**Макет отчета по второму этапу НИР**

**"Компьютерные методы поведенческого фенотипирования лабораторных животных"**

**Аннотация (3 – 5 стр.)**

Конушин (>1 стр.)

Ветров (>1 стр.)

В данной работе предложен новый метод для поиска скрытых закономерностей(паттернов) в последовательностях событий, основанный на вероятностном представлении P-паттернов(probabilistic pattern) в дискретных последовательностях событий. Поиск производится снизу вверх: сначала находятся простые закономерности, потом, путем их соединения, образуются более сложные паттерны. Рассматривается применение данного алгоритма для анализа поведения мышей. Найденные паттерны используются для классификации животных. Проведено сравнение реализованного алгоритма с существующими аналогами, показавшее, что предложенный метод более устойчив к шуму в исходных данных. Рассмотрены новые подходы к исследованию поведения.

Сергиевский (1 стр.)

Мишулина (1 стр.)

Зарайская И.Ю., Долгов О.Н. (0.5 стр.)

**СПИСОК ОСНОВНЫХ ИСПОЛНИТЕЛЕЙ**

**по Государственному контракту от 23 июля 2009 г. № П264 на выполнение поисковых научно-исследовательских работ для государственных нужд**

Организация-Исполнитель: **Федеральное государственное бюджетное образовательное учреждение высшего профессионального образования «Национальный исследовательский ядерный университет «МИФИ»**

**Руководитель проекта:**

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| № п/п | ФИО | дата рождения | ученая степень, ученое звание | место работы | должность | подпись |
| 1. | Мишулина Ольга Александровна | 04.05.39 | к.т.н., доцент | НИЯУ "МИФИ" | доцент |  |

**Основные Исполнители проекта:**

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| 1. | Ветров Дмитрий Петрович | 17.08.81 | к.ф-м.н. | ВМиК МГУ | научный сотрудник |  |
| 2. | Конушин Антон Сергеевич | 17.08.81 | к.ф-м.н. | ВМиК МГУ | научный сотрудник |  |
| 3. | Сергиевский Георгий Максимович | 25.04.41 | к.т.н., доцент | НИЯУ «МИФИ» | доцент |  |
| 4. | Ушаков Вадим Леонидович | 06.04.71 | к.б.н., доцент | НИЯУ «МИФИ» | доцент |  |
| 5 | Зарайская Ирина Юрьевна |  |  |  |  |  |
| 6 | Долгов Олег Николаевич |  |  |  |  |  |
| 7. | Суконкин Илья Николаевич | 14.01.87 | – | НИЯУ "МИФИ" | аспирант |  |
| 8. | Бесчастный Леонид Леонидович | 23.07.89 | – | НИЯУ "МИФИ" | студент |  |
| 9. | Эйдлин Александр Андреевич | 07.04.89 | – | НИЯУ "МИФИ" | студент |  |
| 10. | Синдеев Михаил Сергеевич | 18.09.86 | – | ИПМ им. М.В.Келдыша | аспирант |  |
| 11. | Темлянцев Александр Валерьевич | 20.06.87 |  | ВМиК МГУ | аспирант |  |
| 12. | Вишневский Валерий Викторович | 21.03.89 | – | ВМиК МГУ | студент |  |

**Реферат (2-3 стр.)**

Конушин (0,5 стр.)

Ветров (0,5 стр.)

Сергиевский (0,5 стр.)

Мишулина (0,5 стр.)

Зарайская И.Ю., Долгов О.Н. (0.5 стр.)

**Введение**

Конушин (>0,5 стр.)

Ветров (>0,5 стр.)

Задача поиска закономерностей(стереотипов, паттернов, шаблонов –– здесь синонимы) в поведении животных и людей крайне важна в современной нейробиологии и когнитивных науках. Выделив характерные паттерны, мы, например, можем делать выводы о сложности поведения различных особей, определять изменения в поведении наблюдаемых процессов, другими словами, решив задачу поиска паттернов, мы можем определенным образом измерять поведение особи, или группы особей, становится возможно более наглядно анализировать поведение.

Именно анализ поведения является основным инструментом при исследовании на системном уровне механизмов работы памяти и обучения животных.

Сергиевский (0,5 стр.)

Мишулина (0,5 стр.)

Зарайская И.Ю., Долгов О.Н. (0.5 стр.)

*Во введении промежуточного отчета по этапу НИР должны быть приведены цели и задачи этапа исследований, их место в выполнении НИР в целом.*

# Сокращения и обозначения

|  |  |
| --- | --- |
| ГС | геометрическое состояние |
| ДЖ | действие животного |
| ИД | исследовательское действие |
| ЛЖ | лабораторное животное |
| МВС | модель внешней среды |
| МС | мотивационное состояние |
| ПВ | протокол вычислений |
| ПВФ | протокол во внутренней форме |
| ПП | программа поведения |
| ПС | пространство состояний |
| СВ | состояние вычислений |
| СММ | скрытые марковские модели |
| ССПП | система синтеза программы поведения |
| ФД | физическое действие |
| SMP | symmetric multiprocessor |

# Развитие алгоритмического и программного обеспечения системы видеонаблюдения

# за лабораторными животными

* 1. ***Разработка алгоритмов и программ точного выделения контуров животных в видеопоследовательностях***

**[Конушин А.С., Синдеев М.С.]**

*Приводится описание алгоритмов и результатов тестирования на реальных данных, а также краткая характеристика программной реализации.*

* 1. ***Методы* *отслеживания нескольких лабораторных животных, одновременно находящихся в одной арене***

**[Ветров Д.П., Конушин А.С.]**

*Приводится описание алгоритмов и результатов тестирования на реальных данных, а также краткая характеристика программной реализации.*

1. **Анализ поведения лабораторных животных**
   1. ***Разработка вероятностных методов выделения стереотипии в поведении* [Ветров Д.П., Вишневский В.В.]**

**2.1.1 Определения и обозначения**

Пусть время наблюдения разбито на интервалов. В каждый момент  *периода наблюдения* может произойти некоторое событие (*действие, поведенческий акт, event*) [[1]](#footnote-2) из множества допустимых событий (*event types*). Соответственно, каждому типу события сопоставляется множество моментов времени :

здесь –– количество появлений события в данных.

