

«МОСКОВСКИЙ ФИЗИКО-ТЕХНИЧЕСКИЙ ИНСТИТУТ (национальный
исследовательский университет)»

Физтех-школа прикладной математики и информатики
Кафедра «Интеллектуальные системы»

Филиппова Анастасия Владиславовна

Методы анализа движений человека на основе показаний сигналов с сенсоров

03.03.01 – Прикладные математика и физика

ВЫПУСКНАЯ КВАЛИФИКАЦИОННАЯ РАБОТА БАКАЛАВРА

Научный руководитель:

д.ф.-м.н. Стрижов Вадим Викторович

Консультант:

Нейчев Радослав

Москва

2021

Содержание

Введение	5
1 Методы построения признакового описания	7
1.1 Частотное описание сигнала	7
1.1.1 Дискретное преобразование Фурье	7
1.1.2 Вейвлет-преобразование	8
1.2 Классификация временных рядов	8
2 Постановка задачи	9
2.1 Построение выборки данных	9
2.2 Постановка задачи построения векторного представления активностей по сегменту сигнала	10
3 Обучение и тестирование модели	12
3.1 Стратегия генерации триплетов	12
3.2 Алгоритм тестирования	13
4 Вычислительный эксперимент	14
4.1 Данные	14
4.1.1 Daily and Sports Activities Data Set	14
4.1.2 Crossfit Dataset	14
4.2 Архитектура нейронной сети	15
4.3 Результаты вычислительного эксперимента	16
5 Анализ результатов	18

Аннотация

В работе рассматривается задача построения репрезентативного признакового пространства в задаче распознавания активностей человека. Исходное признаковое описание активностей - временные ряды сигналов, полученных с инерционных датчиков (акселерометр и гироскоп). Предлагается метод оптимального кодирования многомерного временного сигнала с помощью функций, параметризованных нейронной сетью. Это позволяет существенно снизить размерность признакового пространства, а также использовать полученное репрезентативное векторное представление для решения задач классификации и кластеризации. В работе предлагается процедура обучения модели, обеспечивающая быструю сходимость, а также алгоритм тестирования построенного признакового описания активностей. В работе приведены результаты численных экспериментов на данных, полученных с датчиков, закрепленных на запястье людей во время выполнения различных спортивных активностей.

Ключевые слова: *распознавание активностей, репрезентативное признаковое описание, оптимальное векторное представление, классификация активностей, нейронные сети.*

