

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

Задача

Сегментация и классификация временного ряда, генерируемого линейным акселерометром носимого устройства, по типам движения.

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
- Карасикова М. Е., Стрижов В. В. Классификация временных рядов в пространстве параметров порождающих моделей, 2016

Определение: Временной ряд

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

Определение: Сегмент временного ряда

При заданной ширине сегмента и метке времени это вектор

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

Определение: Локально аппроксимирующая модель

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

Тогда оптимальные параметры будут являться 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.

Постановка задачи

Дано

\mathbf{X} — набор сегментов данных акселерометра

\mathbf{y} — метки классов движения (бег, ходьба, подъем и спуск по лестнице)

\mathbf{h} — конечный набор моделей локальной аппроксимации.

\mathbf{h} отображает пространство сегментов \mathbf{X} в *промежуточное пространство* признаков описаний \mathbf{Z} .

Алгоритм классификации

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

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

Локально-аппроксимирующие модели

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

Выбранные модели порождают промежуточное пространство \mathbf{Z}

$$[\mathbf{h}_1 \dots \mathbf{h}_k] : x \mapsto [w_1^* \dots w_k^*]$$

Структурный параметр: порядок m .

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

Оптимальные веса w минимизацией MSE.

$$w_{\text{AR}} = \arg \min_{w \in \mathbb{R}^m} \sum_{i=1}^n \|x_i - \hat{x}_i\|^2$$

Структурный параметр: количество главных собственных значений k . Траекторная матрица:

$$S = \begin{pmatrix} x_1 & x_2 & \dots & x_m \\ x_2 & x_3 & \dots & x_{m+1} \\ \vdots & \vdots & \ddots & \vdots \\ x_{n-m+1} & x_{n-m+2} & \dots & x_n \end{pmatrix}$$

Сингулярное разложение:

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

m образуют k старших собственных значения.

Модели локальной аппроксимации: Вейвлет-модель (FFT)

Структурный параметр: количество главных частот k .
Дискретное преобразование Фурье.

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

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

Признакове описанием сегмента — k частот из прямого преобразования Фурье, соответствующие наибольшим амплитудам.

Модели локальной аппроксимации: SEMOR (Self-Modeling Regression)

Модель Self-Modeling Regression описывается следующим выражением

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

- Параметры w_1, w_2 находятся шагом двухпараметрической линейной регрессии
- w_3, w_4 получаются минимизацией DTW

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