

## System Interpretation of Conceptual Data Model Key Categories

*This paper examines dynamic aspects of entities behavior at domain of interest by means of original entities lifespan models (ELM). After some supplements to extended event driven process chain (eEPC) model the last uses as a basic for ELM. Such ELM enables to determine entities trajectories and objects which compel entities to migrate with the confine of area. Dynamic of entity interprets as property changes due to interaction of an entity with another objects including entities. Any ELM associates with one and only one property which have to be modified. Actions, events and conditions as components of an entity lifespan are connected with objects and timestamps. Thereby all data model categories become interlocking and pursue the aim to derive more detail and exact database. Addition of unified structures for recording ELM and its instances in data model permits to construe some information objects as roles.*

**Keywords:** *entity lifespan, roles, states, actions, timestamps, data model, data structures*

### References

1. **Steimann F.** On the representation of roles in object-oriented and conceptual modeling, *Data and knowledge engineering*, 2000, vol. 35, pp. 83–106.
2. **Jimenez L. J.** REEEM: Reenhancing the entity-relationship model, *Data and knowledge engineering*, 2006, vol. 58, pp. 410–435.
3. **Date C. J., Hugh Darwen, Nikos A.** *Lorentzos Time and relation theory*, Morgan Kaufmann Publishers, 2014, 560 p.
4. **Gregersen H., Jensen C. S.** et al. The *consensus glossery of temporal database concepts*. Temporal Databases: Research and Practice, Springer, 1998, pp. 367–405.
5. **McGinnes S., Kapros E.** Conceptual independence: A design principle for the construction of adaptive information systems, *Information system*, 2015, no. 47, pp. 33–505.
6. **Rodionov A. N.** Kachestvo datalogicheskikh shem. Princip kompaktnosti modelej dannyh i ego prilozheniya II, *Informatika i sistemy upravleniya*, 2012, no. 2, pp. 14–25.
7. **Sheer A.** Modelirovanie biznes-processov, Moscow, Vest'-MetaTehnologija, 2000, 207 p.
8. **Eliferov V. G., Repin V. V.** *Biznes-processy: Reglamentacija i upravlenie*, Moscow, IN-FRA-M, 2011, 319 p.
9. **Vil van der Aalst, Kejs vann Hej.** *Upravlenie potokami rabot: modeli, metody i sistemy*, Moscow, FIZMATLIT, 2007, 316 p.

УДК 519.95

**М. В. Кузнецова**, аспирант,  
Московский физико-технический институт, Москва,  
**В. В. Стрижов**, д-р физ.-мат. наук, вед. науч. сотр.,  
Вычислительный Центр им. А. А. Дородницына ФИЦ ИУ РАН, Москва.

## Локальное прогнозирование временных рядов с использованием инвариантных преобразований\*

*Описан метод построения прогностической модели одномерного временного ряда. Предложено найти сегменты локальной предыстории, похожие на прогнозируемый сегмент. Близкие по заданной функции расстояния сегменты объединены в кластер. Прогностическое значение вычислено путем усреднения значений сегментов кластера. Для повышения качества прогнозирования введено инвариантное преобразование сегментов — преобразование, сохраняющее эквивалентность временных рядов на кластерах. Для преобразования использована функция, построенная с помощью метода динамического выравнивания временных рядов. Предложенный алгоритм проиллюстрирован временными рядами, описывающими движение человека и содержащими показания акселерометра. Построен ретроспективный прогноз временного ряда. В вычислительном эксперименте сравниваются два алгоритма построения прогностической модели: с кластеризацией сегментов временного ряда и с использованием метода ближайшего соседа.*

**Ключевые слова:** *кластеризация, алгоритм ближайшего соседа, путь наименьшей стоимости, динамическое выравнивание, локальное прогнозирование, монотонное преобразование, ретроспективный прогноз, опорный сегмент*

\* Работа выполнена при поддержке РФФИ: проект 16-07-01155.

## Введение

Решается задача построения прогноза временного ряда методом локального прогнозирования [1]. Пусть во временном ряду выделен опорный сегмент — некоторая последняя часть временного ряда. На основании значений этого сегмента необходимо спрогнозировать несколько отсчетов временного ряда. Идея локального метода прогнозирования заключается в следующем. Предполагается, что во временном ряду находятся сегменты, похожие на опорный. Прогноз строится экстраполяцией взвешенного среднего арифметического найденных сегментов.