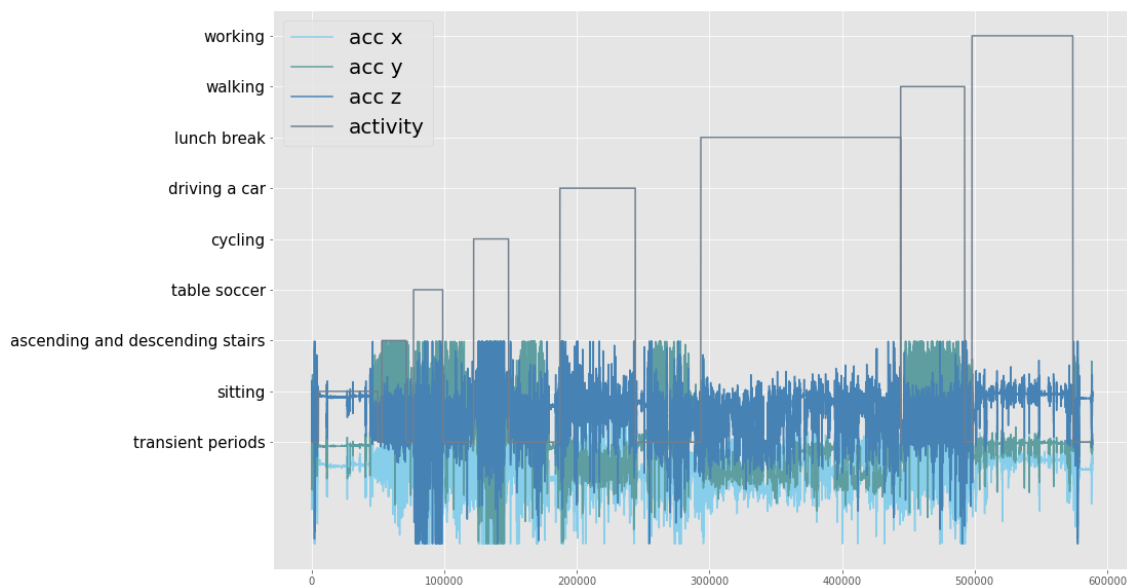
Введение

Анализ повседневной или спортивной активности (HAR) играет важную роль в анализе здоровья и поведения человека. Мониторинг деятельности человека с использованием встроенных сенсорных устройств (например, умных бытовых приборов и мобильных устройств) используется в широком спектре приложений, включая, помимо прочего, безопасность и наблюдение [1], умный дом [2], [3], здравоохранение и взаимодействие человека с компьютером [4]. Распознавание деятельности человека (HAR) использует различные датчики, как автономные, так и встроенные, для сбора необработанных сигналов, описывающих человека и его взаимодействие с окружающей средой. Среди наиболее часто используемых датчиков: акселерометр, гироскоп, фотодетекторы, GPS, а также камеры или изображения. Распознавание активности на основе датчиков, таких как акселерометр и гироскоп [5], [6], [7], направлено на получение более высокого уровня знаний непосредственно о движениях человека. Благодаря растущему распространению смартфонов и носимых устройств с большим количеством сенсоров, исследования в данном направлении в последние годы достигли огромного прогресса.

Однако, несмотря на актуальность задачи, вопрос о построении репрезентативного векторного представления для активности человека является открытым. Сложность данной задачи заключается в том, что одинаковые активности люди выполняют по-разному, следовательно, показания сенсоров разнородные. Существующие классические методы и алгоритмы векторизации не имеют достаточно обобщающей способности для аппроксимации латентного пространства.

Задача построения репрезентативного признанного описания более широкая и сложная, нежели задача классификации временных рядов сигналов, которая решается с высокой точностью [8], [9], [10]. При решении задачи классификации возникают проблемы с активностями, кото-

рых алгоритм не видел в процессе обучения. При построении векторного представления задача модель не ориентируется на конкретные классы, а лишь учится приближать векторные представления для одинаковых активностей и отдалять для различных. В данной работе предлагается реализовать нейросетевой метод построения репрезентативного векторного представления для активностей человека, используя показания акселерометра и гироскопа. Для обучения используется специальная триплет функция потерь (Triplet Loss Function). Также представлены различные методы тестирования построенных векторных представлений (knn, Random Forest classifier, SVM). Показано, что предлагаемый алгоритм работает как для рутинных активностей (положение покоя, ходьба, легкий бег, работа за компьютером), так и для интенсивных занятий спортом (упражнения на пресс, приседания, отжимания и другие кроссфит упражнения). Также представлены различные методы тестирования построенных векторных представлений (knn, Random Forest classifier, SVM).



1 Методы построения признакового описания

1.1 Частотное описание сигнала

Распространенным подходом является переход из временного домена в частотный при описании сигнала. При переходе из временного пространства в частотное анализируется не сам сигнал, а его спектор или спектрограмма. Под спектром сигнала подразумевается набор коэффициентов перед некоторыми базовыми функциями (например, синусоидальными волнами), которые, будучи надлежащим образом скомбинированы, дают изучаемый нами сигнал во временной области.

1.1.1 Дискретное преобразование Фурье

В качестве инструмента спектрального анализа в области обработки дискретных сигналов часто выступает дискретное преобразование Фурье.

Определение 1. Дискретное преобразование Фурье сигнала $\mathbf{x} = [x_1, x_2, \dots, x_N] \in \mathbb{R}^N$ есть $\mathbf{X}_k = \sum_{n=0}^{N-1} x_n \cdot e^{-i2\pi kn/N}$, базисные функции есть $W_N^{nk} = e^{-i2\pi kn/N}$, коэффициенты перед базисными функциями принято называть коэффициентами Фурье.

Дискретное преобразование Фурье также может быть применено к апериодическим сигналам. В работах [3], [2] авторы используют коэффициенты Фурье для сегмента сигнала в качестве признакового описания.

Также в качестве признакового описания можно использовать не непосредственно коэффициенты Фурье сигнала, а спектрограмму.

Определение 2. Спектрограммой называется изображение, показывающее зависимость спектральной плотности мощности сигнала от времени.

Представление сигнала в виде спектрограм широко используется в анализе аудио сигналов [11].

1.1.2 Вейвлет-преобразование

Основным недостатком преобразования Фурье является то, что оно захватывает информацию о глобальной частоте, то есть о частотах, которые сохраняются на протяжении всего сигнала. Альтернативный подход - это вейвлет-преобразование, которое разлагает функцию на набор вейвлетов.