Нечетким паттерном(далее P-Pattern, или просто паттерн) длины назовем упорядоченную последовательность событий , где каждое событие паттерна характеризуется смещением и разбросом от предыдущего события. Будем записывать паттерн в следующем виде:

Здесь и –– смещение и разброс соответствующего события относительно предыдущего(см. Рис. 1). Распределение межточечных расстояний [6, c.138] между событиями в паттерне моделируется нормальным распределением. Если нам не важны параметры смещения разброса событий, то паттерн записываться так: , если нам важна иерархия паттерна, то например так:

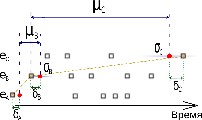


Рисунок 1: Представление нечеткого паттерна .

Далее, чтобы иметь возможность обрабатывать пропуски в реализации паттернов, введем понятие  *функции потерь*, которая определяет “штраф” за пропуск событий в паттерне длины следующим образом:

Здесь является структурным параметром, определяющим максимальный допустимый уровень “нечеткости” паттернов. Если этот параметр велик, то мы, по сути, запрещаем реализациям паттерна иметь пропуски. Если выставить этот параметр слишком малым, то будут обнаруживаться паттерны, не разу полностью не встречающиеся в данных, то есть закономерности могут быть найдены даже в случайных данных. Данный параметр должен выставляться вручную, исходя из априорной информации о типе данных, уровне шума, и сложности поведения.

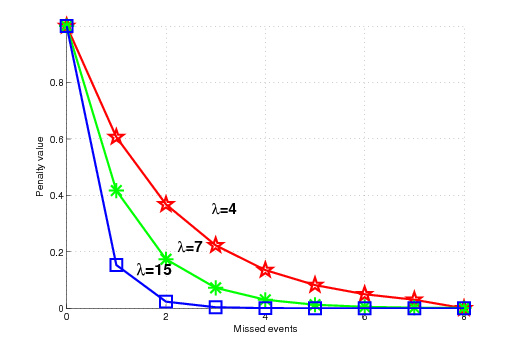


Рисунок 2: Функция потерь для паттерна длины 8 при разных значениях параметра .

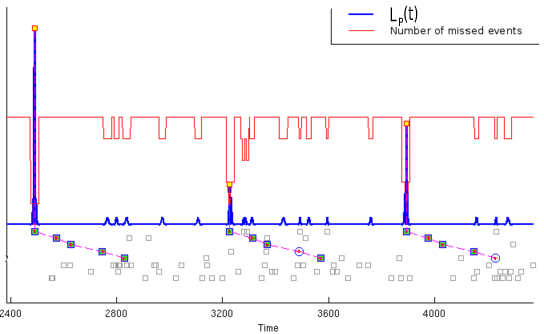


Рисунок 3: Пример функции правдоподобия паттерна. Желтыми маркерами с красной границей изображены максимумы функции правдоподобия: моменты времени, когда мы считаем, что паттерн имеет место. Значение функции правдоподобия в максимуме будет весом -го появления паттерна в данных. В нижней части рисунка закрашенными квадратами показаны присутствующие события, полыми кружками –– пропущенные события в паттерне. Полые серые квадраты соответствуют наблюдаемым поведенческим актам.

Теперь мы можем определить ключевое понятие для представленного метода –– *правдоподобие паттерна*. Правдоподобие паттерна –– это функция, определенная для каждого момента времени наблюдения , показывающая на сколько четко можно говорить, что данный паттерн начинается в данный момент времени . Более формально:

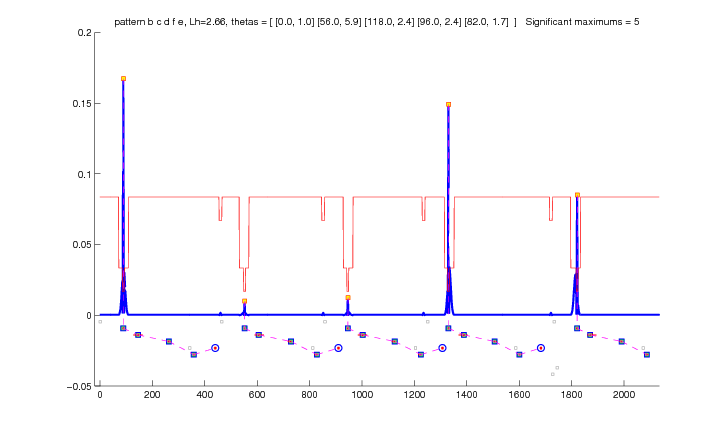
(1)

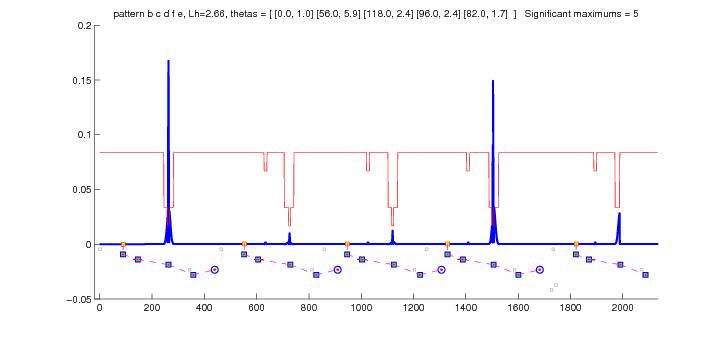
здесь –– расстояния от ожидаемой позиции события в паттерне до ближайшего события в данных(более наглядно см. Рис. 1). Т.е.:

здесь, если событие было пропущено, то соответствующее . Далее, –– количество пропущенных событий в паттерне, а –– множество индексов присутствующих в паттерне событий. Событие считается пропущенным, если , т.е. соответствующее значение больше определенного предела.