Предполагается, что найденные сегменты временного ряда, похожие на опорный, имеют различные растяжения и сжатия относительно оси времени. Поэтому для улучшения качества прогноза предлагается ввести инвариантное преобразование, учитывающее трансформации сегментов. Для иллюстрации свойств предлагаемого алгоритма выбраны временные ряды, описывающие движение человека. Процедура построения прогностической модели включает следующие этапы.

1. *Кластеризация.* Сегменты временного ряда кластеризуют [2—4]. Это делается для выделения схожих групп сегментов и построения в каждой из групп инвариантного преобразования.

2. *Введение преобразования.* Для улучшения качества прогноза и учета растяжений и сдвигов сегментов относительно шкалы времени вводится преобразование сегментов. Предполагается, что во временном ряду найдется эталонный сегмент — сегмент, относительно которого можно выровнять все остальные. Задача выбора эталонного сегмента решается в работе [5]. Выравнивающую функцию строят с помощью пути наименьшей стоимости [6]. В качестве эталонного сегмента в данной работе взят центр кластера или опорный сегмент.

3. *Построение прогностической модели.* Прогноз строят как взвешенное среднее найденных преобразованных сегментов.

4. *Оценка ошибки полученного прогноза.* Для оценки ошибки используют понятие ретроспективного прогноза — прогноз строят на период времени, для которого уже имеются исторические значения. Затем полученные прогнозные значения сравнивают с историческими данными [7].

Сегменты временного ряда кластеризуют одним из методов:  $k$ -means;  $C$ -means [5, 8]; с помощью спектральных методов (вычисляется матрица Грамма) [9],  $EM$ -алгоритма и цепей Маркова [10].

Прогностическую модель строят с учетом преобразований, позволяющих выявить похожие во введенной метрике сегменты временного ряда. В работе [11] проведено сравнение различных метрик: евклидовой, взвешенной евклидовой и метрики Минковского. В работах [12, 13] используется путь наименьшей стоимости. В работе [14] рассмотрены понятие инвариантного преобразования и выбор

наиболее подходящего для решения задачи прогнозирования. В работе [15] введено понятие параметрического инвариантного преобразования. В работе [5, 16] с помощью метода динамического выравнивания введена функция сдвига между кривыми и затем рассмотрено выравнивание кривых относительно некоторой эталонной, в целях получения кривой общей формы. В работе [17] рассматривается парное выравнивание кривых относительно друг друга и получение устойчивой оценки для функции деформации.

Новизна работы заключается во введении понятия инвариантного преобразования сегментов временного ряда, построенного с помощью динамического выравнивания временных рядов.

Данная работа состоит из трех частей. Первая часть — это формальная постановка задачи. Во второй части описывается алгоритм преобразования и прогнозирования. Третья часть — вычислительный эксперимент и исследование эффективности алгоритма.

## 1. Постановка задачи и построение прогностической модели

Для временного ряда требуется построить ретроспективный прогноз — прогноз, который строится в прошлое, т. е. для того периода времени, для которого уже имеются исторические значения.

Временной ряд  $X = \{x_t\}_{t=1}^n$  — последовательность, упорядоченная по времени, где  $n$  — длина временного ряда,  $t$  — номер отсчета.

Данные исходного временного ряда делятся на обучающую  $(x_i, t_i)_{i=1}^{n-L}$  и контрольную выборки  $(x_i, t_i)_{i=n-L+1}^n$ . На основании обучающей выборки требуется построить ретроспективно-прогностическую модель  $\tilde{f} = [\tilde{x}_{n-L+1}, \dots, \tilde{x}_n]$ , минимизирующую функцию ошибки:

$$E = \arg \min_{\tilde{x} \in \mathbb{R}^1} \frac{1}{L} \sum_{j=1}^L |x_{n-j} - \tilde{x}_{n-j}|. \quad (1)$$

