



Кафедра кибернетики (№ 22)

Расширенное содержание пояснительной записки

Разработка алгоритма классификации когнитивных состояний по данным фМРТ на основе анализа межиндивидуальных корреляций

Группа	Б14-506		
Студент	<div></div> <div>(подпись)</div>	<div>Шедько А. Ю.</div> <div>(ФИО)</div>	
Руководитель	<div></div> <div>(подпись)</div>	<div>Трофимов А. Г.</div> <div>(ФИО)</div>	
Научный консультант	<div></div> <div>(подпись)</div>	<div>-</div> <div>(ФИО)</div>	
Оценка руководителя	<div></div> <div>(0-5 баллов)</div>	Оценка консультанта	<div></div> <div>(0-5 баллов)</div>

Москва 2017

МИНИСТЕРСТВО ОБРАЗОВАНИЯ И НАУКИ РОССИЙСКОЙ ФЕДЕРАЦИИ
ФЕДЕРАЛЬНОЕ ГОСУДАРСТВЕННОЕ АВТОНОМНОЕ ОБРАЗОВАТЕЛЬНОЕ УЧРЕЖДЕНИЕ ВЫСШЕГО ОБРАЗОВАНИЯ
**«Национальный исследовательский ядерный университет
«МИФИ»**

Факультет кибернетики и информационной безопасности



КАФЕДРА КИБЕРНЕТИКИ

Задание на УИР

Студенту гр. Б14-506
(группа)

Шедько Андрею Юрьевичу
(ф.и.о.)

ТЕМА УИР

Разработка алгоритма классификации когнитивных состояний по данным фМРТ
на основе анализа межиндивидуальных корреляций

ЗАДАНИЕ

№ п/п	Содержание работы	Форма отчетности	Срок исполнения	Отметка о выполнении Дата, подпись рук.
1.	Аналитическая часть			
1.1.	Изучение и анализ подходов к классификации когнитивных состояний по данным фМРТ (статическим и динамическим) применительно к задачам медицинской диагностики	Пункт ПЗ	1.03.17	
1.2.	Сравнительный анализ методов классификации многомерных данных (линейный дискриминантный анализ, метод опорных векторов, нейросетевые методы) для выбора подходящего набора алгоритмов.	подраздел ПЗ	8.03.17	
1.3.	Сравнительный анализ программных средств визуализации трехмерных данных фМРТ и исследование возможности их использования.	Текст ПЗ	8.03.17	
1.4.	<i>Оформление расширенного содержания пояснительной записки (РСПЗ)</i>	Текст РСПЗ	27.03.17	
2.	Теоретическая часть			
2.1.	Формальная постановка задачи классификации сигналов фМРТ.	подраздел ПЗ	5.03.17	
2.2.	Выбор и разработка показателей точности классификации когнитивных состояний по фМРТ.	Формулы, Вы- ражения	10.03.17	
2.3.	Разработка алгоритма выявления значимых для классификации зон головного мозга на основе анализа межиндивидуальных корреляций.	подраздел ПЗ	14.03.17	
2.4.	Формальное описание алгоритма классификации когнитивных состояний по фМРТ.	рабочие мате- риалы	20.03.17	
2.5.	Формальное описание схемы применения алгоритма для классификации когнитивных состояний в режиме реального времени.	Текст ПЗ	20.03.17	
3.	Инженерная часть			
3.1.	Проектирование программного пакета выполняющего классификацию когнитивных состояний по данным фМРТ на основе анализа межиндивидуальных корреляций	Текст ПЗ	1.04.17	
3.2.	Результаты проектирования оформить с помощью UML диаграммы модели.	UML диа- грамма	1.04.17	

