Анализ свойств процедуры выравнивания временных рядов в задачах метрической классификации

Александра Харь

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

Консультанты: Г. Моргачев, А. Гончаров
Научный руководитель д.ф.-м.н. В. В. Стрижов

Цели исследования

Цель работы

Обосновать применение DTW для метрических методов классификации временных рядов.

Метод анализа

Сравнение качества классификации временных рядов (при помощи SVM) с использованием динамического выравнивания (DTW) и без него. В процессе эксперимента в качестве ядра в SVM используются следующие матрицы: матрица попарных расстояний между временными рядами, полученная при помощи DTW, некоторые ее модификации, а также матрица попарных Евклидовых расстоянийй между временными рядами.

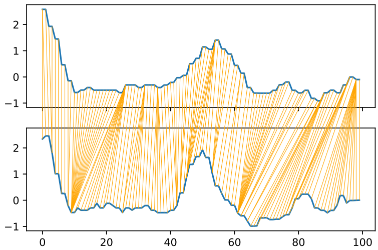
Список литературы

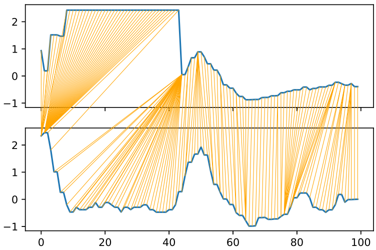
- Hui Ding, Goce Trajcevski, Peter Scheuermann, Xiaoyue Wang, and Eamonn J. Keogh. Querying and mining of time series data: experimental comparison of representations and distance measures. PVLDB, 1(2):1542–1552, 2008.
- Chan P. Fastdtw Salvador S. Toward accurate dynamic time warping in linear time and space.InWorkshop on Mining Temporal and Sequential Data, page 11, 2004.
- Kristin P. Bennett and Colin Campbell. Support vector machines: Hype or hallelujah? SIGKDD Explorations, 2(2):1–13, 2000.

Объекты $\boldsymbol{x} \in \mathbb{R}^n$ – временные ряды, принадлежащие двум непересекающимся классам $Y = \{+1, -1\}$
Линейный классификатор: $a(\boldsymbol{x}) = sign(\langle \boldsymbol{w}, \boldsymbol{x} \rangle - w_0)$, где
$\boldsymbol{x} = [x_1, \ldots, x_n]^T$ – признаковое описание объекта x,
$\boldsymbol{w} = [w_1, \ldots, w_n]^T \in \mathbb{R}^n$ и $w_0 \in \mathbb{R}$ – веса признаков
$\langle \boldsymbol{w}, \boldsymbol{x} \rangle = w_0$ – гиперплоскость, разделяющая классы
задача оптимизации: $\begin{cases} \frac{1}{2}\langle \boldsymbol{w}, \boldsymbol{w} \rangle \to min \\ (\langle \boldsymbol{w}, \boldsymbol{x_j} \rangle - w_0) y_j \geq 1; \quad i = 1, \ldots, \ell. \end{cases}$

Метод динамического выравнивая – DTW используется для определения сходства между двумя временными рядами и введения расстояния между ними. Он находит наилучшее соответствие между двумя временными рядами, если они нелинейно деформированы друг относительно друга – растянуты, сжаты или смещены вдоль оси времени.

Примеры работы DTW (построение выравнивающего пути):





Теоретическая часть: DTW расстояние

Рассмотрим два временных ряда Q и C разной длины:
$Q = q_1, q_2, \ldots, q_i, \ldots, q_n;\ \ C = c_1, c_2, \ldots, c_j, \ldots, c_m$

1. строим матрицу d размера $n \times m$, в которой элемент $d_{ij} = (q_i - c_j)^2$ – расстояние между двумя точками q_i и c_j
2. строим матрицу трансформации D, каждый элемент которой вычисляется следующим образом:
   $D_{ij} = d_{ij} + min(D_{i-1\ j}, D_{i-1\ j-1}, D_{i\ j-1})$
3. строим путь трансформации W – минимизирует общее расстояние между Q и C
   $W = w_1, w_2, \ldots, w_k, \ldots, w_K$, где
   $w_k = (i,j)_k,\ \ d(w_k) = (q_i - c_j)^2$
4. $\rho_{DTW}(Q, C) = \min\left\{\sum_{k=1}^{K} d(w_k)/K\right\}$

Если выборка не является линейно разделимой, то для классификации при помощи SVM необходимо перейти в пространство большей размерности, где она уже будет линейно разделима. Скалярной произведение $\langle \boldsymbol{x}, \boldsymbol{x'} \rangle$, таким образом, везде заменяется на значение функции ядра в соответствующих двух точках $K(\boldsymbol{x}, \boldsymbol{x'})$.

