

Секция 4. ИНФОРМАЦИОННЫЕ И ВЫЧИСЛИТЕЛЬНЫЕ ТЕХНОЛОГИИ

УДК 004.65

Математические методы обработки данных пептидных микрочипов

*Д.С. Анисимов, М.А. Рязанов, А.И. Шановал
(АлтГУ, Барнаул).*

В работе рассматривается подход к обработке многомерных данных пептидных микрочипов. Основными этапами используемой технологии являются предобработка, направленная на уменьшение аппаратных ошибок измерений, уменьшение размерности, направленное на выделение переменных, наилучшим образом описывающих исходные данные, и классификация тестовых данных, результатом которой является определение класса нового объекта, используя множество объектов-образцов. Ввиду малого количества тестовых данных (в работе использовались пробы 25 доноров, из которых 15 – условно-здоровые и 10 с диагнозом рака молочной железы), работа нацелена на апробацию различных алгоритмов обработки, анализ применимости и выявление путей их дальнейшего развития с целью улучшения качества и повышения устойчивости результатов при применении этих алгоритмов. Результатом работы является технология обработки данных пептидных микрочипов и при дальнейшей доработке возможно ее применение на реальных данных с использованием большего количества образцов и большего количества классов.

Ранняя диагностика онкологических заболеваний является основным направлением научных исследований Российско-Американского противоракового центра, созданного в Алтайском государственном университете. Данные исследования проводятся совместно с центром инновационной медицины Аризонского государственного университета и Алтайским краевым онкологическим диспансером.

Основой исследований является обработка данных пептидных микрочипов, разработанных в институте биодизайна университета штата Аризона [1]. В общем виде, пептидный микрочип представляет собой подложку из нейтрального материала (стекла или полимерных мате-

риалов), на которую нанесены, искусственно созданные, последовательности аминокислот (пептиды).

В данной статье рассматривается проблема обработки данных, полученных с микрочипов, состоящих из четырёх независимых блоков, содержащих по 10368 пептидов. Из них 80 «EMPTY» - пептидов, не содержащих аминокислот, 112 «FIDUCIAL» - пептидов, предназначенных для контроля и 10176 уникальных пептидов, имеющих информационный характер.

По окончании биологической части эксперимента по каждому блоку микрочипа получается цифровая характеристика светимости пептидов, включая служебную информацию об условиях проведения эксперимента, минимальная математическая обработка. Общая структура получаемого файла описана в [2].

Предварительным этапом применения математических методов преобработки полученных многомерных данных является процесс их нормализации. Наиболее перспективными методами нормализации светимости пептидов являются методы, изложенные в работах [1, 3, 4]. Нами исследованы следующие алгоритмы нормализации:

- нормализация относительно фона;
- нормализация приведением к медианной светимости;
- квантильная нормализация;
- по методике, изложенной в работе Марины Кретич и Марцела Чиари [4].

В докладе будут представлены результаты вычислительных экспериментов, направленных на исследование методов обработки многомерных данных пептидных микрочипов. Одним из основных является следующий результат: каждому из рассмотренных классификаторов для получения оптимальных оценок нужен свой алгоритм нормализации. Так, с точки зрения чувствительности, специфичности и точности наилучшим методом предварительной обработки имевшихся многомерных данных оказалась нормализация по формуле (4) с последующей их классификацией методом наименьших квадратов.

Изложенный в данной статье подход к обработке данных пептидных микрочипов является основой для проведения дальнейших исследований. Одним из возможных путей развития технологии является применение нестатистических методов анализа и методов интервального анализа, а также методов изложенных в работах [5, 6].

Библиографический список

1. Stafford P., Cichacz Z., Woodbury N., Johnston S.A. Immunosignature System for Diagnosis of Cancer // PNAS. – 2014. DOI: 10.1073/pnas.1409432111.
2. Форматы файлов GenePix (GenePix® File Formats) [Электронный ресурс]. – URL: http://mdc.custhelp.com/app/answers/detail/a_id/18883/~/genepix%C2%AE-file-formats.
3. Sykes K., Legutki J.B., Stratford P. Immunosignaturing: a critical review // Cell Press. – 2012. DOI: 10.1016/j.tibtech.2012.10.012.
4. Cretich M., Chiari M. Peptide Microarrays. Methods and Protocols – Humana Press, 2009. DOI:10.1007/978-1-60327-394-7.
5. Максимов А.В., Оскорбин Н.М. Многопользовательские информационные системы: основы теории и методы исследования. – Барнаул, 2013.
6. Оскорбин Н.М. Математические модели систем с латентными переменными // Известия Алтайского государственного университета. – 2012. – №1.

УДК 004.65

Актуальность разработки геоинформационной вебсистемы для комплексного анализа данных о грозовой активности Горного Алтая

*М.Ю. Беликова, С.Ю. Кречетова, А.А. Перельгин
ГАГУ, г. Горно-Алтайск*

Территория Горного Алтая характеризуется повышенной грозовой активностью. Орографическая сложность и малая освоенность территории ограничивает использование ряда методов, позволяющих получить качественные характеристики грозовой активности. Тем не менее, территория является хорошей лабораторией для фундаментальных исследований закономерностей временного и пространственного распределения грозовой активности.

Данные о закономерностях пространственного распределения гроз необходимы как для решения фундаментальных задач атмосферного электричества, так и для решения практических задач грозозащиты линий электропередач, зданий и сооружений и грозовой пожарной опасности лесов [1].

В настоящее время источниками таких данных являются наблюдения редкой сети гидрометеостанций и единичных на территории Си-