

БИБЛИОГРАФИЧЕСКИЕ ССЫЛКИ

1. Янукович Т. П. Численная модель трехволнового рассеяния Мандельштама-Бриллюэна в оптическом волокне // Оптический журнал. 2002. Т. 69. № 7. С. 49–54.
2. D. Garus, T. Godolla, K. Krebber, F. Schliep. Brillouin optical-fiber frequency-domain analysis for distributed temperature and strain measurements // J. Lightwave Technology. – 1997. – Vol. 15, № 4. – P. 654–662.
3. Hereth R. Stimulierte Brillouin-Streuung in Lichtleitfaser-Ringresonatoren: Dissertation Ruhr-Universitat: Elektronik Nr. 140. Bochum. 1992. 178 p.
4. Янукович Т. П., Поляков А. В. Моделирование распределенного измерителя силы тока на основе деформации оптического волокна // Приборы и методы измерений. 2019. Т. 10. № 3. С. 243–252.

ИССЛЕДОВАНИЕ БИОФИЗИЧЕСКИХ СИСТЕМ С ИСПОЛЬЗОВАНИЕМ АЛГОРИТМОВ ИНТЕЛЛЕКТУАЛЬНОГО АНАЛИЗА ДАННЫХ И ИМИТАЦИОННОГО МОДЕЛИРОВАНИЯ

Н. Н. Яцков, В. В. Апанасович

*Белорусский государственный университет, Минск, Беларусь
E-mail: yatskou@bsu.by*

Предложен комплексный подход на основе методов интеллектуального анализа данных и имитационного моделирования для исследования биофизических систем, позволяющий определить параметры биофизических и оптических процессов в молекулярных соединениях. Рассмотрено применение разработанного подхода при исследовании бимолекулярных соединений с помощью метода флуоресцентной спектроскопии с временным разрешением.

Ключевые слова: биофизическая система; имитационное моделирование; интеллектуальный анализ данных; флуоресцентная спектроскопия.

Развитие биофизических технологий напрямую связано с разработкой эффективных методов и алгоритмов обработки большого объема информации, получаемой с помощью современного высокопроизводительного экспериментального оборудования [1]. Современные вычислительные подходы, например, на основе методов системного анализа, для обработки и количественной интерпретации многомерных экспериментальных данных востребованы в области исследований биофизических систем с помощью методов флуоресцентной спектроскопии [2]. Возможным направлением разработки является интеграция методов интеллектуального анализа данных и имитационного моделирования биофизических процессов при исследовании бимолекулярных соединений в флуоресцентной спектроскопии, позволяющее автоматизировать анализ

больших данных, выделить наиболее информативную часть данных, существенно повысить точность оценки исследуемых характеристик биофизических процессов, углубить знания о физике и сути исследуемых процессов, создать новые инструменты прогнозирования, когда аналитических моделей не существует или вывод аналитических решений затруднен ввиду возрастающей сложности системы, представляемой большими данными [3].

В работе предлагается комплексный подход на основе методов интеллектуального анализа данных и имитационного моделирования для исследования биофизических систем.

Идея подхода состоит в рассмотрении наборов многомерных биофизических данных как единого целого, применении к нему алгоритмов интеллектуального анализа (data mining) или больших данных (big data) с целью выделения наиболее информативных или существенных данных для последующего глубокого изучения с использованием имитационных моделей [4, 5]. Схема исследования объекта, некоторого биофизического процесса или биомолекулярного соединения, с использованием разработанного комплексного подхода представлена на рисунке.