Разобьем временной ряд на сегменты  $\mathbf{S} = \{\mathbf{s}_\beta\}$ , где у каждого сегмента будет свой номер. Пусть  $\mathbf{s}_0 = \{x_t\}_{t=l}^{l+N}$  — опорный сегмент (сегмент с номером 0), последовательность  $N$  отсчетов временного ряда. Требуется построить прогноз следующих  $h$  отсчетов временного ряда. Для построения прогноза предлагается выбрать  $K+1$  сегментов таких, которые удовлетворяют требованию их компактного расположения среди всевозможных сегментов временного ряда. Для построения прогноза присоединим к каждому сегменту его продолжение, состоящее из  $h$  элементов:  $(\mathbf{s} : \tilde{\mathbf{s}})_k$ , где знак  $:$  обозначает присоединение векторов. Вычислим прогноз сег-

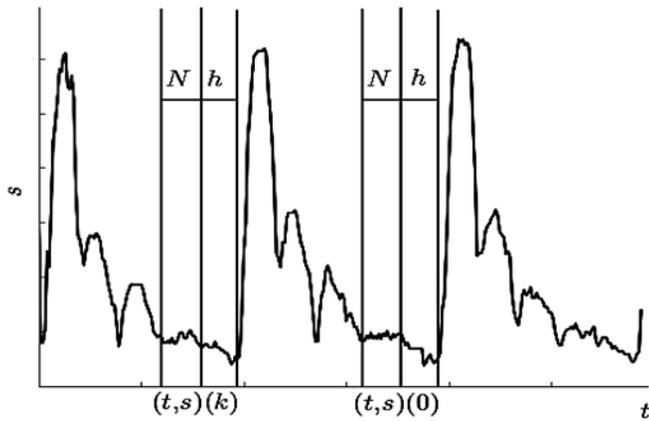


Рис. 1. График, определяемый методом локального прогнозирования

мента с номером  $k = 0$  как усреднение присоединенных  $k$  сегментов:

$$\tilde{s}_0 = \frac{1}{K} \sum_{k=1}^K \tilde{s}_k.$$

В случае, когда для выполнения требования компактного расположения множества сегментов, по которым выполняется прогноз, предполагаются небольшие монотонные преобразования времени, предлагается сделать следующее. Ввести выравнивающую относительно сегмента  $s_0$  функцию  $a$ , и выбирать  $K$  сегментов  $a(s)_k$ , удовлетворяющих требованиям компактности. Рассмотреть соединенные сегменты  $a(s : \tilde{s})_k$ . Прогноз вычислить как усреднение выровненных сегментов:

$$\tilde{s}_0 = \frac{1}{K} \sum_{k=1}^K a(\tilde{s})_k.$$

Рис. 1 иллюстрирует метод локального прогнозирования.

Опишем альтернативные алгоритмы построения прогностической модели.

1. *Многокластерная модель.* Сегменты временного ряда кластеризуются, затем определяется, к какому кластеру относится прогнозируемый сегмент. Находятся ближайшие сегменты в этом кластере. Происходит преобразование сегментов — сегменты выравниваются относительно центра кластера.

2. *Модель ближайшего соседа.* Во временном ряду ищутся сегменты, похожие на опорный сегмент. Найденные сегменты выравниваются относительно опорного сегмента.

Прогноз вычисляется как взвешенное среднее найденных сегментов.

## 2. Кластеризация сегментов

Теперь более подробно опишем каждый из этапов построения прогностической модели. Задано множество сегментов  $S$  и число кластеров  $K$ . Случайным образом выбираются сегменты — исходные центры кластеров  $\{\mu_k^0\}$ ,  $k = 1 \dots K$  (верхний индекс — номер итерации алгоритма). Решим задачу

$$Q = \frac{\sum_{i=1}^m [s_i \in C_k] \rho(s_i, \mu_k)}{\sum_{i=1}^m [s_i \in C_k]} \rightarrow \min,$$

где индикаторная функция  $[\cdot]$  определена как

$$[s_i \in C_k] = \begin{cases} 1, & \text{если } s_i \in C_k; \\ 0, & \text{если } s_i \notin C_k. \end{cases}$$

Выполним следующие шаги: для каждого сегмента  $s_i$  вычислим расстояние  $\rho$  до центра  $\mu_k$  каждого кластера, сегмент относится к тому кластеру, расстояние до центра которого оказалось наименьшим. Затем вычислим центр каждого кластера по формуле

$$\mu_k^h = \frac{1}{|C_k^h|} \sum_{s \in C_k^h} s,$$

где  $C_k^h$  —  $k$ -й кластер на шаге  $h$ , а  $|C_k^h|$  — его мощность. До тех пор, пока не прекратится изменение положения центров кластеров, переносим этот центр кластера  $\mu_k^h$  в найденный вектор  $\mu_k^{h+1}$ . После выполнения алгоритма кластеризации множество сегментов  $S$  разбито на  $K$  кластеров.

## 3. Инвариантное преобразование и путь наименьшей стоимости

Инвариантным преобразованием назовем такое преобразование  $a$  сегмента  $s_i$ , которое сохраняет эквивалентность на кластерах, т. е. если  $s_i \in C_k$  то и  $a(s_i) \in C_k$ .

Во временном ряду задан эталонный сегмент, относительно которого трансформируются остальные сегменты ряда. Сегменты трансформируются с помощью инвариантного преобразования. Выбор эталонного сегмента зависит от динамики временного ряда. Для модели ближайшего соседа эталонным сегментом является опорный. Для построения прогностической модели все сегменты трансформируются относительно опорного сегмента. В многокластерной модели похожие сегменты объединяются в кластер. Сегменты трансформируются относительно эталонного сегмента — центра кластера.

Обозначим  $v$  — эталонный сегмент, а  $\{s_i\}_{i=1}^n$  — сегменты, которые требуется трансформировать относительно эталонного. Каждый сегмент  $s_i$  трансформируем относительно  $v$ . Введем понятие пути наименьшей стоимости между эталонным сегментом  $v$  и сегментом  $s_i$ :

$$s_i = [x_{i1}, \dots, x_{ij}]^T, v = [x_{v1}, \dots, x_{vj}]^T.$$

Зададим матрицу  $\Omega$  с элементами-парами из  $(l \times l)$  — декартова произведения, квадрата множества  $L^2 = \{1, \dots, l\}^2$ . Обозначим путь  $\pi$  в матрице  $\Omega$  — последовательностью

$$\pi = (\pi_1(1), \pi_2(1)), (\pi_1(2), \pi_2(2)), \dots, (\pi_1(N), \pi_2(N)),$$

где  $N$  — длина пути, которая удовлетворяет условию  $l \leq N < 2l - 1$ .

Путь удовлетворяет следующим условиям.

1. *Граничные условия.* Начало и конец пути  $\pi$  находятся на диагонали в противоположных углах  $\Omega$ , т. е.

$$1 = \pi_1(1) \leq \pi_1(2) \leq \dots \leq \pi_1(N),$$

$$1 = \pi_2(1) \leq \pi_2(2) \leq \dots \leq \pi_2(N).$$

2. *Непрерывность.* В шаге пути  $\pi$  участвуют только соседние элементы матрицы, включая соседние по диагонали;  $\pi_1(j+1) \leq \pi_1(j) + 1$  и  $\pi_2(j+1) \leq \pi_2(j) + 1$ .

3. *Монотонность.* Точки  $\pi$  монотонно перемещаются во времени:

$$(\pi_1(j+1) - \pi_1(j)) + (\pi_2(j+1) - \pi_2(j)) \geq 1.$$

*Определение 1.* Стоимостью пути наименьшей стоимости  $\text{Cost}(\pi, \mathbf{s}_i, \mathbf{v})$  между сегментами  $\mathbf{s}_i$  и  $\mathbf{v}$  назовем

$$\text{Cost}(\pi, \mathbf{s}_i, \mathbf{v}) = \frac{1}{N} \sum_{i, i' \in \pi} \Omega_{i, i'}$$

Знаменатель  $N$  нужен для того, чтобы учесть длину пути  $\pi$ .

*Определение 2.* Путь наименьшей стоимости  $\hat{\pi}$  между сегментами  $\mathbf{s}_i$  и  $\mathbf{v}$  — это путь, имеющий наименьшую стоимость среди всех возможных путей между сегментами  $\mathbf{s}_i$  и  $\mathbf{v}$ :

$$\hat{\pi} = \arg \min_{\pi} \text{Cost}(\pi, \mathbf{s}_i, \mathbf{v}).$$

Для получения пути наименьшей стоимости в работе использован метод динамического выравнивания, который рекурсивно находит длину пути наименьшей стоимости по матрице  $\gamma$ , элементы которой определяются следующим образом:

$$\gamma_{i,j} = d_{L_p}(\mathbf{s}_i, \mathbf{v}) + \min(\gamma_{i,j-1}, \gamma_{i-1,j}, \gamma_{i-1,j-1}).$$

*Определение 3.* Назовем проекциями пути наименьшей стоимости  $\pi_x, \pi_y$  отображение подмножества  $\hat{\pi}$  декартова произведения  $L^2$  в его сомножители  $L$ .

Обозначим  $L = \{1 \dots l\}$  — множество индексов сегмента  $\mathbf{s}_i$  и  $P = [1 \dots N]$  — множество индексов пути наименьшей стоимости  $\hat{\pi}$ . Определим  $f$  и  $g$  — сюръективные отображения элементов множества  $P$  в элементы множества  $L$ :

$$f, g : P \rightarrow L.$$

Определим  $F : L \rightarrow 2^P$  как функцию, возвращающую все преобразы индексов пути

$$F(j) = \{p | f(p) = j\}, \text{ для любого } j \in L: F(j) \neq \emptyset.$$

Определим сюръективное отображение  $G : 2^P \rightarrow L$ ,  $G(j) = \{g(p) | p \in j\}$ .

Определим функцию усреднения по значениям полученных индексов

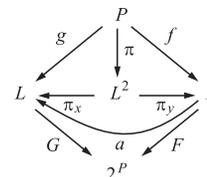
$$\text{avg} : L \rightarrow \mathbb{R},$$

$$\text{где } \text{avg}(G) = \frac{\left( \sum_{j \in G} \mathbf{s}_j \right)}{|G|}.$$

Таким образом, итоговое преобразование  $a$  сегмента  $\mathbf{s}_i$  к эталонному сегменту  $\mathbf{v}$  — композиция:

$$a = \text{avg} \circ G \circ F : \mathbf{s}_i \mapsto \hat{\mathbf{s}}_i.$$

Покажем преобразование  $a$  на следующей коммутативной диаграмме:



## 4. Алгоритмы построения прогностической модели

### 4.1. Многокластерная модель

В каждом кластере  $C_k$  прогностическую модель  $\tilde{f}$  строим путем усреднения преобразованных сегментов.

1. Найдем ближайший к опорному сегменту  $\mathbf{s}$  центр кластера  $\mu_k$ :

$$k^* = \arg \min_{k \in 1 \dots K} \rho(\mu_k, \mathbf{s}).$$

Для этого сегмента требуется построить прогноз.

2. Выполним преобразование  $a$  опорного сегмента  $\mathbf{s}$  и всех сегментов  $\mathbf{s}_{k_i^*}$  в кластере  $C_{k^*}$ .

3. Преобразуемый сегмент находим следующим образом:

$$\tilde{\mathbf{s}}_j = a_s(a_{\hat{\mathbf{s}}_{k_i^*}}(\mathbf{s}_{k_{i+1}^*})).$$

4. Прогноз вычислим как среднее  $m$  найденных преобразованных сегментов  $\tilde{\mathbf{s}}_j$ , где среднее определяем как взвешенное среднее арифметическое. Формула для вычисления весов введена в работе [18]:

$$\tilde{\mathbf{s}} = \frac{\sum_{j=1}^m w_j \mathbf{s}_j}{\sum_{j=1}^m w_j}, \text{ где } w_j = \left( 1 - \frac{\rho_{ij}^2}{\rho_{i_{m+1}}^2} \right)^2, \quad (2)$$

где  $\rho_{i_{m+1}}^2$  — расстояние до  $(m+1)$ -го ближайшего сегмента.

### 4.2. Модель ближайшего соседа

1. Находим во временном ряде  $m$  сегментов  $\mathbf{s}_{i_m}$ , похожих на опорный  $\mathbf{s}$ . В качестве функции расстояния берем стоимость пути наименьшей стоимости.

2. Прогноз вычисляем по формуле (2).

