C-2-13

Toward Crowdsourcing Evaluation of Synthetic Accessibility

Yukino Baba[1]
yukino_baba@mist.i.u-tokyo.ac.jp

Tetsu Isomura[2]
isomura.tetsu@mx.kaiteki-institute.com

Hisashi Kashima[1]
kashima@mist.i.u-tokyo.ac.jp

Hirofumi Nakano[3]
h-nakano@mol.f.u-tokyo.ac.jp

[1] Department of Mathematical Informatics, The University of Tokyo, 7-3-1 Hongo, Bunkyo-ku, Tokyo 113-8656, Japan
[2] The KAITEKI Institute, Inc., 1-1-1 Marunouchi, Chiyoda-ku, Tokyo 100-8251, Japan
[3] Open Innovation Center for Drug Discovery, The University of Tokyo, 7-3-1 Hongo, Bunkyo-ku, Tokyo 113-0033, Japan

Keywords: Synthetic accessibility, Crowdsourcing, Human computation

A rapid method for the assessment of synthetic accessibility for a vast number of chemical compounds is expected to bring about a breakthrough in the drug discovery. Although several computational methods have been proposed for estimating the synthetic accessibility [1-3], they do not meet the requirements of both quick computation time and high estimation accuracy. In the field of computer science, *crowdsourcing* has received considerable attention for being an efficient and scalable approach for solving complicated tasks that are difficult for computers but relatively easier for humans. Using crowdsourcing platforms such as Amazon Mechanical Turk [5], one can easily outsource various professional and non-professional tasks to a large group of people online. Jobs on crowdsourcing are typically decomposed into small and simple tasks that can be solved in parallel by different workers.

As the first step toward creating a crowdsourcing platform where the synthetic accessibility of a number of compounds is assessed by people with different levels of expertise and backgrounds, we investigate the consistency among experienced medicinal chemists in assessing synthetic accessibility. Unlike previous studies [2, 4], we prepared the following three compound groups: Already existing compounds, ones having substructures that are hard to be synthesized, and ones having novel substructures. We picked ten compounds from each group and asked the experienced medicinal chemists to rate the synthetic accessibility of each compound on a five-point scale within several minutes. Results show that there is a high correlation between the assessments given by a subset of the chemists.

[1] Baber, J. C., Feher, M.: Predicting Synthetic Accessibility: Application in Drug Discovery and Development. *Mini Reviews in Medicinal Chemistry*, 4:681-692, 2004.
[2] Boda, K., Seidel, T., Gasteiger, J.: Structure and Reaction based Evaluation of Synthetic Accessibility. *Journal of Computer-Aided Molecular Design*, 21:311-325, 2007.
[3] Ertl, P., Schuffenhauer, A.: Estimation of synthetic accessibility score of drug-like molecules based on molecular complexity and fragment contributions, *Journal of Cheminformatics*, 1:8, 2009.
[4] Takaoka, Y., Endo, Y., Yamanobe, S., Kakinuma, H., Okubo, T., Shimazaki, Y., Ota, T., Sumiya, S., Yoshikawa, K.: Development of a Method for Evaluating Drug-Likeness and Ease of Synthesis Using a Data Set in Which Compounds Are Assigned Scores Based on Chemists' Intuition. *Journal of chemical information and computer sciences*, 43:1269-1275, 2003.
[5] http://mturk.com