

Анализ двигательной активности человека на основе данных с мобильных устройств

И. Р. Юсупов

Московский физико-технический институт
Факультет управления и прикладной математики
Кафедра интеллектуальных систем

Научный руководитель Р. Г. Нейчев

Москва, 2020

Цель

Построить признаковое пространство с помощью моделей локальной аппроксимации для устойчивой классификации сигналов носимых устройств.

Задачи

- Предложить признаковое пространство для временных рядов
- Предложить алгоритм классификации

Исследуемая проблема

Извлечение признаков, на которых алгоритмы классификации выдают более высокое качество.

Постановка задачи классификации

Задан временной ряд, то есть собранный в разные моменты времени статистический материал о значении каких-либо параметров.

$$x = [x^{(1)}, x^{(2)}, \dots, x^{(t)}]$$

где $x^{(i)}$ - вектор значений в момент времени i .

Задана модель локальной аппроксимации $h(x)$, которая отображает пространство временных рядов X в промежуточное пространство признаков описаний Z

Модель классификации

$$X \xrightarrow{h} Z \xrightarrow{a} Y$$

где a - многоклассовый классификатор, Y - множество меток класса.

Модели локальной аппроксимации

Модель	Признаковое пространство
Векторная авторегрессия (VAR)	Веса модели
Анализ сингулярного спектра (SSA)	Вектор сингулярных чисел
Анализ сингулярного спектра (SSA)	Элементарные временные ряды

Векторная авторегрессия (VAR)

Формально векторная авторегрессия – это система уравнений, каждое из которых представляет из себя модель авторегрессии.

$$x_t = W_0 + W_1x_{t-1} + W_2x_{t-2} + \dots + W_px_{t-p} + \varepsilon_t$$

Анализ сингулярного спектра (SSA)

Анализ сингулярного спектра – метод анализа временных рядов, основанный на преобразовании одномерного временного ряда в многомерный ряд с последующим применением к полученному многомерному временному ряду метода главных компонент.

$$X = U\Sigma V^T$$

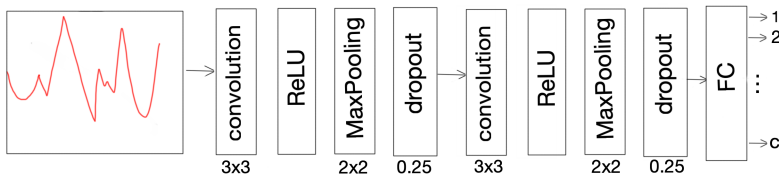
Описание данных

Датасет HAR содержит данные с акселерометра для 6 видов деятельности человека: WALKING, WALKING UP, WALKING DOWN, SITTING, STANDING, LAYING. Датасет содержит 7352 объекта для обучения и 2947 объектов для теста, данные представляют собой многомерные ряды размером (128, 9), то есть каждый многомерный ряд представляет собой 9 одномерных рядов длиной 128.

Классификация без преобразования признакового пространства

Классификатор

В качестве алгоритма для решения задачи многоклассовой классификации использовалась сверточная нейронная сеть



Архитектура нейронной сети

class	precision	recall	f1-score
0	0.99	0.94	0.96
1	0.98	0.94	0.96
2	0.89	1.00	0.94
3	0.85	0.77	0.81
4	0.82	0.87	0.84
5	0.99	0.99	0.99
accuracy			0.92
macro avg	0.92	0.92	0.92
weighted avg	0.92	0.92	0.92

Таблица: Результат классификации сверточной нейронной сети

Признаковое пространство

В качестве признакового пространства использовались веса модели векторной авторегрессии

Классификаторы

В качестве алгоритмов многоклассовой классификации использовались:

- Полносвязная нейронная сеть
- Gradient Boosting
- SVM
- Random Forest
- Adaptive Boostin
- Hist Gradient Boosting

Результаты эксперимента

	accuracy	macro avg precision	macro avg recall	macro avg f1-score
FNN	0.67	0.69	0.67	0.67
Gradient Boosting	0.75	0.77	0.75	0.75
Random Forest	0.73	0.73	0.73	0.73
SVM	0.41	0.44	0.41	0.38
Hist Gradient Boosting	0.76	0.78	0.76	0.76
Adaptive Boosting	0.67	0.68	0.67	0.67

Признаковое пространство

В качестве признакового пространства рассматривались сингулярные числа траекторной матрицы временного ряда.

Классификаторы

В качестве алгоритмов многоклассовой классификации использовались:

- Полносвязная нейронная сеть
- Gradient Boosting
- SVM
- Random Forest
- Adaptive Boostin
- Hist Gradient Boosting

	accuracy	macro avg precision	macro avg recall	macro avg f1-score
FNN	0.74	0.75	0.74	0.74
Gradient Boosting	0.53	0.53	0.54	0.51
Random Forest	0.55	0.60	0.56	0.52
SVM	0.39	0.41	0.42	0.31
Hist Gradient Boosting	0.51	0.65	0.52	0.47
Adaptive Boosting	0.35	0.36	0.35	0.20

Признаковое пространство

В качестве признакового пространства использовались элементарные ряды, восстановленные из разложения траекторной матрицы.

Классификаторы

В качестве алгоритмов многоклассовой классификации использовались:

- Полносвязная нейронная сеть
- Gradient Boosting
- SVM
- Random Forest
- Adaptive Boostin
- Hist Gradient Boosting

	accuracy	macro avg precision	macro avg recall	macro avg f1-score
FNN	0.86	0.86	0.86	0.86
Gradient Boosting	0.80	0.80	0.80	0.80
Random Forest	0.76	0.76	0.75	0.75
SVM	0.77	0.79	0.77	0.77
Hist Gradient Boosting	0.81	0.81	0.81	0.81
Adaptive Boosting	0.44	0.50	0.41	0.36

Результаты эксперимента

Была построена нейросеть с тремя входами, один из которых принимал временные ряды, а остальные – предложенные признаковые пространства. Эксперимент показывает, что предлагаемые признаковые пространства повышают устойчивость классификации.

class	precision	recall	f1-score
0	0.94	0.95	0.95
1	0.98	0.90	0.94
2	0.89	0.99	0.94
3	0.89	0.83	0.86
4	0.87	0.89	0.88
5	1.00	1.00	1.00
accuracy			0.93
macro avg	0.93	0.93	0.93
weighted avg	0.93	0.93	0.93

Результат классификации нейронной сети с тремя входами

- Предложены методы построения признаков пространств
- Продемонстрирована работа известных алгоритмов классификации на предложенных признаков пространствах. Качество сравнимо с качеством классификации моделей, получивших широкое применение в подобных задачах.
- Продемонстрировано, что предлагаемые признаки пространства повышают устойчивость классификации
- Ведется сбор данных акселерометрии, которые не вошли в текущий эксперимент
- Планируется:
 - Расширить методы построения признаков пространств
 - Для увеличения качества классификации попробовать ансамблирование классификаторов