

ОТЗЫВ

официального оппонента Леоненко Василия Николаевича
на диссертационную работу Форгани Маджида Али
«Математическое моделирование антигенного сходства штаммов вируса гриппа
с помощью вейвлет-преобразования»,
представленную на соискание учёной степени кандидата физико-
математических наук
по специальности 05.13.18 - математическое моделирование, численные методы
и комплексы программ

1. Актуальность диссертационной работы

Диссертационная работа Форгани Маджида Али посвящена созданию модели антигенности вируса гриппа и комплекса программ для проведения численных экспериментов с моделью. Указанная модель позволяет прогнозировать сходство между штаммами вируса гриппа и ускорить создание вакцин с учётом циркулирующего штамма.

2. Постановка цели и задач исследования

Целью представленных исследований диссертационной работы является математическое моделирование и прогнозирование антигенного сходства между штаммами вируса гриппа на основе их последовательности белка гемагглютинина.

В диссертации М.А. Форгани решены следующие задачи:

1. Оцифровка последовательности белка на основе физико-химических свойств аминокислоты.
2. Извлечение признаков с помощью вейвлет-преобразования и эвристического алгоритма поиска наискорейшего восхождения к вершине.
3. Разработка численного метода декомпозиции вейвлет-частиц.
4. Поиск наилучшего признака с помощью эвристического алгоритма.
5. Создание модели и комплекса программ «Модель антигенности вируса гриппа» для моделирования и прогнозирования антигенного сходства.

3. Содержание диссертационной работы

Диссертация включает в себя: введение, 3 главы, заключение, 2 приложения и библиографический список из 154 наименования.

Во введении обосновывается актуальность работы, предоставляется обзор развития данной области, рассматриваются различные подходы и их основания, определяется цель, задачи, теоретическая и практическая значимость работы.

В первой главе описываются модель, обоснование выбора типа модели, формулируется математическое представление задачи. Модель, используемая в диссертации, основана на линейной модели смешанных эффектов, которая показала эффективность в аналогичных работах других исследователей.

Во второй главе описываются подходы к решению поставленных задач и приведены алгоритмы, реализующие предложенные методы. Последовательности белка оцифровываются с помощью физико-химических свойств аминокислоты с целью последующего применения методов обработки

сигналов. Определяется метод декомпозиции вейвлет-частиц для извлечения признаков, обосновывается метод и его обобщение в ряде математических утверждений. Данный метод учитывает такое биологическое понятие как эффект соседства, позволяя детально изучать любое изменение в соседних аминокислотах конкретной позиции на признаках, полученных из данной позиции. При использовании метода декомпозиции вейвлет-частиц строится дерево декомпозиции, комбинация данного метода с алгоритмом поиска наискорейшего восхождения к вершине уменьшает трудоемкость извлечения локального оптимального признака. В этой главе также приводится ряд вычислительных экспериментов, которые подтверждают эффективность метода в извлечении признаков.

В третьей главе описываются структура и особенности реализации программного комплекса, представлена схема вычислительной модели. Также описана реализация модели для реальных данных вируса гриппа подтипа H1N1. Результаты сравниваются с результатами других работ. К программному комплексу добавлена визуализация дерева эволюции с дополнительной информацией, полученной с помощью метода декомпозиции вейвлет-частиц.

В заключении приведены выводы по диссертационной работе и намечены перспективы дальнейших исследований.

Оформление диссертационной работы и автореферата в целом отвечает предъявляемым требованиям.

