

Классификация квазипериодических временных сигналов с помощью сферических гармоник

Тихонов Д. М., Стрижсов В. В.

Аннотация: Решается задача построения модели аппроксимации и классификации квазипериодических сигналов. Требуется наименьшая структурная сложность. Структурная сложность — это число настраиваемых параметров модели. Для перехода в фазовое пространство временной ряд векторизуется с помощью метода задержек. Для понижения сложности модели в фазовом пространстве выбирается подпространство. Фазовая траектория в полученном подпространстве аппроксимируется линейной комбинацией сферических гармоник. Полученные коэффициенты используются в качестве признакового описания. Вычислительный эксперимент проведен на измерениях акселерометра мобильного устройства с тремя классами движений человека.

Ключевые слова: временные ряды, аппроксимация, классификация, фазовое пространство, сферические гармоники.

1 Введение

Работа посвящена классификации квазипериодических временных рядов. Примерами таких рядов являются показатели акселерометра и гироскопа во время повторяющейся физической активности, электрокардиограмма.

Классическим подходом является использование метода ближайших соседей в сочетании с различными функциями расстояния [1], либо ансамбль различных типов дискриминативных моделей на одном или нескольких пространствах признаков [2–4]. Эти походы имеют общее свойство: этап преобразования данных, на котором временные ряды преобразуются в новое пространство признаков, например, с помощью шейплет преобразований [2, 3] или функции DTW [4]. Другой широкий класс подходов к решению — это использование различных нейронных сетей на исходных временных рядах [5], например, с помощью операции свертки. Методы объединяющие подход порождения признаков и прямое использования временных рядов основаны на методе задержек и классификации полученных векторных представлений [6]. Последний является ближайшим подходом к решению, описанному в данной работе и альтернативным подходом для сравнения.

В этой работе предлагается модель классификации временных рядов, основанный на модели аппроксимации сферическими гармониками на поверхности сферы. Для этого производится переход в пространство фазовых траекторий или траекторное пространство. Переход осуществляется методом задержек [7]. Метод задержек используется при анализе нестационарных временных рядов. Например, в методе сингулярного спектрального анализа [8] разложения на компоненты и прогноз основаны на траекторной матрице. Она позволяет перейти от скалярного временного ряда к многомерному представлению. Метод задержек так же получили широкое распространение в анализе нелинейных динамических систем [7, 9].

Избыточная размерность траекторного пространства [8, 10, 11] приводит к неустойчивости исследуемых моделей и избыточно сложному описанию временного ряда. Для понижения размерности фазового пространства предлагается использовать метод главных компонент [12, 13].

В выбранном фазовом подпространстве малой размерности фазовая траектория проецируется на p -мерную единичную сферу. Полученную на поверхности сферы функцию

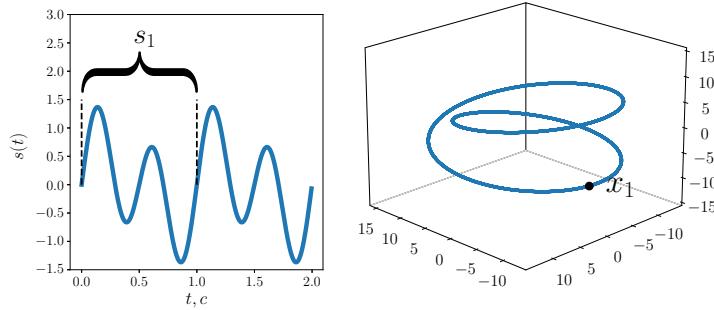


Рис. 1 Слева: сегмент временного ряда, справа: фазовая траектория.

предлагается аппроксимировать линейной комбинацией сферических гармоник. Полученные коэффициенты в дальнейшем используются как признаковое описание в задаче классификации.

На Рис. 1 показан сегмент синтетического временного ряда и его фазовая траектория в пространство размерности три. Фазовая траектория получена с помощью метода задержек и метода главных компонент над матрице задержек.