## 5. Вычислительный эксперимент

Проиллюстрируем предложенный алгоритм построения прогностической модели на данных физической активности человека. Цель: сравнить точность построения прогноза с введением и без введения инвариантного преобразования.

Рассмотрим временные ряды, полученные с акселерометра мобильного устройства. Построим прогноз движения человека. В качестве данных возьмем временной ряд, содержащий показания акселерометра. График временного ряда изображен на рис. 2.

Временной ряд состоит из четырех последовательных действий человека: ходьба вперед, подъем, спуск и опять движение вперед. Диапазон измерений используемого акселерометра —  $\pm 6g$ , диапазон частот — 100 Гц. Изначально временной ряд содержал показания акселерометра в трех пространственных направлениях; для удобства была проведена нормировка ряда.

Сравним алгоритмы построения прогностической модели с кластеризацией и с использованием метода ближайшего соседа с алгоритмом, который не использует никаких преобразований. На рис. 3 (см. четвертую сторону обложки) показаны полученные прогнозы с использованием двух алгоритмов (красным и зеленым цветами) и исторические данные (синим цветом), а также прогноз, полученный без применения преобразований (розовым цветом).

Сравним результаты работы алгоритмов построения прогностической модели с использованием кластеризации и с использованием метода ближайшего соседа. На рис. 4 (см. четвертую сторону обложки) показаны полученные прогнозы с использованием двух алгоритмов (красным и зеленым цветами) и исторические данные (синим цветом).

Алгоритм построения прогностической модели с использованием метода ближайшего соседа показал себя точнее (таблица). Построим прогноз временного ряда по опорному сегменту. На рис. 5 (см. четвертую сторону обложки) зеленым выделен

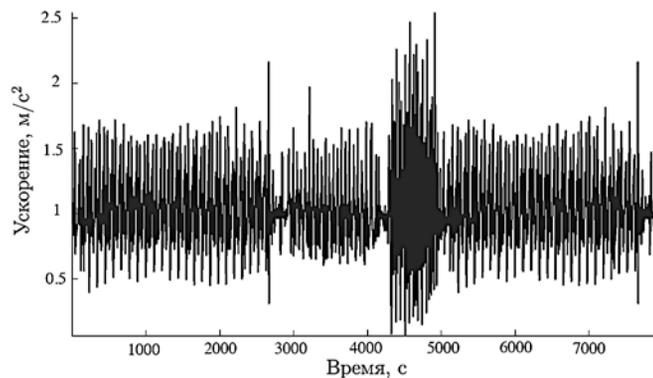


Рис. 2. Графики ходьбы вперед, подъема вверх, спуска вниз, ходьбы вперед

опорный сегмент, синим — исторические данные, красным — полученный прогноз.

На рис. 6 (см. четвертую сторону обложки) показан модуль ошибки (1) на прогнозируемом сегменте по сравнению с историческими данными.

В таблице приведены сравнительные результаты работы алгоритмов построения прогностической модели на исследуемых данных в зависимости от разных длин прогнозируемых сегментов ( $N$  — число экспериментов;  $L$  — средняя длина прогнозируемого сегмента).

Сравнение результатов работы алгоритмов на данных, полученных с акселерометра

Используемый алгоритм	Лучшая MAPE, %	Средняя MAPE, %	$N$	$L$
Многокластерная модель	12,88	15,69	7	100
Модель ближайшего соседа	2,63	5,88	7	100

Точность прогноза зависит от выбора значений параметров — числа кластеров; числа ближайших соседей, длины сегмента, длины предыстории. Оптимальная длина сегмента зависит от природы задачи, например, от частоты данных по времени и характерных времен разных типов процессов. Все эти параметры настраиваются кросс-валидацией в зависимости от данных. Приведенный выше вычислительный эксперимент может быть воспроизведен программным обеспечением [19].

## Заключение

Предложен метод локального прогнозирования временных рядов. Предложено инвариантное преобразование сегментов временного ряда для повышения точности прогноза. Преобразование основано на методе динамического выравнивания временных рядов. Обоснован выбор эталонного сегмента. Реализованы различные алгоритмы построения прогностических моделей с использованием инвариантных преобразований. Алгоритмы протестированы на реальных данных движения человека, сравнительные результаты сведены в таблицу.