4.	Технологическая и практическая часть			
4.1.	Реализация программных модулей для экспериментальных исследований алгоритма классификации когнитивных состояний по фМРТ. с использованием программных сред MATLAB и Scipy.	Исполняемые файлы, исходный текст, подключаемый модуль для ЯП	21.03.17	
4.2.	Описание типов когнитивных состояний и исходных данных для проведения экспериментальных исследований разработанного алгоритма.	Текст ПЗ	15.03.17	
4.3.	Составление плана экспериментальных исследований разработанного алгоритма.	План эксперимента	1.04.17	
4.4.	Исследование точности классификации при различных способах оценки межиндивидуальных корреляций с использованием программных сред MATLAB и Scipy.	Схемы, графики, исходные тексты	10.04.17	
4.5.	Исследование показателей точности классификации, выявление наименее и наиболее разделимых когнитивных состояний и соответствующих зон головного мозга с использованием программных сред MATLAB и Scipy.	Схемы, графики	10.04.17	
5.	<i>Оформление пояснительной записки (ПЗ) и иллюстративного материала для доклада.</i>	Текст ПЗ, презентация	15.05.17	

ЛИТЕРАТУРА

[1]	Дьяконов В. П. MATLAB. Полный самоучитель. – М.// ДМК Пресс, 2012. – 768 с.: ил.
[2]	<i>Pajula Juha, Kauppi Jukka-Pekka, Tohka Jussi.</i> Inter-Subject Correlation in fMRI: Method Validation against Stimulus-Model Based Analysis // PLOS ONE. — 2012. — 08. — Vol. 7, no. 8. — Pp. 1–13.
[3]	<i>Pereira Francisco, Mitchell Tom, Botvinick Matthew.</i> Machine learning classifiers and fMRI: A tutorial overview // NeuroImage. — 2009. — Vol. 45, no. 1, Supplement 1. — Pp. S199 – S209. — Mathematics in Brain Imaging. http://www.sciencedirect.com/science/article/pii/S1053811908012263 .
[4]	<i>Hastie, Trevor, Tibshirani, Robert and Friedman, Jerome.</i> The elements of statistical learning: data mining, inference and prediction – 2 edition – Springer, 2009.
[5]	ГОСТ Р 7.0.53-2007 Система стандартов по информации, библиотечному и издательскому делу. Издания. Международный стандартный книжный номер. Использование и издательское оформление. — М.: Стандартинформ, 2007. — 5 с.
[6]	Буч Г., Рамбо Д., Джекобсон А. Язык UML. Руководство пользователя: Пер. с англ. М.// ДМК, 2007
[7]	<i>Kauppi J. P. et al.</i> Clustering inter-subject correlation matrices in functional magnetic resonance imaging //Information Technology and Applications in Biomedicine (ITAB), 2010 10th IEEE International Conference on. – IEEE, 2010. – С. 1-6.
[8]	<i>Ivezić Ž. et al.</i> Statistics, Data Mining, and Machine Learning in Astronomy: A Practical Python Guide for the Analysis of Survey Data. //Princeton University Press, 2014.
[9]	<i>Pajula Juha.</i> Inter-Subject Correlation Analysis for Functional Magnetic Resonance Imaging: Properties and Validation. Tampere University of Technology. Publication. — Tampere University of Technology, 2016. — 4. — Awarding institution: Tampere University of Technology.

Дата выдачи задания: _____ Руководитель _____ (_____)

(подпись) (фио)

« » 2017г. Студент _____ (_____)

(подпись) (фио)

Реферат

Пояснительная записка содержит страниц (из них XX страниц приложений). Количество использованных источников – XX. Количество приложений – X.

Ключевые слова: Межиндивидуальная корреляция, Машинное обучение, классификация, фМРТ, Кластеризация.

Целью данной работы является описание применения Межиндивидуальной корреляции для кластеризации признаков при анализе неестественных стимулов в фМРТ.

В первой главе проводится обзор и анализ ...

