

Метрическая классификация временных рядов с выравниванием относительно центроидов классов

Гончаров Алексей Владимирович

Московский физико-технический институт

*Курс: Численные методы обучения по прецедентам
(практика, В. В. Стрижов)/Группа 274, весна 2015*

Цели исследования

Исследовать модель метрической классификации, основанной на DTW-методе и сравнить ее с моделью разделяющей классификации.

Проблема

- Разность между значениями временного ряда не может описывать расстояния между двумя объектами.
- Матрица попарных расстояний имеет большую размерность.

Было предложено

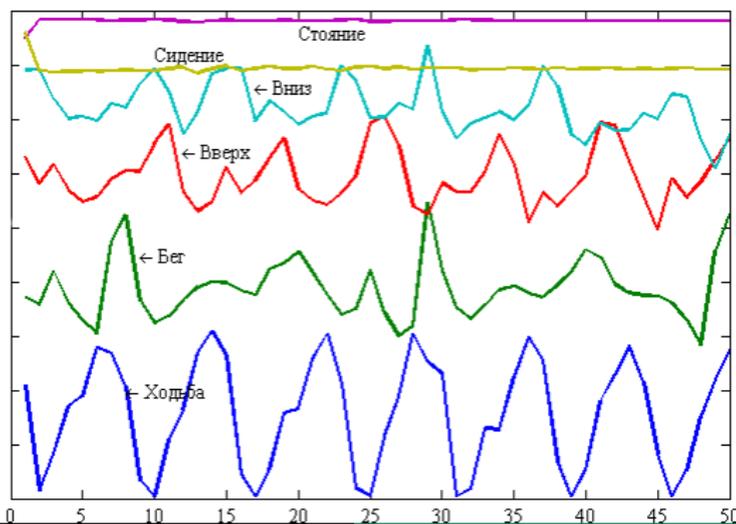
В качестве признаков объекта использовать расстояния, вычисленные с помощью DTW-метода, от этого объекта до других с предварительным выделением эталонных объектов каждого класса.

- *Kwapisz J. R. Weiss G. M. Moore S.* Activity recognition using cell phone accelerometers // SIGKDD Explorations, 2010. Vol.12, No.2. Pp.74–82.
- *Попова М. С., Стрижов В. В.* Выбор оптимальной модели классификации физической активности по измерениям акселерометра // Информатика и ее применения, 2015. Т.9. Вып 1. С. 79–89.
- *Berndt D. J., Clifford J.* Using dynamic time warping to find patterns in time series // In KDD Workshop, 1994.
- *Keogh E. J., Ratanamahatana C. A.* Exact indexing of dynamic time warping. // Knowl. Inf. Syst., 2005. Vol. 7, No. 3.
- *Kwapisz J. R.* 2010. Данные из акселерометра.

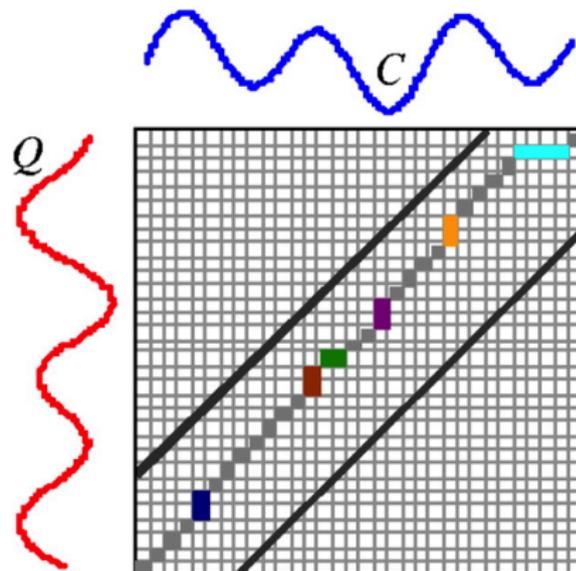
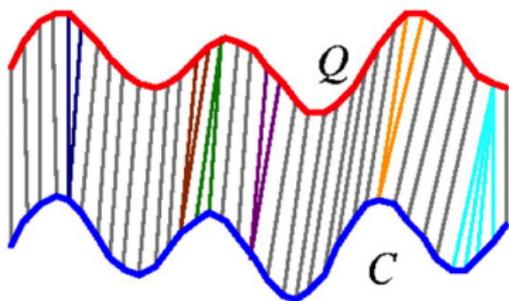
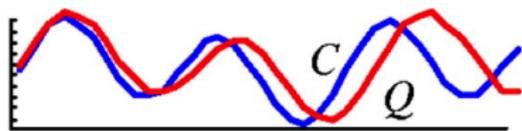
[http : //sourceforge.net/p/mlalgorithms/TSLearning/data/preprocessedla](http://sourceforge.net/p/mlalgorithms/TSLearning/data/preprocessedla)

