

Сравнительный анализ мер сходства,
основанных на преобразовании скользящих
аппроксимаций,
в задачах классификации временных рядов

Алимова И.С.¹, Соловьев В.Д.¹, Батыршин И.З.²

¹Казанский Федеральный Университет
Казань, Россия

²Instituto Politecnico Nacional
Mexico, Мексика

1 декабря 2016

План

- ▶ Временной ряд
- ▶ Меры сходства
- ▶ Алгоритм классификации
- ▶ Набор данных
- ▶ Результаты

Временной ряд

- ▶ Упорядоченная последовательность данных, собранных в разные моменты времени
- ▶ Примеры:
 - ▶ показатели курсов валют, взятые за определенный период времени,
 - ▶ кардиограмма человека,
 - ▶ значения температуры воздуха окружающей среды в течении суток

Классификация временных рядов

- ▶ Цель: выявление структуры ряда и прогнозирование дальнейших значений
- ▶ Область применения:
 - ▶ **экономика** - предсказание падения или роста цен акций компаний,
 - ▶ **нефтедобыча** - выявление взаимодействия скважин для добычи нефти с целью определения возможного расположения новых добывающих и нагнетающих скважин,
 - ▶ **робототехника** - моделирование траектории движения мобильных роботов в окружении людей

Меры сходства

- ▶ Евклидово расстояние
- ▶ Алгоритм Динамической Трансформации (АДТ)

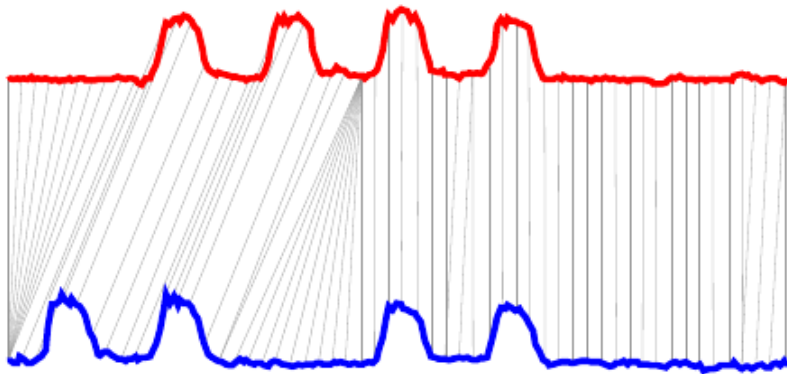
- ▶ Алгоритм преобразования скользящих аппроксимаций (САП трансформ)
- ▶ САП трансформ с динамической трансформацией локальных трендов (САП+ДТ)

Евклидово расстояние

Пусть $x = (x_1, x_2, \dots, x_m)$ и $y = (y_1, y_2, \dots, y_m)$ -временные ряды

$$d_{Eukl} = \sqrt{\sum_{i=0}^m |x_i - y_i|^2}$$

Алгоритм динамической трансформации



- ▶ Учитывает сдвиг временных рядов относительно друг друга по оси X

Алгоритм динамической трансформации

- ▶ $d_{i,j} = |x_i - y_j|, i = 1, \dots, m, j = 1, \dots, m$ - матрица расстояний

Алгоритм динамической трансформации

- ▶ $d_{i,j} = |x_i - y_j|, i = 1, \dots, m, j = 1, \dots, m$ - матрица расстояний
- ▶ $D_{i,j} = d_{i,j} + \min(D_{i-1,j-1}, D_{i-1,j}, D_{i,j-1})$ - матрица трансформаций

