УДК 519.246.8

Л.М. ЛЮБЧИК, д-р техн. наук, проф., ***А.А. МИРОШНИЧЕНКО***

МОНИТОРИНГ ДИНАМИЧЕСКИХ ПРОЦЕССОВ НА ОСНОВЕ СИНГУЛЯРНО-СПЕКТРАЛЬНОГО АНАЛИЗА

Наведено результати застосування методу головних компонент та сингулярно - спектрального аналізу до задач структурного аналізу нестаціонарних часових рядів. Наведено приклади застосування зазначених методів у практичних задачах моніторингу теплоспоживання.

Введение. Задача мониторинга динамических процессов состоит в непрерывном контроле их состояния с целью обнаружения моментов изменения их параметров и характеристик. В частности, мониторинг позволяет обнаруживать моменты возникновения отклонений от нормального хода процесса, а также диагностировать возможные причины их возникновения. Современные системы мониторинга используют различные методы моделирования и идентификации случайных процессов в сочетании с методами статистических решений. При этом основная проблема связана с построением адекватной модели исследуемого процесса и идентификацией ее параметров в реальном масштабе времени с учетом возможных изменений.

Постановка задачи. Задачи мониторинга характеристик временного ряда основаны на обнаружении момента времени изменения его свойств. Построение моделей для мониторинга производственных процессов, диагностики сбоев и отказов сводится к решению задачи о "разладке", то есть обнаружению происходящего в неизвестный момент времени скачкообразного изменения свойств наблюдаемого временного ряда.

Формальная постановка задачи обнаружения разладки [1] состоит в следующем: пусть дана случайная последовательность $x_1, x_2, \dots x_N$, которая в некоторый момент времени t_0 скачкообразно меняет свои свойства, однозначно определяемые вектором параметров θ. Это означает, что при $t \le t_0 - 1$ вектор $\theta = \theta_1$, а при $t \ge t_0$ вектор $\theta = \theta_2$. Необходимо, наблюдая последовательность $x_1, x_2, \dots x_N$, обнаруживать момент разладки t_0.

Как указано в [1] существует типа задач о разладке. Первый, получивший название "Задача о скорейшем обнаружении разладки", описывается следующей постановкой: пусть последовательно во времени поступает информация о случайном процессе, и в некоторый момент времени происходит изменение какой-либо его вероятностной характеристики. Необходимо скорейшим образом обнаружить это изменение после того, как оно произошло, и при этом ограничить некоторой величиной вероятность ложного сигнала тревоги.

Второй тип носит название "Апостериорных задач о разладке". Постановка задачи здесь следующая: пусть предъявлена реализация случайного процесса. Если данная выборка не является статистически однородной, то возникает задача обнаружения моментов времени изменения вероятностных характеристик временного ряда.

В [2] рассмотрены различные постановки задачи о разладке. В [1] приведены различные варианты критериев для построения алгоритмов решения задач о разладке. Наибольший практический интерес представляет алгоритм кумулятивных сумм. Его идея состоит в анализе поведения кумулятивной суммы [1]:

$$S_t = S_{t-1} + ln(\omega(x_t / \theta_2) / \omega(x_t / \theta_1)) \quad (1)$$

где $\omega(x_t / \theta_2)$ - плотность распределения вероятностей случайных величин $x_{t_0},\dots x_t$ с момента возникновения разладки (t_0);

$\omega(x_t / \theta_1)$ - плотность распределения вероятностей случайных величин $x_1,\dots x_{t_0-1}$ до момента возникновения разладки ($t_0 - 1$);

Задача о разладке представляет особый интерес для решения проблем текущего контроля производства. Суть такой задачи может заключаться в последовательном контроле и наискорейшем обнаружении изменения параметров контролируемого производственного процесса при определенном установленном уровне ложных сигналов о разладке. При этом важным является не только наискорейшее обнаружение самого факта сбоя или отказа, но и определение причины их возникновения.

Метод главных компонент. Обнаружение разладок при анализе временных рядов сложной структуры требует их предварительного анализа с целью выделения наиболее существенных компонент, несущих основную информацию об отклонениях параметров. Из анализа зарубежных литературных источников [3-6] видно, что в течение последних лет для такого предварительного анализа широко используется метод PCA (Principal Components Analysis).

