

УДК 519.2

Т.В. Книгницкая

ОЦЕНКА СХОДСТВА ВРЕМЕННЫХ РЯДОВ НА ОСНОВЕ МОДЕЛЕЙ

Ключевые слова: расстояние между временными рядами по моделям, кластеризация, кластер, временной ряд, модель временного ряда, DTW, ERP.

Введение

Большинство задач, основанных на временных рядах как на математических моделях, посвящено подбору оптимальных моделей для временных рядов. Наряду с этой проблемой за счет развития машинного обучения Big Data и теории случайных процессов все больше внимания уделяется кластеризации временных рядов [1]. Решение данной проблемы чаще всего проводят двумя методами.

1. Наивный — рассматривают временной ряд как точку в R^T , где T — длина временного ряда. При этом кластеризация временного ряда превращается в кластеризацию точек в R^T [2].

2. Кластеризация на основе параметров модели. Для всех временных рядов выбирается одна «базовая» модель, например [3–5]. Следующий шаг — предположение о распределении основных параметров модели и вычисления апостериорных оценок параметров временных рядов. На основе данных параметров и проводится кластеризация.

Недостаток первого метода очевиден, поскольку он не учитывает особенности модели временного ряда — наличие корреляции между временными рядами, наличие детерминированных компонент (тренда, сезонной компоненты), вид автокорреляционной функции, спектральную плотность и др. Основной недостаток второго метода, как вероятностного метода, заключается в том, что модель временного ряда фиксированная и выбор начального распределения параметров выбирается с точки зрения простоты вычисления апостериорного распределения. Хотя вторая проблема для второго алгоритма достаточно легко решается с помощью алгоритмов Метрополиса–Гастинга [1], Гиббса [6], однако для небольших временных рядов неудачный подбор начального распределения параметров приводит к плохому приближению апостериорного распределения. Кроме того, методика данного метода основывается на разбиении временного ряда на три основные части:

$$x_t = p_t + s_t + c_t, \quad (1)$$

где x_t — значение временного ряда, p_t — полиномиальный тренд, s_t — значение сезонной компоненты, c_t — случайная составляющая временного ряда, для которой определяется модель временного ряда. Аналогичное разбиение будет использоваться и в данной работе. Следует отметить, что разбиение ряда на компоненты (1) прослеживается также и во многих других моделях случайных процессов — семимартингалы, полумартингалы и тому подобное.

© Т.В. КНИГНИЦКАЯ, 2019

В настоящей работе рассматриваются стационарные временные ряды: математическое ожидание и дисперсия не меняются со временем — постоянные величины; ковариация зависит только от расстояния между измерениями временного ряда и не зависит от их значений. Если же входные временные ряды нестационарны, с помощью взятия дифференциальных разниц их приводят к стационарным временным рядам [7].

1. Обзор исследований

Способы оценки сходства данных временных рядов представляют интерес для специалистов по анализу данных, поскольку временные ряды являются одними из наиболее широко применяемых математических моделей для описания реальных явлений. Одна из критических исследовательских проблем анализа временных рядов — выбор функции расстояния для определения понятия сходства между двумя временными рядами.

Рассматривая непрерывные функции, приходим к выводу, что наиболее распространенной метрикой для временных рядов является евклидово расстояние:

$$d_E = \sqrt{(\bar{x} - \bar{y})(\bar{x} - \bar{y})'}$$

инвариантное при рассмотрении изменений в порядке представления временных рядов. Это означает, что евклидово расстояние в принципе неприменимо для нахождения расстояний в случае многомерных временных рядов и не учитывает корреляцию между наблюдениями, что является существенным недостатком. Для сравнения данных временных рядов, где тренды и эволюции должны учитываться, или, если форма временного ряда описывается последовательностью функций, актуальным является соотношение Пирсона:

$$d_C(\bar{x} - \bar{y}) = 1 - \frac{(\bar{x} - \bar{y})(\bar{x} - \bar{y})'}{\sqrt{(\bar{x} - \bar{x})(\bar{x} - \bar{x})'} \sqrt{(\bar{y} - \bar{y})(\bar{y} - \bar{y})'}}$$

которое также широко используются, хотя не лишено искажений и проблем [8]. Расстояние Махаланобиса

$$d_M(\bar{x}, \bar{y}) = \sqrt{(\bar{x} - \bar{y})C^{-1}(\bar{x} - \bar{y})'}$$

можно рассматривать как эволюционное евклидово расстояние, которое учитывает корреляцию данных. Это расстояние использует матрицу ковариации входных векторов C для «взвешивания» особенностей временного ряда. Расстояние Махаланобиса [9] обычно успешно работает с большими наборами данных с уменьшенными характеристиками. Более детальное описание расстояний можно найти в [10].

