Выбор ансамбля локальных моделей в задачах анализа сигналов

Сайранов Данил Айдарович

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

МФТИ, 2020

Классификация временных рядов

Цель

Предложить способ построения ансамбля моделей локальной аппроксимации для классификации сигналов носимых устройств.

Гипотеза

Ансамбль моделей локальной аппроксимации предпочтительнее в парето-оптимальном смысле универсальной модели (нейросети): точнее, устойчивее, проще.

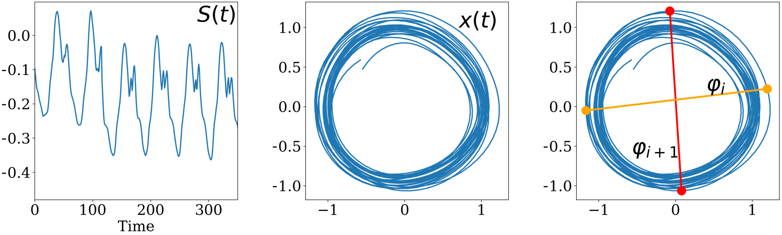
Задача

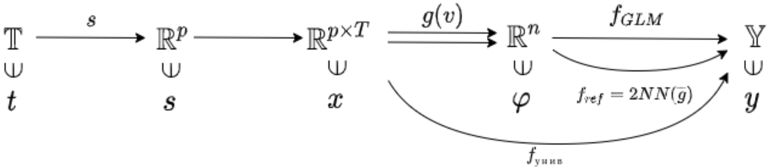
Требуется построить признаковое описание временных рядов на используя параметры моделей локальной аппроксимации.

Метод

Предложить критерий сложности ансамбля для выбора оптимального признакового описания.

Классификация временных рядов





Литература

- **Кузнецов М. П., Ивкин Н. П.**, *Алгоритм классификации временных рядов акселерометра по комбинированному признаковому описанию*, Машинное обучение и анализ данных, 2015.
- **Карасиков М. Е., Стрижов В. В.**, *Классификация временных рядов в пространстве параметров порождающих моделей*, Информатика и её применения, 2016.
- **Иванычев С. Д.**, *Выбор оптимальных моделей локальной аппроксимации для классификации временных рядов*, Выпускная квалификационная работа бакалавра, 2018.
- **A. M. Katrutsa, V. V. Strijov**, *Comprehensive study of feature selection methods to solve multicollinearity problem according to evaluation criteria*, Expert Systems with Applications, 2016
- **Cybenko, G. V.**, *Approximation by Superpositions of a Sigmoidal function*, Mathematics of Control Signals and Systems, 1989.

Постановка задачи

Временной ряд:

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

Сегмент временного ряда:

$$\boldsymbol{x}_i = [S(t_i), S(t_i - d), \dots, S(t_i - (n-1)d))], \boldsymbol{x}_i \in \boldsymbol{X} \equiv \mathbb{R}^n$$

Выборка: $\mathfrak{D} = \{(\boldsymbol{x}_i, y_i)\}_{i=1}^{N}, \;\; y_i \in \{1, \dots, K\}$
$\boldsymbol{X}$ — набор сегментов данных акселерометра,
y — метки классов движения.

Универсальная модель

Многослойная полносвязная нейронная сеть

Модели локальной аппроксимации

Моделью локальной аппроксимации g называется модель, аппроксимирующая временной ряд $x(t)$ на отрезке времени $[t, t - \Delta t]$:

$$g : [t, t - \Delta t] \to \hat{\mathbf{x}}.$$

Параметры $\mathbf{w}$ модели $g = g(\mathbf{w})$ оптимизируются в соответствии функции ошибки:

$$\hat{\mathbf{w}} = \arg\min S(\mathbf{w}) = \|\hat{\mathbf{x}} - \mathbf{x}\|.$$

Признаковое описание φ генерируется методом локальной аппроксимации для каждой временной точки t:

$$\varphi = \varphi(\mathbf{g}, \mathbf{w}, \mathbf{x}).$$

Модели локальной аппроксимации

Метод

Оптимальные параметры $\boldsymbol{g}_i(\boldsymbol{x})$ определяются как

$$\boldsymbol{g}_i(\boldsymbol{x}) = \arg\min_{\boldsymbol{w}\in\mathbb{R}^{n_g}} (g(\boldsymbol{w}, \boldsymbol{x}), \boldsymbol{x})$$

$\boldsymbol{g}_i$ - модель локальной аппроксимации.