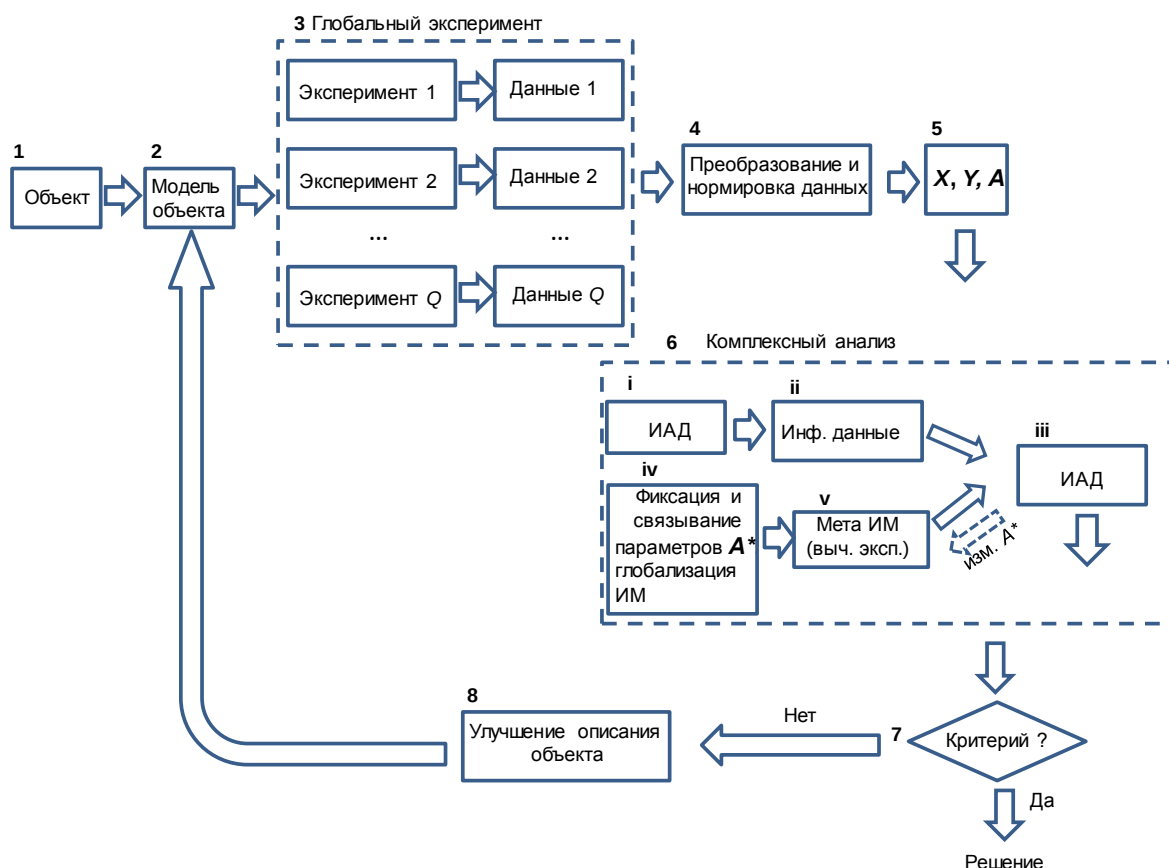


Рис. Схема исследования объекта с использованием разработанного подхода

В блоке 4 осуществляется преобразование данных отдельных экспериментов с целью сведения данных к единому формату, устранению неоднородностей различных экспериментальных искажений, связанных с конкретными экспериментами. Осуществляется фильтрация, нормировка, векторизация или специальные преобразования данных, например – логарифмическая трансформация данных для снижения эффекта влияния аномальных значений и выбросов. Наборы преобразованных данных из различных экспериментов собираются в сводный или объединённый набор данных $\{X, Y\}$ (блок 5) для последующей обработки в ходе комплексного анализа с помощью методов интеллектуального анализа данных и имитационного моделирования. Параметры отдельных экспериментов собираются в единый вектор A , затем уточняются в ходе последующего анализа объединенных данных. Анализ отдельных экспериментов может быть осуществлен независимо или совместно. Достоинством комплексного анализа является объединение данных (выборок) из различных экспериментов в один большой набор, что обеспечивает обобщение и увеличение статистической мощности результатов и, как следствие, повышение точности последующего анализа. В ходе комплексного анализа (блок 6) часть параметров A фиксируется (являются глобальными для проводимых экспериментов), связывается (в случае зависимых экспериментов) или остается свободным для точной оценки с помощью методов оптимизации; с применением методов интеллектуального анализа данных производится сжатие данных с целью исключения неинформативных, избыточных данных или шума, выделяются существенные информативные данные. Методами оптимизации производится оценка свободных параметров A^* глобальной имитационной модели объекта, построенной на основе моделей отдельных экспериментов. Если достигнута желаемая точность соответствия экспериментальных и смоделированных данных, определяется по заданному статистическому критерию в блоке 7, то процесс анализа завершается – предоставляется решение (оцененный набор параметров, точная математическая модель объекта, способная давать прогноз поведения объекта в пределах заданной точности), иначе в блоке 8 осуществляется улучшение описания объекта, включающие углубление детализации формализации поведения объекта, сбор новых данных, изменение моделей, проведение дополнительных экспериментов, изменение параметров объекта, среды или окружения, и переход к блоку 2 для выполнения следующей итерации анализа данных. Представленная схема является общим подходом, конкретная реализация определяется типом решаемой задачи и должна быть сконструирована с учетом специфики проводимого эксперимента.

