

# Порождение признаков с помощью локально-аппроксимирующих моделей\*

Максим Христороубов, В. В. Стрижов, Александра Гальцева,  
Данил Сайранов

khristolyubov.me@phystech.edu

<sup>1</sup>Московский физико-технический институт

В работе рассматривается классификация временных рядов акселерометра. Классификация производится методом порождения признаков с помощью локально аппроксимирующих моделей. Предполагается, что точки временного ряда можно разбить на кластеры, соответствующие разным классам. Предлагается выделить квазипериодические сегменты из временных интервалов точек, принадлежащих одному кластеру. В качестве признаков для классификации использовать параметры моделей, обученных на этих сегментах. Исследуется вопрос информативности порожденных признаков и возможность идентификации по ним владельца прибора или его физических параметров.

**Ключевые слова:** *временной ряд; классификация; кластеризация; сегментация временного ряда; локально аппроксимирующая модель*

## 1 Введение

В статье изучается задача идентификации движений человека по временным рядам. В дополнении к этому исследуется возможность выделения атрибутивных паттернов, которые могут быть использованы для определения личности или физических параметров субъектов данных в дополнении к их деятельности. Классификация временных рядов находит широкое применение в сфере здравоохранения.

Временные ряды являются объектами сложной структуры, требующие предварительной обработки и представления их в удобном для классификации виде. Необходимо отобрать исходный временной ряд в некоторое пространство признаков. Например, в статье [1] временной ряд аппроксимируется моделью, а признаками являются ее параметры. В качестве аппроксимирующей модели берется модель авторегрессии, а так же собственные числа траекторной матрицы, в случае модели сингулярного спектра. В работе [2] проводится разбиение временных рядов на сегменты фиксированной длины, на которых впоследствии обучается локально-аппроксимирующая модель. Для аппроксимации используется линейная модель, модель авторегрессии и коэффициенты преобразования Фурье. В [3] предлагается более разумный способ сегментации, а так же применяется аппроксимация сплайнами. Еще более общий подход к способу сегментации посредством нахождения главных компонент траекторной матрицы, рассмотрен в [6]. В [8] сравниваются между собой перечисленные выше подходы.

Метод кластеризации точек, соответствующих участкам разной деятельности, с помощью метода главных компонента (SSA, алгоритм гусеница [12]) рассмотрен в [11]. На участках, содержащих точки одного кластера, уже можно применять описанные выше методы. Другим подходом к классификации точек временного ряда на основе нейросетей рассмотрены в [9] и [13].

В работе исследуется оптимальное признаковое описание точек при котором можно будет идентифицировать род деятельности человека. Предлагается построить набор

---

\*Работа выполнена при финансовой поддержке РФФИ, проекты № 00-00-00000 и 00-00-00001.

27 локально-аппроксимирующих моделей и выбрать наиболее адекватные. Производится по-  
 28 строение пространства описаний элементарных движений. Новизна работы заключается  
 29 в исследовании зависимости решений задач классификации действий и предсказания па-  
 30 раметров человека.

## 31 2 Постановка задачи

Пусть имеется исходный временной ряд  $\mathbf{d} = \{d_i\}_{i=1}^M \in \mathbb{R}^M$ . Предполагается, что он состоит из последовательности сегментов:

$$\mathbf{d} = [\mathbf{s}_1, \dots, \mathbf{s}_N],$$

32 где  $\mathbf{s}_i \in \mathcal{S}$ ,  $|\mathcal{S}|$  — число различных действий (кластеров). Считается, что периоды  $|\mathbf{s}|$   
 33 сегментов различаются незначительно, причем известен максимальный период  $|\mathbf{s}| \leq T$ .

Требуется решить задачу классификации точек ряда:

$$R : \mathcal{I} \rightarrow Y,$$

34 где  $\mathcal{I} = \{1, \dots, M\}$  — моменты времени, на котором задан временной ряд, а  $Y$  — метки  
 35 классов.