Пространство сегментов $\boldsymbol{X}$ отображается набором функций $\boldsymbol{g} = [\boldsymbol{g}_1, ..., \boldsymbol{g}_{\boldsymbol{k}}]$ в пространство признаковых описаний Φ.

Модели локальной аппроксимации

Модель	**Структурные параметры**
AR-авторегрессия	Порядок
ARMA	Порядок
ARIMA	Порядок
Фурье-модель (FFT)	Кол-во главных частот
Сингулярный спектр (SSA)	Кол-во сингулярных чисел
Self-Modelling Regression (SEMOR)	-

Модели локальной аппроксимации

AR-авторегрессия

Структурный параметр: порядок m,

$$x_t = w_0 + \sum_{i=1}^{m} w_i x_{t-i},$$

где w_0 — константа, $w_1, \ldots, w_m$ — параметры модели.

ARMA

Структурные параметры: порядки m и n,

$$x_t = w_0 + \sum_{i=1}^{m} w_i x_{t-i} + \sum_{i=1}^{n} \theta_i \varepsilon_{t-i},$$

где w_0 — константа, $w_1, \ldots, w_m, \theta_1, \ldots, \theta_n$ — параметры модели, $\varepsilon_{t-1}, \ldots, \varepsilon_{t-n}$ - ошибки.

Модели локальной аппроксимации

Фурье-модель (FFT)

Структурный параметр: количество главных частот из прямого преобразования Фурье,

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

Сингулярный спектр (SSA)

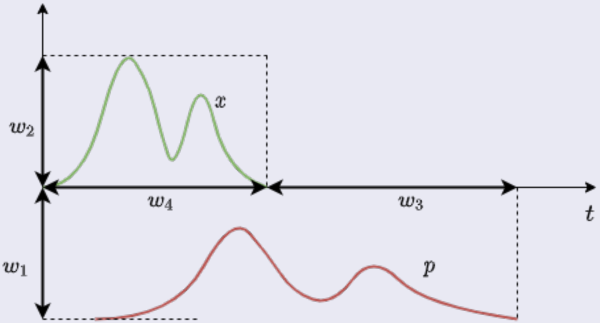
Структурный параметр: количество главных собственных значений.

$$S^\top S = VHV^\top, H = diag(\lambda_1, ..., \lambda_m)$$

Self-Modelling Regression (SEMOR)

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

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



Построение промежуточной выборки. Отбор признаков

Промежуточная выборка

- Для модели локальной аппроксимации $\mathbf{g}_i \in \mathbf{g}$ вычисляем

$$[\varphi_i^1, \dots, \varphi_i^k]^\top = \varphi(\mathbf{g}_i, \mathbf{w}, \mathbf{x})$$

- Конкатенируем векторы параметров φ_i набора моделей.

Отбор признаков

Признаки отбираются с помощью метода QPFS.
Пусть $\mathcal{J} = \{1, \dots, n\}$ — множество индексов признаков, $\mathcal{A} \subseteq \mathcal{J}$ — его подмножество. Решается задача поиска оптимального подмножества признаков:

$$\mathcal{A}^* = \arg\min_{\mathcal{A} \subseteq \mathcal{J}} Q(\mathcal{A} | \mathbf{X}, \mathbf{y}),$$

где Q — критерий качества.