Алгоритм динамической трансформации

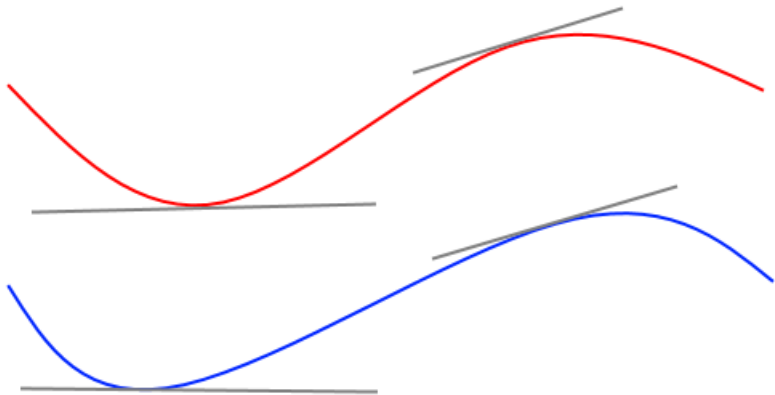
- ▶ $d_{i,j} = |x_i - y_j|, i = 1, \dots, m, j = 1, \dots, m$ - матрица расстояний
- ▶ $D_{i,j} = d_{i,j} + \min(D_{i-1,j-1}, D_{i-1,j}, D_{i,j-1})$ - матрица трансформаций
- ▶ $w_0 = D_{m,m}, w_k = \min(D_{i-1,j-1}, D_{i-1,j}, D_{i,j-1}), k = 1, \dots, m$ - вектор трансформации W

Алгоритм динамической трансформации

- ▶ $d_{i,j} = |x_i - y_j|, i = 1, \dots, m, j = 1, \dots, m$ - матрица расстояний
- ▶ $D_{i,j} = d_{i,j} + \min(D_{i-1,j-1}, D_{i-1,j}, D_{i,j-1})$ - матрица трансформаций
- ▶ $w_0 = D_{m,m}, w_k = \min(D_{i-1,j-1}, D_{i-1,j}, D_{i,j-1}), k = 1, \dots, m$ - вектор трансформации W

$$d_{ADT} = \frac{1}{m} \sum_{i=0}^m w_i$$

САП трансформ



- ▶ Основана на касательных и углах наклона между ними

САП трансформ

▶ $a_i = \frac{6 \sum_{j=0}^{k-1} (2j-k+1)y_{i+j}}{hk(k^2-1)}$ - локальный тренд

САП трансформ

- ▶ $a_i = \frac{6 \sum_{j=0}^{k-1} (2j-k+1)y_{i+j}}{hk(k^2-1)}$ - локальный тренд
- ▶ $\text{coss}_k(X, Y) = \frac{\sum_{i=1}^m a_{yi} a_{xi}}{\sqrt{\sum_{i=1}^m a_{yi}^2 \sum_{j=1}^m a_{xj}^2}}$ - мера ассоциаций для локальных трендов

САП трансформ

- ▶ $a_i = \frac{6 \sum_{j=0}^{k-1} (2j-k+1)y_{i+j}}{hk(k^2-1)}$ - локальный тренд
- ▶ $\text{coss}_k(X, Y) = \frac{\sum_{i=1}^m a_{yi} a_{xi}}{\sqrt{\sum_{i=1}^m a_{yi}^2 \sum_{j=1}^m a_{xj}^2}}$ - мера ассоциаций для локальных трендов

$$d_{SAP} = \frac{1}{|K|} \sum_{k=1}^{m-1} \text{coss}_k(X, Y)$$

▶ $\text{COSS}_{p,q}(a_x, a_y) = \frac{\sum_{i=1}^{m-\max(p,q)+1} a_{yi+q} a_{xi+p}}{\sqrt{\sum_{i=1}^{m-q+1} a_{yi+q} \sum_{j=1}^{m-p+1} a_{xi+p}}}$ - матрица
косинусов

- ▶ $COSS_{p,q}(a_x, a_y) = \frac{\sum_{i=1}^{m-\max(p,q)+1} a_{yi+q} a_{xi+p}}{\sqrt{\sum_{i=1}^{m-q+1} a_{yi+q} \sum_{j=1}^{m-p+1} a_{xi+p}}}$ - матрица косинусов
- ▶ $w_0 = COSS_{mm}, w_k = w_{k-1} + \min(COSS_{i-1,j-1}, COSS_{i-1,j}, COSS_{i,j-1})$
- вектор трансформации косинусов

САП+ДТ

▶ $COSS_{p,q}(a_x, a_y) = \frac{\sum_{i=1}^{m-\max(p,q)+1} a_{yi+q} a_{xi+p}}{\sqrt{\sum_{i=1}^{m-q+1} a_{yi+q} \sum_{j=1}^{m-p+1} a_{xi+p}}}$ - матрица косинусов