2 Постановка задачи классификации сферическими гармониками

Пусть заданы \mathbf{S} — множество временных рядов, \mathbf{Z} — множество номеров классов. Требуется по конечной выборке $\mathbf{S}^b = \{(\mathbf{s}_1, y_1), \dots, (\mathbf{s}_b, y_b)\}$, где $\mathbf{s} = [s_1, \dots, s_N]$ — временной ряд, b — число объектов в выборке, построить отображение

$$y^* : \mathbf{S} \rightarrow \mathbf{Z}. \quad (1)$$

По имеющемуся временному ряду \mathbf{s} строится траекторная матрица

$$\mathbf{H}_s^n = \begin{bmatrix} s_1 & s_2 & \dots & s_{n-1} & s_n \\ s_2 & s_3 & \dots & s_n & s_{n+1} \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ s_{N-n+1} & s_{N-n+2} & \dots & s_{N-1} & s_N \end{bmatrix} = \begin{bmatrix} \mathbf{s}_1 \\ \mathbf{s}_2 \\ \vdots \\ \mathbf{s}_m \end{bmatrix}, \quad m = N - n + 1,$$

где N — длина временного ряда, n — ширина окна, не меньшая, чем предполагаемый период, $\mathbf{s}_t = [s_t, s_{t+1}, \dots, s_{t+n-1}] \in \mathbb{H}_s \subseteq \mathbb{R}^n$ — векторы, образующие фазовую траекторию ряда \mathbf{s} .

Размерность траекторного пространства \mathbb{H}_s^n избыточна. Предлагается снижать размерность $\mathbb{H}_s^n \rightarrow \mathbb{H}_x^p$ с помощью метода главных компонент при $p \ll n$:

$$\mathbf{H}_x^p = \mathbf{H}_s^n \mathbf{U} = \begin{bmatrix} \mathbf{x}_1 \\ \mathbf{x}_2 \\ \vdots \\ \mathbf{x}_m \end{bmatrix}, \quad \mathbf{x}_t \in \mathbb{R}^p, \quad t \in [1, m] \quad (2)$$

где \mathbf{U} — матрица преобразования алгоритма метода главных компонент с количеством компонент равным p , соответствующим наибольшим собственным значениям.

В полученном подпространстве фазовая траектория переводится из декартовых в сферические координаты $\mathbb{H}_x^p \rightarrow \mathbb{S}_r^p$:

$$\varphi : \mathbf{x} \rightarrow \mathbf{r} = [r, \alpha_{p-1}, \dots, \alpha_1], \quad \mathbf{a} = [\alpha_{p-1}, \dots, \alpha_1], \quad \mathbf{S}_r^p = \varphi(\mathbf{H}_x^p).$$

По полученным представлениям точек в пространстве \mathbb{S}^p строится модель фазовой траектории как линейная комбинация сферических гармоник.

Предлагается представить y^* (1) как суперпозицию отображений

$$t \mapsto \mathbf{s} \mapsto \mathbb{H}_s^n \rightarrow \mathbb{H}_x^p \rightarrow \mathbb{S}_r^p \rightarrow \mathbb{W}^{p-1} \rightarrow \mathbf{Z},$$

где \mathbb{W}^{p-1} — пространство весов модели аппроксимации, \mathbb{H}_s^n — фазовое пространство, полученное методом задержек, \mathbb{H}_x^p — фазовое подпространство в декартовых координатах, \mathbb{S}_r^p — фазовое подпространство в сферических координатах.

3 Модель фазовой траектории

В полученном подпространстве \mathbb{S}^p строится модель фазовой траектории

$$f_{\text{sp}} : \mathbb{R}^{|\mathbf{w}_{\text{sp}}|} \times \mathbb{S}_r^p \rightarrow \mathbb{R}. \quad (3)$$

Нормированные значения функции f_{sp} интерпретируются как вероятность принадлежности точки фазового пространства \mathbb{S} к фазовой траектории \mathbf{S} временного ряда \mathbf{s}

$$\pi(\mathbf{a}) \approx f_{\text{sh}}(\mathbf{w}_{\text{sp}}, \mathbf{a}) = \sum_{l_{p-1}=0}^{N_{\text{approx}}} \sum_{l_{p-2}=0}^{l_{p-1}} \cdots \sum_{l_1=-l_2}^{l_2} w_{l_{p-1}, \dots, l_1} Y_{l_{p-1}, \dots, l_1}(\mathbf{a}), \quad (4)$$