Определение 3. *Вейвлет-преобразование — интегральное преобразование, которое представляет собой свертку вейвлет-функции с сигналом:*

$$F(a, b) = \int_{-\infty}^{\infty} f(x)\psi_{(a,b)}^*(x)dx$$

, где $*$ — комплексное сопряжение, а $\psi(x)$ — некоторая функция.

1.2 Классификация временных рядов

Задача анализа движений человека часто ставится как задача классификации временных рядов, в частности задача классификации активности человека [9], [8]. Для решения задачи классификации используются как классические методы машинного обучения (SVM, Random Forest и так далее), так и нейросетевые модели (сверточные и рекуррентные сети). При такой постановке задачи появляются проблемы, связанные с получением информации о действиях и активностях, которых не было непосредственно в обучении. К сожалению, данную проблему можно обойти только изменив постановку задачи. В данной работе будет предложен метод построения признакового описания временного ряда с использованием структуры сигнала временного ряда, но без решения непосредственно задачи классификации.

2 Постановка задачи

2.1 Построение выборки данных

Пусть $\mathbf{a}_t^i = (a_t^x, a_t^y, a_t^z)^T \in \mathbb{R}^3$ - показания акселерометра в момент времени t для i -го эксперимента, $\mathbf{g}_t^i = (g_t^x, a_t^y, a_t^z)^T \in \mathbb{R}^3$ - показания гироскопа в момент времени t . Обозначим за \mathbf{a}^i временной ряд сигнала акселерометра для i -го эксперимента, тогда $\mathbf{a}^i = [\mathbf{a}_0, \mathbf{a}_1, \dots, \mathbf{a}_{T-1}] \in \mathbb{R}^{3 \times T}$, аналогично временной ряд сигнала, полученный с гироскопа : $\mathbf{g}^i = [\mathbf{g}_0, \mathbf{g}_1, \dots, \mathbf{g}_{T-1}] \in \mathbb{R}^{3 \times T}$. Обозначим $\mathbf{x}^i = \begin{bmatrix} \mathbf{a}^i \\ \mathbf{g}^i \end{bmatrix} \in \mathbb{R}^{6 \times T}$ - временной ряд сигнала, соответствующий i -ому эксперименту. Каждой точке последовательности \mathbf{x}^i ставится в соответствие метка активности: $\mathbf{act}^i \in \mathbb{R}^T$, где act_t^i - метка активности, соответствующая i -тому эксперименту в момент времени t . Выборка: пусть $\mathbf{X} = \{\mathbf{x}^i\}_{i=1}^n$ - множество экспериментов. Каждый эксперимент разбивается на пересекающиеся сегменты длины w с фиксированным шагом s : $\mathbf{x}_{t,t+w-1}^i = (\mathbf{x}_t, \dots, \mathbf{x}_{t+w-1}) \in \mathbb{R}^{6 \times w}$ - m -ый сегмент, соответствующий i -ому эксперименту.

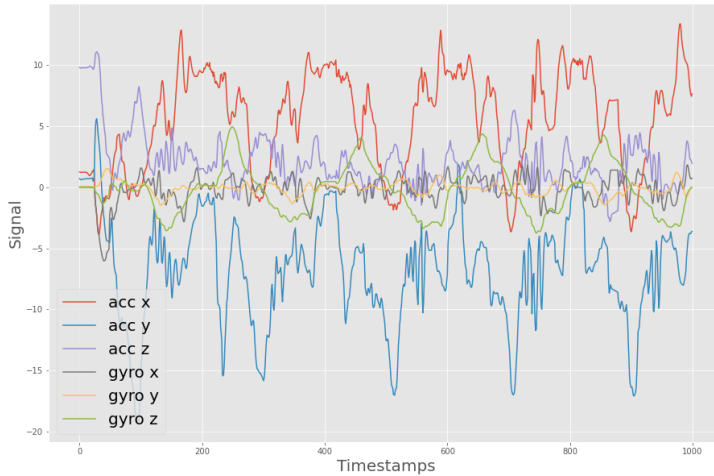


Рис. 1: Сегмент временного ряда сигнала длиной 10 секунд, соответствующий активности "Приседания"

Каждому такому сегменту временного ряда можно поставить в соответствие метку активности, которая характеризует класс движения, которое человек делает в данном временном интервале. Зададим правило для взаимнооднозначного соответствия следующим образом: разобьем после-

довательность меток активностей для каждой временной точки \mathbf{act}^i на сегменты длины w с шагом s (данные параметры такие же, как и при разбиении временного ряда сигнала). Получим сегменты длины w , количество которых, очевидно, равно количеству сегментов ряда сигнала. Метка сегмента сигнала $\mathbf{x}_{t,t+w-1}^i: l_{t,t+w-1}^i = mode(\mathbf{act}_{t,t+w-1}^i)$. Таким образом получаем обучающую выборку: $(\{\mathbf{x}_m^i, l_m^i\}_{m=0}^{\lfloor \frac{T_i}{s} \rfloor})$, где $i \in [0, n]$, n - число экспериментов, T_i - время i -ого эксперимента.

