

Порождение признаков с помощью локально-аппроксимирующих моделей

Максим Евгеньевич Христюбов

Московский физико-технический институт

Курс: Численные методы обучения по прецедентам
(практика, В. В. Стрижов)/Группа 774, весна 2020

Консультант: Сайранов Д.

Цель

Решить задачу классификации моментов временного ряда.

Задача

Требуется предложить способ построения признакового пространства и решить задачу классификации на построенном пространстве.

Метод

Суперпозиция моделей локальной аппроксимации и отбор признаков.

Базовая литература

- Кузнецов М.П., Ивкин Н.П., *Алгоритм классификации временных рядов акселерометров по комбинированному описанию признаков*. Машинное обучение и анализ данных, 2015
- Isachenko R.V., Bochkarev A.M., Zharikov I.N., Strijov V.V., *Feature Generation for Physical Activity Classification*. Artificial Intelligence and Decision Making, 2018, 3 : 20-27.
- Petr Somol, Jana, Novovicova, Pavel Pudil, *Efficient Feature Subset Selection and Subset Size Optimization*. 2010.

Данные

Задан исходный временной ряд с метками классов:

$$\mathfrak{D} = \{(d_i, y_i)\}_{i=1}^M,$$

где $y_i \in Y$, $|Y|$ — количество классов. Предполагается, что ряд квазипериодичен, причем известен максимальный период T .

Модель

Модель будет приближать отображение $R : \mathcal{I} \rightarrow Y$, где $\mathcal{I} = \{1, \dots, M\}$ — моменты времени. Будем искать ее в виде суперпозиции:

$$\hat{R}(k) = b(\varphi(h(k)), \mathbf{w}),$$

где $h : \mathcal{I} \rightarrow \mathbb{X}$ — отображение момента времени в сегмент \mathbf{x}_k ,
 $\varphi : \mathbb{X} \rightarrow W$ — процедура построения признаков описания сегмента,
 \mathbf{w} — вектор параметров модели.

Локально-аппроксимирующая модель

Модель $\hat{\mathbf{x}}$ назовем локально-аппроксимирующей моделью, если она приближает временной ряд на локальном сегменте \mathbf{x} :

$$\hat{\mathbf{x}} = \arg \min_{\hat{\mathbf{x}}} \|\hat{\mathbf{x}} - \mathbf{x}\|_2.$$

Признаковое описание сегмента

В качестве признакового описания $\varphi(\mathbf{x})$ сегмента \mathbf{x} используется вектор параметров локально-аппроксимирующей модели $\hat{\mathbf{x}}(\mathbf{v})$:

$$\varphi(\mathbf{x}) = \arg \min_{\mathbf{v}} \|\hat{\mathbf{x}}(\mathbf{v}) - \mathbf{x}\|_2.$$

Первый этап

Классифицируемые моменты времени задают локальные сегменты x_k временного ряда.

Второй этап

Для полученных сегментов \mathbf{x}_k строится признаковое описание $\varphi(x_k)$ с помощью локально-аппроксимирующих моделей.

Третий этап

После получения признакового описания каждого момента времени, решается задача классификации моментов времени ряда.

Цели

Решить задачи:

- 1) классификации типа активности,
- 2) классификации пола человека.

Изучить влияние решения одной задачи классификации на решение другой. Сравнить распределения параметров моделей, построенных для решения этих задач.

Данные

Для эксперимента берется шесть временных рядов в 39000 временных моментов (780 секунд). Данные снимаются с акселерометра, которым пользуются четыре человека (двое мужчин и две женщины), которые выполняют подъем или спуск по лестнице (два типа деятельности).

Кросс-валидация

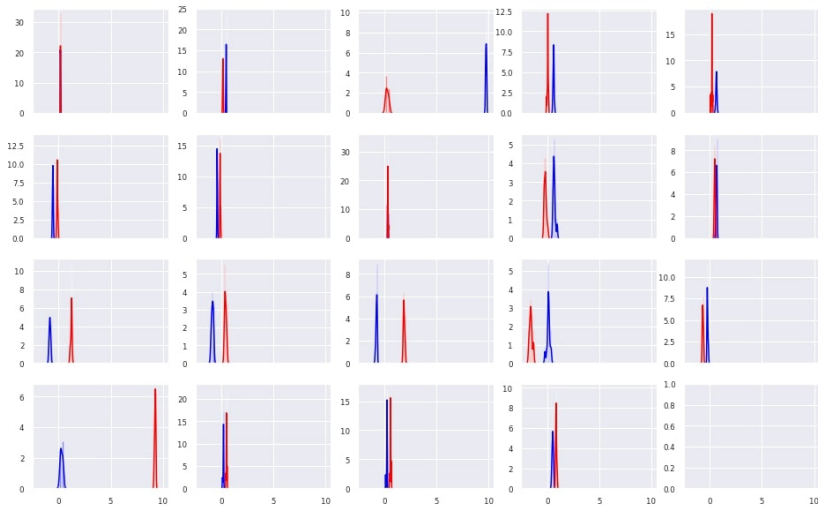
Для оценки обобщающей способности моделей используется k-fold кросс-валидация на $k = 13$ подвыборках.

Среднее и дисперсия доли верных классификаций

	mean(activity)	std(activity)	mean(gender)	std(gender)
LogRegression	0.908	0.147	0.952	0.152
after adding	0.903	0.152	0.950	0.136
SVC	0.934	0.097	0.971	0.074
after adding	0.933	0.101	0.966	0.072

Анализ

Использование результата решения одной задачи классификации существенно не влияет на качество решения другой задачи.



mean(act)	-2.89	-1.37	4.20	1.58	-3.41	-2.86	4.51	4.12	-1.24	9.29	2.35
std(act)	0.33	0.23	0.22	0.21	0.37	0.36	0.39	0.27	0.16	0.55	0.41
mean(act)	2.64	-6.85	-4.02	0.41	1.82	-1.17	1.29	1.26	-0.88	-2.87	
std(act)	0.17	0.54	0.36	0.47	0.24	0.24	0.39	0.44	0.20	0.31	
mean(gen)	1.66	0.41	1.28	-1.62	-1.19	-1.47	1.61	2.17	0.71	1.37	3.70
std(gen)	0.31	0.14	0.31	0.19	0.24	0.23	0.15	0.14	0.11	0.44	0.41
mean(gen)	0.23	-3.98	8.82	-7.39	3.64	-4.78	3.17	3.53	1.36	7.65	
std(gen)	0.18	0.50	0.20	0.33	0.23	0.26	0.18	0.17	0.09	0.17	

Расстояние Кульбака-Лейблера

Расстояние Кульбака-Лейблера между распределениями параметров модели классификации активности и модели классификации пола оказалось равным $KL = 56.20$.

- Рассмотрен способ описания момента времени по локальному сегменту этого момента времени.
- Произведена классификация типа активности и пола человека.
- Сравнили распределения параметров моделей для решения этих задач классификации.

Вывод

Распределения параметров моделей классификации типа активности и пола значительно отличаются.