



Math-Net.Ru

Общероссийский математический портал

М. Е. Карасиков, В. В. Стрижов, Классификация временных рядов в пространстве параметров порождающих моделей, *Информ. и её примен.*, 2016, том 10, выпуск 4, 121–131

DOI: 10.14357/19922264160413

Использование Общероссийского математического портала Math-Net.Ru подразумевает, что вы прочитали и согласны с пользовательским соглашением

<http://www.mathnet.ru/rus/agreement>

Параметры загрузки:

IP: 44.220.184.63

6 октября 2024 г., 22:06:28



КЛАССИФИКАЦИЯ ВРЕМЕННЫХ РЯДОВ В ПРОСТРАНСТВЕ ПАРАМЕТРОВ ПОРОЖДАЮЩИХ МОДЕЛЕЙ*

М. Е. Карасиков¹, В. В. Стрижов²

Аннотация: Работа посвящена задаче многоклассовой признаковой классификации временных рядов. Признаковая классификация временных рядов заключается в сопоставлении каждому временному ряду его краткого признакового описания и позволяет решать задачу классификации в пространстве признаков. Исследуются методы построения пространства признаков временных рядов, при этом временной ряд рассматривается как последовательность сегментов, аппроксимируемых некоторой параметрической моделью, параметры которой используются в качестве их признаков описаний. Построенное признаковое описание сегмента временного ряда наследует от модели аппроксимации такое полезное свойство, как инвариантность относительно сдвига. Для решения задачи классификации в качестве признаков описаний временных рядов предлагается использовать распределения параметров аппроксимирующих сегменты моделей, что обобщает базовые методы, использующие непосредственно сами параметры аппроксимирующих моделей. Проведен ряд вычислительных экспериментов на реальных данных, показавших высокое качество решения задачи многоклассовой классификации. Эксперименты показали превосходство предлагаемого метода над базовым и многими распространенными методами классификации временных рядов на всех рассмотренных наборах данных.

Ключевые слова: временные ряды; многоклассовая классификация; сегментация временных рядов; гиперпараметры аппроксимирующей модели; модель авторегрессии; дискретное преобразование Фурье

DOI: 10.14357/19922264160413

1 Введение

Временным рядом x будем называть конечную упорядоченную последовательность чисел

$$x = [x^{(1)}, \dots, x^{(t)}].$$

Временные ряды являются объектом исследования в таких задачах анализа данных, как прогнозирование, обнаружение аномалий, сегментация [1], кластеризация и классификация [1]. Обзор по задачам и методам анализа временных рядов дается в [2]. Последние годы связаны с ростом интереса к данной области, проявляющемся в непрекращающемся предложении новых методов анализа временных рядов — метрик, алгоритмов сегментации, кластеризации и др.

В данной работе рассматривается задача классификации временных рядов, возникающая во многих приложениях (медицинская диагностика по электрокардиограммам [3] и электроэнцефалограммам [4], классификация типов физической активности по данным акселерометра [5], верификация динамических подписей [6] и т. д.).

Формально задача классификации в общем виде ставится следующим образом. Пусть X — множество описаний объектов произвольной природы, Y — конечное множество меток классов. Предполагается существование целевой функции — отображения $y : X \rightarrow Y$, значения которого известны только на объектах обучающей выборки

$$\mathcal{D} = \{(x_1, y_1), \dots, (x_m, y_m)\} \subset X \times Y.$$

Требуется построить классификатор $a : X \rightarrow Y$ — отображение, приближающее целевую функцию y на множестве X . При $|Y| > 2$ задачу классификации будем называть многоклассовой. Задачей классификации временных рядов будем называть задачу классификации, в которой объектами классификации являются временные ряды.

Задание метрики, или функции расстояния [7], на парах временных рядов позволяет применять метрические методы классификации. При удачном выборе метрики классификация может производиться простейшими метрическими алгоритмами классификации, например методом ближайшего соседа [8]. Данный подход к решению задачи классификации временных рядов чрезвычайно распро-

*Работа выполнена при финансовой поддержке РФФИ (проект 16-37-00485).

¹Московский физико-технический институт, Сколковский институт науки и технологий, karasikov@phystech.edu

²Вычислительный центр им. А. А. Дородницына Федерального исследовательского центра «Информатика и управление» Российской академии наук, strizov@ccas.ru

странен в силу того, что позволяет свести исходную задачу классификации временных рядов к задаче выбора метрики.

Второй подход к решению задачи классификации состоит в построении для каждого временного ряда его информативного признакового описания $f : X \rightarrow \mathbb{R}^n$, позволяющего строить точные классификаторы с хорошей обобщающей способностью. Построение информативного пространства признаков исходных объектов множества X , позволяющего добиться заданной точности классификации и значительно упрощающего последующий анализ, является важнейшим этапом решения задачи классификации. Признаки могут задаваться экспертом. Так, в работе [9] предлагается использовать в качестве признаков статистические функции (среднее, отклонения от среднего, коэффициенты эксцесса и др.). Стоит заметить, что при таком подходе к построению пространства признаков часто удается добиться необходимого качества классификации путем выбора соответствующих конкретной задаче признаков (см., например, [10]), а сам выбор признаков становится важной технической задачей. Другой метод построения пространства признаков заключается в задании параметрической регрессионной или аппроксимирующей модели временного ряда. Тогда в качестве признаков временных рядов будут выступать параметры настроенной модели. В работе [11] в качестве признаков предлагается использовать коэффициенты дискретного преобразования Фурье (DFT) и дискретного вейвлет-преобразования (DWT), а в [12, 13] — модели авторегрессии. Таким образом, при данном методе построения признаковых описаний возникает задача выбора аппроксимирующей модели временного ряда.

В работе исследуются методы классификации временных рядов, использующие в качестве их признаковых описаний параметры аппроксимирующих моделей. Приводится сравнение моделей аппроксимации. Из временного ряда могут извлекаться сегменты — его подпоследовательности, для которых признаковые описания строятся так же, как и для исходных временных рядов. Использование подпоследовательностей позволяет обобщить алгоритмы классификации. Так, в работе [1] предлагается алгоритм классификации временных рядов методом голосования их случайных сегментов (непрерывных подпоследовательностей со случайным начальным элементом). В данной работе предлагается алгоритм классификации временных рядов в пространстве параметров распределений признаков их сегментов, который сравнивается с родственным ему алгоритмом голосования сегментов [1]. В разд. 7 приводятся результаты экс-

периментов на реальных данных, показывающие высокое качество и общность предлагаемого алгоритма в сочетании с методом признаковых описаний временных рядов параметрами аппроксимирующих их моделей.

