Правительство Российской Федерации
Федеральное государственное бюджетное образовательное учреждение
высшего профессионального образования
«Санкт-Петербургский государственный университет»

Кафедра информационно-аналитических систем

Дубатовка Алина Дмитриевна

Автоматическое выделение фаз ментальной активности

Бакалаврская работа

Допущена к защите.

Зав. кафедрой:

д.ф.-м.н., профессор Новиков Б.А.

Научный руководитель:

д.ф.-м.н., профессор Новиков Б.А.

Рецензент:

к.ф.-м.н., доцент Графеева Н.Г.

Санкт-Петербург

2015

SAINT-PETERSBURG STATE UNIVERSITY

Chair of Analytical Information Systems

Alina Dubatovka

Automatic extraction of mental activity phases

Bachelor's Thesis

Admitted for defence.
Head of the chair:
professor Boris Novikov

Scientific advisor:
professor Boris Novikov

Reviewer:
associate professor Natalia Grafeeva

Saint-Petersburg

2015

Оглавление

Введение 5
Глава 1. Постановка Задачи 6
Глава 2. Обзор предметной области 9
Глава 3. Описание используемых методов 10
3.1. Первичное выделение интервалов 10
3.1.1. Без разбиения 10
3.1.2. С помощью равномерного разбиения 10
3.1.3. С помощью поиска локальных экстремумов (экстремумы) 10
3.1.4. С помощью алгоритма change point detection (cpts) 10
3.2. Автоматическое определение порогов 11
3.2.1. Алгоритмы кластеризации 11
3.2.2. Анализ гистограмм 11
3.2.3. Трёхстолбчатая гистограмма 12
3.2.4. Определение минимальной длины рассматриваемых интервалов 13
3.2.5. Определение минимального порога коэффициента линейной регрессии 13
3.3. Постобработка интервалов 14
3.3.1. Склеивание интервалов на основе коэффициента линейной регрессии (склеивание) 14
3.3.2. Анализ коэффициентов линейной регрессии на префиксах (префиксы) 15
3.3.3. Анализ коэффициентов линейной регрессии на суффиксах (суффиксы) 15
3.3.4. Локальный минимум коэффициента линейной регрессии, посчитанного на префиксах (локальный минимум на префиксе) 15
Глава 4. Эксперименты 17
4.1. Эксперимент №1 18
4.2. Эксперимент № 2 19
4.3. Эксперимент №3 20
Глава 5. Графическое представление результатов 21
Глава 6. Результаты 22
6.3. Без выделения + префиксы + локальный минимум на префиксе. 22

6.2. Экстремумы + префиксы + суффиксы 25
6.3. Экстремумы + префиксы + локальный минимум на префиксе: 28
Заключение 32
Список литературы 33

Введение

Потоки данных (например, потоки данных с сенсоров) являются одним из источников слабоструктурированных и неструктурированными данных. Результатом анализа потоковых данных может являться выделение высокоуровневых фактов из исходных данных низкого уровня. Такое отображение низкоуровневых данных в высокоуровневые факторы может использоваться для упрощения интерпретации набора данных и выявления определённых характеристик, которыми потенциально могут обладать данные. Примером может служить выделение в данных фаз с различными заранее заданными свойствами.

Среди затрудняющих факторов анализ реальных потоковых данных можно выделить большие объёмами, непрерывные изменения и необходимость онлайн-обработки, а также наличие неточностей в измерениях: шума и пропусков данных, что может мешать проведению анализа. Для корректной обработки данных эти погрешности должны быть нивелированы: устранены пропуски в данных, выявлены аномальные отклонения, удалены шумы.

В этой работе была поставлена задача анализа измерений частоты сердечных сокращений во время различных умственных нагрузок с целью выделения фаз ментальной активности. В данной работе предложена обобщённая схема алгоритма выделения фаз ментальной активности, описана реализация методов для выделения фаз. Была проведена экспертная оценка работы описанных алгоритмов на реальных данных, в работе представлен сравнительный анализ их применимости для решения данной задачи.

Глава 1. Постановка Задачи

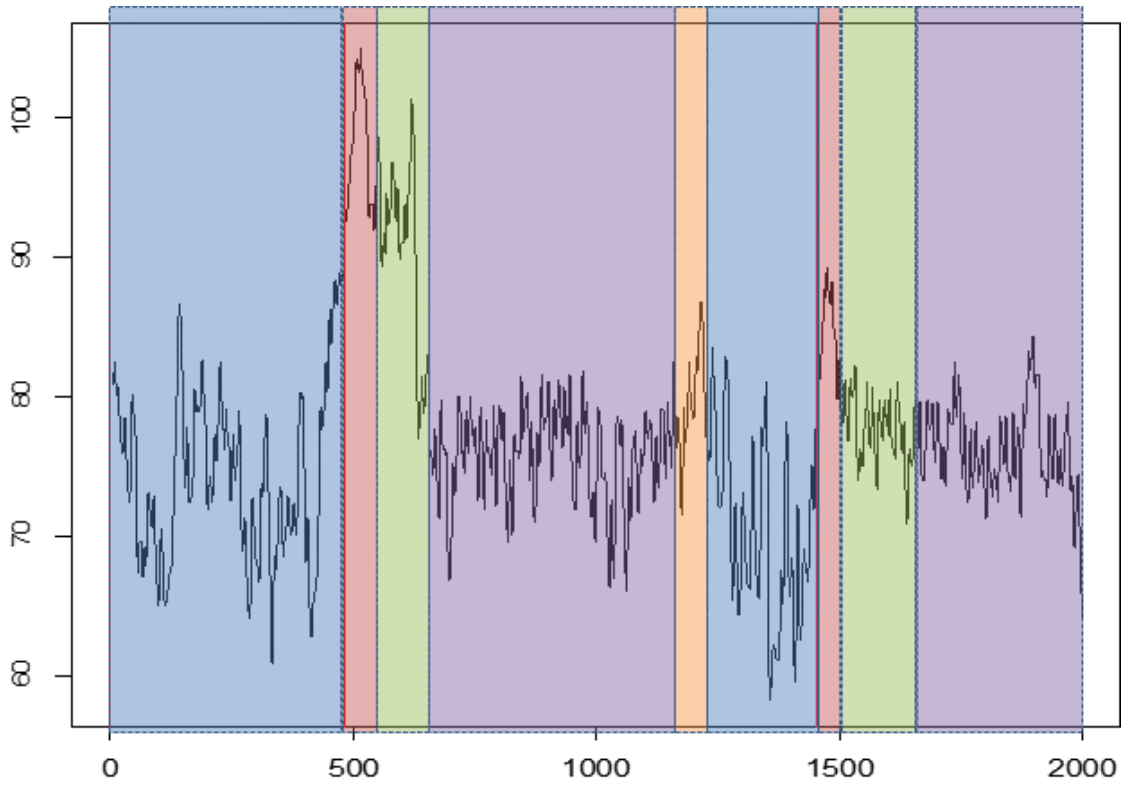
Данные содержат запись ЧСС испытуемых в некоторый промежуток времени, в течение которого они поочерёдно выполняли задания и отдыхали. Таким образом, процесс выполнения заданий может быть разделён на 2 этапа: выполнение задания и покой.

Известно, что в ходе выполнения работы и последующего отдыха человек проходит через следующие фазы:

- Мобилизация (красный)
- Стабилизация (зеленый)
- Работа (фиолетовый)
- Восстановление (оранжевый)
- Отдых (синий)

Важно отметить, что фазы мобилизации, стабилизации и работы составляют этап выполнения задания, а фазы восстановления и отдыха – этап покоя.