Постановка задачи

- $\mathcal{D} = \{(s_i, y_i)\}_{i=1}^m$, $s_i \in \mathbb{R}^n$ — множество временных рядов. Выборка разбита на обучающую \mathcal{D}_l и контрольную \mathcal{D}_t .
- Каждому временному ряду s_i поставлена в соответствие метка класса $y_i \in Y$. Множество классов Y конечно.
- $f = \underset{f}{\operatorname{argmin}} (S(f(s_i), y_i | \mathcal{D}_l)) = \underset{f}{\operatorname{argmin}} (\frac{1}{|\mathcal{D}_l|} \sum_{i=1}^{|\mathcal{D}_l|} [f(s_i) \neq y_i])$.



Выравнивание временных рядов



Определим матрицу $\Omega^{n \times n} : \Omega_{ij} = |s_{1i} - s_{2j}|$.

- Путь π между s_1 и s_2 :

$$\pi = \{\pi_k\} = \{(i_k, j_k)\}, \quad k = 1, \dots, K, \quad i, j \in \{1, \dots, n\} :$$

- $\pi_1 = (1, 1)$ и $\pi_K = (n, n)$
- Если $\pi_k = (a, b)$ и $\pi_{k-1} = (p, q)$, то $a - p \leq 1$, $b - q \leq 1$.
- И хотя бы одно из условий: $a - p \geq 1$, $b - q \geq 1$.
- Стоимость π : $\text{Cost}(s_1, s_2, \pi) = \sum_{(i,j) \in \pi} \Omega_{ij}$.
- Путь наименьшей стоимости (выравнивающий путь) $\hat{\pi}$:

$$\hat{\pi} = \underset{\pi}{\operatorname{argmin}} \text{Cost}(s_1, s_2, \pi).$$

- Центроидом множества векторов $\mathcal{D}_k = \{\mathbf{s}_i | y_i = k\}_{i=1}^m$ по расстоянию ρ назовем вектор $\mathbf{z} \in \mathbb{R}^n$ такой, что:

$$\mathbf{z} = \operatorname{argmin}_{\mathbf{z} \in \mathbb{R}^n} \left(\sum_{\mathbf{s}_i \in \mathcal{D}_k} \rho(\mathbf{s}_i, \mathbf{z}) \right).$$

Рассмотрим произвольное множество векторов $s = \{\mathbf{s}_i\}_{i=1}^n$.

$$\mathbf{z} = \operatorname{argmin}_{\mathbf{s} \in s} \left(\sum_{i=1}^n \left| \left[\frac{1}{n} \sum_{j=1}^n \rho(W^T \mathbf{s}_i, W^T \mathbf{s}_j) - \rho(W^T \mathbf{s}_i, W^T \mathbf{s}) \right] \right| \right),$$

где матрица W — ортогональная матрица, проецирующая вектор на пространство первых k главных компонент множества s .

