

НЕЧЕТКАЯ СЕГМЕНТАЦИЯ ВРЕМЕННЫХ РЯДОВ

П. Н. Зайцев

Воронежский государственный университет

Поступила в редакцию 1.03.2009 г.

Аннотация. Предложены модифицированные алгоритмы для измерения однородности сегментов временного ряда и универсальный алгоритм нечеткой сегментации временного ряда. Показано, что предложенный алгоритм более четко обозначает характерные временные периоды, а «всплески» являются более прогнозируемыми.

Ключевые слова: алгоритм нечеткой сегментации, временной ряд.

Abstract. Modified algorithms of time series homogeneity measurement and universal algorithm of fuzzy clustering based segmentation of time series have been offered. The algorithm shows time periods with more accuracy and have more predictable «bursts».

Keywords: algorithm of fuzzy clustering, time series.

ВВЕДЕНИЕ

Изменения переменных во временном ряде обычно неопределенны и не происходят в один момент времени. Однако непрактично выделять четкие границы сегментов. Например, если люди анализируют визуальную информацию о процессе, они используют такие слова, как «эта точка принадлежит к этой операционной точке, а также к какой-то еще». Хороший пример нечеткой группировки такого вида это определение раннего утра — неизвестно, во сколько оно начинается и когда заканчивается. Нечеткая логика широко используется в различных приложениях, где необходима кластеризация нечетких и неопределенных элементов, также есть много удачных примеров совместного применения нечеткой логики и анализа временных рядов.

Пусть последовательность N наблюдаемых данных $x_1 \dots x_N$ распределенных во времени называется временным рядом. В жизни временные ряды применяются в бизнесе, физике, социальных науках, экономике, инженерии, и т. д. Понятие сегментации временных рядов заключается в следующей статистической проблеме: пусть дан временной ряд T , необходимо найти разбиение этого ряда на s сегментов, которые являются внутренне однородными [1]. В зависимости от применения и цели разбиения необходимо найти стабильные периоды времени, найти точки изменения, или просто сжать

исходный временной ряд в более компактное представление [2]. Хотя во многих реальных приложениях наблюдение за многими переменными производится похожим образом, большинство алгоритмов сегментации используются только для анализа одной временной переменной [3].

Основная цель этой статьи — на основе изучения T^2 и Q -методов [4] разработать модифицированные T^2, Q -алгоритмы и смешанный $T-Q$ -алгоритм, способный не только получить изменяющиеся во времени характеристики многомерных данных (изменение среднего, изменение дисперсии и изменения в корреляционной структуре между переменными), но и учитывающий проблемы, которые уменьшили эффективность этих методов. Основная идея данной статьи — использование T^2, Q и $T-Q$ -методов для измерения однородности сегментов.

Для получения необходимой информации используются такие многомерные статистические инструменты, как метод главных компонент (Principal component analysis - PCA) [5]. Метод главных компонент позволяет отображать данные в пространстве с более низкой размерностью, которое полезно при анализе и визуализации коррелированной многомерной информации [6]. Линейные PCA модели имеют две особенности: они высоко детализированы и они простые в осуществлении. Линейный метод главных компонент может хорошо предсказывать результат для простых временных рядов, но плохо анализирует накопленную информа-

цию об изменениях или когда зависимость между переменными не является линейной. Анализ таких данных требует поиска локальных коррелированных групп [7]. Предлагаемый алгоритм осуществляет группировку данных для того, чтобы обнаружить локальные отношения между похожими переменными. В исследовании использован пример, в котором используются данные производства высокоплотного полиэтилена [8].

1. ОСНОВНАЯ ТЕРМИНОЛОГИЯ

Введем некоторые понятия. Положим, что временной ряд $T = \{x_k | 1 \leq k \leq N\}$ - это конечное множество N выборок, отмеченных моментами времени $t_1 \dots t_N$, $x_k = [x_{1k}, x_{2k}, \dots, x_{nk}]^T$ - вектор наблюдаемых параметров. Пусть сегмент ряда T — это множество последовательных временных точек $S(a, b) = \{x_k | a \leq k \leq b\}$, т.е. x_a, x_{a+1}, \dots, x_b . c -сегментация временного ряда T — это разбиение T на c непересекающихся сегментов $S_T^c = \{S_i(a_i, b_i) | 1 \leq i \leq c\}$, таких, что $a_1 = 1, b_c = N$ и $a_i = b_{i-1} + 1$. Другими словами, c -сегментация разделяет T на c непересекающихся сегментов $s_1 < s_2 < \dots < s_c$.

