QSPR modeling approaches representing conformations

Tomoyuki Miyao
Data-driven Chemistry Laboratory, Division of Materials Science,
Graduate School of Science and Technology
Data Science Center
Nara Institute of Science and Technology

Molecular representations properly handling conformations are prerequisites for property (activity) prediction models. Since thermodynamic properties are determined from a set of locally minimum conformations, multi-instance learning is a natural approach for considering multiple conformations. Nevertheless, property prediction models that rely exclusively on the structural formula or molecular graph sometimes show superior prediction accuracies than those using three-dimensional molecular representations and neural network models taking molecular conformations as input. One reason for not deriving consistent conclusions regarding three or two-dimensional molecular representations is a lack of appropriate benchmark datasets. To evaluate methods and representations in terms of correctly handling the conformation for molecular prediction tasks, the benchmark datasets should contain diverse conformers, the target property should be sensitive to conformation even for a single compound, and the conformer (a set of conformers) deriving the property value for a compound should be known. Herein, we constructed quantum chemistry-related datasets satisfying the three criteria above and conducted a series of property prediction tasks using various molecular representations. In this talk, I would like to share our findings on molecular representations conveying meaningful three-dimensional structural information for property prediction models.