где $\pi(\mathbf{a})$ — функция проекций, определенная ниже, l_{p-1}, \dots, l_1 — индексы, определяющие сферические гармоники и удовлетворяющие условию $l_{p-1} \geq l_{p-2} \dots l_2 \geq |l_1|$, N_{approx} — максимальное значение старшего индекса, $Y_{l_{p-1}, \dots, l_1}(\mathbf{a})$ — вещественные сферические гармоники, определенные ниже, w_{l_{p-1}, \dots, l_1} — весовые коэффициенты $w_{l_{p-1}, \dots, l_1} \in \mathbb{W}^{p-1}$. N_{approx} связан с точностью модели, чем больше значение, тем точнее аппроксимация и больше переобучение.

В качестве базисных функций на поверхности $(p-1)$ -мерной сферы используются сферические гармоники:

$$\mathcal{Y}_{l_{p-1}, \dots, l_1}(\alpha_{p-1}, \dots, \alpha_1) = \left[\prod_{k=2}^{p-1} \bar{P}_{l_k}^{l_{k-1}}(\alpha_k) \right] \frac{1}{\sqrt{2\pi}} \exp(\pm i|l_1|\alpha_1), \quad (5)$$

За функцию $\bar{P}_{l_k}^{l_{k-1}}(\alpha_k)$ обозначается

$$\bar{P}_{l_k}^{l_{k-1}}(\alpha) = c_{l_k}^{l_{k-1}} \cdot (\sin \alpha)^{\frac{-(k-2)}{2}} P_{l_k + \frac{(k-2)}{2}}^{-(l_{k-1} + \frac{(k-2)}{2})}(\cos \alpha), \quad c_{l_k}^{l_{k-1}} = \sqrt{\frac{2l_{k-1} + k - 1}{2} \frac{(l_k + l_{k-1} + k - 2)!}{(l_k - l_{k-1})!}}$$

где $P_\mu^{-\eta}(x)$ — полиномы Лежандра, $c_{l_k}^{l_{k-1}}$ — нормировочный коэффициент. Подробнее о выводе формул и введенных обозначениях в

Сферические гармониками определены в комплексном пространстве. В исследуемом случае нет необходимости в реальной и комплексной части одновременно. Используются только вещественные сферические гармоники. Это упрощает реализацию и сохраняет свойства ортонормированности. Так вещественные сферические гармоники представимы в следующем виде

$$Y_{l_{p-1}, \dots, l_1}(\alpha_{p-1}, \dots, \alpha_1) = \begin{cases} \operatorname{Re}(\mathcal{Y}_{l_{p-1}, \dots, l_1}(\alpha_{p-1}, \dots, \alpha_1)), & \text{если } l_1 \geq 0 \\ \operatorname{Im}(\mathcal{Y}_{l_{p-1}, \dots, l_1}(\alpha_{p-1}, \dots, \alpha_1)), & \text{иначе.} \end{cases} \quad (6)$$

Для расчета w_{l_{p-1}, \dots, l_1} весовых коэффициентов в (4) необходимо определить проекцию фазовой траектории $\pi(\mathbf{a})$ на поверхности сферы \mathbb{S}^{p-1} . Для этого поверхность сферы разбивается на области, как показано слева на Рис. 2.

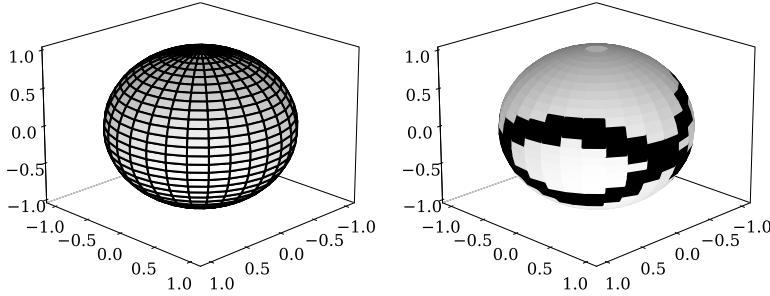


Рис. 2 Слева: области на сфере, справа: полученная проекция.