Обычно целью является нахождение однородных сегментов в данном временном ряде, т.е. сегменты, содержащие элементы со сходными свойствами. В этом случае проблема сегментации может быть описана как проблема группировки с ограничениями: данные со схожими свойствами должны быть определены в одну группу, но с ограничением, что все объекты в этой группе должны быть последовательными по времени. Для того чтобы формализовать эту цель, введем функцию стоимости $\cos t(S(a, b))$, в качестве некоторой меры однородности сегмента. Таким образом, сегмент является однородным, если функция стоимости одинакова для любой точки этого сегмента. На практике сегменты можно формировать так: сегмент является однородным, если функция стоимости для любой точки этого сегмента находится в допустимых пределах, которые определяются исходя из практических соображений. В качестве функции стоимости может выступать произвольная. Например, в [1] в качестве неё рассматривается сумма дисперсий. Обычно, функция стоимости $\cos t(S(a, b))$ определяется отклонением между значениями временного ряда и значениями некоторой функции, которая аппроксимирует его для каждого сегмента данных.

Решение проблемы сегментации опирается на кластеризацию сегментов, взаимосвязанных по времени. В качестве целевой функции выступает стоимость всей c -сегментации:

$$\cos t(S_T^c) = \sum_{i=1}^c \cos t(S_i) \quad (1)$$

Оптимальная c -сегментация должна одновременно определять границы сегментов a_i, b_i и вектор параметров θ_i для моделей сегментов.

Для нахождения отдельных сегментов, функция (1) может быть минимизирована с помощью динамического программирования. К сожалению, применение динамического программирования для многих задач весьма затруднено из-за труднообрабатываемых вычислений. Поэтому эвристические методы оптимизации, такие как восходящий и нисходящий методы [9], метод раздвижного окна [9], часто используются при нахождении хорошего, но условно оптимального c -разбиения.

2. ОПИСАНИЕ АЛГОРИТМОВ

Рассмотрим:

$$\begin{aligned} \cos t(S_T^c) &= \sum_{i=1}^c \sum_{k=s_{i-1}+1}^{s_i} \|x_k - v_i^x\|^2 = \\ &= \sum_{i=1}^c \sum_{k=1}^N \beta_i(t_k) D_{ik}^2(v_i^x, x_k) \end{aligned} \quad (2)$$

где $D_{ik}^2(v_i^x, x_k)$ — представляет расстояние между центром в i -м сегменте и точкой данных, $\beta_i(t_k) = \{0, 1\}$ — определяет принадлежность точки k к i -му сегменту:

$$\beta_i(t_k) = \begin{cases} 1, & \text{если } s_{i-1} < k \leq s_i; \\ 0, & \text{иначе.} \end{cases}$$

Изменение значений переменных временных рядов обычно случайное. Так как практически невозможно определить четкие границы сегментов, для представления $\beta_i(t_k) = \{0, 1\}$ нечетких сегментов временных рядов - $A_i(t_k)$ используются гауссовские функции, выбор которых приводит к следующей компактной формуле количества элементов в i -м сегменте k -го наблюдения:

$$\begin{aligned} A_i(t_k) &= \exp\left(-\frac{1}{2} \frac{(t_k - v_i^t)^2}{\sigma_i^2}\right), \\ \beta_i(t_k) &= \frac{A_i(t_k)}{\sum_{j=1}^c A_j(t_k)} \end{aligned} \quad (4)$$

Для нахождения центров v_i^t и дисперсии σ_i^2 функций представлен алгоритм нечеткой классификации, который является модифицированным алгоритмом вида *Gath-Geva* [10]. Этот алгоритм представляет данные как многомерное распределение Гаусса, таким образом, это минимизирует сумму квадратов между точками данных $z_k = [t_k, x_k^T]^T$ и η_i — прототипом формируемой группы:

$$\sum_{i=1}^c \sum_{k=1}^N (\mu_{ik})^m D_{ik}^2(\eta_i, z_k) = \sum_{i=1}^c \sum_{k=1}^N (\mu_{ik})^m D_{ik}^2(v_i^t, t_k) D_{ik}^2(v_i^x, x_k), \quad (5)$$

где μ_{ik} — степень принадлежности наблюдения $z_k = [t_k, x_k^T]^T$, $m \in \infty$ — экспонента, которая определяет нечеткость групп (обычно $m = 2$).