36 Каждый моменты времени  $i_k$  отобразим с помощью  $g : \mathcal{I} \rightarrow R^T$  в временной сегмент  
 37  $\mathbf{x}_k$  длины  $3T$ , по предположению, содержащий всю информацию о локальном поведении  
 38 ряда:

$$39 \quad g(i_k) = \mathbf{x}_k = \{d_i\}_{i=k-2T}^{k+T}. \quad (1)$$

Полученные сегменты  $\mathbf{x} \in \mathcal{X}$  — объекты сложной структуры, представленные временными рядами. Рассматривается задача классификации, а именно восстановление зависимости

$$y = f(\mathbf{x}),$$

40 где  $y \in Y$  — пространство ответов. Тогда исходная задача классификации представляет  
 41 собой  $R = f \circ g$ .

42 Заданы выборка объектов сложной структуры и ответов  $\mathfrak{D} = \{(\mathbf{x}_i, y_i)\}_{i=1}^m$ . Задача со-  
 43 стоит в нахождении функции  $f$ , минимизирующие суммарные потери на выборке  $\mathfrak{D}$ , при  
 44 заданной функции потерь  $\mathcal{L} : (\mathcal{X}, F, Y) \rightarrow R$ ,  $\mathcal{L}(f(\mathbf{x}_i), y_i)$ , характеризующая ошибку  
 45 классификации функции  $f \in F$  на элементе  $\mathbf{x}_i$ .

Пусть

$$w : \mathcal{X} \rightarrow W = \mathbb{R}^n$$

46 — процедура построения признаковов описания сегмента. Тогда  $W$  — пространство при-  
 47 знаков, в котором производится классификация временных рядов.

Пусть  $b$  — алгоритм многоклассовой классификации:

$$b : W \rightarrow Y$$

Тогда  $f$  ищется в среди композиций вида

$$f = b \circ w$$

48 Для любого  $w$  можно найти оптимальное значение вектора  $\hat{\mu}$  параметров классификатора  
 49  $b(w(x), \mu)$ , минимизирующего функционал качества:

$$50 \quad \hat{\mu} = \arg \min_{\mu} \tilde{Q}(b \circ w, \mathfrak{D}) \quad (2)$$

Оптимальный метод обучения для конкретного способа задания пространства признаков  $w$  и выбранной модели классификации, определяется по скользящему контролю

$$\hat{f}_w = \arg \min_w \widehat{CV}(f_w, \mathfrak{D}),$$

где  $\widehat{CV}(f, \mathfrak{D})$  — внешний контроль качества метода обучения  $f$ ,  $\mathfrak{D} = \mathfrak{L} \sqcup \mathfrak{E}$ :

$$\widehat{CV}(f, \mathfrak{D}) = \frac{1}{r} \sum_{k=1}^r Q(f^*(\mathfrak{L}), \mathfrak{E}) \quad (3)$$

В качестве функционала качества используется

$$Q(f, \mathfrak{L}) = \frac{1}{|\mathfrak{L}|} \sum_{(\mathbf{x}, y) \in \mathfrak{L}} |\{(\mathbf{x}, y) \in \mathfrak{L} | f(\mathbf{x}) = y\}|,$$

а в оценке точности классификации объектов класса  $c \in Y$  используется модифицированный функционал качества

$$Q_c(f, \mathfrak{L}) = \frac{1}{|\mathfrak{L}|} \sum_{(\mathbf{x}, y) \in \mathfrak{L}} \frac{|\{(\mathbf{x}, y) \in \mathfrak{L} | f(\mathbf{x}) = y = c\}|}{|\{(\mathbf{x}, y) \in \mathfrak{L} | y = c\}|}$$

### 3 Порождение признаков

Модель  $\hat{\mathbf{x}}(\mathbf{w}, \mathbf{x})$ , с помощью которой производится порождение признаков временных сегментов, называется локально-аппроксимирующей моделью, в силу локальности рассматриваемого сегмента временного ряда. В качестве признакового описания сегмента предлагается брать вектор оптимальных параметров  $w$  модели:

$$\mathbf{w}(\mathbf{x}) = \arg \min_{\mathbf{w}} \|\hat{\mathbf{x}}(\mathbf{w}, \mathbf{x}) - \mathbf{x}\|_2$$