В последних исследованиях по этой тематике разработана серия методов для нахождения расстояний между многомерными временными рядами, которые можно разделить на два класса. Первый класс включает функции, основанные на нормах L_1 и L_2 [11]. Примерами функций этого класса является динамическая трансформация времени (Dynamic Time Warping distance — DTW) [12] и корректировка расстояния со штрафом (Edit Distance with Real Penalty — ERP) [4].

Второй класс функций расстояний включает в себя методы, которые вычисляют оценку сходства за порогом соответствия ε . Примерами этого класса функций является самая длинная общая подпоследовательность (Longest Common Subsequence — LCSS) [13] и корректировка расстояния в реальной последовательности (Edit Distance on Real Sequence — EDR) [14]. Предыдущие исследования [15, 16] показали, что этот второй класс методов является лучшим при наличии шумов и смещения по времени.

Динамическая трансформация времени (DTW) специально предназначена для сравнения временных рядов [17]. Применение DTW позволяет провести нелинейное отображение двух векторов, минимизируя расстояние между ними. Расстояние DTW может использоваться для векторов разной длины: $X = [x_1, x_2, \dots, x_n]$ и $Y = [y_1, y_2, \dots, y_n]$. С помощью DTW определяется матрица расходов C размерности $n \times m$, которая содержит расстояния (обычно евклидовы) между двумя точками: x_i и y_j . Временная трансформация $W = w_1, w_2, \dots, w_K$, где $\max(m, n) \leq K < m + n - 1$, формируются набором матричных компонент на следующих условиях:

- граничные условия: $w_1 = C(1, 1)$ и $w_K = C(n, m)$;
- условия монотонности: дано $w_k = C(a, b)$ и $w_{k-1} = C(a', b')$, $a \geq a'$ и $b \geq b'$;
- условие величины шага: дано $w_k = C(a, b)$ и $w_{k-1} = C(a', b')$, $a - a' \leq 1$ и $b - b' \leq 1$.

Существует много способов, обеспечивающих выполнение указанных условий. Один из них — расстояние DTW, с помощью которого минимизируется значение временной трансформации во временном ряду:

$$d_w(\vec{x}, \vec{y}) = \min \sqrt{\sum_{k=1}^K w_k}. \quad (2)$$

Основным недостатком (2) остается расчет пути минимальной стоимости. К тому же расстояние DTW не может рассматриваться как метрика, потому что оно не удовлетворяет неравенству треугольника.

Обзор показателей метрик расстояний для кластеризации временных рядов можно найти в [18]. Другой важный способ нахождения расстояний — косинус измерения [5], что хорошо подходит для моделей с разной или переменной длиной или методы сходства Жаккара и Танимото [19], которые также интуитивно можно понимать как комбинацию евклидовых расстояний и корреляций.

Итак, рассмотрим временной ряд X , который определяется последовательностью реальных значений (измерений), где каждое значение x_i взято в конкретный момент времени, т.е. $X = [x_1, x_2, \dots, x_n]$. Имея последовательность X , которая описывает временной ряд, можно нормировать временной ряд, используя среднее μ и стандартное отклонение σ :

$$\text{Norm}(X) = \left[\frac{x_1 - \mu}{\sigma}, \frac{x_2 - \mu}{\sigma}, \dots, \frac{x_n - \mu}{\sigma} \right].$$

Нормализация рекомендуется для того, чтобы расстояние между двумя временными рядами было инвариантно для масштабирования амплитуды и смещения временных рядов.

Проблема использования L_1 -нормы для временных рядов, как указано выше, заключается в том, что она требует, чтобы временные ряды были одинаковой длины и не имели локального сдвига по времени. Если необходимо найти расстояние между двумя временными рядами разной длины, то можно добавить измерения в качестве пропущенных значений (gaps) во временном ряду с меньшей длиной или удалить значение в более длинном временном ряду. Тем самым исчезнет проблема равенства длин и локального сдвига по времени.

