

Порождение признаков с помощью локально-аппроксимирующих моделей*

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

khristolyubov.me@phystech.edu

¹Московский 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 методы 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

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

$$34 \quad \mathbf{d} = [\mathbf{s}_1, \dots, \mathbf{s}_N], \quad (1)$$

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

Требуется решить задачу классификации точек ряда:

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

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

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

$$42 \quad h(k) = \mathbf{x}_k = \{d_i\}_{i=k-2T}^{k+T}. \quad (2)$$

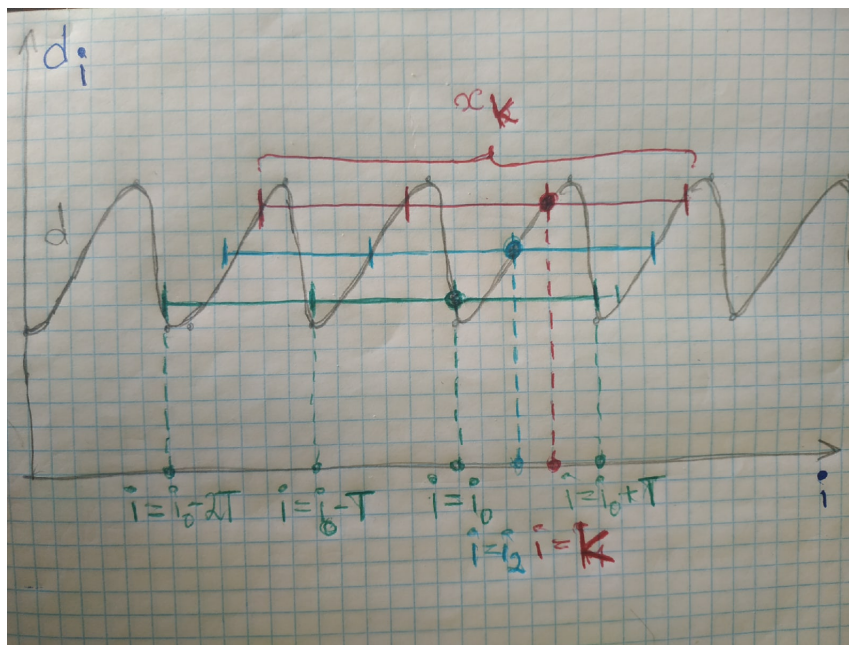


Рис. 1 Временной ряд

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

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

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

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

Пусть

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

— процедура построения признакового описания сегмента. Тогда W — пространство признаков, в котором производится классификация временных рядов.

Пусть b — алгоритм многоклассовой классификации:

$$b : W \rightarrow Y \quad (3)$$

Тогда f ищется в среди композиций вида

$$f = b \circ w$$

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

$$\hat{\mu} = \arg \min_{\mu} \tilde{Q}(b \circ w, \mathfrak{D}) \quad (4)$$

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 , $\mathfrak{D} = \mathfrak{L} \sqcup \mathfrak{E}$:

$$\widehat{CV}(f, \mathfrak{D}) = \frac{1}{r} \sum_{k=1}^r Q(f^*(\mathfrak{L}), \mathfrak{E}) \quad (5)$$

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 g , с помощью которой производится порождение признаков временных сегментов, называется локально-аппроксимирующей моделью, в силу локальности рассматриваемого сегмента 2 временного ряда 1. В качестве признакового описания сегмента предлагается использовать 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 к каждому одномерному ряду, и все полученные векторы признаков \mathbf{w}_l объединяются:

$$\tilde{\mathbf{w}} = \mathbf{w}_1 \times \dots \times \mathbf{w}_L,$$

61 где \times это операция конкатенации, $\mathbf{w} \in W$, $\tilde{\mathbf{w}} \in W^L$

62 3.1 Авторегрессия

Модель авторегрессии AR(2T) предсказывает следующее значение сегмента (2) 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} решается задача минимизации:

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

63 где \mathbf{X} — матрица Ганкеля:

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

64 3.2 Анализ сингулярного спектра

Рассмотрим модель SSA порождения данных. Поставим в соответствие временному сегменту \mathbf{x} его траекторную матрицу \mathbf{X} 1. Ее сингулярное разложение

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

65 $h_1 \dots h_T$ — собственные числа матрицы $\mathbf{X}^\top \mathbf{X}$. Первые десять собственных чисел берется в
66 качестве нового признакового описания $\mathbf{w} = [h_1, \dots, h_{10}]$.