Во второй главе описываются использованные и разработанные/модифицированные методы-/модели/алгоритмы

В третьей главе приводится описание программной реализации и экспериментальной проверки

В приложении А приведены исходные тексты некоторых программ

Содержание

Введение	7
1 Анализ проблематики задач классификации когнитивных состояний	8
1.1 Изучение и анализ подходов к классификации когнитивных состояний по данным фМРТ (статическим и динамическим) применительно к задачам медицинской диагностики	8
1.1.1 SVM	8
1.2 Сравнительный анализ методов классификации многомерных данных	9
1.3 Сравнительный анализ программных средств визуализации трехмерных данных фМРТ и исследование возможности их использования	9
1.3.1 Nilearn	9
1.4 Выводы и постановка задачи курсового проекта	9
2 Алгоритм классификации когнитивных состояний по данным фМРТ на основе метода межиндивидуальных корреляций	11
2.1 Формальная постановка задачи.	11
2.2 Алгоритм определения информативных вокселей фМРТ	11
2.3 Алгоритм формирования вектора характерных признаков сигналов фМРТ для классификации	11
2.4 Показатели точности классификации	11
2.5 Формальное описание схемы применения алгоритма для классификации когнитивных состояний в режиме реального времени.	12
2.6 Выводы	12
3 Разработка программной системы для классификации сигналов фМРТ	13
3.1 Проектирование программного пакета выполняющего классификацию когнитивных состояний по данным фМРТ на основе анализа межиндивидуальных корреляций	13
3.2 Программная реализация системы классификации	13
3.3 Состав и структура реализованного программного обеспечения	13
3.4 Основные сценарии работы пользователя	14
3.5 Сравнение реализованного программного обеспечения с существующими аналогами	14

3.6	Выводы	14
4	Экспериментальные исследования алгоритма классификации сигналов фМРТ	15
4.1	Описание исходных данных	15
4.2	Составление плана экспериментальных исследований разработанного алгоритма .	15
4.3	Исследование точности классификации при различных способах оценки межиндивидуальных корреляций	15
4.4	Исследование показателей точности классификации, выявление наименее и наиболее разделимых когнитивных состояний и соответствующих зон головного мозга .	15
	Заключение	17
	Список литературы	18
	Список литературы	18
	Приложения	19
A	Исходные тексты программ	19

Введение

В настоящее время актуальны проблемы анализа многомерных данных, особенно в медицинских приложениях. Данная работа рассматривает новый подход задаче понижения размерности: матрицу Межиндивидуальных корреляций. Активно публикуются в этой области: Juha Pajula из университета Тампере (в 2016 году защитившего диссертацию по данной теме [5]), Jussi Tohka, Jukka-Pekka Kauppi, Юрия Хассона, впервые описавшего данный метод. Первое упоминание применения метода для задачи кластеризации данных фМРТ можно найти в статье Юрия Хассона и других в 2004[2]. Однако в этой работе рассматривались естественные стимулы (просмотр фильмов) что не соотносится с доступными авторам данными (вербальные и пространственные задачи). Из-за разреженности данных без модификации методы предыдущих исследований не применимы без модификаций.

Таким образом, получим задачу данной работы — использование метода межиндивидуальной корреляции для кластеризации данных в задаче понижения размерности. Также проводится сравнение нового метода с традиционными подходами к данной задаче, не использующими множество испытуемых (Метод главных компонент (PCA), Обобщённая линейная модель (GLM)).

Новизна работы состоит в применении метода ISC для кластеризации в условиях неестественных стимулов.

В первой главе подробно рассматриваются теоретические аспекты задачи понижения размерности, задачи классификации (Метод опорных векторов (SVM), нейронные сети, Линейный дискриминантный анализ (ЛДА)) и специфических для проблемной области (фМРТ) подходов к анализу данных. Также описываются программные средства визуализации трёхмерных данных с примерами их использования. (`nilearn.plotting`[3], `matplotlib3d`[1], NIFTI, MITK[4])

Во второй главе описаны используемые в работе алгоритмы, а именно: кластеризация на основе ISC, формирование вектора признаков, вычисление показателей точности классификации, классификация в режиме реального времени.