По полученному разбиению \mathbb{A}^{p-1} на сфере и по точкам фазовой траектории строится функция проекции $\pi(\mathbf{a})$, $\mathbf{a} \in \mathbb{A}^{p-1}$

$$\pi(\mathbf{a}) = \begin{cases} 1, & \text{если } \mathbf{a} \in \mathbb{A}^{p-1} \cap \mathbf{S}_z^p, \\ 0, & \text{если } \mathbf{a} \in \mathbb{A}^{p-1} \setminus \mathbf{S}_z^p. \end{cases} \quad (7)$$

Пример проекции справа на Рис. 2.

Составляется система относительно весовых коэффициентов w_{l_{p-1}, \dots, l_1}

$$\begin{pmatrix} Y_{l_{p-1}, \dots, l_1}(\mathbf{a}_1) & \dots & Y_{0, \dots, 0}(\mathbf{a}_1) \\ \vdots & \ddots & \vdots \\ Y_{l_{p-1}, \dots, l_1}(\mathbf{a}_d) & \dots & Y_{0, \dots, 0}(\mathbf{a}_d) \end{pmatrix} \begin{pmatrix} w_{l_{p-1}, \dots, l_1} \\ \vdots \\ w_{0, \dots, 0} \end{pmatrix} = \begin{pmatrix} \pi(\mathbf{a}_1) \\ \vdots \\ \pi(\mathbf{a}_d) \end{pmatrix}, \quad (8)$$

где $Y_{l_{p-1}, \dots, l_1}(\mathbf{a}_i)$ — значение сферической гармоники в точке \mathbf{a}_i , d — количество точек в сетке на сфере Рис. 2. В более короткой записи:

$$\mathbf{Y}\mathbf{w}_{\text{sp}} = \boldsymbol{\Pi}. \quad (9)$$

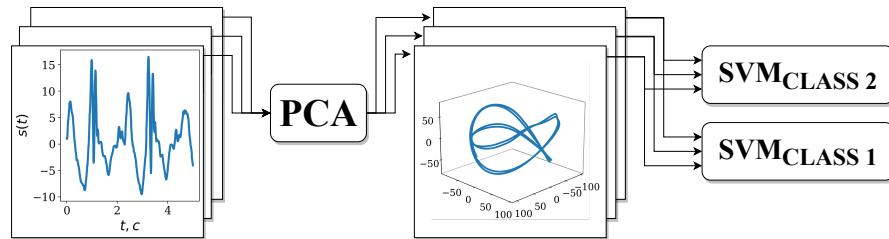
Эта постановка сводит решения к МНК, что ускоряет процедуру расчета весовых коэффициентов. В силу экспоненциального роста числа сферических гармоник при повышении размерности фазового пространства, вводятся регуляризаторы

$$\hat{\mathbf{w}}_{\text{sp}} = \arg \min_{\mathbf{w}_{\text{sp}}} \|\mathbf{Y}\mathbf{w}_{\text{sp}} - \boldsymbol{\Pi}\|^2 + \lambda \|\mathbf{w}_{\text{sp}}\|^2. \quad (10)$$

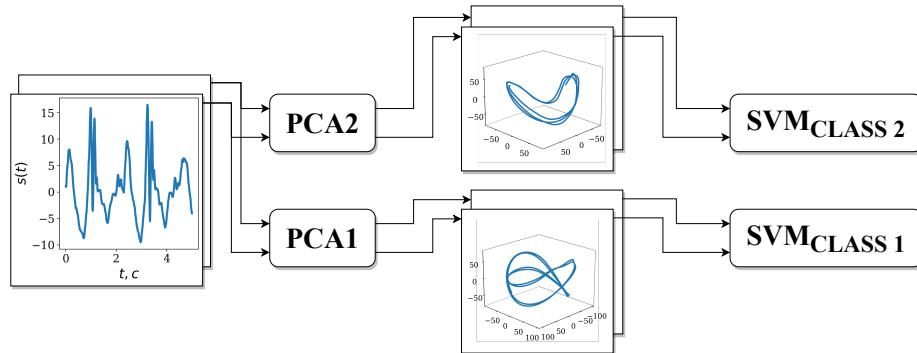
Решение представимо аналитически:

$$\hat{\mathbf{w}}_{\text{sp}} = (\lambda \mathbf{I} - \mathbf{Y}^\top \mathbf{Y})^{-1} \mathbf{Y}^\top \boldsymbol{\Pi}. \quad (11)$$

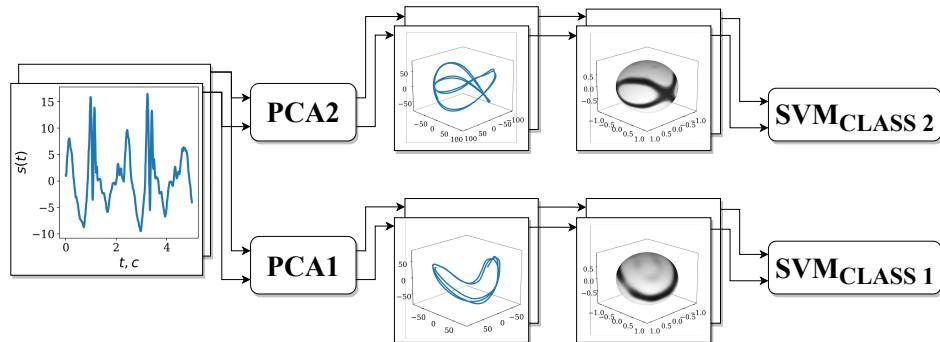
4 Модель классификации весовых коэффициентов



(a) Классификации с совместной матрицей РСА для всех классов



(б) Классификации с отдельной матрицей РСА для каждого класса



(в) Предложенный метод классификации

Рис. 3 Сравнение различных подходов к классификации фазовой траектории

Полученные весовые коэффициенты (11) для каждого $\mathbf{s} \in \mathbf{S}^b$ используются в качестве признакового описания в задаче классификации. Для классификации используется метод опорных векторов с мягким зазором

$$g(\mathbf{w}_{\text{sh}}) = \text{sign}(\mathbf{w}_{\text{svm}}^\top \mathbf{w}_{\text{sh}}). \quad (12)$$

Весовые коэффициенты \mathbf{w}_{svm} (12) оптимизируются

$$\begin{cases} \frac{1}{2} \|\mathbf{w}_{\text{svm}}\|_2^2 + C \sum_{i=1}^{\hat{m}} \xi_i \rightarrow \min_{w_{\text{svm}}, \xi} \\ y_i \cdot \mathbf{w}_{\text{svm}}^\top \mathbf{w}_{\text{sh},i} \geq 1 - \xi_i, \quad i = 1, \dots, b \\ \xi_i \geq 0, \quad i = 1, \dots, b, \end{cases} \quad (13)$$

где C — параметр настройки метода, который позволяет регулировать отношение между максимизацией ширины разделяющей полосы и минимизацией суммарной ошибки, ξ_i — величина, характеризующая ошибку на объектах $w_{\text{svm}} i$.

5 Модель классификации точек фазовой траектории

В качестве альтернативы используется ближайшая по архитектуре модель описанная в [6]. В этой работе классифицируются точки фазовой траектории \mathbf{H}_x^p по стратегии One-Vs-Rest. Решается задача

$$g(\mathbf{x}) = \text{sign}(\mathbf{w}_{\text{svm}}^\top \mathbf{x}), \quad \mathbf{x} \in \mathbf{H}_x^p. \quad (14)$$

Сравниваются два варианта классификации. В первом матрица преобразования метода главных компонент (2) строится для объединения всех фазовых траекторий полученных методом задержек и далее классифицируется для каждого класса отдельным методом опорных векторов как показано на Рис. 3а. Во втором варианте матрица преобразования метода главных компонент строится для каждого класса отдельно и далее классифицируется аналогично первому варианту как показано на Рис. 3б.