2 Постановка задачи

Поставим задачу многоклассовой классификации временных рядов в общем виде. Пусть (X, ρ) — метрическое пространство временных рядов, Y — множество меток классов, $\mathcal{D} \subset X \times Y$ — конечная обучающая выборка.

Пусть S — процедура сегментации:

$$S(x) \subset 2^{\mathbf{S}(x)}, \quad (1)$$

где $\mathbf{S}(x)$ — множество всех сегментов временного ряда $x \in X$; $f(S(x)) \in \mathbb{R}^n$ — процедура построения признакового описания набора сегментов; b — алгоритм многоклассовой классификации:

$$b : \mathbb{R}^n \rightarrow Y. \quad (2)$$

Рассмотрим семейство $A = \{a : X \rightarrow Y\}$ алгоритмов классификации вида

$$a = b \circ f \circ S. \quad (3)$$

Пусть задана функция потерь $\mathcal{L} : X \times Y \times Y \rightarrow \mathbb{R}$ и функционал качества

$$Q(a, \mathcal{D}) = \frac{1}{|\mathcal{D}|} \sum_{(x,y) \in \mathcal{D}} \mathcal{L}(x, a(x), y). \quad (4)$$

В качестве методов обучения $\mu(\mathcal{D}) \in A$ будем использовать следующие:

$$\mu_{f,S}(\mathcal{D}) = \hat{b} \circ f \circ S,$$

где \hat{b} — минимизатор эмпирического риска:

$$\hat{b} = \arg \min_b Q(b \circ f \circ S, \mathcal{D}).$$

Оптимальный метод обучения определяется по скользящему контролю:

$$\mu^* = \arg \min_{f,S} \widehat{CV}(\mu_{f,S}, \mathcal{D}),$$

где $\widehat{CV}(\mu, \mathcal{D})$ — внешний критерий качества метода обучения μ ; при этом исходная обучающая выборка \mathcal{D} случайно разбивается r раз на обучающую и контрольную ($\mathcal{D} = \mathcal{L}_1 \sqcup \mathcal{T}_1 = \dots = \mathcal{L}_r \sqcup \mathcal{T}_r$),

$$\widehat{CV}(\mu, \mathcal{D}) = \frac{1}{r} \sum_{v=1}^r Q(\mu(\mathcal{L}_v), \mathcal{T}_v), \quad (5)$$

где

$$Q(a, \mathfrak{T}) = \frac{1}{|\mathfrak{T}|} \sum_{(x,y) \in \mathfrak{T}} \mathbf{1}\{a(x) = y\}. \quad (6)$$

Средняя точность (precision) классификации объектов класса $c \in Y$ оценивается функционалом скользящего контроля (5) с модифицированным функционалом качества Q :

$$Q_c(a, \mathfrak{T}) = \frac{|\{(x,y) \in \mathfrak{T} \mid a(x) = y = c\}|}{|\{(x,y) \in \mathfrak{T} \mid y = c\}|}. \quad (7)$$

3 Сегментация временных рядов

Определение 1. Сегментом временного ряда $x = [x^{(1)}, \dots, x^{(t)}]$ будем называть любую его непрерывную подпоследовательность $s = [x^{(i)}]_{i=t_0}^{t_1}$, $1 \leq t_0 \leq t_1 \leq t$.

Определение 2. Под сегментацией будем понимать отображение временных рядов во множество их сегментов (1).

Примеры

1. Тривиальная сегментация

$$S(x) = \{x\}, \forall x \in X. \quad (8)$$

2. Случайное выделение сегментов некоторой длины ℓ [1].
3. Важным является случай квазипериодичности временного ряда, когда сам ряд состоит из похожих в определенном смысле сегментов, называемых периодами:

$$x = \left[\underbrace{x^{(1)}, \dots, x^{(t_1)}}_{s^{(1)}}, \underbrace{x^{(t_1+1)}, \dots, x^{(t_2)}}_{s^{(2)}}, \dots, \dots, \underbrace{x^{(t_{p-1}+1)}, \dots, x^{(t)}}_{s^{(p)}} \right].$$

Тогда в качестве процедуры сегментации можно взять разбиение на периоды:

$$S(x) = \{s^{(1)}, \dots, s^{(p)}\}.$$

4 Аппроксимирующая модель сегмента временного ряда

Поскольку сегмент временного ряда сам является временным рядом, в этом разделе слово сегмент будем опускать.

Определение 3. Параметрической аппроксимирующей моделью временного ряда x будем называть отображение

$$g : \mathbb{R}^n \times X \rightarrow X. \quad (9)$$

В слово «аппроксимирующая» вкладывается тот смысл, что модель должна приближать временной ряд в пространстве (X, ρ) , т. е. для некоторого $\mathbf{w} \in \mathbb{R}^n$

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

где

$$\rho(\hat{x}, x) < \varepsilon.$$

При этом естественно взять в качестве признакового описания временного ряда x вектор оптимальных параметров его модели.

Определение 4. Признаковым описанием временного ряда x , порожденным параметрической моделью $g(\mathbf{w}, x)$, назовем вектор оптимальных параметров этой модели:

$$\mathbf{w}_g(x) = \arg \min_{\mathbf{w} \in \mathbb{R}^n} \rho(g(\mathbf{w}, x), x). \quad (10)$$

В качестве аппроксимирующих моделей предлагается использовать следующие.

1. **Модель линейной регрессии.** Пусть задан r -компонентный временной ряд (например, время и три пространственные координаты):

$$x = [x^{(1)}, \dots, x^{(t)}],$$

где

$$x^{(k)} = [x_1^{(k)}, \dots, x_r^{(k)}]^T, \quad k = 1, \dots, t.$$

Рассмотрим модель линейной регрессии одной из компонент временного ряда на остальные компоненты как аппроксимирующую модель:

$$g(\mathbf{w}, x) = [\hat{x}^{(1)}, \dots, \hat{x}^{(t)}],$$

где

$$\hat{x}^{(k)} = [x_1^{(k)}, \dots, x_{r-1}^{(k)}, \hat{x}_r^{(k)}]^T, \quad k = 1, \dots, t,$$

$$\underbrace{\begin{bmatrix} \hat{x}_r^{(1)} \\ \vdots \\ \hat{x}_r^{(t)} \end{bmatrix}}_{\hat{\mathbf{x}}_r} = \underbrace{\begin{bmatrix} x_1^{(1)} & \dots & x_{r-1}^{(1)} \\ \vdots & \ddots & \vdots \\ x_1^{(t)} & \dots & x_{r-1}^{(t)} \end{bmatrix}}_{\mathbf{X}} \underbrace{\begin{bmatrix} w_1 \\ \vdots \\ w_{r-1} \end{bmatrix}}_{\mathbf{w}}.$$

Тогда, выбрав в качестве ρ евклидово расстояние, по определению 4 получим признаковое описание объекта x :

$$\begin{aligned} \mathbf{w}_g(x) &= \arg \min_{\mathbf{w} \in \mathbb{R}^n} \|\mathbf{x}_r - \hat{\mathbf{x}}_r\|_2^2 = \\ &= \arg \min_{\mathbf{w} \in \mathbb{R}^n} \|\mathbf{x}_r - \mathbf{X}\mathbf{w}\|_2^2 = (\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T \mathbf{x}_r. \quad (11) \end{aligned}$$

2. Модель авторегрессии AR(p).