4. Новизна исследований и полученных результатов

Основные результаты, представленные в диссертации являются новыми, они обобщают и дополняют результаты предыдущих работ по данной теме. К основным научным результатам можно отнести следующее:

1. Предложен вычислительный метод декомпозиции вейвлет-частиц. Этот метод был применен к исследованию антигенности вируса гриппа. Также он может быть использован при извлечении признаков в аналогичных задачах моделирования фенотипа на основе генетических данных.
2. Получены утверждения о вычислении вейвлет-частиц на конкретном уровне, что может дать возможность подсчитывать частицы данного уровня независимо от частиц предыдущих уровней. Таким образом, можно подсчитать частицы в глубине дерева декомпозиции без подсчета частиц на предыдущих уровнях.
3. Реализован программный комплекс, основанный на вычислительных и эвристических математических методах, таких как вейвлет-преобразование и алгоритм поиска наискорейшего восхождения к вершине. В качестве дополнительной возможности комплекса программы была добавлена визуализация дерева эволюции (в частности, с использованием виртуальной реальности). Эта возможность позволяет детально интерпретировать изменения свойств аминокислоты с эволюцией вируса.

На защиту соискателем выносятся следующие основные **результаты**:

1. Математическая модель антигенности вируса гриппа, основанная на линейной модели смешанных эффектов и теории вейвлета.

2. Вычислительный алгоритм и его оптимизация для декомпозиции значения физико-химического свойства аминокислоты на основе эффекта соседства и методологии вейвлет-пакета.

3. Программно-аппаратный комплекс «Модель антигенности вируса гриппа», обеспечивающий возможности моделирования, прогнозирования антигенного сходства вирусов, выделения участков с различными свойствами антигенности, а также визуализации результатов.

Преимуществом модели соискателя по сравнению с существующими подходами является полноценный учёт изменения свойства аминокислоты в процессе эволюции и учёт регионального влияния замен аминокислот. Эффективность разработанной модели демонстрируется снижением средней ошибки по сравнению с принятой соискателем за эталон моделью из работы Harvey et al., 2016: 0.6 единиц титра против 0.9 единиц для модели Харви.

В ходе работы соискателем продемонстрированы знания как в области математического моделирования, так и в предметных областях, относящихся к модели - вирусологии и генетике.

Результаты работы представлялись на семинарах и конференциях, в том числе международных. По результатам диссертации опубликовано восемь работ, из них четыре - в журналах, рекомендованных ВАК.

6. Замечания

К содержанию диссертационной работы имеются следующие замечания:

1. Есть нарекания к логике изложения информации в некоторых частях диссертационной работы. А именно:
 - а. Не совсем понятна роль раздела «Содержательная постановка задачи» (стр.10-14) в главе «Введение». Представляется целесообразным сокращение объёма этого раздела и разделение оставшегося текста между предыдущим («Краткое описание цели, решения и результатов исследования») и последующим («Современное состояние исследований в области антигенной эволюции») разделами.
 - б. Кажется излишним в текущем его виде раздел «Краткое описание цели, задач и подходов работы» главы «Заключение», который и по названию, и по содержанию в значительной степени дублирует упомянутый раздел «Краткое описание цели, решения и результатов исследования» главы «Введение».
 - в. В тексте работы имеет место дублирование информации. Например, задачи работы и актуальность приводятся дважды одними и теми же словами на стр. 36-37 (раздел «Актуальность, цель, задачи и научная новизна работы» главы «Введение») и стр. 142, 144 (раздел «Краткое описание цели, задач и подходов работы» главы «Заключение»).
2. В ряде предложений соискателем используются неточные термины, например:
 - а. стр. 12: "**Иногда** высокочастотный компонент также содержит важную информацию и необходимо его тоже декомпозировать"

- b. здесь же: "Для этого автор предлагает использовать **весьма известный** метод вейвлет–преобразования"
- c. стр 26: "**Безусловно**, если можно посчитать, численно представить, и добавить эффект соседства в модели, можно улучшить результаты моделирования биологических функций"
- d. стр 40-41: "Хотя, **как известно**, корреляция в данном отношении может быть не слишком высокой."

Следует сопровождать подобные утверждения количественными оценками («иногда» – как часто?) и/или ссылками на публикации («безусловно», «как известно» - из чего следует?)