▶ $w_0 = COSS_{mm}, w_k = w_{k-1} + \min(COSS_{i-1,j-1}, COSS_{i-1,j}, COSS_{i,j-1})$
- вектор трансформации косинусов

$$d_{SAP+DT} = \frac{1}{m} \sum_{k=1}^m w_k$$

Выбор оптимального размера окна для меры САП+ДТ

- ▶ **Входные данные:** S - обучающий набор данных
- ▶ **Результат:** $bestk$ - размер оптимального окна k
- ▶ S разбивается на S_{contr} и S_{train}

$k = 2$

$bestk = 2$

$maxAcc = 0$

Для $k \leq 30$:

acc = классификация(S_{contr} , S_{train})

Если acc > maxAcc:

maxAcc = acc

bestk = k

Алгоритм классификации

- ▶ **Входные данные:** T - тестовый набор временных рядов, S - обучающий набор данных, M - мера сходства
- ▶ **Результат:** k - кол-во верно классифицированных рядов
- ▶ $k = 0$

Для каждого $t \in T$:

Находится $s \in S$ такой, что $M(t,s)$ - минимальна

Если номера классов s и t совпадают:

$$k = k + 1$$

Качественный анализ классификации

Точность (Acc) классификации меры для набора данных:

$$Acc = \frac{k}{|T|}$$

k - количество верно классифицированных рядов, $|T|$ - количество рядов в тестовом наборе данных

Набор данных

- ▶ Временные ряды из 43 наборов данных из коллекции UCR
- ▶ Каждый набор данных содержит обучающий и тестовый файлы
- ▶ В каждом из файлов записаны временные ряды с номером класса к которому они принадлежат

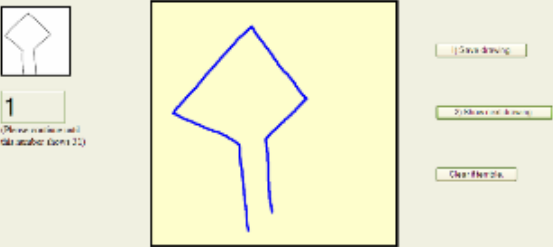
Набор временных рядов GUN POINT



- ▶ Траектории движения центра ладони руки человека по оси X, когда он достаёт пистолет из кобуры на бедре, целится в мишень и кладет пистолет обратно

Набор временных рядов Symbols

Draw the picture displayed.
Do not draw outside the box.



1
(Please continue until the number shows 2)

Save drawing
Print my drawing
Clear the plot

- ▶ Координаты оси X рисунка символа

Оценка точности мер сходства

Мера	Точность
АДТ	0,91
Евклидово расстояние	0,9
САП+ДТ	0,88
САП	0,86

Таблица 1: Точность рассматриваемых мер по всем наборам коллекции, полученное макро усреднением.

Оценка точности мер сходства

Набор данных	Евклидово расстояние	АДТ	САП	САП + ДТ
50Words	0,8	0,71	0,65	0,67(+3%)
Car	0,8	0,73	0,65	0,67(+2%)
ECGFiveDays	0,81	0,77	0,68	0,97(+29%)
FaceFour	0,78	0,84	0,82	0,84(+2%)
FacesUCR	0,85	0,94	0,78	0,87(+9%)
Gun_Point	0,93	0,87	0,84	0,94(+10%)
MoteStrain	0,88	0,89	0,83	0,86(+3%)
SonyAIBORobotSurface	0,71	0,72	0,72	0,78(+6%)
SonyAIBORobotSurfaceII	0,87	0,85	0,85	0,88(+3%)
Symbols	0,9	0,95	0,88	0,94(+6%)
Trace	0,8	0,98	0,71	0,95(+24%)
TwoLeadECG	0,84	0,97	0,76	0,92(+16%)
uWaveGestureLibrary_X	0,79	0,72	0,66	0,71(+5%)
WordsSynonyms	0,72	0,68	0,59	0,63(+4%)

Таблица 2: Значения точности рассматриваемых мер для наборов данных для временных рядов из коллекции UCR.