Формально идея метода анализа главных компонент (PCA) представленная в [7, 8, 9, 10] состоит в следующем.

Пусть X - матрица данных размерностью $(m \times n)$. Столбцы матрицы X - переменные (признаки объектов или измеряемые параметры процесса). Предположим, что матрица X приведена к стандартной форме и все ее столбцы имеют нулевое среднее значение и дисперсию равную единице. Далее по матрице X находят корреляционную матрицу параметров. После чего находят собственные числа корреляционной матрицы, упорядоченные в порядке убывания: $\lambda_1 \geq .. \geq \lambda_i \geq \dots \geq \lambda_n$ (в этом случае дисперсия i- ой главной компоненты равна i-му собственному числу λ_i) и соответствующие им собственные векторы $\omega_1,\dots \omega_i,\dots \omega_n$, образующие ортогональный базис. Из этих векторов составляется ортогональная матрица $W = (\omega_1,\dots \omega_i,\dots \omega_n)$, связывающая параметры процесса и главные компоненты [7, 8]:

$$x_i = \sum_{j=1}^{n} \omega_{ij} f_j \qquad (i, j = 1, \ldots, n) \tag{2}$$

Здесь x_i - i -я измеряемая переменная; f_i - i -я главная компонента; ω_{ij} - вес j -ой компоненты в i -ой переменной.

Модель, полученную при учете только k первых собственных векторов, связанных с систематическим изменением данных, можно считать моделью неполного порядка:

$$x_i = \sum_{j=1}^{k} \omega_{ij} f_j + \xi_i \tag{3}$$

где ξ_i - помехи или шум.

Одним из недостатков построения модели для мониторинга на основе метода главных компонент является слабая чувствительность такой модели к обнаружению небольших отклонений в процессе. Так, очень слабое отклонение параметров процесса от стандартных значений может быть не зафиксировано, поскольку в саму модель неполного порядка заложен вектор, характеризующий помехи или шум.

Для решения такой проблемы предлагается более детальный анализ структуры временного ряда, составленного на основе данных о протекании процесса с помощью сингулярно-спектрального анализа.

Сингулярно - спектральный анализ. Для исследования структуры временных рядов часто применяется метод сингулярно-спектрального анализа - Singular Spectrum Analysis (SSA) [11, 12]. Метод SSA основан на выделении главных компонент анализируемого временного ряда [11, 12].

Алгоритм метода сингулярно - спектрального анализа выглядит следующим образом: исследуется временной ряд $F_N = (f_0, \ldots, f_{N-1})$ длины N. Первым следует этап вложения, то есть формирование траекторной матрицы. Основным шагом здесь является выбор целого числа L - длины окна, $2 \le L \le N-1$. При исследовании структуры временного ряда рекомендуется использовать несколько вариантов длинны окна [12].

После выбора параметра L, значениями исходного временного ряда последовательно заполняют столбцы траекторной матрицы X. Столбцы матрицы X имеют вид:

$$X_j = \left(f_{j-1}, \ldots, f_{j+L-2}\right)^{\mathrm{T}}, \; 1 \le j \le K \tag{4}$$

где $K = N - L + 1$.

Далее следует этап сингулярного разложения. Вычисляется матрица $R = XX^T$. Находятся собственные числа $\lambda_1, \ldots, \lambda_L$ $\left(\lambda_1 \ge \ldots \ge \lambda_L \ge 0\right)$ и соответствующие этим числам собственные векторы $U_1, \ldots, U_L$ матрицы R.

Вычисляют $\sqrt{\lambda_i}$ - сингулярные числа и находят факторные векторы:

$$V_i = \frac{1}{\sqrt{\lambda_i}} X^T U_i, \quad (i = 1,\ldots,d), \quad d = max\{\, i, \lambda_i > 0\} \tag{5}$$

Получаем сингулярное разложение траекторной матрицы X :

$$X = \sum_{i=1}^{d} X_i = \sum_{i=1}^{d} \sqrt{\lambda_i} U_i V_i^T \;. \tag{6}$$

Для данного сингулярного разложения набор $\sqrt{\lambda_i}, U_i, V_i$ - это i-я собственная тройка; вектор $Z_i = \sqrt{\lambda_i} V_i$, – вектор i-х главных компонент; матрицы X_i - элементарные (имеют ранг, равный 1). Каждое собственное число λ_i характеризует вклад матрицы X_i в сингулярное разложение.