- 3. К приведённому на стр. 11 тексту, - "существует линейное отношение между заменами аминокислоты антигенных участков в последовательности НА и логарифмом значения титра", - следует добавить информацию о том, какой именно числовой показатель характеризует замены аминокислоты (размерность, способ измерения, интерпретация).
- 4. У большого количества рисунков (стр.55, 58, 67, 78, 79, 87, 119, 120, 121, 122, 123, 130, 138) подписи к осям не читаются из-за слишком мелкого шрифта либо отсутствуют, аналогичная проблема со значениями, отложенными на осях. Некоторые диаграммы (например, рис. 5.1. на стр. 148) выглядят размытыми: их следовало бы привести в векторный формат или хотя бы увеличить их разрешение.
- 5. Неудачное использование разнонаправленных стрелок на схемах на стр. 11 и 47 затрудняет понимание того, в каком направлении следует читать указанные схемы.
- 6. В тексте присутствуют грамматические и пунктуационные ошибки, а также опечатки. Есть нарекания к ряду использованных терминов, например:
 - a. стр. 29: «база данных AAindex была кластрирована» - очевидно, имелось в виду «кластеризована»
 - b. стр. 43: «матрица предиктов» - возможно, более правильно было бы написать «матрица прогнозирования»
 - c. стр. 71: «это транспозированный столбец матрицы восстановления» - очевидно, имелось в виду «транспонированный»
- 7. Большое количество терминов, связанных с вирусологией и генетикой, в работе не определяется либо используется до их определения, что усложняет понимание текста. Особенно это характерно для текста Приложения 1. Например:
 - a. стр. 150: «Известно, что в сравнении с НА, НА имеет высокую аминокислотную замену в своих **эпитопных областях** [138]».
 - b. здесь же: «Белок НА вируса гриппа синтезируется как один **полипептид** (НА0), который затем расщепляется на два полипептида (НА1 и НА2)»

- с. здесь же: «Ножка (см. рис.5.2) является **иммуно-субдоминантом** по сравнению с **глобулярным участком**».
- d. стр. 154: «Он представляет собой анализ связывания на основе способности НА к **агглютинации эритроцитов** и дополнительной способности антисыворотки к тем же или близким штаммам блокировать эту агглютинацию».

Возможно, некоторые из указанных терминов вводились в тексте где-то ранее, но предоставленный для рецензирования электронный вариант рукописи без возможности текстового поиска затрудняет обнаружение определений.

7. Заключение

Отмеченные замечания не носят принципиального характера и не снижают значимость результатов диссертационной работы. Форгани М.А. выполнено серьезное исследование предметной области, проанализированы существующие подходы и методы, выполнено сравнение с новейшей моделью. Текст диссертации представляет собой самостоятельную научно-квалификационную работу на актуальную тему, не содержит материала без ссылки на автора и (или) источник заимствования, а также не содержит результатов научных работ, выполненных в соавторстве, без ссылок на соавторов.

Считаю, что диссертационная работа соответствует требованиям Положения о порядке присуждения научных степеней, тема и содержание диссертации соответствует паспорту специальности, а её автор заслуживает присуждения степени кандидата физико-математических наук по специальности 05.13.18 - математическое моделирование, численные методы и комплексы программ.

Официальный оппонент,

Доцент Института дизайна и урбанистики
ФГАОУ ВО «Санкт-Петербургский национальный
исследовательский университет информационных
технологий, механики и оптики»,
кандидат физико-математических наук

Леоненко Василий Николаевич

4 февраля 2019 г.

Адрес:

Университет ИТМО,
197101, Санкт-Петербург, Кронверкский пр., д. 49
e-mail: vnleonenko@corp.ifmo.ru
тел.: +7 (812) _ _ _ _ _

Подпись В. Н. Леоненко заверяется



Леоненко В. Н.

Удмилаева Р. И.
04.02.2019