Задан временной ряд

$$x = [x^{(1)}, \dots, x^{(t)}], x^{(k)} \in \mathbb{R}, k = 1, \dots, t.$$

Выберем в качестве модели аппроксимации авторегрессионную модель порядка p :

$$g(\mathbf{w}, x) = [\hat{x}^{(1)}, \dots, \hat{x}^{(t)}],$$

где

$$\hat{x}^{(k)} = \begin{cases} x^{(k)}, & k = 1, \dots, p; \\ w_0 + \sum_{i=1}^p w_i x^{(k-i)}, & k = p+1, \dots, t. \end{cases}$$

Далее признаковое описание определяется аналогично случаю линейной регрессии (11).

3. Дискретное преобразование Фурье. Задан временной ряд

$$x = [x^{(0)}, \dots, x^{(t-1)}], x^{(k)} \in \mathbb{C}, k = 0, \dots, t-1.$$

Взяв в качестве аппроксимирующей модели обратное преобразование Фурье

$$g(\mathbf{w}, x) = [\hat{x}^{(0)}, \dots, \hat{x}^{(t-1)}],$$

где

$$\hat{x}^{(k)} = \frac{1}{t} \sum_{j=0}^{t-1} (w_{2j} + iw_{2j+1}) e^{(2\pi i/t)kj},$$

$$k = 0, \dots, t-1, \quad (12)$$

получим, что признаковым описанием временного ряда x является прямое преобразование:

$$\mathbf{w}_g(x) = [w_0, \dots, w_{2t-1}], \quad (13)$$

где

$$w_{2k} + iw_{2k+1} = \sum_{j=0}^{t-1} x^{(j)} e^{-(2\pi i/t)kj},$$

$$k = 0, \dots, t-1.$$

Переписывая (12) в матричном виде, заметим, что, как и в предыдущих случаях, параметры модели \mathbf{w} эквивалентно находятся при помощи линейной регрессии временного ряда на столбцы матрицы Фурье. Выбор лишь некоторых комплексных амплитуд соответствует регрессии временного ряда на соответствующие столбцы матрицы Фурье. Случай дискретного вейвлет-преобразования аналогичен.

Заметим, что в первых двух случаях используются билинейные аппроксимирующие моде-

ли $g(\mathbf{w}, x)$, а в третьем — линейная. Приведенные примеры демонстрируют большую общность построения пространства признаков при помощи моделей типа (9) и решения оптимизационной задачи (10). Вообще говоря, при $|X| \geq 2$ любая процедура построения признаковых описаний $\mathbf{f} : X \rightarrow \mathbb{R}^n$ задается эквивалентно решением оптимизационной задачи (10) при выборе соответствующей пары (g, ρ) .

5. Распределения признаков сегментов

Объединим идеи, изложенные в предыдущих разделах. Согласно аппроксимирующей модели (9) получим для каждого сегмента $s^{(k)} \in S(x) = \{s^{(1)}, \dots, s^{(p)}\}$ временного ряда x его признаковое описание $\mathbf{w}^{(k)} := \mathbf{w}_g(s^{(k)})$, решив оптимизационную задачу (10). Тогда всему набору сегментов $S(x)$ будет соответствовать выборка:

$$\mathbf{F} = (\mathbf{w}^{(1)}, \dots, \mathbf{w}^{(p)}). \quad (14)$$

Примем гипотезу простоты выборки (14).

Гипотеза 1. Выборка $\mathbf{F} = (\mathbf{f}^{(1)}, \dots, \mathbf{f}^{(p)})$ — простая, т. е. случайная, независимая и однородная, где $\mathbf{w}^{(k)} \sim P_0$.

Пусть имеется параметрическое семейство распределений $\{P_\theta\}_{\theta \in \Theta}$. Будем рассматривать вероятностную модель, в которой объект x зависит от случайного параметра θ .

Гипотеза 2. $p(x|\theta, y) = p(x|\theta)$.

Тогда

$$p(x, y) = p(\mathbf{F}, y) = \int_{\Theta} p(\mathbf{F}, \theta, y) d\theta = \int_{\Theta} p(\mathbf{F}|\theta) p(\theta, y) d\theta.$$

При этом распределение $p(\theta, y)$ предлагается оценивать на этапе обучения, где признаковыми описаниями объектов x_i задачи классификации являются оценки параметров θ_i :

$$\hat{\theta}_i = T(x_i) = T(\mathbf{F}_i).$$

Получив оценку $\hat{p}(\theta, y)$, находим оценку плотности $\hat{p}(x, y)$:

$$\hat{p}(x, y) = \int_{\Theta} p(\mathbf{F}|\theta) \hat{p}(\theta, y) d\theta,$$

по которой строится байесовский классификатор.

В алгоритмической постановке задачи классификации получим $\hat{p}(y|\theta) = \delta(a(\theta), y)$ и

$$\hat{p}(\theta, y) = \delta(a(\theta), y)p(\theta).$$

Тогда

$$\begin{aligned} \hat{p}(x, y) &= \int_{\Theta} p(\mathbf{F}|\theta) \hat{p}(\theta, y) d\theta = \\ &= \int_{\Theta} p(\mathbf{F}|\theta) \delta(a(\theta), y) p(\theta) d\theta = \\ &= \int_{a^{-1}(y)} p(\mathbf{F}|\theta) p(\theta) d\theta = \int_{a^{-1}(y)} p(\theta|\mathbf{F}) p(\mathbf{F}) d\theta. \end{aligned}$$

Приближая распределение $p(\theta|\mathbf{F})$ вырожденным $\delta(\theta - T(\mathbf{F}))$, получим

$$\begin{aligned} \hat{p}(y|x) &= \int_{a^{-1}(y)} p(\theta|\mathbf{F}) d\theta = \int_{a^{-1}(y)} \delta(\theta - T(\mathbf{F})) d\theta = \\ &= \delta(a(T(\mathbf{F})), y). \end{aligned}$$

Таким образом, задача классификации временных рядов свелась к задаче классификации оценок параметров распределений семейства $\{P_{\theta}\}_{\theta \in \Theta}$.