В третьей главе рассматриваются программные аспекты реализации алгоритмов описанных в предыдущей главе.

В заключительной главе описывается характер экспериментальных данных и количественные показатели точности работы системы. Также проводится исследование эффективности различных показателей точности классификации применительно к конкретным экспериментальным данным.

1. Анализ проблематики задач классификации когнитивных состояний

Аннотация. В первой главе подробно рассматриваются теоретические аспекты задачи понижения размерности, задачи классификации (Метод опорных векторов (SVM), нейронные сети, Линейный дискриминантный анализ (ЛДА)) и специфических для проблемной области (фМРТ) подходов к анализу данных. Также описываются программные средства визуализации трёхмерных данных с примерами их использования. (`nilearn.plotting[3]`, `matplotlib3d[1]`, `NIFTI`, `MITK[4]`)

1.1 Изучение и анализ подходов к классификации когнитивных состояний по данным фМРТ (статическим и динамическим) применительно к задачам медицинской диагностики

1.2 Сравнительный анализ методов классификации многомерных данных

Аннотация. Рассмотрим такие методы как: Метод опорных векторов (SVM), нейронные сети, Линейный дискриминантный анализ (ЛДА)

Вначале дадим общее определение задачи классификации. Пусть входные данные имеют вид $\{(x_i, c_i) | i \in \overline{1, P}\}$, где P — количество примеров, $x_i \in \mathcal{X}$, $c_i \in \mathcal{C}$, \mathcal{X} — пространство входных признаков, \mathcal{C} — конечное множество классов. Часто множество \mathcal{C} является 2-элементным, в этом случае классификация называется *бинарной*. Требуется построить функцию, которая для каждого $x \in \mathcal{X}$ даёт правильное значение c .

1.2.1 SVM

Рассмотрим метод опорных векторов, для чего вначале дадим определение метода.

Метод опорных векторов (англ. *SVM, support vector machine*) — набор схожих алгоритмов обучения с учителем, использующихся для задач классификации и регрессионного анализа. Основная идея метода — перевод исходных векторов в пространство более высокой размерности и поиск разделяющей гиперплоскости с максимальным зазором в этом пространстве. Две параллельных гиперплоскости строятся по обеим сторонам гиперплоскости, разделяющей классы. Разделяющей гиперплоскостью будет гиперплоскость, максимизирующая расстояние до двух па-

параллельных гиперплоскостей. Алгоритм работает в предположении, что чем больше разница или расстояние между этими параллельными гиперплоскостями, тем меньше будет средняя ошибка классификатора.

1.3 Сравнительный анализ программных средств визуализации трехмерных данных фМРТ и исследование возможности их использования

Аннотация. В разделе описаны различные программные компоненты для визуализации нейроданных.

1.3.1 Nilearn

Nilearn makes it easy to use many advanced machine learning, pattern recognition and multivariate statistical techniques on neuroimaging data for applications such as MVPA (Mutli-Voxel Pattern Analysis), decoding, predictive modelling, functional connectivity, brain parcellations, connectomes.

Nilearn can readily be used on task fMRI, resting-state, or VBM data.

For a machine-learning expert, the value of nilearn can be seen as domain-specific feature engineering construction, that is, shaping neuroimaging data into a feature matrix well suited to statistical learning, or vice versa.

Данная библиотека предоставляет

1.4 Выводы и постановка задачи курсового проекта

Это всегда последний пункт. Здесь, по-первых, приводятся, по пунктно, основные выводы из проделанного анализа. Например:

1. Выполнен сравнительный анализ таких-то формальных систем с точки зрения применимости к решению такой-то задачи. Ни одна из проанализированных напрямую не подходит, поэтому требуется разработать вариацию на основе системы такой-то.
2. Были проанализированы варианты программных архитектур на основе систем. С учетом требований к поддержке больших объемов данных и высоких требований к потенциалу модернизруемости, была выбрана за основу такая-то архитектура.
3. Сравнительный анализ таких-то библиотек показал, что библиотека X проще в использовании, но менее производительна, в то время как библиотека Y обеспечивает высокую производительность, но и требует значительных трудозатрат для использования. В связи с такими-то соображениями были принято решение использовать такую-то библиотеку.