Особенность метода состоит в том, что дисперсия состоит из двух частей:

$$D_{ik}^2(\eta_i, z_k) = D_{ik}^2(v_i^t, t_k) D_{ik}^2(v_i^x, x_k).$$

Первая из них $D_{ik}^2(v_i^t, t_k)$ — это отклонение k -й точки от центра i -го сегмента v_i^t во времени.

$$1/D_{ik}^2(v_i^t, t_k) = \frac{1}{\sqrt{2\pi\sigma_i^2}} \exp\left(-\frac{1}{2} \frac{(t_k - v_i^t)^2}{\sigma_i^2}\right), \quad (6)$$

где центр стандартного отклонения функции Гаусса определяется следующим образом:

$$v_i^t = \frac{\sum_{k=1}^N (\mu_{ik})^m t_k}{\sum_{k=1}^N (\mu_{ik})^m}, \quad (7)$$

$$\sigma_i^2 = \frac{\sum_{k=1}^N (\mu_{ik})^m (t_k - v_i^t)^2}{\sum_{k=1}^N (\mu_{ik})^m}.$$

Вторая часть дисперсии $D_{ik}^2(v_i^x, x_k)$ определяется расстоянием между прототипом группы и пространством особенности:

$$1/D_{ik}^2(v_i^x, x_k) = \frac{\alpha_i \sqrt{\det(A_i)}}{(2\pi)^{r/2}} \times \exp\left(-\frac{1}{2} (x_k - v_i^x)^T (A_i) (x_k - v_i^x)\right), \quad (8)$$

где α_i — априорная вероятность группы и v_i^t — координаты центра группы в пространстве особенности:

$$\alpha_i = \frac{1}{N} \sum_{k=1}^N \mu_{ik},$$

$$v_i^x = \frac{\sum_{k=1}^N (\mu_{ik})^m x_k}{\sum_{k=1}^N (\mu_{ik})^m}, \quad (9)$$

r — ранг A_i нормы отклонения, соответствующая i -й группе. Норма отклонения может быть определена многими способами. Разумно применить к переменным масштабирование, так чтобы переменные с большим отклонением не преобладали в группе. Масштабирование производится делением на значение стандартных отклонений, однако лучший способ — использовать статистическое отклонение, которое также приспособливается к корреляции между переменными. В этом случае A_i — обратная к матрице нечеткой ковариации: $A_i = F_i^{-1}$, где

$$F_i = \frac{\sum_{k=1}^N (\mu_{ik})^m (x_k - v_i^t)(x_k - v_i^t)^T}{\sum_{k=1}^N (\mu_{ik})^m}. \quad (10)$$

Когда переменные в матрице сильно зависят друг от друга, матрица ковариации F_i может оказаться плохо обусловленной и не иметь обратной матрицы. В [4] эта проблема была решена. Первый из двух методов, которые использованы в [4], основан на пропорции между минимальным и максимальным значением собственного значения матрицы ковариации. Второй метод основан на добавлении единой матрицы весов к посчитанной матрицы ковариации. Оба метода позволяют получить обратную матрицу, но ни один из них не выдает потенциальную информацию о скрытой структуре данных.

Метод главных компонент (PCA) основан на проектировании коррелированной, многомерной информации на гиперплоскость, которая очень удобна для визуализации и анализа данных. Отображение использует только несколько первых r ненулевых собственных значений и соответствующих им собственных векторов $F_i = U_i \Lambda_i U_i^T$. Таким образом, матрица ковариации раскладывается на Λ_i - матрицу с собственными значениями, расположенными на главной диагонали в убывающем порядке, и на U_i — матрицу собственных векторов в столбцах, соответствующих собственным значениям, $y_{ik} = \Lambda_{ip}^{-\frac{1}{2}} U_{ip}^T x_k$.