Предлагается применять метод k ближайших соседей.
 $\mathbf{u} \in X$. Пронумеруем $\{\mathbf{x}_1, \dots, \mathbf{x}_l\} = X^l$:

$$\eta(\mathbf{u}, \mathbf{x}_u^{(1)}) \leq \eta(\mathbf{u}, \mathbf{x}_u^{(2)}) \leq \dots \leq \eta(\mathbf{u}, \mathbf{x}_u^{(l)}),$$

где $\eta(\mathbf{u}, \mathbf{x}_u^{(i)})$ — евклидово расстояние по матрице \mathbf{X} .

$$h_u(u, X^l, k) = \operatorname{argmax}_{y \in Y} \sum_{i=1}^k [y_u^{(i)} = y].$$

- Блок выделения центроидов.

$$g : Y \rightarrow \mathcal{D}'.$$

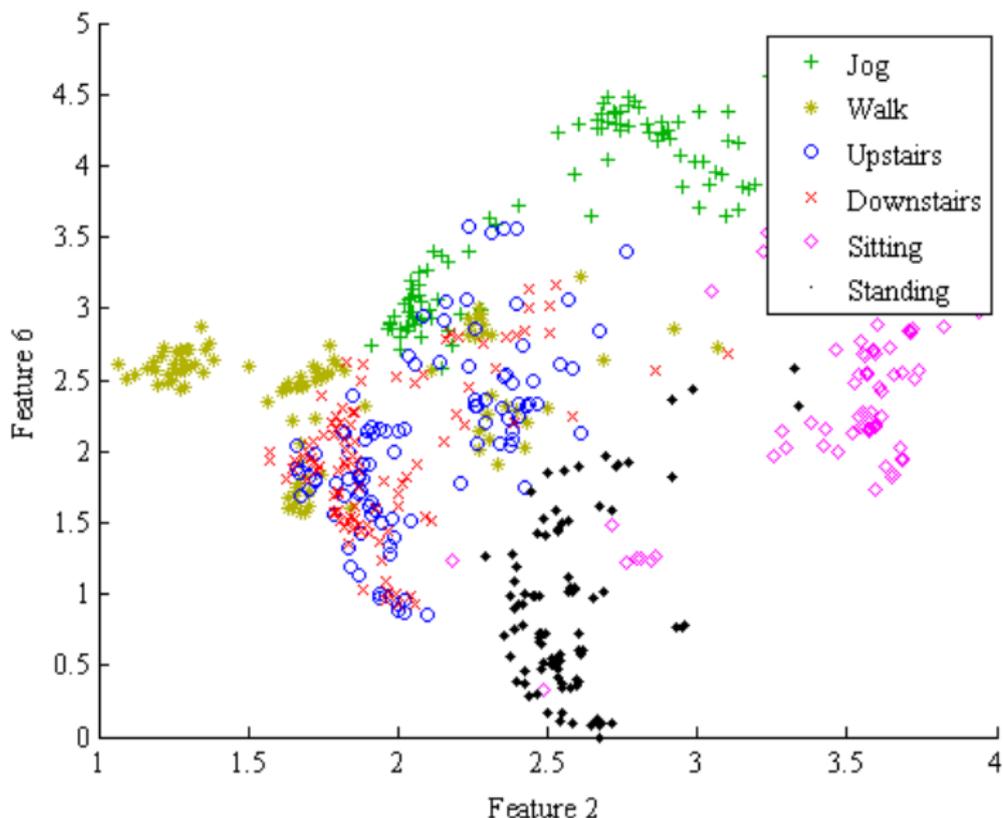
- Блок построения матрицы попарных расстояний.

