



УДК 658.512.011

I.I.JUMANOV, Z.T.BEKMURODOV

**METHODS OF OPTIMIZING DATA PROCESSING BASED ON FUZZY CORRECTION OF
TIME SERIES ELEMENTS AND VARIABLE IDENTIFICATION MODELS**

Тасодифий вақтли қатор (ТВҚ) идентификациясини таъминловчи статистик, динамик, нотиниқ модель, нейрон тармоғи хоссаларидан фойдаланиш асосида маълумотларга оптимал ишлов бериш муаммоси талқин этилган. Нотиниқ тенглама, мантиқий хулосалаш, ташиқи муҳит омили, жараён ностационарлиги таъсирларини инобатга олувчи алгоритмлар бўйича ТВҚни сифатли идентификациялаш усули таклиф этилган. Нотиниқ мантиқ, t ва S -нормалар бўйича экстремумни излаш услублари асосида ўзгарувчи қийматини мослаш ва сошлаш ҳамда идентификацияловчи умумлашган алгоритм яратилган. Хос моделни аниқлаш; параметрли ва таркибий идентификациялаш; излашни мақбуллаштириш; моделни ўргатиш; “киришлар ва чиқишлар” боғлиқлигини топиш; билимлар базасини шакллантириш ва нотиниқ қоидалар, лингвистик ўзгарувчилар, тегишлилик функциялари танламаларидан фойдаланиш; нотўғри маълумотларни ва хатоликни нотиниқ назорат қилиш ҳамда ўзгарувчи қийматларини мослашни амалга оширувчи алгоритмлар яратилган. Жорийлаштирилган дастурий мажмуа маълумотга ишлов бериш аниқлигини юқори даражада ҳамда энг кам харажатларда амалга оширади.

Таянч сўзлар: тасодифий вақтли қатор, ностационар объект, нотиниқ тенглама, мантиқий хулосалаш, нотиниқ идентификациялаш, ростлаш, монанд модель, оптималлаштириш, ўзгарувчилар қийматларини тўғрилаш, дастурий мажмуа.

Сформулирована проблема оптимальной идентификации и обработки случайных временных рядов (СВР) на основе использования свойства статистических, динамических, нечетких моделей. Предложен метод качественной идентификации СВР, включающий алгоритмы нечетких уравнений, логических выводов, учета воздействий факторов внешней среды и нестационарности процессов. Разработан обобщенный алгоритм идентификации СВР с регулированием и корректировкой значений переменных на основе правил нечеткой логики, способов поиска экстремумов по t -норм и S -норм. Спроектированы инструменты для оптимальной обработки данных путем определения адекватной модели; параметрической и структурной идентификации объектов; оптимизации поиска; обучения моделей; идентификации зависимости «входов и выходов»; формирования и использования базы знаний, а также наборов нечетких правил, лингвистических переменных, функций принадлежности и алгоритмов регулирования значений переменных. Разработаны методы нечеткой коррекции искаженной информации путем контроля погрешности идентификации СВР, а также реализован программный комплекс, обеспечивающий высокой точности обработки данных со значительно меньшими затратами.

Ключевые слова: случайный временный ряд, нестационарный объект, нечеткие уравнения, логический вывод, нечеткая идентификация, регулирование, адекватная модель, оптимизация, корректировка значений переменных, программный комплекс.

The problem of optimal identification and processing of random time series (RTS) based on the use of the property of statistical, dynamic, fuzzy models is formulated. A method for the qualitative identification of RTS is proposed, including algorithms for fuzzy equations, logical conclusions, taking into account the effects of environmental factors and the nonstationarity of processes. A generalized algorithm for identifying the RTS with regulation and correction of variables values based on fuzzy logic rules, ways of searching for extrema by t norms and S -norms is developed. Designed tools for optimal data processing by determining the appropriate model; parametric and structural identification of objects; search optimization; learning models; identification of the dependence of "inputs and outputs"; formation and use of knowledge base, as well as sets of fuzzy rules, linguistic variables, membership functions and algorithms for regulating the values of variables. Methods of fuzzy correction of distorted information by means of the error control of RTS identification have been developed, and a software package has been implemented to providing high accuracy of data processing with significantly lower costs.

Keywords: random time series, non-stationary object, fuzzy equations, logical conclusion, fuzzy identification, regulation, adequate model, optimization, correction of variables values, software complex.

Актуальность темы. В существующей литературе за последние 10-15 лет подтверждается актуальность исследований, направленных на совершенствование и развитие методов повышения достоверности передачи и обработки данных в системах управления производственно – технологическими процессами на основе адекватной идентификации случайных временных рядов (СВР) нестационарных объектов с помощью статистических, динамических, нечетких моделей и нейронных сетей [1,2].