2.2 Постановка задачи построения векторного представления активностей по сегменту сигнала

Рассматривается множество параметрических моделей \mathfrak{F} :

$$\mathfrak{F} = \{\mathbf{f}: (\mathbf{w}, \mathbf{X}_m) \rightarrow \mathbf{v}\},$$

где $\mathbf{X}_m \in \mathbb{R}^{wx6}$ – исходный сегмент, $\mathbf{v} \in \mathbb{R}^d$, – векторное представление сегмента сигнала, \mathbf{w} – параметры модели. Для нахождения оптимального отображения из пространства \mathbb{R}^{6xw} в пространство \mathbb{R}^d (нахождение параметров нейронной сети) решается задача оптимизации:

$$\mathbf{f} = arg \min_{\mathbf{f} \in \mathfrak{F}} \mathcal{L},$$

где \mathcal{L} целевая функция потерь.

В работе в качестве целевой функции потерь используется функция ошибки триплетов.

Определение 4. *Триплетом называется тройка сегментов временного ряда сигнала: $(\mathbf{X}_a, \mathbf{X}_p, \mathbf{X}_n)$, где сегмент \mathbf{X}_a называется якорем и обладает некоторой активностью a , сегмент \mathbf{X}_p отличен от первого, но обладает той же активностью, что и первый сегмент. \mathbf{X}_n – сегмент другой активности.*

Функция ошибки триплетов (англ. Triplet Loss Function) для одного триплета:

$$L(a, n, p) = \max(\|\mathbf{f}(\mathbf{X}_a) - \mathbf{f}(\mathbf{X}_p)\|^2 - \|\mathbf{f}(\mathbf{X}_a) - \mathbf{f}(\mathbf{X}_n)\|^2 + \text{margin}, 0),$$

Тогда целевая функция потерь есть сумма функции ошибки триплетов по всем тройкам:

$$\mathcal{L} = \sum_{a,n,k} [L(a, n, k)].$$

В процессе обучения

Пусть k - число возможных активностей. Следовательно, $\forall j \in [0, N] l_j \in [0, \dots, k - 1]$, где $N = \sum_{i=0}^n \lfloor \frac{T_i}{s} \rfloor$ - общее число сегментов в полученной выборке.

Функция \mathbf{f} должна удовлетворять следующему требованию:

$$\forall a, p, n \text{ таких, что } a = p \neq n \hookrightarrow \|\mathbf{f}(\mathbf{X}_a) - \mathbf{f}(\mathbf{X}_p)\|^2 < \|\mathbf{f}(\mathbf{X}_a) - \mathbf{f}(\mathbf{X}_n)\|^2,$$

где $a, p, n \in \mathbb{L}$.

То есть расстояние между векторными представлениями сегментов, имеющих одинаковую метку активности должно быть меньше, чем до любого сегмента, обладающего другой меткой.

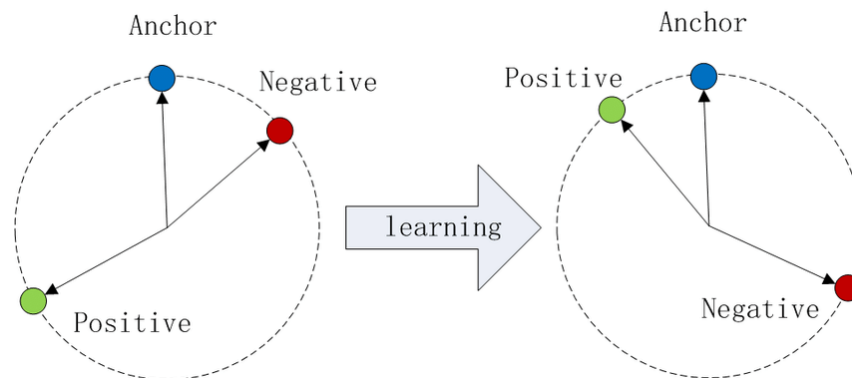


Рис. 2: Триплет лосс функция

3 Обучение и тестирование модели

3.1 Стратегия генерации триплетов

Число сегментов временного ряда велико, так как каждый временной ряд (каждый эксперимент) разбивается на пересекающиеся сегменты длиной 5–10 секунд при общей длине эксперимента 60–90 минут. Использовать все возможные триплеты в процессе обучения не оптимально, так как в таком случае время обучения будет выше. Следовательно, необходимо использовать оптимальную стратегию генерации триплетов в процессе обучения, которая будет обеспечивать быструю сходимость.