Далее пишется постановка задачи, на основе выданного задания. Это должен быть связный текст в объеме до 1-1,5 страниц. В этом разделе необходимо раскрыть цели и задачи УИРа/диплома.

2. Алгоритм классификации когнитивных состояний по данным фМРТ на основе метода межиндивидуальных корреляций

В этой главе описываются разработанные/модифицированные модели/методы/ алгоритмы, или/и описывается применение известных стандартных методов. Также, в конце главы обычно приводится общая архитектура программной системы, вытекающая из описанной теории. Приведенные ниже заголовки подразделов так же весьма примерные и сильно зависят от особенностей конкретной работы.

Формулы и их части необходимо набирать в математическом режиме (символ $\$$). Во избежание переноса длинных формул между строками их стоит размещать по центру колонки, например,

$$S_{abc} = (\lambda xyz.xz(yz))abc = ac(bc),$$

и, если абзац после формулы продолжается, необходимо использовать `\noindent`.

Для набора правил вывода можно использовать пакет `mathpartir.sty`. Правила вывода могут быть вынесены в виде рисунка (см. рис. 2.1).

2.1 Формальная постановка задачи.

Аннотация. Суть алгоритма: посмотреть какие воксели действуют схожим образом для каждого типа стимулов. Для этого применим метод Межиндивидуальных корреляций.

Основные Определения и Описание данных

2.2 Алгоритм определения информативных вокселей фМРТ

2.3 Алгоритм формирования вектора характерных признаков сигналов фМРТ для классификации

2.4 Показатели точности классификации

ROC, AUC

$$\begin{array}{ccc}
\frac{M \rightarrow M'}{NM \rightarrow NM'} & (\mu) & \frac{M \rightarrow M'}{MN \rightarrow M'N} \quad (\nu) \qquad \frac{M \rightarrow M'}{\lambda x.M \rightarrow \lambda x.M'} \quad (\xi)
\end{array}$$

Рис. 2.1 – Правила редукции

2.5 Формальное описание схемы применения алгоритма для классификации когнитивных состояний в режиме реального времени.

2.6 Выводы

Необходимо перечислить, какие теоретические результаты были получены с указанием степени новизны. Например: «Была разработана такая-то модель. Она представляет собой адаптированную версию модели X, в которой уравнение Z заменено на уравнение Z'». Еще пример: «Была предложена такая-то архитектура, она отличается от типовой в том-то и том-то. Это позволяет избежать таких-то проблем.». При этом следует заниматься «высасыванием из пальца»: «Поставленная задача является типовой; для ее решения применены стандартные средства (перечислить, какие).».

3. Разработка программной системы для классификации сигналов фМРТ

В этой главе описывается, что и как было запрограммировано, отлажено, протестировано, и что в результате получилось. Большинство работ должны содержать приведенные ниже разделы. Но нужно учитывать, что точный состав этой главы, как и других глав, зависит от специфики работы.

Фрагменты программного кода в тексте необходимо выделять при помощи команды `\verb`. Многострочные листинги должны оформляться при помощи пакета `listings`.

3.1 Проектирование программного пакета выполняющего классификацию когнитивных состояний по данным фМРТ на основе анализа межиндивидуальных корреляций

3.2 Программная реализация системы классификации

Аннотация. Здесь будет описана реализация

- алгоритмов быстрой (параллельной загрузки) примеров для обучения и кластеризации.

UML

В этом разделе обосновывается выбор инструментальных средств; одним из критериев выбора могут быть какие-либо требования к разрабатываемой системе, и если этих требований много, они могут быть выделены в отдельный раздел, или же в приложение. Этот пункт не пишется, если в аналитической главе был раздел, посвященный сравнительному анализу и выбору инструментальных средств.