В работе рассмотрено применение комплексного подхода при исследовании бимолекулярных соединений с помощью метода флуоресцентной спектроскопии с временным разрешением [6].

БИБЛИОГРАФИЧЕСКИЕ ССЫЛКИ

1. Cheng M., Cao L., Ning K. Microbiome Big-Data Mining and Applications Using Single-Cell Technologies and Metagenomics Approaches Toward Precision Medicine // Front. Genet. 2019. V. 10:972.
2. Яцков Н. Н., Скакун В. В., Апанасович В. В. Метод исследования флуоресценции молекулярных соединений с использованием алгоритмов интеллектуального анализа данных // Прикладные проблемы оптики, информатики, радиофизики и физики конденсированного состояния: Матер. 5-ой Междунар. науч.-практ. конф. НИУ «Ин-т прикл. физ. проблем им. А.Н. Севченко» БГУ, 2019. С. 125–127.
3. Яцков Н. Н., Апанасович В. В. Комплексный анализ данных при исследовании сложных биомолекулярных систем // Квантовая электроника: Матер. XII Междунар. конф. Мн.: БГУ, 2019. С. 282–283.
4. Яцков Н. Н. Интеллектуальный анализ данных: пособие. Мн.: БГУ, 2014. 151 с.
5. Yatskou M. Computer simulation of energy relaxation and -transport in organized porphyrin systemsю Ponsen & Looijen Printing Establishment. Wageningen. The Netherlands, 2001. 176 p.
6. Яцков Н. Н., Скакун В. В., Апанасович В. В. Метод обработки кинетических кривых затухания флуоресценции с использованием алгоритмов интеллектуального анализа данных // ЖПС. 2020. Т.87, №2. С. 311–322.

ОБРАБОТКА ГИСТОГРАММ ЧИСЛА ФОТООТСЧЕТОВ В ЭКСПЕРИМЕНТАХ ФЛУОРЕСЦЕНТНОЙ ФЛУКТУАЦИОННОЙ СПЕКТРОСКОПИИ

Н. Н. Яцков, В. В. Скакун, В. В. Апанасович

*Белорусский государственный университет, Минск, Беларусь
E-mail: yatskou@bsu.by*

Перспективным направлением исследований в флуоресцентной спектроскопии является анализ флуктуаций интенсивности флуоресценции во времени с использованием гистограмм числа фотоотсчетов. Ключевым элементом подхода является применение метода главных компонент для проверки гипотезы о разделимости кластеров многомерных экспериментальных данных. В работе исследуются возможности использования нормировки локального взвешенного сглаживания данных в пространстве первых двух главных компонент и метода независимых компонент для интерпретации независимых факторов, объясняющих вариацию в данных.

Ключевые слова: флуоресцентная флуктуационная спектроскопия; гистограмма числа фотоотсчетов; имитационное моделирование; метод главных компонент; нормировка локального взвешенного сглаживания; метод независимых компонент.