

Выбор оптимальных моделей локальной аппроксимации для классификации временных рядов

Сергей Дмитриевич Иванычев

Московский физико-технический институт
Физтех-школа прикладной математики и информатики
Факультет управления и прикладной математики
Кафедра «Интеллектуальные системы»

Научный руководитель: д.ф.-м.н. В.В. Стрижов

Выпускная квалификационная работа бакалавра

Москва 2018

Классификация временных рядов

Цель

Предложить способ построения набора локально аппроксимирующих моделей для устойчивой классификации сигналов носимых устройств.

Гипотеза

Суперпозиция локально аппроксимирующих моделей доставляет более высокое качество при меньшей сложности чем универсальные модели.

Прямая задача

Требуется выбрать такой набор моделей локальной аппроксимации, что порождающая выборка в промежуточном пространстве является *простой*.

Обратная задача

Оптимизировать структурные параметры выбираемых моделей по порождающей выборке с целью получения выборки с оптимальными свойствами.

- Кузнецов М. П., Ивкин Н. П., *Алгоритм классификации временных рядов акселерометра по комбинированному признаковому описанию*, 2015.
- Карасиков М. Е., Стрижов В. В. *Классификация временных рядов в пространстве параметров порождающих моделей*, 2016.
- Артемов А. В., *Математические модели временных рядов с трендом в задачах обнаружения разладки*, 2016.

Постановка задачи классификации

Временной ряд

$$S : T \rightarrow \mathbb{R} \text{ где } T = \{t_0, t_0 + d, t_0 + 2d \dots\}.$$

Сегмент временного ряда

$$\mathbf{x}_i = (S(t_i), S(t_i - d), S(t_i - 2d), \dots, S(t_i - (n-1)d)), \quad \mathbf{x}_i \in X \equiv \mathbb{R}^n.$$

X — набор сегментов данных акселерометра

y — метки классов движения (бег, ходьба, подъем и спуск по лестнице)

h — конечный набор моделей локальной аппроксимации.

Постановка задачи классификации

Локально аппроксимирующая модель

$$g_i(w, x) \in X, \text{ где } w \in \mathbb{R}^{n_g}.$$

Оптимальные параметры определяются образом

$$\mathbf{h}_i(x) = \arg \min_{w \in \mathbb{R}^{n_g}} \rho(g(w, x), x).$$

\mathbf{h}_i — локально аппроксимирующая модель сегмента.

$\mathbf{h} = [\mathbf{h}_1 \dots \mathbf{h}_k] : x \mapsto [w_1^* \dots w_k^*]$ отображает пространство сегментов \mathbf{X} в *промежуточное пространство* признаков описаний \mathbf{Z} .

Алгоритм классификации

$$T \rightarrow \mathbf{X} \xrightarrow{\mathbf{h}} \mathbf{Z} \xrightarrow{a} Y$$

Где \mathbf{h} набор моделей локальной аппроксимации, $a(\cdot, \gamma)$ — алгоритм многоклассовой классификации.

Локально-аппроксимирующие модели

Модель	Структурные параметры
SEMOR	-
AR-авторегрессия	порядок
Фурье-модель FFT	количество главных частот
Вейвлет-модель SSE	количество сингулярных чисел

AR-авторегрессия

Структурный параметр: порядок m .

$$g_{\text{AR}}(w, x) = \hat{x}, \text{ где } \hat{x}_i = \begin{cases} x_k & \text{при } k \in [1, m] \\ w_0 + \sum_{i=1}^m w_i x_{k-i} & \text{при } k \in [m+1, n] \end{cases}$$

Фурье-модель (SSA)

Структурный параметр: количество главных собственных значений k . Сингулярное разложение траекторной матрицы:

$$S^T S = V H V^T, H = \text{diag}(\lambda_1 \dots \lambda_m)$$

w образуют k старших собственных значений.

