

Классификация суперпозиций движений физической активности

Александр Денисович Проскурин
Евгений Александрович Белых

Московский физико-технический институт

Курс: Численные методы обучения по прецедентам
(практика, В. В. Стрижов)/Группа 594, весна 2018

Цель исследования

Найти способ распознать сложные движения человека, являющиеся суперпозицией более простых движений, используя данные акселерометра телефона.

Проблема

Данные являются непериодическими временными рядами, поэтому одной из задач является поиск оптимального способа сегментации и описания временного ряда, а также построение метрического пространства описаний физических движений.

Новизна

Предлагается рассматривать движение как суперпозицию действий, а не однородный процесс.

- *Карасиков М.Е., Стрижов В.В.* Классификация временных рядов в пространстве параметров порождающих моделей // Информатика и ее применения, 2016
- *Кузнецов М.П., Ивкин Н.П.* Алгоритм классификации временных рядов акселерометра по комбинированному признаковому описанию // Машинное обучение и анализ данных, 2015
- *Фадеев И.В.* Выбор иерархических моделей в авторегрессионном прогнозировании // Магистерская диссертация, 2013, Московский физико-технический институт

Постановка задачи

Пусть $D(X, Y)$ — это обучающая выборка, где (X, ρ) образует метрическое пространство временных рядов, Y — это метки временных рядов.

Пусть F — это функция построения множества признаков временного ряда:

$$F : X \rightarrow \mathbb{R}^n$$

Пусть G — многоклассовый классификатор, который переводит признаки в метки:

$$G : \mathbb{R}^n \rightarrow Y$$

Итоговый алгоритм — это композиция некоторого G и F :

$$a = G \circ F$$

Пусть задана некоторая функция потерь $L : X \times Y \times Y \rightarrow \mathbb{R}$, тогда функционал качества имеет вид:

$$Q(a, D) = \frac{1}{|D|} \sum_{(x,y) \in D} L(x, a(x), y).$$

Предлагается в качестве функции потерь использовать идентификатор

$$Q(a, D) = \frac{1}{|D|} \sum_{(x,y) \in D} I(a(x) = y)$$

В методе обучения будем сначала фиксировать F , а после оптимизировать функцию G , тогда наш оптимальный алгоритм принимает следующий вид $a_F = \hat{G} \circ F$, где

$$\hat{G} = \operatorname{argmin}_G (Q(G \circ F, D))$$

Чтобы оценить эффективность метода обучения, мы будем разбивать нашу выборку r раз на тестовую и тренировочную ($D = A_1 \cup B_1 = A_2 \cup B_2 = \dots = A_r \cup B_r$), в таком случае наш критерий качества a_F будет:

$$QV(a_F, D) = \frac{1}{r} \sum_{i=1}^r Q(a_F(A_i), B_i)$$

Итоговая цель — найти

$$\hat{a} = \operatorname{argmin}_F (QV(a_F, D))$$

Рассмотрим параметрическую модель, которая будет приближать реальные значения нашего временного ряда:

$$g(w, X) \rightarrow X, \text{ где } w \in \mathbb{R}^n.$$

В качестве параметрической модели рассмотрим:

- Авторегрессионную модель $AR(p)$:

Пусть $x = [x_1, x_2, \dots, x_t]$ — временной ряд, где $x_i \in \mathbb{R}$. Тогда

$$g(w, x) = [\hat{x}_1, \hat{x}_2, \dots, \hat{x}_t],$$

$$\text{где } \hat{x}_k = \begin{cases} x_k, & k = 1, \dots, p, \\ w_0 + \sum_{i=1}^p w_i \cdot x_{k-i}, & k = p+1, \dots, t. \end{cases}$$

- Преобразование Фурье

Для завершения построения классификатора временных рядов, необходимо построить классификатор G по обучающей выборке $\{(F(x), y) \mid (x, y) \in D\}$.

