https://doi.org/10.26160/2572-4347-2023-18-4-12

ИДЕНТИФИКАЦИЯ ДИНАМИЧЕСКИХ СИСТЕМ НА ОСНОВЕ ОБРАБОТКИ ЭКСПЕРИМЕНТАЛЬНЫХ ДАННЫХ МЕТОДОМ ГЕНЕТИЧЕСКОГО ПРОГРАММИРОВАНИЯ

***Чжан Лэл*[1]*, Филимонов Н.Б.*[1,2]**

[1]*Московский государственный технический университет им. Н.Э. Баумана;*

[2]*Московский государственный университет им. М.В. Ломоносова, Москва, Россия*

Ключевые слова: идентификация модели динамических систем, экспериментальные данные, метод генетического программирования, система Лоренца, компромисс между точностью и сложностью модели.

Аннотация. В работе рассматривается задача идентификации динамических систем на основе обработки экспериментальных данных методом генетического программирования. Приводится пример построения по результатам наблюдения эмпирической модели системы Лоренца. Обсуждается вопрос компромисса между точностью и сложностью идентифицированных математических моделей.

IDENTIFICATION OF DYNAMIC SYSTEMS BASED ON EXPERIMENTAL DATA PROCESSING USING GENETIC PROGRAMMING

***Zhang Lele*[1]*, Filimonov N.B.*[1,2]**

[1]*Bauman Moscow State Technical University;*

[2]*Lomonosov Moscow State University, Moscow, Russia*

Keywords: identification of dynamic systems model, experimental data, genetic programming method, compromise between accuracy and complexity of the model.

Abstract. The problem of the identification of the dynamic systems based on the processing of experimental data by the method of genetic programming is considered in the paper. An example of the construction of an empirical model of the Lorentz system based on the results of observation is given. The question of a compromise between the accuracy and complexity of the identified mathematical models is discussed.

Введение

Одним из интенсивно развивающихся направлений современной теории управления является идентификация систем, связанная с проблемой построения математических моделей в виде совокупности математических соотношений, адекватно отражающих основные свойства системы. Математическая модель, являясь образом («эквивалентом») некоторой реальной системы (явления, процесса) находит самое широкое применение в современной науке и технике, позволяя как исследовать систему, так и управлять ею [1].

Различают теоретические и эмпирические математические модели. Теоретические модели систем получают путем изучения свойств системы и протекающих в ней процессов. Однако это не всегда возможно из-за незнания механизма функционирования системы и теории протекающих в ней процессов. В этих случаях обычно используют т.н. *эмпирические модели* систем, которые получают посредством обработки результатов наблюдения (измерений) внешних проявлений свойств и процессов, протекающих в системе [2]. Данные модели лишены физического смысла, но удовлетворительно описывают экспериментальные данные, характеризующие свойства реальных систем, и являются единственно возможной альтернативой строгим математическим моделям.

В зависимости от априорной информации о системе различают задачи построения математических моделей системы в узком смысле (задача параметрической идентификации) и в широком смысле (задача структурной идентификации). В настоящей работе обсуждается задача структурно-параметрической идентификации динамических систем по результатам наблюдений.

Задача идентификации динамических систем в парадигме «черного ящика»

В современной теории идентификации, согласно концепции «черного ящика» (Н. Винер и У.Р. Эшби) [3], эмпирические математические модели системы строятся на основе изучения наблюдений результатов специально спланированного эксперимента – реакций системы на возмущения внешней или внутренней среды. При этом предполагается, что устройство и механизм функционирования системы неизвестны, недоступны для наблюдения. В этом случае система рассматривается как «черный ящик» без детализации его «внутренности», а модель системы рассматривается как некоторая интерполяционная зависимость, полученная на основе экспериментальных наблюдаемых данных непосредственно на самой системе или на ее физической модели. Следует заметить, что переход к эмпирическим моделям предполагает отказ от теоретических математических моделей и именно поэтому эмпирические модели более разнообразны и включают в себя различные по форме математические зависимости.

Рассмотрим следующую задачу структурно-параметрической идентификации динамической системы.

Положим, что дана нелинейная динамическая система n-го порядка, описываемая некоторым неизвестным дифференциальным уравнением в форме Коши с неизвестными параметрами:

$$\frac{d}{dt}\mathbf{x} = \mathbf{f}(\mathbf{x}, \boldsymbol{\theta}, t)\,,\ \mathbf{x}(t_0)\,,$$

где $\mathbf{x} \in R^n$ – вектор состояния системы; $t > t_0$ – время, $\mathbf{f}(\cdot)$ – нелинейная вектор-функция неизвестных характеристик системы; $\boldsymbol{\theta} \in R^k$ – вектор неизвестных параметров; $\mathbf{x}(t_0)$ – вектор начального состояния системы.

