Построение метрического пространства многоиндексных временных рядов

Моргачев Глеб Игоревич

Московский физико-технический институт
Факультет управления и прикладной математики
Кафедра интеллектуальных систем

Научный руководитель: д.ф.-м.н. В.В. Стрижов

Цель работы

Построить метрическое пространство с инвариантной функцией, выравнивающей оси временных рядов и провести в нем классификацию сегментов.

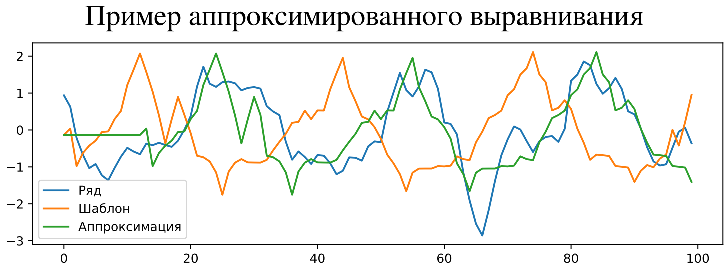
Проблема

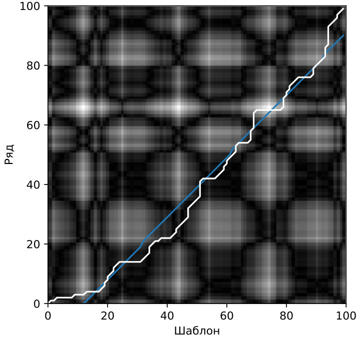
Исследуется проблема сложности нахождения выравнивающего пути для пространственно-временных рядов.

Решение

Предлагается снизить стоимость вычисления за счет наложения ограничений на класс моделей, которому принадлежит выравнивающий путь.

Аппроксимация пути линейной моделью





- модель пути имеет 2 параметра;
- обучение проводится с помощью генетического алгоритма.

Литература

- Rakthanmanon et al. Searching and mining trillions of time series subsequences under dynamic time warping.
  In *Proceedings of the 18th ACM SIGKDD*, 2012

Монотонная нейросеть

- Antoine Wehenkel and Gilles Louppe. Unconstrained monotonic neural networks.
  CoRR, abs/1908.05164, 2019

Полносвязная нейронная сеть, монотонная по одному из входных параметров. Монотонность обеспечивается моделированием производной искомой функции по времени.

Постановка задачи

Задано $\mathfrak{D} = \{(\mathbf{x}_i, y_i), i = 1, \dots N\}$ – N пар объект-ответ.
Объекты $\mathbf{x}_i$ - временные ряды длины n, $\mathbf{x}_i \in \mathbb{R}^n$.
Ответ y_i - метка класса, $y_i \in \{1, \dots, k\}$.

Функция расстояния

Задано множество функций расстояния между парами временных рядов $\mathbf{R} = \{\mathbb{R}^n \times \mathbb{R}^n \to \mathbb{R}_+\}$, выравнивающих оси времени рядов друг относительно друга. Где путь выравнивания является функцией одного из l классов.

Задача

Найти эффективно-вычислимую функции расстояния ρ между временными рядами, оптимальную относительно внешних критериев качества Q_1, Q_2.

$$\rho_i = \underset{\rho \in \mathbf{R}}{\operatorname{argmax}}\, Q_i.$$

Задача метрической классификации

Модель классификации

В пространстве объектов с заданной функией расстояния ρ рассматривается многоклассовая классификация методов k ближайших соседей

$$f_\rho : \mathbb{R}^n \to \{1, \ldots, k\}.$$

Метрика качества

Критерий качества – accuracy:

$$Q_i = \frac{1}{N} \sum_{i=1}^{N} [f_\rho(\mathbf{x}_i) = y_i].$$

Задача классификации в пространстве представлений

Модель классификации

Методом DBA решается промежуточная задача получения шаблона $\mathbf{p}_j \in \mathbb{R}^n,\ j \in \{1, \dots, k\}$ каждого из классов.
В качестве представления объекта $\mathbf{x}_i$ выступает вектор расстояний до каждого из шаблонов:

$$\mathbf{h}_i = [\rho(\mathbf{x}_i, \mathbf{p}_1), \dots, \rho(\mathbf{x}_i, \mathbf{p}_k)] \,.$$

Модель классификации

В пространстве представлений с евклидовой функций расстояния рассматривается классификация методом опорных векторов:

$$f : \mathbb{R}^k \to \{1, \dots, k\}.$$

В качестве метрики качества Q_2 также выступает accuracy.