Рассмотрим расстояние коррективы строки:

$$\text{dist}(y_i, x_i) = \begin{cases} 0, & \text{если } y_i = x_i, \\ 1, & \text{если } y_i \text{ или } x_i \text{ пропущенно} \\ 1 & \text{в противном случае,} \end{cases} \quad (3)$$

где y_i, x_i — элементы строк.

При переходе от последовательностей до временных рядов сложность заключается в том, что элементы y_i и x_i являются не символами, а реальными значениями. Для анализа временных рядов необходимо учитывать корреляции между значениями ряда. Расстояние EDR позволяет учесть реальные значения посредством введения параметра δ , который будем называть порогом:

$$dist_{edr}(y_i, x_i) = \begin{cases} 0, & \text{если } |y_i - x_i| \leq \delta, \\ 1 & \text{в противном случае.} \end{cases} \quad (4)$$

Сформулируем недостатки расстояния EDR: каждая разница между элементами временных рядов равна единице (второй случай в формуле (4)); использование порога δ . Этой проблемы не существует для расстояния DTW, так как используется L_1 -норма между двумя непропущенными элементами.

Расстояние DTW [20] отличается от EDR двумя основными моментами, которые описываются следующей формулой:

$$dist_{dtw}(y_i, x_i) = \begin{cases} |y_i - x_i|, & \text{если } y_i, x_i \text{ не пропущены,} \\ |y_i - x_{i-1}|, & \text{если } x_i \text{ пропущено,} \\ |x_i - y_{i-1}|, & \text{если } y_i \text{ пропущено.} \end{cases} \quad (5)$$

Основная причина, почему DTW не удовлетворяет неравенству треугольника, заключается в том, что когда необходимо интерполировать пропущенное измерение, повторяется (копируется) предыдущий элемент. Таким образом, как показано во втором и третьем случаях формулы (5), разница между элементом и пропущенным значением зависит от y_{i-1} или x_{i-1} .

Рассмотрим расстояние ERP, так как оно использует «штраф» между двумя не пропущенными значениями и постоянной величиной вычисления расстояний для пропущенных значений. Таким образом, метод ERP использует следующую формулу для вычисления расстояния:

$$dist_{erp}(y_i, x_i) = \begin{cases} |y_i - x_i|, & \text{если } y_i, x_i \text{ не пропущены,} \\ |y_i - g|, & \text{если } x_i \text{ пропущено,} \\ |x_i - g|, & \text{если } y_i \text{ пропущено,} \end{cases} \quad (6)$$

где g — некоторая константа, «штраф» за пропущенное значение или усредненное значение соседних значений временного ряда. Константа g может быть выбрана как среднее двух или нескольких соседних значений временного ряда. Специалист по анализу данных с помощью анализа графика временного ряда может сделать вывод, сколько значений необходимо рассмотреть для заполнения пропущенных измерений.

Основываясь на формуле (6), расстояние ERP между двумя временными рядами определяется формулой (9) и обозначается $ERP(Y, X)$. Тщательное сравнение формул (7)–(9) показывает, что расстояние ERP можно рассматривать как комбинацию L_1 -нормы и EDR.

Сложность алгоритмов DTW и ERP составляет $O(NM)$, где N и M — длины двух входных последовательностей. Без потери общности, предполагая, что $N \geq M$, сложность алгоритма составляет $O(N^2)$. Алгоритмы имеют ту же сложность при рассмотрении многомерных временных рядов в пространстве.

$$DTW(Y, X) = \begin{cases} 0, & \text{àñèè } m = n = 0, \\ \infty, & \text{àñèè } n = 0 \text{ èèè } m = 0, \\ dist_{dtw}(y_1, x_1) + A \text{ èâ÷â } [38]; \end{cases} \quad (7)$$

$$A = \min\{DTW(Re\ st(Y), Re\ st(X)), DTW(Re\ st(Y), X), DTW(Y, Re\ st(X))\},$$

$$EDR(Y, X) = \begin{cases} n, & \text{àñèè } m = 0, \\ m, & \text{àñèè } n = 0, \\ EDR(Re\ st(Y), Re\ st(X)), \text{ àñèè } dist_{edr}(y_1, x_1) = 0, \\ B \end{cases} \quad (8)$$

$$B = \min \left\{ \begin{array}{l} EDR(Re\ st(Y), Re\ st(X)) + dist_{edr}(y_1, x_1), \\ EDR(Re\ st(Y), X) + dist_{edr}(y_1, gap), \\ EDR(Y, Re\ st(X)) + dist_{edr}(gap, x_1) \end{array} \right\};$$

$$ERP(Y, X) = \begin{cases} \sum_{i=1}^n |x_i - g|, & \text{àñèè } m = 0, \\ \sum_{i=1}^m |y_i - g|, & \text{àñèè } n = 0, \\ C \end{cases} \quad (9)$$

$$\tilde{N} = \min \left\{ \begin{array}{l} ERP(Re\ st(Y), Re\ st(X)) + dist_{erp}(y_1, x_1), \\ ERP(Re\ st(Y), X) + dist_{erp}(y_1, gap), \\ ERP(Y, Re\ st(X)) + dist_{erp}(gap, x_1). \end{array} \right\}$$