На картинке приведен график ЧСС одного из испытуемых, на котором обозначены интервалы времени, соответствующие приведённым выше фазам.

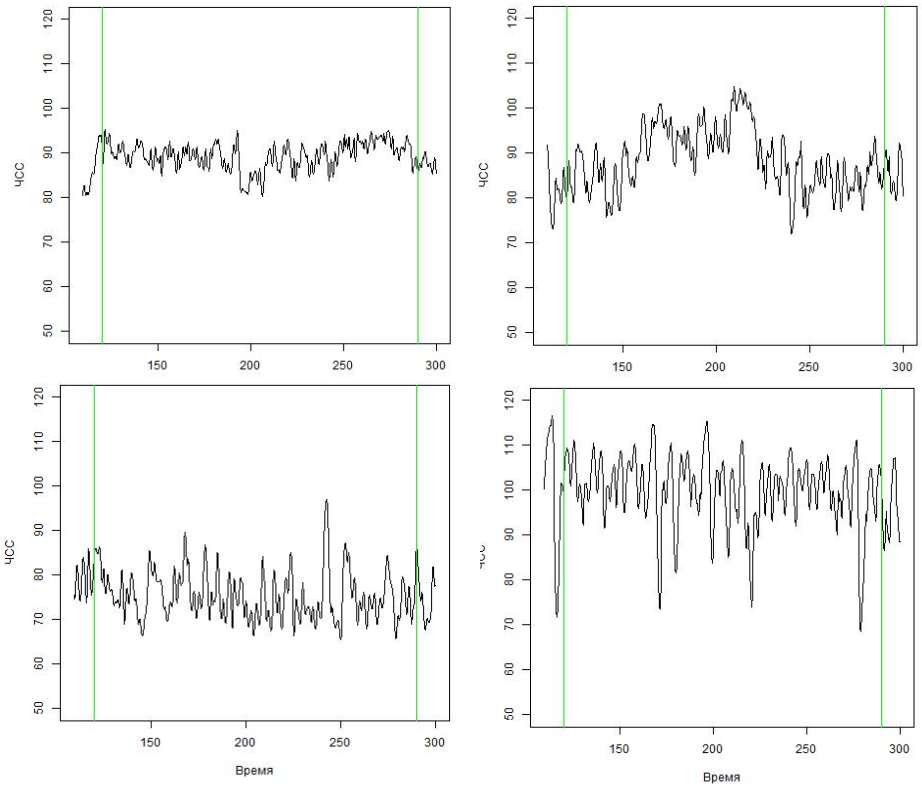


От специалистов в предметной области известно, что данные фазы имеют следующие нечётко определённые свойства:

- Отдых (синий)
    - Относительно низкая средняя ЧСС (по сравнению с другими фазами)
    - Относительно большая вариация ЧСС (по сравнению с другими фазами)
    - Заканчивается в момент получения задания (с возможной краткосрочной задержкой или опережением)
- Мобилизация (красный)
    - Стабильный подъем ЧСС во времени (заметный по сравнению с обычными колебаниями)
    - Начинается в момент получения задания (с возможной краткосрочной задержкой или опережением)
    - Среднее значение ЧСС выше, чем на фазе отдыха
    - Вариация ЧСС ниже, чем на фазе отдыха
- Стабилизация (зеленый)
    - Снижение ЧСС во времени
    - Среднее значение ЧСС выше, чем на фазе отдыха
    - Вариация ЧСС ниже, чем на фазе отдыха
- Работа (фиолетовый)
    - Среднее значение ЧСС выше, чем на фазе отдыха
    - Вариация ЧСС ниже, чем на фазе отдыха
    - Заканчивается в момент прекращения работы над заданием (с возможной краткосрочной задержкой)
- Восстановление (оранжевый)
    - Снижение ЧСС во времени
    - Среднее значение ЧСС выше, чем на фазе отдыха
    - Вариация ЧСС ниже, чем на фазе отдыха

- Начинается в момент прекращения работы над заданием (с возможной краткосрочной задержкой)

Первые эксперименты по автоматическому выделению фаз были проведены на этапах выполнения заданий, поскольку фазы мобилизации, стабилизации и работы являются более значимыми с точки зрения практического использования. На рисунке ниже представлены примеры поведения ЧСС в ходе выполнения задания разными людьми (прямые проведены в момент начала и конца задания).



Задача выделения фаз затрудняется индивидуальностью поведения ЧСС как для каждого человека, так и при разных видах умственной активности. Это приводит к малому количеству данных для анализа, а также необходимости самонастройки алгоритмов для выделения фаз в режиме онлайн без предварительного обучения.

Глава 2. Обзор предметной области

Задача выделения фаз ментальной активности относится к классу задач изучения активности мозга, большинство из которых исследуются на основе анализа данных ЭКГ, ЭЭГ, ВЧСС [1]. Как и к любым потоковым данным, к обработке вышеописанных данных существует несколько различных подходов: вейвлет-преобразования, методы машинного обучения, поиск точек смены режима. Выбор того или иного подхода определяется природой обрабатываемых данных и особенностями задачи.

Так, например, обработка данных ЭКГ и ЭЭГ обычно связана с вейвлет-преобразованием [2, 3]. Однако подобные преобразования не могут быть применены к ЧСС из-за быстро осциллирующих данных, а также из-за ошибок и провалов в данных, к которым очень чувствительно преобразование Фурье.

Алгоритмы машинного обучения, являющиеся распространённым подходом к анализу данных, могут быть применены и для изучения ментальных процессов, например, для классификации активности мозга [2, 4]. Однако все используемые методы являются примерами обучения с учителем, а значит, требуют большого количества размеченных заранее данных, что не позволяет применять их для индивидуальной обработки ЧСС человека, поскольку характер поведения ЧСС различается как между людьми, так и в зависимости от типа выполняемого задания.

Помимо вышеперечисленных, алгоритмы поиска точек смены режима (change point detection) [5] могли бы быть полезны в задаче выделения фаз (поскольку поведение данных на границе двух фаз меняется). Однако они рассчитывают на более или менее плавное поведение тренда, что неверно для рассматриваемых данных.

Глава 3. Описание используемых методов

Процесс поиска границ фаз состоит из 3 основных этапов: первичного выделения интервалов, автоматического определения порогов и постобработки первичных интервалов. Преимуществом такого подхода является его возможность адаптироваться к данным конкретного человека без предварительного обучения на других людях, что позволяет учесть психофизиологические особенности различных людей.

3.1. Первичное выделение интервалов

Из всех исходных точек выбираются наиболее подходящие на роль границ фаз, то есть весь исходный временной отрезок разбивается на первичные отрезки, которые и будут обрабатываться далее.

3.1.1. Без разбиения

Первоначальное выделение интервалов не производится, все точки в равной степени претендуют на роль границ фаз и рассматриваются при постобработке.

3.1.2. С помощью равномерного разбиения

Исходный отрезок разбивается на достаточно большое количество интервалов равной длины – первичные интервалы.

3.1.3. С помощью поиска локальных экстремумов (экстремумы)

В данном случае под локальным экстремумом понимается экстремум по симметричному «окну» некоторой длины. В описываемых экспериментах использовалось «окно» длины 21 (10 точек слева и 10 – справа). Полученные точки локальных экстремумов впоследствии рассматривались как первоначальные концы интервалов.

3.1.4. С помощью алгоритма change point detection (cpts)

Данные усредняются по окну длины 10, затем считаются частичные разности полученных «сглаженных» данных (разность значений, стоящих на расстоянии 2 друг от друга), далее к полученным значениям применяется

алгоритм change point detection, чувствительный к изменению среднего значения и вариации, который находит несколько (не более 50) точек смены поведения частичных разностей. Полученные точки и принимаются за первичные концы интервалов.

