

ОТЗЫВ

официального оппонента на кандидатскую диссертацию Форгани Маджид Али «Математическое моделирование антигенного сходства штаммов вируса гриппа с помощью вейвлет-преобразования», представленную к защите на соискание ученой степени кандидата физико-математических наук по специальности 05.13.18 – Математическое моделирование, численные методы и комплексы программ.

Диссертация М.А. Форгани состоит из списка основных обозначений и соглашений, введения, трех глав, заключения, двух приложений и списка литературы. Общий объем работы – 174 страниц. Библиография содержит 154 наименований.

В диссертации изучается прикладная задача математического моделирования о предсказании на парах аминокислотных последовательностей (t, r) заданного белка значений титра $H_{t,r}$ – экспериментальной величины, характеризующей антигенное сходство между t и r .

На данный момент решить эту задачу с помощью физико-химико-биологических моделей не представляется возможным, а лабораторные измерения сложны и дорогостоящи. Поэтому большое значение приобретают статистические методы регрессионного анализа на основе учета косвенных признаков. Наличие хорошей статистической модели позволит значительно ускорить поиск близкого к неизвестному штамму вируса t референтного штамма r , а, значит, упростить разработку вакцины. Это подчеркивает актуальность и практическую значимость работы. Данный и другие аспекты прикладной значимости проблемы, а также подходы к ее решению излагаются во введении.

Автор разработал решение на основе открытых данных для белка гемагглютинаина вируса гриппа А(H1N1) [95]. За основу взята модель линейной регрессии для логарифма титра из работы W.T. Harvey и др. (2016) [1]:

$$\log_2 H_{t,r} = k_0 + \sum_j k_j \alpha_j(t,r) + \sum_l m_l \delta_l(t,r) + f_t + f_r + \varepsilon.$$

Здесь базисные функции $\alpha_j(t,r)$ – 0-1 функции, зависящие от близости символьных представлений последовательностей t и r , $\delta_l(t,r)$ – 0-1 функции разделимости t и r филогенетическим деревом белка гемагглютинаина. С точки зрения цифровой обработки сигналов анализ осуществляется во временной области.

Автор предложил и в этом состоит научная новизна привлечь к решению задачи спектральный анализ на основе дискретного вейвлет-преобразования аминокислотной последовательности f , записанной в чи-

словом виде $f = (f_1, \dots, f_x, \dots, f_N)$, где f_x – числовая векторная характеристика i -й аминокислоты. Эта характеристика строится автором на основе свойств аминокислот, приведенных в открытой базе данных AAindex [35]. Таких свойств порядка 500, поэтому автор применил кластерный анализ для построения составных свойств (индексов), по которым определяется вектор-значение f_x . В зависимости от технических возможностей для оцифровки аминокислотной последовательности можно использовать всю базу данных или индексы, полученные кластеризацией.

Математическая модель автора, приведенная в первой главе, имеет следующий вид:

$$\log_2 H_{t,r} = \beta_0 + \sum_l \beta_l \zeta_l^*(t,r) + \varepsilon,$$

где базисные функции $\zeta_l^*(t,r)$ (или признаки) нетривиально выражаются через расстояние между вейвлет-разложениями последовательностей t и r .

Численным аспектам вычисления $\zeta_l^*(t,r)$ посвящена глава 2. Вначале к аминокислотной последовательности применяется заранее выбранное дискретное вейвлет-преобразование, позволяющее записать ортогональную декомпозицию $f_x = \sum_{p=1}^M f_{p,x}$. Далее автор, следуя идеям R.R. Coifman и M.V. Wickerhauser (1992), предложил применить вейвлет-преобразование уже к каждой компоненте $(f_{1,x}, \dots, f_{p,x}, \dots, f_{M,x})$: $f_{p,x} = \sum_{p'=1}^M f_{p',p,x}$. Продолжая эту процедуру получают последовательности $f_{p_n \dots p_1, x} =: f_{p,x}$, которые автор назвал вейвлет-частицами. Их сумма по всем индексам p_i по-прежнему равна f_x . Такое избыточное разложение автор мотивирует увеличением вариантов представления сигнала и возможностью выбора при последующем анализе наиболее адекватного представления $\sum_p f_{p,x}$. Автор для этого применяет алгоритмы на деревьях.