Заметим, что правдоподобие паттерна можно считать не только начиная с первого события, но и, например, с конца паттерна. Для упрощения вычислений, значение правдоподобия паттерна с события можно рассчитывать по следующей формуле:

На рисунке 4 видно, что правдоподобие, посчитанное с -го события имеет максимумы в момент наступления -го события в текущем паттерне. Данный факт будет активно использоваться при вычислении межточечных расстояний между паттернами и в процессе редукции паттернов-дубликатов.



 Рисунок 4: Функция правдоподобия одного и того же паттерна, подсчитанная, начина с первого события (сверху) и с третьего события (снизу).

**2.1.2** **Поиск “значимых” максимумов функции правдоподобия**

Определенная выше функция правдоподобия P-Паттерна имеет много локальным максимумов, однако при визуальном рассмотрении, обычно видно несколько ярко выраженных максимумов, в точках, которые можно считать началом вхождения паттерна. На шаге поиска новых закономерностей нам придется иметь дело с конкретными моментами времени, в которых имело место начало паттерна. Поэтому требуется определить, какие именно точки максимумов действительно соответствую началам паттернов.

Предположим, что в определенный момент времени имело место начала модельного P-Паттрена(без пропусков событий) длины . Запишем функцию правдоподобия:

При таких условиях вычислим математическое ожидание :

Таким образом, нами получено характеристическое значение правдоподобия P-Паттерна длины с параметрами разброса . После чего, при поиске “значимых” максимумов отсекаются все локальные максимумы, значение которых меньше , где –– заданное число от до , в наших экспериментах обычно подходит значение . Если среди оставшихся максимумов есть точки, которые удалены друг на друга меньше длины паттерна, то берется наибольший из них. Этот шаг объясняется тем, что один P-Паттерн не может иметь место в один и тот же момент времени.

**2.1.3** **Процедура конструирования P-Паттернов**

Рассмотрим два P-Паттерна (левый) и (правый). Пусть и –– значения в “значимых” максимумах функций правдоподобия и соответственно. Заметим, что правдоподобие левого паттерна отсчитывается с конца, так как мы ищем связь между концом левого паттерна и началом правого. Также пусть, и –– индексы этих “значимых” максимумов. и –– количество таких максимумов у паттернов и соответственно. Определим множество межточечных расстояний:

Для каждого расстояния из этого множества введем соответствующий вес:

где

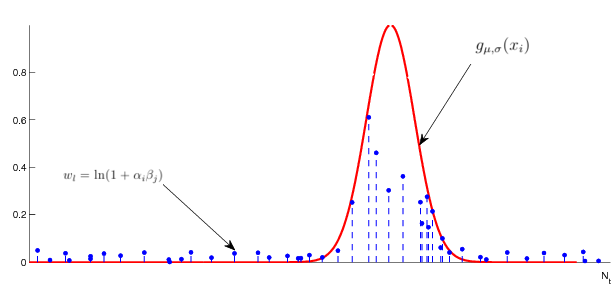


Рисунок 5: Пример распределения межточечных расстояний и статистической модели связи.

Примем гипотезу, что координаты “значимых” максимумов распределены независимо и равномерно на всем наблюдаемом промежутке, т.е. . Тогда очевидно, что плотность распределение введенных выше межточечных расстояний имеет следующий вид:

Введем статистическую модель связи между паттернами(проверяемые параметры связи и фиксированы):

Рассмотрим следующую сумму:

(2)

где . О распределении значений мы не можем делать никаких предположений, поэтому в вычислениях будем использовать выборочные моменты.

Для оценки распределения случайной величины проведем требуемые вычисления.

Далее, считаем параметры случайной величины . Здесь и –– выборочное среднее и дисперсия весов, соответственно.

Запишем дисперсию произведения независимых случайных величин:

Тогда по Центральной Предельной Теореме:

Теперь можно провести сравнение подсчитанного на реальных данных значения с правой односторонней квантилью уровня . Если односторонняя гипотеза о равномерном распределении событий с моделью связи отвергается, то из паттернов и конструируется новый паттерн длины :

Если существует несколько пар и , для которых отвергается гипотеза, то для конструирования P-Паттернов берутся непересекающиеся[[2]](#footnote-3) параметры соответствующие максимальным значениям .

**2.1.4** **Процедура редукции множества P-Паттернов**

Так же как и в методе Магнуссона, наш алгоритм может конструировать P-Паттерны являющиеся дубликатами, или неполными копиями уже обнаруженных P-Паттернов. Далее дадим более полное описание этой проблемы.

**Паттерны-дубликаты**:

Один и тот же паттерн может быть сконструирован из разных подпаттернов. Например P-Паттерн может быть получен как путем соединения паттернов и , так и паттернов и . Обычно такие P-Паттерны представляют одну и ту же поведенческую закономерность, но из-за сложного процесса соединения паттернов, функции правдоподобия этих двух паттернов-дубликатов могут отличаться.