Когда гиперплоскость PCA модели имеет достаточное число пространств, дальность информации от гиперплоскости является величиной незначительной, поэтому проекция данных на p -размерную гиперплоскость не является причиной существенных ошибок в преобразовании

$$Q_{ik} = (x_k - \hat{x}_k)^T (x_k - \hat{x}_k) = x_k^T (I - U_{ip} U_{ip}^T) x_k. \quad (11)$$

Хотя зависимость между переменными может быть описана линейной моделью достаточно точно, в некоторых случаях возможно распределить данные вокруг некоторых отдельных центров в линейном подпространстве. Для подсчета дальности данных от центра в линейном подпространстве используется следующая величина

$$T_{ik}^2 = y_{ik}^T y_{ik}. \quad (12)$$

T^2 и Q часто используются для слежения за многомерными системами и для исследования ошибок и причин, вызывающих их. На рис. 1 показаны эти параметры при наличии двух переменных и одного главного компонента.

Рассмотрим центральные идеи этих методов:

Q-метод: Измерение отдаленности основано на Q-реконструкции ошибки. Для измерения нормы отдаленности A_i строится матрица из самых маленьких $n-p$ собственных значений и собственных векторов матрицы F_i и находится обратная матрица.

$$A_i^{-1} = U_{i,n-p} \Lambda_{i,n-p} U_{i,n-p}^T. \quad (13)$$

T^2 метод: измерение отдаленности основано на T^2 модели. A_i определяется p наибольшими собственными значениями и собственными векторами матрицы F_i

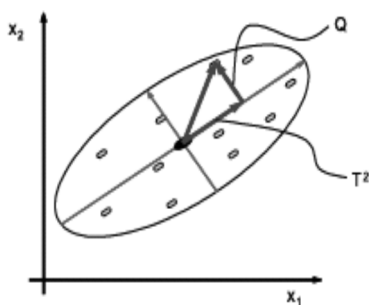


Рис. 1. Измерение отклонения PCA моделью

$$A_i^{-1} = U_{ip} \Lambda_{ip} U_{ip}^T. \quad (14)$$

Когда p равно рангу матрицы F_i , данный метод идентичен методу измерения отдаленности, основанному на матрице нечеткой ковариации, где методом Мура-Пенроуза находится обратная F_i матрица.

Оптимальные параметры прототипа группы $\eta_i = \{v_i^x, A_i, v_i^t, \sigma_i^2, \alpha_i\}$ определяются минимизацией функции (5) при следующих условиях, вводимых для того, чтобы пространство разбиения было нечетким,

$$U = [\mu_{ik}]_{c \times N} \mid \mu_{ik} \in [0, 1], \forall i, k; \quad (15)$$

$$\sum_{i=1}^c \mu_{ik} = 1, \forall k;$$

$$0 < \sum_{k=1}^N \mu_{ik} < N, \forall i.$$

Самым популярным методом решения этой проблемы является альтернативная, выборочная оптимизация, которая состоит из применения пикарового [11] повторения до выполнения условий первого приближения для стационарной точки (5), которая может быть найдена наложением условия (15) к J использованием множителей Лагранжа [11]

$$\bar{J}(Z, U, \eta, \lambda) = \sum_{i=1}^c \sum_{k=1}^N (\mu_{ik})^m D_{ik}^2(z_k, \eta_i) + \sum_{k=1}^N \lambda_i \left(\sum_{i=1}^c \mu_{ik} - 1 \right) \quad (16)$$

и приравниваем градиентов \bar{J} , соответствующих Z, U, η, λ , к нулю.

В итоге можно описать алгоритм следующим образом:

Инициализация. Дан временной ряд T , со свойствами c , выбирается допустимое отклонение (допуск прерывания) $\varepsilon > 0$, и создается матрица разбиения $U = [\mu_{i,k-1}]$ со случайными числами.

Повторение для $l = 1, 2, \dots$

Шаг 1. Вычисляется η_i параметров по формулам (7), (9) и (13) или (14).