3.2. Автоматическое определение порогов

После первичного выделения интервалов необходимо определить различные пороговые значения для дальнейшей обработки. Так, например, требуется определить минимальную длину рассматриваемых на следующем этапе интервалов, поскольку первоначально мы могли выделить очень короткие интервалы из-за того, что данные на них «скачут» или просто не соответствуют общей тенденции остальных данных, то есть эти интервалы можно считать выбросами. Кроме того, нужно понять, какие значения коэффициента линейной регрессии соответствуют возрастанию (а значит, мобилизации), а какие – убыванию, т.е. стабилизации. Для этого нужны нижний и верхний (иногда – только верхний) пороги коэффициента линейной регрессии фазы работы, поскольку она характеризуется близким к нулю коэффициентом линейной регрессии. Определение пороговых значений коэффициента линейной регрессии

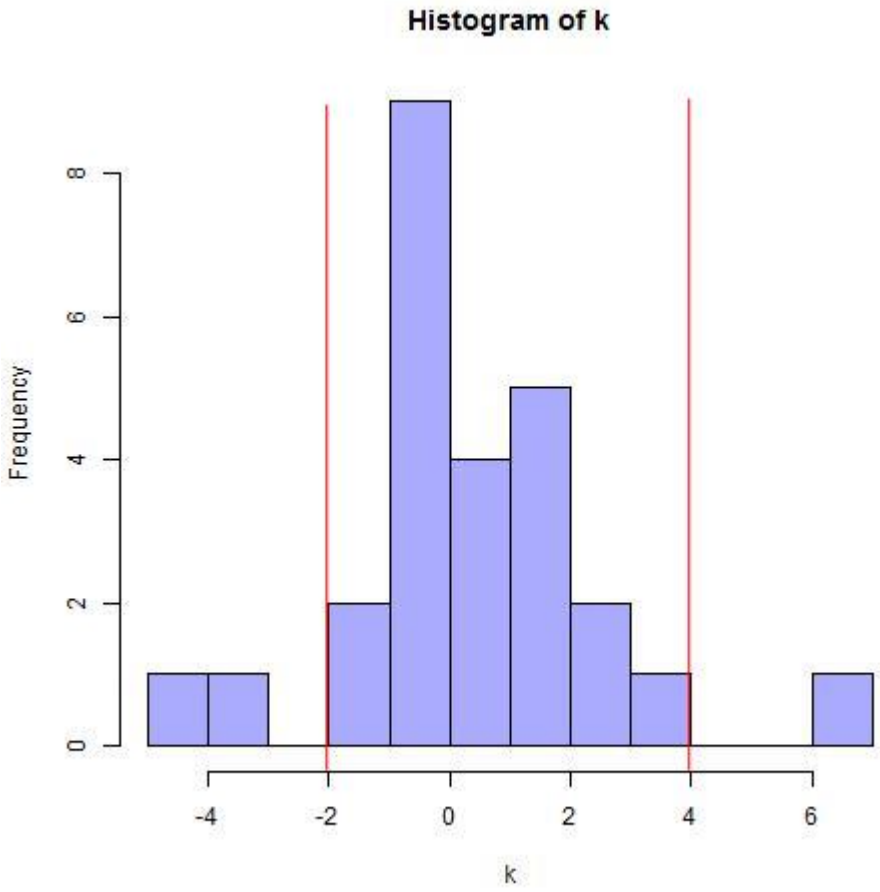
3.2.1. Алгоритмы кластеризации

Нижний и верхний порог коэффициента линейной регрессии вычисляются с помощью кластеризации значений коэффициента на всех первичных префиксах (префиксах с концами в выделенных на первом этапе точках) на 3 кластера. Границы между полученными кластерами принимаются в качестве верхнего и нижнего порогов.

3.2.2. Анализ гистограмм

Строится гистограмма значений коэффициента линейной регрессии на всех первичных префиксах. Затем на ней находится максимум, поскольку самое распространенное значение, как правило, близко к нулю и соответствует фазе работы – самой продолжительной фазе. После этого все промежутки,

соседние с найденным и имеющие ненулевую высоту, объединяются в один промежуток, концы которого считаются нижним и верхним порогами коэффициента линейной регрессии фазы работы соответственно. На приведённой ниже гистограмме значений коэффициента регрессии для случая выделения первичных интервалов алгоритмом change point detection красными прямыми обозначены нижняя и верхняя границы фазы работы.

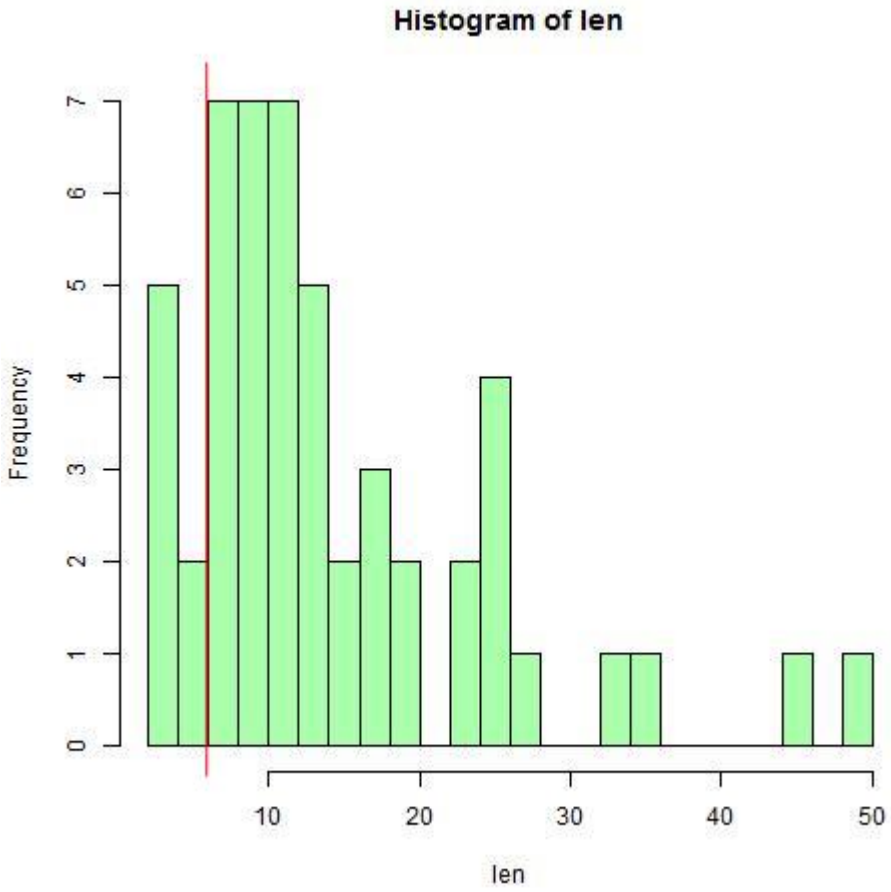


3.2.3. Трёхстолбчатая гистограмма

На основе значений коэффициента линейной регрессии на всех первичных интервалах строится гистограмма, состоящая ровно из 3 столбцов. Считается, что значения в левом столбце соответствуют фазе стабилизации, в правом – мобилизации, поэтому в качестве нижнего и верхнего порогов коэффициента линейной регрессии рассматриваются левая и правая границы среднего столбца соответственно.

