

Разработанная нейросетевая модель, алгоритм и методика ее создания и использования внедрены в практику производства промышленной продукции в условиях ПАО «АВЕРС» как эффективный инструмент поддержки принятия организационно-технических решений при управлении качеством продукции.

Таблица 6 – Степень влияния входов модели на показатель Y^{i+3}

Вход	Y_{avr}^i	X_1^{i-1}	X_4^i	Y^{i-3}	X_2^i	X_6^{i-1}	X_5^i	X_3^{i-1}	Y^i
Ранг	1	2	3	4	5	6	7	8	9
Влияние	0,255	0,246	0,188	0,159	0,151	0,127	0,073	0,056	0,037

Выводы. Установлена зависимость уровня качества от значений технико-экономических показателей, характеризующихся сложными изменяющимися во времени взаимосвязями и разработана адаптивная нейросетевая модель, позволяющая с достоверностью 98% прогнозировать изменение уровня качества на установленный период упреждения прогноза.

Модель может быть использована как эффективный инструмент поддержки принятия организационно-технических решений при выборе рациональных значений технико-экономических показателей, обеспечивающих высокий уровень качества промышленной продукции различного целевого назначения.

Список литературы: 1. Бобровников, Г. Н. Прогнозирование в управлении техническим уровнем и качеством продукции [текст] / Г. Н. Бобровников, А. И. Клебанов. – М.: Изд-во стандартов, 1984. – 232 с. 2. Федин, С. С. Оценка и прогнозирование качества промышленной продукции с использованием адаптивных систем искусственного интеллекта: [монография] / С. С. Федин, Н. А. Зубрецкая. – К.: Интерсервис, 2012. – 206 с. 3. Базы данных. Интеллектуальная обработка информации [текст] / В. В. Корнеев, А. Ф. Гареев, С. В. Васютин, В. В. Райх. – М.: Нолидж, 2000. – 352 с.

Поступила в редколлегию 01.06.2013

УДК 004.8:658.5

Нейросетевое прогнозирование уровня качества промышленной продукции/ Н.А. Зубрецкая // Вісник НТУ «ХПІ». Серія: Нові рішення в сучасних технологіях. – Х: НТУ «ХПІ», – 2013. - № 38 (1011). – С.81-86. – Бібліогр.:5 назв.

Розроблено адаптивну нейромережну модель, що дає можливість отримувати прогноз рівня якості промислової продукції в залежності від впливу сукупності техніко-економічних показників виробництва.

Ключові слова: нейромережна модель, прогнозування, техніко-економічні показники, рівень якості.

The adaptive neural network model for quality forecasting of industrial products depending on retarded impact of technical and economic manufacturing indexes was developed.

Keywords: neural network model, forecasting, technical and economic indexes, quality level.

УДК 004.89

Д. А. МАШОШИН, студент, ХНУРЭ, Харьков

АДАПТИВНАЯ ОБРАБОТКА НЕСТАЦИОНАРНЫХ ВРЕМЕННЫХ РЯДОВ НА ОСНОВЕ НЕЧЕТКОГО ПОДХОДА

На основе объединения нечеткого пакетного способа обработки и сегментации временных рядов с рекуррентными процедурами обработки текущих значений, предложен онлайн метод сегментации многомерных временных рядов, который применим для обнаружения однородных сегментов в реальном режиме времени на основе потоковых данных.

Введение. Временные ряды являются совокупностью наблюдений случайного процесса, а задача сегментации достаточно часто возникает во многих приложениях, связанных с их обработкой. Для обработки таких данных некорректно выделять четкие границы сегментов. В таких задачах используются алгоритмы нечеткой сегментации. Нечеткие алгоритмы сегментации широко применяются в различных приложениях, где необходима кластеризация нечетких и неопределенных элементов, перекрывающихся групп и расплывчатых объектов, но они не могут быть непосредственно применены к временным рядам, так как кластеры могут быть смежными во времени.

В сфере обработки временных рядов также существуют задачи, для которых необходима обработка значений в режиме реального времени с учетом постоянно поступающих новых входных данных. Основными такими задачами является мониторинг работы системы в процессе переходов для сокращения бракованной продукции, не соответствующей спецификации, определение нарушений и раннего предупреждения отклонений в технологическом процессе или роста неисправностей, а также мгновенное выявление образцов для анализа, прогноза и оптимизации. С учетом непрерывно обновляющихся входных данных, обычные алгоритмы пакетной обработки не подходят для такого вида задач ввиду их вычислительной сложности при пересчете новых значений. Для того, чтобы избавиться от необходимости заново рассчитывать обратную матрицу при подключении/удалении в модель новых членов, необходимо корректировать ковариационную матрицу, полученную на предыдущем шаге.