Шаг 2. Вычисляется $D_{ik}^2(\eta_i, z_k)$ и находится отклонение (5) при помощи (6) и (8)

Шаг 3. Обновление матрицы разбиения

$$\mu_{ik}^l = \frac{1}{\sum_{j=1}^c (D_{ik}/D_{jk})^{2/(m-1)}}, \quad (17)$$

$$1 \leq i \leq c, \quad 1 \leq k \leq N.$$

До тех пор, пока $\|U(t) - U(t-1)\| < \varepsilon$.

3. ВЫЧИСЛИТЕЛЬНЫЙ ЭКСПЕРИМЕНТ, ВЫВОДЫ

Цель вычислительного эксперимента состоит в том, чтобы показать, как предложенный алгоритм способен обнаружить значащие формы графика в многомерных накопленных данных процесса. В качестве примера взято производство полиэтилена со средней и высокой плотностями (MDPE, HDPE) [7]. Завод принадлежит крупнейшей компании Венгрии, производящей полимерные продукты. Проблема производства состоит в том, что необходимо выпускать около десяти видов продукции, согласно требованиям рынка. Следовательно, существует проблема минимизации времени переключения, потому что в течение переключения будет производиться продукция ненадлежащего качества. Сложность проблемы исходит оттого, что более десяти переменных процесса должны учитываться. Измерение значений переменных происходит каждые 15 секунд. Выделены следующие переменные $x_k: x_{k,1}$ — интенсивность выпуска полимерного продукта (PE), скорости потока входного отверстия гексена (начальная скорость потока гексена) — $C6_{in}$, этилена — $C2_{in}$, водорогена — $H2_{in}$, растворитель изобутана — IB_{in} , катализатор — Kat , концентрации этилена — C_2 , гексена — C_2 и водорогена — H_2 , жидкий раствор в реакторе — $slurry$ и температура в реакторе — T .

Рассмотрим интервал времени в 125 часов, который включает в себя, по крайней мере, 3 отрезка времени (сегмента): перемещение (изменение) продукта около 25-го часа, спокойный (неизменный) период до 100-го часа и 30-и часовой активный период (см. рис. 2).

Расстояние между точками данных и центрами групп может быть посчитано двумя способами, соответствующие двум разным методам измерения расстояния принципиального анализа компонентов: Q -реконструкция ошибок и T^2 -распределение. Эти методы дают разные результаты, которые могут использоваться для различных целей.

Нечеткий метод PCA-Q чувствителен к числу главных компонентов. С увеличением p ошибка реконструкции уменьшается. В случае $p = n = 11$ реконструированная ошибка становится равной нулю во всем диапазоне исходных данных. Если $p \rightarrow 0$, то ошибка реконструкции будет большая для всего временного ряда. В этих двух крайних случаях сегментация не основана на внутренних отношениях между переменными и появляются равноотстоящие сегменты. Когда число скрытых переменных находится в диапазоне 3,...8, получаются весьма четкие (надежные) сегменты, алгоритм находит изменение класса производства на 25-м часе (рис. 3).

Нечеткий метод PCA-T² также чувствителен к числу принципиальных компонент. Увеличение пространства модели напрямую не уменьшает эффективность метода, потому что

