

ТОПОЛОГИЧЕСКИЕ ДЕСКРИПТОРЫ В ЗАДАЧЕ «СТРУКТУРА-СВОЙСТВО» ДЛЯ ПРОГНОЗИРОВАНИЯ БИОЛОГИЧЕСКОЙ АКТИВНОСТИ НА ОБОБЩЕННЫХ ДЕРЕВЬЯХ РЕШЕНИЙ

Научный руководитель – Кумсков Михаил Иванович

Васильева Варвара Олеговна

Студент (специалист)

Московский государственный университет имени М.В.Ломоносова,
Механико-математический факультет, Кафедра вычислительной математики, Москва,
Россия

E-mail: var-bara@mail.ru

Рассмотрена задача QSAR-моделирования, описана постановка задачи, которая складывается из двух самостоятельных подзадач:

• Задача построения матрицы Молекула-Дескриптор, • Задача построения модели, предсказывающей численное значение свойства для нового молекулярного графа (М-графа). Задача QSAR-моделирования поставлена как поиск зависимости между столбцами матрицы Молекула-Дескрипторы целевым признаком. Представлен способ топологического построения дескрипторов и его вариации на основе маркирования вершин молекулярных графов. Целью работы является исследование Метода Группового Учета Аргументов в качестве алгоритма прогнозирования свойств соединений и выявление его достоинств и недостатков. В экспериментальной части получены результаты работы МГУА в задаче классификации на различных выборках (BZR, COX-2, ER_LIT) с различными параметрами алгоритма. Проведены эксперименты на сбалансированных выборках (прогнозирование на всей выборке) и на несбалансированных (прогнозирование в кластерах). Были рассмотрены различные алгоритмы кластер-анализа, среди которых k-means (показывал хороший результат, но был неустойчивым) и DBSCAN(устойчивый, всегда получался хотя бы один “хороший” кластер).

В дальнейшем планируется реализовать модифицированный МГУА. Также планируется рассмотреть выборки со значительно большим числом молекул и применить описанные методы на них.

Исследование выполнено при поддержке РФФИ - проект 19-07-00752 и Междисциплинарной научно-образовательной школы Московского университета «Мозг, когнитивные системы, искусственный интеллект»

Литература

1. Кумсков М.И. "Методология прогнозирования свойств химических соединений и ее программная реализация". Автореферат докторской диссертации. М., Вычислительный центр РАН. <https://istina.msu.ru/dissertation/3541505/> 2. «Прогнозирование свойств молекулярных графов на основе RBF нейронных сетей» - Сборник работ под ред. М.И.Кумскова, <https://istina.msu.ru/collections/354167635/> 3. Свитанько И.В. «Моделирование в направленном синтезе веществ с заданными свойствами. - Автореферат докторской диссертации. М., ИОХ РАН, 2018 - https://zioc.ru/files/Автореферат_Свитанько.pdf 4. Davis, A.M. The use of the grid program in the 3-D QSAR analysis of a series of calcium channel agonists / A.M. Davis, N.P. Gensmantel, E.Johansson, D.P. Marriott // J. Med. Chem. - 1994. - Vol. 37. - P. 963-972 5. Klebe, G. Comparative molecular similarity index analysis (CoMSIA) to study hydrogen-bonding properties and to score combinatorial libraries / G.Klebe, U. Abraham. // Journal of Computer-Aided Molecular Design. -1999. - Vol. 13. - P. 1-10 6. Жохова, Н.И.

Метод непрерывных молекулярных полей в поиске количественных соотношений <структура - активность> / Н.И.Жохова, И.И. Баскин, Д.К. Бахронов и др. // Докл. РАН. - 2009. - Т.429. - № 2. - С. 127-140