Статистические и динамические методы идентификации СВР, рассматриваемые в традиционных системах включают механизмы регулирования значений переменных моделей, извлечения и использования наиболее типичных статистических характеристик таких, как математического ожидания, среднеквадратического отклонения, законов распределения, автокорреляционных функций, коэффициентов парных, взаимных, множественных корреляционных связей, а также коэффициентов динамических характеристик объектов.

Для установления их значений проводятся обширные экспериментальные исследования и расчеты по обширной априорной статистике измерения переменных [3,4].

В работах [5] предложена методика гибридной идентификации СВР, построения прозрачных и эффективных алгоритмов для оптимизации обработки данных, контроля значений погрешности статистических и динамических моделей идентификации, когда искажения информации обуславливаются из-за ошибок, допускаемых на этапах ввода, передачи, формирования и обработки данных.

Большую значимость представляет разработка методов повышения достоверности передачи и обработки данных на основе методов, использующих свойства и особенности алгоритмов нечеткой логики и поиска экстремумов по способам t -норм и s -норм, а также учитывающих воздействие факторов внешней среды, априорную недостаточность, параметрическую неопределенности и нестационарность переменных.

Настоящая работа посвящена разработке методов повышения достоверности информации на основе решения следующих задач:

а) устранение искажений в информации, обусловленных ошибками на этапах ввода, передачи и формирования данных путем применения пороговых и адаптивных границ контроля погрешностей статистической и динамической идентификации СВР;

б) фильтрация случайных всплесков, шумовых составляющих, нестационарности в последовательности измерений СВР на основе контроля погрешности нечеткой модели идентификации СВР с использованием механизмов регулирования значений переменных моделей на основе алгоритмов нечеткой логики;

в) нечеткая идентификация СВР с использованием механизма регулирования параметров компонентов нейронной сети (НС), вычислительной схемы нейро-нечеткой сети (ННС), а также с учетом условий априорной недостаточности и параметрической неопределенности.

Следует отметить, что методы повышения достоверности обработки данных, устраняющих искажения в информации на этапах ввода, передачи и формирования информации применением пороговых и адаптивных алгоритмов контроля погрешностей статистической и динамической идентификации СВР изложены в работах [5].

Основное внимание ниже уделяется представлению результатов исследований и разработок, относящихся к задачам пунктов б) и в).

Оптимизация нечеткой идентификации СВР на основе нечеткой логики. В исследованиях, направленных на повышение достоверности передачи и обработки данных нестационарных объектов и применение статистических и динамических моделей большое

внимание уделяется построению инструментария уменьшения погрешности идентификации СВР с помощью механизмов контроля и корректировки искаженных элементов, регулирования значений переменных моделей на основе привлечения сложных рекуррентных выражений, использования дифференциальных и разностных уравнений.

Для нечеткой идентификации СВР в условиях недостаточности априорных сведений и параметрической неопределенности исследование и разработка методов повышения достоверности обработки данных связана с расширением решений задач оптимизации, адекватной идентификации СВР, контроля погрешности, регулирования переменных на основе привлечения нечеткой логики и учета воздействия внешней среды, а также свойства нестационарности случайных процессов.

Методы нечеткой идентификации, корректировки искаженных элементов СВР, контроля погрешности, регулирования на основе нечеткой логики используют лингвистическую информацию, базы данных (БД) и базы знаний (БЗ) нечетких правил и реализуют типичные вычислительные схемы нечетких моделей.

В связи с этим проектирование упрощенных и эффективных инструментариев оптимизации достоверности обработки данных на основе мягких вычислений достигается на основе решения следующих взаимосвязанных задач:

- определение адекватной модели идентификации СВР, функциональных зависимостей «входы и выход», а также формирование нечетких БЗ, лингвистических переменных, набора нечетких правил, способов увязки нечетких правил;

- формирование функций принадлежности (ФП) терм множеств лингвистических переменных входов и выхода, построение алгоритмов регулирования переменных моделей на основе правил нечетких выводов;

- проведение параметрической и структурной идентификации, оптимизации поиска экстремумов, обучения нечетких алгоритмов;

- определение рационального набора обучающих данных, нечетких правил контроля погрешности идентификации с целью повышения точности обработки данных;

- проектирование программного комплекса идентификации и обработки данных, которые в отличие от существующих используют возможности, свойства статистических, динамических, нечетких моделей, алгоритмов нечеткой логики, а также особенности информации нестационарных объектов.