3.2.4. Определение минимальной длины рассматриваемых интервалов

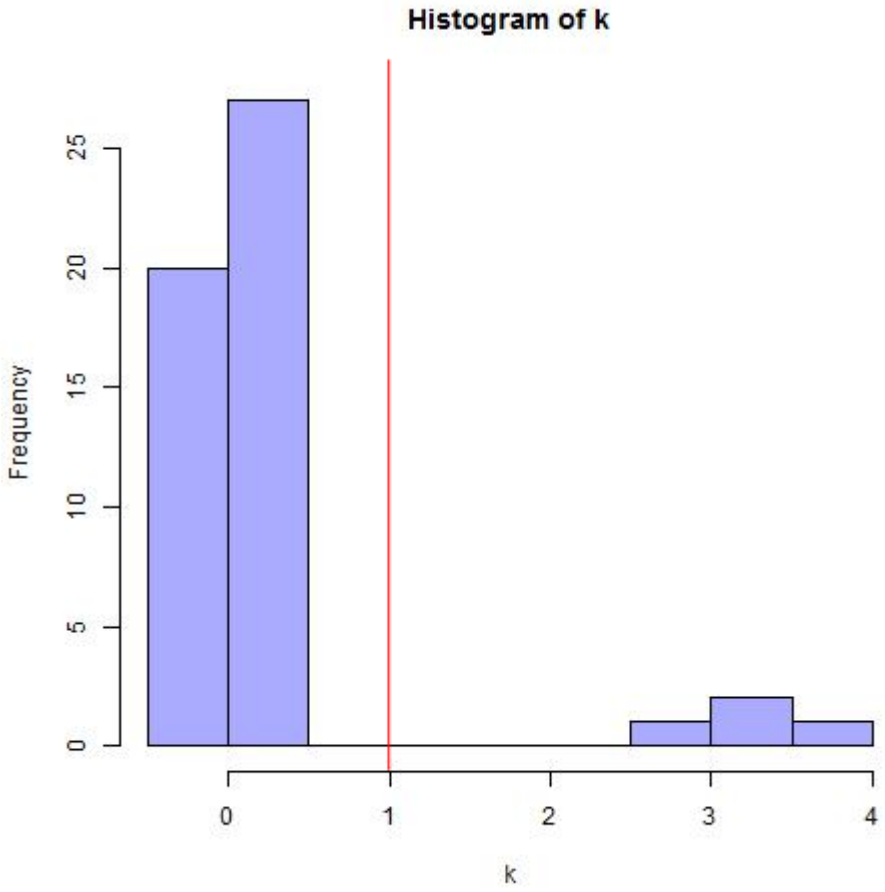
В качестве нижней границы длины рассматриваемых интервалов берётся самый левый локальный минимум гистограммы (его правая граница) длин отрезков, полученных на фазе выделения первичных интервалов. Ниже приведена описанная гистограмма для первичных интервалов, выделенных алгоритмом change point detection, с проведённой нижней границей длины интервалов для дальнейшей обработки.



3.2.5. Определение минимального порога коэффициента линейной регрессии

Строится гистограмма коэффициентов линейной регрессии на первичных префиксах. Затем на ней находится точка максимума, как правило, это небольшое положительное число, соответствующее «длинным» префиксам с концами во второй половине задания. После этого справа от точки максимума вычисляется точка минимума гистограммы, правая граница

которой и принимается в качестве порогового значения. Ниже приведён пример гистограммы с проведённой границей коэффициента линейной регрессии, соответствующей фазе мобилизации, построенной для первичных интервалов, выделенных алгоритмом change point detection.



3.3. Постобработка интервалов

На данном этапе первичные интервалы, полученные на первом этапе, обрабатываются с применением порогов, вычисленных автоматически на втором этапе, для окончательного выделения границ фаз.

3.3.1. Склеивание интервалов на основе коэффициента линейной регрессии (склеивание)

На полученных первичных отрезках высчитывается коэффициент линейной регрессии. Далее все интервалы, превышающие по длине порог, найденный на предыдущем этапе, раскрашиваются в зависимости от величины коэффициента регрессии в 3 цвета: красный (коэффициент не превосходит

нижнего порога), зелёный (превышает верхний) и синий. Используемые пороги также получены автоматически на предыдущем этапе. После этого соседние интервалы, покрашенные в одинаковый цвет, объединяются в один, коэффициент линейной регрессии пересчитывается, что может привести к изменению цвета полученного интервала. Этот процесс завершается, когда интервалы перестают склеиваться.

3.3.2. Анализ коэффициентов линейной регрессии на префиксах (префиксы)

На первичных префиксах вычисляется коэффициент линейной регрессии. Предполагаемый разрез мобилизация-стабилизация проведён на правой границе префикса, коэффициент линейной регрессии на котором больше порогового значения, определяемого автоматически на предыдущем этапе, и значение ЧСС на правом конце которого максимально.

3.3.3. Анализ коэффициентов линейной регрессии на суффиксах (суффиксы)

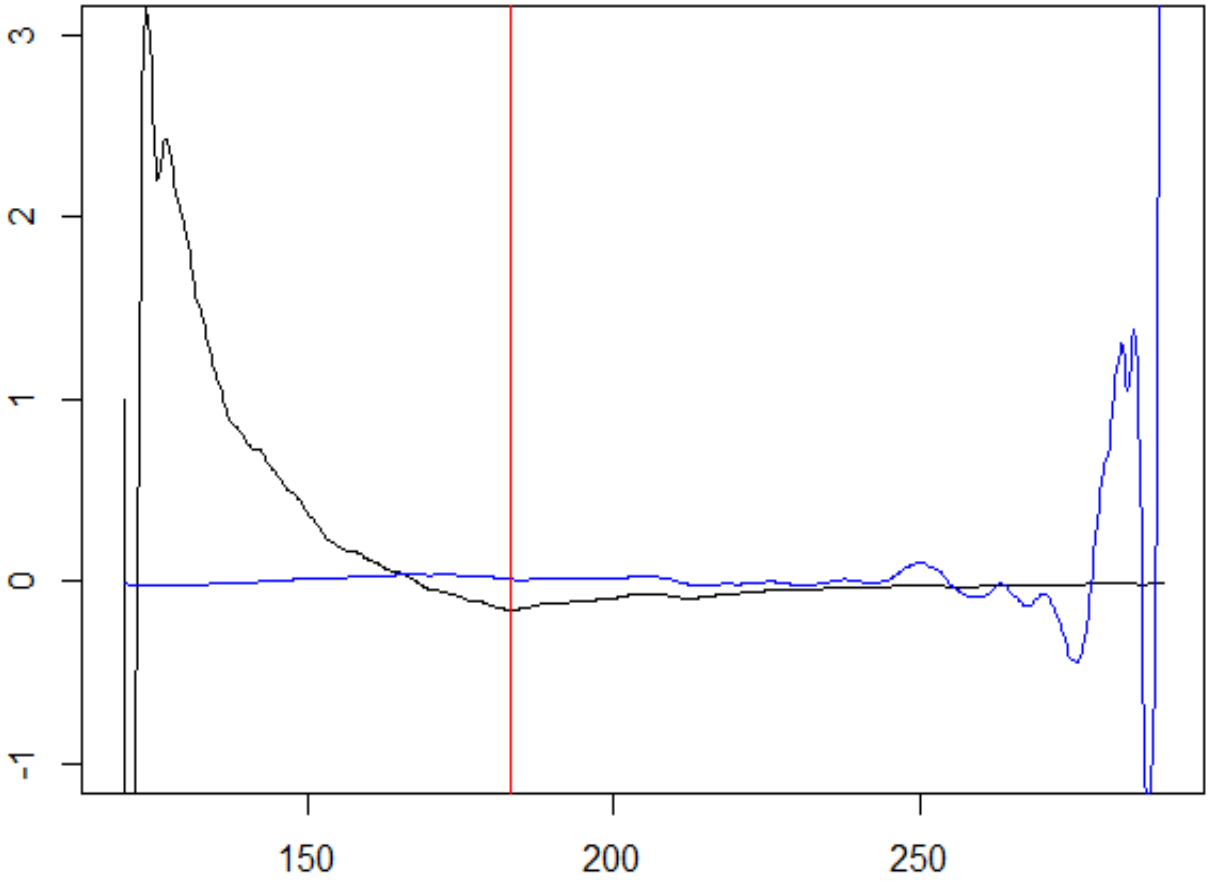
На суффиксах с левыми концами в точках, полученных на этапе первичного выделения интервалов, вычисляется коэффициент линейной регрессии. Предполагаемый разрез стабилизация-работа проводится на левой границе суффикса, коэффициент линейной регрессии на котором минимален по модулю.