Анализ литературных данных и постановка проблемы. Понятие сегментации временных рядов заключается в следующей статистической проблеме: пусть дан временной ряд $T = \{x_k \mid 1 \leq k \leq N\}$, представленный конечным множеством N выборок, отмеченных моментами времени t_1, \dots, t_N , необходимо найти разбиение этого ряда на множество последовательных временных точек $S(a, b) = \{a \leq k \leq b\}, x_a, x_{a+1}, \dots, x_b$, называемых границами сегментов. c -сегментация временного ряда T – это разбиение сегмента T на c непересекающихся сегментов $S_T^c = S_i(a_i, b_i) \mid 1 \leq i \leq c$, которые являются внутренне однородными, таким образом, чтобы $a_1 = 1, b_c = N$ и $a_i = b_{i-1} + 1$. Т.е. необходимо найти стабильный период времени, для определения изменения сегментов. 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. Хотя во многих реальных приложениях наблюдение за многими переменными производится похожим образом, большинство алгоритмов сегментации используются только для анализа одной нестационарной переменной.

Целью процедуры сегментации 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 формализации этого, необходимо ввести функцию качества $cost(S(a, b))$, как меру однородности сегмента. Таким образом,

сегмент является однородным, если эта функция одинакова или находится в допустимых пределах для любой точки этого сегмента, при этом функция может выбираться достаточно произвольно.

Чаще всего, функция качества $cost(S(a,b))$ определяется отклонением между значениями временного ряда и значениями некоторой функции (константной, линейной или полиномиальной), которая аппроксимирует его для каждого сегмента данных. Оптимальная c -сегментация должна одновременно определять границы сегментов и вектор параметров для моделей сегментов. Решение проблемы сегментации опирается на кластеризацию сегментов, взаимосвязанных по времени. В качестве целевой функции обычно выступает стоимость всей c -сегментации:

$$cost(S_T^c) = \sum_{i=1}^c cost(S_i) . \quad (1)$$

Для нахождения отдельных сегментов, функция качества может быть минимизирована с помощью динамического программирования, но, к сожалению, его применение весьма затруднено для многих реальных задач из-за громоздких вычислений.

Цель и задачи исследования. Основная цель этой работы – на основе изучения методов нечеткой сегментации временных рядов разработать модифицированные алгоритмы, способные не только получать изменяющиеся во времени потоковые характеристики многомерных данных (изменение среднего, изменение дисперсии и изменения в корреляционной структуре между переменными), но и учитывающие проблемы, которые уменьшают эффективность этих методов. Для получения необходимой информации используются такие многомерные статистические инструменты, как метод главных компонент (Principal component analysis – PCA). Метод главных компонент позволяет отображать данные в пространство с более низкой размерностью, которое полезно при анализе и визуализации коррелированной многомерной информации. Предлагаемый алгоритм осуществляет группировку данных для того, чтобы обнаружить локальные отношения между похожими переменными.

Описание алгоритмов. Рассмотрим функцию качества [1]:

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

где $D_{ik}^2(v_i^x, x_k)$ – расстояние между центром в i -м сегменте и точкой данных,

$b_i(t_k) = \{0,1\}$ – определяет принадлежность точки k к i -му сегменту:

$$\beta_i(t_k) = \begin{cases} 1, & s_{i-1} < k \leq s_i, \\ 0, & \text{иначе.} \end{cases} \quad (3)$$

Изменение значений переменных временных рядов обычно случайно, 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:

$$A_i(t_k) = \exp\left(-\frac{1}{2} \frac{(x_k - v_i^x)^2}{\sigma_i^2}\right), \quad (4)$$

$$\beta_i(t_k) = \frac{A_i(t_k)}{\sum_{j=1}^c A_j(t_k)}. \quad (5)$$

Для нахождения центров кластеров v_i^t и дисперсии s_i^2 введем онлайн алгоритм нечеткой кластеризации, который является модифицированным алгоритмом вида Gath-Geva [1]. Этот алгоритм представляет данные на основе многомерного распределения Гаусса, таким образом, это минимизируется сумма квадратов между точками данных $z_k = [t_k, x_k^T]^T$ и η_i – прототипом формируемой группы:

$$J = \sum_{i=1}^c \sum_{k=1}^N (\mu_{i,k})^m D^2(z_k, \eta_i), \quad (6)$$

где $\mu_{i,k}$ – степень принадлежности наблюдения $z_k = [t_k, x_k^T]^T$,

$m \in [1, \infty)$ – показатель, который определяет нечеткость групп (обычно $m = 2$).

