МЕТОД SSA ДЛЯ ПРОГНОЗИРОВАНИЯ ВРЕМЕННЫХ РЯДОВ

И. В. Фадеев

1. Описание алгоритма

1.1. **Постановка задачи.** Наблюдается система функций дискретного аргумента $(f_i^{(k)})_{i=1}^N$, где k = 1, ..., s. s — число временных рядов, k — номер ряда, N — длина временного ряда, i — номер отсчета. Требуется разложить ряд в сумму компонент (используя метод главных компонент, см. описание алгоритма), интерпретировать каждую компоненту, и построить продолжение ряда $(f_i^{(k)})_{i=1}^{N+M}$ по выбранным компонентам.

1.2. **Построение матрицы наблюдений.** Рассмотрим сначала одномерный временной ряд $(f_i)_{i=1}^N$. Выберем n такое, что $0 < n \leq N-1$ — время жизни многомерной гусеницы. Пусть $\sigma = N - n + 1$ — длина гусеницы. Построим последовательность из n векторов в R^σ следующего вида:

$$Y^{(l)} \in R^\sigma,$$
$$Y^{(l)} = (f_{i+l-1})_{i=1}^\sigma$$

Обозначим

$$Z = (Y^{(1)}, \ldots, Y^{(n)}) :$$

Будем называть Z нецентрированной матрицей наблюдений, порождённой гусеницей со временем жизни n.

В случае многомерного временного ряда матрицей наблюдения называется столбец из матриц наблюдений, соответствующих каждой из компонент.

Проводимый в дальнейшем анализ главных компонент может проводиться как по центрированной, так и по нецентрированной выборкам. Для упрощения выкладок рассмотрим простейший нецентрированный вариант.

1.3. **Анализ главных компонент.** Рассмотрим ковариационную матрицу полученной выборки:

$$C = \frac{1}{n} Z Z^T.$$

Выполним её svd-разложение:

$$C = V \Lambda V^T,$$

Date: 22 апреля 2011 г.

Научный руководитель В. В. Стрижов.

где $\Lambda = diag(\lambda_1, \ldots, \lambda_\tau)$ — диагональная матрица собственных чисел, $V = (v^{(1)}, \ldots, v^{(\tau)})$, $(v^{(i)})^T v^{(j)} = \delta_{ij}$ — ортогональная матрица собственных векторов.

Далее рассмотрим систему главных компонент:

$$U = V^T Z, U = (U^{(1)}, \ldots, U^{(\tau)})^T.$$

После проведения анализа главных компонент обычно предполагается проведение операции восстановления исходной матрицы наблюдений по некоторому поднабору главных компонент, т. е. для $V' = (v^{(i_1)}, \ldots, v^{(i_r)})$ и $U' = V'^T Z$ вычисляется матрица $Z' = V'U'$.

Далее восстанавливаются исходные последовательности. В одномерном случае i-ая компонента восстановленного ряда есть среднее значение по i-ой диагонали восстановленной матрицы наблюдений Z'.

В многомерном случае усреднение проводится с учётом того, что матрица наблюдений состоит из подматриц, соответствующих каждой компоненте ряда:

$$f'^{(k)}_m = \left\{ \begin{array}{ll} \frac{1}{m}\sum_{i=1}^{m} x_i^{(m-i+1,k)} & 1 \le m \le \sigma, \\ \frac{1}{\sigma}\sum_{i=1}^{\sigma} x_i^{(m-i+1,k)} & \sigma \le m \le n, \\ \frac{1}{N-m+1}\sum_{i=1}^{N-m+1} x_{i+m-n}^{(n-i+1,k)} & n \le m \le N. \end{array} \right\}$$

1.4. **Прогноз.** Числовой ряд $(f_i)_{i=1}^{N+1}$ называется продолжением ряда $(f_i)_{i=1}^{N}$, если порождаемая им при гусеничной обработке выборка лежит в той же гиперплоскости, что и у исходного ряда. Пусть у нас есть некоторый набор выбранных главных компонент $i_1, i_2, \ldots, i_r$. Определим

$$w = \begin{pmatrix} v_\sigma^{(i_1)} & v_\sigma^{(i_2)} & \ldots & v_\sigma^{(i_r)} \\ v_{2\sigma}^{(i_1)} & v_{2\sigma}^{(i_2)} & \ldots & v_{2\sigma}^{(i_r)} \\ \vdots & \vdots & \ddots & \vdots \\ v_\tau^{(i_1)} & v_\tau^{(i_2)} & \ldots & v_\tau^{(i_r)} \end{pmatrix}$$

и

$$V_* = \begin{pmatrix} v_1^{(i_1)} & v_1^{(i_2)} & \ldots & v_1^{(i_r)} \\ \vdots & \vdots & \ddots & \vdots \\ v_{\sigma-1}^{(i_1)} & v_{\sigma-1}^{(i_2)} & \ldots & v_{\sigma-1}^{(i_r)} \\ v_{\sigma+1}^{(i_1)} & v_{\sigma+1}^{(i_2)} & \ldots & v_{\sigma+1}^{(i_r)} \\ \vdots & \vdots & \ddots & \vdots \\ v_{2\sigma-1}^{(i_1)} & v_{2\sigma-1}^{(i_2)} & \ldots & v_{2\sigma-1}^{(i_r)} \\ \vdots & \vdots & \ddots & \vdots \\ v_{\tau-1}^{(i_1)} & v_{\tau-1}^{(i_2)} & \ldots & v_{\tau-1}^{(i_r)} \end{pmatrix}$$

Также положим

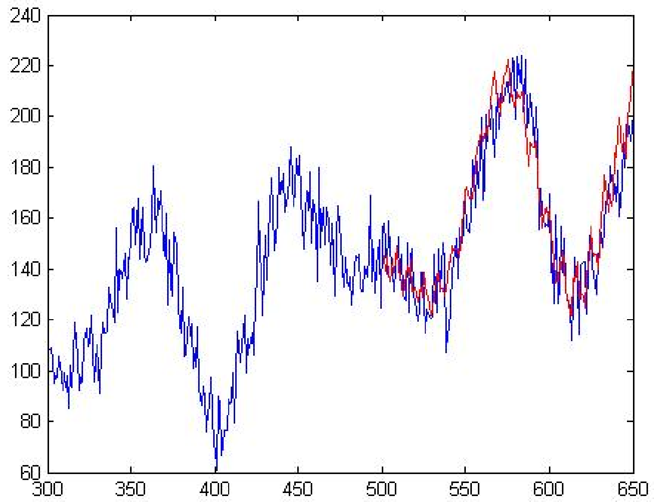
$$Q = \left(f^{(1)}_{N-\sigma+2}, \ldots, f^{(1)}_{N}, f^{(2)}_{N-\sigma+2}, \ldots, f^{(2)}_{N}, \ldots, f^{(s)}_{N-\sigma+2}, \ldots, f^{(s)}_{N}\right)^T$$

Тогда прогнозируемые значения системы в точке $N+1$ вычисляются по формуле:

$$f_{N+1} = w(V_*^T V_*)^{-1} V_*^T Q.$$

2. Тестовый прогноз

Протестируем алгоритм на тестовом временном ряде, синтезируемом как сумма линейной компоненты (тренда), двух сезонных компонент (синусоид разной частоты) и нормального шума. На графике представлен исходный временной ряд (синий) и прогнозируемый с помощью алгоритма SSA (красный):



Московский физико-технический институт, ФУПМ, каф. «Интеллектуальные системы»