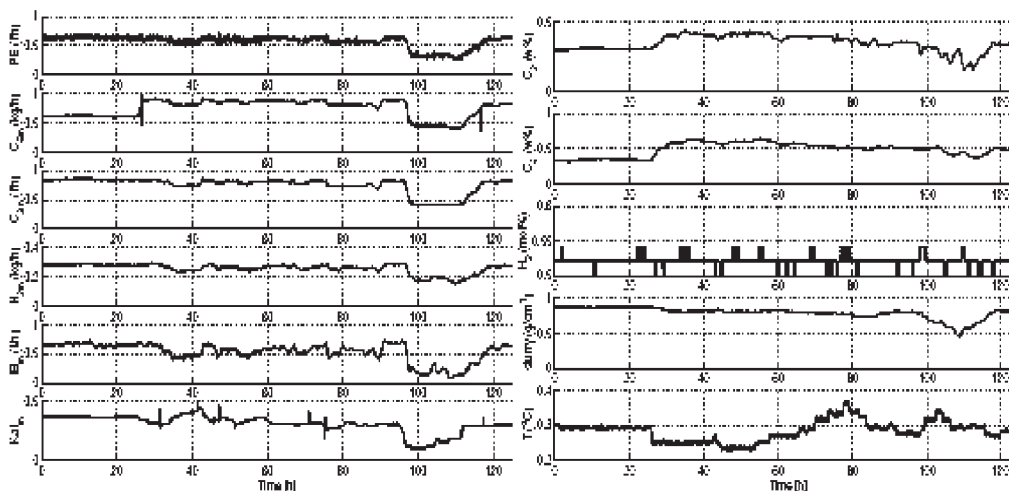


Рис. 2. Пример временного ряда изменения вида продукции

размеры матрицы нечеткой ковариации стремятся к нулю. Следовательно, практичным является использование обеих PCA моделей одновременно.

Алгоритм основан на одновременном определении нечетких множеств, которые представляют собой сегменты во времени и гиперплоскости локальных PCA моделей, используемых для измерения однородности сегментов. Были определены два типичных приложения PCA моделей. Сегменты Q -реконструкции ошибки согласуются с изменениями корреляции между

переменными, в то время как T^2 -распределение сегментирует временной ряд методом, основанном на дрейфе центра исследуемого региона. Описанный выше алгоритм был применен к мониторингу за производством полиэтилена высокой плотности. Когда для группировки использовалась Q -реконструкция ошибки, высокая производительность была достигнута, когда несколько главных компонент были определены как параметры групп. Когда производилась группировка, основанная на T^2 расстояниях от центра группы, было необходимо оп-

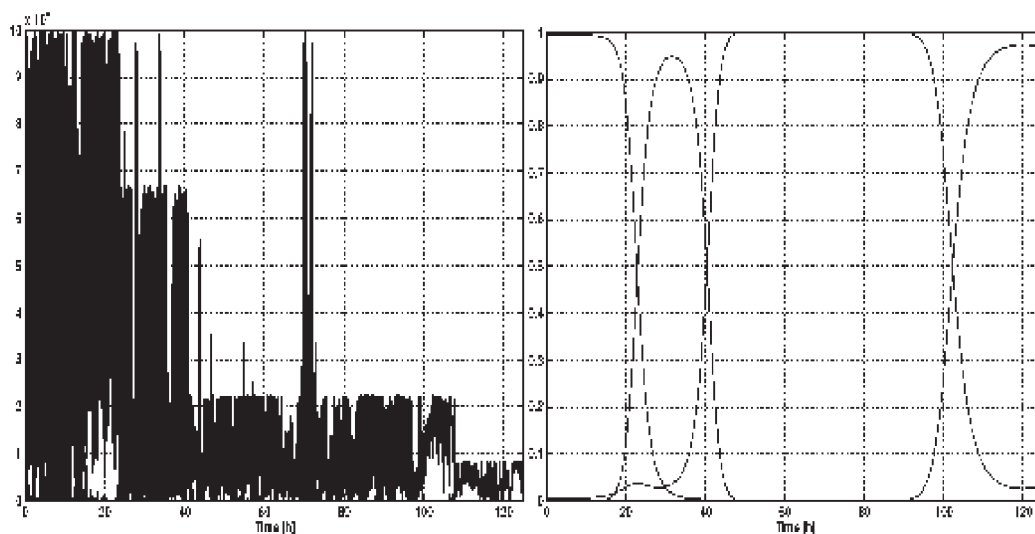


Рис. 3. Нечеткая PCA-Q сегментация: (а) $1/D_{ik}(v_i^x, x_k)$ отклонения, (б) нечеткие сегменты временного ряда, $\beta_i(t_k)$

