Классификация временных рядов в пространстве параметров порождающих моделей

Карасиков Михаил

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

Научный руководитель: д.ф.-м.н. В. В. Стрижов

Москва, 2015 г.

Описание исследования

Исследуется

Задача построения пространства признаков для многоклассовой классификации временных рядов

Проблема

Построение краткого интерпретируемого признакового описания временных рядов

Цели исследования:

- построение алгоритма многоклассовой классификации, использующего в качестве признаков временных рядов параметры моделей временных рядов и их распределения,
- обобщение методов классификации временных рядов, использующих явное признаковое описание,
- повышение качества решения задач классификации временных рядов.

Литература

- Time-series data mining / P. Esling, C. Agon // ACM Comput. Surv. — 2012.— December.— Vol. 45, no. 1.— Pp. 12:1–12:34.
- Segment and combine approach for non-parametric time-series classification / P. Geurts, L. Wehenkel // Knowledge Discovery in Databases: PKDD 2005.— Springer, 2005.— Pp. 478–485.
- Human activity recognition using smart phone embedded sensors: A linear dynamical systems method / W. Wang, H. Liu, L. Yu, F. Sun // Neural Networks (IJCNN), 2014 International Joint Conference on.— 2014.—July.— Pp. 1185–1190. //doi.acm.org/10.1145/1964897.1964918.

Постановка задачи

Дано: $X = \left\{ x_i = [x_i^{(1)}, \dots, x_i^{(t)}] :\ i \in \mathcal{I} \right\}$ — временные ряды,
Y — множество меток классов,
$\mathfrak{D} \subset X \times Y$ — обучающая выборка.

Модель классификации: $a = b \circ \mathbf{f} \circ S$, где
S — алгоритм сегментации $S:\ x \mapsto \{s^{(1)}, \dots, s^{(p)}\}$,
где $s^{(k)} = [x^{(t_k)}, \dots, x^{(t_{k+1}-1)}]$.
$\mathbf{f}$ — признаковое описание набора сегментов,
b — алгоритм многоклассовой классификации.

Метод обучения $\mu:\ \mathfrak{D} \mapsto a$
выбирается по скользящему контролю:

$$\mu^* = \arg\min_{\mu} \widehat{CV}(\mu, \mathfrak{D}).$$

Модель сегмента временного ряда

Сегменты временных рядов описывается моделями вида

$$g(\mathbf{w}, x) \in X, \text{ где } \mathbf{w} \in \mathbb{R}^n.$$

Параметры настроенной модели определяются по формуле

$$\mathbf{w}_g(x) = \operatorname*{arg\,min}_{\mathbf{w}\in\mathbb{R}^n} \rho\left(g(\mathbf{w}, x), x\right).$$

Исследуются

- Модель линейной регрессии
- Модель авторегрессии (AR)
- Фурье-модель
- Вейвлет-модель

. Пусть задан многокомпонентный временной ряд

$$x = [\boldsymbol{x}^{(1)}, \dots, \boldsymbol{x}^{(t)}], \text{ где } \boldsymbol{x}^{(k)} = [x_0^{(k)}, \dots, x_n^{(k)}]^\mathsf{T}.$$

Модель линейной регрессии одной из компонент на остальные:

$$g(\mathbf{w}, x) = [\hat{\boldsymbol{x}}^{(1)}, \dots, \hat{\boldsymbol{x}}^{(t)}], \text{ где } \hat{\boldsymbol{x}}^{(k)} = [\hat{x}_0^{(k)}, x_1^{(k)}, \dots, x_n^{(k)}]^\mathsf{T},$$

$$\hat{\boldsymbol{x}}_0 = \begin{bmatrix} \hat{x}_0^{(1)} \\ \vdots \\ \hat{x}_0^{(t)} \end{bmatrix} = \underbrace{\begin{bmatrix} x_1^{(1)} & \dots & x_n^{(1)} \\ \vdots & \ddots & \vdots \\ x_1^{(t)} & \dots & x_n^{(t)} \end{bmatrix}}_{X} \underbrace{\begin{bmatrix} w_1 \\ \vdots \\ w_n \end{bmatrix}}_{\mathbf{w}}.$$

