

МИНИСТЕРСТВО ОБРАЗОВАНИЯ И НАУКИ РОССИЙСКОЙ ФЕДЕРАЦИИ
МОСКОВСКИЙ ФИЗИКО-ТЕХНИЧЕСКИЙ ИНСТИТУТ (государственный университет)
ФАКУЛЬТЕТ УПРАВЛЕНИЯ И ПРИКЛАДНОЙ МАТЕМАТИКИ
КАФЕДРА «ИНТЕЛЛЕКТУАЛЬНЫЕ СИСТЕМЫ»
ПРИ ВЫЧИСЛИТЕЛЬНОМ ЦЕНТРЕ ИМ. А. А. ДОРОДНИЦЫНА РАН

Иванычев Сергей Дмитриевич

**Выполнимость гипотезы простоты выборки для
комбинированных признаков описаний в задаче
классификации временных рядов**

010900 — Прикладные математика и физика

ВЫПУСКНАЯ КВАЛИФИКАЦИОННАЯ РАБОТА БАКАЛАВРА

Научный руководитель:
д.ф.-м.н. Стрижов Вадим Викторович

Москва

2018 г.

Содержание

1	Введение	4
2	Обзор литературы	6
3	Постановка задачи прогнозирования	7
3.1	Комбинированное признаковое описание	8
4	Построение признакового пространства	9
4.1	Локально-аппроксимирующие модели	9
4.1.1	SE MOR	9
4.1.2	AR-авторегрессия	10
4.1.3	Фурье-модель	10
4.1.4	Вейвлет-модель	11
4.2	Построение пространства Z	11
5	Алгоритм многоклассовой классификации	12
5.1	Выбор моделей локальной аппроксимации	12

Аннотация

Ключевые слова: *прогнозирование временных рядов; объекты сложной структуры.*

1 Введение

В работе рассматривается задача классификации временных рядов в задаче распознавания действий человека по временным рядам, порождаемым датчиками носимых устройств, например, с акселерометра, гироскопа, альтиметра смартфонов или умных браслетов. Существует несколько подходов к классификации временных рядов: среди них можно выделить машины опорных векторов [1, 2], рекуррентные [3] и глубокие [4] нейронные сети или решающие деревья. Классификация временных рядов является частным случаем классификации объектов сложной структуры. Из-за того, что подобные задачи возникают во многих областях, например, в обработке сигналов, биологии, финансах, метеорологии, существует довольно много техник ее решения.

В нашей работе нас интересует решение задачи классификации временных рядов путем построения промежуточного признакового пространства [5]. Этот метод применим не только к задаче классификации рядов с носимых устройств, так как к объектам сложной структуры можно свести соответствующие ряды из других задач. В общем случае подход с промежуточным признаковым пространством разделим на два этапа.

- На первом этапе для сегментов временных рядов, которые выступают в роли объектов (которые, вообще говоря, могут быть различной длины и даже частоты дискретизации) вычисляются некоторые статистики или добываются некоторые экспертные оценки. В результате на каждый объект мы имеем некоторый набор численных показателей.
- Над вторичным пространством этих показателей (то есть преобразованными объектами) работает некоторый алгоритм классификации (например ...), который обучается на "вторичной" выборке.

Эти этапы зависимы, так как классификатор, используемый во втором этапе может потребовать от обучающей выборки выполнимость некоторых гипотез и, в частности, гипотезы простоты выборки, что может быть обеспечено только корректным первым этапом. Выполнимость гипотезы простоты выборки, находящейся в промежуточном пространстве, необходима для корректной работы алгоритмов классификации.

В нашей работе мы рассматриваем при каких условиях отображение объектов сложной структуры порождает *простую* выборку, то есть случайную, однородную и независимую, а также предлагаем пути построения соответствующей выборки.

2 Обзор литературы

Алгоритм классификации временных рядов по их признаковому описанию а также базовые подходы к генерации признаков описаний в задаче определения типа движения рассмотрены в работе [5]. Альтернативный подход к генерации признаков описаний, основанный на описании сегментов оптимальными параметрами аппроксимирующих моделей описан в работе [6].

