Development of new algorithm for the classification of chemical compound using SMILES

강성신, 강정원*
고려대학교
(jwkang@korea.ac.kr*)

A lot of experimental data sets captured from published literatures are collected in a database. Usually, these data sets in the database are classified by only data type such as critical property and saturated property. However, group contribution model such as UNIFAC needs the data classification by functional group of chemical compound for calculation of the model parameters. In this study, new algorithm for the classification is developed matching patterns of SMILES (Simplified Molecular-Input Line-Entry System) string. The algorithm is composed of three classes; atomic class, main class and sub class. The atomic class and the main class are determined by the number of specific atoms and the functional group, respectively. Additionally, sub class shows additional information such as unsaturated bond and multifunctional groups. The proposed algorithm is available to most chemical compounds and in this study 6178 chemical compounds are classified into 14 atomic classes, 40 main classes and 207 sub classes.