Тогда, выбрав в качестве ρ евклидово расстояние, получим

$$\mathbf{w}_g(x) = \underset{\mathbf{w} \in \mathbb{R}^n}{\arg\min}\, \rho\left(g(\mathbf{w}, x), x\right) = \left(X^\mathsf{T} X\right)^{-1} X^\mathsf{T} \boldsymbol{x}_0.$$

Модель авторегрессии AR(p)

Задан временной ряд

$$x = [x^{(1)}, \dots, x^{(t)}],\ x^{(k)} \in \mathbb{R},\ k = 1, \dots, t.$$

Модель аппроксимации — авторегрессионная модель порядка p:

$$g(\mathbf{w}, x) = [x^{(1)}, \dots, x^{(p)}, \hat{x}^{(p+1)}, \dots, \hat{x}^{(t)}],$$

где

$$\hat{x}^{(k)} = w_0 + \sum_{i=1}^{p} w_i x^{(k-i)},\ k = p+1, \dots, t.$$

Признаковое описание $\mathbf{w}_g(x)$ определяется аналогично случаю линейной регрессии.

Параметры моделей сегментов

Предлагаются две схемы решения исходной задачи.

- Принцип голосования: обучение алгоритма b на новой обучающей выборке $\widehat{\mathfrak{D}}$, составленной из сегментов временных рядов исходной обучающей выборки $\mathfrak{D}$:

$$\widehat{\mathfrak{D}} = \{(\mathbf{w}_g(s), y) : \ (x, y) \in \mathfrak{D}, \ s \in S(x)\}$$

  и последующая классификация

$$a(x; S, g, b) = h\left(\{b(\mathbf{w}_g(s)) : \ s \in S(x)\}\right).$$

- Классификация в пространстве гиперпараметров моделей (параметров распределений параметров моделей).

$\mathbf{w}_g \circ S$ дает множество наборов параметров модели:

$$W(x; S, g) = \{\mathbf{w}_g(s) : \ s \in S(x)\}.$$

Гипотеза порождения временного ряда

Сегменты временного ряда $s \in S(x)$ описываются моделью $g(\mathbf{w}, s)$ со случайными параметрами $\mathbf{w}$ из параметрического семейства распределений $\{\mathsf{P}_{\boldsymbol{\theta}}\}_{\boldsymbol{\theta} \in \Theta}$.

Предлагается в качестве признакового описания временного ряда использовать оценку вектора параметров распределения:

$$\mathbf{f}(x; S, g, \Theta) = \operatorname*{arg\,max}_{\boldsymbol{\theta} \in \Theta} \mathcal{L}\left(\boldsymbol{\theta} | W(x; S, g)\right).$$

Тогда получим алгоритм классификации временных рядов:

$$a(x) = b(\mathbf{f}(x; S, g, \Theta)).$$

Сведение задачи многоклассовой классификации к бинарным

- One-vs-All approach:

$$a(x) = \operatorname*{arg\,max}_{i=1,\dots,N} f_i(x), \quad f_i(x) = \begin{cases} \geqslant 0, & \text{если } y(x) = i, \\ < 0, & \text{если } y(x) \neq i. \end{cases}$$

- One-vs-One approach:

$$a(x) = \operatorname*{arg\,max}_{i=1,\dots,N} \sum_{\substack{j=1,\dots,N \\ j \neq i}} f_{ij}(x), \quad f_{ij}(x) = \begin{cases} +1, & \text{если } y(x) = i, \\ -1, & \text{если } y(x) = j. \end{cases}$$

- Error-Correcting Output Codes approach:

$$a(x) = \operatorname*{arg\,min}_{i=1,\dots,N} \sum_{j=1}^{F} L(M_j^i f_j(x)),$$

где $M \in \{-1, 0, +1\}^{N \times F}$ — матрица, строки которой состоят из кодов меток классов Y, а L — функция потерь.

