

*Л. И. Сучкова***Алгоритмическое обеспечение мониторинга
нештатных состояний объекта контроля на основе
многомерных паттернов***L. I. Suchkova***Algorithmic Support for Monitoring the Supernumerary
States of Control Object Based on Multidimensional Patterns**

Рассматривается многомерный паттерн поведения как основа логико-лингвистической модели, позволяющей описывать взаимовлияния контролируемых параметров при мониторинге. Приведены архитектура системы мониторинга и алгоритм оценки адекватности модели.

Ключевые слова: логико-лингвистическая модель, паттерн поведения, мониторинг, адекватность, прогнозирование состояний.

DOI 10.14258/izvasu(2013)1.2-23

Современные системы автоматизированного мониторинга должны обрабатывать разнородные данные, полученные из различных источников — результаты измерений контролируемых параметров, полученные с датчиков; граничные значения контролируемых параметров, установленные согласно условиям протекания процессов в соответствии с нормативами; лингвистические экспертные характеристики значений параметров. При этом представление данных и результаты мониторинга должны быть ориентированы на удобство восприятия информации оператором с учетом требований высокоскоростной обработки данных и реализации возможности прогнозирования. Эти аспекты требуют разработки новых алгоритмических подходов к идентификации текущего состояния контролируемых параметров и, что особенно важно, к идентификации штатных ситуаций.

Рассмотрим систему мониторинга, в которой осуществляется контроль значений группы параметров, и формирование, соответственно, группы временных рядов Y_i , $i = 1, \dots, m$, где m — количество рядов в группе. Каждый i -тый ряд представлен отсчетами y_{iq} , $q = 0, 1, 2, \dots, t_{\text{мек}}$, где $t_{\text{мек}}$ соответствует текущему моменту времени. Так как каждый контролируемый параметр может измеряться или вычисляться с различной периодичностью, то количество доступных для анализа отсчетов для каждого ряда группы может отличаться.

Анализ литературных источников свидетельствует о том, что наиболее приемлемым для анализа групп временных рядов является гибридный подход, в настоящее время интенсивно развиваемый и отраженный в работах С. М. Ковалева, Н. Г. Ярушкиной, Т. В. Афанасьевой, И. Перфильевой, И. Батыршина и др. [1–3]. Гибридный подход к исследованию групп

The paper observes the multidimensional pattern of behaviour as a basis of the logic-linguistic model allowing us to describe interferences of controllable parameters at monitoring. The architecture of monitoring system and algorithm of estimation of model adequacy are resulted.

Key words: logic-linguistic model, behaviour pattern, monitoring, adequacy, forecasting of states.

временных рядов сочетает в себе представления и алгоритмы, характерные для различных представлений моделей рядов и различных методов их анализа, в том числе интеллектуального, включающего нейронные сети, нечеткие вычисления, генетические алгоритмы, экспертные правила, естественно-языковые средства. В настоящей работе предлагается новый подход для идентификации закономерностей изменения поведения в группе временных рядов в системе мониторинга, основанный на нечетких лингвистических паттернах. Понятие перцептивного паттерна, основанного на восприятии понятий человеком, ранее рассматривалось в работах И. Батыршина, Л. Шереметова, J. Kasprzyk, A. Ultsch, F. Mörchen, F. Höppner и др. [2, 4, 5]. В перцептивных паттернах допускается оперирование с числовой и лингвистической информацией, с трендами изменения значений ряда и его производных. Перцептивные паттерны описываются правилами вида «если T есть T_v , то A есть A_v », где $v = 1, \dots, n_v$; T_v — временной интервал; A_v — лингвистическое описание геометрической формы или лингвистический терм. Временной ряд представляется последовательностью примитивных паттернов восприятия A_1, \dots, A_{N_v} , где A_i есть описание формы тренда или текст. Примитивный паттерн P_{pat} определяется по комбинации знаков первой и второй производных, и из последовательности примитивных паттернов формируются темпоральные эпизоды временного ряда.

Нечетко-темпоральная модель С. М. Ковалева, устанавливающая зависимость между продолжительностями и типами тренда ряда, позволяет формализовать описание последовательности трендов и интерпретировать правила в системе автоматизированного анализа, однако в ней не предусмотрен механизм

обнаружения закономерностей поведения группы временных рядов для последующего формирования высказываний и правил. Кроме того, несмотря на интуитивно понятные представления о динамике изменения контролируемых параметров, примитивные паттерны геометрических форм тренда неудобны для анализа группы временных рядов и прогнозирования нештатных ситуаций, так как количество сочетаний темпоральных эпизодов рядов группы может быть очень велико, что затрудняет и делает громоздким описание подобных соотношений.