График временных рядов из набора данных GUN POINT

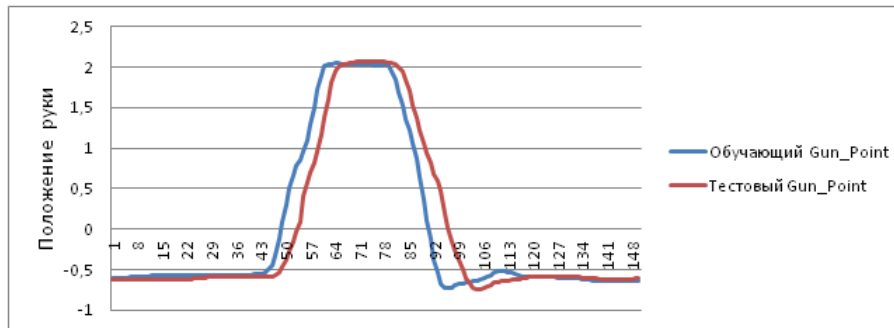


График временных рядов из набора данных ECGFiveDays

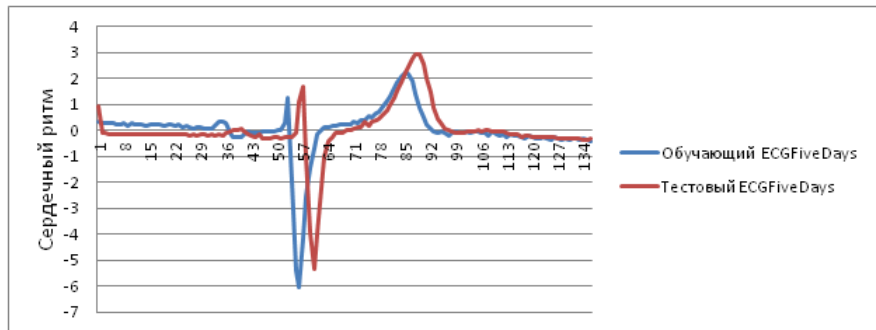
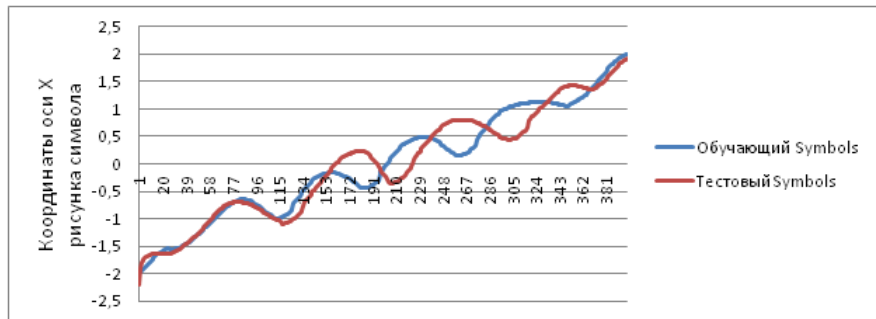


График временных рядов из набора данных Symbols



Статистическая оценка эффективности меры САП+ДТ

- ▶ Критерий Вилкоксона
- ▶ **Гипотеза:** мера САП+ДТ показывает более высокую точность, чем мера САП
- ▶ $T_{exp} = 35$, при $T_{krit} = 43$, для $p = 0.01$
- ▶ $T_{exp} < T_{krit} \Rightarrow$ гипотеза достоверна

Оценка производительности алгоритмов

Мера	Точность
Евклидово расстояние	$O(m)$
АДТ	$O(m^2)$
САП+АДТ	$O(m^2)$
САП	$O(m^2)$

Таблица 3: Оценки производительности алгоритмов для вычисления рассматриваемых мер сходства для временного ряда длины m .

Заключение

- ▶ Проведен сравнительный анализ меры сходства временных рядов, САП трансформ с Евклидовой и АДТ мерами
- ▶ Предложен алгоритм, улучшающий точность меры САП трансформ на временных рядах сдвинутых относительно друг друга по оси X

Спасибо за внимание!