Для порождения признакового описания  $\hat{\mathbf{w}}$  для  $L$ -мерного временного ряда последовательно применяется операция порождения признаков в точке к каждому одномерному ряду, а потом все полученные векторы признаков  $\mathbf{w}_l$  объединяются:

$$\hat{\mathbf{w}} = \mathbf{w}_1 \oplus \dots \oplus \mathbf{w}_L,$$

где  $\oplus$  это операция конкатенации,  $\mathbf{w} \in W$ ,  $\hat{\mathbf{w}} \in W^L$

#### 3.1 Авторегрессия

Модель авторегрессии  $AR(T)$  предсказывает следующее значение сегмента (1) как линейную комбинацию  $2T$  предыдущих:

$$\hat{x}^{(k)} = w_0 + \sum_{j=1}^{2T} w_j x^{(k-j)}$$

Для нахождения оптимального вектора параметров  $\mathbf{w}$  нужно решить задачу минимизации:

$$\mathbf{w}(\mathbf{x}) = \arg \min_{\mathbf{w}} (\|\mathbf{x} - \hat{\mathbf{x}}\|_2^2 = \arg \min_{\mathbf{w}} \|\mathbf{x} - \mathbf{X}\mathbf{w}\|_2^2 = (\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T \mathbf{x},$$

58 где  $\mathbf{X}$  — матрица Ганкеля:

$$\mathbf{X} = \begin{pmatrix} x^{(1)} & x^{(2)} & \dots & x^{(T)} \\ x^{(2)} & x^{(3)} & \dots & x^{(T+1)} \\ \vdots & \vdots & \ddots & \vdots \\ x^{(2T+1)} & x^{(2T+2)} & \dots & x^{(3T)} \end{pmatrix}$$

### 59 3.2 Анализ сингулярного спектра

Рассмотрим модель SSA порождения данных. Поставим в соответствие временному сегменту  $\mathbf{x}$  траекторную матрицу  $\mathbf{X}$ . Ее сингулярное разложение

$$\mathbf{X}^T \mathbf{X} = \mathbf{V} \mathbf{H} \mathbf{V}^T, \mathbf{H} = \text{diag}(h_1, \dots, h_T).$$

60  $h_1 \dots h_T$  — собственные числа матрицы  $\mathbf{X}^T \mathbf{X}$ . Первая пара собственных чисел берется в  
61 качестве нового признаков описания  $\mathbf{w} = (h_1, h_2)$ .

### 62 3.3 Дискретное преобразование Фурье

К временному сегменту применяется дискретное преобразование Фурье:

$$w_k = \sum_{n=1}^{2T} x^{(k)} e^{-\frac{2\pi i}{2T} kn}$$

63 и в качестве признаков  $\mathbf{w}$  берутся все его коэффициенты.

## 64 4 Вычислительный эксперимент

65 Цель эксперимента: исследовать влияние использования результата решения одной из  
66 задач классификации на решение другой задачи. Исследование зависимости между реше-  
67 ниями задачи классификации типа активности и задачи классификации пола.

68 Условия измерения данных: данные — это измерения акселерометра и гироскопа,  
69 встроенных в мобильное устройство iPhone 6s, хранящегося в переднем кармане брюк  
70 участника. Временные ряды содержат значения ускорения человека и углы ориентацию  
71 телефона для каждой из трёх осей — всего шесть временных рядов. Частота дискрети-  
72 зации составляет 50 Гц. Метками классов служат: подъем по лестнице вверх, спуск по  
73 лестнице вниз, ходьба, бег трусцой, сидение, лежание. Данные собраны с 24 участников,  
74 для каждого из которых известны рост, вес, возраст и пол. Данные собирались в условиях  
75 проведения эксперимента: участникам выдавали телефон и просили выполнять одно из  
76 шести действий в течении двух минут.

77 Для эксперимента берется шесть временных рядов в 39000 временных моментов (780  
78 секунд). Данные снимаются с четырех человек (двое мужчин и две женщины), которые  
79 выполняют подъем или спуск по лестнице (два типа деятельности).