3.3 Состав и структура реализованного программного обеспечения

Нужно охарактеризовать реализованное ПО: является ли оно настольной программой для Windows, или веб-приложением в форме сайта/веб-сервиса, или модулем/подключаемой библиотекой, или Также нужно перечислить, из чего оно состоит: какие исполняемые файлы и их назначение, конфигурационные файлы, файлы баз данных, требования к программному и аппаратному окружению, и т.п.

Если реализованное приложение достаточно обширно, этот раздел может быть разделен на несколько: один с общим описанием, и по одному на подсистемы самого верхнего уровня.

3.4 Основные сценарии работы пользователя

Нужно помнить, что пользователем может быть не только «менеджер» или «человек в белом халате», но и другой программист. Последнее относится, в первую очередь, к реализованным библиотекам. Для «обычных» приложений нередко бывают пользователи нескольких категорий — например, обычный пользователь и администратор. Для каждой категории нужно описать, как выполняются основные функции, предпочтительно, с помощью серии скрин-шотов. Однако считается плохим тоном вставлять длинную вереницу из скрин-шотов: если их много, большую часть нужно выносить в приложение. Для *этого* раздела нормальной является плотность скрин-шотов из расчета: 1 страница скрин-шотов на 1-2 страницы текста.

3.5 Сравнение реализованного программного обеспечения с существующими аналогами

В сравнении должно быть отражено, чем полученное ПО выгодно (и невыгодно) отличается от прочих ближайших аналогов. Практика показывает, что аналоги есть всегда. А если нет аналогов, значит есть частичные решения, которые реализуют какие-то части функционала вашей системы. Тут тоже может быть относительно много таблиц и графиков.

3.6 Выводы

Следует перечислить, какие практические результаты были получены, а именно: какое программное или иное обеспечение было создано. В число результатов могут входить, например, методики тестирования, тестовые примеры (для проверки корректности/оценки характеристик тех или иных алгоритмов) и др. По каждому результату следует сделать вывод, насколько он отличается от известных промышленных аналогов и исследовательских прототипов.

4. Экспериментальные исследования алгоритма классификации сигналов фМРТ

4.1 Описание исходных данных

Воксели и Мозги во времени На рисунке 4.1 представлен результат работы следующего набора команд:??

4.2 Составление плана экспериментальных исследований разработанного алгоритма

Что хотим

Параметры точности

Вопросы,отв на кот хотим получить

При каком числе вокселей лучше точность

4.3 Исследование точности классификации при различных способах оценки межиндивидуальных корреляций

4.4 Исследование показателей точности классификации, выявление наименее и наиболее разделимых когнитивных состояний и соответствующих зон головного мозга

