**Разработка методов выделения стереотипии в поведении животных**

Проведенная НИР по разработке методов выделения стереотипии животных, по сути, состояла из четырех основных этапов:

* анализ существующих методов, программная реализация одного из них,
* разработка новой модели вероятностных паттернов,
* параллельная реализация методов поиска закономерностей,
* разработка способов применения предложенного метода для экспериментов с реальными данными.

Первый этап заключался в анализе существующих методов исследования поведения. Была разработана программная реализация одного из самых широко распространенных на сегодняшний день методов, предложенного М.С. Магнуссоном –– алгоритм поиска Т-паттернов. Полученная программная реализация была испытана на реальных поведенческих данных, результаты работы программы полностью согласовывались с оригинальным алгоритмом Магнуссона(Noldus THEME).

Исходный алгоритм поиска Т-паттернов обладает рядом недостатков. Во-первых, сама модель Т-паттернов сильно чувствительна к шуму в исходных данных и не может обрабатывать пропуски событий внутри паттерна, а реальные данные сегментации поведения по данным видеонаблюдения, могут иметь достаточно много неточностей. Во-вторых, Т-паттерны сильно специфичны особи, в поведении которой они были найдены. Более неформально, если у двух разных особей имеются реальные одинаковые поведенческие паттерны, то мы не сможем автоматически определять по виду Т-паттернов, что разные Т-паттерны соответствуют какой-то одной определенно стереотипии поведения. Третий недостаток, отчасти являющийся следствием предыдущего, связан с тем, что из-за дискретной природы Т-паттернов, затруднена процедура редукции множества паттернов, из-за чего на выходе программы остается много похожих Т-паттернов, среди которых эксперт должен выделять значимые.

Анализ приведенных выше недостатков привел к созданию нового алгоритма поиска паттернов, основывающегося на уже хорошо зарекомендовавшем себя методе поиска Т-паттернов. Была предложена новая вероятностная модель P-паттерна(probability pattern), которая, по построению, могла обрабатывать пропуски в закономерностях. Основная идея метода осталась прежней: поиск производится снизу вверх, сначала находятся простые закономерности, потом, путем их соединения, образуются более сложные паттерны. Однако из-за новой модели паттерна пришлось полностью пересмотреть этапы конструирования, редукции, и поиска паттернов.

В целом, 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. Однако если алгоритм поиска Т-паттернов имеет квадратичную сложность от количества событий в данных, то разработанный метод поиска P-паттернов имеет кубическую от числа событий сложность.

Для возможности применять метод поиска вероятностных паттернов на практике, была разработана параллельная версия алгоритма для архитектуры графических устройств(GPU). Параллельная GPU-версия программы, на практике, при использовании рядовых GPU, работает примерно в 40 раз быстрее последовательной версии.

Был проведен анализ работы алгоритма на реальных данных поведения мышей. Имелась разметка поведения двух групп животных: здоровая контрольная группа, и “гиппокампальная” особей с поврежденной функцией гиппокампа. Было решены следующие задачи.

* Выделение поведенческих паттернов. Анализ экспертов подтвердил, что найденные паттерны имеют четкую, характерную структуру и имеют биологический смысл.
* Классификация особей на группы по поведению. Получено качество классификации порядка 90–95% правильных ответов, против 65–70%, при использовании метода Т-паттернов.
* Поиск характеристических определенной группе животных паттернов.

Представленный метод решает поставленные перед ним задачи и производит качественный поиск закономерностей как в синтетических временных рядах, так и в реальных поведенческих данных. В открытый доступ выложена свободная, документированная, эффективная реализация представленного метода. Главным преимуществом метода поиска P-паттернов является их вариабельность: если на то есть предпосылки, то P-паттерны найденные в поведении одной особи будут также найдены в поведении другой особи. Данный факт позволяет описывать поведение на основе найденных паттернов, и использовать стандартные алгоритмы машинного обучения для решения, например, задач классификации, кластеризации, или восстановления регрессии.

В течение НИР были написаны и опубликованы три статьи и доклады на конференциях:

* Публикация в сборнике научной сессии НИЯУ МИФИ –2010 “Материалы избранных научных трудов по теме: Актуальные вопросы нейробиолгоии, нейроинформатики и когнитивных исследований”.   
  В.В. Вишневский, Д.П. Ветров. Обзор методов генерации паттернов во временных рядах.
* Международная конференция Measuring Behavior 2010, публикация в сборнике докладов и устное выступление.   
  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.
* Публикация в сборнике докладов 15-ой Всероссийской конференции “Математические методы распознавания образов 2011”. Статья принята к публикации.  
  В.В. Вишневский, Д.П. Ветров. Вероятностный подход к поиску поведенческих паттернов.

Также по данной теме на оценку “отлично” была защищена дипломная работа на факультете вычислительной математики и кибернетики Московского государственного университета им. М.В. Ломоносова. К печати готовится статья в один из журналов из списка ВАК.