2. Алгоритм поиска оптимальных моделей

Как отмечалось выше, важную роль в исследовании временных рядов играет кластеризация временных рядов. В качестве меры сходства при кластеризации является следующая мера [21]:

$$HM(G_1, \dots, G_m) = \sum_{k=1}^m \frac{2}{n_k - 1} \sum_{i < j \in G_k} d(M_i, M_j), \quad (10)$$

где $d(M_i, M_j)$ — расстояние между моделями временных рядов i и j . Данное расстояние, как основной объект работы, вычислим ниже. Итак, главной проблемой для определения меры сходства $HM(G_1, \dots, G_m)$ является определение расстояния между временными рядами $d(M_i, M_j)$.

2.1. Постановка задачи. Рассмотрим N временных рядов, функционирование которых отслеживается в течение T периодов, т.е. данные имеют следующий вид:

$$\begin{pmatrix} x_{11} & \dots & x_{1T} \\ x_{12} & \dots & x_{2T} \\ \dots & \dots & \dots \\ x_{N1} & \dots & x_{NT} \end{pmatrix}. \quad (11)$$

Базовой моделью временного ряда, как указывалось выше, является $ARMA(p, q)$ -модель:

$$\Phi(L) = \phi_0 + \dots + \phi_p L^p,$$

$$\Theta(L) = \theta_0 + \dots + \theta_q L^q,$$

причем $\phi_0 = \theta_0 = 1$; L — лаговый оператор.

Таким образом, наряду с работой [21], где определялась одна модель для всех временных рядов, в данном случае нужно проверить

$$N_{total} = ((P+1)(Q+1))^N \quad (12)$$

моделей, что обычно не представляется возможным при большом количестве моделей, т.е. при больших P , Q и N . Учитывая вычислительную сложность поставленной задачи, в настоящей работе предложен алгоритм поиска расстояния между двумя временными рядами, основываясь на их моделях, с использованием только $ARMA(P, Q)$ для всех временных рядов.

В большинстве работ «расстояние» между моделями, которое характеризует временные ряды x_1 и x_2 , определяется следующим образом:

$$d(M_1, M_2) = \sqrt{\sum_{t=1}^T (x_{1t} - x_{2t})^2},$$

т.е. рассматривается L_2 -норма.

Данная мера позаимствована из наивного подхода и не учитывает моделей, согласно которым функционируют временные ряды. Для определения расстояния между моделями используется понятие $MA(\infty)$ представления [7]. Согласно утверждению разд. 1 [22] произвольный стационарный временной ряд x_t , заданный с помощью $ARMA(p, q)$ модели (4), однозначно можно записать в виде $MA(\infty)$ -представления:

$$x(t) = \tilde{\Theta}(L) \varepsilon_t, \quad (13)$$

где

$$\tilde{\Theta}(L) = L + \tilde{\theta}_1 L + \dots + \tilde{\theta}_k L^k + \dots, \quad (14)$$

причем для модели, которая задает стационарный временной ряд:

$$\sum_{k=1}^{\infty} |\tilde{\theta}_k|^2 < \infty.$$

Используя представление (14) для $ARMA(p, q)$ -модели, определим расстояние между моделями M_1 и M_2 следующим образом:

$$d(M_1, M_2) = \sum_{k=1}^{\infty} |\tilde{\theta}_k^1 - \tilde{\theta}_k^2|. \quad (15)$$

Для моделирования предыдущей формулы на примере следует отметить, что ряд $\sum_{k=1}^{\infty} |\tilde{\theta}_k^1 - \tilde{\theta}_k^2|$ сходимый. Поэтому можно пренебречь элементами ряда после некоторого K . Данное расстояние будет интерпретироваться именно как расстояние между моделями, а не как расстояние между данными.