3.3.4. Локальный минимум коэффициента линейной регрессии, посчитанного на префиксах (локальный минимум на префиксе)

На первичных префиксах вычисляется коэффициент линейной регрессии. Граница фаз стабилизации и работы проводится в точке правого конца префикса, значение коэффициента на котором является локальным минимумом среди коэффициентов линейной регрессии на полученных префиксах. В случае, когда локальных минимумов несколько, выбирается самый левый из них, попадающих в интервал выполнения задания

(поскольку минимум до начала работы, скорее всего, предшествует росту ЧСС на фазе мобилизации и относится к фазе отдыха).

Ниже в качестве иллюстрации приведён график значения коэффициента линейной регрессии на префиксах (чёрная кривая) и суффиксах (синяя кривая) с концами в каждой точке. Красная прямая проведена в точке локального минимума коэффициента регрессии на префиксе, что соответствует результату работы метода «Локальный минимум на префиксе».



Глава 4. Эксперименты

Исходными данными для работы являются показания частоты сердечных сокращений (ЧСС) 63 испытуемых, каждый из которых выполнял по 5 однотипных заданий. Показания снимались 4 раза в секунду, а всё испытание длилось 1375 секунд. Испытуемые были разделены на три группы:

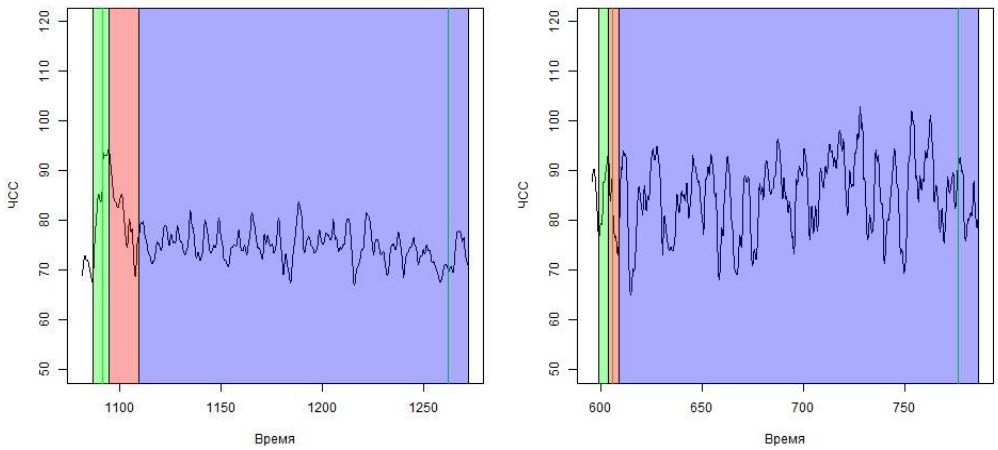
- E0 – 15 «экспертов», специально подготовленных и имеющих большой опыт в выполнении заданий;
- U0 – 25 «новичков», выполняющих задания впервые;
- U1 – 23 «новичка после тренировки», то есть немного потренировавшиеся, однако, не готовившиеся специально.

Процесс выполнения заданий может быть разделён на 2 этапа: выполнение задания и покой. Границы данных этапов известны, поскольку известны моменты получения и прекращения заданий, представленные в таблице временных рамок нагрузки.

Начало	**Окончание**	**Тип**
0	120	Покой
120	290	Нагрузка тип 1
290	363	Покой
363	533,5	Нагрузка тип 2
533,5	606	Покой
606	776,5	Нагрузка тип 3
776,5	849	Покой
849	1020	Нагрузка тип 4
1020	1092	Покой
1092	1262,5	Нагрузка тип 5
1262,5	1375	Покой

4.1. Эксперимент №1

Поскольку из-за вариативности на всех этапах предложенного подхода число всевозможных алгоритмов довольно велико, необходимо первичное тестирование на меньшем объёме данных для сокращения множества рассматриваемых методов при дальнейших экспериментах. В ходе данного эксперимента рассматривались всевозможные комбинации реализаций этапов вышеописанного алгоритма, которые применялись к данным первого эксперимента всех испытуемых из группы экспертов с целью выявления наиболее удачных подходов для дальнейшей работы. В качестве данных были выбраны именно показания экспертов, так как у этой группы наиболее чётко выражены фазы активности в силу того, что они являются самыми тренированными из всех испытуемых и меньше всех подвержены стрессу. Это позволяет оценивать результаты работы алгоритмов самостоятельно – исходные фазы, как правило, видны непрофессионалу. Поэтому можно достаточно визуального анализа графика с выделенными алгоритмом фазами, чтобы оценить его работу, без привлечения дополнительных экспертных оценок. Ниже для сравнения приведены 2 графика с результатами работы одного из приведённых ниже алгоритмов. Левый является записью ЧСС эксперта, а правый – новичка, на котором фазы не смог выделить даже профессионал.



4.2. Эксперимент № 2

В ходе описанного эксперимента для дальнейшей работы были выбраны следующие алгоритмы, краткое описание которых приводится ниже.

4.2.1. Без выделения + префиксы + локальный минимум на префиксе: первичного выделения интервалов не производится, граница мобилизация-стабилизация определяется методом префиксов (пороговое значение коэффициента линейной регрессии определяется автоматически), а граница стабилизация-работа находится методом поиска локального минимума коэффициента регрессии на префиксах;

4.2.2. Экстремумы + префиксы + суффиксы: первичные концы находятся в точках локальных экстремумов, граница мобилизация-стабилизация определяется методом префиксов (пороговое значение коэффициента линейной регрессии определяется автоматически), а граница стабилизация-работа – методом суффиксов;

4.2.3. Экстремумы + префиксы + локальный минимум на префиксе: первичные концы находятся в точках локальных экстремумов, граница мобилизация-стабилизация определяется методом префиксов (пороговое значение коэффициента линейной регрессии определяется автоматически), а граница стабилизация-работа находится методом поиска локального минимума коэффициента регрессии на префиксах;

4.2.4. Cpts + склеивание интервалов: первичные интервалы выделяются с помощью алгоритма change points, далее полученные интервалы склеиваются, все необходимые пороги (минимальная длина интервала и границы коэффициента регрессии) определяются автоматически.

В описываемом эксперименте данные алгоритмы были применена ко всем имеющимся данным: по пять заданий для испытуемых каждой из трёх групп – для более детального анализа методов с применением экспертных оценок. В результате первый алгоритм (Без выделения + префиксы + локальный минимум на префиксе) показал, по мнению экспертов, наилучшие

результаты, в то время как четвёртый (Cpts + склеивание интервалов) не справился совсем.