**Неполные копии паттернов**:

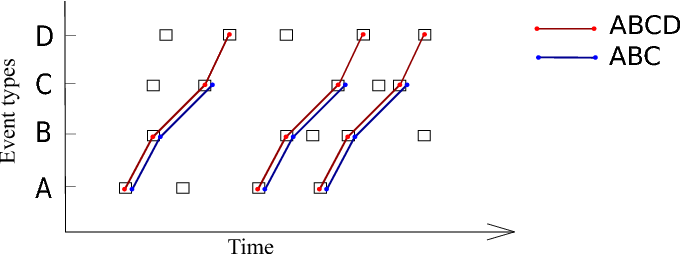
Конструируя паттерны из подпаттернов, некоторые подпаттерны могут появляться в данных только в составе более сложных составных паттернов(см. Рис. 6). Такие подпаттерны не имеет смысла рассматривать отдельно, поэтому их надо удалять из множества найденных P-Паттернов.   
  


Рисунок 6: Здесь –– неполная копия паттерна . Заметьте, что правдоподобия и отсчитанное от первого события будут похожими.

Для удаления таких P-Паттернов предлагается анализировать коэффициент корреляции функций правдоподобия. Пусть –– вектор-столбец значений функции правдоподобия, отсчитанной от -го события во всех моментах времени наблюдения.

–– коэффициент корреляции между двумя P-Паттернами. Чем он ближе к , тем два паттерна более близки друг к другу.

Процедура редукции паттернов выглядит следующим образом: перебираем все пары из множества найденных паттернов. Если все поведенческие акты, присутствующие в паттерне также присутствуют в с учетом порядка, и

тогда паттерн удаляется из множества найденных паттернов. Отметим, что в данном методе не удаляются псевдопаттерны(паттерны состоящие из одного события), так как они могут быть необходимы для конструирования закономерностей на следующих шагах.

**2.1.5** **Алгоритм поиска P-Паттернов**

После описания основных шагов конструирования и редукции P-Паттернов, можно описать сам алгоритм поиска:

1. Инициализировать текущее множество паттернов псевдопаттернами.

2. Для всевозможных пар паттернов из текущего множества для которых не было произведено попытки их слития, провести процедуру конструирования паттернов. Сконструированные паттерны, которые встречаются в данных не менее раз, добавить в текущее множество.

3. Для всевозможных пар паттернов из текущего множества, провести процедуру редукции паттернов.

4. Если текущее множество паттернов изменилось, перейти к п.2.

Очевидно, что описанный выше метод остановится, так как на каждом шаге будут произведены попытки сконструировать паттерны все бльшей длины, а одни и те же паттерны не проверяются больше одного раза.

Предложенный в данной работе алгоритм обладает рядом параметров, которые не настраиваются автоматически. Эти параметры должны быть выставлены вручную исследователями, исходя из априорных сведений о типе исходных данных, или ожидаемого вида поведенческих закономерностей. Список этих структурных параметров с объяснениями и значениями по-умолчания представлен в Таблицу 1.

|  |  |  |  |
| --- | --- | --- | --- |
| **Параметр** | **Возможные значения** | **Значения поумолчанию** | **На что влияет** |
|  |  |  | Уровень значимости паттерна |
|  |  | 3 | Минимальное количество появлений паттерна в данных |
|  |  | 8 | Допустимая степень нечеткости паттерна |
|  |  | 0.6 | Минимальная степень похожести паттернов для удаления |
|  |  | 0.4 | Чувствительность к отклонению от ожидаемого правдоподобия |

Таблица 1: Параметры алгоритма поиска P-Паттернов.

Отметим, что предложенный алгоритм является отчасти переборным, а самыми вычислительно-сложными процедурами являются подсчет функции правдоподобия для каждого паттерна, отсчитанная для каждого события, и подбор параметров смещения и разброса при попытке конструирования паттернов.

**2.1.6** **Параллельная реализация алгоритма поиска “нечетких” закономерностей(P-Паттерны)**

Алгоритм поиска P-Паттернов является еще более вычислительно-сложным, чем алгоритм поиска Т-Паттернов, поэтому для его применения на реальных данных крайне важна параллельная реализация. Для распараллеливания алгоритма была использована технология NVIDIA CUDA, позволяющая создавать программы, в том числе, для современных игровых видеокарт, которые широко распространены и имеют один из самых выгодных показателей FLOPS(FLoating point OPerations per Second) за доллар.

Подробное описание архитектуры GPU от NVIDIA и технологии CUDA можно найти, например, в [11, 12]. Отметим лишь, что логически все вычисления на GPU разбиваются на блоки(*blocks*), блоки, в свою очередь, разбиваются на нити(потоки, *threads*). Нити из одного блока могут синхронизироваться и обмениваться данными через высокоскоростную shared-память. Нити из одного блока физически выполняются на одном мультипроцессоре  *half-warp'*ами(группы по 16 нитей), что нужно учитывать при доступе к различным видам памяти. Также существуют доступные и крайне эффективные [11, 12] схемы одного из важнейших параллельных примитивов: параллельной редукции. Крайне важно избегать ветвлений внутри  *half-warp'*ов, так как мультипроцессор исполняет одновременно одни и те же операции для нескольких разных данных(схема SIMD).

При этом, между нитями на CPU и нитями на GPU есть принципиальные различия [12]:

• нити на GPU обладают крайне “небольшой стоимостью” –– их создание и управление требует минимальных ресурсов (в отличии от CPU);

• для эффективной утилизации возможностей GPU нужно использовать многие тысячи отдельных нитей (для CPU обычно нужно не более 10-20 нитей)

В нашей реализации на GPU были перенесены две самые вычислительно-сложные процедуры: конструирование паттернов, вычисление правдоподобия паттерна. Предложенная конкретная реализация проектировалась с учетом спецификаций устройств NVIDIA Compute Capability 1.1. Данная спецификация определяет, например, доступные инструкции чипа, объем типов памяти, количество регистров.

