

ОТЗЫВ

на выпускную квалификационную работу магистра

Ивановой Е.В.

“Задачи анализа спектров тандемной масс-спектрометрии”

В последние годы методы тандемной масс-спектрометрии стали одними из основных способов изучения структуры веществ, поставленных “на поток”. Данные методы обладают существенными преимуществами перед альтернативами как по точности полученных результатов, так и по требованиям, предъявляемым к изучаемым образцам.

Тем не менее, тандемная масс-спектрометрия не лишена недостатков. Получаемые данные (т.н. “масс-спектры”) как правило зашумлены, содержат явно выраженные систематические ошибки (т.н. baseline в виде медленно меняющейся компоненты) и т.п. Данные ошибки обусловлены физико-химическими свойствами процесса и поэтому использование масс-спектров сопряжено с необходимостью предварительной обработки данных масс-спектрометрии с целью выделения истинного сигнала.

В рамках магистерской работы были изучены существующие алгоритмы фильтрации масс-спектров с целью дальнейшего интегрирования в алгоритм Дерепликатор, разработанный в лаборатории “Центр алгоритмических биотехнологий” СПбГУ совместно с коллегами из Университета Калифорнии в Сан-Диего. Целью данного исследования являлся отбор алгоритмов, которые с одной стороны имеют наименьшее количество параметров (или включают автоматическую процедуру подбора параметров), а с другой – позволяют массово обрабатывать большое количество масс-спектров. В итоге в рамках работы была предложена модификация алгоритма MassSpecWavelet, учитывающая дополнительную специфику данных, полученных на последних поколениях масс-спектрометров.

Второй задачей, рассмотренной в рамках магистерской работы, являлось исследование подходов кластеризации масс-спектров,

согласованных со аминокислотной последовательностью пептидов. Данная задача имеет большой практический интерес для задач широкомасштабного поиска спектров в базах данных пептидов: предварительная кластеризация базы потенциально может уменьшить вычислительную трудоемкость этой процедуры в разы.

В работе рассмотрены несколько известных метрик и расстояний (например матрица Вассерштайна между вероятностными мерами) и проверено, насколько они подходят для задачи кластеризации масс-спектров, представленных в виде гистограмм, исследованы результаты кластеризации на реальных данных.

В качестве недостатков работы следует отметить явную незаконченность части про кластеризацию, так например, полностью отсутствуют какие-либо результаты моделирования. Кажется, что автору несколько не хватило времени.

Тем не менее, представленные недостатки не умаляют общей ценности работы; считаю, что работа может быть оценена на оценку “отлично”.

к.ф.-м.н. доцент А.И. Коробейников