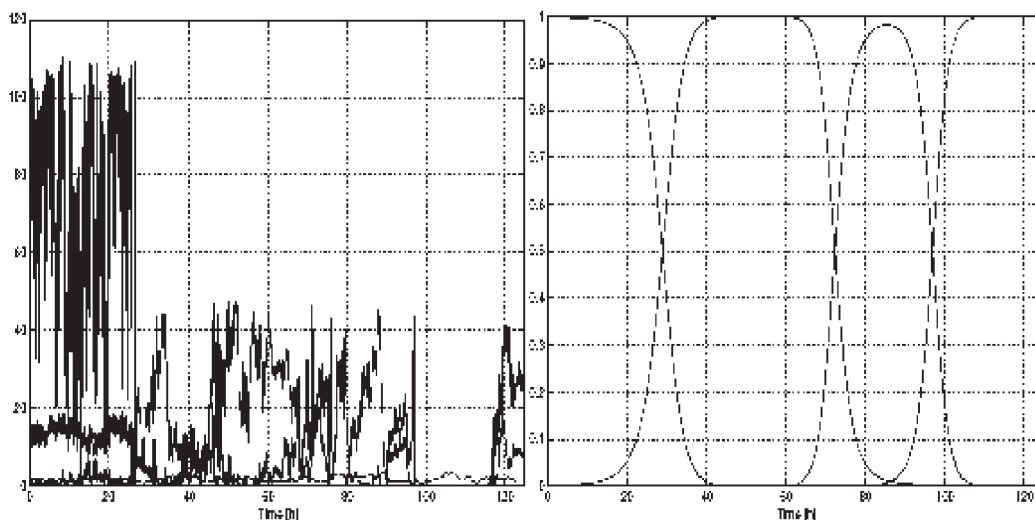


Рис. 4. Нечеткая PCA- T^2 сегментация: (а) $1/D_{ik}(v_i^x, x_k)$ отклонения, (б) нечеткие сегменты временного ряда, $\beta_i(t_k)$

ределить желательное выполнение разбиения для получения хороших результатов сегментации. Результаты предполагают, что предложенный алгоритм может быть применен для выделения типичных эксплуатационных состояний и анализа изменения класса продукта.

$T-Q$ -метод — это алгоритм, который использует оба метода одновременно.

$T-Q$ -алгоритм основан на структурном «взаимодействии» обоих алгоритмов, которое базируется на нескольких принципах:

1. Определение p , основанное на желаемой точности PCA моделей:

$$\sum_{j=1}^{p-1} \text{diag}(\Lambda_{ij}) / \sum_{j=1}^n \text{diag}(\Lambda_{ij}) < q \leq \sum_{j=1}^p \text{diag}(\Lambda_{ij}) / \sum_{j=1}^n \text{diag}(\Lambda_{ij}). \quad (18)$$

Как показывает рисунок 4, когда $q = 0.99$, алгоритм также находит правильную сегментацию.

2. Основываясь на этом, вычисляются «диапазоны эффективности» методов. С помощью этих диапазонов можно учесть слабые стороны обоих методов и свести их к минимуму, тем самым, повысив общую эффективность алгоритма.

В результате численного эксперимента установлено, что нечеткие модифицированные методы PCA- T^2 и PCA- Q чувствительны к числу принципиальных компонент и выбору функции принадлежности. В программной реализации рассматривались функции четырех основных видов:

$$\mu(x) = \frac{1}{1 + e^{-a_1(x-c_1)}} - \frac{1}{1 + e^{-a_2(x-c_2)}} \quad (1)$$

$$\mu(x) = \frac{1}{1 + \left| \frac{x-c}{a} \right|^{2b}} \quad (2)$$

$$\mu(x) = e^{-\frac{(x-b)^2}{2c^2}} \quad (3)$$

$$\mu(x) = \frac{1}{1 + e^{-a(x-c)}} \quad (4)$$

В процессе эксперимента выяснено, что в случае использования T^2 и Q -метода лучше всего использовать функции (4) и (3). За счет правильного выбора функций принадлежности так же повышается эффективность и скорость алгоритма.

Главными недостатками алгоритмов являются: для Q -алгоритма - наличие «критических» точек, примером которых может служить всплеск в районе 70-го часа (см. рис.3), для T^2 -алгоритма — большое время работы алгоритма. Общими недостатками можно назвать чувствительность к числу главных компонент и выбору функций принадлежности. $T-Q$ -алгоритм работает гораздо быстрее, менее чувствителен к числу главных компонент, а так же более четко обозначает характерные временные периоды (рис. 5), что важно для мониторинга производства.

Всплеск в районе 70-го часа является гораздо более показательным и информативным, то есть четко указывающим на изменение в структуре активности производства. В отличие от рис.