6 Эксперимент по классификации сферическими гармониками

Задан набор данных, измеренный датчиками акселерометра телефона. Он собирается с частотой дискретизации 50 Гц. Двадцать четыре участника разного пола, возраста, веса и роста выполнили три типа повторяемых физических активностей: движение по лестнице, ходьба, бег. Более подробное описание набора данных в [14]. Участники имеют разные типы походки и, как следствие, разные фазовые траектории. Поэтому были отобраны три человека со схожими показателями роста и веса.

Алгоритм классификации схож с вариантом на Рис. 3б. Основное отличие в аппроксимация фазовой траектории линейной комбинацией сферических гармоник как показано на Рис. 3в.

Размер предыстории выбирается равным $n = 150$, что советует трем секундам. Период движений в данных не превосходит двух с половиной секунд. Значение старшего индекса, используемого для построения модели сферических (4) гармоник, выбирается равным $N_{\text{approx}} = 10$ для размерности фазового пространства \mathbf{H}_x^p равного трем, $N_{\text{approx}} = 6$ для размерности четыре и $N_{\text{approx}} = 3$ для размерности пять. Для пространств большей размерности эксперимент не проводился из-за малого объема данных и вычислительной неэффективности подсчета сферических гармоник для размерности пространства равной шести и более.

Таблица 1 F-score классификации на трех пользователях с одинаковым типом походки

Метод Размерность p	Ходьба			Бег			Лестница		
	3	4	5	3	4	5	3	4	5
Пользователь 1	0.98	0.99	0.95	0.97	0.99	0.8	0.86	0.84	0.92
Пользователь 2	0.99	0.97	0.57	0.98	0.98	0.91	0.95	0.68	0.86
Пользователь 3	0.96	0.93	0.58	0.89	0.64	0.82	0.85	0.65	0.86

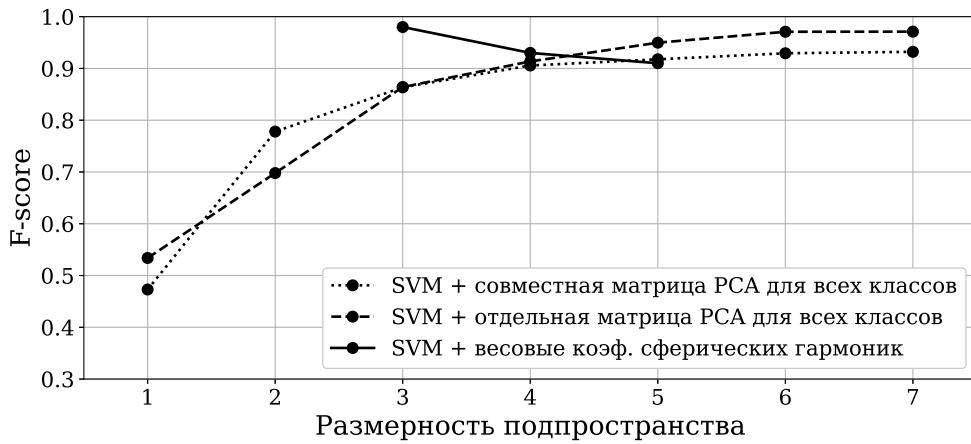


Рис. 4 F-score на ходьбе, беге, спуску и подъему по лестнице для одного пользователя

На Рис. 4 показано качество классификации F-score движении одного человека. Видно, что для эффективной классификации достаточно трехмерных сферических гармоник. При применении методов, обученных на данных первого пользователя, для классификации временных рядов записанных с второго и третьего пользователей качество падает в зависимости от типа движения. Так бег и ходьба классифицируются лучше, чем движение по лестнице. В таблице 1 представлена качество классификации для нескольких пользователей. При этом из-за увеличения размерности пространства падает общее качество классификации. Средние значения представлены на Рис. 5 и в таблице 2.