Бинарная классификация: $Y = \{\pm 1\}$

Пусть $\mathbf{f}(x) \in \mathbb{R}^n$ — признаковое описание временного ряда x.

Тогда для решения задачи бинарной классификации временных рядов необходимо задать метод обучения $\mu_b : \mathfrak{D} \mapsto f$.

Например,

- SVM с линейным ядром:

$$f(x; \boldsymbol{w}) = \mathsf{sign}(\boldsymbol{w}^\mathsf{T}\mathbf{f}(x) - w_0),$$

  где $\boldsymbol{w}$ и w_0 — решения оптимизационной задачи

$$\frac{1}{2C}\|\boldsymbol{w}\|^2 + \sum_{(x,y)\in\mathfrak{D}} \Big(1 - y(\boldsymbol{w}^\mathsf{T}\mathbf{f}(x) - w_0)\Big)_+ \to \min_{\boldsymbol{w}\in\mathbb{R}^n,\ w_0\in\mathbb{R}};$$

- регуляризованная лог. регрессия (RLR):

$$\frac{1}{2C}\|\boldsymbol{w}\|^2 + \sum_{(x,y)\in\mathfrak{D}} \log\left(1 + e^{-y\boldsymbol{w}^\mathsf{T}\mathbf{f}(x)}\right) \to \min_{\boldsymbol{w}\in\mathbb{R}^n}.$$

Приложение

В качестве приложения рассматривается задача классификации физической активности по данным с акселерометра.

Особенности

- Классификация физической активности людей с разными физическими характеристиками
- Форма временного ряда существенно зависит от характеристик человека
- Во временных рядах допускаются аномалии

Предположение

Форма временного ряда сохраняются для конкретного человека и типа физической активности

Цели эксперимента:

1. демонстрация качества предлагаемого алгоритма
2. изучение зависимости качества классификации от
   - алгоритма сегментации
   - модели сегмента

Dataset WISDM

Временной ряд из трех компонент: $\mathbf{x} = [x_t^k]^{k=1,2,3}$.

Признаки (по 31 на каждый временной ряд)

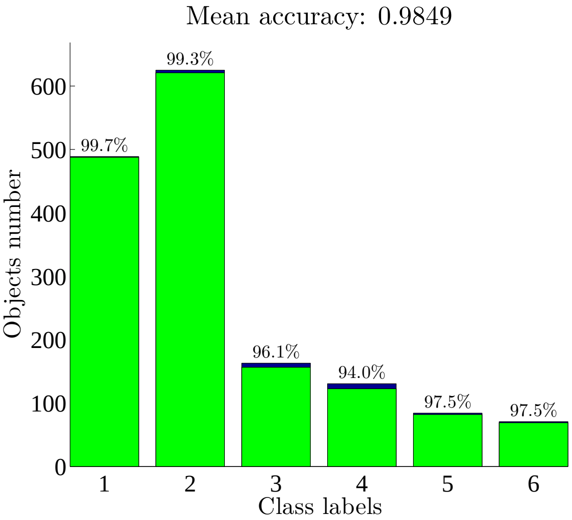
1. 7 коэффициентов авторегрессии $AR(6)$:

$$\underset{w_0,\dots,w_6}{\text{arg min}} \sum_t \left(x_t^k - w_0 - \sum_{i=1}^{6} w_i x_{t-i}^k \right)^2 .$$

2. Статистики:

- $\bar{x}^k = \frac{1}{T} \sum_t x_t^k$,
- $\frac{1}{T} \sum_t |x_t^k - \bar{x}^k|$,
- $\sqrt{\frac{1}{T-1} \sum_t (x_t^k - \bar{x}^k)^2}$,
- $\frac{1}{T} \sum_t \|\mathbf{x}_t\|$.