4.3. Эксперимент №3

По результатам экспертных оценок было решено оставить для рассмотрения три первых алгоритма, описанных в предыдущем пункте. Для этих подходов ставились эксперименты с различными способами пересчёта порогов между заданиями для каждого испытуемого. Поскольку пороги во многом зависят от физиологических особенностей человека, можно предположить, что они будут не сильно меняться от задания к заданию. Это позволит использовать данные о предыдущих заданиях для более точной оценки последующих.

В ходе данного эксперимента были рассмотрены следующие способы определения необходимых пороговых значений для каждого испытуемого.

4.3.1. Пересчёт порогового значения при каждом новом эксперименте: все пороги на каждом задании вычисляются заново, не используя информации о результатах предыдущих заданий;

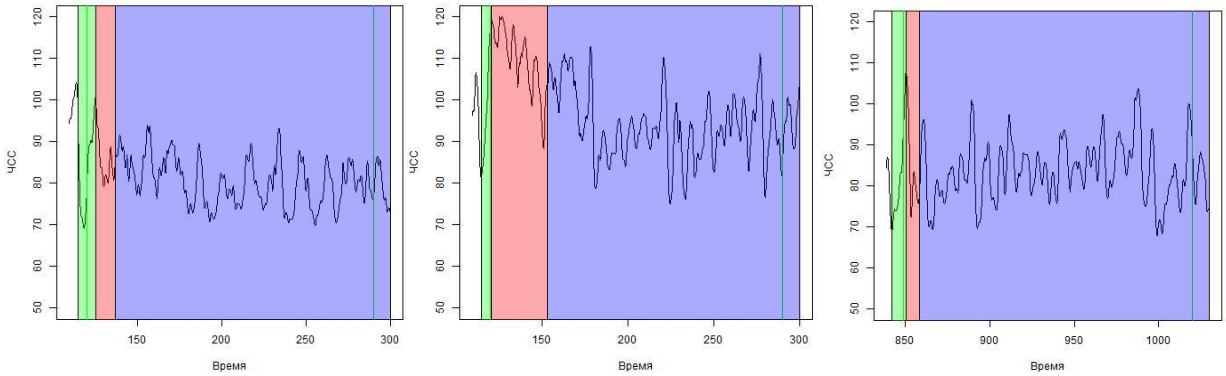
4.3.2. Использование порогового значения, определённого в ходе первого задания, для разметки всех заданий: значения необходимых порогов вычисляются один раз по результатам обработки данных первого задания для каждого испытуемого. Далее эти пороги используются при выделении фаз всех заданий этого испытуемого.

4.3.3. Использование для разметки задания порога, полученного в ходе обработки предыдущего задания: для выделения фаз данного задания использовались пороги, полученные по результатам обработки предыдущего задания рассматриваемого испытуемого (исключение составляет первое задание – для него порог получается в ходе его автоматической обработки).

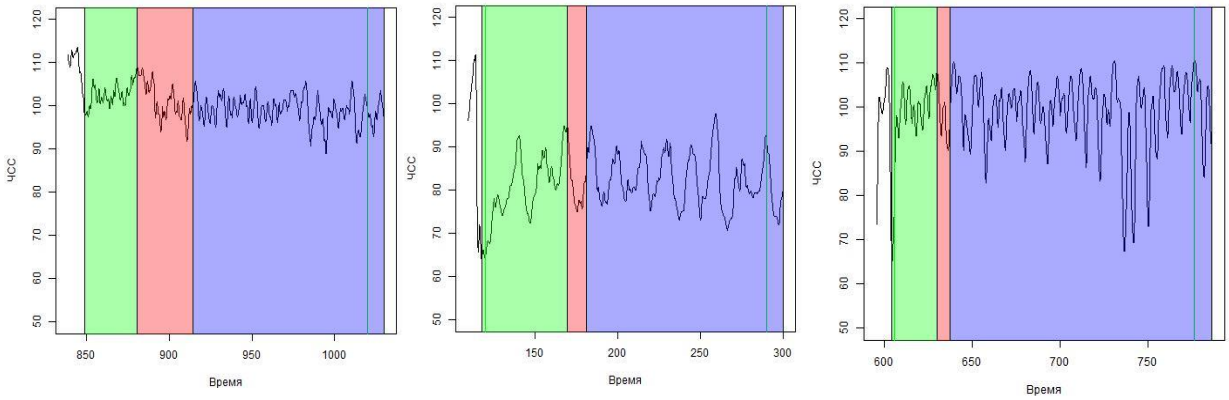
Глава 5. Графическое представление результатов

Ниже представлены графические примеры работы алгоритма «Без выделения + префиксы + локальный минимум на префиксе» – лучшего из приведённых алгоритмов. Графики разделены на 3 группы: примеры, где фазы удалось выделить правильно, примеры, где фазы выделены неправильно и примеры, где эксперты не смогли выделить фазы.

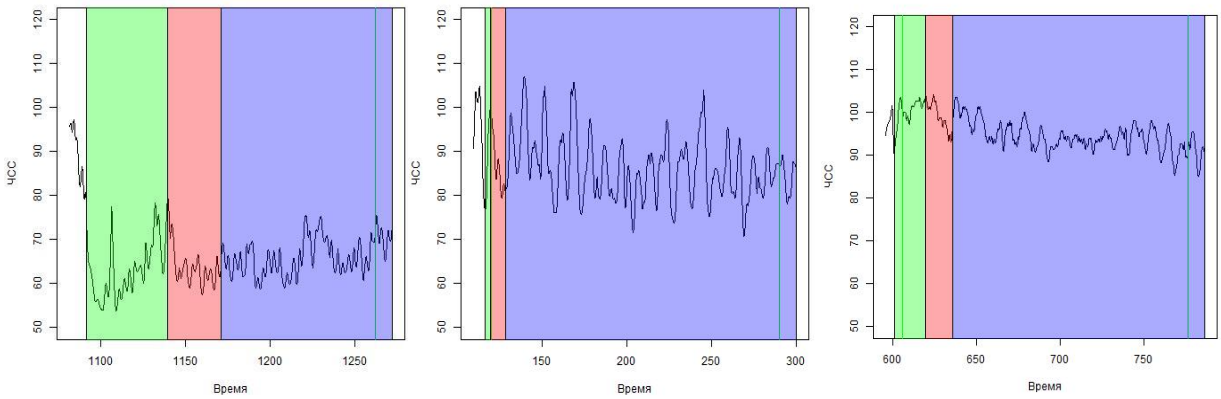
Правильно выделенные фазы



Неправильно выделенные фазы



Невыделенные фазы



Глава 6. Результаты

В экспериментах рассматривались данные ЧСС 63 испытуемых, каждый из которых выполнял по 5 различных заданий. Испытуемые были разделены на три группы:

- E0 – 15 «экспертов», специально подготовленные и имеющие большой опыт в выполнении заданий
- U0 – 25 «новичков», выполняющих задания впервые
- U1 – 23 «новичка после тренировки», то есть немного потренировавшиеся, однако, не готовившиеся специально

Ниже представлены результаты работы описанных алгоритмах как в целом, так и на данных всех трёх групп испытуемых, на каждом из пяти заданий, а также для каждого отдельного испытуемого.