Для параллельной версии конструирования паттернов использовался следующий подход: разные блоки соответствуют разным  *парам* тестируемых параметров и для двух паттернов. В shared-память каждого блока загружается множество межточечных расстояний, после чего эффективным образом вычисляется статистика (2) для каждой данной пары параметров. Данная статистика сравнивается с фиксированным порогом и принимается решение о добавлении паттерна в текущее множество.

Для параллельной версии процедуры вычисления правдоподобия использовался следующий подход: моменты времени, в которых вычисляется правдоподобие разбиваются на участки, соответствующие блокам, каждый участок, в свою очередь, состоит из моментов времени, каждый из которых соответствует нити внутри блока. Далее, производится процесс редукции для поиска ближайшего элемента паттерна и по формуле (1).

Вычислительные эксперименты(см. Рис. 7, 8) проводились на устройстве NVIDIA GeForce 8800GTX с 128 потоковыми процессорами, вычисления проводились с одинарной точностью(single precision floating-point format). Для процедуры конструирования паттернов удалось достичь ускорения работы в 15­–20 раз, а для процедуры вычисления правдоподобия было получено ускорение в 120–140 раз, что соответствует максимальному порядку ускорения задач для данной архитектуры. Общее ускорение работы программы, в среднем, 40 раз. Например, при разметке поведения на 80 поведенческих актов(), последовательная версия на используемом оборудовании работает 423 сек., параллельная версия –– 11 сек. Скорее всего, для современных GPU(GeForce 8800GTX был выпущен в 2006-ом году), имеющих большее количество потоковых процессоров и shared-памяти, возможно создать еще более эффективную реализацию, которая будет проводить вычисления уже с двойной точностью(double precision floating-point format).

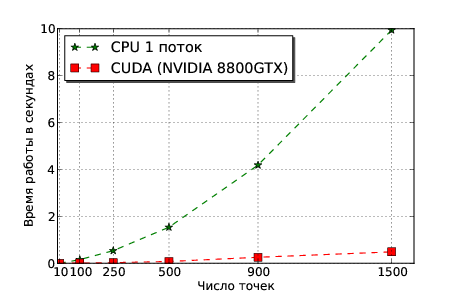
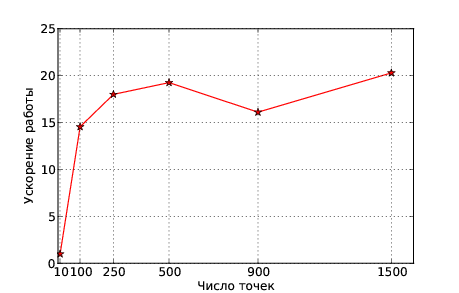
 

Рисунок 7: Время работы и ускорение процедуры конструирования P-Паттернов на GPU и CPU.

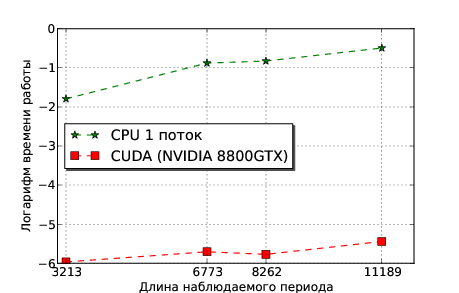
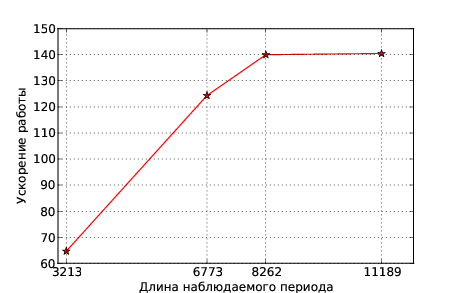
 

Рисунок 8: Время работы и ускорение вычисления правдоподобия P-Паттернов на GPU и CPU.

**2.1.7** **Эксперименты на реальных данных**

Описанный ниже эксперимент демонстрирует способ применения предложенного метода на реальных поведенческих данных. Целью эксперимента является анализ того, как влияет отсутствие гиппокампа на поведение. Гиппокамп –– один из древнейших отделов головного мозга млекопитающих, его функции связывают с механизмами работы памяти(консолидация информации из краткосрочной в долгосрочную память), обучением, пространственной навигацией [15, c. 744].

Мы располагали данными о поведении мелких мышеобразных грызунах –– рыжих полевках. Исследуемые особи разбиваются на две группы по 12 в каждой. Грызуны из первой группы(контрольной) –– нормальные, здоровые особи, грызуны из второй группы –– гиппокамп данных животных разрушали путем введения в эту структуру лидокаина, растворенного в искусственной спинномозговой жидкости (2 мкл. 4% раствора), лидокаин обладает местно-анестезирующим действием, блокирует потенциал-зависимые натриевые каналы, что препятствует генерации импульсов клетками. Используемая доза блокатора вызывала длительную инактивацию клеток структуры, что в дальнейшем приводило к их отмиранию и разрушению части структуры. Данные видеонаблюдения за грумингом(процесс умывания, очистки поверхности тела животным) каждой отдельной особи были размечены на поведенческие акты.

