

Порождение признаков с помощью локально-аппроксимирующих моделей

Абдурахмон Садиев

Московский физико-технический институт

Курс: Численные методы обучения по прецедентам
(практика, В. В. Стрижов)/Группа 674, весна 2019

Цель работы

Задача

Построить систему локально-аппроксимирующих моделей, которая порождает пространство признаков.

Проблематика задачи

Использование значений сигнала в качестве признакового описания приводит к появлению моделей высокой сложности и низкого качества.

Предлагаемое решение

Требуется построить набор локально-аппроксимирующих моделей и выбрать наиболее адекватные.

Порождение признаков

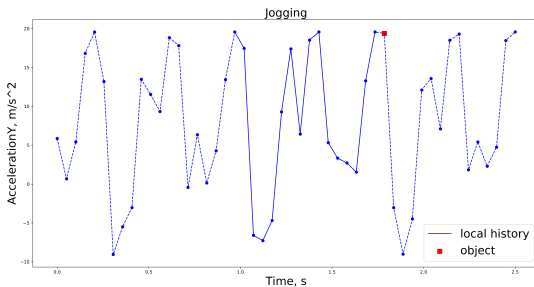
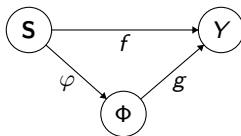


Рис.: Сегментация. Порождение признаков.



$$\hat{f} = g \circ \varphi$$

Постановка задачи

Данные

Пусть задана выборка :

$$\mathcal{D} = \{(\mathbf{s}_i, y_i) \mid i = 1, \dots, m; \mathbf{s}_i = [\mathbf{s}_i(1), \dots, \mathbf{s}_i(T)] \in \mathbf{S} \subset \mathbb{R}^{n \times m}\},$$

где $\mathbf{s}_i(t) \in \mathbb{R}^n$, $y_i \in Y$ - пространство ответов, $|Y| = K \in \mathbb{N}$, m - количество элементов в выборке.

Модель

Модель будет приближать отображение $f : \mathbf{S} \rightarrow Y$ и будем искать ее в виде суперпозиции:

$$\hat{f}(\mathbf{s}) = g(\varphi(\mathbf{s}), \mathbf{w}) \tag{1}$$

где $\varphi : \mathbf{S} \rightarrow \Phi$. $\Phi \subset \mathbb{R}^p$ - пространство признаков, \mathbf{w} - вектор параметров модели.

Локально-аппроксимирующие модели

Модели $\varphi_j \in \mathcal{F}$, где $j \in \{1, \dots, r\}$, а r - количество моделей в наборе \mathcal{F} .
Оптимальный набор моделей находится решением оптимизационной задачи

$$\mathcal{P}_{opt} = \arg \min_{\mathcal{P} \subset \mathcal{F}} \min_{\mathbf{w} \in \mathbb{R}^p} \mathcal{L}[g(\mathcal{P}, \mathbf{w})] \quad (2)$$

- Anastasia Motrenko and Vadim Strijov. Extracting fundamental periods to segment biomedical signals. IEEE J. Biomedical and Health Informatics, 20(6):1466–1476, 2016
- М. Е. Карасиков В. В. Стрижов. Классификация временных рядов в пространстве параметров порождающих моделей. Информ. и ее примен., 10(4):121–131, 2016.
- М. П. Кузнецов Н. П. Ивкин. Алгоритм классификации временных рядов акселерометра по комбинированному признаковому описанию. Машинное обучение и анализ данных., 1(11):1471–1483, 2015.

Нулевой этап

Сегментируем временной ряд s_i . Выбираем, как это сделать.

Первый этап

Для каждого сегмента временной ряда s_i строим признаковое описание. Для этого есть множество разных алгоритмов.

Второй этап

После получения признакового описания сегментов временного ряда решаем задачу многоклассовой классификации.

Вычислительный эксперимент

Выборка

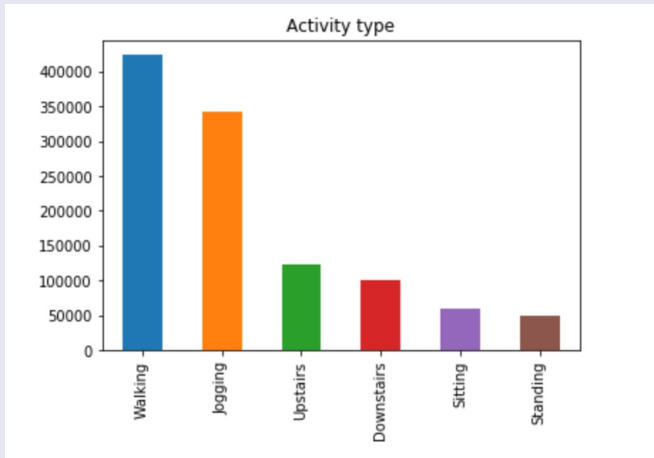


Рис.: Количество измерений для каждого класса

Выборка

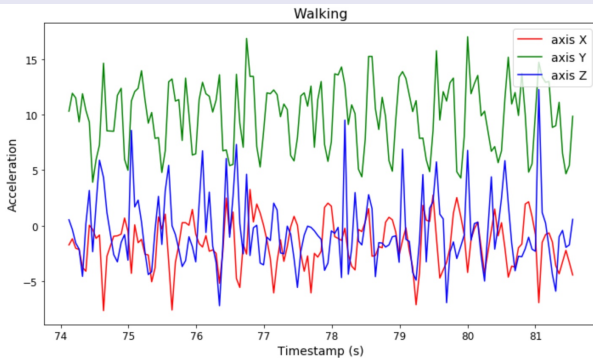


Рис.: Пример ряда из выборки

Вычислительный эксперимент

Результаты

	all	Jogging	Upstairs	Standing	Walking	Downstairs	Sitting
lr_all_feat_sampled_	0.946450	0.993762	0.965094	0.995514	0.970561	0.970071	0.997897
lr_fft_sampled_	0.764071	0.970702	0.880704	0.987734	0.830308	0.866055	0.992640
rf_all_feat_sampled_	0.969090	0.991869	0.976870	0.998528	0.987594	0.985211	0.998108
rf_fft_sampled_	0.848812	0.977641	0.911264	0.995865	0.894372	0.922128	0.996355
svm_all_feat_sampled_	0.964183	0.994883	0.978972	0.995654	0.984229	0.979253	0.995374
svm_fft_sampled_	0.851616	0.978902	0.909161	0.996355	0.903834	0.918343	0.996636

	all	Jogging	Upstairs	Standing	Walking	Downstairs	Sitting
lr_expert_sampled	0.834653	0.981566	0.900610	0.989697	0.895984	0.912385	0.989066
lr_ssa_sampled_20	0.795612	0.986963	0.879933	0.990678	0.846569	0.897175	0.989907
rf_expert_sampled	0.945469	0.990187	0.966426	0.998528	0.966286	0.971473	0.998037
rf_ssa_sampled_20	0.931871	0.986753	0.961940	0.995234	0.957805	0.967057	0.994953
svm_expert_sampled	0.923460	0.989837	0.953739	0.991379	0.961730	0.959136	0.991098
svm_ssa_sampled_20	0.865003	0.991379	0.912666	0.992640	0.906077	0.935025	0.992220

Рис.: Результат работы

Резюме

- Создали стандарт локально-аппроксимирующих моделей.
- Мы научились классифицировать виды деятельности человека по созданному набору признаков.
- Если сегментировать более разумно, то можно получить более высокие результаты.
- Классификации на наборе признаков, созданных разными моделями, имеет высокое качество.