2.2. Сравнение расстояний. Для сравнения расстояний между временными рядами в настоящей работе используется метод Монте–Карло [23]. В качестве моделей для сравнения рассмотрим L_1 -расстояние, DTW- и ERP-методы. Основные характеристики, которые обуславливают смену расстояний и правильность определения моделей временного ряда, — наличие выбросов, структурное изменение временного ряда [24] и длина временного ряда. В следующих вычислениях будет исследоваться влияние наличия выбросов и длины временного ряда на погрешность в вычислении между временными рядами.

Рассмотрим влияние наличия аутлайеров на расстояние между временными рядами и расстояниями между их моделями. Для этого рассмотрим генерацию двух временных рядов с $T = 150$ исследованиями, первый из которых моделируется $ARMA(p, q)$ моделью, второй — это тот же временной ряд, в котором $\alpha\%$ значений заменены выбросами*. В качестве оригинального временного ряда взята $ARMA(2, 1)$ модель:

$$x_t = 0,3x_{t-1} - 0,4x_{t-2} + \varepsilon_t + 0,1\varepsilon_{t-1}.$$

Из рис. 1 видно, что при наличии большого количества выбросов, классические методы линейно увеличивают расстояния между временными рядами, в то же время расстояние по моделям ведет себя как логарифмическая функция. При этом следует заметить, что при моделировании рассмотрены небольшие выбросы, которые незначительно выходят за пределы интервала $(Q_1 - 1,5IQR, Q_1 + 1,5IQR)$. При больших выбросах разница между временными рядами будет расти линейно с коэффициентом, равным средней величине выброса.

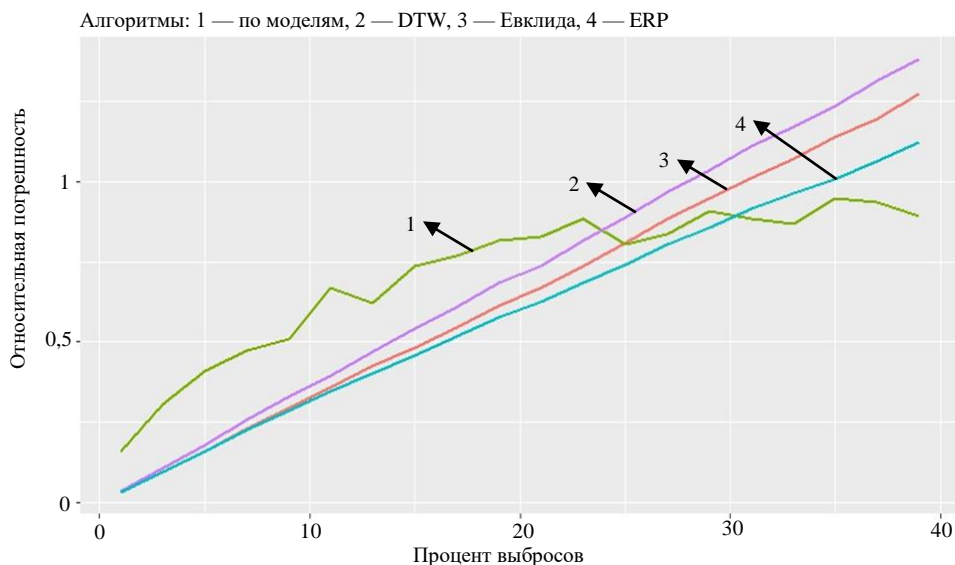


Рис. 1

Наряду с данной зависимостью рассмотрим также зависимость расстояний от величины временного ряда, т.е. от величины T (рис. 2).

* Под выбросом будем понимать значение, выходящее за пределы интервала $(Q_1 - 1,5IQR, Q_1 + 1,5IQR)$, где $IQR = Q_3 - Q_1$, Q_3, Q_1 — 1- и 2-й квартили выборки.

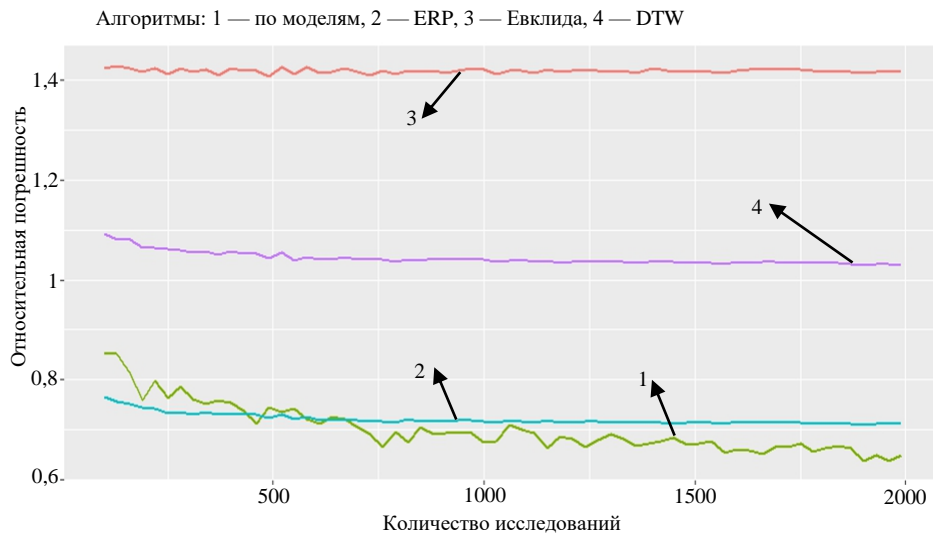


Рис. 2

Таким образом, из рисунка следует, что при больших T расстояния между временными рядами для одинаковых стационарных моделей практически постоянны для всех методов, однако для оценки расстояния по моделям эта константа значительно меньше, чем соответствующая константа для всех других методов.

Заключение

В настоящей работе рассмотрено расстояние между временными рядами, основанное не на данных, а на моделях временных рядов. С помощью метода Монте–Карло, показано, что данное расстояние более устойчиво к выбросам и дает более точные результаты для более длинных временных рядов (при больших T). Кроме того, сложность алгоритма вычисления данного расстояния для N временных рядов составляет $O(T * N^2)$, в то же время аналогичная сложность для алгоритмов DTW, ERP, EDR составляет $O(T^2 * N^2)$. Несомненно, использование моделей — один из наиболее удобных инструментов исследования сходства процессов. Кроме того, для анализа удобно использовать усредненные эволюции и предельные эволюции в схеме диффузионной аппроксимации [25]. Также за счет устойчивости к выбросам данное расстояние можно использовать при кластеризации для построения более устойчивых к шумам кластеров.

Т.В. Книгніцька

ОЦІНКА ПОДІБНОСТІ ЧАСОВИХ РЯДІВ НА ОСНОВІ МОДЕЛЕЙ

Визначення міри відстані між часовими рядами є відправною точкою для багатьох завдань інтелектуального аналізу даних, таких як кластеризація та класифікація. Кластеризація є основним методом навчання без вчителя,

який використовується для розбиття даних на групи на основі внутрішніх і апіорних невідомих характеристик, властивих даним. При розбитті даних на кластери виникає потреба правильного вибору метрики подібності між об'єктами. Описано основні існуючі алгоритми пошуку відстані між часовими рядами, які добре описують дану проблему для малих часових рядів та за відсутності викидів. Викиди, властиві реальним процесам, призводять до неправильної кластеризації, а отже, до прийняття неправильних рішень. Запропоновано розглянути відстань між часовими рядами у вигляді відстані між моделями (ARIMA) даних часових рядів. За наявності великої кількості викидів класичні методи лінійно збільшують відстані між часовими рядами, в той же час запропонована у статті відстань за моделями поводить себе як логарифмічна функція. Показано, що при зростанні кількості вимірювань відносні похибки для всіх класичних методів залишаються майже незмінними. Відносна похибка для оцінки відстані за моделями значно менша та спадає при зростанні кількості вимірювань. Основним досягненням статті є визначення відстані між часовими рядами, яка ґрунтується на понятті моделі, і порівняння цієї відстані з відповідними класичними найбільш вживаними методами. Використовуючи метод Монте–Карло, показано, що запропонована відстань більш стійка до викидів і дає більш точні результати для часових рядів з великою кількістю спостережень. Крім того, складність алгоритму обчислення відстані на основі моделей менша за аналогічну обчислювальну складність існуючих алгоритмів (DTW, ERP, Евкліда). Немає сумнівів, що використання моделей є одним з найбільш зручних інструментів дослідження подібності процесів. Крім того, для аналізу з врахуванням даного алгоритму зручно використовувати усереднені еволюції і граничні еволюції в схемі дифузійної апроксимації. Також за рахунок стійкості до викидів граничних еволюцій введена відстань можна використовувати при кластеризації для побудови більш стійких до шумів кластерів.

Ключові слова: відстань між часовими рядами за моделями, кластеризація, кластер, часовий ряд, модель часового ряду, DTW, ERP.

T.V. Knignitskaya

ESTIMATE OF TIME SERIES SIMILARITY BASED ON MODELS

Determining the measure of the distance between time series is the starting point for many data mining tasks such as clustering and classification. Clustering is the main method of teaching without a teacher, which is used to divide data into groups based on the internal and a priori unknown characteristics inherent in the data. When dividing data into clusters, the need arises to select the similarity metric between objects. The paper describes the main existing algorithms for the “distance” searching between time series, which well describe this problem for small time series and under the absence of outliers. Outliers inherent in real processes lead to improper clustering, and, consequently, to wrong decisions making. It is proposed to consider the distance between time series in the form of the distance between models (ARIMA) of these time series. In the presence of a large number of outliers, classical methods linearly increase the distances between time series, while the distance proposed in the article according to the models behaves as a logarithmic function. It is shown that with an increase in the number of measurements, the relative errors for all classical methods remain almost unchanged. At the same time, the relative error for estimating the distance by the models is much smaller and decreases with an increase in the number of measurements. The main achievement of the article is the determination of the distance between time series, based on the concept of a model, and the comparison of this distance with the corresponding classical methods most commonly used. Using the Monte Carlo method, it has been shown that the proposed distance is more resistant to outliers and gives more accurate results for

time series with a large number of observations. In addition, the complexity of the algorithm for calculating distances based on models is less than the analogous computational complexity of existing algorithms (DTW, ERP, Euclidean distance). There is no doubt that the use of models is one of the most convenient tools for studying the similarity of processes. In addition, for analysis taking into account this algorithm, it is convenient to use the averaged evolutions and the limiting evolutions in the diffusion approximation scheme. Also, due to the resistance to outliers of limiting evolutions, the entered distance can be used in clustering to build more noise-resistant clusters.

Keywords: distance between time series by models, clustering, cluster, time series, time series model, DTW, ERP.

1. Aue A., Horváth L. Structural breaks in time series. *Journal of Time Series Analysis*. 2013. **34**, N 1. P. 1–16.
2. Keogh E., Zhu Q., Hu B., Hao Y., Xi X., Wei L. Ratanamahatana CA: The UCR Time series classification/clustering homepage, 2011.
3. Akaike H. Time series analysis and control through parametric models. *Applied Time Series Analysis*. New York : Academic Press, 1978.
4. Shapiro S.S., Francia R.S. An approximate analysis of variance test for normality. *J. Amer. Stat. Assoc.* 1972. **67**. P. 215–216.
5. Magidson J., Vermunt J.K. Latent class factor and cluster models, bi-plots and related graphical displays. In *Sociological Methodology*. Cambridge: Blackwell. 2001.
6. Liao T.W. Clustering of time series data — A survey. *Pattern Recognit.* 2005. P. 1857–1874.
7. Канторович Г.Г. Анализ временных рядов (курс лекций). *Экономический журнал ВШЭ*. 2003. 1(202). P. 85–116.
8. Rodgers J.L., Nicewander W.A. Thirteen ways to look at the correlation coefficient. *Am. Stat.* 1988. **42**. P. 59–66.
9. Maeschalck R.D., Jouan-Rimbaud D., Massart D. The mahalanobis distance. *Chemom. Intell. Lab. Syst.* 2000. **50**. P. 1–18.
10. Félix Iglesias, Wolfgang Kastner. Analysis of similarity measures in times series clustering for the discovery of building energy patterns. *Energies*. 2013. **6**(2). P. 579–597.
11. Yam Khoon. L1, L2, Kalman filter and time series analysis in deformation analysis, Singapore.
12. Berndt D.J., Clifford J. Using dynamic time warping to find patterns in time series. In *AAAI-94 Workshop on Knowledge Discovery in Databases*. 1994. P. 359–370.
13. Vlachos M., Kollios G., Gunopulos D. Discovering similar multidimensional trajectories. *ICDE*. 2002. P. 673–684.
14. Stramer O., Brockwell P.J., Tweedie R.L. Existence and stability of continuous time threshold ARMA processes. *Statistica Sinica*. 1995.
15. Lei Chen, Raymond Ng. On the Marriage of Lp-norms and Edit Distance. In *VLDB*. 2004. P. 792–803.
16. Lei Chen, M. Tamer Özsu, Vincent Oria. Robust and fast similarity search for moving object trajectories. In *SIGMOD*. 2005. P. 491–502.
17. Pole A., West M., Harrison J. Applied Bayesian Forecasting and Time Series Analysis. *Chapman and Hall*. New York. 1994.
18. Yildirim Ilker. Bayesian inference: Gibbs sampling. *Department of Brain and Cognitive Sciences, University of Rochester*. August, 2012. P. 1–6.
19. Lipkus A. A proof of the triangle inequality for the Tanimoto distance. *J. Math. Chem.* 1999. **26**. P. 263–265.
20. Usue Mori, Mendiburu A., Lozano J. Distance measures for time series in R: The TSdist Package. *The R Journal*. 2016. **8**(2). P. 451–459.
21. Luis E. Nieto-Barajas, Alberto Contreras-Crist. A Bayesian nonparametric approach for time series clustering. *Bayesian Analysis*. 2014. **9**. P. 147–170.
22. Brockwell P.J., Davis R.A. Introduction to time series and forecasting. NY : Springer, 2012. P. 500.
23. Dick J., Kuo F.Y., Peters G.W., Sloan I. Monte Carlo and quasi-Monte Carlo methods 2012. NY : Springer, 2014.