В качестве оценок параметров θ предлагается брать оценки максимального правдоподобия:

$$\begin{aligned} \hat{\theta} = T(x) &= \arg \max_{\theta \in \Theta} \mathcal{L}(\theta | x) = \arg \max_{\theta \in \Theta} p(\mathbf{F}|\theta) = \\ &= \arg \max_{\theta \in \Theta} \prod_k p(\mathbf{w}^{(k)} | \theta). \end{aligned}$$

Заметим, что в частном случае тривиальной сегментации (8) и семейства вырожденных распределений оценка $\hat{\theta}$ является исходным признаковым описанием. Таким образом, предложенный подход к построению признакового описания временного ряда

$$\mathbf{f} : x \mapsto \hat{\theta}$$

является достаточно общим и при этом хорошо интерпретируется.

6 Алгоритм классификации

Для завершения построения классификатора временных рядов (3) построим многоклассовый классификатор b (2) по обучающей выборке $\{(\mathbf{f}(x), y) | (x, y) \in \mathcal{D}\}$.

Сведем задачу многоклассовой классификации к задачам бинарной классификации при помощи стратегий One-vs-All и One-vs-One.

В данной работе для решения задач бинарной классификации, где $Y = \{-1, +1\}$, берутся различные модификации SVM (support vector machine).

7 Вычислительный эксперимент

Вычислительный эксперимент проводился на данных для задачи классификации типов физической активности человека.

7.1 Датасет WISDM

Датасет (набор данных) WISDM [5] содержит показания акселерометра для 6 видов человеческой активности. Необработанные данные, представляющие собой последовательность размеченных показаний акселерометра (по тройке чисел на каждый отсчет времени с интервалом в 50 мс), были разбиты на временные ряды длиной по 200 отсчетов (10 с). Распределение полученных временных рядов по классам приведено в табл. 1.

Таблица 1 Распределение временных рядов по классам. Набор данных WISDM

| Классы | Число объектов |
|----------------------------|----------------|
| 1. Jogging (бежит) | 1624 |
| 2. Walking (идет) | 2087 |
| 3. Upstairs (поднимается) | 549 |
| 4. Downstairs (спускается) | 438 |
| 5. Sitting (сидит) | 276 |
| 6. Standing (стоит) | 231 |

7.1.1 Ручное выделение признаков

Выбор признаков. Каждая компонента временного ряда описывалась ее средним, стандартным отклонением, средним модулем отклонения от среднего, гистограммой с 10 областями равной ширины. Полученные признаки для каждой компоненты объединялись, и к ним добавлялся признак средней величины ускорения. Таким образом, каждый временной ряд описывался 40 признаками.

Классификатор. Задача многоклассовой классификации сводилась к задаче бинарной классификации при помощи подхода One-vs-One. В качестве бинарного классификатора использовался SVM с RBF (radial basis function) ядром и параметрами $C = 8,5$ и $\gamma = 0,12$.

Результаты. На диаграмме рис. 1 демонстрируется качество классификации при усреднении по $r = 50$ случайным разбиениям исходной выборки на тестовую и контрольную в пропорции 7 к 3.

Как видно из табл. 2, классы 2, 3 и 4 недостаточно хорошо отделяются друг от друга.

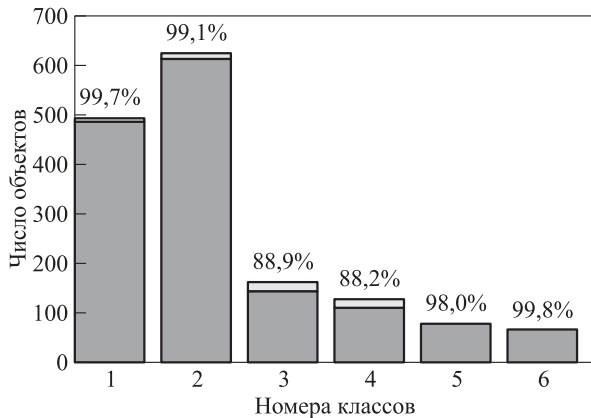


Рис. 1 Набор данных WISDM. Средняя точность 0,9726 — вычисляется по формуле (6). Средние точности классификации для каждого класса вычисляются по формуле (7)

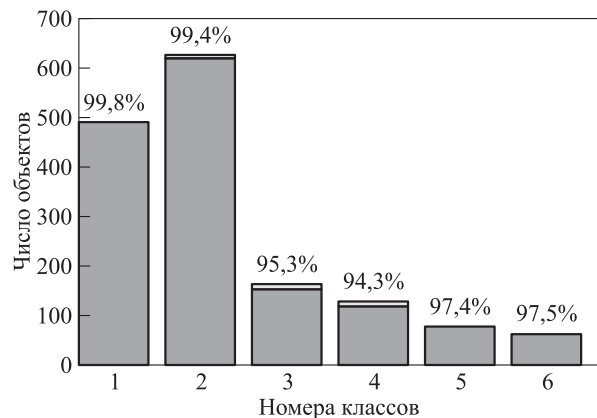


Рис. 2 Точность классификации для параметров модели авторегрессии в качестве признаков описаний

Таблица 2 Усредненная матрица неточностей. Ручное выделение признаков. Набор данных WISDM

| Класс объекта | Предсказанный класс | | | | | |
|---------------|---------------------|-------------|-------------|-------------|-------------|-------------|
| | 1 | 2 | 3 | 4 | 5 | 6 |
| 1 | 1,00 | 0,00 | 0,00 | 0,00 | 0,00 | 0,00 |
| 2 | 0,00 | 0,99 | 0,01 | 0,00 | 0,00 | 0,00 |
| 3 | 0,03 | 0,04 | 0,89 | 0,04 | 0,00 | 0,00 |
| 4 | 0,02 | 0,05 | 0,05 | 0,88 | 0,00 | 0,00 |
| 5 | 0,01 | 0,00 | 0,00 | 0,00 | 0,98 | 0,00 |
| 6 | 0,00 | 0,00 | 0,00 | 0,00 | 0,00 | 1,00 |

