

МИНИСТЕРСТВО ОБРАЗОВАНИЯ И НАУКИ РОССИЙСКОЙ ФЕДЕРАЦИИ
МОСКОВСКИЙ ФИЗИКО-ТЕХНИЧЕСКИЙ ИНСТИТУТ (государственный университет)
ФИЗТЕХ-ШКОЛА ПРИКЛАДНОЙ МАТЕМАТИКИ И ИНФОРМАТИКИ
КАФЕДРА «ИНТЕЛЛЕКТУАЛЬНЫЕ СИСТЕМЫ»

Сайранов Данил Айдарович

Выбор ансамбля локальных моделей в задачах анализа сигналов

03.04.01 — Прикладные математика и физика

ВЫПУСКНАЯ КВАЛИФИКАЦИОННАЯ РАБОТА МАГИСТРА

Научный руководитель:

д. ф.-м. н. Стрижов Вадим
Викторович

Москва

2020

Оглавление

Введение	4
Обзор литературы	6
1. Постановка задачи классификации временных рядов	7
2. Построение признакового пространства на основе моделей локальной аппроксимации	10
3. Отбор признаков и классификация объектов промежуточного пространства	14
4. Генерация обучающей выборки	16
5. Сравнение универсальной и обобщенно-линейной моделей	19
Заключение	25
Список литературы	25

Аннотация

В работе рассматривается многоклассовая классификация временных рядов акселерометра мобильного устройства. Осуществляется классификация признакового описания временных рядов, построенного с помощью моделей локальной аппроксимации. Предлагается построение ансамбля локально-аппроксимирующих моделей, параметры которых являются признаковым описанием временных рядов. Построен метод отбора признаков рядов на основе их коллинеарности. Показано, что классификация, основанная на локальных моделях, устойчивее, точнее и проще в парето-оптимальном смысле, чем универсальная нейросетевая модель.

Ключевые слова: *прогнозирование временных рядов, модели локальной аппроксимации.*

Введение

В работе рассматривается задача классификации временных рядов в задаче распознавания активности человека по сигналам, регистрируемым датчиками носимых устройств: телефоны, умные браслеты. В качестве датчиков могут выступать, например, акселерометр, гироскоп или альтиметр. На данный момент основными подходами к классификации временных рядов являются: машины опорных векторов, рекуррентные или глубокие нейронные сети, решающие деревья.

Помимо них, в силу широкой распространенности задачи классификации временных рядов, существует множество других подходов и техник ее решения. Так же, стоит отметить, что классификация временных рядов — частный случай классификации объектов сложной структуры.

В данной работе основное внимание уделено решению задачи классификации путем построения промежуточного признакового пространства, которое было бы проще нейросетевого подхода, устойчивее, а также обеспечивало приемлемое качество. Такой метод применим не только в задаче распознавания активности человека, но и в других задачах классификации временных рядов, так как к объектам сложной структуры можно отнести и другие временные ряды, используемые в других задачах.

Основная идея в разделении решения на три этапа:

1. На первом этапе из временных рядов извлекаются сегменты, описывающие один период какого-либо действия. После этого

каждому сегменту ставится в соответствие некоторый вещественный вектор, компонентами которого являются некоторые статистики, экспертные оценки или параметры некоторых моделей. В результате чего каждый объект будет описываться некоторым набором численных показателей, то есть будет построено отображение сегментов в промежуточное признаковое пространство

2. Так как некоторые признаки могут быть избыточными и скоррелированными между собой, то это может привести к неустойчивости решения дальнейшей задачи классификации. Поэтому вторым этапом необходимо отфильтровать скоррелированные признаки, которые не несут в себе полезной информации, тем самым уменьшив сложность модели.
3. Полученное признаковое описание данных формирует новую выборку, на которой решается задача классификации активности.

Обзор литературы

В работах [1, 2, 3] рассмотрены алгоритмы классификации временных рядов на основе их признакового описания и отмечены некоторые базовые подходы к генерации признаковых описаний в задаче распознавания активности человека.

В [4, 5] рассмотрены подходы к классификации временных рядов с помощью машины опорных векторов.

Для решения задачи может быть задействован нейросетевой подход. Так, в [6, 7] используются рекуррентные и глубокие сверточные нейронные сети.

1. Постановка задачи классификации временных рядов

Пусть имеется некоторый временной ряд S , являющийся отображением из множества временных меток T во множество действительных чисел, то есть:

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

Сегментом \mathbf{x}_i длины $n \in \mathbb{N}$ временного ряда S будем называть набор

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

Рассматривается задача классификации, то есть восстановление отображения

$$f : \mathbf{X} \rightarrow \mathbb{Y},$$

где $\mathbb{Y} = \{1, 2, \dots, K\}$ — множество меток классов движения.

Задана выборка $\mathfrak{D} = \{(\mathbf{x}_i, y_i)\}_{i=1}^N$. Задача заключается в нахождении функции f , минимизирующей суммарные потери на выборке \mathfrak{D} при заданной функции потерь

$$\mathcal{L}(f(\mathbf{x}_i), y_i) = \sum_{k=1}^K [y_i = k] \log P(y_i = k | \mathbf{x}_i, \theta)$$

Таким образом решается задача оптимизации

$$\hat{\theta} = \arg \min_{\theta \in \Theta} \sum_{i=1}^N \mathcal{L}(f(\mathbf{x}_i), y_i)$$

1.1. Промежуточное признаковое пространство

Пусть G — множество функций вида $g : X \rightarrow \mathbb{R}^{n_g}$. Иными словами, это множество отображений, которые каждому объекту сложной структуры ставят в соответствие вектор действительных чисел. Примером множества G могут быть множество статистик, экспертных оценок, а также множество моделей локальной аппроксимации.

Пусть имеется конечное подмножество локально-аппроксимирующих моделей $\{g_1, \dots, g_k\} \in G$, и пусть каждая функция g_i является отображением в пространство $\mathbb{R}^{n_{g_i}}$. Тогда строится отображение $g : X \rightarrow \Phi$, $\Phi \equiv \mathbb{R}^{n_g}$, где

$$n_g = \sum_{i=1}^k \dim(\text{Im}(g_i)) = \sum_{i=1}^k n_{g_i}.$$

Таким образом Φ — промежуточное признаковое пространство объектов сложной структуры. Заметим, что отображение строится таким образом, что первые n_{g_1} компонент признакового описания соответствуют образу отображения g_1 , следующие n_{g_2} соответствуют g_2 и так далее.

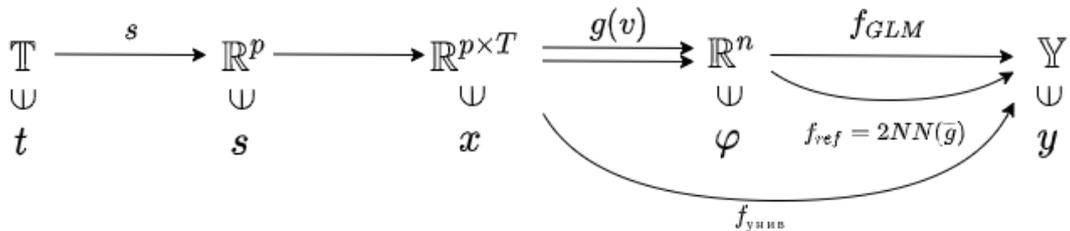


Рис. 1. Схема решения задачи классификации временных рядов

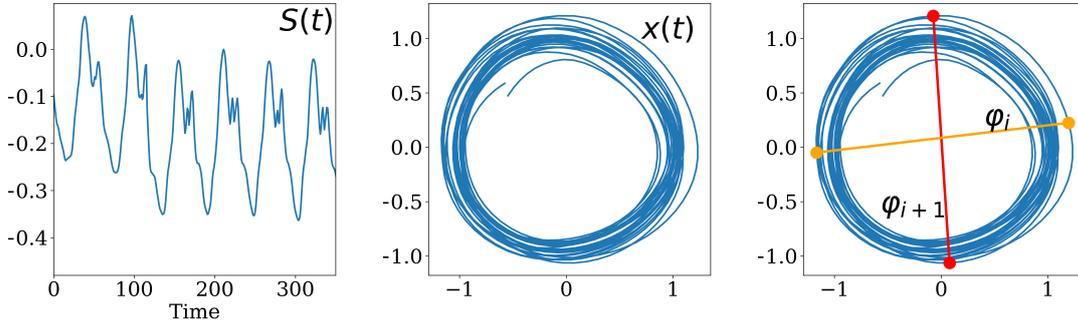


Рис. 2. Слева направо: временной ряд $S(t)$, фазовая траектория сегментов $x(t)$, признаковое описание сегментов.

Имея пространство Φ , для решения задачи классификации функцию f можно искать в семействе суперпозиций $a(g(\cdot), \gamma)$, где g — отображение в признаковое пространство Φ , $a(\cdot, \gamma)$ — некоторый алгоритм машинного обучения, параметризованный вектором гиперпараметров γ .

Решение задачи разбивается на два этапа:

1. Построение признакового описания объектов $\varphi = \{g(x_i)\}_{i=1}^l$ такого, что минимизируется функционал ошибок каждой модели локальной аппроксимации

$$\arg \min_{\mathbf{w} \in W} L_g(\mathbf{X}, \mathbf{w}) = \arg \min_{\mathbf{w} \in W} \sum_{i=1}^l \sum_{k=1}^n \|g(\mathbf{w}, \mathbf{x}_i) - \mathbf{x}_i\|_2^2.$$

2. Оптимизация функции ошибки обобщенной линейной модели, обучающей выборкой которой является набор (φ, \mathbf{y}) ,

$$\arg \min_{\theta \in \Theta} L_{GLM}(\varphi, \mathbf{y}, \theta) = \arg \min_{\theta \in \Theta} \left(- \sum_{i=1}^l \sum_{k=1}^K [y_i = k] \log P(y_i = k | \mathbf{x}_i, \theta) \right)$$

2. Построение признакового пространства на основе моделей локальной аппроксимации

2.1. Модели локальной аппроксимации

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

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

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 :

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

Модели авторегрессии

Модель AR-авторегрессии порядка m записывается следующим образом:

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

где w_0 — константа, w_1, \dots, w_m — параметры модели. Оптимальные параметры \mathbf{w}^* модели авторегрессии определяются минимизацией функции ошибки, например среднеквадратичной ошибки (MSE):

$$\mathbf{w}^* = \arg \min_{\mathbf{w} \in \mathbb{R}^m} \sum_{i=1}^n \|x_i - \hat{x}_i\|_2^2.$$

Модель авторегрессионного скользящего среднего действует модель скользящего среднего порядка q :

$$x_t = w_0 + \underbrace{\sum_{i=1}^m w_i x_{t-i}}_{AR(m)} + \underbrace{\sum_{i=1}^q \theta_i \varepsilon_{t-i}}_{MA(q)},$$

где w_0 — константа, $w_1, \dots, w_m, \theta_1, \dots, \theta_q$ — параметры модели, $\varepsilon_{t-1}, \dots, \varepsilon_{t-q}$ — ошибки.

Интегрированная модель авторегрессионного скользящего среднего является расширением моделей $ARMA(m, q)$ для нестационарных временных рядов.

Фурье-модель (FFT)

С помощью прямого преобразования Фурье находятся n первых комплексных амплитуд синусоидальных сигналов сегмента S . Так как амплитуды комплексные, то по ним можно вычислить и амплитуду и фазу:

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

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

Таким образом, аппроксимирующей моделью является обратное преобразование Фурье:

$$x_t = \frac{1}{N} \sum_{k=1}^n X_k \exp\left(\frac{2\pi i}{n}kj\right),$$

где X_k — k -я комплексная амплитуда.

Признаковым описанием сегмента является вектор $\mathbf{w}_{FFT} = [w_1, \dots, w_n] \in \mathbb{R}^n$.

Сингулярный спектр (SSA)

Из временного ряда S строится траекторная матрица H таким образом, что k -я строка траекторной матрицы получается сдвигом окна длины m на $k - 1$ элемент вправо:

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

Далее, к траекторной матрице применяется сингулярное разложение:

$$H = U\Sigma V^*,$$

где Σ — прямоугольная матрица, содержащая сингулярные числа $\sigma_1, \dots, \sigma_m$ на диагонали, а матрицы U и V — унитарные матрицы, состоящие из левых и правых сингулярных векторов соответственно [8].

В качестве признакового описания сегментов временных рядов используется вектор \mathbf{w}_{SSA} , образованный сингулярными числами

$$\mathbf{w}_{SSA} = \{\sigma_1, \dots, \sigma_m\}.$$

Self-Modeling regression

Структурным параметром модели SEMOR является некоторый временной ряд $p(t)$ (в общем случае — функция формы), называемый профилем модели. Такой подход предполагает, что входной временной ряд x и профиль модели p имеют одинаковые формы изгиба, то есть ряд x можно аппроксимировать профилем p . Модель можно описать выражением

$$x_t = w_1 + w_2 p(w_3 + w_4 t)$$

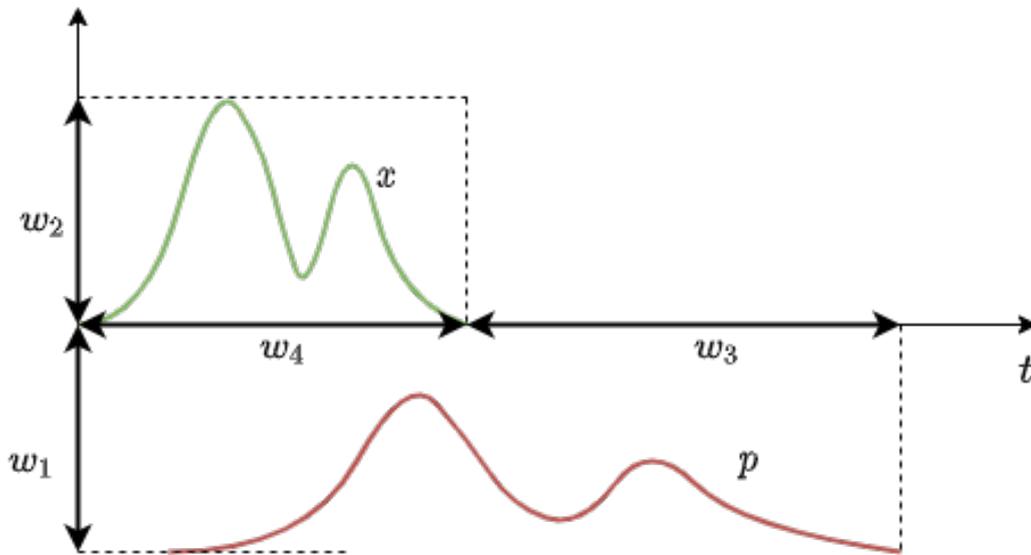


Рис. 3. Вычисление параметров модели SEMOR.

Поиск параметров модели разбивается на два шага. На первом шаге с помощью алгоритма динамической трансформации временной шкалы определяются параметры w_3 и w_4 . Далее, параметры w_1, w_2 находятся из шага двухпараметрической линейной регрессии.

Помимо описанных выше параметров, признаковое описание модели содержит параметр w_5 , определяющий качество аппроксимации временного ряда x профилем модели p , в данном случае:

$$w_5 = \sum_{t \in T} |x_t - (w_1 + w_2 p(w_3 + w_4 t))|.$$

3. Отбор признаков и классификация объектов промежуточного пространства

3.1. Отбор параметров

Образы моделей локальной аппроксимации, которые образуют промежуточное признаковое пространство, могут быть скоррелированы, что может привести к потере устойчивости и снижению качества классификации. Таким образом ставится задача минимизации числа признаков в промежуточном признаковом пространстве. Для решения этой задачи используется метод отбора признаков с помощью дискретного квадратичного программирования (QPFS)[9].

Пусть $\mathcal{J} = \{1, \dots, n_g\}$ — множество индексов признаков, $\mathcal{A} \subseteq \mathcal{J}$ — его подмножество. Решается задача поиска оптимального подмножества признаков:

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

где $Q : \mathcal{A} \rightarrow \mathbb{R}$ — критерий качества подмножества индексов признаков.

3.2. Многоклассовая классификация

Для классификации после отбора признаков применяется многомерная логистическая регрессия, позволяющая определить вероятность принадлежности объекта классу.

Algorithm 1 Алгоритм отбора признаков и оптимизации параметров обобщенно-линейной модели

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: **конец для**
 - 5: инициализировать $\mathbf{Z} = \mathbf{z}[1]$
 - 6: для $i = 2, \dots, l$
 - $\mathbf{Z} := (\mathbf{Z}, \mathbf{z}[i])$
 - 7: **конец для**
 - 8: $\mathbf{A} = \text{QPFS}(\mathbf{Z})$
 - 9: $\mathbf{Z} = \mathbf{Z}[\mathbf{A}]$
 - 10: $\hat{\theta} = \arg \min(L(f(\mathbf{Z}), \mathbf{Y}))$
-

3.3. 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 \mathbf{w} и параметризованная функция $G(\mathbf{x}, \sigma, \mathbf{w}) : I_n \rightarrow \mathbb{Y}$ такая, что $|G(\mathbf{x}, \mathbf{w}) - f(\mathbf{x})| < \varepsilon$ для всех $\mathbf{x} \in I_n$, где $\mathbf{w}_i \in \mathbb{R}^n$, $G = \sum_i \sigma(\mathbf{w}_i^\top \mathbf{x})$.

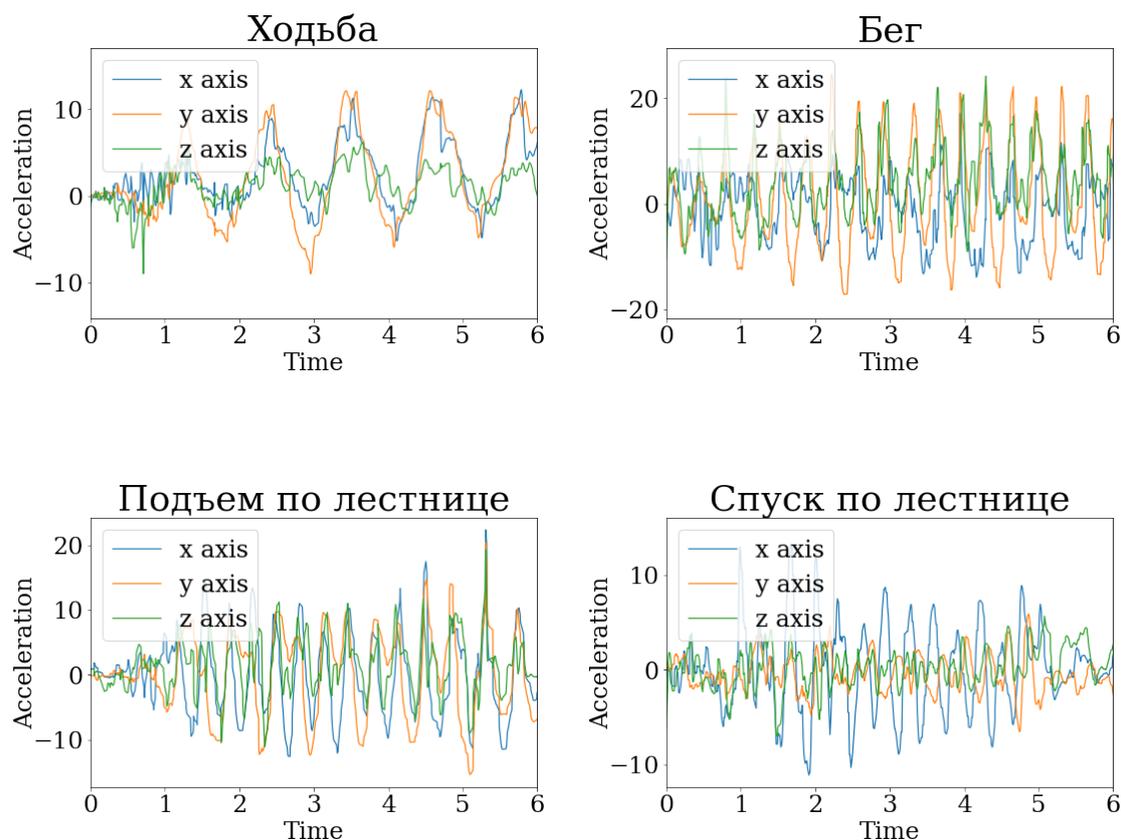
Теорема [10] гласит о том, что искусственная нейронная сеть прямой связи с одним скрытым слоем может аппроксимировать любую непрерывную функцию многих переменных с любой точностью. Это позволяет ввести понятие сложности модели как количество нейронов на скрытом слое эквивалентной ей нейронной сети:

$$\text{Complexity}(\mu) = \#|\text{нейронов на скрытом слое}|.$$

4. Генерация обучающей выборки

Для проведения эксперимента были использованы два набора данных: выборка UCI Human Activity Recognition [3] и собранные вручную данные с линейного акселерометра на телефоне.

Выборка UCI Human Activity Recognition состоит из 6 типов активности: ходьба, подъем по лестнице, спуск по лестнице, стояние, сидение, лежание - каждая из активностей снята с 30 человек в возрасте 19-48 лет. Данные разделены на последовательные участки по 2.56 секунды (128 считываний на участок), перекрывающие друг друга на 50%. Так же к данным изначально применен фильтр с частотой отсечения 0.3 Гц.



Собранные вручную данные состоят из четырех типов активности: ходьба, бег, подъем по лестнице и спуск по лестнице. Данные записывались приложением Physics Toolbox Suite с частотой 100 Гц.

4.1. Сегментация

Так как изначально данные представляют собой временные ряды, содержащие большое количество периодов, необходимо провести их сегментацию. Предполагаем, что частота повторения движений во время активности в среднем остается постоянной.

Назовем точку t временного ряда x_t локальным максимумом радиуса R , если

$$x_t \geq x_i, \forall i \in [i - R, i + R].$$

Сегментация происходит в три этапа:

1. Поиск всех локальных максимумов радиуса R

$$I = \{i \mid S(i) = \max_{j \in N([i-R, i+R])} S(j)\}.$$

2. Формирование набора сегментов X , где

$$X = \left\{ [S(i), S(i+1), \dots, S(j)] \mid i = I(k), j = I(k+1) \forall k \in \overline{1, |I| - 1} \right\}.$$

3. Фильтрация сегментов на основе гистограммы их длины.

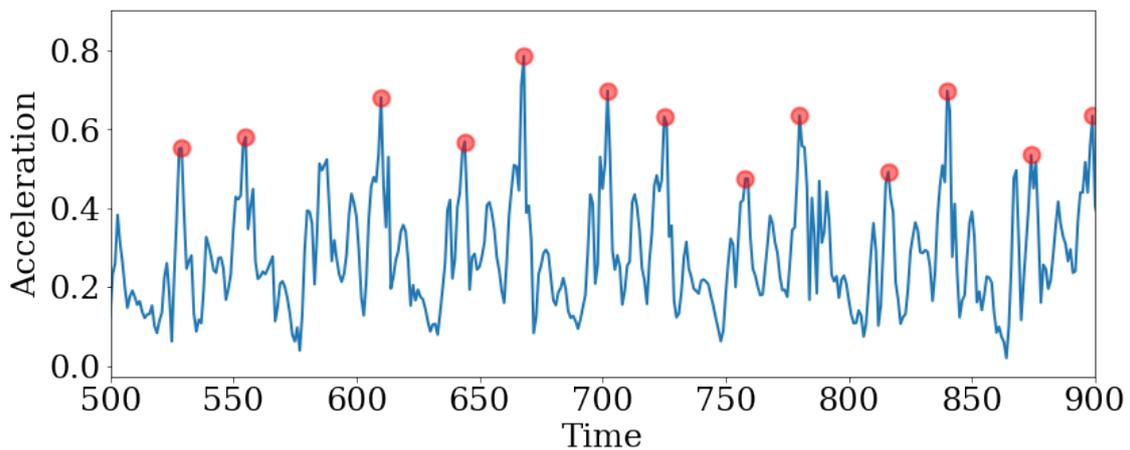


Рис. 4. Пример выделения локальных максимумов (отмечены красным) радиуса $R = 21$ на временном ряде ходьбы

4.2. Нормализация

Элементы полученного набора сегментов содержат различное количество временных точек. Необходимо привести все сегменты к одинаковой размерности, то есть нормализовать их. Для этого используются кубические сплайны.

Кубическим сплайном называется функция $S(x)$, которая:

- На каждом отрезке $[x_{i-1}, x_i]$ является многочленом степени не выше третьей.
- Имеет непрерывные первую и вторую производные на всем отрезке $[x_0, x_n]$, то есть

$$S(x) \in C^2[x_0, x_n]$$

- В точках x_i выполняется равенство $S(x_i) = y_i$, то есть сплайн интерполирует временной ряд.

Таким образом, для нормализации сегментов сначала находится сегмент с наименьшей длиной m . Затем каждый сегмент интерполируется кубическим сплайном, после чего вычисляются m равноудаленных точек внутри отрезка $[0, |x_t|]$ так, что первая и последняя точки являются началом и концом отрезка соответственно. С помощью нормализации все сегменты приводятся к одинаковой длине и строится выборка (x, y) .

5. Сравнение универсальной и обобщенно-линейной моделей

В вычислительном эксперименте использовался ансамбль из следующих моделей локальной аппроксимации:

1. модель AR-авторегрессии с параметром 5,
2. фурье-модель с выбором 4 главных частот,

3. модель сингулярного спектра с окном длины 10,
4. модель SEMOR с профилем ходьбы,
5. модель SEMOR с профилем бега,
6. модель SEMOR с профилем подъема по лестнице,
7. модель SEMOR с профилем спуска по лестнице.

Порядок параметров моделей в признаковом описании сохранен.

5.1. Построение промежуточного признакового пространства

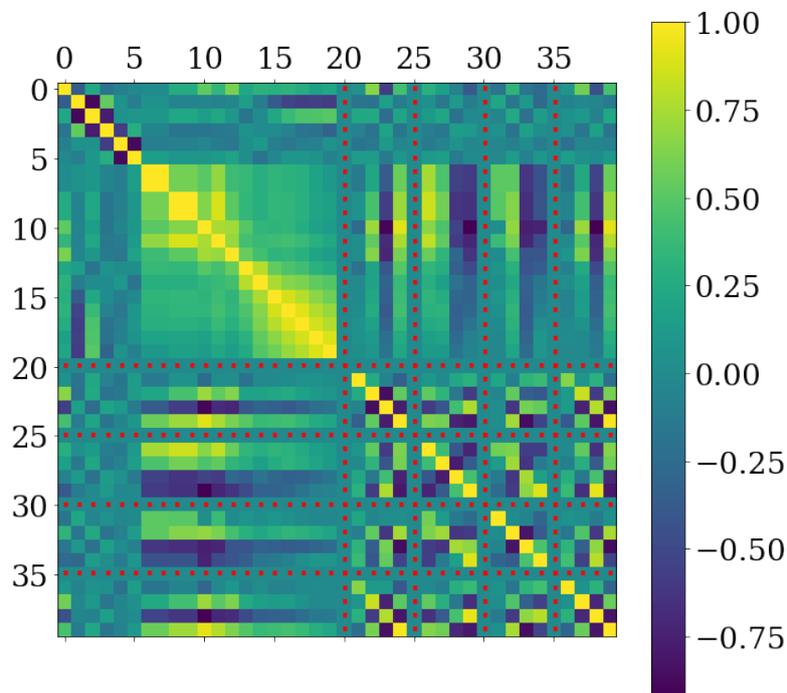


Рис. 5. Матрица корреляции признаков до удаления нулевых признаков. Красными точками отмечены нулевые признаки, которые впоследствии удаляются.

В ходе построения признакового пространства на основе параметров моделей локальной аппроксимации было замечено, что часть признаков всегда принимала одни и те же значения. Параметр w_1 каждой из SEMOR-моделей всегда принимал нулевое значение, соответственно являлся неинформативным в контексте промежуточного признакового пространства. Таким образом перед отбором признаков проводится процедура удаления всех нулевых признаков.

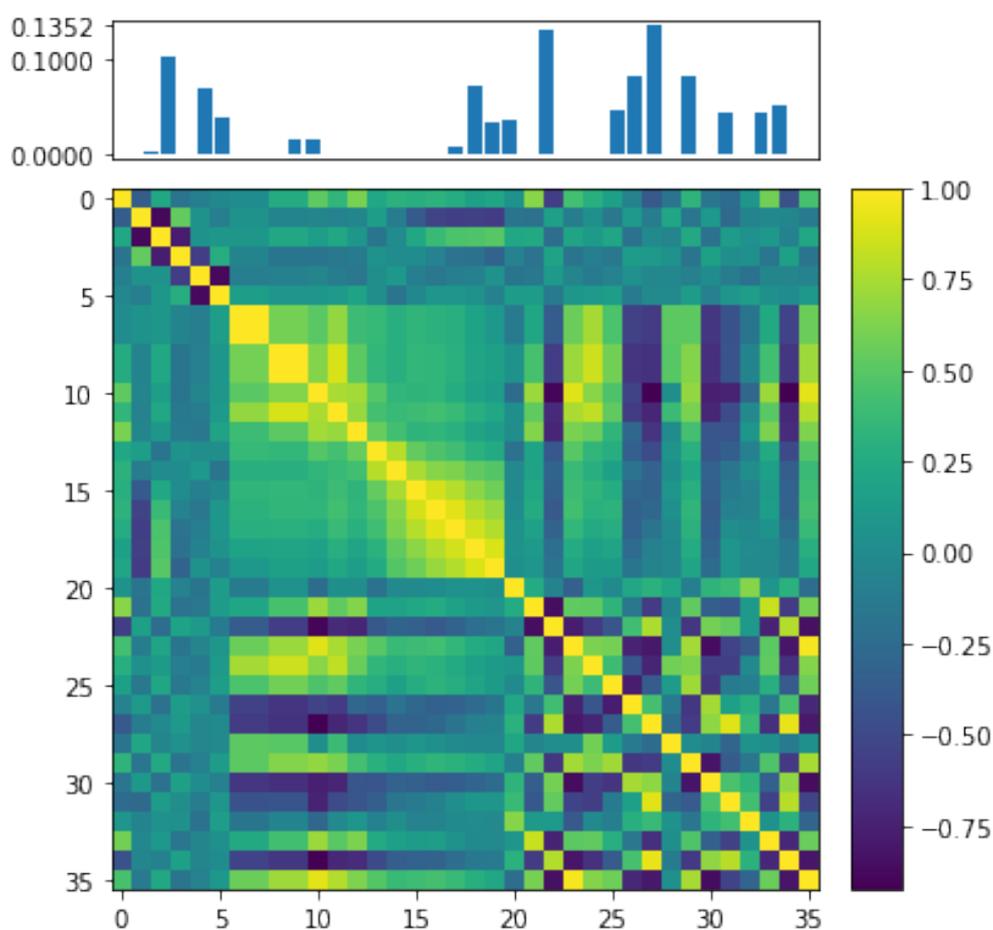


Рис. 6. Отбор признаков. Сверху: Нормализованная значимость признаков. Снизу: Матрица корреляции признаков.

Отбор признаков производится с помощью алгоритма QPFS, принимающего на вход матрицу признакового описания объектов. На выходе алгоритма возвращается нормализованная значимость для каждого из признаков.

5.2. Сравнение универсальной модели и обобщенной линейной модели

На полученной после отбора признаков выборке обучается обобщенная линейная модель GLM. Сложность модели в данном случае описывается количеством выбранных признаков. Для того, чтобы оценить качество обучения модели GLM сложности C выбираются C наиболее значимых признаков.

В качестве универсальной модели взята двухслойная нейронная сеть, обученная на выборке. Сложность такой модели описывается числом нейронов на скрытом слое.

Перед обучением выборка данных каждой модели 10 раз случайным образом делится на обучающую и тестовую в соотношении 9 : 1. На полученном наборе выборок вычисляется средняя точность, усредненная функция потерь и среднеквадратичное отклонение функции потерь.

Для сравнения двух разных подходов классификации временных рядов, обе модели обучаются в одном диапазоне сложности $C \in [1; 35]$

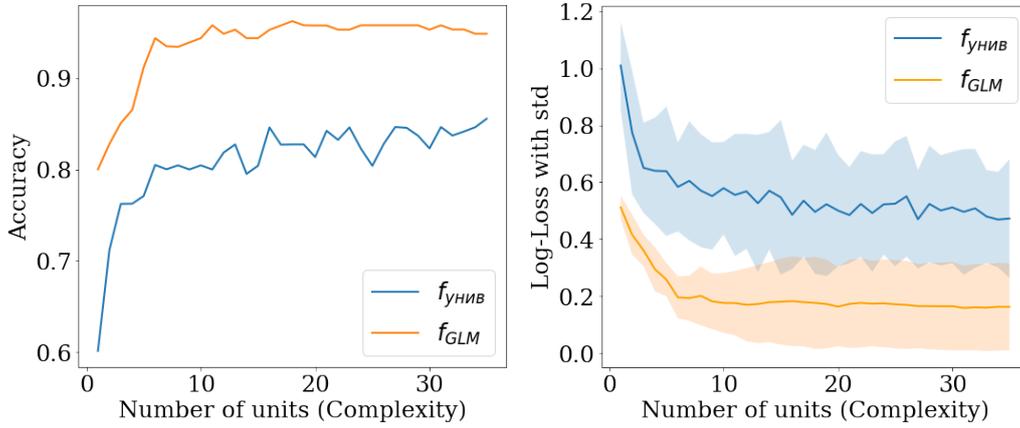


Рис. 7. Сравнение качества обучения обобщенной-линейной модели и универсальной двухслойной нейросети в зависимости от заданной сложности. Слева: Точность классификации временных рядов универсальной моделью и моделью GLM в зависимости от их сложности. Справа: Функция ошибки от сложности моделей со среднеквадратичным отклонением.

Как можно заметить на Рис. 7, обобщенно-линейная модель, обученная на параметрах локально-аппроксимирующих моделей при одинаковой сложности обучается гораздо лучше универсальной модели.

Для оценки устойчивости моделей наложим на выборку (\mathbf{X}, \mathbf{y}) гауссовский шум:

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

Проведем аналогичные эксперименты на зашумленной выборке. Качество классификации обеих моделей снизилось (Рис. 8). Для оценки устойчивости моделей необходимо проанализировать, насколько сильно упало качество относительно незашумленных данных. На Рис. 9 показано, что падение качества на зашумленной выборке у

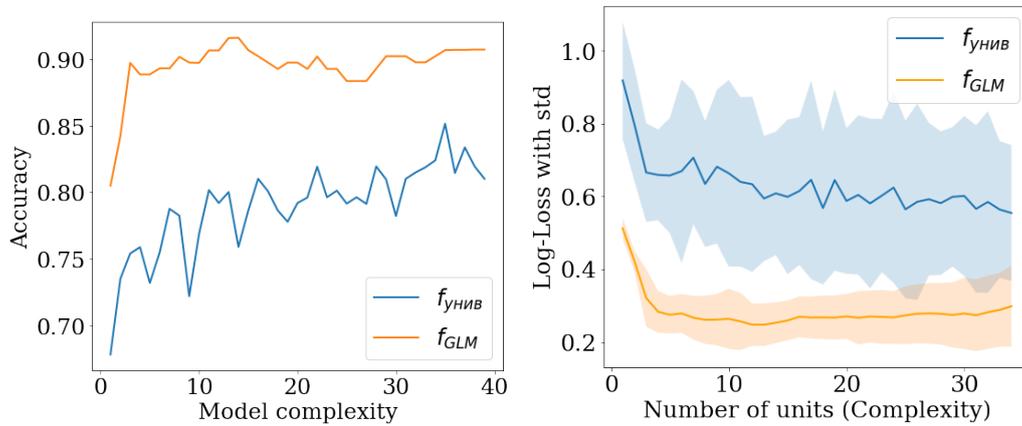


Рис. 8. Сравнение качества обучения обобщенно-линейной модели и универсальной двухслойной нейросети на зашумленных данных. Слева: Точность классификации временных рядов в зависимости от сложности моделей. Справа: Функция ошибки от сложности моделей со среднеквадратичным отклонением.

обобщенно-линейной модели ниже, чем у универсальной. Из чего следует заключение, что модель GLM при той же сложности является более устойчивой.

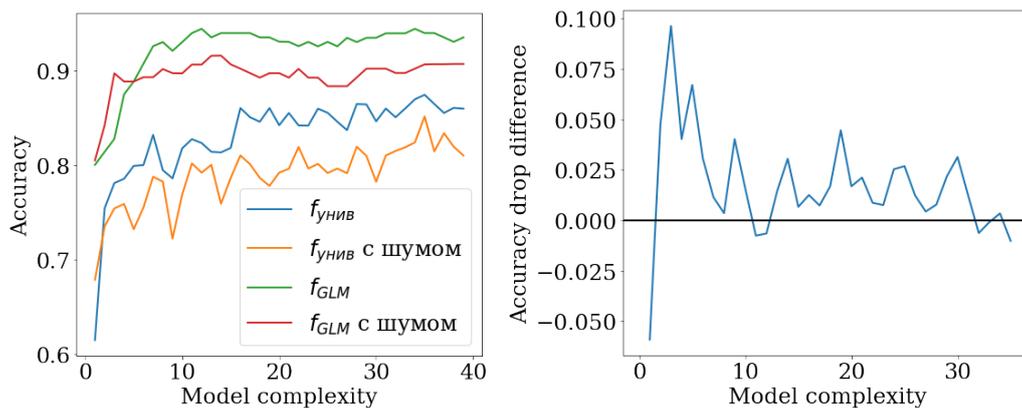


Рис. 9. Слева: Точность классификации временных рядов универсальной моделью и моделью GLM на различных данных. Справа: Разность падения качества модели GLM и универсальной модели.

Заключение

Вычислительные эксперименты показали, что построение промежуточного признакового пространства, порожденного параметрами ансамбля моделей локальной аппроксимации, в задаче классификации временных рядов обеспечивает более точную модель и устойчивую модель при одинаковой сложности. Классификация, основанная на локальных моделях, менее склонна к переобучению, в отличие от универсальной модели. Таким образом, ансамбль локальных моделей предпочтительнее в парето-оптимальном смысле универсальной модели: точнее, устойчивее, проще.

Список литературы

1. Кузнецов МП and Ивкин НП. Алгоритм классификации временных рядов акселерометра по комбинированному признаковому описанию. *Машинное обучение и анализ данных*, pages 1471–1483, 2015.
2. Карасиков МЕ and Стрижов ВВ. Классификация временных рядов в пространстве параметров порождающих моделей . *Информ. и её примен.*, page 121–131, 2016.
3. Davide Anguita, Alessandro Ghio, Luca Oneto, Xavier Parra, and Jorge L. Reyes-Ortiz. Human activity recognition on smartphones using a multiclass hardware-friendly support vector machine. In José Bravo, Ramón Hervás, and Marcela Rodríguez, editors, *Ambient Assisted Living and Home Care*, pages 216–223, Berlin, Heidelberg, 2012. Springer Berlin Heidelberg.
4. Argyro Kampouraki, George Manis, and Christophoros Nikou. Heartbeat time series classification with support vector machines. In *IEEE Transactions on Information Technology in Biomedicine*, volume 13, pages 512–518, 2009.
5. Damian Eads, Daniel Hill, Sean Davis, Simon Perkins, Junshui Ma, Reid Porter, and James Theiler. Genetic algorithms and support vector machines for time series classification. *Proceedings of SPIE - The International Society for Optical Engineering*, 4787:74–85, 2002.
6. Michael Hüsken and Peter Stagge. Recurrent neural networks for time series classification, 2003.

7. Yi Zheng, Qi Liu, Enhong Chen, Yong Ge, and J. Leon Zhao. Time series classification using multi-channels deep convolutional neural networks. In *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, volume 8485 LNCS, pages 298–310, 2014.
8. H. Hassani. Singular Spectrum Analysis: Methodology and Comparison. *Journal of Data Science*, 5:239–257, 2007.
9. V. V. Strijov A. M. Katrutsa. Comprehensive study of feature selection methods to solve multicollinearity problem according to evaluation criteria. *Expert Systems with Applications*.
10. G. V. Cybenko. Approximation by Superpositions of a Sigmoidal function. *Mathematics of Control Signals and Systems*.