80 `data.png`

81 В силу большой размерности и того, что тип деятельности не изменяется часто, целе-  
82 сообразно классифицировать временной ряд только в каждый десятый момент времени  
83 (десять раз в секунду), а остальные точки относить к тому же классу, которому принад-  
84 лежит ближайшая точка. Это позволяет уменьшить размер выборки для классификации  
85 с 39000 точек до 3900, если классифицировать каждую десятую точку.

На соответствующих моментам времени сегментах строятся локально-аппроксимирующие модели, чьи параметры используются в качестве признаков. Сравниваются информативность признаков, порожденных моделью авторегрессии (с 50 членами), коэффициенты ряда Фурье (100 для каждого ряда), 2 сингулярных числа SSA разложения.

Отбор оптимальных признаков производится сначала по критерию Пирсона (остается 8% признаков), потом оптимальные 20 выбираются методом перебора Sequential Backward Floating Selection [15] для выбранной модели классификации. Алгоритм перебора состоит из двух шагов:

1) Выбор признака  $w^- = \arg \max_{w^- \in w} \max_f \widehat{CV}(f, \mathfrak{D}(w^k - w^-))$ , без которого модель классификации, полученная на оставшихся признаках дает максимальное значение  $\widehat{CV}(f^*, \mathfrak{D}(w^k - w^-))$ , где  $\widehat{hatCV}$  это (3), и отбрасывание  $w^-$  из вектора признаков  $w^{k-1} = w^k - w^-$ .

2) Если существует признак  $w^+$ , добавив который, можно получить большее значение функции, чем она была на выборке с признаками, которые были до предыдущего шага  $\max_f \widehat{CV}(f, \mathfrak{D}(w^{k-1} + w^-)) < \max_f \widehat{CV}(f, \mathfrak{D}(w^{k-1} + w^+))$ , то добавить этот признак к вектору признаков  $w^k = w^{k-1} + w^+$ .

Эти шаги повторяются, пока не останется 20 оптимальных признаков, то есть пока не будет достигнуто равенство  $k = 20$ .

Для классификации используется две модели для обучения.

1) Логистическая регрессия [16] с  $l_2$  регуляризацией, для которой  $\tilde{Q}$  в (2) задается выражением:

$$\tilde{Q}(w, \mathfrak{D}) = \frac{1}{2} w^T w + \sum_{i=1}^{|\mathfrak{D}|} \ln(1 + \exp(-y_i(X_i^T w + c)))$$

2) C-Support Vector Classification [17], минимизирующий функцию:

$$\begin{aligned} \min_{w, b, \zeta} \quad & \frac{1}{2} w^T w + C \sum_{i=1}^n \zeta_i \\ \text{subject to} \quad & y_i(w^T \varphi(x_i) + b) \geq 1 - \zeta_i, \\ & \zeta_i \geq 0, i = 1, \dots, n \end{aligned} \tag{4}$$

где  $K(x_i, x_j) = \varphi(x_i)^T \varphi(x_j)$  — ядро.

Для исследования зависимости между решением задач классификации активности и пола находятся общие признаки, попавшие в векторы оптимальных параметров для решения обеих задач классификации. Однако, в силу высокой размерности и мультиколлинеарности возможна ситуация, когда для решения обеих задач используют одну и ту же информацию, хотя векторы оптимальных признаков не имеют общих компонент. Поэтому для проверки наличия причинно-следственной зависимости между активностью и полом производится обучения на расширенной матрице признаков: сначала решается одна задача классификации, а потом полученный вектор решения используется как признак для решения второй задачи.

Таблица 1 Результаты экспериментов

	mean(activity)	std(activity)	mean(gender)	std(gender)	Common features
Log Regression	0.888	0.162	0.928	0.212	0
after adding	0.882	0.170	0.955	0.112	
SVC	0.911	0.143	0.962	0.110	0
after adding	0.883	0.191	0.960	0.087	

Результаты экспериментов показывают, что использование полового признака при определении типа деятельности не улучшает точности модели, а использование типа активности при определении пола значительно уменьшает ошибки и увеличивает устойчивости модели логистической регрессии. Ошибки модели C-Support Vector Classification не уменьшаются, но увеличивается устойчивость.

