

**РАЗРАБОТКА ПРОГРАММНОГО ПАКЕТА NFRAG,
ПРЕДНАЗНАЧЕННОГО ДЛЯ РАСЧЕТА ОБОБЩЕННЫХ
МОЛЕКУЛЯРНЫХ ДЕСКРИПТОРОВ НА ЯЗЫКЕ
HASKELL**

Соснин Сергей Борисович

Младший научный сотрудник

Институт физиологически активных веществ РАН, Черногловка, Россия

E-mail: serg.sosnin@gmail.com

Методы моделирования количественных соотношений структура/активность и структура/свойство (Quantitative Structure-Activity/Property Relationships, QSAR/QSPR) приобретают все большую популярность в задачах поиска и отбора новых перспективных структур лекарственных соединений. Одним из интересных вариантов данного метода является т.н. «фрагментный подход», в рамках которого исходные молекулярные структуры представляются в виде ненаправленных взвешенных, помеченных графов. На базе исходных молекулярных графов производится расчет «фрагментных дескрипторов», которые отражают присутствие или количество определенных структурных фрагментов (подграфов) в молекулярном графе, соответствующем структуре соединения. Далее выполняется построение предсказательных моделей при помощи методов статистического машинного обучения.

Однако в настоящее время все существующие решения в области генерации фрагментных дескрипторов имеют существенные ограничения: использование достаточно простых фрагментов ограниченного размера (линейные цепочки, иногда также простые циклы и разветвления), жестко заданные схемы генерации и обобщения фрагментных дескрипторов, что в ряде случаев затрудняет построение моделей высокого качества и их интерпретацию. Вследствие вышесказанного разработка программного обеспечения для генерации множества обобщенных фрагментных дескрипторов представляется актуальной задачей. Вместе с этим использование парадигмы «программирования без изменяемого состояния» в рамках функционального подхода в ряде случаев позволяет упростить проектирование и реализацию алгоритмов, распараллеливание вычислений, а также облегчить тестирование программ. Мы поставили задачу разработки ядра упрощенной библиотеки хемоинформатики, на ее основе создание программного обеспечения для генерации обобщенных фрагментных дескрипторов с использованием функционального языка

программирования Haskell (в полной мере соответствующего парадигме «программирование без изменяемого состояния»). Также, нашей задачей являлся анализ преимуществ и недостатков разработки алгоритмов и программного обеспечения для нужд хемоинформатики с использованием чистого функционального подхода. По результатам работы был создан программный пакет HFRag, полностью написанный на чистом функциональном языке программирования Haskell. Для представления графов в форме удобной для реализации функциональных алгоритмов нами используется библиотека FGL (Functional Graph Library). Мы реализовали алгоритм поиска множества всех связанных подграфов, ограниченных сверху по количеству вершин (алгоритм Рюкера [1]), а также алгоритмы нахождения изоморфных подграфов методом Ульмана и VF2, произвели сравнение производительности реализаций. Оценка корректности работы алгоритмов проводилась при помощи метода комбинаторного тестирования с использованием подхода, предлагаемого библиотекой QuickCheck. С использованием библиотеки комбинаторного разбора Parsec реализованы методы разбора химических структур формата SMILES (Simplified Molecular Input Line Entry Specification), а также языка шаблонов химических структур SMARTS. Показана целесообразность применения парадигмы комбинаторного разбора для построения парсеров химических структур и шаблонов химических структур данных форматов. С использованием программного пакета HFRag построена модель прогноза мутагенной активности органических соединений.

Литература

1. Rucker G., Rucker C. // MATCH Commun. Math. Comput. Chem. 2000. N 41 P. 145-149.