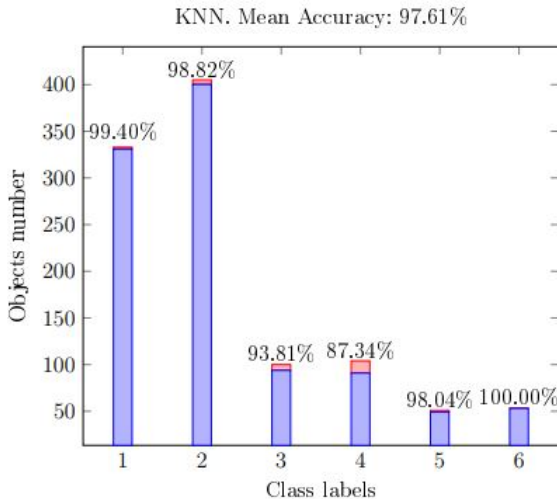
В качестве классификатора используем следующие модели:

- KNN
- Random forest
- Logistic regression

В вычислительном эксперименте предполагается использовать следующие данные:

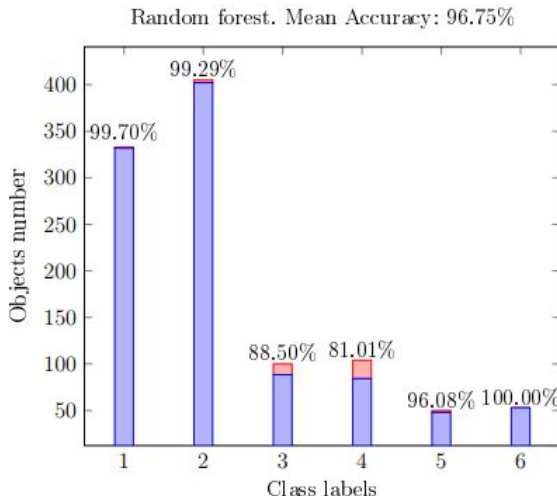
- Датасет WISDM для базового вычислительного эксперимента
- Собранные самостоятельно для проверки полученного итогового алгоритма

Результаты при использовании классификатора KNN:

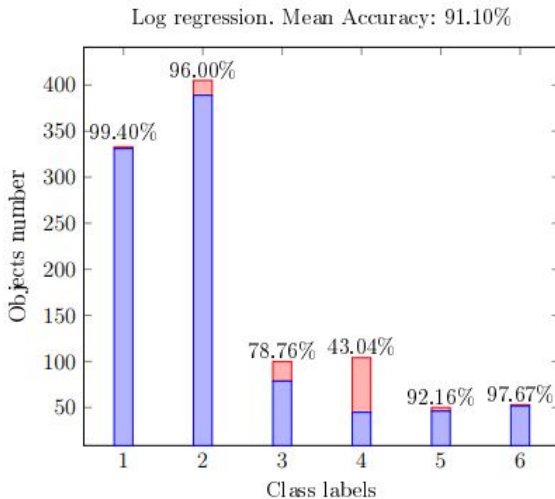


Датасет WISDM. Авторегрессионная модель

Результаты при использовании классификатора Random forest:



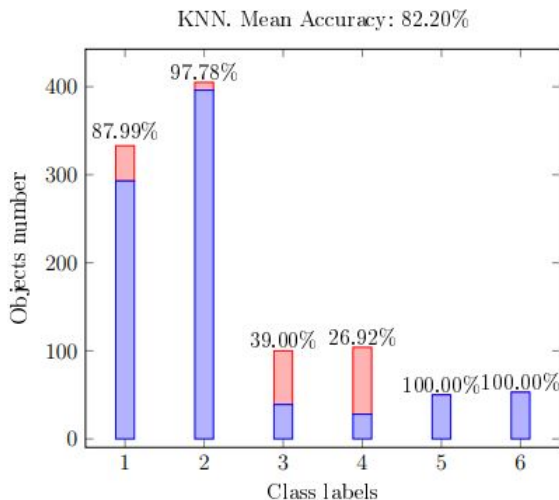
Результаты при использовании классификатора Logistic regression:



Как видно, все три классификатора показывают хорошие результаты и классифицируют данные с точностью более 91%. В то же время, наилучшие результаты показывает KNN, при этом все классификаторы показывают наихудшие результаты для классов «Upstairs» и «Downstairs». При этом, все алгоритмы можно улучшать с помощью, например, добавления дополнительных признаков или перебора параметров классификаторов.

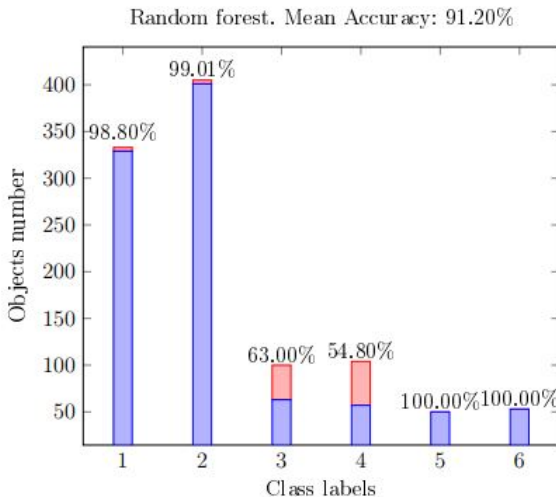
Датасет WISDM. Преобразование Фурье

Результаты при использовании классификатора KNN:



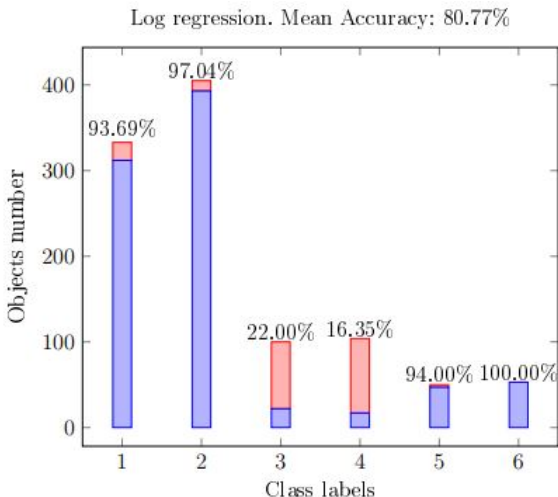
Датасет WISDM. Преобразование Фурье

Результаты при использовании классификатора Random forest:



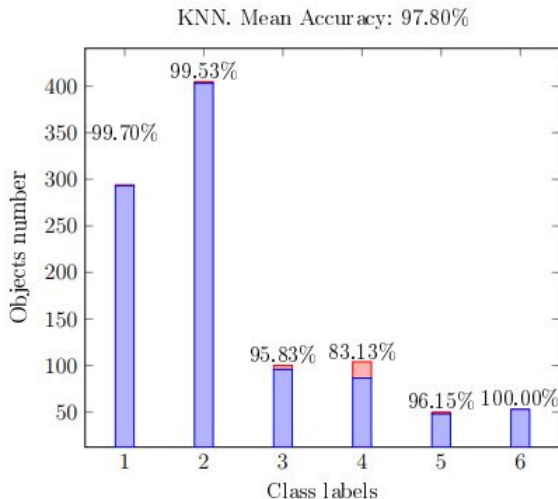
Датасет WISDM. Преобразование Фурье

Результаты при использовании классификатора Logistic regression:



Датасет WISDM. Комбинирование линейного преобразования рядов и модели авторегрессии

Результаты при использовании классификатора KNN:



Датасет WISDM. Комбинирование линейного преобразования рядов и модели авторегрессии

Результаты при использовании классификатора Random forest:

