Теоретическая обоснованность применения метрических методов классификации с использованием динамического выравнивания (DTW) к пространственно-временным объектам

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 решения

Проверка качества классификации временных рядов (при помощи SVM) с использованием DTW и без него

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.
- Eamonn Keogh and M. Pazzani. Scaling up dynamic time warping to massive datasets. In Proceedings 3rd European Conference on Principles and Practice of Knowledge Discovery in Databases, pages 1–11, 1999.
- 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.
- Mercer. Functions of positive and negative type and their connection with the theory of integral equations.Philos. Trans. Royal Soc. (A), 83(559):69–70, November 1909.

Постановка задачи

Объекты $X \in \mathbb{R}^n$ - временные ряды, принадлежащие двум непересекающимся классам $Y = \{+1, -1\}$

Линейный классификатор:

$a(x) = sign(\sum_{i=1}^{n} w_i x_i - w_0) = sign(\langle w, x \rangle - w_0)$, где

$x = (x_1, \ldots, x_n)$ – признаковое описание объекта x,

$w = (w_1, \ldots, w_n) \in \mathbb{R}^n$ и $w_0 \in \mathbb{R}$ – веса признаков

$\langle w, x \rangle = w_0$ – гиперплоскость, разделяющая классы

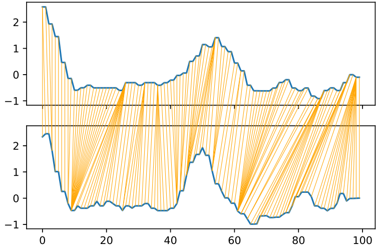
задача оптимизации: $\begin{cases} \frac{1}{2}\langle w, w \rangle \to min \\ (\langle w, x^j \rangle - w_0) y^j \geq 1 \quad i = 1, \ldots, l \end{cases}$

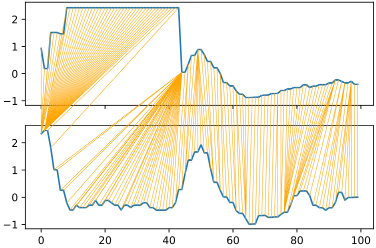
Если выборка не является линейно разделимой, то для классификации при помощи SVM необходимо перейти в пространство большей размерности, где она уже будет линейно разделима. Скалярной произведение $\langle x, x' \rangle$, таким образом, везде заменяется на значение функции ядра в соответствующих двух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

Теорема Мерсера:

Функция $K(x, x')$ является ядром тогда и только тогда, когда:
- $K(x, x') = K(x', x)$
- матрица $K = ||K(x_i, x_j)||_{i,j}$ неотрицательно определена: $v^T K v \geq 0 \;\; \forall v \in \mathbb{R}^n$

Примеры выравнивающих путей между двумя временными рядами (DTW)





Эксперимент

При помощи алгоритма DTW вычисляется матрица D попарных расстояний между временными рядами, при помощи теоремы Мерсера выявляется, что она не является ядром.

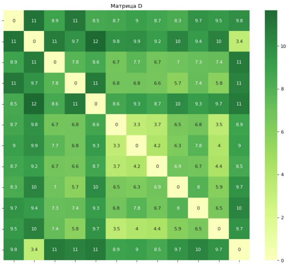
Модификация матрицы D

• 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,
$M_+ = max(M, 0)$, $A = QM_+Q^T$ – 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.

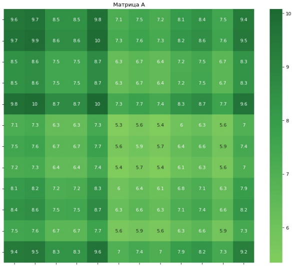
Эксперимент

На небольшом объеме данных матрицы выглядят следующим образом

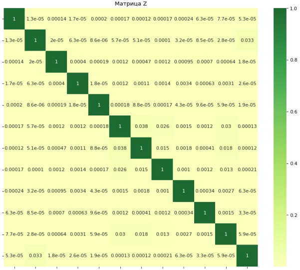
Матрица D



Матрица A



Матрица Z



Затем мы работаем уже с большими объемами данных и считаем аналогично матрицы D, A, Z. Затем используем эти матрицы, а также обыкновенную норму l_2 в качестве "ядра"в алгоритме SVM для классификации данных.

Оценка точности классификации

- метрика l_2 : 0.182
- матрица D : 0.227
- матрица A : 0.523
- матрица Z : 0.864

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