Для нечеткой идентификации СВР предлагается следующая модифицированная вычислительная схема:

- фазификатор преобразует фиксированное значение вектора входной переменной X в нечеткие элементы \tilde{X} ;

- экстракция БЗ, которая содержит нечеткие правила лингвистических переменных о зависимостях «входы и выход»;

- генерация нечетких выводов, определяющие значения выходной переменной \tilde{Y} в соответствии нечетким значением \tilde{X} , а также оцениваются параметры ФП термов - лингвистических переменных;

- дефазификатор преобразует значение выхода из нечеткого множества \tilde{Y} в реальные числа Y .

Корректировки искаженных элементов СВР проводятся на основе алгоритмов нечеткой логики Мамдани и Сугэно, отличие которых заключается в формате задания БЗ.

Для нелинейных зависимостей «входы - выход» инструмент регулирования значений элементов СВР производится на основе лингвистических высказываний типа «ЕСЛИ - ТО». Тогда для выводов Мамдани оценки выхода Y задаются в виде [6]:

ЕСЛИ $(x_1 = a_{1,j1})$ И $(x_2 = a_{2,j1})$ И...И $(x_n = a_{n,j1})$

ИЛИ $(x_1 = a_{1,j2})$ И $(x_2 = a_{2,j2})$ И...И $(x_n = a_{n,j2})$

...

ИЛИ $(x_1 = a_{1,jk_j})$ И $(x_2 = a_{2,jk_j})$ И...И $(x_n = a_{n,jk_j})$

ТО $y = d_j, i = \overline{1, m}$,

где $a_{i,jp}$ - оценки лингвистического термина входной переменной $x_i, i = \overline{1, n}$, которые записываются в строку матрицы с номером jp ($p = \overline{1, k_j}$), $j = \overline{1, m}$; k_j - количество строк-конъюнкций, по которым оценивается лингвистический терм d_j выхода y ; m - количество термов лингвистической переменной y .

Из БД извлекается набор данных, представляемых для идентификации СВР в виде

$$(X_i, Y_j), j = \overline{1, m},$$

где $X_i = (x_{i1}, x_{i2}, \dots, x_{in})$ - входной n - мерный вектор; $Y_j = (y_1, y_2, \dots, y_m)$ - вектор выхода вычислительной схемы нечеткой идентификации СВР.

Нечеткая корректировка искаженных элементов СВР, проведение адекватной идентификации, контроля погрешности, регулирования переменных вычислительных схем нечеткой модели, повышение достоверности обработки данных требует выделения эффективного набора данных для адаптивной идентификации нестационарных объектов.

Вычисляется выходная величина по адаптированной вычислительной схеме в виде

$$Y_j^f = b_{m0} + b_{m1}x_1^j + \dots + b_{mn}x_n^j,$$

где Y_j^f - выходная величина предиктора СВР, полученная на основе X_i и представляемая в виде вектора B .

Вычислительная схема позволяет также вычислить погрешности сравнения с целью минимизации их значений в виде

$$D_r = \sum_{r=1}^M (X_r - Y_r^f) \rightarrow \min,$$

где r - строка входной матрицы X_r .

Матрица вектора B задается в виде

$$B = (b_{ij}), i = \overline{1, m}, j = \overline{0, n},$$

Реакции БЗ, представляются в виде операций \cup и \cap [7]

$$\bigcup_{p=1}^{k_j} \bigcap_{i=1}^n (x_i = a_{i,jp}) \rightarrow y = d_j, j = \overline{1, m}.$$

В качестве инструмента корректировки значения искаженной СВР используются механизмы определения типа, границ ФП $\mu_{jp}(x_i)$ и интервалы принадлежности элементов нечетким множествам для каждого входных и выходного лингвистических термов в виде нечетких уравнений [8]

$$\tilde{y} = \text{agg} \left(\int_{j=1, m}^{\bar{y}} \text{imp}(\mu_{d_j}(X^*), \mu_{d_j}(y)) / y \right),$$

где *imp* - импликация, реализуемая как операция нахождения минимума; *agg* - агрегирование нечетких множеств, реализуемое операцией нахождения максимума.

Элементы входной матрицы X_r регулируются на основе результатов нечетких выводов

$$Y_r^f = \frac{\sum_{i=1}^m \mu_{d_i}(X_r) \cdot d_i}{\sum_{i=1}^m \mu_{d_i}(X_r)},$$

где $d_i \in D_r$ результат нечетких выводов, формируемый по i -му нечеткому правилу.

