Погружение временных рядов с высокой волатильностью в метрическое пространство

A Preprint

Эйнуллаев Алтай
Кафедра интеллектуальных систем
Московский физико-технический институт
Долгопрудный
einullaev.ae@phystech.edu

Яковлев Константин
Кафедра интеллектуальных систем
Московский физико-технический институт
Долгопрудный
iakovlev.kd@phystech.edu

Abstract

Рассматривается задача прогнозирования финансовых временных рядов. Основными особенностями таких временных рядов являются высокая волатильность и высокая попарная ковариация. Классическим подходом к решению задачи является выполнение прогноза в исходном пространстве. Новый метод заключается в переходе в пространство попарных расстояний между временными рядами, осуществлении прогноза в нем и переходе обратно в исходное пространство. Для его реализации необходимо ввести функцию расстояния между временными рядами, которая должна удовлетворять определенным свойствам. В данной статье изучаются эти свойства и проводятся сравнения различных метрик на основе численных экспериментов.

Keywords Временные ряды · Метрика · Ковариация

1 Introduction

В текущей статье исследуется задача погружения временных рядов в метрическое пространство. Набору временных рядов ставится в соответствие матрица попарных расстояний и появляется возможность перейти от прогнозирования набора временных рядов к прогнозированию матрицы попарных расстояний. При этом выбор функции расстояния осуществляется так, чтобы по полученной матрице расстояний можно было восстановить прогноз для набора временных рядов.

В статистике, обработке сигналов и многих других областях под временным рядом понимаются последовательно измеренные через некоторые (зачастую равные) промежутки времени данные. [1] Прогнозирование временных рядов заключается в построении модели для предсказания будущих событий основываясь на известных событиях прошлого, предсказания будущих данных до того как они будут измерены.

Одними из хорошо известных, классических методов прогнозирования временных рядов являются экспоненциальное сглаживание (англ. Exponential Smoothing) [2], LSTM (англ. Long Short-Term Memory) [3], ARIMA (англ. autoregressive integrated moving average) [4]. Главным отличием исследуемого метода от вышеперечисленных является то, что временные ряды прогнозируются при помощи прогнозирования матрицы попарных расстояний.

В качестве простейшей метрики рассматривается ковариация между временными рядами. [5] Таким образом, для набора временных рядов получаем матрицу ковариации. Стоит заметить, что матрица ковариации (матрица попарных расстояний) вычисляется в каждый момент времени. Альтернативные варианты метрики выбираются из класса ядер [6].

Численные эксперименты проводятся на трех видах данных: синтетические, сигналы коры головного мозга, финансовые временные ряды. Эксперимент состоит из выполнения прогноза временного ряда при помощи прогнозирования матрицы попарных расстояний. В качестве прогностической модели выбирается линейная регрессия с использованием SSA (англ. Singular Spectrum Analysis) [7]. По результатам экспериментов проводится анализ точности прогноза и его устойчивости в зависимости от выбранной метрики и вида данных. Цель эксперимента состоит в оптимальном выборе функции попарных расстояний для выполнения прогноза.

2 Problem statement

Пусть $\mathrm{X} = \{\mathbf{x} = [x_1, \dots, x_\mathrm{m}]^T | x_i \in R\}$ — множество временных рядов, заданных своей реализацией. Обозначим через $\mathbf{Y} \in R^{n \times m}$ заданный набор из n временных рядов:

$$\mathbf{Y} = [\mathbf{x}^{(1)}, \dots, \mathbf{x}^{(n)}]^T. \tag{1}$$

Через $\mathbf{Y}_t \in R^{n \times t}$ обозначим $t < m$ первых столбцов $\mathbf{Y}$:

$$\mathbf{Y}_t = [\mathbf{x}^{(1)}_{1:t}, \dots, \mathbf{x}^{(n)}_{1:t}]^T. \tag{2}$$

.

Определим функцию расстояния между временными рядами: $d : \mathrm{X} \times \mathrm{X} \to R$, удовлетворяющую условиям Мерсера [8]:

1. $d(\mathbf{x}^{(1)}, \mathbf{x}^{(2)}) = d(\mathbf{x}^{(2)}, \mathbf{x}^{(1)})\ \forall \mathbf{x}^{(1)}, \mathbf{x}^{(2)} \in \mathrm{X}$.

2. $\forall n \in N, \forall \mathbf{x}^{(1)}, \dots, \mathbf{x}^{(n)} \in X$ матрица $\Sigma \in S_n$, составленная из попарных расстояний между элементами, является неотрицательно определенной.

Обозначим расстояние между временными рядами $\mathbf{x}^{(i)} = [x^{(i)}_1, \dots, x^{(i)}_t], \mathbf{x}^{(j)} = [x^{(j)}_1, \dots, x^{(j)}_t]$ следующим образом:

$$d_t(\mathbf{x}^{(i)}, \mathbf{x}^{(j)}) = d_t(i, j) \tag{3}$$

Таким образом, в каждый момент времени t набору временных рядов $\mathbf{Y}_t$ поставлена в соответствие матрица попарных расстояний $\Sigma_t \in \mathcal{S}^+_n$ (симметричная, неотрицательно определенная матрица):

$$\mathbf{Y}_t \Rightarrow \Sigma_t = \begin{pmatrix} d_t(\mathbf{x}^{(1)}, \mathbf{x}^{(1)}) & d_t(\mathbf{x}^{(1)}, \mathbf{x}^{(2)}) & \dots & d_t(\mathbf{x}^{(1)}, \mathbf{x}^{(n)}) \\ d_t(\mathbf{x}^{(2)}, \mathbf{x}^{(1)}) & d_t(\mathbf{x}^{(2)}, \mathbf{x}^{(2)}) & \dots & d_t(\mathbf{x}^{(2)}, \mathbf{x}^{(n)}) \\ \vdots & \vdots & \ddots & \vdots \\ d_t(\mathbf{x}^{(n)}, \mathbf{x}^{(1)}) & d_t(\mathbf{x}^{(n)}, \mathbf{x}^{(2)}) & \dots & d_t(\mathbf{x}^{(n)}, \mathbf{x}^{(n)}) \end{pmatrix} \tag{4}$$

Дадим описание двум используемым прогностическим моделям.

2.1 Авторегрессия

Пусть имеем реализацию матрицы попарных расстояний из m компонент: $\Sigma = [\Sigma_1, \Sigma_2, \dots, \Sigma_\mathrm{m}]$. Пусть $\Sigma_{\mathrm{m}+1}$ - прогнозируемая матрица. Авторегрессионная модель:

$$\hat{\Sigma}_{\mathrm{m}+1} = \sum_{i=1}^{m} a_i \Sigma_i + \epsilon_{\mathrm{m}+1} \tag{5}$$

где $a_1, \dots, a_\mathrm{m} \in R$ — параметры модели (коэффициенты авторегрессии), $\epsilon_{\mathrm{m}+1} \in R^{\mathrm{m} \times \mathrm{m}}$ — белый шум. Обозначим через $mathbfa \in R^\mathrm{m}$ вектор составленный из параметров модели. Вектор оптимальных параметров модели $\hat{\mathbf{a}} \in R^\mathrm{m}$ определяется с помощью решения задачи оптимизации:

$$\hat{\mathbf{a}} = \arg \min_{\mathbf{a} \in R^{m \times m}} \|\Sigma_{\mathrm{m}+1} - \hat{\Sigma}_{\mathrm{m}+1}\|^2_F. \tag{6}$$

2.2 Многомерная гусеница (MSSA)

Пусть имеем реализацию матрицы попарных расстояний из m компонент: $\Sigma = [\Sigma_1, \Sigma_2, \dots, \Sigma_{\mathrm{m}}]$. Определим L - ширина окна, $\mathrm{K} = \mathrm{m} - \mathrm{L} + 1$. В силу того, что матрица симметричная, то для каждой компоненты матрицы попарных расстояний, где $i \leqslant j$, построим матрицу траекторий (матрица Ганкеля):

$$\mathbf{H}_{ij} = \begin{pmatrix} d_1(\mathbf{x}^{(i)}, \mathbf{x}^{(j)}) & d_2(\mathbf{x}^{(i)}, \mathbf{x}^{(j)}) & \dots & d_{\mathrm{K}}(\mathbf{x}^{(i)}, \mathbf{x}^{(j)}) \\ d_2(\mathbf{x}^{(i)}, \mathbf{x}^{(j)}) & d_3(\mathbf{x}^{(i)}, \mathbf{x}^{(j)}) & \dots & d_{\mathrm{K}+1}(\mathbf{x}^{(i)}, \mathbf{x}^{(j)}) \\ \vdots & \vdots & \ddots & \vdots \\ d_{\mathrm{L}}(\mathbf{x}^{(i)}, \mathbf{x}^{(j)}) & d_{\mathrm{L}+1}(\mathbf{x}^{(i)}, \mathbf{x}^{(j)}) & \dots & d_{\mathrm{m}}(\mathbf{x}^{(i)}, \mathbf{x}^{(j)}) \end{pmatrix}. \tag{7}$$