Рассмотрим возможные варианты:

Определение 5. *Сложными положительными примерами будем называть пары сегментов $(\mathbf{X}_a, \mathbf{X}_p)$, для которых справедливо:*

$$\arg \max_{\mathbf{X}_p} d(\mathbf{f}(\mathbf{X}_a), \mathbf{f}(\mathbf{X}_p))$$

.

Аналогично сложные отрицательные примеры это пары $(\mathbf{X}_a, \mathbf{X}_n)$, для которых справедливо:

$$\arg \min_{\mathbf{X}_n} d(\mathbf{f}(\mathbf{X}_a), \mathbf{f}(\mathbf{X}_n))$$

.

Во время обучения для каждого якоря (каждого сегмента) в обучающей выборке предлагается выбирать k первых сложных положительных примеров и k первых сложных отрицательных примеров. Полученные таким образом $2k$ сегментов предлагается использовать для генерации триплетов.

3.2 Алгоритм тестирования

На этапе тестирования необходимо проверить обобщающую способность модели как для активностей, которые были на этапе обучения, так и для тех активностей, на которых модель не обучалась. В связи с этим предлагается разделить валидационную выборку на две части: на трейновую и тестовую. На трейновой части валидационной выборки происходит обучение классификатора для тестирования модели.

В качестве классификаторов использовались KNN, Random Forest, Logistic Regression. В процессе обучения классификаторов фиксировалось число тренировочных семплов, чтобы можно было проанализировать зависимость качества классификации от размера тренировочной выборки. Далее рассчитывались следующие метрики: F_1 и *accuracy-score* вычислялись только для специальных классов, после чего происходило усреднение полученных значений метрики.

На изображении 3 представлена диаграмма, описывающая алгоритм обучения и тестирования модели.

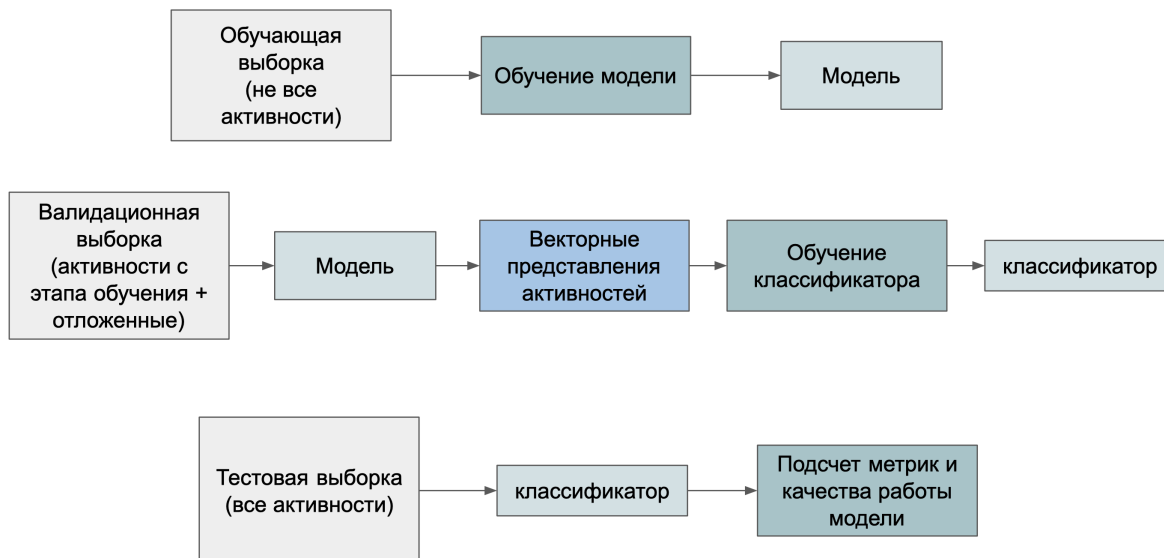


Рис. 3: Диаграмма обучения и тестирования

4 Вычислительный эксперимент

Эксперимент проводился на двух наборах данных: Daily and Sports Activities Dataset и Crossfit Dataset.

