O 10

Solubility prediction by deep learning of quantum information

Tonglei Li

Department of Industrial and Physical Pharmacy at Purdue University, USA

It takes more than a decade and billions of dollars to put a new drug product on the market. The essence of drug discovery and development is to characterize molecular interactions and realize wanted interactions by molecular engineering. The astronomical drug development process is attested by dismal failure rates moving from hits to leads and from leads to final products. The recent advance in artificial intelligence (AI) software and hardware facilitates the application of machine and deep learning (ML/DL) to mine molecular data. The power of AI has yet to be fully unleashed in accelerating drug development timelines and cutting costs. A major hurdle is molecular description for learning. Because the electronic structures of a molecule bear the ground truth of molecular interactions, it is advantageous to utilize electronic attributes in AI models. We have recently developed a novel route of capturing the quantum information (QI) of a molecule, coined Manifold Embedding of Molecular Surface (MEMS). MEMS is readily computable and can be further featurized as input for machine learning. We have applied to predict water solubilities of small molecules based on the input of training data published in the recent Solubility Challenges. Our approach appeared to outperform many reported ML/DL efforts of solubility prediction. Our study further indicated the critical roles of training data and model architecture, in addition to the input representation of molecules.