Для регулирования параметров вычислительной схемы нечеткой модели ФП для всей последовательности элементов СВР задается, как

$$\begin{aligned} \mu_{d_i}(X_r) &= \mu_{i1}(x_{r1}) \cdot \mu_{i1}(x_{r2}) \cdot \mu_{i1}(x_{r3}) \cdot \dots \cdot \mu_{i1}(x_{rn}) \vee \\ &\vee \mu_{i2}(x_{r1}) \cdot \mu_{i2}(x_{r2}) \cdot \mu_{i2}(x_{r3}) \cdot \dots \cdot \mu_{i2}(x_{rn}) \vee \dots \\ &\dots \vee \mu_{im}(x_{r1}) \cdot \mu_{im}(x_{r2}) \cdot \mu_{im}(x_{r3}) \cdot \dots \cdot \mu_{im}(x_{rn}), \end{aligned}$$

Результат механизма регулирования вычислительных схем оценивается в виде

$$y_r^f = \sum_{i=1}^m \beta_{ir} d_i = \sum_{i=1}^m (\beta_{ir} \cdot b_{i0} + \beta_{ir} \cdot b_{i1} \cdot x_{r1} + \beta_{ir} \cdot b_{i2} \cdot x_{r2} + \dots + \beta_{ir} \cdot b_{in} \cdot x_{rn}),$$

где $\beta_{ir} = \frac{\mu_{d_i}(X_r) \cdot d_i}{\sum_{i=1}^m \mu_{d_i}(X_r)}$ - переменная, значение которой определяется i нечеткими правилами.

Вычислительные схемы позволяют сформировывать следующие матрицы переменных:

$Y^f = (y_1^f, y_2^f, \dots, y_M^f)^T$ – результаты, полученные по нечеткой идентификации с механизмом регулирования на основе нечеткой логики;

$Y = (y_1, y_2, \dots, y_M)^T$ – результаты, полученные по модели нечеткой идентификации без механизма регулирования значений переменных вычислительных схем. Результат оптимизации обработки данных оценивается по функционалу

$$E = (Y - Y^f)^T \cdot (Y - Y^f) \rightarrow \min.$$

В работе важное внимание уделено вопросу проектирования оптимальной структуры лингвистических входных и выходной переменных. В результате чего, механизм корректировки искаженных значений СВР приобретает следующие особенности:

- термы выходной переменной d_j по Мамдани задаются синглтонами – нечеткими аналогами реальных чисел;
- степень принадлежности одного из элементов универсального нечеткого множества будет равна единице, а остальные нулю;
- заключения БЗ нечетких правил по Сугэно задаются функциями, в которых первоначальные коэффициенты принадлежности терм - входных переменных, соответствующему нечеткому множеству равны нулю;
- БЗ нечетких правил включают виды посылок, свойств нечетких множеств, заключения в виде четкой линейной функции.

Результаты исследований реализованы в виде программного комплекса идентификации и обработки данных нестационарных объектов.

Анализ результатов тестирования программного комплекса. Для тестирования программного комплекса в системе мониторинга производственно-технологическими комплексами формируются следующие массивы:

X_r - входных данных;

I - термы входных и выходной лингвистических переменных;

O - вектор переменных, параметры и границы ФП; интервалы принадлежности терм - лингвистических переменных, соответствующих нечетким множествам;

B - типы нелинейных зависимостей «входов и выходов» в заключениях нечетких правил.

I_0, I_T - наборы обучающих и тестирующих данных;

R_S - классы, сформированные по числу нечетких правил и наборов обучения;

X_i^S, y_i^S - входы и выходы, определяемые по классу $R_S, X_i \in R_S, i = 1, 2, \dots, I_S$;

I_S - число измерений СВР в наборе обучения;

m_i^S - вектор класса по нечетким правилам R_S ;

$m_{A_S}(X_i) = m_i^S$ степень принадлежности входного элемента классу $R_S, i = 1, 2, \dots, r$;

A_S - нечеткое множество принадлежности элемента лингвистического терма по S -му нечеткому правилу;

u_p^* - выходное значение.

Комплекс включает следующие основные программные модули:

- предварительная обработка данных (сегментация, выделение контура объекта, идентификация и аппроксимация, оценка погрешности);
- формирование обучающих и тестовых наборов данных;
- расчет по вычислительным схемам и обучение нечеткой модели;
- генерация логических выводов;
- регулирование и корректировка значений переменных;
- оценка точности (достоверности) обработки данных.

Доказана эффективность нелинейных идентификаторов нечетких СВР по инструментариям пакета Fuzzy Logic Toolbox.

