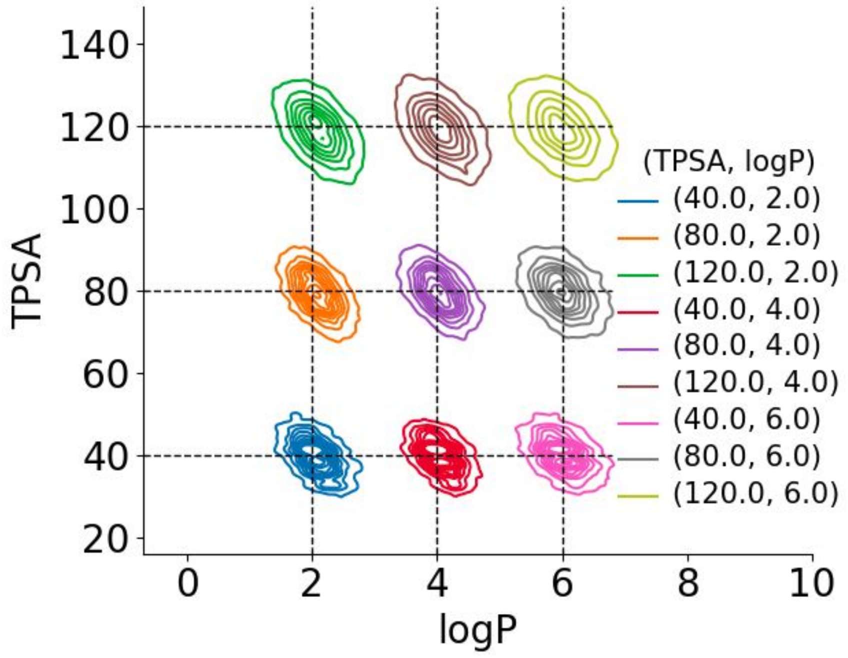
ABSTRACTS

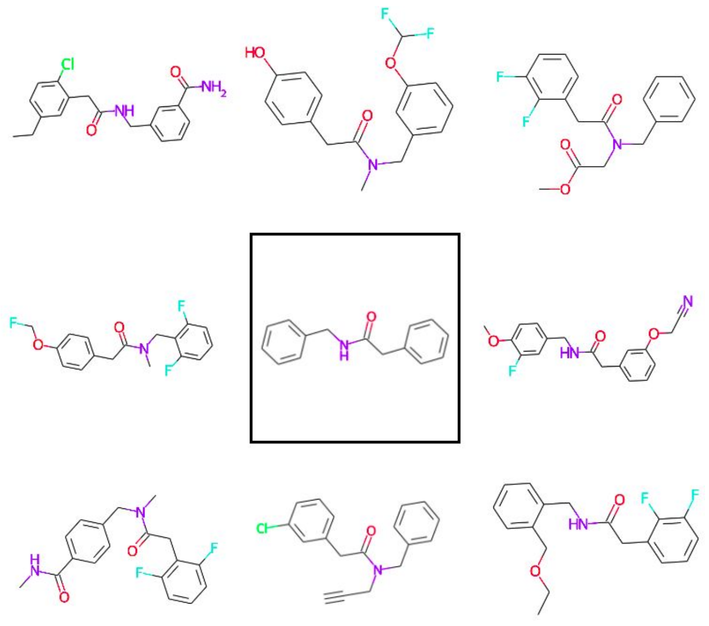
The representation of molecules in SMILES notation as a string of characters enables the usage of state-of-the-art models in Natural Language Processing, such as the Transformers, for drug discovery. We propose an interpretable transformer decoder for conditional molecular generation. Our proposed model - LigGPT - is capable of generating valid, unique and novel molecules having desired values of physicochemical properties such as logP, QED, TPSA, SAS while also maintaining the underlying scaffold structure.



OBJECTIVE

It has been postulated that the total number of potential drug like candidates can range from 10^{23} to 10^{60} molecules. However, only about 10^{8} molecules have been synthesized at least once. This calls for generative models that can help traverse that space. Moreover, many chemical and biological processes require the molecules to have certain values of physicochemical properties, thus it is necessary for generation to be conditional. So, the objective was to build an interpretable state-of-the-art model for conditional generation of molecules.





METHOD

We use two benchmark datasets, MOSES and Guacamol for training and evaluation.Each molecule is first represented as SMILES. SMILES tokenizer is then used to tokenize it. Addition of embeddings of SMILES tokens, position tokens and segment tokens is provided as input to the model and the model is then trained on the next token prediction task. For conditional training, we concatenate the property vector at the start of the SMILES vector of the molecule and ensure that every token can attend to the conditional vector.

Results

- LigGPT has the best performance on the Guacamol dataset and comparable performance on the MOSES dataset.
- LigGPT can generate molecules satisfying multiple properties simultaneously.
- It can also generate molecules similar to the desired scaffold .

Authors: Viraj Bagal, Rishal Aggarwal, P.K. Vinod, U. Deva Priyakumar

Research Center Name: HAI Research Center