Строим блочную матрицу траекторий:

$$\mathbf{H} = [\mathbf{H}_{11} : \mathbf{H}_{12} : \mathbf{H}_{13} : \mathbf{H}_{1n} : \mathbf{H}_{22} : \dots : \mathbf{H}_{nn}] \tag{8}$$

Сингулярное разложение симметричной, положительно полуопределенной матрицы $\mathbf{H}\mathbf{H}^T = \mathbf{U}\Lambda\mathbf{U}^T$, где Λ - диагональная матрица $\mathrm{L} \times \mathrm{L}$ собственных значений $\lambda_1 \geq \dots \geq \lambda_{\mathrm{L}}$, $\mathbf{U} = [\mathbf{u}_1, \dots, \mathbf{u}_{\mathrm{L}}]$ - ортогональная матрица собственных векторов матрицы $\mathbf{H}\mathbf{H}^T$. Пусть $d = \max\{i | \lambda_i > 0, i \in \{1, \dots, \mathrm{L}\}\}$. Тогда исходную матрицу траекторий можно представить в виде:

$$\mathbf{H} = \sum_{i=1}^{d} \sqrt{\lambda_i} \mathbf{u}_i \mathbf{v}_i^T = \mathbf{H}^{(1)} + \dots + \mathbf{H}^{(d)} = \sum_{i=1}^{r} \mathbf{H}^{(i)} + \sum_{i=r+1}^{d} \mathbf{H}^{(i)}, \tag{9}$$

где $\mathbf{v}_i = \mathbf{H}^T \mathbf{u}_i / \sqrt{\lambda_i}$ $\mathbf{H}_r = \sum\limits_{i=1}^{r} \mathbf{H}^{(i)}$ — сигнальные компоненты, а $\mathbf{H} - \mathbf{H}_r$ — шумовые. Полученную матрицу можно представить в виде:

$$\mathbf{H}_r = [\widetilde{\mathbf{H}}_{11} : \widetilde{\mathbf{H}}_{12} : \dots : \widetilde{\mathbf{H}}_{nn}] \tag{10}$$

где каждая из матриц $\widetilde{\mathbf{H}}_{ij} \in R^{\mathrm{m} - \mathrm{L} + 1 \times \mathrm{L}}$. Восстановим исходные расстояния с помощью антидиагонального усреднения элементов соответствующих матриц:

$$\widetilde{d}_t(i,j) = \frac{1}{2t-1} \sum_{k,l : k+l-1=t} \widetilde{h}_{kl}^{(ij)}, t \in \{1, \dots, m\} \tag{11}$$

где $\widetilde{h}_{kl}^{(ij)}$ — элемент на пересечении k-ой строки и l-го столбца матрицы $\mathbf{H}_{ij}$. Обозначим через $\mathbf{u}_j^{\triangledown} \in R^{\mathrm{L}-1}$ первые L - 1 компоненты собственного вектора $\mathbf{u}_j$, а $\pi_j \in R$ — последнюю компоненту вектора $\mathbf{u}_j$, $(j = 1, \dots, r)$. Определим $v^2 = \sum\limits_{j=1}^{r} \pi_j^2$ и $\mathbf{p} \in R^{\mathrm{L}-1}$:

$$\mathbf{p} = \frac{1}{1 - v^2} \sum_{j=1}^{r} \pi_j \mathbf{u}_j^{\triangledown}. \tag{12}$$

При $v^2 < 1$ возможен прогроз:

$$[\widetilde{d}_{\mathrm{m}+1}(1,1), \widetilde{d}_{\mathrm{m}+1}(1,2), \dots, \widetilde{d}_{\mathrm{m}+1}(n,n)]^T = \mathbf{p}^T \mathbf{Z} \tag{13}$$