3 Постановка задачи прогнозирования

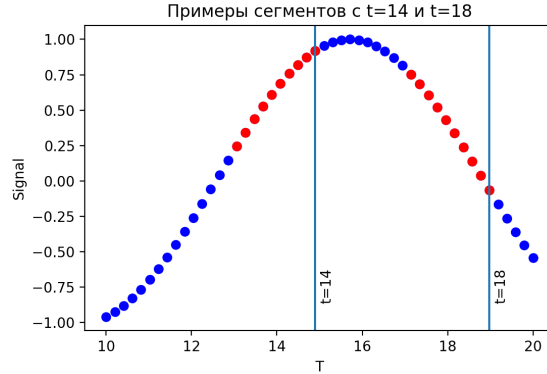
Рассмотрим некоторый временной ряд, то есть функцию определенную на множестве временных меток.

$$S : T \rightarrow \mathbb{R} \text{ где } T = \{t_0, t_0 + d, t_0 + 2d \dots\}, |T| < \infty$$

Зададим некоторую ширину сегмента $n \in \mathbb{N}$, тогда объектом s_i мы назовем набор

$$x_i = s(t_i) = (S(t_i), S(t_i - d), S(t_i - 2d), \dots, S(t_i - (n - 1)d)), \quad x_i \in X \equiv \mathbb{R}^n$$

Рис. 1: Пример сегментов временного ряда с $n = 10$.



Необходимо восстановить зависимость $y = f(x)$, $f : X \rightarrow \{1, 2, \dots K\}$. Для этого задана обучающая выборка

$$\mathfrak{D} = \{(x_i, y_i)\}_{i=1}^l, \quad y_i \in \{1, 2, \dots K\}$$

а также функция потерь (multinomial log-loss)

$$L(f(x_i), y) = \sum_{i=1}^l \sum_{k=1}^K [y_i = k] \log P(y_i = k | x_i, \theta) \quad (3.1)$$

Таким образом мы решаем задачу оптимизации

$$\hat{\theta} = \arg \min_{\theta \in \Theta} \sum_{i=1}^l L(f(x_i), y_i)$$

3.1 Комбинированное признаковое описание

Пусть H — множество функций вида $h : X \rightarrow \mathbb{R}^m$, где $m = m(g)$, то есть это множество отображений пространства объектов сложной структуры в пространство действительных чисел некоторой размерности (для каждой функции размерность может быть своя). В H могут лежать например

- Множество моделей локальной аппроксимации сигнала
- Множество статистик
- Множество экспертных оценок каждого из сложных объектов

Возьмем конечный поднабор этих функций

$$\mathbf{h} = [h_1 \dots h_k], \{h_1 \dots h_k\} \subset H$$

Обозначим сумму размерностей образов функций из набора как

$$n_{\mathbf{h}} \triangleq \dim(\text{Im}(h_1)) + \dim(\text{Im}(h_2)) + \dots + \dim(\text{Im}(h_k))$$

Тогда \mathbf{h} индуцирует отображение $\mathbf{h} : X \rightarrow Z \subset \mathbb{R}^{n_g}$, причем в векторах образа первые $\dim(\text{Im}(h_1))$ компонент соответствуют образу h_1 , следующие $\dim(\text{Im}(h_2))$ соответствуют h_2 и так далее. Z называется *признаковым пространством* объектов сложной структуры X . Тогда, мы можем искать f в семействе суперпозиций $a(\mathbf{h}(\cdot), \gamma)$, где

$$X \xrightarrow{\mathbf{h}} Z \xrightarrow{a} Y$$

- \mathbf{h} — это признаковое отображение
- $a(\cdot, \gamma)$ — параметрическое отображение Z в $\{1, 2, \dots, K\}$, которое соответствует некоторому алгоритму машинного обучения, параметризованного вектором гиперпараметров γ .

В таком подходе функция потерь теперь определена на отображении a , то есть

$$L(f(x), y) = L(a(g(x), \gamma), y)$$

Итак, задача решается следующим образом:

- Поиск и вычисление отображения $Z = \{\mathbf{h}(x_i)\}_{i=1}^l$.
- Определение оптимальных параметров γ в задаче оптимизации

$$\hat{\gamma} = \arg \min_{\gamma} \sum_{i=1}^l L(\mathbf{h}(x_i, \gamma), y_i)$$

Основное допущение, принимаемое в данном методе является допущение о том, что выборка в признаковом пространстве объектов является простой. В данной работе мы рассматриваем, для каких признаков пространств это допущение справедливо, а также предлагаем способы построения таких выборок.