Введем понятие нечеткого многомерного прогнозирующего паттерна поведения группы временных рядов $Y_p, I = 1, \dots, m$, где m — количество рядов в группе, фиксируемых в системе мониторинга. Нечеткий многомерный прогнозирующий паттерн поведения представляет собой пятерку вида $\langle TP, Num, LP, RP, MP \rangle$, где TP — временной дискрет измерений для паттерна; Num — вектор-столбец размерности s , хранящий номера рядов в паттерне, LP и RP являются матрицами, описывающими поведение группы рядов соответственно до и после текущего момента времени:

$$Num = \{num_p\} \quad LP = \begin{pmatrix} x_{j_1 l_1} & \dots & x_{j_1 l_n} \\ \vdots & & \vdots \\ x_{j_s l_1} & \dots & x_{j_s l_n} \end{pmatrix},$$

$$RP = \begin{pmatrix} z_{j_1 l_{n+1}} & \dots & z_{j_1 l_{n+k}} \\ \vdots & & \vdots \\ z_{j_s l_{n+1}} & \dots & z_{j_s l_{n+k}} \end{pmatrix},$$

где s может принимать значения от 1 до $2 * m$, причем num_1, \dots, num_m соответствуют временным рядам Y_p , $num_{m+1}, \dots, num_{2 * m}$ соответствуют рядам Y'_i , содержащим разности значений $y_{iq+1} - y_{iq}$ рядов Y_i . Каждая строка j_d ($d = 1, \dots, s$) матриц LP и RP соответствует временному ряду с номером num_d вектора Num , столбцы l_p, \dots, l_n соответствуют n временным отсчетам, предшествующим текущему времени $t_{mex} = t_n$, столбцы l_{n+1}, \dots, l_{n+k} соответствуют прогнозу на k отсчетов вперед. Значения $x_{j_d l_h}$ и $z_{j_d l_{h+g}}$ в матрицах LP и RP могут представлять собой:

— лингвистические термы для лингвистических переменных, соответствующих ряду num_d , в том числе термы для обозначения нештатных ситуаций на объекте контроля;

— четкие значения переменных, вычисляемых по отсчетам ряда Y_{num_d} в момент времени l_h , в простейшем случае значения самих отсчетов ряда;

— неопределенные значения (по-значения), соответствующие для матрицы LP незначимому для анализа или отсутствующему в ряде значению, а для матрицы RP — неопределенному значению прогноза.

Назовем матрицы LP и RP соответственно левой и правой частью паттерна. Каждый нечеткий многомерный прогнозирующий паттерн характеризуется маркером MP , обозначающим принадлежность прогнозируемых значений одной из трех областей в пространстве состояний:

— S -области штатных значений, когда прогнозируемые лингвистические значения информативных сигналов соответствуют нормальному функционированию объектов мониторинга (S -маркер);

— K -области критических значений, когда лингвистические значения информативных сигналов соответствуют пограничному лингвистическому терму, предшествующему возникновению нештатных ситуаций (например, терм «очень высокое» для лингвистических переменных «давление на 2 трубопроводе» может являться пограничным значением для нештатных ситуаций «взрыв») (K -маркер);

— NS -области нештатных состояний, когда в матрице RP хотя бы одно значение является нештатной ситуацией (NS -маркер).

Рассмотрим укрупненный алгоритм прогнозирования нештатных ситуаций на базе нечетких многомерных паттернов поведения. Он включает в себя следующие этапы:

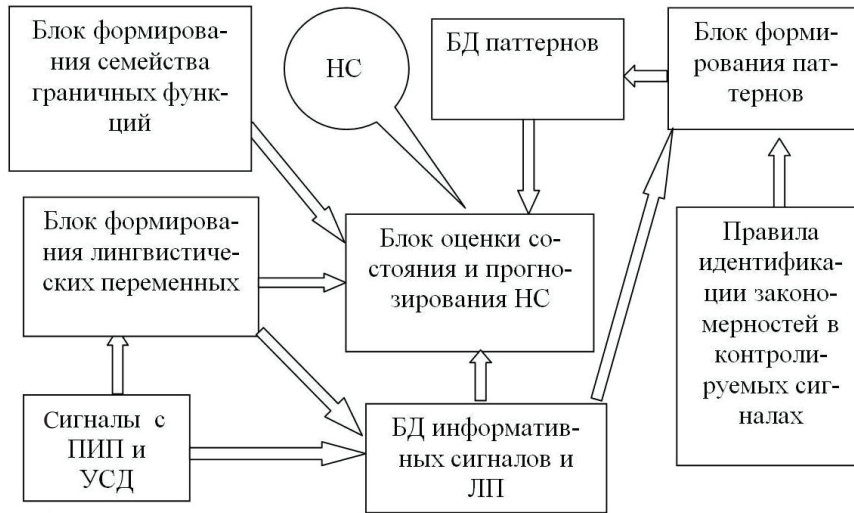
- Получение информационных сигналов с первичных измерительных преобразователей.
- Формирование лингвистических переменных и базы данных.
- Сравнение текущего состояния контролируемых сигналов и их истории с нечеткими паттернами из базы паттернов.
- Если обнаружен паттерн с S -маркером, то перейти к п. 1, иначе к п. 5.
- Если обнаружен паттерн с K -маркером, то уточнить прогноз нештатных ситуаций путем дефазификации пограничных лингвистических термов для сигналов, где спрогнозированы эти термы, и последующего сравнения экстраполированного четкого значения сигнала с прогнозируемым значением граничной функции для четких значений.
- Если в результате уточненного прогноза нештатных ситуаций для паттерна с K -маркером не идентифицируется, то перейти к п. 1, иначе сообщить оператору о прогнозируемой нештатной ситуации.

Для реализации алгоритма прогнозирования нештатных ситуаций разработана автоматизированная система контроля, архитектура которой приведена на рисунке. Система контроля включает в себя следующие компоненты:

- первичные измерительные преобразователи и устройства сбора данных;
- блок формирования лингвистических переменных;
- база данных для хранения параметров информативных сигналов и лингвистических переменных;

- блок формирования семейства граничных функций;
- правила для формирования многомерных нечетких паттернов поведения и идентификации закономерностей в рядах Y_i и Y_i' ;
- блок формирования паттернов;

- база данных для хранения паттернов;
- блок оценки текущего состояния и истории контролируемых сигналов и прогнозирования нештатных ситуаций на базе многомерных нечетких паттернов поведения.



Архитектура системы контроля

Рассмотрим более подробно алгоритмы обработки сигналов и лингвистических переменных в блоках системы контроля. Блок формирования лингвистических переменных сопоставляет каждому из рядов Y_i и Y_i' свою лингвистическую переменную α_i (или α_i'), имеющую имя, соответствующее понятию естественного языка, и набор из kt_{α_i} термов-значений, каждому из которых соответствует своя функция принадлежности LR-типа $f_{\alpha_i c}(u), c=1, \dots, kt_{\alpha_i}$, характеризуемая значениями a_l, a', a'', a_r , а u принимает четкие значения отсчетов рядов Y_i или Y_i' . В каждый момент времени для четких значений группы рядов Y_i и $Y_i', I=1, \dots, m$ осуществляется фаззификация, т.е. вычисление значения функций принадлежности термов и формирование лингвистического термина с максимальным значением функции принадлежности. Будем считать, что для четкого значения u результатом функции фаззификации $Fuz(u)$ является кортеж $\langle c, u \rangle$, где $\chi = \max_l f_{\alpha_i l}(u), l=1, \dots, kt_{\alpha_i}, v = f_{\alpha_i \chi}(u)$; α_d — лингвистические переменные для d -того ряда ($d=1, \dots, 2*m$); c — номер термина; u — максимальное значение степени принадлежности для функций принадлежности лингвистических переменных α_d .

Значения лингвистических переменных и четкие значения информативных сигналов хранятся в базе данных для предоставления возможности оперирования в паттернах поведения данными истории фиксируемых сигналов и их изменений. Выходными данными блока формирования лингвистических переменных являются нечеткие временные ряды, состоящие из термов.

Для уточнения прогноза для паттерна поведения с K -маркером применяется метод, основанный на параболической экстраполяции четких значений ряда и граничных функций. Так же, как и для нечетких термов, для четких значений ряда, хранящего параметры сигнала, можно выделить S-, K- и NS-области изменения значений сигнала, соответствующие штатной, критичной и нештатной ситуации на объекте мониторинга. Выявление переходов значений за границы областей состояния объекта подразумевает определение для четких значений временных рядов граничных функций, являющихся кусочно-непрерывными на заданных временных промежутках. Для каждого ряда Y_i и $Y_i', I=1, \dots, m$ используется два граничных слоя.

Блок оценки состояния и прогнозирования нештатных ситуаций в каждый момент времени осуществляет оценку степени соответствия паттернам базы текущего состояния объектов контроля, описываемого лингвистическими терминами и четкими значениями.

