Spiral Mining using Attributes from 3D Molecular Structures

Takashi Okada, Masumi Yamakawa, and Hirotaka Niitsuma

Abstract: Active responses from analysis play an essential role in the knowledge discovery of SAR (structure activity relationship) from drug data. Experts often think of hypotheses, and they want to reflect these ideas to the attribute generation and selection process, Authors have analyzed SAR of dopamine agonists and antagonists using the cascade model. The presence or absence of linear fragments in molecules constitute the core attributes in the mining. In this paper, we generated attributes indicating the presence of hydrogen-bonds from 3D coordinates of molecules. Various improvements on the fragment expressions are also introduces following the suggestions given by chemists. Attribute selection from the generated fragments is another key step of the mining. Close interactions between chemists and system developers has enabled spiral mining, where the results of analysis are incorporated to the development of new functions in the mining system. All these efforts has led to the success of SAR mining.