Сингулярное разложение можно интерпретировать по аналогии с методом главных компонент в статистике, как разложение всех столбцов матрицы X по базису из их главных векторов.

Следующий этап - группировка. Целью группировки является разделение временного ряда на аддитивные компоненты. Процедура группировки позволяет разделить все множество индексов $(i = 1,\ldots,d)$ на m непересекающихся подмножеств: $I_1,\ldots,I_m$. Каждому подмножеству I соответствует результирующая матрица X_I, определяющаяся как сумма матриц, входящих в подмножество:

$$X = X_{I_1} + X_{I_2} + \ldots + X_{I_m} \,. \tag{7}$$

Эта процедура основана на анализе собственных чисел λ_i а так же векторов U_i и V_i. Анализируя форму сингулярных векторов одной из собственных троек, можно предсказать поведение соответствующей компоненты ряда, восстановленной с помощью этой тройки. Чем больше сингулярное число, тем больше вклад соответствующей компоненты [12].

Последним этапом метода SSA является диагональное усреднение. На этом этапе проводится восстановление рядов $F_N^{(j)}$ по сгруппированным матрицам X_{I_j}, каждая матрица переводится в новый ряд длины N. Значения ряда задаются как средние значения элементов матрицы вдоль соответствующих побочных диагоналей.

В результате применения метода сингулярно - спектрального анализа, исходный ряд $F_N = (f_0,\ldots,f_{N-1})$ раскладывается на сумму m рядов $F_N^{(j)}$, то есть на аддитивные составляющие.

Пример практического применения. В данной работе по описанной методике исследовались данные потребления электроэнергии и природного газа с целью обнаружения моментов изменения поведения потребителей. Использовались данные потребления электроэнергии и природного газа различными районами города Харькова за период сентябрь-октябрь 1996 года приведенные в [13] и представленные на рис.1. Для анализа структуры временного ряда и выделения главных компонент был использован программный продукт [14].

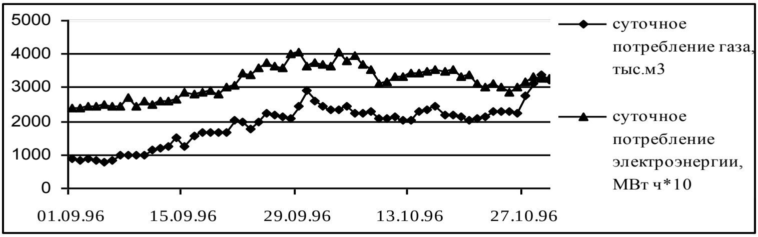

Рис. 1. Суточное потребление электроэнергии и газа одним из районов г. Харькова.

При сингулярном разложении ряда данных потребления природного газа были выделены следующие главные компоненты (рис. 2):

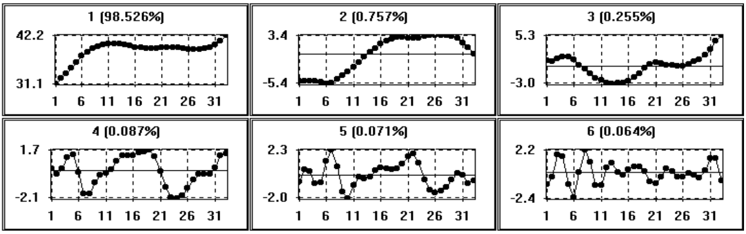

Рис. 2. Первые главные компоненты, полученные при разложении ряда данных потребления природного газа.

Результаты, полученные при сингулярном разложении ряда данных потребления электроэнергии, представлены на рис. 3.

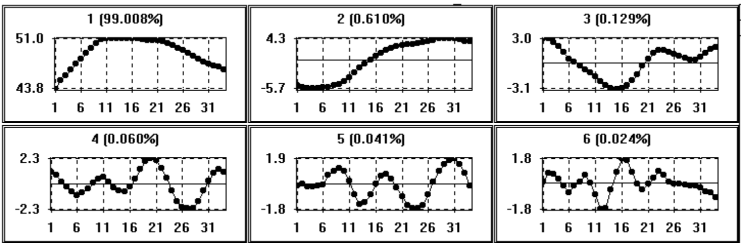

Рис. 3. Первые главные компоненты, полученные при разложении ряда данных потребления электроэнергии.