6.3. Без выделения + префиксы + локальный минимум на префиксе

	Доля правильно выделенных фаз среди всех экспериментов	Доля фаз, выделить которые мануально не удалось, среди всех экспериментов	Доля правильно выделенных фаз среди экспериментов с экспертной оценкой
E0	0.573	0.147	0.672
U0	0.544	0.136	0.63
U1	0.583	0.322	0.859
Total	0.565	0.206	0.712

	Доля правильно выделенных фаз среди всех экспериментов	Доля фаз, выделить которые мануально не удалось, среди всех экспериментов	Доля правильно выделенных фаз среди экспериментов с экспертной оценкой
Задание 1			
E0	0.6	0.067	0.643
U0	0.52	0.2	0.65
U1	0.478	0.304	0.688
Total	0.524	0.206	0.66
Задание 2			
E0	0.6	0.133	0.538
U0	0.52	0.12	0.591

U1	0.478	0.478	0.916
Total	0.492	0.254	0.66
Задание 3			
E0	0.533	0.133	0.615
U0	0.6	0.08	0.652
U1	0.696	0.217	0.889
Total	0.619	0.143	0.722
Задание 4			
E0	0.333	0.4	0.556
U0	0.6	0.08	0.652
U1	0.609	0.348	0.933
Total	0.54	0.254	0.723
Задание 5			
E0	0.933	0	0.933
U0	0.48	0.2	0.6
U1	0.652	0.261	0.882
Total	0.65	0.175	0.788

	Доля правильно выделенных фаз среди всех экспериментов	Доля фаз, выделить которые мануально не удалось, среди всех экспериментов	Доля правильно выделенных фаз среди экспериментов с экспертной оценкой
E0_1	1	0	1
E0_2	0.6	0	0.6
E0_3	0.4	0.2	0.5
E0_4	0.6	0.2	0.75
E0_5	0.4	0.4	0.667
E0_6	0.6	0.2	0.75
E0_7	0.4	0.2	0.5
E0_8	0.8	0	0.8
E0_9	1.0	0	1
E0_10	0.4	0.4	0.667
E0_11	0.8	0	0.8
E0_12	0.6	0.2	0.75
E0_13	0.2	0.2	0.25
E0_14	0.4	0.2	0.5
E0_15	0.4	0	0.4

	Доля правильно	Доля фаз, выделить которые	Доля правильно выделенных

	выделенных фаз среди всех экспериментов	мануально не удалось, среди всех экспериментов	фаз среди экспериментов с экспертной оценкой
U0_1	0.6	0.2	0.75
U0_2	0.2	0.2	0.25
U0_3	0.4	0	0.4
U0_4	1	0	1
U0_5	1	0	1
U0_6	0.6	0	0.6
U0_7	0.6	0	0.6
U0_8	0.2	0.4	0.333
U0_9	0.4	0	0.4
U0_10	0.6	0	0.6
U0_11	0.6	0	0.6
U0_12	0.6	0.2	0.75
U0_13	0.6	0.2	0.75
U0_14	0.8	0	0.8
U0_15	0.6	0	0.6
U0_16	0.2	0.2	0.25
U0_17	0.8	0	0.8
U0_18	0.6	0	0.6
U0_19	0.8	0	0.8
U0_20	0.8	0	0.8
U0_21	0.2	0.2	0.25
U0_22	0.4	0.4	0.667
U0_23	0.4	0.6	1
U0_24	0.2	0.4	0.333
U0_25	0.4	0.4	0.667

	Доля правильно выделенных фаз среди всех экспериментов	Доля фаз, выделить которые мануально не удалось, среди всех экспериментов	Доля правильно выделенных фаз среди экспериментов с экспертной оценкой
U1_1	0.4	0.4	0.66
U1_2	1	0	1
U1_3	0.6	0	0.0
U1_4	1.0	0	1
U1_5	0.6	0.4	1
U1_6	1	0	1

U1_7	0.8	0.2	1
U1_8	0.4	0.4	0.667
U1_9	0.6	0.4	1
U1_10	0	0.8	0
U1_11	1	0	1
U1_12	0.6	0.2	0.75
U1_13	0.4	0.4	0.667
U1_14	0.2	0.8	1
U1_15	1	0	1
U1_16	0.4	0.4	0.667
U1_17	0.6	0.4	1
U1_18	0.4	0.4	0.667
U1_19	0.6	0.2	0.75
U1_20	0.4	0.6	1.00
U1_21	0	0.8	0
U1_22	0.6	0.4	1
U1_23	0.8	0.2	1

6.2. Экстремумы + префиксы + суффиксы

По оценкам экспертов, данный алгоритм работает хуже, чем описанный выше алгоритм. Сильные различия проявляются при выделении границы фаз стабилизации и работы, поскольку алгоритм суффиксов находит её значительно правее, т.е. позже, чем она должна быть. Это происходит из-за того, что коэффициент линейной регрессии на суффиксах из-за сильно осциллирующих данных зачастую достигает минимума по модулю слишком рано, в середине фазы работы, а не на реальной границе, как можно было бы предположить, исходя из свойств фаз, что приводит к низким результатам этого метода при выделении фазы стабилизации.

	Доля правильно выделенных фаз среди всех экспериментов	Доля фаз, выделить которые мануально не удалось, среди всех экспериментов	Доля правильно выделенных фаз среди экспериментов с экспертной оценкой
E0	0.08	0.147	0.094
U0	0.072	0.136	0.083
U1	0.07	0.322	0.103

Total	0.073	0.206	0.092

	Доля правильно выделенных фаз среди всех экспериментов	Доля фаз, выделить которые мануально не удалось, среди всех экспериментов	Доля правильно выделенных фаз среди экспериментов с экспертной оценкой
Задание 1			
E0	0.067	0.067	0.071
U0	0.04	0.2	0.05
U1	0.087	0.304	0.125
Total	0.063	0.206	0.08
Задание 2			
E0	0.067	0.133	0.077
U0	0.08	0.12	0.091
U1	0.043	0.478	0.083
Total	0.063	0.254	0.085
Задание 3			
E0	0.067	0.133	0.077
U0	0.16	0.08	0.174
U1	0.13	0.217	0.167
Total	0.127	0.143	0.148
Задание 4			
E0	0.133	0.4	0.222
U0	0.04	0.08	0.043
U1	0	0.348	0
Total	0.048	0.254	0.064
Задание 5			
E0	0.067	0	0.067
U0	0.04	0.2	0.05
U1	0.087	0.261	0.118
Total	0.063	0.175	0.077

	Доля правильно выделенных фаз среди всех экспериментов	Доля фаз, выделить которые мануально не удалось, среди всех экспериментов	Доля правильно выделенных фаз среди экспериментов с экспертной оценкой
E0_1	0	0	0
E0_2	0	0	0
E0_3	0.6	0.2	0.75
E0_4	0	0.2	0
E0_5	0	0.2	0

E0_6	0	0.2	0
E0_7	0	0.2	0
E0_8	0.2	0	0.2
E0_9	0.2	0	0.2
E0_10	0	0.4	0
E0_11	0	0	0
E0_12	0.2	0.2	0.25
E0_13	0	0.2	0
E0_14	0	0.2	0
E0_15	0	0	0

	Доля правильно выделенных фаз среди всех экспериментов	Доля фаз, выделить которые мануально не удалось, среди всех экспериментов	Доля правильно выделенных фаз среди экспериментов с экспертной оценкой
U0_1	0	0.2	0
U0_2	0	0.2	0
U0_3	0	0	0
U0_4	0.2	0	0.2
U0_5	0.2	0	0.2
U0_6	0.2	0	0.2
U0_7	0	0	0
U0_8	0.2	0.4	0.333
U0_9	0.2	0	0.2
U0_10	0	0	0
U0_11	0.2	0	0.2
U0_12	0	0.2	0
U0_13	0	0.2	0
U0_14	0.4	0	0.4
U0_15	0	0	0
U0_16	0	0.2	0
U0_17	0	0	0
U0_18	0	0	0
U0_19	0	0	0
U0_20	0	0	0
U0_21	0	0.2	0
U0_22	0	0.4	0
U0_23	0	0.6	0