Алгоритм кластеризации GG может быть рассмотрен, как основывающийся на вероятностных принципах. В таком случае функция расстояния $D^2(z_k, h_i)$ обратно пропорциональна вероятности $p(z_k | h_i)$ того, что точка z_k принадлежит i -тому кластеру. Данные описываются нормальным распределением случайных величин со значением центров кластера v_i и ковариационной матрицей \mathbf{F}_i . Алгоритм GG представляет собой смесь гауссовских распределений:

$$p(z_k | \eta) = \sum_{i=1}^c p(z_k | \eta_i) p(\eta_i), \quad (7)$$

где $p(z | \eta_i)$ – распределение, порожденное i -тым кластером, представленное гауссовской функцией:

$$p(z | \eta_i) = \frac{1}{(2\pi)^{\frac{n+1}{2}} \sqrt{\det(\mathbf{F}_i)}} \exp\left(-\frac{1}{2}(z - v_i)^T (\mathbf{F}_i)^{-1} (z - v_i)\right) \quad (8)$$

и $p(\eta_i)$ – безусловная кластерная вероятность, η_i представляет собой параметры i -того кластера, $\eta_i = \{p(\eta_i), v_i, \mathbf{F}_i | i = 1, \dots, c\}$.

Так как время не зависит от x_k переменных, предложенный алгоритм кластеризации основывается на следующей функции расстояния $D^2(z_k, \eta_i)$:

$$p(z_k | \eta_i) = \frac{1}{D^2(z_k, \eta_i)} = \frac{\alpha_i}{p(\eta_i)} \underbrace{\frac{1}{\sqrt{2\pi\sigma_{i,t}^2}} \exp\left(-\frac{1}{2} \frac{(t_k - v_i^t)^2}{\sigma_{i,t}^2}\right)}_{p(t_k | \eta_i)} \times \underbrace{\frac{1}{(2\pi)^{\frac{r}{2}} \sqrt{\det(\mathbf{A}_i)}} \exp\left(-\frac{1}{2}(x_k - v_i^x)^T (\mathbf{A}_i)^{-1} (x_k - v_i^x)\right)}_{p(x_k | \eta_i)}, \quad (9)$$

которая состоит из трех частей. Первая, α_i – априорная вероятность кластера, вторая – отклонение k -й точки от центра i -го сегмента v_i^t во времени. Третья представляет собой расстояние между прототипом кластера и пространства признаков, координаты центра группы в пространстве признаков.

Норма отклонения \mathbf{A}_i может быть определена многими способами. Разумно применить к переменным масштабирование так, чтобы переменные с большим отклонением не преобладали в группе. Масштабирование производится путем

делением на значение стандартных отклонений, однако лучший способ – использовать статистическое отклонение, которое также приспособляется к корреляции между переменными. В этом случае \mathbf{A}_i – обратная к матрице нечеткой ковариации $\mathbf{A}_i = \mathbf{F}_i$, где

$$\mathbf{F}_i = \frac{\sum_{k=1}^N (\mu_{i,k})^m (x_k - v_i^x)(x_k - v_i^x)^T}{\sum_{k=1}^N (\mu_{i,k})^m}. \quad (10)$$

Когда переменные в матрице сильно зависят друг от друга, матрица ковариации \mathbf{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 матрице ковариации. Оба метода позволяют получить обратную матрицу, но ни один из них не выдает потенциальную информацию о скрытой структуре данных.