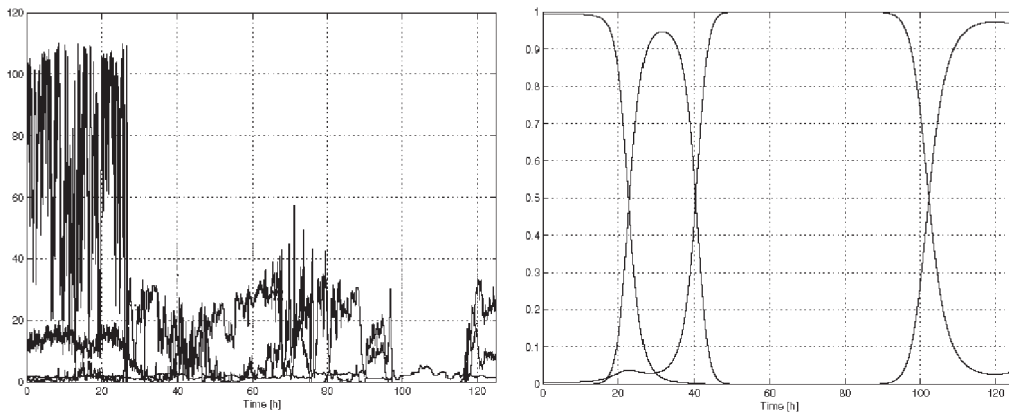


Рис. 5. Нечеткая PCA- $T-Q$ сегментация: (а) $1/D_{ik}(v_i^x, x_k)$ отклонения, (б) нечеткие сегменты временного ряда, $\beta_i(t_k)$

З и нечеткой РСА-Q сегментации, всплеск является прогнозируемым: если на рис. 3 он резкий и «неожиданный», то на рис. 5 ему предшествует планомерное возрастание показателя. Эта особенность метода может быть использована для прогнозирования подобных состояний производства, и, следовательно, может сделать алгоритм основной для принятия своевременных решений.

СПИСОК ЛИТЕРАТУРЫ

1. *Vasko K., Toivonen H.* Estimating the number of segments in time series data using permutation tests, IEEE International Conference on Data Mining (2002) 466—473.
2. *Last M., Klein Y., Kandel A.* Knowledge discovery in time series databases, IEEE Transactions on Systems, Man, and Cybernetics 31 (1) (2000) 160—169.
3. *Kivikunnas S.* Overview of process trend analysis methods and applications, ERUDIT Workshop on Applications in Pulp and Paper Industry (1998).
4. *Babuska R., van der Veen P. J., Kaymak U.* Improved covariance estimation for gustafson-kessel clustering, IEEE International Conference on Fuzzy Systems (2002) 1081—1085.
5. *Tipping M.E., Bishop C.M.* Mixtures of probabilistic principal components analysis, Neural Computation 11 (1999) 443—482.
6. *Stephanopoulos G., Han C.* Intelligent systems in process engineering: A review, Comput. Chem. Engng. 20 (1996) 743—791.
7. *Chakrabarti K., Mehrotra S.* Local dimensionality reduction: A new approach to indexing high dimensional spaces, Proceedings of the 26th VLDB Conference Cairo Egypt (2000).
8. *Abonyi J.* Fuzzy Model Identification for Control, Birkhauser Boston, 2003.
9. *Keogh E., Chu S., Hart D., Pazzani M.* An online algorithm for segmenting time series, IEEE International Conference on Data Mining, 2001.
10. *Abonyi J., Babuska R., Szeifert F.* Modified Gath-Geva fuzzy clustering for identification of takagi-sugeno fuzzy models, IEEE Transactions on Systems, Man, and Cybernetics 32(5) (2002) 612—321.
11. *Hoppner F., Klawonn F., Kruse R., Runkler T.* Fuzzy Cluster Analysis, Wiley, Chichester, 1999.

Зайцев Павел Николаевич — аспирант, Воронежский государственный университет, т. 8-903-850-06-03, e-mail: dolphin_land@mail.ru

Zaitsev P.N. — Postgraduate student, Voronezh State University, T. 8-903-850-06-03, e-mail: dolphin_land@mail.ru.