

## ПАРАЛЛЕЛЬНАЯ РЕАЛИЗАЦИЯ МЕТОДА ПОИСКА ЗАКОНОМЕРНОСТЕЙ В ПОСЛЕДОВАТЕЛЬНОСТЯХ СОБЫТИЙ

**Вишневский Валерий Викторович**

*E-mail: valera.vishnevskiy@yandex.ru*

*Кафедра математических методов прогнозирования*

*Научный руководитель: к.ф.-м.н., н.с. Ветров Дмитрий Петрович*

В данной дипломной работе предлагается оригинальный алгоритм для поиска скрытых закономерностей(паттернов) в последовательностях событий, основанный на вероятностном представлении паттернов в дискретных последовательностях событий, рассматривается применение данного алгоритма для анализа поведения мышей.

Задача поиска закономерностей в поведении животных и людей крайне важна в современной нейробиологии и когнитивных науках. Выделив характерные паттерны, можно, например, делать выводы о сложности поведения различных особей, определять изменения в поведении наблюдаемых процессов, другими словами, решив задачу поиска паттернов, можно определенным образом *измерять* поведение. Анализ поведения является основным инструментом при исследовании механизмов работы памяти и обучения на системном уровне .

Используемый на сегодняшний день метод поиска Т-паттернов, предложенный М.С. Магнуссоном в [1] имеет два ключевых недостатка: сильная чувствителен к шуму; полученные Т-Паттерны специфичны особи, в поведении которой они были найдены, т.е., на практике, невозможно определять степень «похожести» поведения разных особей.

Разработанный в данной дипломной работе метод, с одной стороны, основывается на зарекомендовавшей себя и верифицированной идее Магнуссона: поиск паттернов производится снизу вверх, сначала находятся простые закономерности, потом, путем их соединения, образуются более сложные паттерны. В виду введенной вероятностной модели Р-паттерна(probabilistic pattern) были полностью пересмотрены этапы конструирования, редукции, и поиска паттернов. В рамках данной научной работы были получены оценки на уровень «четкости» паттернов, разработан статистический критерий для определения информативных закономерностей(более подробно см. [2, 3]).

В [2, 3] показано, что разработанный подход расширяет метод поиска Т-паттернов: гарантировано нахождение, как минимум, тех же закономерностей, которые нашел метод Магнуссона. Более того, показано, что предложенный метод может обрабатывать шум и пропуски в исходных данных, и чаще находит более длинные и значимые закономерности.

Метод поиска Р-паттернов имеет сложность  $O(n^3)$ (алгоритм поиска Т-паттернов —  $O(n^2)$ ) от числа событий  $n$ . Была разработана параллельная версия алгоритма для GPU, работающая примерно в 40 раз быстрее последовательной версии. В итоге, характерная 12-ти минутная сегмента-

ция поведения лабораторной мыши обрабатывается примерно 11 секунд.

В заключении работы был проведен анализ работы алгоритма на реальных данных поведения мышей. Имелась разметка поведения двух групп животных: здоровая контрольная группа, и группа особей с поврежденной функцией гиппокампа. Были решены следующие задачи.

- *Поиск закономерностей.* Эксперты подтвердили, что найденные Р-паттерны имеют четкую, нетривиальную структуру и полностью согласуются с современными представлениями о поведении.
- *Классификация особей на группы по поведению.* Для Р-паттернов получено качество классификации порядка 90–95% правильных ответов, для Т-паттернов примерно 65–70%.
- *Поиск характеристических определенной группе животных паттернов.* При этом оказалось, что в «гиппокампальной» группе была особь, имеющая много паттернов характерных контрольной группе. Данный факт согласовался с экспериментом: после гистологического анализа выяснилось, что гиппокамп одной из мышей не удалось полностью парализовать.

Таким образом, разработанный в рамках дипломной работы метод поиска Р-паттернов, благодаря вариабельности найденных закономерностей между разными особями, позволяет описывать поведение конкретного животного, вектором «откликов» на найденные паттерны, после чего, могут быть использованы стандартные алгоритмы машинного обучения для решения, например, задач классификации, кластеризации, или восстановления регрессии. Подчеркнем, что данный подход не имеет аналогов в области анализа поведения. Похожий подход *мешка слов* (*bag-of-words*) активно и успешно используется в области компьютерного зрения, что дает повод рассчитывать на успешность схожих идей в области анализа поведения.

Одна из нетривиальных задач, для решения которой можно использовать предложенный метод — оценка «сложности» поведения и проверка гипотезы соответствия сложности поведения филогенетическому дереву.

### Литература

1. M.S. Magnusson. *Discovering hidden time patterns in behavior: T-patterns and their detection.* — Behavior Research Methods, Instruments, Computers 2000.
2. V.V. Vishnevskiy, D.P. Vetrov. *The Algorithm for Detection of Fuzzy Behavioral Patterns.* — Proceedings of Measuring Behavior 2010, ISBN 978-90-74821-86-5.
3. В.В. Вишнеvский. *Параллельная реализация метода поиска закономерностей в последовательностях событий.* — Дипломная работа. Факультет ВМК МГУ имени М. В. Ломоносова, 2011.