Вейвлет-модель (FFT)

Структурный параметр: k частот из прямого преобразования Фурье, соответствующие наибольшим амплитудам

$$w_{2j} = \operatorname{Re} \sum_{k=1}^n x_k \exp\left(-\frac{2\pi i}{n} kj\right), \quad w_{2j+1} = \operatorname{Im} \sum_{k=1}^n x_k \exp\left(-\frac{2\pi i}{n} kj\right)$$

Self-Modeling Regression

$$g(x, w) = w_1 + w_2 p(w_3 + w_4 t)$$

$$w_{\text{SEMOR}} = [\hat{w}_1, \hat{w}_2, \hat{w}_3, \hat{w}_4, \rho]$$

Алгоритм 5.1 Алгоритм обучения

Вход: $\mathbf{X}, \mathbf{Y}, \mathbf{h}, l, l_h$;

Выход: $\hat{\theta}$;

1: инициализировать $\mathbf{u} := \mathbf{y}_1$ (вектор матрицы \mathbf{Y})

2: $\mathbf{z} = []$

3: для $i = 1, \dots, l_h$

$\mathbf{z}[i] := \mathbf{h}[i](\mathbf{X})$

4: инициализировать $\mathbf{Z} = \mathbf{z}[1]$

5: для $i = 2, \dots, l_h$

$\mathbf{Z} := (\mathbf{Z}, \mathbf{z}[i])$

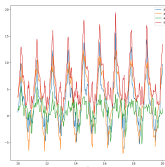
6: $\hat{\theta} = \arg \min(L(f(\mathbf{Z}), \mathbf{Y}))$

Решение задачи: генерация данных

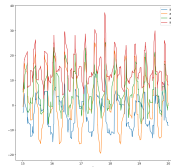
Данные с акселерометра: 4 типа движения, частота дискретизации 100 Гц.

Сегментация: локальные экстремумы с окном + квантиль по длине сегментов.

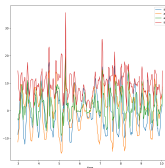
Нормализация: приведение к одной размерности с помощью кубических сплайнов.



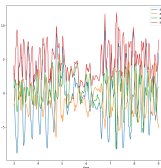
(a) Ходьба



(b) Бег



(c) Вверх по лестнице



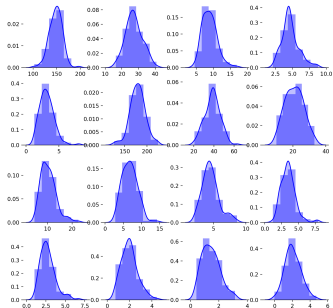
(d) Вниз по лестнице

Решение задачи: проверка простоты выборки

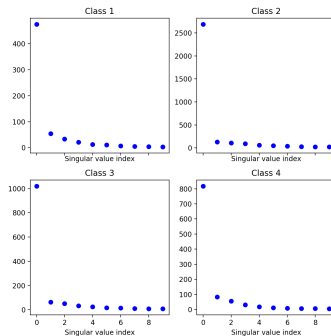
Данные: сегменты временного ряда акселерометра.

Тесты простоты выборки: (T -тест) $\mathbb{E}\varepsilon = 0$, $D\varepsilon = \text{const}$, а также

Анализ унимодальности распределений



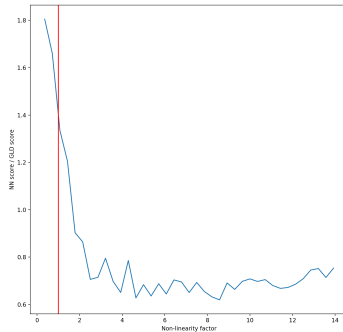
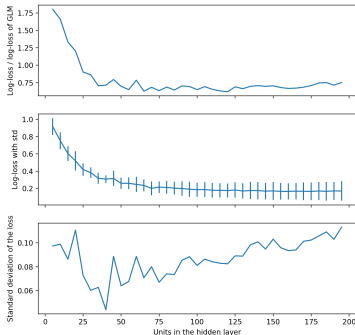
Анализ спектра выборки



Решение задачи: order of non-linearity

GLM: Отбор признаков из промежуточного пространства + GLD vs. нейронная сеть.

Non-linearity factor: Отношение сложности универсальной модели и сложности GLM.



- Выборка в промежуточном пространстве простая, а аппроксимирующие ее линейная модель являются адекватной.
- GLM адекватнее разделяет выборку чем универсальная модель, то есть при одинаковой сложности обеспечивает более высокое качество и меньше переобучается.