Сложность модели

Теорема Цыбенко

Пусть $\sigma(\xi)$ любая непрерывная сигмоидная функция, например, $\sigma(\xi) = 1/(1 + e^{-\xi})$. Тогда, если дана любая непрерывная функция действительных переменных f на $I_n = [0, 1]^n$ и $\varepsilon > 0$, тогда существуют векторы $\boldsymbol{w}$ и параметризованная фукнция $G(\boldsymbol{x}, \sigma, \boldsymbol{w}) : I_n \to \mathbb{Y}$ такая, что $|G(\boldsymbol{x}, \boldsymbol{w}) - f(\boldsymbol{x})| < \varepsilon$ для всех $\boldsymbol{x} \in I_n$, где $\boldsymbol{w}_i \in \mathbb{R}^n$, $G = \sum\limits_i \sigma(\boldsymbol{w}_i^\top \boldsymbol{x})$.

Определим сложность модели, как минимальное число нейронов на скрытом слое полносвязной двухслойной нейросети, которое позволяет аппроксимировать данные с заданным качеством L^*:

$$Comp_{L^*}(\mathfrak{D}) = \min\left(N | L = L^*\right)$$

Эксперимент

Данные

Данные сняты с акселерометра.
4 типа движения: ходьба, бег, подъем по лестнице, спуск по лестнице
Частота дискретизации: 100 Гц.

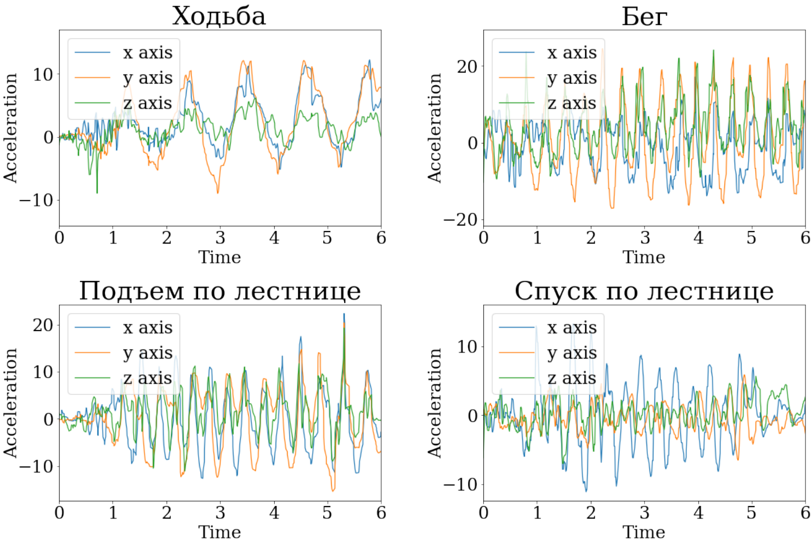
Сегментация

Локальные экстремумы с окном и ограничения на длину сегментов

Нормализация

Данные приведены к одной размерности с помощью кубических сплайнов

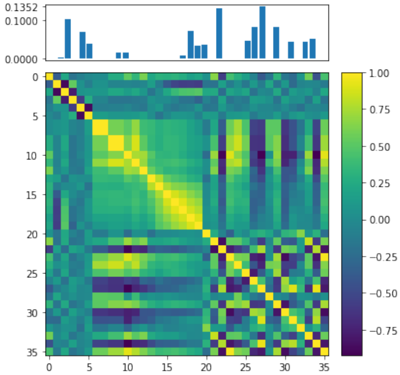
Эксперимент



Эксперимент. Отбор признаков

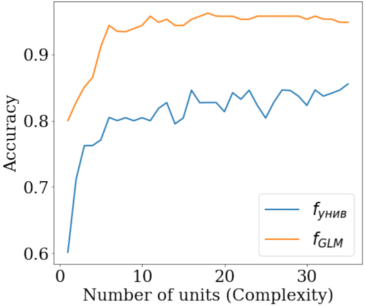
Отбор признаков производится с помощью метода QPFS:

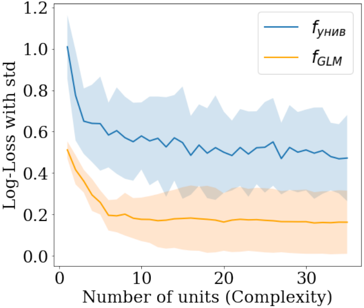
- Признаки сортируются по их значимости.
- Выбирается k наиболее значимый признаков.
- Оценивается качество обобщенной линейной модели (GLM) на выделенных признаках.