67 3.3 Дискретное преобразование Фурье

К временному сегменту применяется дискретное преобразование Фурье:

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

68 или же в матричном виде $\mathbf{w} = \mathbf{Z}\mathbf{x}$, где матрица \mathbf{Z} :

$$\mathbf{Z} = \begin{pmatrix} e^{-\frac{2\pi i}{T}} & e^{-\frac{2\pi i}{T} \cdot 2} & \dots & e^{-\frac{2\pi i}{T} \cdot T} \\ e^{-\frac{2\pi i}{T} \cdot 2} & e^{-\frac{2\pi i}{T} \cdot 4} & \dots & e^{-\frac{2\pi i}{T} \cdot 2T} \\ \vdots & \vdots & \ddots & \vdots \\ e^{-\frac{2\pi i}{T} \cdot 3T} & e^{-\frac{2\pi i}{T} \cdot 3T \cdot 2} & \vdots & e^{-\frac{2\pi i}{T} \cdot 3T \cdot T} \end{pmatrix}.$$

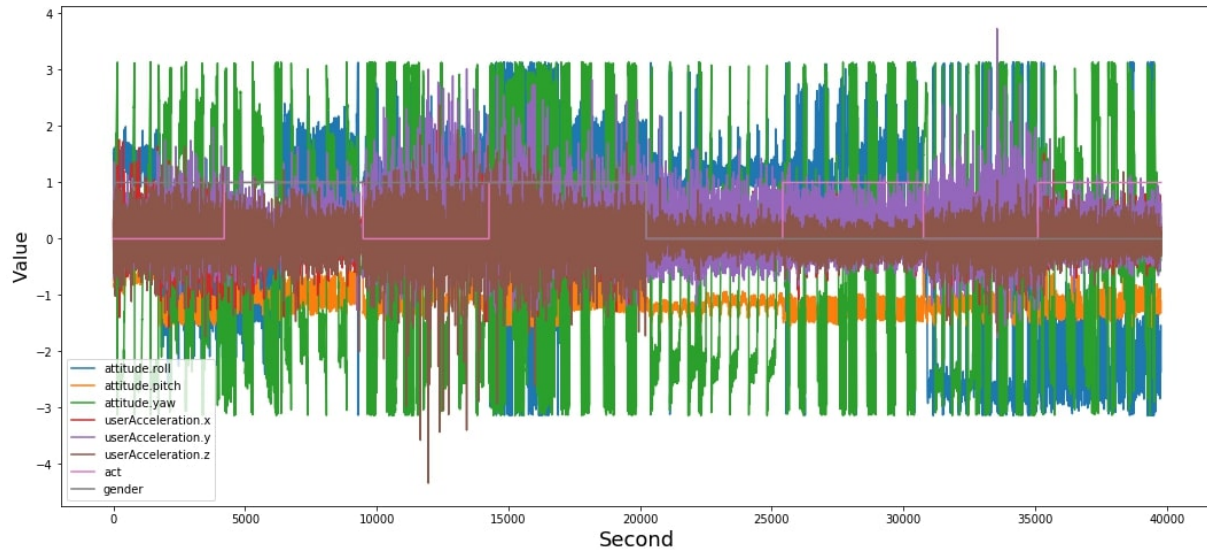
69 В качестве признаков \mathbf{w} берутся T коэффициентов Фурье.

4 Вычислительный эксперимент

В качестве вычислительного эксперимента решаются две задачи: классификация типов активности человека и классификация пола человека.

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

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



Производится классификация не всех моментов времени, а только каждый десятый момент времени (пять в секунду), а класс остальных моментов времени определяется классом ближайшего классифицированного момента времени.. Это позволяет уменьшить размер выборки для классификации с 39000 точек до 3900, если классифицировать каждую десятую точку.

На соответствующих моментах 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), 10 сингулярных числа SSA разложения.

Отбор оптимальных признаков производится методом Sequential Backward Floating Selection [15] для выбранной модели классификации.

Для классификации используются две модели b (3).