Определение центроида (DBA)

Для множества временных рядов $\mathbf{X} \subset \mathbb{R}^n$ центроидом по расстоянию ρ называется вектор $\mathbf{c} \in \mathbb{R}^n$, минимизирующия расстояние до всех объектов множества:

$$\mathbf{c} = \operatorname*{argmin}_{\mathbf{c} \in \mathbb{R}^n} \sum_{\mathbf{x} \in \mathbf{X}} \rho(c, \mathbf{x}).$$

В случае DBA функция расстояния ρ порождена алгоритмами выравнивания.

Классы моделей выравнивающих путей

Рассматриваются следующие классы параметризованных моделей, среди которых ищется выравнивающий путь:

- монотонно-возрастающий полином степени k;
- монотонная нейросеть.

Задача локальной аппроксимации шаблоном

Рассматриваются временные ряды $\mathbf{x}$, $\mathbf{y}$. Ряд $\mathbf{y}$ выступает в качестве шаблона, которым аппроксимируется ряд $\mathbf{x}$.

$$\hat{w} = \underset{w}{\operatorname{argmin}} ||\mathbf{x} - \mathbf{y}[g(w, t)]||_2,$$

здесь функция g переводит временную ось ряда $\mathbf{x}$ во временную ось шаблона.

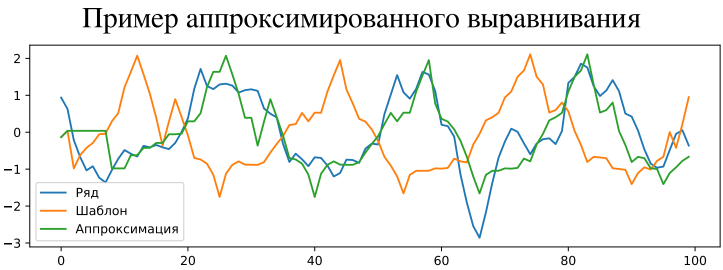
Класс полиномов

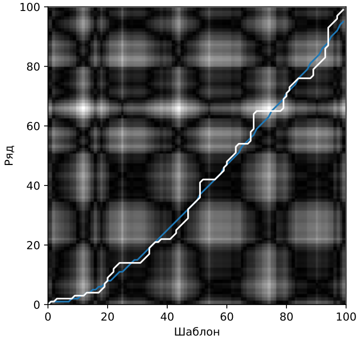
Для выбора выравнивающего пути из класса полиномов степени k, необходимо выбрать $k+1$ параметров.
Их выбор проводится с помощью генетического алгоритма.
Монотонность получаемой функции достигается с помощью отбора монотонных экземпляров из популяции.

Утверждение (Моргачев 2021)

Сложность итерации $\mathcal{O}(qs)$, s – размер популяции, q – количество точек для оценки качества аппроксимации.

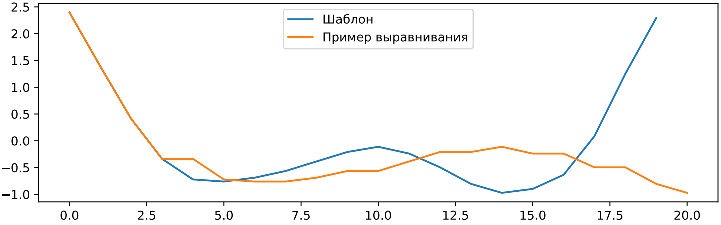
Аппроксимация пути полиномами

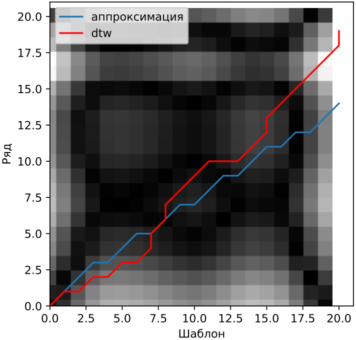




- рассматривается полином порядка 4;
- модель пути имеет 5 параметра;
- обучение проводится с помощью генетического алгоритма.