4.1 Данные

4.1.1 Daily and Sports Activities Data Set

Данный датасет содержит показания акселерометра и гироскопа, установленных на разных частях тела человека, в том числе на запястье (часы). В датасете присутствуют 19 активностей для 10 испытуемых. Каждый эксперимент содержит одну активность и длится 6 минут. Активности описаны в Таблице 1.

4.1.2 Crossfit Dataset

Датасет содержит только интенсивные спортивные активности (подробнее Таблица 2). Во время выполнения упражнений браслет на запястье испытуемых записывает показания сенсоров (акселерометра и гироскопа). Каждое упражнение выполняется с 10 – 20 повторениями. Данный датасет далее будет использоваться для решения задачи построения дескрипторов активностей.

Activity	Label
sitting	1
standing	2
lying on back	3
lying on right side	4
ascending stairs	5
descending stairs	6
standing in an elevator still	7
moving around in an elevator	8
walking in a parking lot	9
walking on a treadmill with a speed of 4 km/h in flat position	10
walking on a treadmill with a speed of 4 km/h in 15 deg inclined positions	11
running on a treadmill with a speed of 8 km/h	12
exercising on a stepper	13
exercising on a cross trainer	14
cycling on an exercise bike in horizontal position	15
cycling on an exercise bike in vertical positions	16
rowing	17
jumping	18
playing basketball	19

Таблица 1: Daily sports dataset activities

4.2 Архитектура нейронной сети

Мы использовали ResNet-18 без последнего слоя при проведении вычислительного эксперимента. Такая модификация известной архитектуры очень мало отличается от ResNet-18, и ей требуется в 3,7 раз меньше памяти, поэтому мы решили использовать ее как более практичный вариант. Входной тензор 6 — : Архитектура полученной сети проиллюстрирована на

Activity	Label
box jumps	0
squats	1
dead lifts	2
pull ups	3
KB press	4
push ups	5
burpees	6
wall balls	7
KB squart press	8
crunches	9

Таблица 2: Crossfit dataset activities

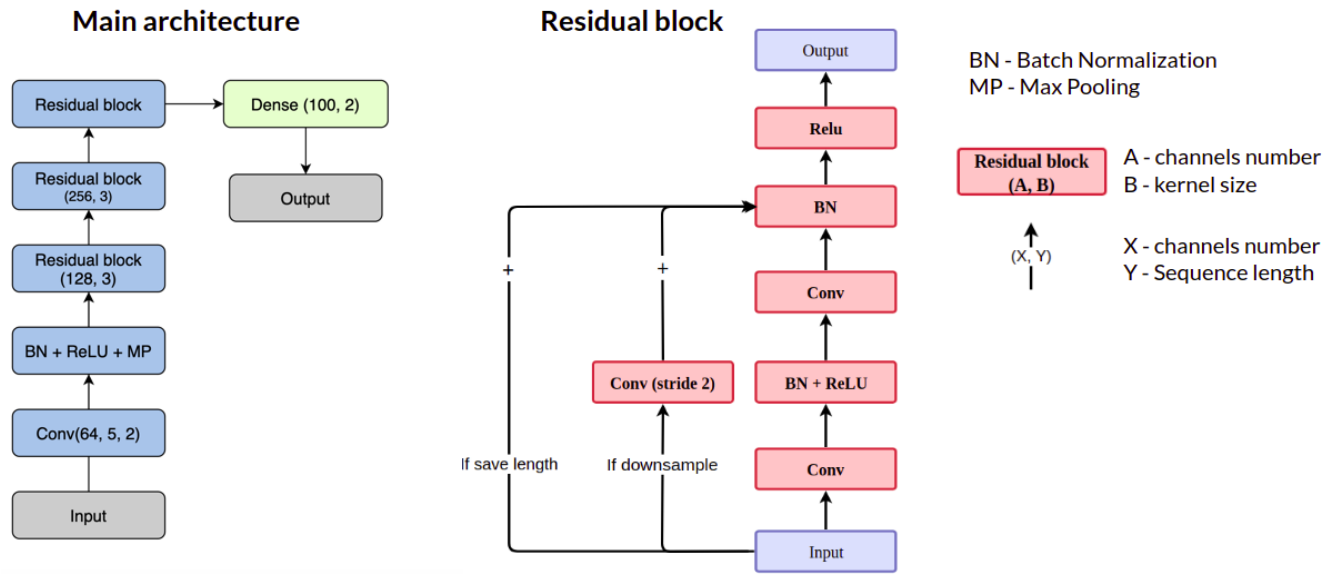


Рис. 4: Архитектура нейронной сети

4.3 Результаты вычислительного эксперимента

На таблицах 3 и 4 представлены соответственно результаты работы модели на датасетах Crossfit Activities Dataset и Daily Sports activities. Как видно из таблиц полученные признаковые описания активностей имеют