Без сегментации, классификатор RBF SVM ($\gamma = 0.8$, $C = 4$), подход One-vs-All, 50 случайных разбиений в отношении 7 к 3.



Class labels:

1. Jogging
2. Walking
3. Upstairs
4. Downstairs
5. Sitting
6. Standing

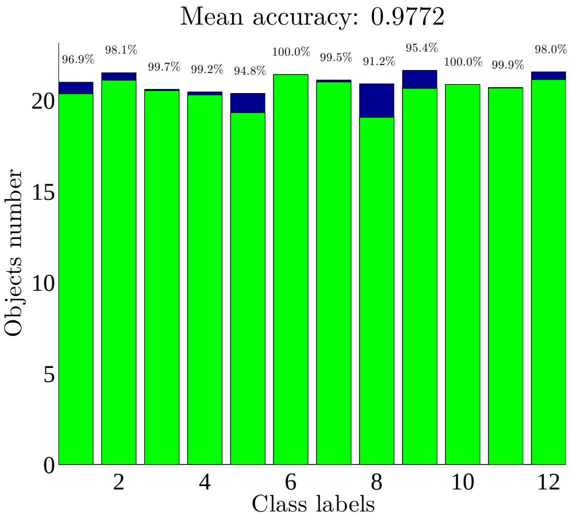
Dataset USC-HAD

Выборка содержит: показания акселерометра и гироскопа для 12 типов физической активности.

Частота исходной выборки 100 Hz.
Приведем данные к частоте 10 Hz.

Будем решать задачу многоклассовой классификации, классификатор RBF SVM ($C = 16$, $\gamma = 0.04$), подход One-vs-All, 100 случайных разбиений в пропорции 7 к 3.

В качестве признаков берутся коэффициенты модели авторегрессии $AR(10)$ и первые 5 коэффициентов модели Фурье.



Class labels:

1. walk forward
2. walk left
3. walk right
4. go upstairs
5. go downstairs
6. run forward
7. jump up and down
8. sit and fidget
9. stand
10. sleep
11. elevator up
12. elevator down

Segments Voting vs Normal Distribution, случайная сегментация

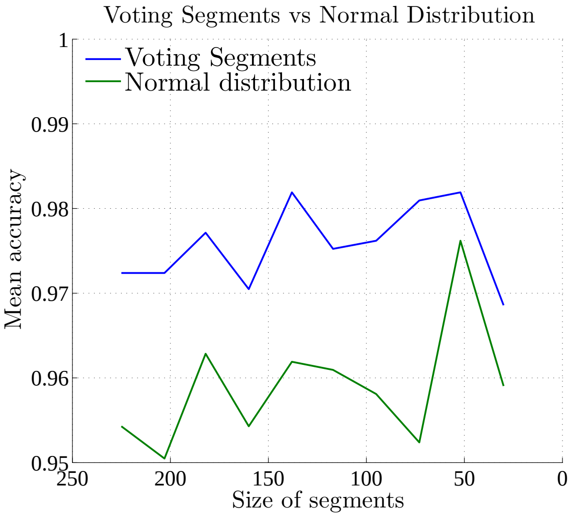


Рис.: На первых 10 классах. One-vs-One, Linear SVM ($C = 0.25$).

Заключение

- Метод классификации временных рядов в пространстве параметров моделей показывает высокие результаты.
- Использование модели авторегрессии не накладывает ограничений на алгоритм сегментации.
- При случайной сегментации голосование сегментов оказывается лучше алгоритма классификации в пространстве гиперпараметров.

Результаты, выносимые на защиту

- Предложен алгоритм построения пространства признаков.
- Предложены алгоритмы классификации временных рядов.
- Выполнена программная реализация и проведены численные эксперименты, показавшие повышение качества решения задачи классификации типов физической активности.