В итоге, функции $\zeta_l^*(t,r)$ выражается через $\text{dist}(t_{p,x}, r_{p,x})$, где dist – евклидова или другая метрика. Применяются методы кластерного анализа для уменьшения числа признаков, чтобы избежать переобучения модели. Достоверность выкладок обосновывается сформулированными и доказанными утверждениями, а также проверяется вычислительными экспериментами.

Особенности программной реализации метода, выполненной в MATLAB, даются в главе 3. Автор сравнивает полученную им точность предсказания титра с результатами авторов работы [1] и заявляет, что достигнутая им точность сопоставима и даже превосходит точность, полученную в [1].

Достоинством диссертации также является широкий охват литературы по данной проблеме. Также приятна структура, где глава 1 – это модель, глава 2 – численный метод и глава 3 – программная реализация.

Однако имеются замечания, связанные с оформлением диссертации.

1. В диссертации много стилистических, грамматических и орфографических ошибок. Во многих местах идет абзацный отступ там, где его не должно быть. Укажу только один пример: стр. 66, 3-я строка снизу, фраза «где, ».
2. Диссертацию тяжело читать, в том числе из-за обилия аббревиатур. Многие из них приведены в списке сокращений, хотя встречаются один раз, например, «АСС» на стр. 46. С другой стороны аббревиатура «АР» встречается много раз, но ее нет в списке сокращений и я не нашел определения в тексте. По смыслу это «АСС».
3. Список сокращений содержит только одну формулу, а введение ни одной, хотя во многих местах проще было написать формулу, чем пояснять нестрогим текстом. Это странно для диссертации на степень кандидата физико-математических наук. Во введении слишком много биологии и мало математического моделирования.
4. Название глав можно было бы дать информативнее.
5. Обозначения первой главы не повторяются в главе 2, а глава 3 не использует обозначения первых глав. В результате, например, очень сложно понять, как определяются базисные функции модели $\zeta_l^*(t, r)$. Нет итоговой расчетной формулы для них и явно не указано, как изменяется индекс l . Это недостаток.
6. Основная формула (1.2.5) модели записана с огрехами: локальные переменные x и h не нужно указывать в $\zeta_l^*(x, h, t, r)$, поскольку левая часть не зависит от x и h . Свободный коэффициент k_0 надо было обозначить, например, β_0 , иначе идет пересечение с обозначениями модели [1]. Лучше было назвать его свободным коэффициентом или смещением, но не «пересечением» как на стр. 47.
7. Не нужно было приводить в главе выдержки из теории всплесков на оси. А вот из теории дискретных всплесков, имеющей особенности и применяемой в итоге в работе, стоило бы.
8. Термин «вейвлет-частица» лучше было бы назвать «вейвлет-последовательность (или множество) координаты f_x ».
9. Термин «векторное расстояние» на стр. 81 неудачен.
10. Литература оформлена вразнобой.
11. На некоторых рисунках, например, рис. 2.11–2.13, нет координатных осей, что усложняет их понимание.

Автореферат правильно и полно отражает содержание диссертации, а ее основные результаты являются новыми и с надлежащей полнотой опубликованными в ведущих рецензируемых журналах из списка ВАК и приравненных к ним. Содержание диссертации соответствует специальности, по которой она рекомендуется к защите.

Диссертация Форгани Маджид Али соответствует требованиям п. 9 «Положения о присуждении ученых степеней», удовлетворяет требованиям ВАК к диссертациям, представленным на соискание ученой степени

кандидата физико-математических наук по специальности 05.13.18 – Математическое моделирование, численные методы и комплексы программ.

Автор диссертации Форгани Маджид Али заслуживает присуждения ученой степени кандидата физико-математических наук по специальности 05.13.18 – «Математическое моделирование, численные методы и комплексы программ».

Официальный оппонент, доктор физико-математических наук по специальности 01.01.01 – «Вещественный, комплексный и функциональный анализ», профессор кафедры прикладной математики и информатики Института прикладной математики и компьютерных наук ФГБОУ ВО «Тульский государственный университет», 300012, г. Тула, пр. Ленина, д. 92

9

9

Горбачев Дмитрий Викторович

«23» 01 2019 г.

E-mail: dvgmail@mail.ru

Подпись Д.В. Горбачева заверяю,

Начальник ОК "11"

Е.Ю. Меркулов
23.01.2019.