Теорема Мерсера:

Функция $K(\boldsymbol{x}, \boldsymbol{x'})$ является ядром тогда и только тогда, когда:
- $K(\boldsymbol{x}, \boldsymbol{x'}) = K(\boldsymbol{x'}, \boldsymbol{x})$
- матрица $K = [K(\boldsymbol{x_i}, \boldsymbol{x_j})]_{i,j}$ неотрицательно определена:

$\boldsymbol{v}^T K \boldsymbol{v} \geq 0, \quad \forall \boldsymbol{v} \in \mathbb{R}^n.$

Эксперимент

При помощи алгоритма DTW вычисляется попарное расстояние между временными рядами, эти расстояния записываются в матрицу D. В таком случае она будет симметричной, с нулями на диагонали. При помощи теоремы Мерсера выявляется, что D не является ядром (не является неотрицательно определенной). Так как для перехода в пространство большей размерности в алгоритме SVM для более качественной классификации обычно используют ядро, то мы попробуем модифицировать D, чтобы она стала таковым, затем сравним точности классификации при использовании первоначальной D и модифицированных матриц.

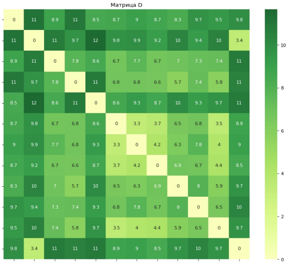
Модификация матрицы D

Модифицируем D двумя способами:

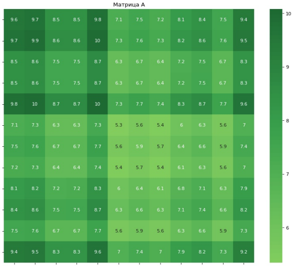
- Ищем ближайшую к матрице D неотрицательно определенную матрицу A следующим образом:
  так как D – симметрична, то $D = QMQ^T$, где M – диагональная с собственными числами D на диагонали
  $M_+ = [max(M_{ij}, 0)]_{ij}$
  $A = QM_+Q^T$ – неотрицательно определенная
- Создаем RBF-ядро $K(x, x') = exp(-\gamma \cdot \rho_{dtw}(x, x'))$, где $\rho_{dtw}(x, x')$ - расстояние между рядами x и x'.
  Получаем матрицу Z, она уже является ядром, что было проверено на эксперименте при помощи теоремы Мерсера.

На небольшом объеме данных матрицы выглядят следующим образом:

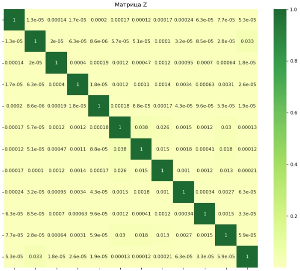
Матрица D



Матрица A



Матрица Z



Эксперимент

Используем эти матрицы, а также обыкновенную норму l_2 в качестве ядра в алгоритме SVM для классификации данных и оцениваем точность классицикации.

Оценка точности классификации

- метрика l_2 : 0.182
- матрица D : 0.327
- матрица A : 0.523
- матрица Z : 0.864

Метрика l_2 дает худший результат, следующая по качеству – матрица D попарных расстояний между рядами, а затем уже две ее модификации – матрицы A и Z, причем при помощи RBF-ядра Z получена лучшая классификация.

Заключение

- Произведена классификация временных рядов при помощи ядрового SVM с различными ядрами: метрика l_2, матрица расстояния между временными рядами D, ее аппроксимация до ядра A, и RBF ядро Z.
- Проведено сравнение данных классификаций.
- Использование матрицы D попарных расстояний улучшает точность классификации по сравнению с l_2 на 15%.
- Модифицированные версии A и Z матрицы D улучшают точность классификации еще на 20 и 55 % соответственно.