Для каждой из 24-ех особей был произведен поиск паттернов. Далее, для всех пар особей вычислялось число  *соответствий* паттернов, найденных у -ой особи в поведении -ой особи:

где –– множество паттернов, найденное в поведении особи , –– длина паттерна , –– количество найденных вхождений паттерна в поведении особи , –– скобки Аверсона. Параметр определяет минимальную длину паттернов, соответствия которых мы ищем, данный параметр имеет смысл ставить бльшим, или равным 3, так как короткие паттерны зачастую являются шумовыми. Параметр может принимать значения из , однако самыми осмысленными являются значения 0 и 1. При число соответствий между особями и равно сумме числа вхождений всех паттернов(длиннее чем ) особи в поведении особи . При число соответствий между особями и равно количеству  *разных паттернов* особи , встречающихся в поведении особи . Отметим, что при число соответствий более инвариантно к длине периода наблюдения за животным.

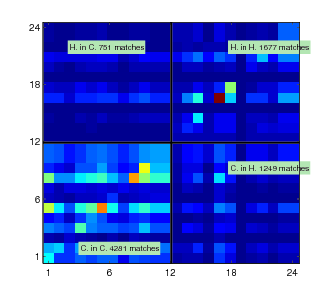


Рисунок 9: Таблица соответствий для , . По вертикали отложены значения , по горизонтали –– (неформально: по вертикали  *откуда* берутся паттерны, по горизонтали ––  *где* ищутся вхождения этих паттернов; например, в ячейке записано число соответствия паттернов третей особи в поведении десятой). Матрица, вообще говоря, не должна быть симметричной.

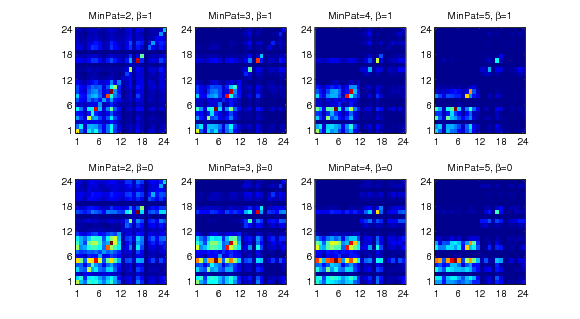


Рисунок 10: Разные способы подсчета числа соответствий.

На Рис. 9 и 10 видно, что исследуемые группы хорошо разделимы по введенной выше таблице соответствий.

Используя таблицу соответствий паттернов можно провести классификацию особей. В задаче классификации объектом возьмем вектор

здесь –– общее число особей(в нашем случае 24). Таким образом, объект, соответствующий особи является вектором <<откликов>> на паттерны -ой особи в поведении других особей.

Мы рассмотрели два метода решения задачи классификации: метод опорных векторов(SVM), метод ближайших соседей(NN). Также было проведено сравнение качества классификации, используя P-Паттерны и Т-Паттерны. Отметим, что в методе ближайших соседей упорядочивание соседей проводилось не по Евклидовой метрике между введенными выше векторами, а по количеству соответствий() паттернов данного объекта с объектами из обучения .

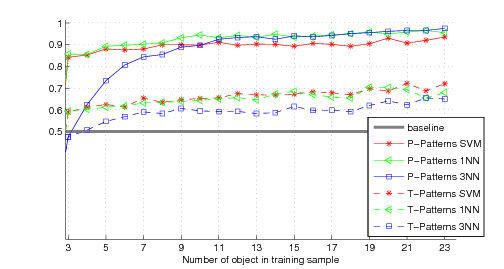


Рисунок 11: Средняя доля правильных ответов классификации методами SVM, 1NN, 3NN, с разными способами поиска паттернов, в зависимости от размера обучающей выборки. Параметры подсчета числа соответствий: , данные значения были взяты для обеспечения <<чистоты>> эксперимента.

Анализируя Рис. 11, можно сделать следующие выводы:

• классификация при использовании метода поиска P-Паттернов производится на порядок точнее, чем при использовании существующего метода поиска Т-Паттернов;

• для классификации особей, используя P-Паттерны, можно использовать любой из рассмотренных методов классификации, имея в виду, что метод ближайших соседей при имеет низкое качество классификации на короткой обучающей выборке;

• при классификации с использованием P-Паттернов доля правильных ответов, в среднем, равна 92%.

Также, в качестве эксперимента, в исходные данные были добавлены еще 2 группы: рандомизированные данные (события равномерно распределены по интервалу наблюдения) и данные с искусственными паттернами. Таблица соответствий и качество классификации показано на Рис. 12. Таким образом, мы продемонстрировали, что:

• предложенным методом можно проводить классификацию на несколько классов;

• группа особей с нарушением работы гиппокампа может быть отличена от рандомизированных данных, несмотря на то, что априорно, поведение таких особей “проще” и содержит меньше паттернов.

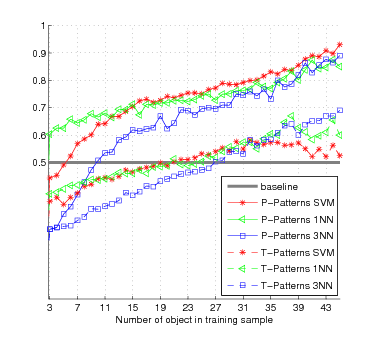
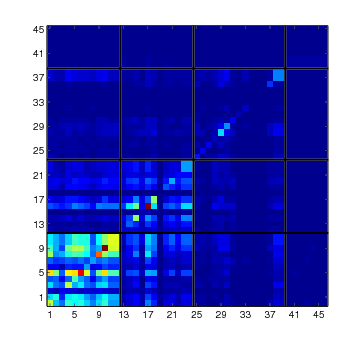
 

Рисунок 12: Справа: таблица соответствий , для , . Группы(слева на право): контрольная(12 элементов), мыши без гиппокампа(12 элементов), рандомизированные данные второй группы(15 элементов), искусственные паттерны(7 элементов). Слева: средняя доля правильных ответов классификации методами SVM, 1NN, 3NN.