## Список литературы

1. **McNames J.** Innovations in local modeling for time series prediction: Ph. D. thesis, Stanford University, 1999. 161 p.
2. **Hartigan J. A., Wong M. A.** Algorithm as 136: A k-means clustering algorithm // Applied statistics. 1979. Vol. 28, N. 1. P. 100—108.
3. **Loochach R., Garg K.** Effect of distance functions on simple k-means clustering problem // International Journal of Computer Applications. 2012. Vol. 49, N. 6. P. 7—9.
4. **Воронцов К. В.** Лекции по алгоритмам кластеризации и многомерного шкалирования. URL: <http://www.ccas.ru/voron/download/Clustering.pdf> (дата обращения: июнь 12, 2014)
5. **Wang K., Gasser T.** Aligment of curves by dynamic time warping // The Annals of Statistics. 1997. Vol. 25, N. 3. P. 1251—1276.
6. **Keogh E. J., Pazzani M. J.** Derivative dynamic time warping // In First SIAM International Conference on Data Mining, 2001.

7. Айвазян С. А., Мхитарян В. С. Прикладная статистика. Основы эконометрики. Т. 2. М.: ЮНИТИ-ДАТА, 2001. 432 с.
8. Abonyi J., Feil B., Nemeth S., Arva P. Fuzzy Clustering Based Segmentation of Time-Series // 5th International Symposium on Intelligent Data Analysis. Berlin. 2003. Vol. 2810. P. 275—285.
9. Jebara T., Song Y., Thadani K. Spectral Clustering and Embedding with Hidden Markov Models // Lecture Notes in Computer Science. 2007. Vol. 4701. P. 164—175.
10. Coviello E., Antoni B. Chan, Gert R. G. Lanckriet. Clustering Hidden Markov Models with Variational HEM // Journal of Machine Learning Research. 2014. Vol. 15. P. 697—747.
11. Варфоломеева А. А. Локальные методы прогнозирования с выбором метрики // Машинное обучение и анализ данных. 2012. Т. 1, №. 3. С. 367—375.
12. Iglesias F., Kastner W. Analysis of Similarity Measures in Times Series Clustering for the Discovery of Building Energy Patterns // Energies. 2013. Vol. 6. P. 579—597.
13. Романенко А. А. Выравнивание временных рядов: прогнозирование с использованием DTW. URL: <http://rlu.ru/5K7n> (дата обращения: июнь 12, 2014).
14. Цыганова С. В. Локальные методы прогнозирования с выбором преобразования // Машинное обучение и анализ данных. 2012. Т. 1, №. 3. С. 311—317.
15. Кононенко Д. С. Оценка параметров инвариантных преобразований в задачах прогнозирования временных рядов. Магистерская диссертация, Московский физико-технический институт, 2013, 24 с. URL: <http://rlu.ru/5K7g> (дата обращения: июнь 22, 2014).
16. Arribas-Gil A., Muller H.-G. Pairwise dynamic time warping for event data // Computational Statistics Data Analysis. 2014. Vol. 69. P. 255—268.
17. Tang R., Muller H.-G. Pairwise curve synchronization for functional data // Biometrika. 2008. Vol. 95, N. 4. P. 875—889.
18. Магнус Я. Р., Катышев П. К., Пересецкий А. А. Эконометрика. Т. 2. М.: Дело, 2014, 576 с.
19. Кузнецова М. В. Алгоритм локального прогнозирования временных рядов. URL: <http://rlu.ru/5K79> (дата обращения: декабрь 25, 2015).