высокое качество для каждого из методов тестирования.

classifier	F1	accuracy
kNN	89.19 ± 0.2	87.55 ± 0.2
SVM	88.44 ± 0.2	87.20 ± 0.2
Random Forest	88.69 ± 0.2	87.21 ± 0.2

Таблица 3: Crossfit dataset activities

classifier	F1	accuracy
kNN	92.39 ± 0.1	93.55 ± 0.1
SVM	90.34 ± 0.1	91.23 ± 0.1
Random Forest	89.83 ± 0.1	89.54 ± 0.1

Таблица 4: Daily Sports activities

5 Анализ результатов

На изображении 5 представлена матрица неточностей для построенных векторных представлений активностей. По матрице можно сделать вывод, что модель строит репрезентативные векторные представления для классов, которых не было в обучении: приседания и жим гантели вверх (Squats и Pull ups). По матрице неточностей также можно сделать вывод, что модель часто ошибается, выбирая между активностями приседания и становая тяга (Squats и Dead lifts). Такие ошибки возникают, так как активности похожи по своей природе: человек выполняет очень похожие движения руками, тем самым расстояние между построенными векторными представлениями одного класса получаются близкими с векторными представлениями другого класса.

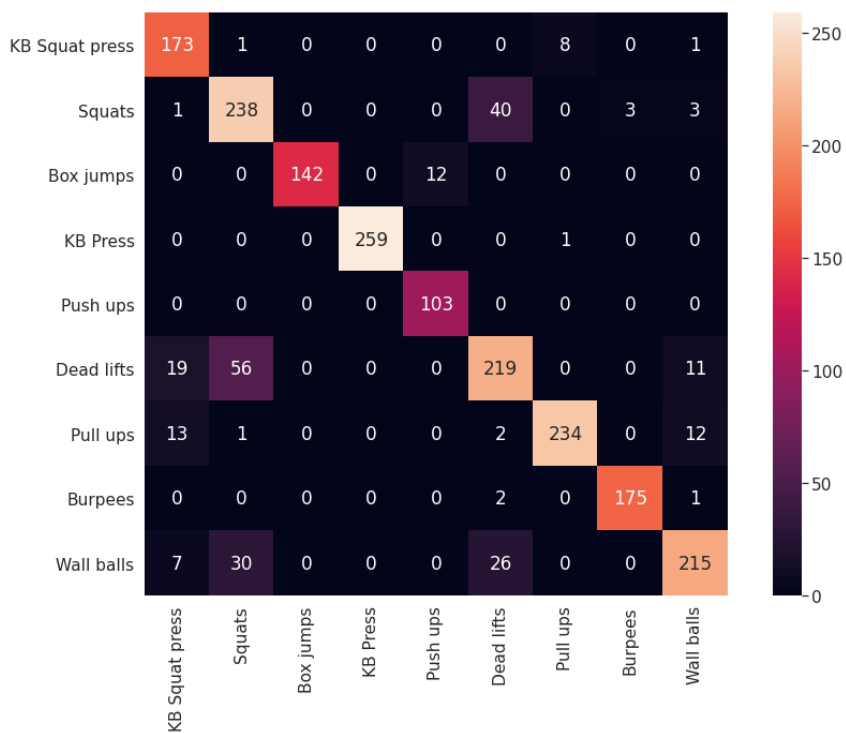


Рис. 5: Матрица неточности

На изображении 6 представлены построенные T-SNE проекции в двумерное пространство для полученных векторных представлений.