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:

$$\begin{aligned} U &= [\mu_{i,k}]_{c \times N} \mid \mu_{i,k} \in [0,1], \forall i, k, \\ \sum_{i=0}^c \mu_{i,k} &= 1, \\ \forall k; 0 < \sum_{k=1}^N \mu_{i,k} < N, \forall i. \end{aligned} \quad (11)$$

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

$$\bar{\mathbf{J}}(Z, U, \eta, \lambda) = \sum_{i=1}^c \sum_{k=1}^N (\mu_{i,k})^m D^2(z_k, \eta_i) + \sum_{k=1}^N \lambda_k (\sum_{i=1}^c \mu_{i,k} - 1) \quad (12)$$

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

Явным недостатком данного решения является его сложность в использовании онлайн-овых процедур кластеризации. Для решения данной проблемы используются рекуррентные процедуры обращения матриц и нахождения определителя матрицы, что является наиболее ресурсоемкой задачей.

Метод рекуррентного обращения позволяет со сравнительно малыми вычислительными затратами пересчитывать матрицу, обратную к A при добавлении к случайной сетке дополнительных узлов. Благодаря этому, можно постепенно наращивать объем случайной сетки и останавливать вычислительный процесс сразу же по достижении требуемой точности.

Лемма Шермана-Моррисона описывает обращение $A + uv^T$, при известном разложении на множители матрицы A :

$$(A + uv^T)^{-1} = A^{-1} - \frac{A^{-1}uv^T A^{-1}}{1 + v^T A^{-1}u}. \quad (13)$$

На основе этой формулы, а также леммы Вудбери можно рассчитать формулу рекуррентного вычисления обратной матрицы и определителя:

$$A^{-1}(k+1) = A^{-1}(k) - \frac{A^{-1}(k)x(k+1)x^T(k+1)A^{-1}(k)}{1+x(k+1)A(k)x^T(k+1)}. \quad (14)$$

В итоге алгоритм можно описать следующим образом:

Инициализация. Дан временной ряд T со свойствами c , выбирается допустимое отклонение (допуск прерывания) $e > 0$ и инициализируются $W_i, v_i^x, \sigma_{i,x}^2, \mu_{i,k}$.

Повторение до тех пор, пока не дана точка выхода или не превышено время простоя:

Сбор поступившей информации и формирование новых входных параметров.

Если обратная матрица известна:

Вычисление новой обратной матрицы по формулам (13) и (14):

$$F_{i+1} = \frac{1}{d} F_i - \frac{F_i x_k x_k^T F_i}{d + x_k^T F_i x_k},$$

где d – параметр забывания.

Повторение для l от 1 до N :

Шаг 1. Вычисляется η_i параметров кластера:

- априорная вероятность группы (кластера):

$$\alpha_i = \frac{1}{N} \sum_{k=1}^N \mu_{i,k}; \quad (15)$$

- центры кластеров:

$$v_i^x = \frac{\sum_{k=1}^N (\mu_{i,k})^m (x_k - W_i(y_{i,k}))}{\sum_{k=1}^N (\mu_{i,k})^m}; \quad (16)$$

- пересчитываем W_i :

$$\tilde{W}_i = F_i W_i (\sigma_{i,x}^2 I + M_i^{-1} W_i^T F_i W_i)^{-1} \quad (17)$$

- норма отклонения A_i (матрица $n \times n$):

$$A_i = \sigma_{i,x}^2 I + \tilde{W}_i \tilde{W}_i^T; \quad (18)$$

- центр кластера и стандартное отклонение:

$$v_i^t = \frac{\sum_{k=1}^N (\mu_{i,k})^m t_k}{\sum_{k=1}^N (\mu_{i,k})^m} \quad (19)$$

$$\sigma_{i,t}^2 = \frac{\sum_{k=1}^N (\mu_{i,k})^m (x_k - v_i^x)(x_k - v_i^x)^T}{\sum_{k=1}^N (\mu_{i,k})^m} \quad (20)$$

Шаг 2. Вычисляется расстояние $D^2(z_k, h_i)$ при помощи (9).

Шаг 3. Обновление матрицы разбиения:

$$\mu_{i,k}^{(l)} = \frac{1}{\sum_{j=0}^c (D(z_k, \eta_i) / D(z_k, \eta_j))^{2/(m-1)}}, \quad 1 \leq i \leq c, 1 \leq k \leq N, \quad (21)$$

$$A(k) = \sum_{i=1}^k x(i)x^T(i). \quad (22)$$

до тех пор, пока $\|U(l) - U(l-1)\| < \varepsilon$.