Таблица 3 Усредненная матрица неточностей. Признаки, порожденные моделью авторегрессии. Набор данных WISDM

| Класс объекта | Предсказанный класс | | | | | |
|---------------|---------------------|-------------|-------------|-------------|-------------|-------------|
| | 1 | 2 | 3 | 4 | 5 | 6 |
| 1 | 1,00 | 0,00 | 0,00 | 0,00 | 0,00 | 0,00 |
| 2 | 0,00 | 0,99 | 0,00 | 0,00 | 0,00 | 0,00 |
| 3 | 0,01 | 0,02 | 0,95 | 0,02 | 0,00 | 0,00 |
| 4 | 0,00 | 0,02 | 0,04 | 0,94 | 0,00 | 0,00 |
| 5 | 0,01 | 0,00 | 0,00 | 0,00 | 0,97 | 0,01 |
| 6 | 0,01 | 0,00 | 0,00 | 0,00 | 0,01 | 0,97 |

7.1.2 Модель авторегрессии

Признаковое описание. Во втором эксперименте в качестве признаков описаний временных рядов использовались все статистические функции, что брались в первом эксперименте, за исключением гистограммы, вместо которой использовалось 7 коэффициентов модели авторегрессии AR(6) (см. (2)). Таким образом, каждый временной ряд описывался 31 числом. Также проводилась предварительная нормализация признаков.

Классификатор. Задача многоклассовой классификации сводилась к задаче бинарной классификации при помощи подхода One-vs-All. В качестве бинарного классификатора использовалась SVM с RBF-ядром и параметрами $C = 8$ и $\gamma = 0,8$.

Результаты. На диаграмме рис. 2 и в табл. 3 показано качество классификации при усреднении по $r = 50$ случайным разбиениям исходной выборки на тестовую и контрольную в отношении 7 к 3.

Несмотря на неравномерное распределение объектов по классам, использование признакового

описания, порожденного моделью авторегрессии, позволяет значительно повысить качество классификации. Точность построенного классификатора минимальна для 4-го класса — Downstairs — и составляет 94,3%.

7.2 Датасет USC-HAD

Датасет USC-HAD [14] содержит показания акселерометра для 12 типов физической активности человека:

- 1) walk forward (идет вперед);
- 2) walk left (идет влево);
- 3) walk right (идет вправо);
- 4) go upstairs (подъем по лестнице);
- 5) go downstairs (спуск по лестнице);
- 6) run forward (бежит вперед);
- 7) jump up and down (делает прыжок);
- 8) sit and fidget (сидит);
- 9) stand (стоит);

- 10) sleep (спит);
- 11) elevator up (поднимается в лифте);
- 12) elevator down (спускается в лифте).

Выборка содержит примерно по 70 шестикомпонентных временных рядов для каждого класса, а средняя длина временного ряда — 3300. Частота записи измерений сенсора 100 Гц.

7.2.1 Модель авторегрессии и Фурье

Признаковое описание. Исходные временные ряды приводились к частоте 10 Гц при помощи осреднения.

В качестве признаков описаний преобразованных временных рядов брались статистические функции, описанные в п. 7.1.1, за исключением гистограммы. Также для каждой компоненты отдельно и для модуля результирующего ускорения и поворота добавлялось по 11 параметров авторегрессионной модели AR(10) (см. (2)). Затем проводилась нормализация признаков и добавлялись коэффициенты Фурье (13) с индексами 3–12. Таким образом, каждый 6-компонентный временной ряд описывался 128 признаками.

Классификатор. Задача многоклассовой классификации сводилась к задаче бинарной классификации при помощи подхода One-vs-One. В качестве бинарного классификатора использовалась SVM с RBF-ядром и параметрами $C = 10$ и $\gamma = 0,13$.

Результаты. На диаграмме рис. 3 показано качество классификации при усреднении по $r = 500$ случайным разбиениям исходной выборки на тестовую и контрольную в отношении 7 к 3.

Из табл. 4 видно, что использование коэффициентов Фурье значительно повысило качество классификации. Хуже всего класс 8 (sit and fidget) отделяется от класса 9 (stand). Точность классификации для него составляет 92,2%.

7.2.2 Классификация голосованием и классификация в пространстве распределений параметров

Рассмотрим алгоритм классификации в сочетании с процедурой сегментации временных рядов. В качестве процедуры сегментации $S(x)$ (см. (1)) будем использовать выделение сегментов фиксированной длины. Решим задачу классификации для первых 10 классов (за исключением «elevator up» и «elevator down», которые плохо отделяются друг от друга при малой длине сегментов) двумя алгоритмами.

В алгоритме голосования классификатор $b : \mathbb{R}^n \rightarrow Y$ обучается на новой обучающей выборке для сегментов исходных временных рядов

$$\mathcal{D}_S = \{(\mathbf{w}_g(s), y) : (x, y) \in \mathcal{D}, s \in S(x)\}.$$

Далее производится голосование $\hat{y} = \arg \max_y \sum_{s \in S(x)} 1 [b(\mathbf{w}_g(s)) = y]$.

Алгоритм классификации в пространстве гиперпараметров (распределений параметров аппрокси-

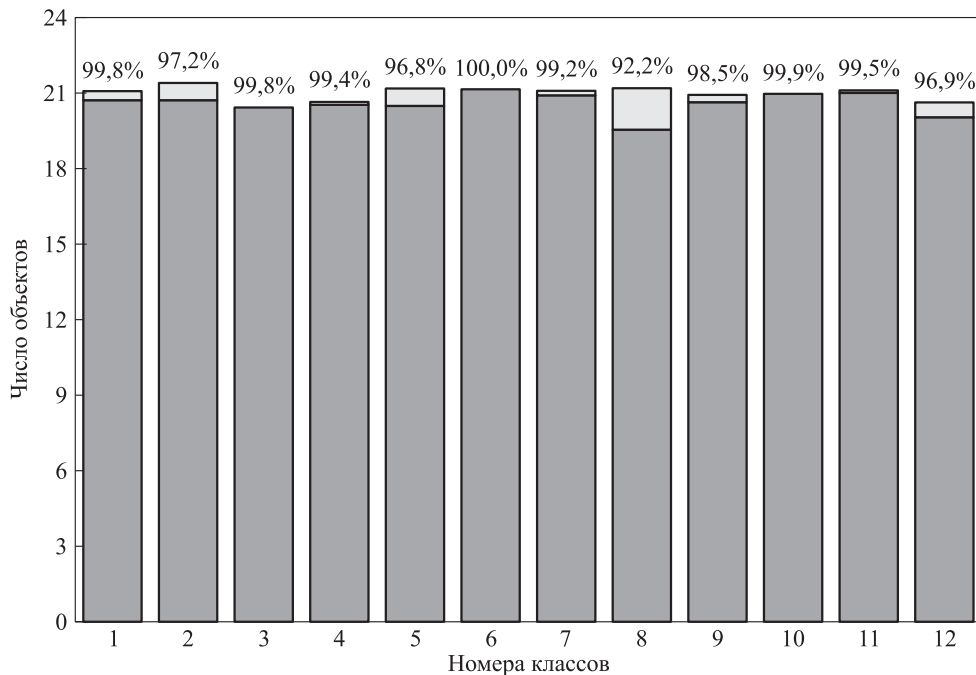


Рис. 3 Точность классификации для параметров модели авторегрессии в качестве признаков описаний

