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

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 выьборке с целью получения выборки с оптимальными свойствами.

Литература

- 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

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

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* признаковых описаний $\mathbf{Z}$.

Алгоритм классификации

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

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

Построение промежуточного пространства

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.

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

Фурье-модель (SSA)

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

$$S^\mathsf{T} S = VHV^\mathsf{T}, H = \mathrm{diag}(\lambda_1 \ldots \lambda_m)$$

w образуют k старших собственных значения.

Модели локальной аппроксимации: Фурье-модель (SSA)

Вейвлет-модель (FFT)

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

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

Self-Modeling Regression

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

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