Вычисление степени соответствия текущего состояния объекта контроля нечеткому многомерному паттерну поведения осуществляется по следующему алгоритму:

1. Пока не проанализированы все паттерны базы, выбор из БД паттернов текущего паттерна $PatB$ с минимальным временным дискретом $TP = \tau$.

2. Формирование по БД информативных сигналов и лингвистических переменных матрицы O , содержащей термы и четкие значения, выбранные с дискретом τ . Строки матрицы O соответствуют рядам

с номерами из вектора-столбца Num паттерна $PatB$, соседние столбцы — отсчетам рядов в моменты времени, отличающиеся на t .

3. Пока не проанализированы все ряды паттерна, для каждого ряда, термы и четкие значения которого присутствуют в паттерне $PatB$, по формуле (1) вычисляется абсолютная погрешность $ErrR_d$ соответствия данных измерения ряда d его поведению в d -той строке левой части LP паттерна $PatB$:

$$ErrR_d = \frac{1}{n^*} \sum_{j=1}^n Err_{dj}, \quad (1)$$

где n^* — количество значений, отличных от *non-significant* и *not-defined*, в d -той строке паттерна Err_{dj} — абсолютная погрешность соответствия измеренного значения величине x_{dj} матрицы LP паттерна $PatB$.

Err_{dj} вычисляется в зависимости от типа значения в паттерне:

1. Если x_{dj} имеет тип *non-significant*, то $Err_{dj} = 0$.
2. Если x_{dj} представляет собой терм лингвистических переменных, то погрешность одного измерения

$$Err_{dj} = \frac{|NomTerm(x_{dj}) - NomTerm(o_{dj})|}{kt_d}, \text{ где функция}$$

$NomTerm(ter)$ вычисляет номер термина ter в диапазоне от 1 до kt_d для лингвистических переменных α_d , соответствующей ряду с номером d . Если требуется учет степени принадлежности, то погрешность одного измерения вычисляется с учетом значения функции принадлежности термина по данным измерений в матрице O и степени принадлежности дефаззифицированного термина из паттерна, полученного с применением функции дефаззификации $DeFuz$:

$$Err_{dj} = \frac{\left| f_{\alpha_d NomTerm(x_{dj})}(DeFuz(x_{dj})) \cdot NomTerm(x_{dj}) - f_{\alpha_d NomTerm(o_{dj})}(y_{dj}) \cdot NomTerm(o_{dj}) \right|}{kt_d}.$$

3. Если x_{dj} представляет собой четкое значение, то погрешность одного измерения вычисляется как разность по модулю между четкими значениями:

$$Err_j = |y_{dj} - o_{dj}|.$$

4. Если $ErrR_d$ для любого ряда паттерна превышает заданную погрешность $ErrZ$ соответствия текущей ситуации на объекте контроля паттерну поведения, то остальные ряды паттерна не анализируются и осуществляется переход к п. 1.

5. Вычисление абсолютной погрешности $ErrG$ соответствия данных измерений для группы временных рядов левой части паттерна осуществляется по формуле $ErrG = \max\{ErrR_d\}, d = 1, \dots, s$, где d — номер ряда в векторе Num паттерна $PatB$.

6. Сравнение $ErrG$ с заданной погрешностью соответствия $ErrZ$. Если $ErrG < ErrZ$, то считается, что ситуация на объекте контроля соответствует левой части

паттерна и анализируется маркер паттерна MP , в зависимости от типа которого прогнозируется нештатная ситуация с ее возможным уточнением для K -паттерна.

Таким образом, нечеткий многомерный паттерн поведения является моделью прогнозирования для состояния объекта контроля, адекватность которой оценивается путем сравнения прогнозируемых и фактических термов и четких значений для рядов, входящих в паттерн. В качестве критерия адекватности модели рассмотрим среднюю относительную погрешность $MAPE_p$ применения паттерна P , являющуюся усреднением по количеству применений kol средней относительной погрешности $MAPE_p^r$ при однократном r -том применении паттерна: $MAPE_p = \frac{1}{kol} \sum_{r=1}^{kol} MAPE_p^r$. Средняя относительная погрешность при однократном применении паттерна вычисляется по формуле

$$MAPE_p^r = \frac{1}{s \cdot k} \sum_{d=1}^s \sum_{j=1}^k \Delta_{dj} \cdot 100\%,$$

где s — размерность вектора Nom паттерна; k — количество прогнозируемых отсчетов — столбцов в матрице RP паттерна; Δ_{dj} — относительная погрешность единичного прогноза в паттерне.