4 Построение признакового пространства

В качестве элементов из пространства функций H мы будем использовать локально-аппроксимирующие модели сегментов.

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

Тогда параметры настроенной модели будут являться образом отображения h

$$h(x) = \arg \min_{w \in \mathbb{R}^{n_g}} \rho(g(w, x), x)$$

4.1 Локально-аппроксимирующие модели

4.1.1 SEMOR

Пусть наряду с сегментом x мы имеем некоторый временной p , который содержит меньшее число элементов. Модель предполагает, что изгиб временного ряда p повторяет форму x , поэтому мы аппроксимируем временной ряд x сегментом p .

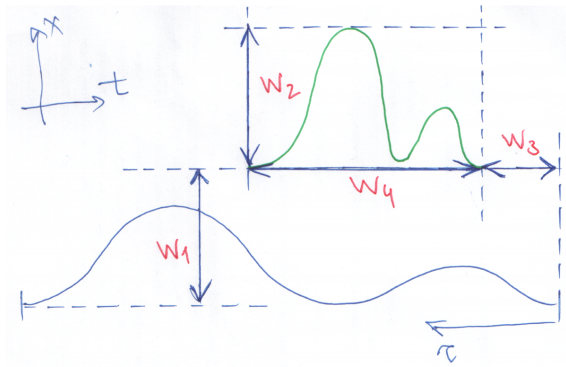
$$w = [w_1, w_2, w_3, w_4]$$

Модель Self-Modeling Regression описывается следующим выражением

$$g(x, w) = w_1 + w_2 p(w_3 + w_4 t)$$

Где в общем случае p — функция формы, однако в частном случае это может быть заранее заданный временной ряд. Параметры w_1, w_2 находятся шагом двухпараметрической линейной регрессии, w_3, w_4 получаются минимизацией DTW, где из

Рис. 2: Иллюстрация параметров метора SEMOR.



наклона пути, результирующего в minimum edit distance мы получаем w_4 , а w_3 — это смещение этой прямой относительно нуля. Следуя приведенной схеме мы получаем следующие оптимальные веса

$$w_{\text{SEMOR}} = [\hat{w}_1, \hat{w}_2, \hat{w}_3, \hat{w}_4]$$

4.1.2 AR-авторегрессия

Является моделью авторегрессии порядка m .

$$g_{\text{AR}}(w, x) = [\hat{x}_1 \dots \hat{x}_n], \text{ где } \hat{x}_i = \begin{cases} x_k & \text{при } k \in [1, m] \\ w_0 + \sum_{i=1}^m w_i x_{k-i} & \text{при } k \in [m+1, n] \end{cases}$$

Оптимальные веса w определяются минимизацией некоторой функции ошибки, например MSE

$$w_{\text{AR}} = \arg \min_{w \in \mathbb{R}^m} \sum_{i=1}^n \|x_i - \hat{x}_i\|^2$$

4.1.3 Фурье-модель

С помощью эмбединга построим траекторную матрицу S для сегмента $x \in \mathbb{R}^n$

$$S = \begin{pmatrix} x_1 & x_2 & \dots & x_m \\ x_2 & x_3 & \dots & x_{m+1} \\ \vdots & \vdots & \ddots & \vdots \\ x_{n-m+1} & x_{n-m+2} & \dots & x_n \end{pmatrix}$$

Каждая следующая строка траекторной матрицы S получается сдвиганием окна длины m в сегменте на единицу. m — единственный параметр SSE. Применяем сингулярное разложение к матрице $X^\top X$

$$X^\top X = V H V^\top$$

Где $H = \text{diag}(\lambda_1 \dots \lambda_m)$ — диагональная матрица собственных значений. Вектор, образованный этими собственными значениями используются в качестве признакового описания сегмента x .