где $\mathbf{Z} \in R^{(\mathrm{L}-1) \times (\frac{n(n+1)}{2})}$, $\mathbf{Z} = [\mathbf{z}^{(11)}, \dots, \mathbf{z}^{(nn)}]$, $\mathbf{z}^{(ij)} = [\widetilde{d}_{\mathrm{m}-\mathrm{L}+2}(i,j), \dots, \widetilde{d}_{\mathrm{m}}(i,j)]^T$. Таким образом, получаем:

$$\hat{\Sigma}_{\mathrm{m}+1} = [\widetilde{d}_{\mathrm{m}+1}(i,j)]_{i,j=1}^{n} \tag{14}$$

где $\widetilde{d}_{\mathrm{m}+1}(i,j) = \widetilde{d}_{\mathrm{m}+1}(j,i)$, при $i > j$.

2.3 Задача прогнозирования набора временных рядов

В качестве критериев качества прогноза временных рядов используются

$$\text{MAPE} = \frac{1}{m}\sum_{i=1}^{m}\frac{|x_t^{(i)} - \hat{x}_t^{(i)}|}{|x_t^{(i)}|}, \tag{15}$$

$$\text{MSE} = \frac{1}{m}\sum_{i=1}^{m}(x_t^{(i)} - \hat{x}_t^{(i)})^2. \tag{16}$$

Также выбираются $p < m$ временных рядов из набора так, чтобы отказ от прогнозирования соответствующих $m - p$ временных рядов существенно повышал качество MAPE, MSE.

2.4 Выбор функции попарных расстояний

Пусть $\mathcal{F} = \{d^{(1)}, \dots, d^{(s)}\}$ — множество функций попарных расстояний, из которого нужно выбрать оптимальный вариант. Пусть $\hat{\Sigma}(d)$ — спрогнозированная, а $\Sigma(d)$ — точная матрица попарных расстояний при использовании $d \in \mathcal{F}$. Тогда оптимальный по точности выбор функции попарных расстояний:

$$d_{\text{acc}} = \arg\min_{d\in\mathcal{F}} \sigma^2(d) = \arg\min_{d\in\mathcal{F}} \|\hat{\Sigma}(d) - \Sigma(d)\|_{\text{F}}^2. \tag{17}$$

Пусть $\widetilde{\mathbf{x}}^{(i)} = \mathbf{x}^{(i)} + \varepsilon_i$, $i = 1, \dots, n$ — зашумленный исходный набор временных рядов, где $\varepsilon_i \in R^{\text{N}}$. Пусть $\widetilde{\sigma}^2(d)$ — точность прогноза матрицы попарных расстояний зашумленного набора временных рядов при использовании $d \in \mathcal{F}$. Тогда наболее устойчивый выбор функции попарных расстояний:

$$d_{\text{stable}} = \arg\min_{d\in\text{F}} |\widetilde{\sigma}^2(d) - \sigma^2(d)| \tag{18}$$

Список литературы

[1] Robert H Shumway, David S Stoffer, and David S Stoffer. Time series analysis and its applications, volume 3. Springer, 2000.

[2] Everette S. Gardner Jr. Exponential smoothing: The state of the art. Journal of Forecasting, 1985.

[3] J. Schmidhuber S. Hochreiter. Long short-term memory. Neural Computation, 1997.

[4] Jenkins Box. Time Series Analysis: Forecasting and Control. San Francisco: Holden-Day, 1970.

[5] Steven Diamond Stephen Boyd, Enzo Busseti and Ronald N. Kahn. Multi-period trading via convex optimization. Foundations and Trends in Optimization, 2017.

[6] John Shawe-Taylor and Nello Cristianini. Kernel methods for pattern analysis. Cambridge university press, 2004.

[7] Robert Vautard, Pascal Yiou, and Michael Ghil. Singular-spectrum analysis: A toolkit for short, noisy chaotic signals. Physica D: Nonlinear Phenomena, 58(1-4):95–126, 1992.

[8] Benyamin Ghojogh, Ali Ghodsi, Fakhri Karray, and Mark Crowley. Reproducing kernel hilbert space, mercer's theorem, eigenfunctions, nystr\"om method, and use of kernels in machine learning: Tutorial and survey. arXiv preprint arXiv:2106.08443, 2021.