Пример аппроксимированного выравнивания





- рассматривается трехслойная нейросеть;
- размерность скрытых слоев $[200, 100, 100]$.

Эксперимент

Цель

Оценить потери качества при использовании аппроксимированного DTW в качестве функции расстояния, по сравнению с классическим алгоритмом выравнивания DTW.

Данные

В качестве данных для тестирования использовались следующие выборки:

- распознавание активностей человека по показаниям акселерометра телефона;
- распознавание жестов по показаниям акселерометра контроллера **Wii**;
- распознавание используемых устройств в квартирах по данным о потреблении электроэнергии.

Метрическая классификация

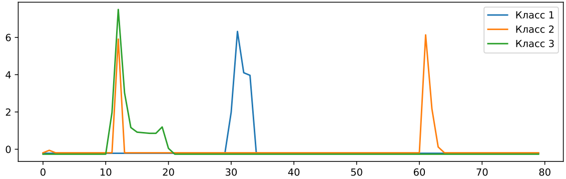
Функция расстояния	Скорость, ит./с	Q_1
Классификация активности человека		
DTW	18.2	$\mathbf{0.71 \pm 0.02}$
DTW_{p_1}	12.2	0.53 ± 0.06
DTW_{p_4}	10.3	0.56 ± 0.04
NN	**90.0**	0.51 ± 0.06
Распознавание жестов		
DTW	2.2	$\mathbf{0.46 \pm 0.04}$
DTW_{p_1}	**4.4**	0.43 ± 0.06
DTW_{p_4}	4.3	0.41 ± 0.04
NN	—	—
Распознавание электроприборов		
DTW	**2.4**	$\mathbf{0.71 \pm 0.02}$
DTW_{p_1}	2.2	0.40 ± 0.06
DTW_{p_4}	2.2	0.29 ± 0.04
NN	—	—

Классификация в пространстве представлений

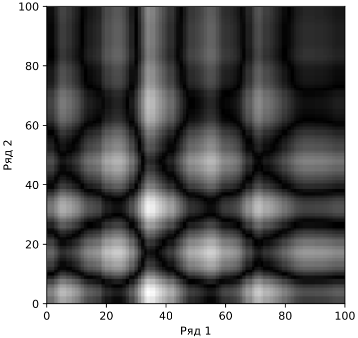
Функция расстояния	Скорость, ит./с	Q_2
Классификация активности человека		
DTW	18.2	$\mathbf{0.61 \pm 0.05}$
DTW_{p_1}	12.2	0.48 ± 0.04
DTW_{p_4}	10.3	0.51 ± 0.02
NN	**90.0**	0.41 ± 0.06
Распознавание жестов		
DTW	1.8	0.38 ± 0.04
DTW_{p_1}	**2.3**	$\mathbf{0.40 \pm 0.03}$
DTW_{p_4}	**2.3**	0.37 ± 0.03
NN	–	–
Распознавание электроприборов		
DTW	**2.2**	$\mathbf{0.44 \pm 0.11}$
DTW_{p_1}	2.1	0.30 ± 0.02
DTW_{p_4}	**2.2**	0.26 ± 0.09
NN	–	–

Анализ ошибки

Объекты из датасета используемых устройств



Расстояний между точками для распознавания жестов



Выносится на защиту

1. Предложена функция расстояния между временными рядами;
2. Изучены различные модели аппроксимации выравнивающего пути между временными рядами;
3. Проведено сравнение предложенных алгоритмов выравнивания в задачах классификации и получения шаблонов классов.