**2.1.8** **Характерные паттерны**

Среди найденных паттернов можно выделить именно *характерные* для определенного класса паттерны. Неформально, паттерн является характерным для заданного класса, если он присутствует в поведении многих особей этого класса и редко выявляется в поведении животных из других классов. Для формализации данного понятия можно использовать Статистическое, энтропийное, или эмпирическое определение информативности [17].

**2.1.9** **Основные результаты. Обсуждение**

Реализованный в данной дипломной работе метод решает поставленные перед ним задачи и производит качественный поиск закономерностей как в синтетических временнх рядах, так и в реальных поведенческих данных. Предложенный метод поиска P-Паттернов имеет 2 главных недостатка:

• из-за статистической основы метода, при малых объемах данных, некоторые паттерны могут быть приняты за шум. С этим эффектом можно бороться путем более тщательного подбора параметров алгоритма.

• Алгоритм имеет экспоненциальную сложность и, несмотря на 140-кратное ускорение, будет крайне долго работать на очень длинных данных(на несколько порядков длиннее, чем стандартная разметка поведения мышей). Имеющиеся в нашем распоряжении поведенческие данные обрабатывались не дольше двух минут.Если данная проблема будет актуальна, ее можно решить, выставив ограничение на максимальную длину связи между событиями в паттерне.

В данной работе алгоритм поиска закономерностей рассматривался в контексте применения его для анализа поведения. Но, очевидно, заменив понятие поведенческого акта на какое-то абстрактное событие(маркер), мы можем искать закономерности в различных потоках данных. Например, событиями могут быть: повышения и понижения курсов валют в анализе поведения рынка; аминокислоты, кодоны, или азотистые основания при анализе структуры ДНК; всплеск активности отдельного нейрона при анализе спайковой активности нейронных культур; новостные тренды(в виде ключевых слов) при анализе закономерностей в политике и обществе. Например, алгоритм Т-Паттернов и ранее был использован для анализа структуры ДНК [13] и стратегии футбольных команд во время матчев [14].

* 1. ***Разработка методов композитного анализа поведения***

**[Ветров Д.П., Темлянцев А.В.]**

*Приводятся алгоритмы обучения и сегментации сигналов, а также результаты испытаний на модельных данных.*

* 1. ***Разработка языка описания внешней среды и внутреннего представления протокола видеонаблюдения для программной реализации модели мотивационного поведения животного* [Сергиевский Г.М.]**

*Приводится спецификация языка описания МВС. Дается техническое описание программы трансляции и результатов ее тестирования. Излагаются рекомендации по формированию программы модели поведения.*

1. **Автоматическая сегментация фаз поведения животного**
   1. ***Проведение исследований в интересах выбора состава* *информативных признаков для распознавания фаз поведения***

***и состояний животного***

**[Мишулина О.А., Суконкин И.Н., Бесчастный Л.Л.]**

*Излагается методический подход к проведению исследований, приводятся экспериментальные результаты и даются рекомендации по составу информативных признаков.*

* 1. ***Разработка алгоритмов автоматической сегментации***

***фаз поведения животного***

**[Мишулина О.А., Суконкин И.Н., Эйдлин А.А.]**

*Излагаются алгоритмы автоматической сегментации векторных временных рядов и результаты их исследования на модельных и реальных данных. Приводится качественный анализ сегментированных фаз поведения.*

1. **Результаты биологических экспериментов**
   1. ***Описание биологических экспериментов и данных системы видеорегистрации***

*Приводится содержание биологических задач первой серии экспериментов. Дается описание данных системы видеорегистрации.*

* 1. ***План проведения биологических экспериментов***

*Рассматриваются задачи и план проведения биологических экспериментов на весь срок проекта.*

*Общие требования к отчету* (*из Госконтракта*)

*Отчет должен содержать процесс теоретических и (или) экспериментальных исследований, включая определение характера и содержания теоретических исследований, методы исследований, методы расчета, обоснование необходимости проведения экспериментальных работ, принципы действия разработанных объектов, их характеристики, обобщение и оценку результатов исследований, включающих оценку полноты решения поставленной задачи и предложения по дальнейшим направлениям работ, оценку достоверности полученных результатов и их сравнение с аналогичными результатами отечественных и зарубежных работ, обоснование необходимости проведения дополнительных исследований, отрицательные результаты, приводящие к необходимости прекращения дальнейших исследований.*

**Заключение (2 стр.)**

Конушин (0,4 стр.)

Ветров (0,4 стр.)

Основные результаты работы:

* Разработан метод для поиска закономерностей в последовательностях событий.
* Для поиска поведенческих закономерностей предложенный метод дает результаты, как минимум, не хуже, чем широко зарекомендовавшие себя методы.
* Представлены свободные, документированные реализации методов поиска предложенного метода поиска P-Паттернов, а также метода поиска Т-Паттернов.
* Разработаны эффективные параллельные версии методов поиска P-Паттернов и Т-Паттернов.
* Пиковое ускорение для параллельной реализации поиска Т-Паттернов на 4-х ядерном CPU –– раза.
* Пиковое ускорение для параллельной реализации процедур поиска P-Паттернов на NVIDIA GeForce8800GTX –– 20 раз(этап конструирования) и 140 раз(этап подсчета правдоподобия).
* Для реальных экспериментов с мышами, предложенный метод выделяет паттерны, по которым можно определить из какой группы была взята наблюдаемая особь. То есть, по данным о поведения особей решена задача классификации.

