

Комплексный анализ данных при исследовании сложных биомолекулярных систем

Н.Н. Яцков, В.В. Апанасович

Белорусский государственный университет, Минск

E-mail: yatskou@bsu.by

Стремительное развитие компьютерных технологий, производства новейших высокопроизводительных приборов, вычислительных систем, экспериментального оборудования и облачных хранилищ информации дало толчок активному развитию технологий интеллектуального анализа данных (ИАД). Появление ИАД связано с необходимостью обработки сверхбольших объемов информации, накапливаемой в современных хранилищах, и поиска новых знаний или закономерностей, не поддающихся обнаружению стандартными методами обработки информации или экспертным путем [1, 2]. Использование методов имитационного моделирования для решения задач оптимизации решений [3] в ходе анализа больших данных открывает новые возможности перед аналитиками, исследователями и инженерами в области высокотехнологичных междисциплинарных исследований (биоинформатика, биомедицина, фармацевтика, пищевая промышленность, энергетика, аэрокосмические системы). В числе основных задач – разработка перспективных методов анализа, интерпретации и хранения биофизической информации, создание инфраструктуры для приема, обработки, распространения результатов экспериментальных исследований. Сложность и разнообразие методов интеллектуального анализа данных [4], применяемых в тандеме с алгоритмами имитационного моделирования, требуют создания нового подхода для решения задач комплексного анализа биофизической информации в современных экспериментах и сверхбольших хранилищах данных.

В работе предлагается системный подход на основе методов имитационного моделирования и интеллектуального анализа больших данных для исследования биомолекулярных систем и решения задач оптимизации биофизических экспериментов.

Особенностью подхода является применение алгоритмов имитационного моделирования для воспроизведения процессов в исследуемых системах, что позволяет: 1) повысить точность оценки исследуемых характеристик биофизических процессов; 2) углубить знания о физике и сути исследуемых процессов; 3) создать новые инструменты прогнозирования, когда аналитических моделей не существует или вывод аналити-

ческих решений затруднен ввиду возрастающей сложности системы, представляемой большими данными. Интеграция усовершенствованных алгоритмов ИАД способствует более точному определению знаний и закономерностей, получаемых в результате анализа больших экспериментальных данных, которые невозможно определить классическими методами статистического анализа и ИАД (основанными на использовании аппроксимационных моделей, не учитывающих физику процессов).

Схема исследования объекта, некоторого биофизического процесса или биомолекулярного соединения, с использованием разработанного комплексного подхода представлена на рисунке.

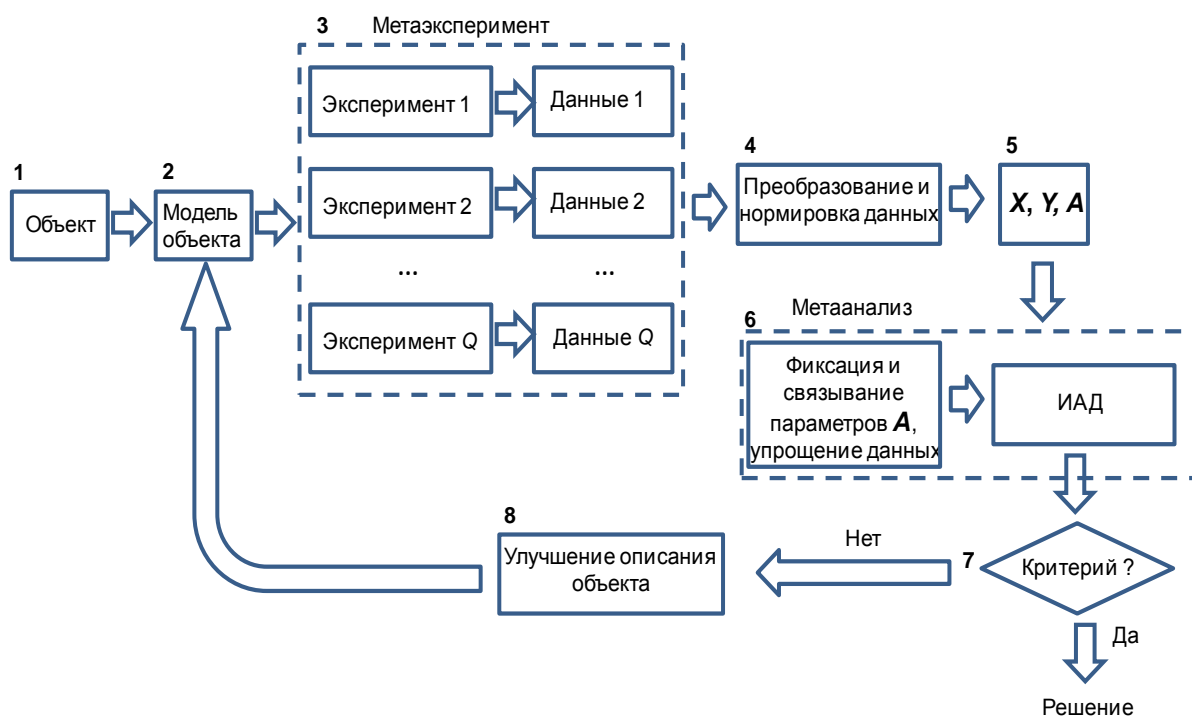


Рис. Схема исследования объекта с использованием разработанного комплексного подхода

В работе представлен пример применения комплексного подхода при исследовании биофизических соединений с помощью методов флуоресцентной спектроскопии.

1. Lynch C. // Nature. 2008. V. 455. P. 28–29.
2. Voit E.O. // PLoSComputBiol. 2019. Vol. 15 (9):e1007279.
3. Yatskou M. Computer simulation of energy relaxation and -transport in organized porphyrin systemsю Ponsen & Looijen Printing Establishment. Wageningen. The Netherlands, 2001. 176 p.
4. Яцков Н.Н. Интеллектуальный анализ данных: пособие. Мн.: БГУ, 2014. 151 с.