Пусть в результате проведения некоторых экспериментов над системой проводятся наблюдения входных переменных (начальное состояние $\mathbf{x}(t_0)$) и выходных переменных (текущее состояние $\mathbf{x}(t)$) системы в дискретные фиксированные моменты времени $t = t_i$, $i = \overline{1, m}$, сосредоточенные в матрице $\mathbf{X}$:

$$\mathbf{X} = (\mathbf{x}(t_1), \mathbf{x}(t_2), ..., \mathbf{x}(t_m))^{\mathrm{T}}.$$

По результатам наблюдения, на основании экспериментальных данных $\mathbf{X}$ требуется определить структуру (вид) и параметры системы, т.е. функцию $\mathbf{f}(\mathbf{x},\boldsymbol{\theta})$, ставящую в соответствие входным переменным $\mathbf{x}(t_0)$ выходные переменные $\mathbf{x}(t)$.

Для оценки степени адекватности модели системы ее оригиналу, т.е. бли-зости расчетных данных $\mathbf{x}(t_i)$ и экспериментальных данных $\tilde{\mathbf{x}}(t_i)$ введем в рассмотрение ошибки измерения (невязки) в виде

$$\varepsilon(t_i) = \mathbf{x}(t_i) - \tilde{\mathbf{x}}(t_i),$$

а в качестве критерия точности идентификации системы – среднюю абсолютную ошибку (англ. mean absolute error, MAE):

$$MAE = \frac{1}{m} \sum_{i=1}^{m} \left| \varepsilon(t_i) \right|.$$

В результате, поставленную задачу идентификации системы можно рассматривать как задачу нахождения эмпирической математической модели динамической системы, т.е. функции $\mathbf{f}(\mathbf{x},\boldsymbol{\theta})$, обеспечивающей экстремум *MAE*:

$$MAE \to \min.$$

Поставленная задача, как и все задачи структурно-параметрической идентификации систем, является одной из наиболее трудных и малоизученных задач (см., напр., [4, 5]). При этом, в последнее время для решения данных задач все большую популярность находят методы эволюционных вычислений и, в частности, метод ГП (см., например, работы [6-9]).

Особенности применения метода ГП в задачах идентификации динамических систем

Метод ГП, предположенный John Koza в 1992 г. [10], является разновидностью эволюционных вычислений, относится к классу методов символьной регрессии и основан на дарвиновском естественном отборе и выживании наиболее приспособленных. В методе ГП, опирающимся на концепцию генетического алгоритма, автоматически «выращиваются» компьютерные программы для решения оптимизационной задачи без каких-либо априорных знаний о предметной области. Задача заключается в нахождении математического выражения из пространства возможных выражений, которые могут быть составлены из набора доступных функций и терминалов. При этом наиболее распространенной структурой для представления особей (потенциальных решений задачи) является древовидное представление формы генома [11, 12]. Здесь программы представляются в виде синтаксических деревьев: листья дерева соответствуют терминалам, а внутренние узлы – функциям. Деревья режутся оператором кроссинговер, а образованные поддеревья обмениваются между собой.

Применение ГП к рассматриваемой задаче структурно-параметрической идентификации динамических систем основано на ее сведении к задаче символьной регрессии, т.е. к одновременному поиску оптимальный структуры и параметров математической модели системы путем перебора различных произвольных суперпозиций функций («программ») из некоторого набора. Общая структура алгоритма ГП содержит следующие традиционные шаги: создание популяции особей – случайных потенциальных решений; оценка приспособленности решений по значению фитнес-функции; создание популяции нового поколения особей; выбор наилучших решений; генетические операции репродукции, скрещивания и мутации; проверка условий остановки эволюции, принятие решения об окончании, либо о повторении итерации поиска.

Для использования метода ГП в прикладных задачах оптимизации доступно множество программ, например, HeuristicLab университета прикладных наук в Верхней Австрии, ECJ на языке Java, gplearn на языке Python. В настоящей работе решение задачи идентификации динамической системы методом ГП проводилось в среде Python на основе пакета DEAP (англ. Distributed Evolutionary Algorithms in Python).

Пример идентификации системы Лоренца методом ГП

В качестве примера рассмотрим задачу идентификации эмпирической модели динамической системы Лоренца (E. Lorenz), описываемой системой обыкновенных дифференциальных уравнений вида

$$\begin{cases} \dot{x} = a(y - x), \\ \dot{y} = cx - xz - y, \\ \dot{z} = xy - bz. \end{cases}$$

Полагаем, что модель динамики данной системы неизвестна и является «черным ящиком». Приведенные уравнения ее динамики с параметрами $a = 10$, $b = 28$, $c = 8/3$ используем для получения экспериментальных данных (результатов наблюдения) изменений вектора состояния $\mathbf{X} = (x, y, z)^{\mathrm{T}}$ и его скорости $\dot{\mathbf{X}} = (\dot{x}, \dot{y}, \dot{z})^{\mathrm{T}}$ путем интегрирования (от $t = 0$ до $t = 20\text{с}$ с шагом $\Delta t = 0.001\text{с}$) методом Рунге-Кутта в среде Matlab. Обычно экспериментальный материал делят на две части: одну используют для проведения идентификации модели, другую оставляют для процедуры верификации модели с целью проверки эффективности повторения неиспользованных при идентификации данных. В нашем случае первая половина наблюдений (от $t = 0$ до $t = 10\text{с}$) используется для построения модели системы, а вторая половина наблюдений (от $t = 10\text{с}$ до $t = 20\text{с}$) используется для сравнения с результатами, полученными на построенной модели.