Таблица 2 Среднее качество классификации на трех пользователях с одинаковым типом походки

Метод Размерность p	F-score		
	3	4	5
SVM + весовые коэф. сферических гармоник	0.90	0.82	0.81
SVM + совместная матрица PCA для всех классов	0.79	0.80	0.89
SVM + отдельная матрица PCA для каждого класса	0.78	0.81	0.81

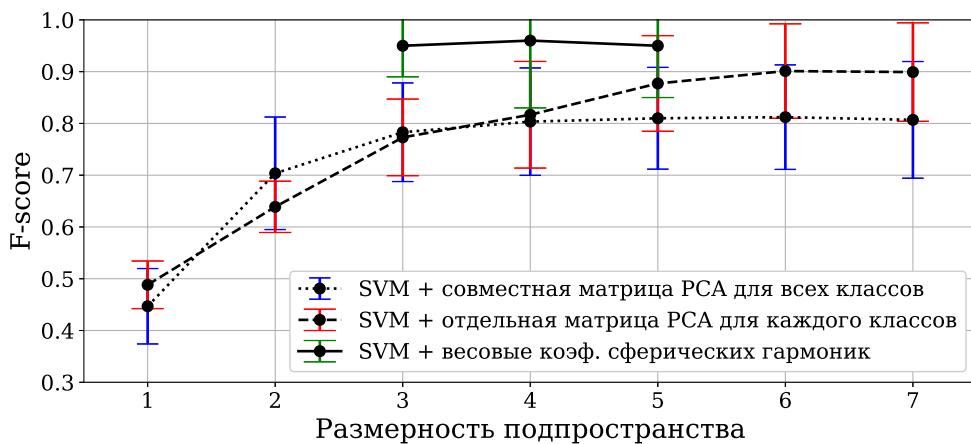
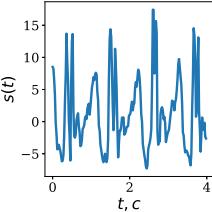
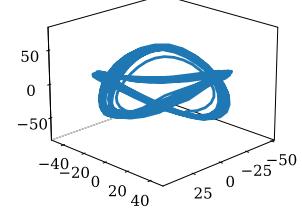
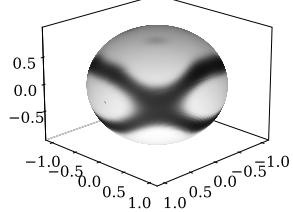
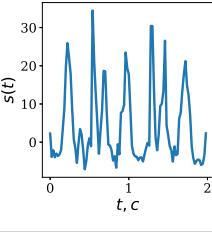
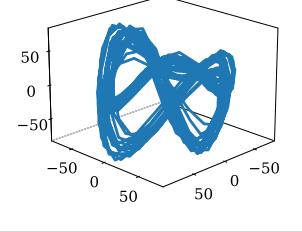
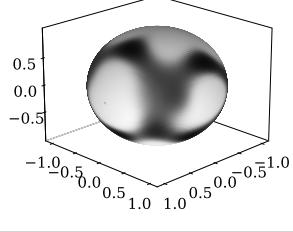
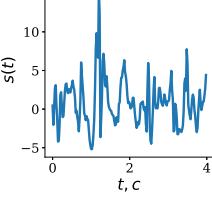
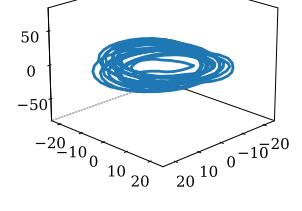
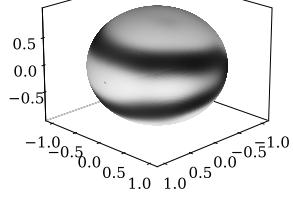


Рис. 5 Средний и СКО F-score на ходьбе, беге, спуску и подъему по лестнице

Таблица 3 Таблица сравнений разных классов движений
 Временной ряд Фазовая траектория Модель

Ходьба 		
Бег 		
Лестница 		

В таблице 3 показаны результаты работы алгоритма. Слева фрагмент временного ряда записанного с акселерометра. По центру полученные фазовые траектории в трехмерном пространстве. Справа модели сферических гармоник. Видно, что разные классы временных рядов имеют разные траектории и разные модели аппроксимации.