1) Логистическая регрессия [16] с l_2 регуляризацией, для которой \tilde{Q} в (4) 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:

$$\begin{aligned} \min_{w,b,\zeta} \frac{1}{2} w^T w + C \sum_{i=1}^n \zeta_i, \\ \text{subject to } y_i(w^T \varphi(x_i) + b) \geq 1 - \zeta_i, \\ \zeta_i \geq 0, i = 1, \dots, n, \end{aligned} \quad (6)$$

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

99 Отдельно решаются задачи классификации активности и пола человека. После этого
100 для проверки наличия причинно-следственной зависимости между активностью и полом
101 производится обучения на расширенной матрице признаков: сначала решается одна зада-
102 ча классификации, а потом полученный вектор решения используется как признак для
103 решения второй задачи.

Таблица 1 Среднее и дисперсия верных классификаций при кросс-валидации

	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	

104 Результаты показывают, что использование пола при определении типа деятельности
105 не улучшает точности модели, а использование типа активности при определении пола зна-
106 чительно уменьшает ошибки и увеличивает устойчивость модели логистической регрес-
107 сии. Ошибки модели C-Support Vector Classification не уменьшаются, но увеличивается
108 устойчивость.

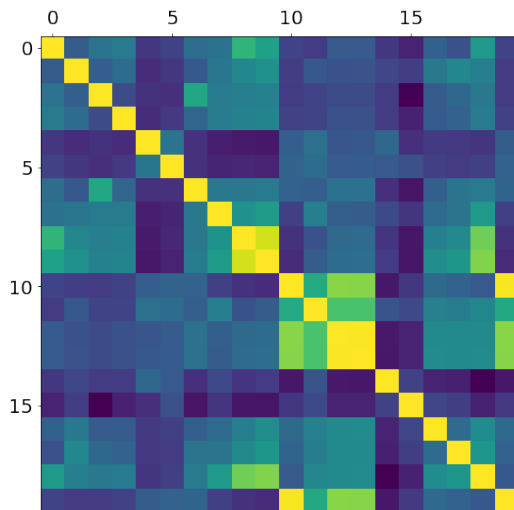


Рис. 2 Корреляционная матрица оптимальных признаков

Первые 13 признаков — это оптимальные признаки для решения задачи классификации активности, последние 7 признаков — оптимальные в задаче классификации пола.

5 Дополнение

Алгоритм отбора Sequential Backward Floating Selection состоит из двух шагов:

1) Выбор признака $w^- = \arg \max_{w^- \in w} \max_f \widehat{CV}(f, \mathcal{D}(w^k - w^-))$, без которого модель классификации, полученная на оставшихся признаках дает максимальное значение $\widehat{CV}(f^*, \mathcal{D}(w^k -$

$w^-)$, где \widehat{CV} это (5), 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 $\max_f \widehat{CV}(f, \mathcal{D}(w^{k-1} + w^-)) < \max_f \widehat{CV}(f, \mathcal{D}(w^{k-1} + w^+))$, то добавить этот признак к вектору признаков $w^k = w^{k-1} + w^+$.

Эти шаги повторяются, пока не останется K оптимальных признаков, то есть пока не будет достигнуто равенство $K=1$.
 khrishtolyubov.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*
- [12] D.L. Danilov and A.A. Zhiglovsky. 1997. *Main components of time series: method "Gesensitsa"(St. Petersburg)*
- [13] Y.G. Cinar and H. Mirisae. 2018. Period-aware content attention RNNs for time series forecasting with missing values *Neurocomputing* 312, 177–186

- 153 [14] Malekzadeh, Mohammad and Clegg, Richard G. and Cavallaro, Andrea and Haddadi, Hamed.
154 2019. *Mobile Sensor Data Anonymization* pages 49–58. Proceedings of the International Con-
155 ference on Internet of Things Design and Implementation doi: [http://dx.doi.org/10.1145/](http://dx.doi.org/10.1145/3302505.3310068)
156 3302505.3310068.
- 157 [15] Petr Somol, Jana, Novovicova, Pavel Pudil. 2010. Efficient Feature Subset Selection and Subset
158 Size Optimization doi: <http://dx.doi.org/10.5772/9356>.
- 159 [16] Joanne Peng, Kuk Lida Lee, Gary M. Ingersoll. 2002. An Introduction to Logistic Regression
160 Analysis and Reporting doi: <http://dx.doi.org/10.1080/00220670209598786>.
- 161 [17] Quanzhong Liu, Chihau Chen, Yang Zhang, Zhengguo Hu. 2011. Feature selection for support
162 vector machines with RBF kernel doi: <http://dx.doi.org/10.1007/s10462-011-9205-2>.

163 *Received*