U0_24	0	0.4	0
U0_25	0.2	0.4	0.333

	Доля правильно выделенных фаз среди всех экспериментов	Доля фаз, выделить которые мануально не удалось, среди всех экспериментов	Доля правильно выделенных фаз среди экспериментов с экспертной оценкой
U1_1	0	0.4	0
U1_2	0	0	0
U1_3	0	0	0
U1_4	0.2	0	0.2
U1_5	0	0.4	0
U1_6	0.2	0	0.2
U1_7	0.2	0.2	0.25
U1_8	0	0.4	0
U1_9	0	0.4	0
U1_10	0.2	0.8	1
U1_11	0	0	0
U1_12	0	0.2	0
U1_13	0.2	0.4	0.333
U1_14	0	0.8	0
U1_15	0.2	0	0.2
U1_16	0	0.4	0
U1_17	0.2	0.4	0.333
U1_18	0.2	0.4	0.333
U1_19	0	0.2	0
U1_20	0	0.6	0
U1_21	0	0.8	0
U1_22	0	0.4	0
U1_23	0	0.2	0

6.3. Экстремумы + префиксы + локальный минимум на префиксе

Данный метод отличается от первого тем, что в качестве первичных интервалов рассматриваются точки локального экстремума, что и приводит к снижению точности – не всегда границы фаз проходят в точках локальных

экстремумов и не всегда экстремумы значения ЧСС и коэффициента линейной регрессии совпадает.

	Доля правильно выделенных фаз среди всех экспериментов	Доля фаз, выделить которые мануально не удалось, среди всех экспериментов	Доля правильно выделенных фаз среди экспериментов с экспертной оценкой
E0	0.467	0.147	0.547
U0	0.344	0.136	0.398
U1	0.357	0.322	0.526
Total	0.378	0.206	0.476

	Доля правильно выделенных фаз среди всех экспериментов	Доля фаз, выделить которые мануально не удалось, среди всех экспериментов	Доля правильно выделенных фаз среди экспериментов с экспертной оценкой
Задание 1			
E0	0.533	0.067	0.571
U0	0.4	0.2	0.5
U1	0.304	0.304	0.438
Total	0.397	0.222	0.51
Задание 2			
E0	0.533	0.133	0.615
U0	0.24	0.12	0.273
U1	0.304	0.478	0.583
Total	0.286	0.333	0.429
Задание 3			
E0	0.333	0.133	0.385
U0	0.438	0.217	0.556
U1	0.435	0.174	0.526
Total	0.397	0.19	0.49
Задание 4			
E0	0.266	0.4	0.444
U0	0.4	0.08	0.434
U1	0.438	0.348	0.667
Total	0.349	0.38	0.564
Задание 5			
E0	0.667	0	0.667
U0	0.2	0.2	0.25
U1	0.304	0.261	0.418
Total	0.349	0.254	0.468

	Доля правильно выделенных фаз среди всех экспериментов	Доля фаз, выделить которые мануально не удалось, среди всех экспериментов	Доля правильно выделенных фаз среди экспериментов с экспертной оценкой
E0_1	0.6	0	0.6
E0_2	0.4	0	0.4
E0_3	0.6	0.2	0.75
E0_4	0.4	0.2	0.5
E0_5	0.2	0.4	0.333
E0_6	0.4	0.2	0.5
E0_7	0.6	0.2	0.75
E0_8	0.4	0	0.4
E0_9	0.8	0	0.8
E0_10	0	0.4	0
E0_11	0.4	0.0	0.4
E0_12	0.6	0.2	0.75
E0_13	0.4	0.2	0.5
E0_14	0.4	0.2	0.5
E0_15	0.8	0	0.8

	Доля правильно выделенных фаз среди всех экспериментов	Доля фаз, выделить которые мануально не удалось, среди всех экспериментов	Доля правильно выделенных фаз среди экспериментов с экспертной оценкой
U0_1	0.6	0.2	0.75
U0_2	0.4	0.2	0.5
U0_3	0.4	0	0.4
U0_4	0.2	0	0.2
U0_5	1	0	1
U0_6	0.4	0	0.4
U0_7	0.4	0	0.4
U0_8	0.4	0.4	0.667
U0_9	0.4	0	0.4
U0_10	0.2	0	0.2
U0_11	0	0	0
U0_12	0.6	0.2	0.75
U0_13	0	0.2	0
U0_14	0.4	0	0.4
U0_15	0.4	0	0.4

U0_16	0	0.2	0
U0_17	0	0	0
U0_18	0.4	0	0.4
U0_19	0.6	0	0.6
U0_20	0.8	0	0.8
U0_21	0.2	0.2	0.25
U0_22	0	0.4	0
U0_23	0.2	0.6	0.5
U0_24	0.4	0.4	0.667
U0_25	0.2	0.4	0.333

	Доля правильно выделенных фаз среди всех экспериментов	Доля фаз, выделить которые мануально не удалось, среди всех экспериментов	Доля правильно выделенных фаз среди экспериментов с экспертной оценкой
U1_1	0	0.4	0
U1_2	0.8	0	0.8
U1_3	0.4	0	0.4
U1_4	0.8	0	0.8
U1_5	0.4	0.4	0.667
U1_6	0.6	0	0.6
U1_7	0.4	0.2	0.5
U1_8	0	0.4	0
U1_9	0.6	0.4	1
U1_10	0.2	0.8	1
U1_11	1	0	1
U1_12	0.2	0.2	0.25
U1_13	0.2	0.4	0.333
U1_14	0	0.8	0
U1_15	0.6	0	0.6
U1_16	0.4	0.4	0.667
U1_17	0.2	0.4	0.333
U1_18	0.4	0.4	0.667
U1_19	0	0.2	0
U1_20	0.2	0.6	0.5
U1_21	0	0.8	0
U1_22	0.6	0.4	1
U1_23	0.2	0.2	0.25

Заключение

В данной работе были предложены различные алгоритмы для обработки потоковых данных, содержащих измерения частоты сердечных сокращений (ЧСС) при выполнении работы, для выделения фаз ментальной активности с учётом индивидуальных психофизиологических особенностей каждого испытуемого. Для решения этой задачи применялись различные алгоритмы стандартных пакетов анализа данных, а также реализованные самостоятельно алгоритмы. Предложена собственная трёхфазная схема решения данной задачи без предварительно размеченных данных с учётом только общих знаний о ментальных фазах и позволяющая выделять фазы ментальной активности с учётом физиологических и психологических особенностей каждого конкретного человека. Проведено сравнение нескольких вариантов реализации этой схемы путём запуска алгоритмов на реальных данных с последующей оценкой их качества специалистами в рассматриваемой предметной области. Наилучшие результаты показал метод, основанный на анализе поведения коэффициента линейной регрессии, аппроксимирующей значения ЧСС, на префиксе.

Список литературы

1. Heart rate variability: standards of measurement, physiological interpretation and clinical use. Task Force of the European Society of Cardiology and the North American Society of Pacing and Electrophysiology. Circulation. 1996;93:1043-65.
2. Akshansh Gupta, R. K. Agrawal, Baljeet Kaur: A three phase approach for mental task classification using EEG. ICACCI 2012: 898-904
3. Mental Stress Assessment of ECG Signal using Statistical Analysis of Bio-Orthogonal Wavelet Coefficients
4. Monitoring of mental workload levels during an everyday life office-work scenario. Burcu Cinaz, Bert Arnrich, Roberto La Marca, and Gerhard Tröster. *Personal and Ubiquitous Computing17(2):229-239* (*2013*)
5. Bayesian Online Changepoint Detection. Ryan P. Adams and David J.C. MacKay. *University of Cambridge, Cambridge, UK.* 2007