Сверху: Нормализованная значимость признаков.
Снизу: Матрица корреляции признаков.

Параметры двухслойной нейронной сети оптимизируются на выборке $(\boldsymbol{X}, \boldsymbol{y})$.
Параметры GLM — на выборке $(\varphi, \boldsymbol{y})$.

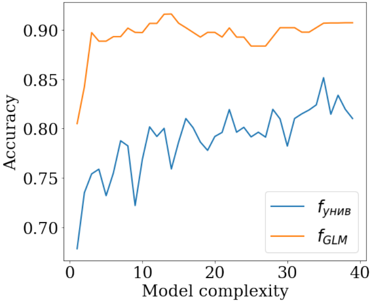


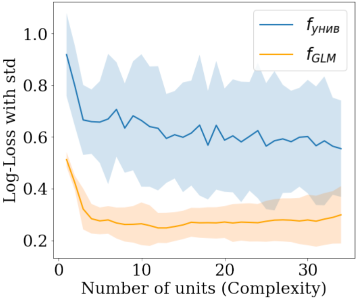


Эксперимент. Сравнение моделей на зашумленных данных

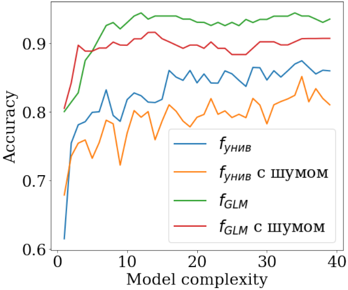
На выборку $(\boldsymbol{X}, \boldsymbol{y})$ накладывается гауссовский шум:

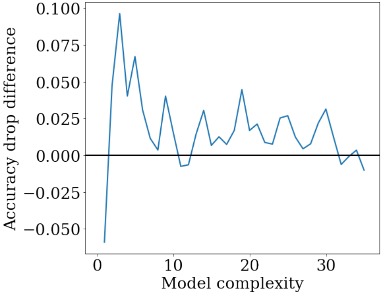
$$\widetilde{\boldsymbol{X}} = \boldsymbol{X} + \varepsilon, \varepsilon \in \mathcal{N}(0, 1)^{n \times N}.$$





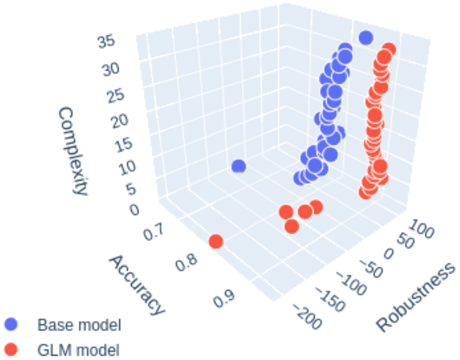
Проанализируем падение качества моделей при зашумлении данных. Как видно из графиков, модель GLM устойчивее универсальной модели.





Эксперимент. Сравнение моделей

Полученные результаты изобразим в трехмерном пространстве «Точность-Устойчивость-Сложность».
Полученные данные демонстрируют, что обобщенно-линейная модель обученная на промежуточных признаках предпочтительнее универсальной.



На защиту выносятся

- Метод построения ансамбля локально-аппроксимирующих моделей.
- Алгоритм выбора признаков в ансамбле локально-аппроксимирующих моделей с помощью дискретного квадратичного программирования.
- Критерий сложности модели классификация.
- Программная реализация (библиотека локального моделирования)

Ссылки

- Проект на github:
  https://github.com/livenrlty/Local-Approximation-Models