$$\rho : \mathcal{D} \times \mathcal{D}' \rightarrow \mathbf{X}.$$

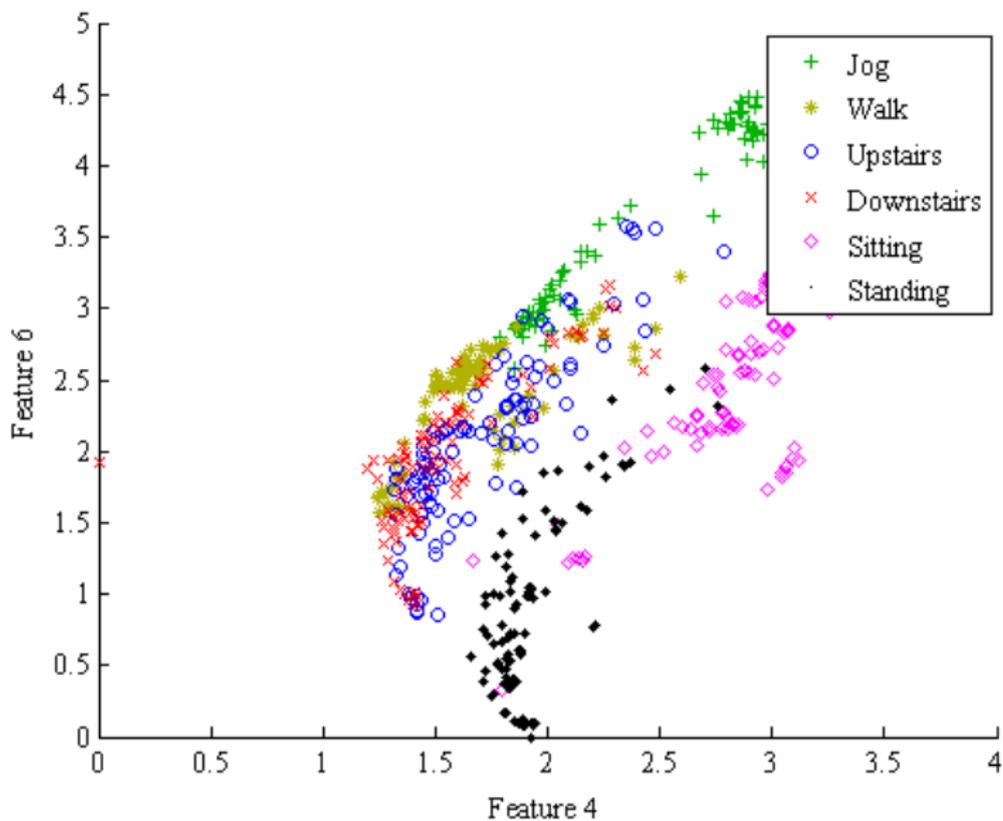
- Блок классификации объекта по построенной матрице.

$$a(\mathcal{D}, k) = h[\rho(\mathcal{D} \times g(Y)), \rho(\mathcal{D}' \times g(Y)), k].$$