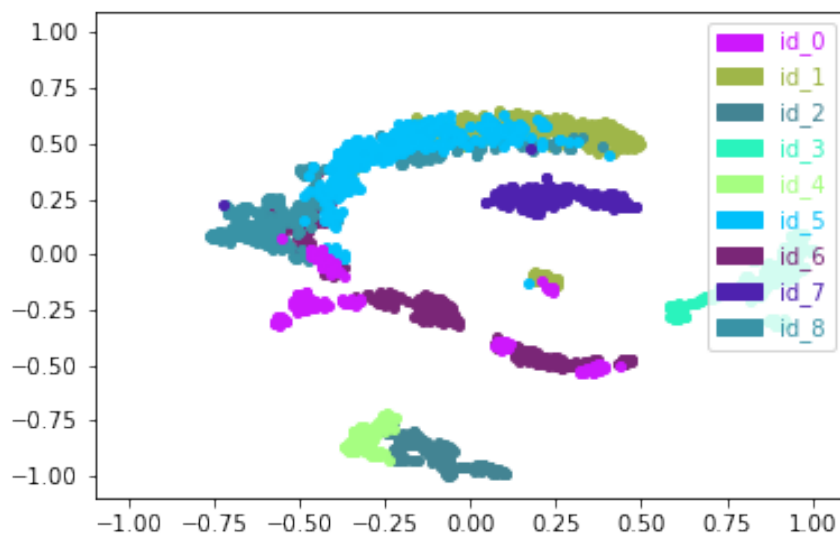


Рис. 6: TSNE карты полученных векторных представлений

6 Заключение

В данной работе были рассмотрены существующие методы построения признакового описания, а также различные методы постановки задачи исследования движений человека. В работе был предложен метод построения признакового описания временного ряда сигналов акселерометра и гироскопа, полученного с носимых человеком устройств, во время повседневной и спортивной активностей. Предлагаемый алгоритм анализа движений человека пользуется структурой сигнала без решения задачи классификации. В процессе обучения предлагается использовать функцию потерь, использование которой позволяет обучать модель таким образом, чтобы приближать одинаковые активности и отдалять различные. Таким образом, мы получаем признаковое описание сигнала, не обзновываясь непосредственно на класс активности, к которому принадлежит данные временной ряд, а лишь на структуру самого сигнала. Также в работе был предложен алгоритм тестирования построенных векторных представлений. По результатам проведенных экспериментов можно сделать вывод, что построенное признаковое пространство является информативным, то есть признаковое представления сигналов репрезентативно и его можно использовать для решений задач кластеризации или классификации.

Список литературы

- [1] Lulu Chen, Hong Wei, and James Ferryman. A survey of human motion analysis using depth imagery. *Pattern Recognition Letters*, 34(15):1995–2006, 2013. Smart Approaches for Human Action Recognition.
- [2] Sandeep Pirbhulal, Heye Zhang, Md Eshrat E Alahi, Hemant Ghayvat, Subhas Chandra Mukhopadhyay, Yuan-Ting Zhang, and Wanqing Wu. A novel secure iot-based smart home automation system using a wireless sensor network. *Sensors*, 17(1):69, 2017.
- [3] Franco Cicirelli, Giancarlo Fortino, Andrea Giordano, Antonio Guerrieri, Giandomenico Spezzano, and Andrea Vinci. On the design of smart homes: A framework for activity recognition in home environment. *Journal of medical systems*, 40(9):1–17, 2016.
- [4] Pedro Canotilho Ribeiro, José Santos-Victor, and P Lisboa. Human activity recognition from video: modeling, feature selection and classification architecture. In *Proceedings of International Workshop on Human Activity Recognition and Modelling*, pages 61–78. Citeseer, 2005.
- [5] Nishkam Ravi, Nikhil Dandekar, Preetham Mysore, and Michael L Littman. Activity recognition from accelerometer data. In *Aaai*, volume 5, pages 1541–1546. Pittsburgh, PA, 2005.
- [6] Jindong Wang, Yiqiang Chen, Shuji Hao, Xiaohui Peng, and Lisha Hu. Deep learning for sensor-based activity recognition: A survey. *Pattern Recognition Letters*, 119:3–11, 2019.
- [7] G De Leonardis, Samanta Rosati, Gabriella Balestra, Valentina Agostini, E Panero, Laura Gastaldi, and Marco Knaflitz. Human activity recognition by wearable sensors: Comparison of different classifiers for real-time applications. In *2018 IEEE International Symposium on*

- Medical Measurements and Applications (MeMeA)*, pages 1–6. IEEE, 2018.
- [8] Ankita Jain and Vivek Kanhangad. Human activity classification in smartphones using accelerometer and gyroscope sensors. *IEEE Sensors Journal*, 18(3):1169–1177, 2017.
- [9] Raihani Mohamed, Mohammad Noorazlan Shah Zainudin, Md Nasir Sulaiman, Thinagaran Perumal, and Norwati Mustapha. Multi-label classification for physical activity recognition from various accelerometer sensor positions. *Journal of Information and Communication Technology*, 17(2):209–231, 2018.
- [10] Kangjae Lee and Mei-Po Kwan. Physical activity classification in free-living conditions using smartphone accelerometer data and exploration of predicted results. *Computers, Environment and Urban Systems*, 67:124–131, 2018.
- [11] Lonce Wyse. Audio spectrogram representations for processing with convolutional neural networks. *arXiv preprint arXiv:1706.09559*, 2017.