Анализируя полученные результаты можно сделать вывод, что выделенные главные компоненты 2 и 3, для ряда данных потребления природного газа и ряда данных потребления электроэнергии, имеют схожую структуру и возможно отражают один и тот же фактор или явление. Восстановление рядов данных суточного потребления газа и суточного потребления электроэнергии по двум компонентам представлено на рис. 4.

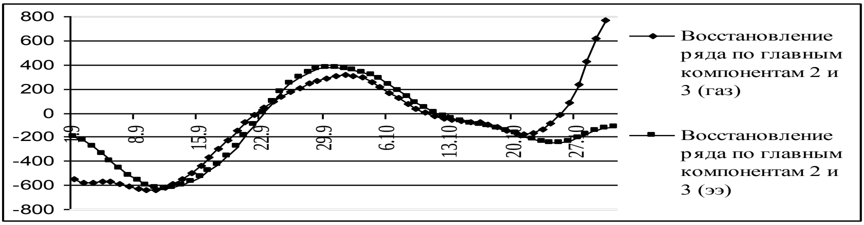

Рис. 4. Восстановление рядов потребления газа и электроэнергии по 2-ой и 3-й главным компонентам.

Анализируя полученные данные можно предположить, что эти компоненты в двух исследуемых рядах отражают "Состояние тепловой комфортности потребителя", и его изменение вызывает изменения в структуре временных рядов данных потребления энергоресурсов. На основе выделенных компонент может осуществляться обнаружение указанных моментов изменения структуры энергопотребления.

Выводы. На основании вышеизложенного можно сделать вывод о том, что предварительный анализ структуры временного ряда методом главных компонент (PCA) и сингулярно-спектрального анализа (SSA) при решении задач о разладке позволит выделить наиболее существенные компоненты ряда, ответственные за изменение его свойств, и тем самым повысить точность обнаружения моментов изменения свойств ряда и диагностировать возможные причины отклонений в процессе.

Список литературы: 1. Никифоров И.В. Последовательное обнаружение изменения свойств временных рядов. М.: Наука, 1983.- 199с. 2. Ширяев А.Н. Статистический последовательный анализ. М.: Наука, 1969, 232 с. 3. A. Singhal and D.E. Seborg, "Pattern matching in historical batch data using PCA", IEEE Control Systems Magazine, 2002, pp.53-63. 4. C. Ündey and Ali Çinar, "Statistical monitoring of multistage, multiphase batch processes", IEEE Control Systems Magazine, 2002, pp.40-52. 5. E. Martin, J. Morris, and S. Lane, "Monitoring process manufacturing performance", IEEE Control Systems Magazine, 2002, pp.26-39. 6. T. Kourti, "Process analysis and abnormal situation detection: from theory to practice", IEEE Control Systems Magazine, 2002, pp.10-25. 7. *Браверман Э.М., Мучник И.Б.* Структурные методы обработки эмпирических данных. - М.: Наука, 1983. - 464с. 8**.** *Горькова К.А., Абрамов Ю.Ш.* Факторный анализ (Метод главных компонент). Учебное пособие. - Л.: ЛФЭИ, 1981. - 66с. 9. *Жуковская В.М., Мучник И.Б.* Факторный анализ в социально-экономических исследованиях. - М.: Статистика, 1976. - 151с. 10. *Харман Г.* Современный факторный анализ/ Пер. с англ. - М: Статистика, 1972. - 488с. 11. Главные компоненты временных рядов: метод «Гусеница» // Под. ред. *Д.Л. Данилова, А.А. Жиглявского.* СПб: Пресском, 1997. 307 с. 12. *Голяндина Н.Э.* Метод «Гусеница»-SSA: анализ временных рядов: Учеб. пособие. СПб: Изд-во СПбГУ, 2004. 76 с. 13. *Вороновский Г.К.* Усовершенствование практики оперативного управления крупными теплофикационными системами в новых экономических условиях. - Х: Изд-во "Харьков", 2002. - 240с. 14**.** http://www.gistatgroup.com

Поступила в редколлегию 15.10.05