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

**Введение**

Анализ взаимодействия генов берет свое начало в 19 веке [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. Так, например, группа ученых обнаружила взаимодействующие гены, ответсвенные за развитие шизофрении [2].

Набор взаимодействующих генов удобно визуально представлять в виде сети, в которой узлами являются гены, а взаимодействия обозначаются направленными отрезками. Выявление структуры генной сети является одной из задач анализа взаимодействий генов. Разнообразие генов и разнообразие их взаимодействий делают эту задачу вычислительно трудоемкой, что требует больших затрат при решении без использования машинных средств. Прямое моделирование процесса взамодействий не всегда является возможным ввиду ограниченности экспериментальных данных полностью описывающих структуру, поведение и функции того или иного гена. Классом эффективных алгоритмов являются алгоритмы, работающие 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. Такой подход позволяет уменьшить трудоемкость задачи, снизить человеческие ресурсы, необходимые для ее решения.

Помимо парных взаимодействий[REFERENCE], в генных сетях присутсвуют взаимодействия более высокого порядка [REFERENCE](согласованного взаимодействия более чем двух элементов) [3]. 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 алгоритм, учитывающий взаимодействия третьего порядка с временными задержками на основе алгоритма MIDER[REFERENCE]. 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].

В работе показано, что доработка алгоритма MIDER учетом тройных взаимодействий, а также изменение алгоритма учета временных задержек на алгоритм DTW увеличивает точность алгоритма на [Insert] процентов. Данные для тестирования были взяты с открытой площадки Dream Challenge. Сложность алгоритма была уменшена в [INSERT] раз за счет применения регуляризации Шмидта, что позволило увеличить количество элементов в исследуемых сетях до [INSERT]. Данный алгоритм может быть эффективно применён как для исследования реальных биологических сетей, так и для других задач анализа взаимодействий.

**Заключение**

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

Доработки алгоритма MIDER учетом тройных взаимодействий, алгоритмом DTW позволили 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. 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].

Реализация алгоритма доступна по ссылке [REFERENCE]. Данная утилита, реализованная в среде MatLab, позволяет сократить трудоемкость задачи выявления структуры генной сети, сократить трудозатраты ученых-генетиков, представляет ценность как инстурмент в решении задач выявления структуры генных сетей.

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. |