**Разработка программного модуля для нахождения тройных взаимодействий в биологических сетях с учётом временных задержек**

**Введение**

Анализ взаимодействия генов берет свое начало в 19 веке в работе [1]. Однако только в конце 20 – начале 21 века развитие экспериментальных подходов и алгоритмов выявления генных взаимодействий позволило восполнить исследовательский интерес, в таких работах как [REFERENCE]. Данная область позволяет найти взаимосвязь между биологическими процессами, протекающими в организме, с болезнями, как, например, в работе [2], в которой нашли взаимодействующие гены, ответсвенные за развитие шизофрении.

Набор взаимодействующих генов удобно визуально представлять в виде сети, в которой узлами являются гены, а взаимодействия обозначаются направленными отрезками. Такие сети называют генными сетями. Задача выявления структуры генной сети является одной из задач данной области. Структура может быть восстановлена экспертом, при условии наличия экспериментальных данных о парном взаимодействии всех генов. Данный подход является неоптимальным с точки зрения трудозатрат, не может быть применен к большим системам (объемом 100 и менее элементов), а также не учитывает одновременные взаимодействия трех и более элементов, что проявляется в реальных биологических системах [REFERENCE]. Альтернативным подходом было бы прямое моделирование взаимодействия генов. Но на практике, далеко не всегда имеются экспериментальные данные, полностью описывающие структуру, поведение и функции того или иного гена, что необходимо для прямого моделирования. Также, разнообразие взаимодействий генов, разнообразие генов делают эту задачу вычислительно трудоемкой и немасштабиремой на сети большого объема. Для решения данных проблем используются экспериментальные установки, позволяющие собрать большое количество данных[REFERENCE], а также алгоритмы, способные выявлять взаимодействия[REFERENCE].

Эффективные алогритмы выявления структуры сети работают по принципу “обратной разработки”[REFERENCE]. Имея набор взаимодействующих генов и результат их взаимодействия в виде изменения концентрации продуктов генной сети во времени(белков, протеинов, и т.д.), задача состоит в нахождении структуры сети, которая приводит к наблюдаемым результатам.

Одним из главных этапов нахождения структуры биологических сетей является выявление как парных взаимодействий[REFERENCE], так и взаимодействий более высокого порядка [REFERENCE](согласованного взаимодействия более чем двух элементов) [3]. Также, в более ранних трудах было показано[REFERENCE], что алгоритмы, учитывающие взаимодействие порядка более двух, дают более точные результаты структуры сети. Особый интерес представляет нахождение взаимодействий с учётом временных задержек, так как в реальных биологических процессах, реакция того или иного компонента сети может проявляться с задержкой[REFERENCE]. На текущий момент, существуют алгоритмы, учитывающие взаимодействия третьего порядка[REFERENCE], а также алгоритмы, учитывающие парные взаимодействия с временными задержками.

В данной работе предлагается инструмент, учитывающий взаимодействия третьего порядка с временными задержками. За основу был принят алгоритм MIDER[REFERENCE], который умеет выявлять парные взаимодействия с учетом временных задержек. Данный алгоритм был доработан учетом тройных взаимодействий, а также был изменён алгоритм учёта временных задержек[REFERENCE]. Были предложены методы оптимизации задачи нахождения тройных взаимодействий. Показаны результаты анализа реальной биологической сети из [3], предложена реализация алгоритма на MatLab, доступная по ссылке [REFERENCE].

**Заключение**

В данной работе был создан программный комплекс для нахождения тройных взаимодействий в биологических сетях с учётом временных задержек. На текущий момент, этот алгоритм является первым, учитывающим тройные взаимодействия, а также временные задержки.

За основу был взят алгоритм MIDER, его доработка учетом тройных взаимодействий позволила улучшить его точность, что было продемонстрировано на площадке DREAM Challenge, на испытаниях [REFERENCE]. На каждом из испытаний, продемонстрированная точность была минимум на [INSERT] выше, чем алгоритм без учёта тройных взаимодействий. Применение алгоритма DTW для учёта временных задержек также позволило улучшить точность. Данный алгоритм может быть эффективно применён как для исследования реальных биологических сетей, так и для других задач анализа взаимодействий, на выборках объёмом не более [INSERT] элементов. Предложенные методы оптимизации, такие как регуляризация Шмидта, выделение значимого подпространства, позволяют уменьшить вычислительную сложность алгоритма минимум в [INSERT] раза, и дальнейшие изыскания в области оптимизации алгоритма позволят увеличить возможное количество элементов сети до порядка [INSERT].

Нахождение взаимодействий более высокого порядка(четырёх и более) является перспективной задачей, поскольку учёт таких взаимодействий позволяет получить более точную картину сети. В особенности, если учитываются временные задержки. Дальнейшая работа в этой области будет направлена на увеличение порядка учитываемых взаимодействий, а также на оптимизацию существующих методов.

# Works Cited

|  |  |
| --- | --- |
| [1] | G. Mendel, "Versuche über Pflanzen-Hybriden," *Proceedings of the Natural History Society of Brünn,* 1866. |
| [2] | J. C. B. X. T. S. B. J. A. G. M. K. &. D. V. Sarah R Gilman, "Schizophrenia Gene Networks Found, with Link to Autism," NATURE NEUROSCIENCE, 2012. |
| [3] | Y.-C. H. C.-M. C. G. S. S. Ci-Ren Jiang, "Inferring Genetic Interactions via a Data-Driven Second Order Model," National Center for Biotechnology Information, 2012. |