$$w_{\text{SSA}} = [\lambda_1 \dots \lambda_m]$$

4.1.4 Вейвлет-модель

В качестве аппроксимирующей модели сегмента берем обратное дискретное преобразование Фурье, то есть признаковым описанием сегмента является прямое дискретное преобразование Фурье.

$$w_{2j} = \text{Re} \sum_{k=1}^n x_k \exp\left(-\frac{2\pi i}{n} k j\right), j = 1 \dots n$$

$$w_{2j+1} = \text{Im} \sum_{k=1}^n x_k \exp\left(-\frac{2\pi i}{n} k j\right), j = 1 \dots n$$

$$w_{\text{FFT}} = [w_1 \dots w_{2n}] \in \mathbb{R}^{2n}$$

4.2 Построение пространства Z

Итак, набор g состоит из локально аппроксимирующих моделей, каждая из которых отображает сегмент x в оптимальные параметры модели. Имеем следующее отображение

$$[h_1 \dots h_k] : x \mapsto [w_1^* \dots w_k^*]$$

Взятые в совокупности эти отображения порождают промежуточное признаковое пространство Z , элементы которого являются конкатенацией векторов $h_1(x), h_2(x) \dots h_k(x)$.

$$g : X \rightarrow Z$$

5 Алгоритм многоклассовой классификации

Алгоритм обучения выглядит следующим образом. После преобразования пространства \mathbf{X} с l объектами в промежуточное пространство \mathbf{Z} с помощью $l_{\mathbf{h}}$ моделей локальной аппроксимации \mathbf{h} , мы применяем для многоклассовой классификации некий алгоритм машинного обучения. В нашем случае лучше всего подходит (для определения вероятностей принадлежности классу) многомерная логистическая регрессия.

Алгоритм 5.1 Алгоритм обучения

Вход: $\mathbf{X}, \mathbf{Y}, \mathbf{h}, l, l_{\mathbf{h}}$;

Выход: $\hat{\theta}$;

1: инициализировать $\mathbf{u} := \mathbf{y}_1$ (вектор матрицы \mathbf{Y})

2: $\mathbf{z} = []$

3: для $i = 1, \dots, l_{\mathbf{h}}$

$\mathbf{z}[i] := \mathbf{h}[i](\mathbf{X})$

4: инициализировать $\mathbf{Z} = \mathbf{z}[1]$

5: для $i = 2, \dots, l_{\mathbf{h}}$

$\mathbf{Z} := (\mathbf{Z}, \mathbf{z}[i])$

6: $\hat{\theta} = \arg \min(L(f(\mathbf{Z}), \mathbf{Y}))$

Применение алгоритма к новым данным выглядит аналогично. Объекты переносятся в промежуточное пространство с помощью тех же моделей \mathbf{h} , а затем считаются вероятности принадлежности объектов из \mathbf{Z} каждому из классов с помощью оптимального вектора параметров $\hat{\theta}$ из обученной логистической регрессии.

5.1 Выбор моделей локальной аппроксимации

Так как в классе моделей локальной аппроксимации могут присутствовать модели, образы которых от одинаковых объектов будут скоррелированы, качество классификации будет различаться от совокупности моделей в \mathbf{h} . Поэтому строит задача максимизации качества на скользящем контроле. Базовым алгоритмом для выбора промежуточного пространства параметров является минимизация мультиномиального log-loss на кросс-валидации.

Алгоритм 5.2 Применение алгоритма

Вход: $\mathbf{X}, \mathbf{h}, l, l_{\mathbf{h}}, \hat{\theta}$;

Выход: $\hat{\mathbf{y}}$;

- 1: инициализировать $\mathbf{u} := \mathbf{y}_1$ (вектор матрицы \mathbf{Y})
 - 2: $\mathbf{z} = []$
 - 3: для $i = 1, \dots, l_{\mathbf{h}}$
 - 4: $\mathbf{z}[i] := \mathbf{h}[i](\mathbf{X})$
 - 5: инициализировать $\mathbf{Z} = \mathbf{z}[1]$
 - 6: для $i = 2, \dots, l_{\mathbf{h}}$
 - 7: $\mathbf{Z} := (\mathbf{Z}, \mathbf{z}[i])$
 - 8: инициализировать $\hat{\mathbf{y}}$
 - 9: для $i = 1, \dots, l$
 - 10: $\hat{\mathbf{y}}[i] = \arg \max_{k \in K} P(f(\mathbf{X}[i]) = y_k | \hat{\theta})$
-

$$\text{CV}(\mu, \mathbf{X}, \mathbf{Y}) = \frac{1}{|N|} \sum_{n \in N} Q_{\mu}(X_n^l, X_n^k)$$

