DEEP LEARNING APPROACH TO NORMAL BOILING POINT PREDICTION FROM EXPERIMENTAL DATA WITH *uncertainty*

Hormazabal Rodrigo, 양대륙[1,†], 강정원[1], 장지웅[2]
고려대학교; [1]고려대학교 화공생명공학과; [2]금오공과대학교
(dryang@korea.ac.kr[†])

Physical properties measurements are normally time consuming and expensive, which is the first limiting factor to effective process design when dealing with new compounds. Many approaches have been used to develop predictive models from molecular structural information. However, since molecular interactions at a quantum mechanical level are hard to quantify and yet to be fully understood, these deterministic approaches fail to work on a wide range of compounds and present rather big deviations for particular types of molecules. Also, methods based on quantum calculations often take a long time to be generated. We propose a machine learning centered pipeline that involves a supervised learning molecular classification step followed by a Graph Convolutional Network(GNU) in order to predict Normal Boiling Point Temperatures from structural information. Particularly, our approach differs from other applications to this task, by the fact that the training data utilized includes uncertainty values for every particular measurement. To leverage this information, we present a modified loss function in order to account for the "trustability" of each data point and its ground truth value.