24. Casini A. Structural breaks in time series. *Oxford Research Encyclopedia of Economics and Finance*. 2017. P. 1–38.
25. Tsarkov Ye.F., Yasinsky V.K., Malyk I.V. Stability in impulsive systems with Markov perturbations in averaging scheme. 2. Averaging principle for impulsive Markov systems and stability analysis based on averaged equations. *Cybernetics and Systems Analysis*. 2011. **47**. P. 44–54.
26. Argiento R., Cremaschi A., Guglielmi A. A Bayesian nonparametric mixture model for cluster analysis. Technical report QuadernoImati CNR. Milano: 3-MI, 2012.
27. Berndt D.J., Clifford J. A dynamic programming approach. *In Advances in Knowledge Discovery and Data Mining*. 1996. P. 229–248.
28. Brockwell P.J. Continuous-time ARMA processes. *Handbook of Statistics*. Elsevier: Amsterdam. 2001. **19**. P. 249–276.
29. Brockwell P.J. On continuous time threshold ARMA processes. *Journal of Statistical Planning and Inference*. 1994. **39**, N 2. P. 291–303.
30. Davis R.A., Dunsmuir W.T. Maximum likelihood estimation for MA(1) processes with a root on or near the unit circle. *Econometric Theory*. 1996. **12**. P. 1–29.
31. Fox E., Sudderth E.B., Jordan M.I. & Willsky A.S. bayesian nonparametric inference of switching dynamic linear models. *IEEE Transactions on Signal Processing*. 2011. **59**. P. 1569-1585.
32. Ghosh A., Mukhopadhyay S., Roy S., Bhattacharya S. Bayesian inference in nonparametric dynamic state-space models, 2012.
33. Gray H.L., Kelley G.D. & McIntire D.D. A new approach to ARMA modeling. *Comm. Stat.* 1978. B7. P. 1–77.
34. Grunwald G.K., Hyndman R.J., Hamza K. Some properties and generalizations of nonnegative Bayesian time series models. *Technical Report*, Statistics Dept., 1994.
35. Huang A. Similarity measures for text document clustering. *In Proceedings of the 6th New Zealand Computer Science Research*. Student Conference, Christchurch, New Zealand, 14–18 April, 2008. P. 49–56.
36. Keogh E. Exact indexing of dynamic time warping. In Proceedings of the 28th International Conference on Very Large Data Bases, Hong Kong, China, 20–23 August 2002. P. 406–417.
37. Ljoi A., Mena, R.H. Controlling the reinforcement in Bayesian nonparametric mixture models. *Journal of the Royal Statistical Society*. 2007, Series B 69. P. 715–740.
38. Navarro D., Perfors A. The Metropolis-Hastings algorithm. COMPSCI 3016: Computational Cognitive Science, 2012.
39. Pitman J., Yor M. The two-parameter Poisson-Dirichlet distribution derived from a stable subordinator. *The Annals of Probability*. 1997. **25**. P. 855–900.
40. Zhou C., Wakefield J. A Bayesian mixture model for partitioning gene expression data. *Biometrics*. 2006. **62**. P. 515–525.

Получено 10.04.2018
После доработки 10.04.2018