Таблица 4 Усредненная матрица неточностей. Признаки, порожденные моделью авторегрессии. Набор данных USC-HAD

| Класс объекта | Предсказанный класс | | | | | | | | | | | |
|---------------|---------------------|-------------|-------------|-------------|-------------|-------------|-------------|-------------|-------------|-------------|-------------|-------------|
| | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 |
| 1 | 0,99 | 0,00 | 0,00 | 0,00 | 0,01 | 0,00 | 0,00 | 0,00 | 0,00 | 0,00 | 0,00 | 0,00 |
| 2 | 0,01 | 0,97 | 0,01 | 0,00 | 0,01 | 0,00 | 0,00 | 0,00 | 0,00 | 0,00 | 0,00 | 0,00 |
| 3 | 0,00 | 0,00 | 1,00 | 0,00 | 0,00 | 0,00 | 0,00 | 0,00 | 0,00 | 0,00 | 0,00 | 0,00 |
| 4 | 0,00 | 0,00 | 0,00 | 0,99 | 0,01 | 0,00 | 0,00 | 0,00 | 0,00 | 0,00 | 0,00 | 0,00 |
| 5 | 0,00 | 0,00 | 0,00 | 0,01 | 0,97 | 0,02 | 0,00 | 0,00 | 0,00 | 0,00 | 0,00 | 0,00 |
| 6 | 0,00 | 0,00 | 0,00 | 0,00 | 0,00 | 1,00 | 0,00 | 0,00 | 0,00 | 0,00 | 0,00 | 0,00 |
| 7 | 0,00 | 0,00 | 0,00 | 0,00 | 0,00 | 0,00 | 0,99 | 0,00 | 0,00 | 0,00 | 0,00 | 0,00 |
| 8 | 0,00 | 0,00 | 0,00 | 0,00 | 0,00 | 0,00 | 0,00 | 0,92 | 0,08 | 0,00 | 0,00 | 0,00 |
| 9 | 0,00 | 0,00 | 0,00 | 0,00 | 0,00 | 0,00 | 0,00 | 0,01 | 0,99 | 0,00 | 0,00 | 0,00 |
| 10 | 0,00 | 0,00 | 0,00 | 0,00 | 0,00 | 0,00 | 0,00 | 0,00 | 0,00 | 1,00 | 0,00 | 0,00 |
| 11 | 0,00 | 0,00 | 0,00 | 0,00 | 0,00 | 0,00 | 0,00 | 0,00 | 0,00 | 0,00 | 1,00 | 0,00 |
| 12 | 0,00 | 0,00 | 0,00 | 0,00 | 0,00 | 0,00 | 0,00 | 0,00 | 0,02 | 0,00 | 0,01 | 0,97 |

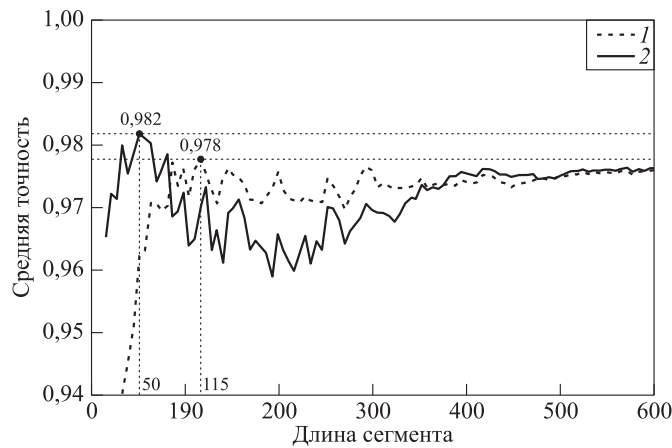


Рис. 4 Зависимость средней точности классификации от длины сегментов: 1 — голосование; 2 — гиперпараметры. Набор данных USC-HAD, первые 10 классов. Точность классификации вычисляется по формуле (6)

мирующих моделей был описан в разд. 5). В эксперименте использовалось семейство нормальных распределений с диагональной ковариационной матрицей.

Задача многоклассовой классификации решалась при помощи подхода One-vs-One бинарными классификаторами SVM с RBF-ядром и параметрами $C = 100$ и $\gamma = 0,017$.

На графике рис. 4 приведены результаты для средней точности решения задачи многоклассовой классификации обоими алгоритмами.

Из графика можно видеть, что оба алгоритма позволяют повысить качество классификации, причем алгоритм классификации в пространстве гиперпараметров при длине сегмента 50 достигает качества 98,2% и показывает результат выше, чем алгоритм голосования.

Объединим результаты из последних двух экспериментов. Будем обучать два классификатора. Пер-

вый классификатор a_1 — One-vs-One SVM с RBF-ядром и параметрами $C = 10$ и $\gamma = 0,13$ — будет разделять классы 11, 12 и первые десять классов для исходных временных рядов. Второй классификатор a_2 — One-vs-One SVM с RBF-ядром и параметрами $C = 100$ и $\gamma = 0,017$ — классификатор в пространстве гиперпараметров, описанный в предыдущем эксперименте.

Итоговый классификатор выглядит следующим образом:

$$a(x) = \begin{cases} a_1(x), & a_1(x) \in \{11, 12\}; \\ a_2(x) & \text{иначе.} \end{cases} \quad (15)$$

Результаты. На диаграмме рис. 5 и в табл. 5 демонстрируется качество классификации построенного классификатора (15) при усреднении по $r = 500$ случайным разбиениям исходной выборки на тестовую и контрольную в отношении 7 к 3.