Графики, таблицы

ROC, AUC

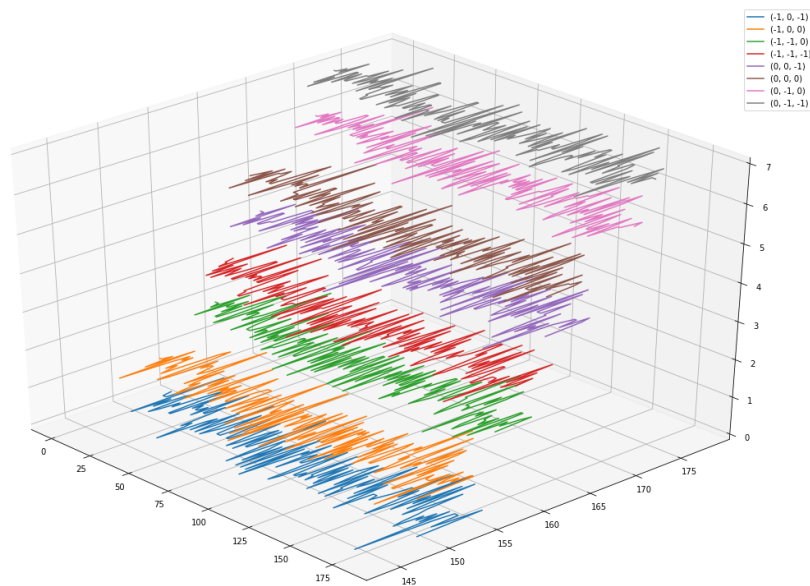


Рис. 4.1 – Пример активности локальной окрестности вокселя в течение ≈ 47 сек

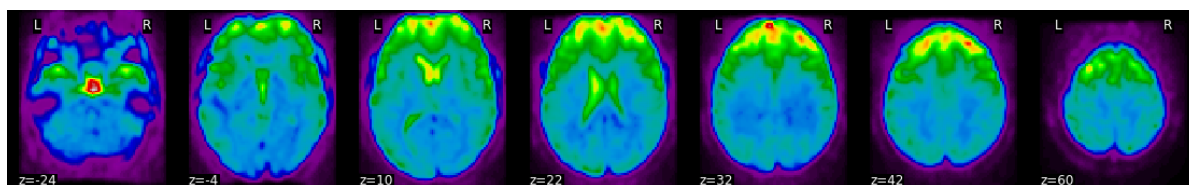


Рис. 4.2 – Разрезы мозга по z -координате

Заключение

В заключении в тезисной форме необходимо отразить результаты работы:

- аналитические (что изучено/проанализировано);
- теоретические;
- инженерные (что спроектировано);
- практические (что реализовано/внедрено).

Примерная формула такая: по каждому указанному пункту приводится по 3-5 результатов, каждый результат излагается в объеме до 5 фраз или предложений.

Также есть смысл привести предполагаемые направления для будущей работы.

Общий объем заключения не должен превышать 1,5 страниц (1 страницы для УИРов).

Список литературы

1. *Hunter J. D.* Matplotlib: A 2D graphics environment // *Computing In Science & Engineering*. — 2007. — Vol. 9, no. 3. — Pp. 90–95.
2. Intersubject Synchronization of Cortical Activity During Natural Vision / Uri Hasson, Yuval Nir, Ifat Levy et al. // *Science*. — 2004. — Vol. 303, no. 5664. — Pp. 1634–1640. <http://science.sciencemag.org/content/303/5664/1634>.
3. Machine learning for neuroimaging with scikit-learn / Alexandre Abraham, Fabian Pedregosa, Michael Eickenberg et al. // *Frontiers in Neuroinformatics*. — 2014. — Vol. 8. — P. 14. <http://journal.frontiersin.org/article/10.3389/fninf.2014.00014>.
4. The Medical Imaging Interaction Toolkit (MITK)—a toolkit facilitating the creation of interactive software by extending VTK and ITK / Ivo Wolf, Marcus Vetter, Ingmar Wegner et al. // *Proc. of SPIE* Vol. — Vol. 5367. — 2004. — P. 17.
5. *Pajula Juha.* Inter-Subject Correlation Analysis for Functional Magnetic Resonance Imaging: Properties and Validation. Tampere University of Technology. Publication. — Tampere University of Technology, 2016. — 4. — Awarding institution: Tampere University of Technology.

Приложение А. Исходные тексты программ

```
def ind(i,l):
    return (i % l -1, (i % (l**2) // l)-1,
            i % (l**3) // l**2 -1)
def geti(x,i):
    return list(map(lambda el: el[i], x))
import matplotlib.pyplot as plt
from mpl_toolkits.mplot3d import Axes3D
fig = plt.figure()
ax = fig.add_subplot(111, projection='3d')
for (i,j) in zip(list(np.argsort([np.mean(geti(x,i))
for i in range(len(r)**3)])),range(len(r)**3)):
    ax.plot(range(183),geti(x,i),j, label =
            "{}".format(ind(i,len(r))));
```

Листинг А.1 – Код для иллюстрации