Выводы. В статье была рассмотрена модификация алгоритма онлайн-нечеткой сегментации многомерных временных рядов. Алгоритм основан на непрерывной идентификации нечетких множеств, которые представляют собой сегменты во времени. Алгоритм обнаруживает наличие смежных кластеров во времени и в состоянии обнаружить изменения в скрытой структуре многомерных временных рядов. Благодаря рекуррентным формулам обращения матриц, становится возможным использование этого алгоритма в ситуации постоянно изменяющихся параметров.

Список литературы: 1. *Abonyi, J.* Modified Gath–Geva clustering for fuzzy segmentation of multivariate time-series [Text] / *Abonyi J., Babuska R., Szeifert F* // Fuzzy Sets and Systems. – 2005. – №149. – P. 39–56. 2. *Keogh, E.* An Online Algorithm for Segmenting Time Series. [Text] / *Keogh, E., Chu, S. Hart, D., Pazzani, M.* // California, USA: IEEE International Conference on Data Mining. – 2001. – P. 248–296. 3. *Малозёмов, В. Н.* Рекуррентный вариант метода наименьших квадратов. [Текст] / *Малозёмов В. Н., Рыбин С. В.* // М.: Радио и связь – 2004. – 294 с. 4. *Babuska, R.* Improved covariance estimation for Gustafson-Kessel clustering [Text] / *Babuska R., van der Veen P. J., Kaymak U.* // IEEE International Conference on Fuzzy Systems. – 2002 – P. 1081–1085. 5. *Перминов, Г. И.* Метод анализа многомерных временных рядов с использованием корректировки предварительно рассчитанной обратной матрицы: исследование в сравнении с другими методами Data Mining [Текст] / *Перминов Г. И.* // Бизнес-информатика. Методы анализа информации. 2008. – №1 – С. 36-44. 6. *Dobos, L.* Application of on-line multivariate time-series segmentation for process monitoring and control [Text] / *Dobos L., Abonyi J.* // Chemical Engineering Science. 2012. – V.75, №4. – P. 96–105. 7. *Chun Yuan Deng* A generalization of the Sherman–Morrison–Woodbury formula [Text] / *Chun Yuan Deng* // Applied Mathematics Letters. 2011. – V.24, №6 – P. 1561–1564. 8. *Малозёмов, В. Н.* Формулы Фробениуса, Шермана-Моррисона и близкие вопросы [Текст] / *Малозёмов В. Н., Монако М. Ф., Петров А. В.* // Журнал вычислительной математики и математической физики. 2002. – Т.42, №10 – С. 1459-1465. 9. *Xu, R.* Clustering [Text] / *Rui Xu, Donald C. Wunsch II* // New Jersey, USA: A John Wiley & Sons, Inc. – 2009. – P.364. 10. *Daniel, L.* Fuzzy Granular Evolving Modeling for Time Series Prediction [Text] / *Daniel L., Fernando G., Rosangela B., Pyramo C.* // IEEE International Conference on Fuzzy Systems. – 2011. – P. 2794-2801.

Поступила в редколлегию 01.06.2013

УДК 004.89

Адаптивная обработка нестационарных временных рядов на основе нечеткого подхода / Д. А. Машошин // Вісник НТУ «ХПІ». Серія: Нові рішення в сучасних технологіях. – Х: НТУ «ХПІ», – 2013. – № 38 (1011). – С.86-92. – Бібліогр.: 11 назв.

На основі об'єднання нечіткого пакетного способу обробки і сегментації часових рядів з рекуррентними процедурами обробки поточних значень, запропонован метод онлайн сегментації багатовимірних часових рядів, який можна застосувати для виявлення однорідних сегментів в реальному режимі часу на основі потокових даних.

Ключові слова: часові ряди, нечітка кластеризація, рекуррентні процедури, онлайн кластеризація.

By combining the fuzzy batch mode processing and segmentation of time series with recurrent processing procedures current values offered online segmentation method of multivariate time series, which is useful for the detection of homogeneous segments in real time based on the data stream.

Keywords: time series, fuzzy clustering, recursive procedures, online clustering.

УДК 044.03

В. М. ЛЕВЫКИН, д-р техн. наук, проф., зав.каф., ХНУРЭ, Харьков;
О. С. ГНИДЕНКО, аспирант, ХНУРЭ, Харьков

РАЗРАБОТКА КАТЕГОРИЙНОЙ МОДЕЛИ РЕАЛИЗАЦИИ ЖИЗНЕННОГО ЦИКЛА ПРОЦЕССА ВЫБОРА МЕРОПРИЯТИЙ