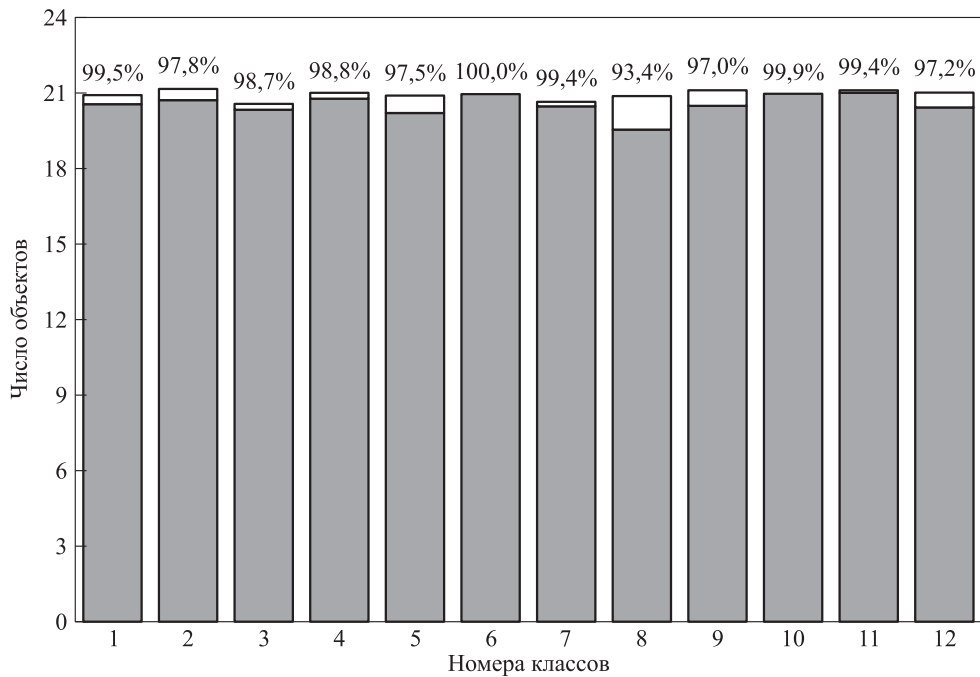


Рис. 5 Точность классификации для гиперпараметров в качестве признаков описаний. Набор данных USC-HAD

Таблица 5 Усредненная матрица неточностей. Признаки — гиперпараметры. Набор данных USC-HAD

| Класс объекта | Предсказанный класс | | | | | | | | | | | |
|---------------|---------------------|-------------|-------------|-------------|-------------|-------------|-------------|-------------|-------------|-------------|-------------|-------------|
| | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 |
| 1 | 1,00 | 0,00 | 0,00 | 0,00 | 0,00 | 0,00 | 0,00 | 0,00 | 0,00 | 0,00 | 0,00 | 0,00 |
| 2 | 0,01 | 0,98 | 0,01 | 0,00 | 0,00 | 0,00 | 0,00 | 0,00 | 0,00 | 0,00 | 0,00 | 0,00 |
| 3 | 0,00 | 0,00 | 0,99 | 0,01 | 0,00 | 0,00 | 0,00 | 0,00 | 0,00 | 0,00 | 0,00 | 0,00 |
| 4 | 0,00 | 0,00 | 0,00 | 0,99 | 0,01 | 0,00 | 0,00 | 0,00 | 0,00 | 0,00 | 0,00 | 0,00 |
| 5 | 0,01 | 0,01 | 0,00 | 0,00 | 0,97 | 0,01 | 0,00 | 0,00 | 0,00 | 0,00 | 0,00 | 0,00 |
| 6 | 0,00 | 0,00 | 0,00 | 0,00 | 0,00 | 1,00 | 0,00 | 0,00 | 0,00 | 0,00 | 0,00 | 0,00 |
| 7 | 0,00 | 0,00 | 0,00 | 0,00 | 0,00 | 0,00 | 0,99 | 0,00 | 0,00 | 0,00 | 0,00 | 0,00 |
| 8 | 0,00 | 0,00 | 0,00 | 0,00 | 0,00 | 0,00 | 0,00 | 0,93 | 0,06 | 0,00 | 0,00 | 0,00 |
| 9 | 0,00 | 0,00 | 0,00 | 0,00 | 0,00 | 0,00 | 0,00 | 0,03 | 0,97 | 0,00 | 0,00 | 0,00 |
| 10 | 0,00 | 0,00 | 0,00 | 0,00 | 0,00 | 0,00 | 0,00 | 0,00 | 0,00 | 1,00 | 0,00 | 0,00 |
| 11 | 0,00 | 0,00 | 0,00 | 0,00 | 0,00 | 0,00 | 0,00 | 0,00 | 0,00 | 0,00 | 0,99 | 0,00 |
| 12 | 0,00 | 0,00 | 0,00 | 0,00 | 0,00 | 0,00 | 0,00 | 0,00 | 0,02 | 0,00 | 0,01 | 0,97 |

8 Заключение

В работе показано, что метод признакового описания временного ряда оптимальными параметрами аппроксимирующих его моделей дает высокое качество решения задачи классификации. Предложенный метод вычислительно эффективен и не требователен к памяти вычислительного устройства.

В работе также предложен алгоритм классификации временных рядов в пространстве распределений параметров моделей, порождающих их сегменты. Он обобщает предыдущий метод клас-

сификации временных рядов и позволяет производить более тонкую настройку алгоритма классификации.

Литература

1. Geurts P., Wehenkel L. Segment and combine approach for non-parametric time-series classification // Knowledge discovery in databases: PKDD 2005. — Berlin—Heidelberg: Springer, 2005. P. 478—485.
2. Esling P., Agon C. Time-series data mining // ACM Comput. Surv., 2012. Vol. 45. No. 1. Article 12. P. 1—34.