## 5 Заключение

Желательно, чтобы этот раздел был, причём он не должен дословно повторять аннотацию. Обычно здесь отмечают, каких результатов удалось добиться, какие проблемы остались открытыми.

khristolyubov.me@phystech.edu

## References

- [1] N. P. Ivkin, M. P. Kuznetsov.. 2015. Time series classification algorithm using combined feature description. . *Machine Learning and Data Analysis* (11):1471–1483.
- [2] V. V. Strijov, M. E. Karasikov. 2016. Feature-based time-series classification *Informatics* doi: <http://dx.doi.org/10.3114/S187007708007>.
- [3] D.A. Anikeev, G.O. Penkin, V.V. Strijov. 2018. Local approximation models for human physical activity classification // *Informatics* doi: <http://dx.doi.org/10.14357/19922264190106>.
- [4] V.V. Strijov, R.V. Isachenko.. 2016. Metric learning in multiclass time series classification problem. *Informatics and Applications* (10(2)):48–57.
- [5] V.V. Strijov, Andrew Zadayanchuk, Maria Popova.. 2016. Selection of optimal physical activity classification model using measurements of accelerometer. *Information Technologies* (22(4)):313–318.
- [6] Strijov V.V., Motrenko A.P.. 2016. Extracting fundamental periods to segment human motion time series. *Journal of Biomedical and Health Informatics* 20(6):1466 – 1476.
- [7] Strijov V.V., Ignatov A.. 2015. Human activity recognition using quasiperiodic time series collected from a single triaxial accelerometer. *Multimedia Tools and Applications* pages 1–14.
- [8] Isachenko R.V., Bochkarev .., Zharikov I.N., Strijov V.V.. 2018. Feature Generation for Physical Activity Classification. *Artificial Intelligence and Decision Making* 3 : 20-27.
- [9] Dafne van Kuppevelt, Joe Heywood, Mark Hamer, Séverine Sabia, Emla Fitzsimons, Vincent van Hees. 2019. Segmenting accelerometer data from daily life with unsupervised machine learning. *PLOS ONE* doi: <http://dx.doi.org/10.5255/UKDA-SN-8156-3>.
- [10] Andrea Mannini, Angelo Maria Sabatini. 2010. Machine Learning Methods for Classifying Human Physical Activity from On-Body Accelerometers *PubMed* doi: <http://dx.doi.org/10.3390/s100201154>.
- [11] Grabovoy A.V., Strijov V.V. 2020. Quasiperiodic time series clustering for human activity recognition *Lobachevskii Journal of Mathematics*

- 
- 150 [12] D.L. Danilov and A.A. Zhiglovsky. 1997. *Main components of time series: method "Gesens-*  
 151 *itsa"* (St. Petersburg)
- 152 [13] Y.G. Cinar and H. Mirisaee. 2018. Period-aware content attention RNNs for time series fore-  
 153 casting with missing values *"Neurocomputing* 312, 177–186
- 154 [14] Malekzadeh, Mohammad and Clegg, Richard G. and Cavallaro, Andrea and Haddadi, Hamed.  
 155 2019. *Mobile Sensor Data Anonymization* pages 49–58. Proceedings of the International Con-  
 156 ference on Internet of Things Design and Implementation doi: [http://dx.doi.org/10.1145/](http://dx.doi.org/10.1145/3302505.3310068)  
 157 3302505.3310068.
- 158 [15] Petr Somol, Jana, Novovicova, Pavel Pudil. 2010. Efficient Feature Subset Selection and Subset  
 159 Size Optimization doi: <http://dx.doi.org/10.5772/9356>.
- 160 [16] Joanne Peng, Kuk Lida Lee, Gary M. Ingersoll. 2002. An Introduction to Logistic Regression  
 161 Analysis and Reporting doi: <http://dx.doi.org/10.1080/00220670209598786>.
- 162 [17] Quanzhong Liu, Chihau Chen, Yang Zhang, Zhengguo Hu. 2011. Feature selection for support  
 163 vector machines with RBF kernel doi: <http://dx.doi.org/10.1007/s10462-011-9205-2>.

Received

164