Реализованы программы параметрической идентификации, заданных в виде нелинейных зависимостей на базе нечеткой гибридной модели в среде MATLAB. Настройка комплекса проводится как в командном режиме, так и в диалоговом режиме.

Для параметрической идентификации на основе Мамдани разработана программа, которая выполняет следующие функции:

- образование m -сценария, вызывающего функцию нелинейной оптимизации;
- образование m -функции, вычисляющей невязку в заданных значениях переменных.

Количество лингвистических переменных принято равным шестнадцати. Из них одиннадцать – коэффициенты концентрации ФП входных и выходной переменных термов; две – координаты максимумов центров ФП термов «средний» входных переменных; три – координаты максимумов центров ФП не крайних термов выходных переменных: «ниже среднего», «средний» и «выше среднего».

Среднеквадратическая ошибка параметрической идентификации СВР на контрольной выборке из 1000 точек по условному параметру составляет 4.61.

Идентификации на основе Сугэно выполняется применением алгоритмов решеточного разбиения и субтрактивной кластеризации.

Первый алгоритм позволяет сформировать БЗ все возможных нечетких правил.

Второй алгоритм генерирует нечеткие правила, соответствующих областям наибольшей концентрации точек ФП. Каждая входная лингвистическая переменная оценивается колоколообразной ФП.

Настраивались 24 переменных, из которых:

– 3 коэффициента для каждого из 4-х правил БЗ;

– 3 коэффициента для каждого из 4-х термов входных переменных.

Количество настраиваемых переменных для Сугэно с двумя входами и одним выходом является минимальным. При этом среднеквадратическая ошибка структурной идентификации на контрольной выборке из 1000 точек составляет 1.81.

В третьей части экспериментального исследования проведено тестирование вычислительных схем нечеткой модели, операторов выработки нечетких выводов, расширения или редукции состава нечетких правил и формирования БЗ с скорректированными нечеткими правилами. Реализованы вычислительные схемы моделей Мамдани, Сугено и полинома 4-го порядка.

Определено, что погрешность нечетких моделей значительно ниже, чем погрешности методов полиномиальной идентификации СВР.

References:

1. Barsegyan A. A., Kupriyanov M. S., Stepanenko V. V. Tehnologii analiza danny'h: Data Mining, Visual Mining, Text Mining, OLAP, SPb.: BHV- Peterburg, 2007. - 384 s.
2. Han j., Kamber M., Pei j. Data mining: concepts and techniques. - 3rd ed. - Morgan Kaufmann / Elsevier, 2012. - 744 p.
3. Jumanov I.I., SHarapova N.A. Sistema analiza i obrabotki danny'h o kadrah vuzov na osnove nechetkogo regul'yatora s fiksirovannoy bazoy danny'h // «Himicheskaya tehnologiya. Kontrol' i upravlenie» - TGTU, Tashkent, 2014- № 2 (56) . - s. 53-61
4. Jumanov I.I. Sistema analiza i obrabotki danny'h o kolichestvenny'h i kachestvenny'h harakteristikah VUZa na osnove apparata myagkih vy'chisleniy // «Nauchny'y vestnik SamGU». - Samarkand: SamGU, 2014 - №1 (83) - 35-42 b.
5. Djumanov O.I., Kholmonov S.M Methods and algorithms of selection the informative attributes in systems of adaptive data processing for analysis and forecasting // “Applied Technologies and Innovations”, Prague Development Center. - Prague, 2012. - Volume 8, November 2012. - pp.45-55.
6. Makary'chev P. P., Afonin A. YU. Operativny'y i intellektual'ny'y analiz danny'h. Penza, Izd-vo PGU, 2010. - 156 s.
7. Rotshteyn A. P. Intellektual'ny'e tehnologii identifikacii: nechetkaya logika, geneticheskie algoritmy', neyronny'e seti. - Vinnica, 1999.- 320 s.
8. Mityushkin YU. I., Mokin B. I., Rotshteyn A. P. Soft - Computing: identifikaciya zakonomernostey nechetkimi bazami znaniy. - Vinnica, 2002.- 145 s.

Жуманов Исроил Ибрагимович – доктор технических наук, профессор кафедры «Информационных технологий» Самаркандского государственного университета.

Тел.: (+99866)2293558 (м.), E-mail: olimjondi@mail.ru;

Бекмуродов Зоҳид Толибович – ассистент кафедры «Информационных технологий»

Самаркандского государственного университета.

Тел.: +998937255222 (м.), E-mail: zohid3788737@mail.ru.