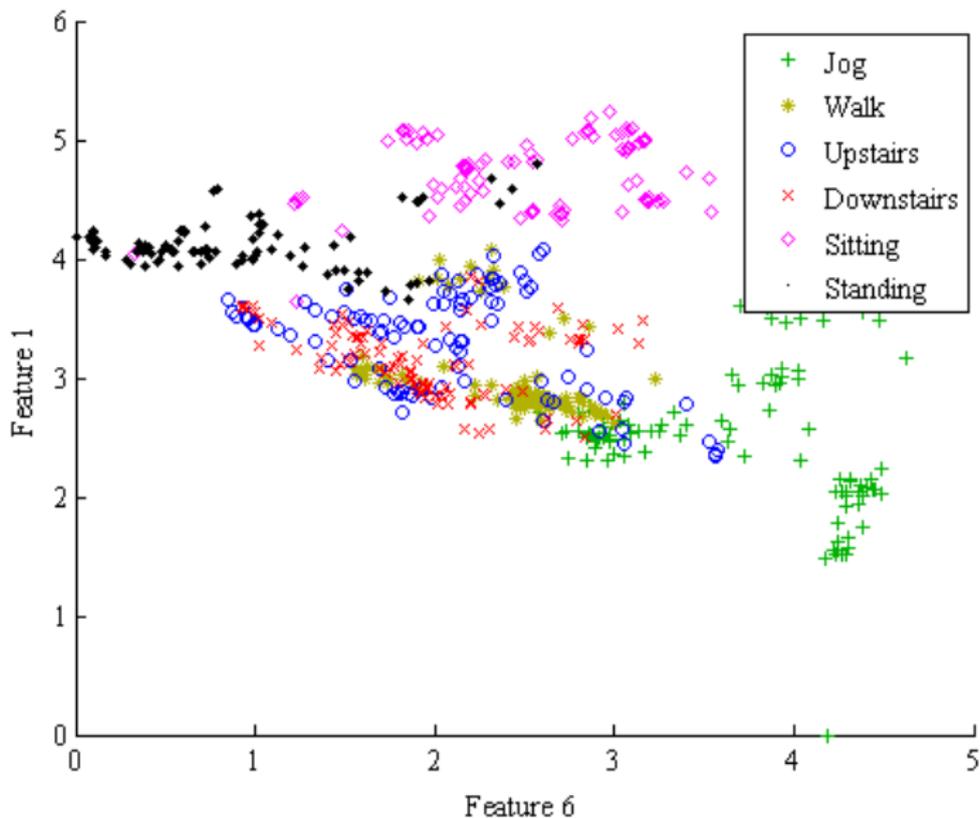
Вычислительный эксперимент



Вычислительный эксперимент



Вычислительный эксперимент



k определяется по критерию скользящего контроля:
 $LOO(k, \mathcal{D}') = \sum_{i=1}^l [h_{u_i}(u_i, \mathcal{D}' \setminus \{u_i\}, k) \neq y_i] \rightarrow \min_{k \in \mathbb{N}}$.

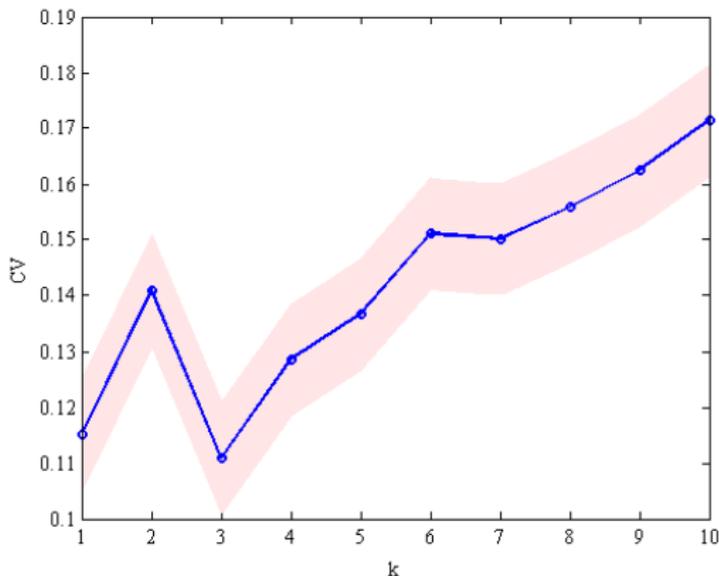


График зависимости величины ошибки от величины k .

В таблице приведены результаты вычислительного эксперимента в сравнении с работой алгоритма разделяющей классификации

	Бег	Ходьба	Вверх	Вниз	Сидение	Стояние
Точность (метр. классиф.), %	98	92	75	88	94	94
Точность (раздел. классиф.), %	98	96	75	72	100	89

- Исследован DTW-метод построения расстояния между временными рядами.
- Исследовано влияние выделения эталонных объектов на алгоритм классификации.
- Выявлены проблемные классы в задаче распознавания движений человека.