3. Basil T., Lakshminarayan C. Automatic classification of heartbeats // 22nd European Signal Processing Conference Proceedings, 2014. P. 1542–1546.
4. Alomari M. H., Samaha A., AlKamha K. Automated classification of l/r hand movement eeg signals using advanced feature extraction and machine learning // Int. J. Adv. Comput. Sci. Appl., 2013. Vol. 4. No. 6. P. 207–212.
5. Kwapisz J. R., Weiss G. M., Moore S. A. Activity recognition using cell phone accelerometers // ACM SigKDD Explorations Newsletter, 2011. Vol. 12. No. 2. P. 74–82.
6. Gruber C., Coduro M., Sick B. Signature verification with dynamic rbf networks and time series motifs // 10th Workshop (International) on Frontiers in Handwriting Recognition. La Baule, 2006. P. 455–460.
7. Ding H., Trajcevski G., Scheuermann P., Wang X., Keogh E. Querying and mining of time series data: Experimental comparison of representations and distance measures // Proc. VLDB Endow, 2008. Vol. 1. No. 2. P. 1542–1552. doi:10.14778/1454159.1454226.
8. Jeong Y. S., Jeong M. K., Omitaomu O. A. Weighted dynamic time warping for time series classification // Pattern Recogn., 2011. Vol. 44. No. 9. P. 2231–2240. doi:10.1016/j.patcog.2010.09.022.
9. Nanopoulos A., Alcock R., Manolopoulos Y. Feature-based classification of time-series data // Int. J. Comput. Res., 2001. Vol. 10. P. 49–61.
10. Wiens J., Horvitz E., Guttag J. V. Patient risk stratification for hospital-associated c. diff as a time-series classification task // Adv. Neur. Inform. Proc. Syst., 2012. Vol. 25. P. 467–475.
11. Mörchen F. Time series feature extraction for data mining using dwt and dft, 2003. Unpubl.
12. Kini B. V., Sekhar C. C. Large margin mixture of ar models for time series classification // Appl. Soft Comp., 2013. Vol. 13. No. 1. P. 361–371.
13. Кузнецов М. П., Ивкин Н. П. Алгоритм классификации временных рядов акселерометра по комбинированному признаковому описанию // Машинное обучение и анализ данных, 2015. Т. 1. № 11. С. 1471–1483.
14. Zhang M., Sawchuk A. A. USC-HAD: A daily activity dataset for ubiquitous activity recognition using wearable sensors // ACM Conference (International) on Ubiquitous Computing Workshop on Situation, Activity and Goal Awareness. — Pittsburgh, PA, USA, 2012.

Поступила в редакцию 10.05.16

FEATURE-BASED TIME-SERIES CLASSIFICATION

M. E. Karasikov^{1,2} and V. V. Strijov³

¹Moscow Institute of Physics and Technology, 9 Institutskiy Per., Dolgoprudny, Moscow Region 141700, Russian Federation

²Skolkovo Institute of Science and Technology, Skolkovo Innovation Center, Building 3, Moscow 143016, Russian Federation

³A. A. Dorodnicyn Computing Center, Federal Research Center “Computer Science and Control” of the Russian Academy of Sciences, 44-2 Vavilov Str., Moscow 119333, Russian Federation

Abstract: The paper is devoted to the multiclass time series classification problem. The feature-based approach that uses meaningful and concise representations for feature space construction is applied. A time series is considered as a sequence of segments approximated by parametric models, and their parameters are used as time series features. This feature construction method inherits from the approximation model such unique properties as shift invariance. The authors propose an approach to solve the time series classification problem using distributions of parameters of the approximation model. The proposed approach is applied to the human activity classification problem. The computational experiments on real data demonstrate superiority of the proposed algorithm over baseline solutions.

Keywords: time series; multiclass classification; time series segmentation; hyperparameters of approximation model; autoregressive model; discrete Fourier transform

DOI: 10.14357/19922264160413

Acknowledgments

The work was supported by the Russian Foundation for Basic Research (project 16-37-00485).

References

1. Geurts, P., and L. Wehenkel. 2005. Segment and combine approach for non-parametric time-series classification. *Knowledge Discovery in Databases: PKDD 2005*. Berlin–Heidelberg: Springer. 478–485.
2. Esling, P., and C. Agon. 2012. Time-series data mining. *ACM Comput. Surv.* 45(1):12:1–12:34.
3. Basil, T., and C. Lakshminarayan. 2014. Automatic classification of heartbeats. *22nd European Signal Processing Conference Proceedings*. 1542–1546.
4. Alomari, M. H., A. Samaha, and K. AlKamha. 2013. Automated classification of l/r hand movement eeg signals using advanced feature extraction and machine learning. *Int. J. Adv. Comput. Sci. Appl.* 4(6):207–212.
5. Kwapisz, J. R., G. M. Weiss, and S. A. Moore. 2011. Activity recognition using cell phone accelerometers. *ACM SigKDD Explorations Newsletter* 12(2):74–82.
6. Gruber, C., M. Coduro, and B. Sick. 2006. Signature verification with dynamic rbf networks and time series motifs. *10th Workshop (International) on Frontiers in Handwriting Recognition*. La Baule. 455–460.
7. Ding, H., G. Trajcevski, P. Scheuermann, X. Wang, and E. Keogh. 2008. Querying and mining of time series data: Experimental comparison of representations and distance measures. *Proc. VLDB Endow* 1(2):1542–1552. doi: 10.14778/1454159.1454226.
8. Jeong, Y.S., M.K. Jeong, and O.A. Omitaomu. 2011. Weighted dynamic time warping for time series classification. *Pattern Recogn.* 44(9):2231–2240. doi: 10.1016/j.patcog.2010.09.022.
9. Nanopoulos, A., R. Alcock, and Y. Manolopoulos. 2001. Feature-based classification of time-series data. *Int. J. Comput. Res.* 10:49–61.
10. Wiens, J., E. Horvitz, and J. V. Guttag. 2012. Patient risk stratification for hospital-associated c. diff as a time-series classification task. *Adv. Neur. Inform. Proc. Syst.* 25:467–475.
11. Mörchen, F. 2003. Time series feature extraction for data mining using dwt and dft. Unpubl.
12. Kini, B. V., and C. C. Sekhar. 2013. Large margin mixture of ar models for time series classification. *Appl. Soft Comp.* 13(1):361–371.
13. Kuznetsov, M. P., and N. P. Ivkin. 2015. Algoritm klasifikatsii vremennykh ryadov akselerometra po kombinirovannomu priznakovomu opisaniyu [Time series classification algorithm using combined feature description]. *Mashinnoe obuchenie i analiz dannykh* [Machine Learning and Data Analysis] 1(11):1471–1483.
14. Zhang, M., and A. A. Sawchuk. 2012. USC-HAD: A daily activity dataset for ubiquitous activity recognition using wearable sensors. *ACM Conference (International) on Ubiquitous Computing Workshop on Situation, Activity and Goal Awareness*. Pittsburgh, PA.

Received May 10, 2016

Contributors

Karasikov Mikhail E. (b. 1992) — student, Moscow Institute of Physics and Technology, 9 Institutskiy Per., Dolgoprudny, Moscow Region 141700, Russian Federation; student, Skolkovo Institute of Science and Technology, Skolkovo Innovation Center, Building 3, Moscow 143016, Russian Federation; karasikov@phystech.edu

Strijov Vadim V. (b. 1967) — Doctor of Science in physics and mathematics, leading scientist, A. A. Dorodnicyn Computing Center, Federal Research Center “Computer Science and Control” of the Russian Academy of Sciences, 44-2 Vavilov Str., Moscow 119333, Russian Federation; strijov@ccas.ru