M. V. Kuznetsova<sup>1</sup>, Graduate Student, V. V. Strijov<sup>2</sup>, Leading Researcher

<sup>1</sup> Moscow Institute of Physics and Technology, Moscow;

<sup>2</sup> Computing Centre of the Russian Academy of Sciences, Moscow

## Local Forecasting of Time Series with Invariant Transformations

*The paper describes a univariate time series forecasting model. It proposes to find — segments of local history, which are similar to the forecasted segment. A distance function is used to cluster segments. The forecast is the average of the value of time series from this cluster. To improve the quality of forecast the paper proposes an invariant transformation of segments. This transformation holds the equivalence of time series respect to clusters. The transformation is a function, constructed by the dynamic time warping procedure. The retrospective forecasting procedure calculates the accuracy of the forecasting model. Accelerometer time series of a person's motion are used in computational experiment. It compares two constructing forecasting models. The first one clusters segments, the second one uses k-nearest neighbor algorithm to select similar segments.*

**Keywords:** clustering, k-nearest neighbor algorithm, dynamic time warping, warping path, dynamic aligning, local forecasting, monotonic transformation, retrospective forecast, supported segment

### References

1. McNames J. Innovations in local modeling for time series prediction: Ph. D. thesis, Stanford University, 1999, 161 p.
2. Hartigan J. A., Wong M. A. Algorithm as 136: A k-means clustering algorithm, *Applied statistics*, 1979, vol. 28, no. 1, pp. 100—108.
3. Loochach R., Garg K. Effect of distance functions on simple k-means clustering problem, *International Journal of Computer Applications*, 2012, vol. 49, no. 6, pp. 7—9.
4. Vorontsov K. V. Lektsii po algoritmam klasterizatsii i mnogomernogo shkalirovaniya, URL: <http://www.ccas.ru/voron/download/Clustering.pdf> (accessed: June 12, 2014).
5. Wang K., Gasser T. Aligment of curves by dynamic time warping, *The Annals of Statistics*, 1997, vol. 25, no. 3, pp. 1251—1276.
6. Keogh E. J., Pazzad M. J. Derivative dynamic time warping, *In First SIAM International Conference on Data Mining*, 2001.
7. Aivazjan S. A., Mhitarjan V. S. *Prikladnaya statistika. Osnovy ekonometriki*, Т. 2, Moscow, YuNITI-DATA, 2001. 432 p.
8. Abonyi J., Feil B., Nemeth S., Arva P. Fuzzy Clustering Based Segmentation of Time-Series, *5th International Symposium on Intelligent Data Analysis*, Berlin, 2003, vol. 2810, pp. 275—285.
9. Jebara T., Song Y., Thadani K. Spectral Clustering and Embedding with Hidden Markov Models, *Lecture Notes in Computer Science*, 2007, vol. 4701, pp. 164—175.
10. Coviello E., Antoni B. Chan, Gert R. G. Lanckriet. Clustering Hidden Markov Models with Variational HEM, *Journal of Machine Learning Research*, 2014, vol. 15, pp. 697—747.
11. Varfolomeeva A. A. Lokalnye metody prognozirovaniya s vyborom metriki, *Journal of Machine Learning and Data Analysis*, 2012, vol. 1, 3, pp. 367—375.
12. Iglesias F., Kastner W. Analysis of Similarity Measures in Times Series Clustering for the Discovery of Building Energy Patterns, *Energies*, 2013, vol. 6, pp. 579—597.
13. Romanenko A. A. *Vyiravnivanie vremennykh ryadov: prognozirovanie s ispolzovaniem DTW*, URL: <http://rlu.ru/5K7n> (accessed: June 12, 2014).
14. Tsyganova S. V. Lokalnye metody prognozirovaniya s vyborom preobrazovaniya, *Journal of Machine Learning and Data Analysis*, 2012, vol. 1, 3, pp. 311—317.
15. Kononenko D. S. *Otsenka parametrov invariantnykh preobrazovaniy v zadachah prognozirovaniya vremennykh ryadov*: Masters thesis, Moscow Institute of Physics and Technology, 2013, 24 p. URL: <http://rlu.ru/5K7g> (accessed: June 22, 2014).
16. Arribas-Gil A., Muller H.-G. Pairwise dynamic time warping for event data, *Computational Statistics Data Analysis*, 2014, vol. 69, pp. 255—268.
17. Tang R., Muller H.-G. Pairwise curve synchronization for functional data, *Biometrika*, 2008, vol. 95, no. 4, pp. 875—889.
18. Magnus Ya. R., Kartushev P. K., Peresetskii A. A. *Ekonometrika*, Moscow, Delo, 2004. 576 p.
19. Kuznetsova M. V. *Algorithm of local modelling*, Available at <http://rlu.ru/5K79> (accessed: December 25, 2015).