

Оптимальное порождение моделей локальной аппроксимации в задаче классификации временных рядов

Терехов Олег Игоревич

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 2019

Классификация временных рядов

Цель

Предложить метод выбора моделей локальной аппроксимации, порождающих простую выборку, для решения задачи многоклассовой классификации временных рядов.

Задача

Требуется оптимизировать структурные параметры выбираемых моделей по порождаемой выборке с целью получения выборки с оптимальными свойствами.

Метод

Сравнение сложности универсальной модели в виде двухслойной нейронной сети и базовой модели с использованием критерия нелинейности выборки.

- 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.
- Иванов С. Д., *Выбор оптимальных моделей локальной аппроксимации для классификации временных рядов*, Выпускная квалификационная работа бакалавра, 2018.

Постановка задачи классификации

Задан временной ряд

$\mathbf{s} : \mathbb{T} \rightarrow \mathbb{R}^p$, где $\mathbb{T} = \{0, \dots, T\}$ — моменты времени,
 p — количество сенсоров.

Требуется найти отображение

$f : t \rightarrow \hat{y}$, где $\mathbb{Y} = \{1, 2, \dots, K\}$ — множество меток классов.

Определен сегмент временного ряда

$\mathbf{x}_t = \text{vec}(\mathbf{s}_t, \dots, \mathbf{s}_{t-T})$.

Задана выборка

$\mathcal{D} = \{(\mathbf{x}_i, y_i, t_i)\}_{i=1}^I$,

$\mathbf{x} \in \mathbb{R}^{p \times T}$ — точка фазового пространство сегментов данных с сенсоров, $y_i \in \{1, 2, \dots, K\}$ — метки классов движения.

Локальная аппроксимация

Определение

Моделью локальной аппроксимации называется модель, аппроксимирующая временной ряд $\mathbf{x}(t)$ в промежутке времени $[t, t - T]$:

$$g : [t, t - T] \rightarrow \hat{\mathbf{x}}.$$

Или поэлементно $g : \tau \mapsto \mathbf{s}$.

Метод

Оптимальные параметры $\hat{\mathbf{v}}$ модели локальной аппроксимации $g = g(\mathbf{v})$ определяются как $\hat{\mathbf{v}}(\mathbf{x}) = \arg \min_{\mathbf{v}} S(\mathbf{v}) = \|\mathbf{x} - \hat{\mathbf{x}}_g\|$, где $S(\mathbf{v})$ — функция потерь. Порожденные признаки \mathbf{z} — функция от решения задачи аппроксимации.

Задан конечный набор моделей локальной аппроксимации \mathfrak{G} :

$g \in \{g_1, \dots, g_k\}$, где $g_j(\mathbf{v}, \mathbf{x}) \in \mathfrak{G} : [t, t - T] \mapsto \hat{\mathbf{x}}$

Структурные параметры

Модель	Структурные параметры
Self-Modeling Regression (SEMOR)	-
Сплайны	порядок
Модель авторегрессии (AR)	порядок
Фурье-модель (FFT)	кол-во главных частот
Сингулярный спектр (SSA)	кол-во сингулярных чисел

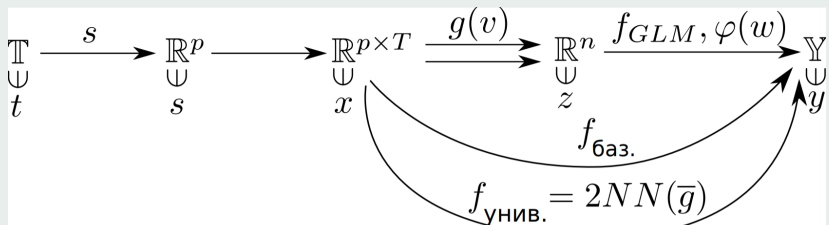
Классификаторы

Модель "Цыбенко"

В качестве универсальной модели многоклассовой классификации будем использовать двухслойную нейронную сеть, в которой последний слой представляет собой обобщенную линейную модель.

Базовая модель

Многослойная полносвязная нейронная сеть.



Модель классификации

$$\mathbf{T} \rightarrow \mathbb{R}^p \rightarrow \mathbb{R}^{p \times T} \xrightarrow{\mathbf{g}} \mathbb{R}^n \xrightarrow{\phi} \mathbb{Y},$$

\mathbf{g} — набор моделей локальной аппроксимации,
 $\phi(\cdot, \gamma)$ — многоклассовый классификатор.

- 1 Наращиваем сложность модели "Цыбенко" до достижения заданной точности аппроксимации S^* , полученной с помощью базовой модели.
- 2 Выбираем лучшую по сложности модель.

Определение структуры модели "Цыбенко"

Пусть $\mathcal{A} \in \mathcal{J}$ — структура модели "Цыбенко": $\mathcal{A} = \{\mathbf{z}_i\}$, где $i \in \mathcal{I} \subset \mathcal{J}$, \mathcal{J} — подмножество индексов выбранных на текущей итерации моделей локальной аппроксимации.

Оптимальная структура $\mathcal{A}^* = \arg \min_{\mathcal{A} \in \mathcal{J}} \text{Comp}(\mathcal{A})$

Th. 1989, 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 $\epsilon > 0$, тогда существуют векторы \mathbf{w} и параметризована функция $G(\mathbf{x}, \sigma, \mathbf{w}) : I_n \rightarrow \mathbb{Y}$ такая, что $|G(\mathbf{x}, \mathbf{w}) - f(\mathbf{x})| < |\epsilon|$ для всех $\mathbf{x} \in I_n$, где $\mathbf{w}_i \in \mathbb{R}^n$, $G = \sum_i \sigma(\mathbf{w}_i^T \mathbf{x})$.

Критерий нелинейности

Введем обозначение $Comp(\mu)$ критерия нелинейности, определяемого как количество нейронов скрытого слоя модели "Цыбенко".

Оптимизация функции потерь:

- 1 Для каждой модели $g_i \in \mathfrak{G}$ вычисляем

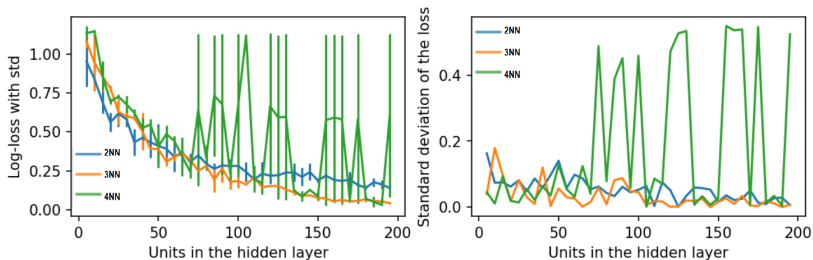
$$[\mathbf{z}_i^1 \dots \mathbf{z}_i^k]^\top = [g_1(\mathbf{x}_i) \dots g_k(\mathbf{x}_i)]$$

- 2 Конкатенируем вектора параметров $\mathbf{z}_i = (\mathbf{z}_i^1 \dots \mathbf{z}_i^k)$, то есть $\mathbf{z}_i = \mathbf{g}(\mathbf{x}_i)$. Получили выборку в промежуточном пространстве \mathbb{R}^n .
- 3 Минимизируем функции потерь обобщенной линейной модели

$$\hat{\theta} = \arg \min_{\theta \in \Theta} L(f(\mathbf{Z}), \mathbf{y}).$$

Оптимальная базовая модель

На выборке (\mathbf{X}, \mathbf{y}) оптимизируем параметры нейронной сети с 1, 2 и 3-мя скрытыми слоями (2NN, 3NN, 4NN). Выбираем в качестве базовой модели двухслойную нейронную сеть. Получаем высокую точность при высокой сложности модели.



Исходные данные

Набор сегментированных временных рядов, соответствующих 4 классам: бег, ходьба, подъем и спуск по лестнице.

Результаты

Предлагается перейти от анализа ансамбля моделей к оценке критерия нелинейности каждой модели в отдельности.

g \ iter	0	1	2	3	4	5	6	7	8	9	10
ar_4	2	5	5	2	3	4	4	2	2	5	3
fft_5	4	1	1	1	3	4	4	3	3	4	2
semor_down	3	3	3	3	1	1	1	2	1	0	1
semor_run	1	3	3	3	2	3	3	5	2	3	3
semor_up	2	1	1	3	3	1	1	2	2	3	3
semor_walk	3	1	1	4	2	2	2	3	3	2	4
ssa5	3	3	3	4	2	3	3	2	2	2	1
comp	139	132	126	126	124	124	102	102	97	93	90

- Предложен критерий нелинейности выборки и основанный на данном критерии метод выбора локально-аппроксимирующих моделей, порождающих простую выборку.
- Проведены эксперименты для нахождения способа построения универсальной модели (модели "Цыбенко").
- Для исследуемых классов временных рядов выбраны оптимальные модели локальной аппроксимации, а также выбрана оптимальная структура универсальной модели.
- Планируется:
 - расширить библиотеку локальных моделей аппроксимации
 - сравнить предложенный в работе алгоритм с другими алгоритмами оптимизации структуры модели "Цыбенко".