Δ_{dj} вычисляется в зависимости от типа прогнозируемого значения z_{dj} в матрице RP паттерна:

- а) если в матрице RP z_{dj} является четким, то

$$\Delta_{dj} = \left| \frac{y_{dj} - z_{dj}}{y_{dj}} \right|,$$

где z_{dj} — прогнозируемое значение в паттерне (четкое); y_{dj} — фактическое четкое значение контролируемого параметра;

- б) если прогнозируется нечеткий терм z_{dj} , то без учета степени принадлежности относительная погрешность единичного прогноза в паттерне вычисляется по формуле

$$\Delta_{dj} = \left| \frac{\chi - NomTerm(z_{dj})}{kt_d} \right|,$$

где χ — номер нечеткого термина w_{dj} , полученного в результате фаззификации фактического четкого значения y_{dj} ; z_{dj} — прогнозируемый терм в паттерне; kt_d — количество термов для лингвистической переменной α_d , соответствующей ряду с номером d и d -той строке матрицы RP паттерна; $NomTerm$ — функция, вычисляющая номер термина-параметра;

- в) если прогнозируется нечеткий терм z_{dj} , то с учетом степени принадлежности ν относительная погрешность единичного прогноза в паттерне вычисляется по формуле

$$\Delta_{dj} = \left| \frac{\chi \cdot \nu - f_{\alpha_d NomTerm(z_{dj})}(DeFuz(z_{dj})) \cdot NomTerm(z_{dj})}{kt_d} \right|,$$

где ν — значение функции принадлежности для фактического термина w_{dj} .

Закономерности динамики и взаимовлияния временных рядов группы с учетом концептов длительности и следования описываются на внутреннем языке, порождаемом разработанной контекстно-свободной грамматикой. Для проверки применимости этих закономерностей к реальным данным измерений используется блок формирования паттернов, включающий интерпретатор, работающий по принципу рекурсивного спуска. В результате интерпретации описания закономерностей по реальным данным измерений вычисляется степень соответствия между прогнозируемыми значениями в описании и фактическими значениями, по которой можно оценить правильность паттерна поведения группы временных рядов. Для интеграции паттерна поведения в систему мониторинга необходимо определить его маркер, причем для паттернов с К-маркером для уточнения прогноза требуется задание нижней и верхней границ области допустимых значений контролируемых сигналов.

Предложенные алгоритмы интеллектуального мониторинга состояния объекта контроля практически реализованы в программе на С# [6], выполнена тестовая проверка паттернов для системы температурного контроля на предприятии, осуществляющем хранение сельскохозяйственной продукции.

Несомненным достоинством модели прогнозирования нештатных ситуаций на основе нечеткого многомерного паттерна поведения является замена сложной физико-математической модели реальных наблюдаемых процессов на логико-лингвистическую модель, позволяющую оперировать качественными понятиями, удобными не только для описания взаимовлияния контролируемых параметров, но и для представления результатов мониторинга. Использование паттернов в системе контроля позволяет отказаться от постоянной корректировки ее алгоритмов и программного обеспечения в связи с появлением новых паттернов.

Библиографический список

1. Ковалев С. М. Гибридные нечетко-временные модели временных рядов в задачах анализа и идентификации слабо формализованных процессов // Сборник трудов IV Международной научно-практической конференции. — М., 2007. — Т. 1.
2. Battyshin I., Kacprzyk J., Sheremetov L., Lotfi A. Zadeh. Perception-based Data Mining and Decision Making in Economics and Finance. — Springer, 2007.
3. Ярушкина Н. Г., Афанасьева Т. В., Перфильева И. Г. Интеллектуальный анализ временных рядов: учеб. пособие. — Ульяновск, 2010.
4. Mörchen F., Ultsch A. Mining Hierarchical Temporal Patterns in Multivariate Time Series [Электронный ре-

сурс]. — URL: <http://citeseerx.ist.psu.edu/viewdoc/download>.

5. Höppner F. Discovery of temporal patterns — learning rules about the qualitative behavior of time series. In: Proc. of the 5th European Conference on Principles and Practice of Knowledge Discovery in Databases, Lecture Notes in Artificial Intelligence 2168. — Springer, 2001.

6. Сучкова Л. И. Построение прогнозирующих нечетких паттернов для группы регистрируемых в системах мониторинга сигналов (Detection Forecast Fuzzy Patterns): св-во о гос. регистрации программы для ЭВМ №2012619879. Зарегистрировано в Реестре программ для ЭВМ 31.10.2012 г.