Система, разработанная и реализованная в данной работе, позволит биологам решать ряд практически важных задач, возникающих при исследовании поведения: выделение поведенческих закономерностей животных, анализ вариабельности и сложности поведения, выявление отличительных черт поведения среди разных групп животных, анализ спайковой активности нейронов. Решение этих задач позволит проводить качественный анализ влияния медицинских препаратов на поведение животных, связи различных анатомических структур мозга и поведения.

Дальнейшая работа в данной области будет связана, во-первых, с исследованием возможности расширения множества задач, которые можно решить предложенным методом. Во-вторых, планируется разработать удобную для биологов графическую среду для исследования поведенческих закономерностей.

Сергиевский (0,4 стр.)

Мишулина (0,4 стр.)

Зарайская И.Ю., Долгов О.Н. (0.4 стр.)

*Заключение должно содержать:*

*- краткие выводы по результатам выполнений НИР или отдельных ее этапов;*

*- оценку полноты решений поставленных задач;*

*- разработку рекомендаций и исходных данных по конкретному использованию результатов НИР;*

*- оценку технико-экономической эффективности внедрения;*

*- оценку научно-технического уровня выполненной НИР в сравнении с лучшими достижениями в данной области.*

**Список использованных источников**

**Литература к разделу 1**



**Литература к разделу 2**

1. P. Martin, P. Bateson. Measuring Behaviour: An Introductory Guide. –– Cambridge University Press, second edition, 1993.
2. M.S. Magnusson. Discovering hidden time patterns in behavior: T-patterns and their detection. –– Behavior Research Methods, Instruments, Computers 2000, 32 (I), 93–I IO.
3. 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.
4. R. Stoop, B. Arthur. Periodic orbit analysis demonstrates genetic constraints, variability, and switching in Drosophila courtship behavior. // Chaos – 2008 –vol.18/2.
5. P. Cvitanovic. Periodic orbits as the skeleton of classical and quantum chaos. // Physica D: Nonlinear Phenomena – 1991 – vol.51 – Issues 1-3 – Pp. 138–151.
6. А.М. Шурыгин. Математические методы прогнозирования. –– Горячая Линия – Телеком, 2009. ISBN 978-5-9912-0062-2.
7. N. Raichma, E. Ben-Jacob. Identifying repeating motifs in the activation of synchronized bursts in cultured neuronal networks. // Journal of Neuroscience Methods – 2007 – vol.170 – Pp. 96–-110.
8. J.-M. Fellous, P.H.E. Tiesinga, P.J. Thomas, T.J. Sejnowski. Discovering spike patterns in neuronal responses. // The Journal of Neuroscience – March 24, 2004 – vol.24(12) – Pp. 2989-–3001.
9. Sheng Ma, J.L. Hellerstein. Mining partially periodic event patterns with Uuknown periods. –– IBM T.J. Watson Research Center Hawthorne, NY 10532
10. В.В. Вишневский, Д.П. Ветров. Поиск скрытых поведенческих паттернов. –– Курсовая работа. ВМиК МГУ 2009.
11. M. Harris. Optimizing parallel reduction in CUDA. –– [PDF] – (http://developer.download.nvidia.com/compute/cuda/1\_1/Website/projects/reduction/doc/reduction.pdfreduction.pdf).
12. Б.А. Боресков, А.В. Харламов Основы работы с технологией CUDA. –– ДМК Пресс, 2010 – ISBN 978-5-94074-578-5.
13. M.S. Magnusson. Analyzing complex real-time streams of behavior: repeated patterns in behavior and DNA. – vol.3 – Ethologie humaine. Levallois-Perret – France, 2003 – ISBN 2-7237-0025-9.
14. G.K. Jonsson, S.H. Bjarkadottir, B. Gislason. Detection of real-time patterns in sports: interactions in football. –– L'?thologie appliqu?e aujourd'hui. – vol.3 – Ethologie humaine. Levallois-Perret – France, 2003 – ISBN 2-7237-0025-9.
15. M.F. Bear, B.W. Connors, M.A. Paradiso. Neuroscience: Exploring the Brain. –– Lippincott Williams & Wilkins – 3d editin, 2006.
16. A.L. Vyssotski, G. Dell'Omo, I.I. Poletaeva, D.L. Vyssotski. Long-term monitoring of hippocampus-dependent behavior in naturalistic settings: m utant mice lacking neurotrophin receptor TrkB in the forebrain show spatial learning but impaired behavioral flexibility. // Hippocampus, 12 – Pp. 27-–38 doi: 10.1002/hipo.10002.
17. К.В. Воронцов. Лекции по логическим алгоритмам классификации. –– 2010.

**Литература к разделу 3**



**Литература к разделу 4**



**Данные по исполнителям**

|  |  |  |
| --- | --- | --- |
| № п/п | ФИО | **Список публикаций** |
| 1. | Ветров Дмитрий Петрович |  |
| 2. | Конушин Антон Сергеевич |  |
| 3. | Сергиевский Георгий Максимович |  |
| 4. | Ушаков Вадим Леонидович |  |
| 5 | Зарайская Ирина Юрьевна |  |
| 6 | Долгов Олег Николаевич |  |
| 7. | Суконкин Илья Николаевич |  |
| 8. | Бесчастный Леонид Леонидович |  |
| 9. | Эйдлин Александр Андреевич |  |
| 10. | Синдеев Михаил Сергеевич |  |
| 11. | Темлянцев Александр Валерьевич |  |
| 12. | Вишневский Валерий Викторович |  |

1. Чаще всего понимается, что в этот момент времени имеет место  *начало* действия [↑](#footnote-ref-2)
2. [↑](#footnote-ref-3)