Для отбора моделей локальной аппроксимации мы будем по очереди включать и выключать модель из вектора функций \mathbf{h} . То есть мы решаем задачу максимизации функционала качества на множестве подмножеств пространства моделей локальной аппроксимации.

$$\text{CV}(\mathbf{h}, \mathbf{X}, \mathbf{Y}) \rightarrow \max_{\mathbf{h} \subset \mathbf{h}_{\text{all}}}$$

Где под \mathbf{h}_{all} понимается вся совокупность моделей локальной аппроксимации, которые у нас есть. Если количество моделей не очень большое, то подойдет такой базовый алгоритм.

Алгоритм 5.3 Отбор моделей локальной аппроксимации

Вход: $\mathbf{X}, \mathbf{Y}, \mathbf{h}_{\text{all}}, l, l_{\mathbf{h}}, N$;

Выход: \mathbf{h}_{best} ;

```

1:  $\mathbf{h}_{\text{best}} := 0$ 
2:  $S_{\text{best}} := -\infty$ 
3: для  $\mathbf{h}$  in Combinations( $\mathbf{h}_{\text{all}}$ )
4:    $\mathbf{s} = []$ 
5:   для  $n = 1, \dots, N$ 
6:      $\mathbf{X}_{\text{train}}, \mathbf{X}_{\text{test}} = \text{Split}(\mathbf{X}, n)$ 
7:      $\mathbf{Y}_{\text{train}}, \mathbf{Y}_{\text{test}} = \text{Split}(\mathbf{Y}, n)$ 
8:      $\theta = \text{Learn}(\mathbf{X}_{\text{train}}, \mathbf{Y}_{\text{train}}, \mathbf{h})$ 
9:      $\mathbf{Y}_{\text{predicted}} = \text{Predict}(\mathbf{X}_{\text{test}}, \mathbf{h}, \theta)$ 
10:     $\mathbf{s}[i] = \text{LogLoss}(\mathbf{Y}_{\text{predicted}}, \mathbf{Y}_{\text{test}})$ 
11:     $S_{\text{mean}} = \frac{1}{|N|} \sum \mathbf{s}[i]$ 
12:    если  $S_{\text{best}} < S_{\text{mean}}$  то
13:       $S_{\text{best}} := S_{\text{mean}}$ 
14:       $\mathbf{h}_{\text{best}} := \mathbf{h}$ 

```

Список литературы

- [1] Argyro Kampouraki, George Manis, and Christophoros Nikou. Heartbeat time series classification with support vector machines. In *IEEE Transactions on Information Technology in Biomedicine*, volume 13, pages 512–518, 2009.
- [2] Damian Eads, Daniel Hill, Sean Davis, Simon Perkins, Junshui Ma, Reid Porter, and James Theiler. Genetic algorithms and support vector machines for time series classification. *Proceedings of SPIE - The International Society for Optical Engineering*, 4787:74–85, 2002.
- [3] Michael Hüskén and Peter Stagge. Recurrent neural networks for time series classification, 2003.
- [4] Yi Zheng, Qi Liu, Enhong Chen, Yong Ge, and J. Leon Zhao. Time series classification using multi-channels deep convolutional neural networks. In *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, volume 8485 LNCS, pages 298–310, 2014.

- [5] Кузнецов МП and Ивкин НП. Алгоритм классификации временных рядов акселерометра по комбинированному признаковому описанию. *Машинное обучение и анализ данных*, pages 1471–1483, 2015.
- [6] Карасикова МЕ and Стрижов ВВ. Классификация временных рядов в пространстве параметров порождающих моделей . *Информ. и её примен.*, page 121–131, 2016.
- [7] H. Hassani. Singular Spectrum Analysis: Methodology and Comparison. *Journal of Data Science*, 5:239–257, 2007.