

ОТЗЫВ

на выпускную квалификационную работу магистра

Абрамовой А.Н.

“Метод Монте-Карло по схеме марковской цепи для оценки вероятности редких событий в задачах биоинформатики”

Начиная с пенициллина, пептидные натуральные продукты (ПНП) оставили яркий след в области фармакологии: многие антибиотики, противовирусные и противоопухолевые средства, иммунодепрессанты и токсины являются ПНП. Недавнее открытие теиксобактина (начало 2015 г.) снова вернуло повышенный интерес к натуральным продуктам после длительного периода спада в усилиях обнаружить новые антибиотики.

Старт проекта Global Natural Products Social (GNPS) Molecular Networking объединил в одном месте более миллиарда масс-спектров натуральных продуктов, собранных в более ста лабораториях по всему земному шару. В то время как эти спектры определенно содержат новые ПНП, в том числе чрезвычайно полезные с медицинской точки зрения, их обнаружение остается сложной вычислительной задачей.

В частности, для такого обнаружения необходимо быстро и точно вычислять значимость вхождений масс-спектров в базу данных. Проблема стоит особенно серьезно для циклических и разветвленных пептидов, когда стандартные комбинаторные подходы перестают работать. Как оказалось, на данную задачу можно посмотреть как задачу оценки вероятностей редких событий (типичные значения этой вероятности в задачах поиска ПНП составляют 10^{-50} - 10^{-10}), которая достаточно изучена методами современной вычислительной статистики.

В рамках магистерской диссертации с использованием новейших результатов метода Монте-Карло на Марковских цепях была построена оценка искомой вероятности. Помимо этого был предложен способ вычисления точности получаемой оценки посредством оценивания ее

дисперсии, и, что не менее важно, способ определения требуемого объема выборки для достижения наперед заданной точности оценки.

Свойства полученных оценок были проверены на модельных и реальных выборках. Было показано, что дисперсия полученных оценок заметно меньше, чем дисперсия обычных оценок по методу Монте-Карло. Более того, выбор подходов к построению оценок в работе был осуществлен с учетом их эффективной реализации с точки зрения вычислительной сложности.

К сожалению, не все свойства оценок удалось проверить теоретически. Так, необходимые условия для наличия функциональной центральной предельной теоремы (ФЦПТ) для получающейся марковской цепи были проверены только для частного случая дискретного пространства состояний. В общем случае непрерывного пространства состояний удалось показать только эргодичность по Харрису (стоит отметить, что проверка достаточных условий для ФЦПТ является достаточно сложной теоретической задачей).

Помимо этого стоит отметить отдельную тезисность и небрежность изложения, что в некоторых случаях заставляет прикладывать некоторые усилия, чтобы убедиться в правильности получаемых автором результатов.

По итогам магистерской работы подготовлена статья на конференцию WABI (Workshop on Algorithms in Bioinformatics).

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

к.ф.-м.н. доцент А.И. Коробейников