

## Тестирование инвариантами в применении к сложным системам в биоинформатике

Научный руководитель – Хританков Антон Сергеевич

*Якушева Софья Федоровна*

*Студент (магистр)*

Московский физико-технический институт, Москва, Россия

*E-mail: yakusheva.sf@phystech.edu*

В данной работе исследуется проблема проверки правильности работы сложных комплексов программ в биоинформатике. Особенностью проблемы является то, что для многих задач биоинформатики нет возможности указать тестовые критерии (test oracles)[1], поскольку их вычисление слишком ресурсозатратно или же вовсе не представляется возможным. Проблема. Инструменты и системы инструментов биоинформатики, как и всякое программное обеспечение, необходимо тестировать. Для валидации таких систем часто используют тестирование на некоторых специально отобранных образцах, и критерием прохождения теста является попадание итоговой метрики качества в некоторых интервал значений, устанавливаемый комиссией экспертов. Как правило, число эталонных образцов невелико, а оценка метрик достаточно субъективна, что мешает поиску ошибок.

Метод. Для подобных задач успешно применяется метод тестирования инвариантами (metamorphic testing) [2]. Основная идея этого метода - не проверка на правильность каждого конкретного теста, а проверка выполнения тестовых инвариантов - выводимых из свойств решения задачи соотношений между результатами работы. Это позволяет проверять программы при отсутствии правильных ответов для тестов, а также автоматизировать процесс проверки и генерировать большие тестовые наборы [2]. Биоинформатика является одной из перспективных областей применения тестирования инвариантами из-за приближенного или неполного решения задач, а также ограниченности количества тестовых образцов [3].

Постановка задачи. Задача данного исследования состоит в разработке методики автоматического тестирования комплексов инструментов биоинформатики с помощью инвариантов, подготовке и реализации подходящих инвариантов и реализации в программном коде, и из проверки в эксперименте по сравнительному анализу здоровых тканей и тканей раковых опухолей.

Решение. Методика тестирования основана на применении инвариантов, ранее использовавшихся для проверки инструментов биоинформатики, тестирования различных типов систем и приложений [4] и анализа похожих задач из области машинного обучения (обучение с учителем, обучение с подкреплением), а также новых инвариантов, разработанных с учётом особенностей составных частей проверяемой системы. Комплекс программ запускался на наборах сгенерированных тестов, полученные ответы проверялись на соблюдение инвариантов. Схема решения приведена на рис.1.

### Источники и литература

- 1) Barr E.T., Harman M., McMin P., Shahbaz M., Yoo S. The oracle problem in software testing: A survey. IEEE transactions on software engineering. 41(5). 2014. P. 507-525.
- 2) Chen T., Kuo F.-C., Liu H., Poon P.-L., Towey D., Tse T.H., Zhou Z.Q. Metamorphic Testing: A Review of Challenges and Opportunities. ACM Computing Surveys. 51. 4:1-4:27. 10.1145/3143561.

- 3) Chen T.Y, Ho J.W.K., Liu H., Xie X. An innovative approach for testing bioinformatics programs using metamorphic testing. BMC Bioinformatics 10. 2009. Article 24. Shahri M.P et al. Metamorphic Testing for Quality Assurance of Protein Function Prediction Tools.
- 4) Zhou Z., Sun L., Chen T.Y, Towey D. “Metamorphic relations for enhancing system understanding and use,” IEEE Transactions on Software Engineering, 2018.

### Иллюстрации

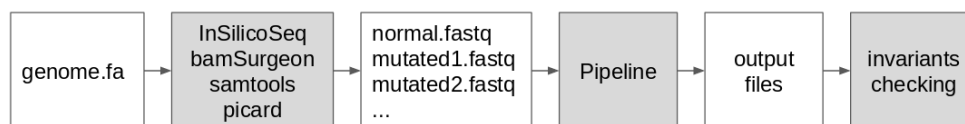


Рис. 1. Схема эксперимента