7 Заключение

В работе решалась задача классификации квазипериодических временных рядов, а также задача построения модели фазовой траектории. Для снижения размерности траекторного подпространства к фазовой траектории применялся метод главных компонент. Фазовая траектория, представленная в сферических координатах, аппроксимировалась линейной комбинацией сферических гармоник.

Эксперимент проводился на трех временных рядах: показатели акселерометра во время ходьбы, бега и спуска по лестнице для трех пользователей с близкими показателями роста и веса. Качество классификации высокое при оценке на пользователе, на котором обучались модели. Модели фазовых траекторий на поверхности сферы показали, что такая аппроксимация сохраняет геометрическую структуру последовательности точек, а также содержит информацию о математическом ожидании и дисперсии фазовой траектории. Это хорошо видно в таблице 3 на рисунках справа. Темно серая область на поверхности сферы представляет собой дисперсию фазовой траектории, более черная — математическое ожидание. Свойства модели аппроксимации будут отдельно исследовать в будущих работах. Количество моделей остается небольшим по сравнению с нейронными сетями или ближайшими соседями. Каждый этап классификации имеет явную интерпретацию.

Литература

1. Anthony Bagnall, Jason Lines, Aaron Bostrom, James Large, and Eamonn Keogh. The great time series classification bake off: a review and experimental evaluation of recent algorithmic advances. *Data Mining and Knowledge Discovery*, 31, 05 2017.
2. Jon Hills, Jason Lines, Edgaras Baranauskas, James Mapp, and Anthony Bagnall. Classification of time series by shapelet transformation. *Data Min. Knowl. Discov.*, 28(4):851–881, jul 2014.
3. Aaron George Bostrom and A. Bagnall. Binary shapelet transform for multiclass time series classification. *Trans. Large Scale Data Knowl. Centered Syst.*, 32:24–46, 2017.
4. Rohit Kate. Using dynamic time warping distances as features for improved time series classification. *Data Mining and Knowledge Discovery*, 30, 05 2015.
5. Jindong Wang, Yiqiang Chen, Shuji Hao, Xiaohui Peng, and Lisha Hu. Deep learning for sensor-based activity recognition: A survey. *Pattern Recognition Letters*, 119:3–11, 2019. Deep Learning for Pattern Recognition.
6. Jordan Frank, Shie Mannor, and Doina Precup. Activity and gait recognition with time-delay embeddings. *Proceedings of the Twenty-Fourth AAAI Conference on Artificial Intelligence, AAAI*, 2010.
7. D. Lai and G. Chen. Dynamical systems identification from time-series data: A hankel matrix approach. *Mathematical and Computer Modelling*, 24(3):1–10, 1996.
8. Zhigljavsky A. A. Golyandina N., Nekrutkin V. Analysis of time series structure: Ssa and related techniques. *Chapman and Hall*, 2002.
9. Floris Takens. Detecting strange attractors in turbulence. pages 366–381, 1981.
10. Anastasia Motrenko and Vadim Strijov. Extracting fundamental periods to segment biomedical signals. *IEEE journal of biomedical and health informatics*, 20, 08 2015.
11. K. Usmanova, Yu Zhuravlev, K. Rudakov, and Vadim Strijov. Approximation of quasiperiodic signal phase trajectory using directional regression. *Moscow University Computational Mathematics and Cybernetics*, 44:196–202, 10 2020.
12. Kenneth Ezukwoke and Samaneh Zareian. Kernel methods for principal component analysis (pca) a comparative study of classical and kernel pca. 12 2019.
13. Bernhard Scholkopf, Alex Smola, and Klaus-Robert Muller. Nonlinear component analysis as a kernel eigenvalue problem. *Neural Computation*, 10:1299–1319, 07 1998.
14. Mohammad Malekzadeh, Richard G. Clegg, Andrea Cavallaro, and Hamed Haddadi. Protecting sensory data against sensitive inferences. In *Proceedings of the 1st Workshop on Privacy by Design in Distributed Systems*, W-P2DS’18, pages 2:1–2:6, New York, NY, USA, 2018. ACM.