Для идентификации нелинейной динамической системы Лоренца, трактуемой как «черный ящик», воспользуемся алгоритмом ГП с основными параметрами, приведенными в таблице 1.

Табл. 1. Параметры настройки алгоритма ГП

Количество поколений	20
Размер популяции	5000
Терминальное множество	T={x, y, z, R}
Функциональное множество	F={+, -, *, /, sin, cos}
Метод инициализации	Комбинированный метод (halfandhalf)
Вероятность кроссинговера	0.7
Вероятность мутации	0.1
Фитнесс-функция	Средняя абсолютная ошибка

При этом, модель системы строится последовательно для трех уравнений относительно переменных x, y, z. Уравнения с меньшими значениями ошибки *MAE* имеют больше шансов на выживание и вхождение в новую популяцию.

В итоге, эмпирическая модель системы Лоренца имеет вид следующей системы уравнений

$$\begin{cases} \dot{x} = 10.004(y - x), \\ \dot{y} = 28.004x - xz - y, \\ \dot{z} = xy - 2.666z, \end{cases}$$

полученных с ошибками *MAE*: 0.0137707/0.0257665/0.00703833 соответственно.

Верификация полученной модели системы Лоренца представлена на рисунке 1. Здесь линия для t < 10с соответствует динамике оригинала системы, а линия для t > 10с соответствует прогнозируемой динамике, полученной на ее модели. Сравнение результатов показывает высокую степень адекватности модели системы ее оригиналу, т.е. точность идентификации модели.

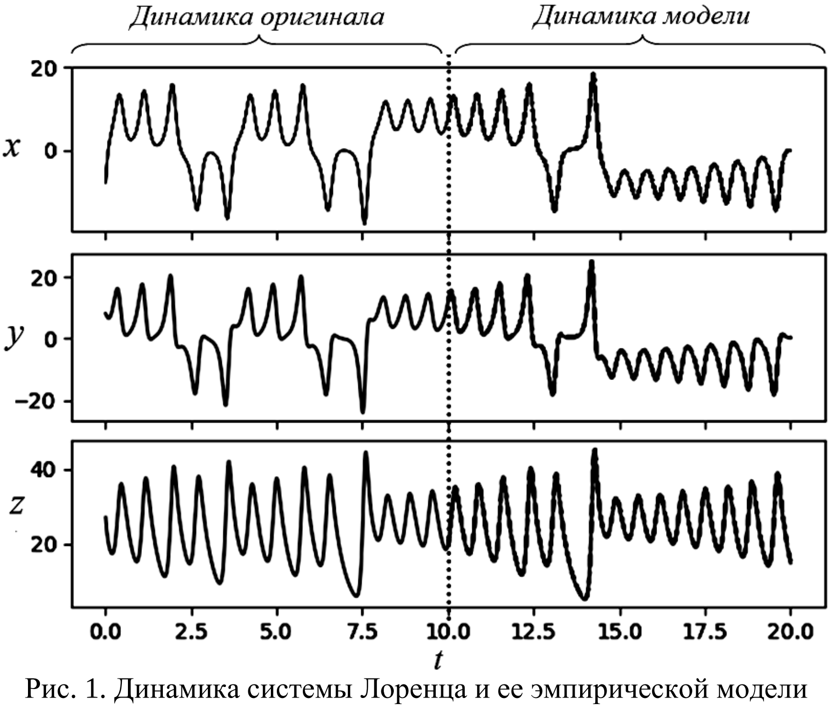

Рис. 1. Динамика системы Лоренца и ее эмпирической модели

О компромиссе между точностью и сложностью идентифицированных моделей

При решении рассматриваемой задачи идентификации динамической системы методом ГП можно получить результаты, характеризующие точность и сложность полученной математической модели в процессе эволюции. Для размера популяции, равной 1000 и 5000, они представлены на рисунке 2 и 3 соответственно. При этом оказывается справедливо известное «проклятие размерности»: по мере увеличения числа эволюций получаемая модель становится более точной и, в то же время, более сложной. Действительно, более сложные модели всегда обеспечивают более высокие точности, чем простые. Однако, при построении моделей рекомендуется придерживаться определенных принципов [13], включая, наряду с *принципом адекватности*, также и *эвристический принцип простоты*: при прочих равных условиях предпочтительнее модель минимальной сложности. На рисунке 4 представлены показатели точности и сложности модели системы Лоренца по переменной *x*. Здесь точность оценивается ошибкой модели, а сложность – обратным числом терминов в ее уравнении.

Парето-фронт точности и сложности построенной эмпирической модели системы Лоренца, представленный на рисунке 4 сплошной линией, показывает несколько простых уравнений (например, при сложности модели, равной -5), которые достаточно точно предсказывают модель системы. При этом точность модели быстро возрастает, а затем, при некоторой минимальной сложности, увеличивается лишь незначительно даже для сложных уравнений.

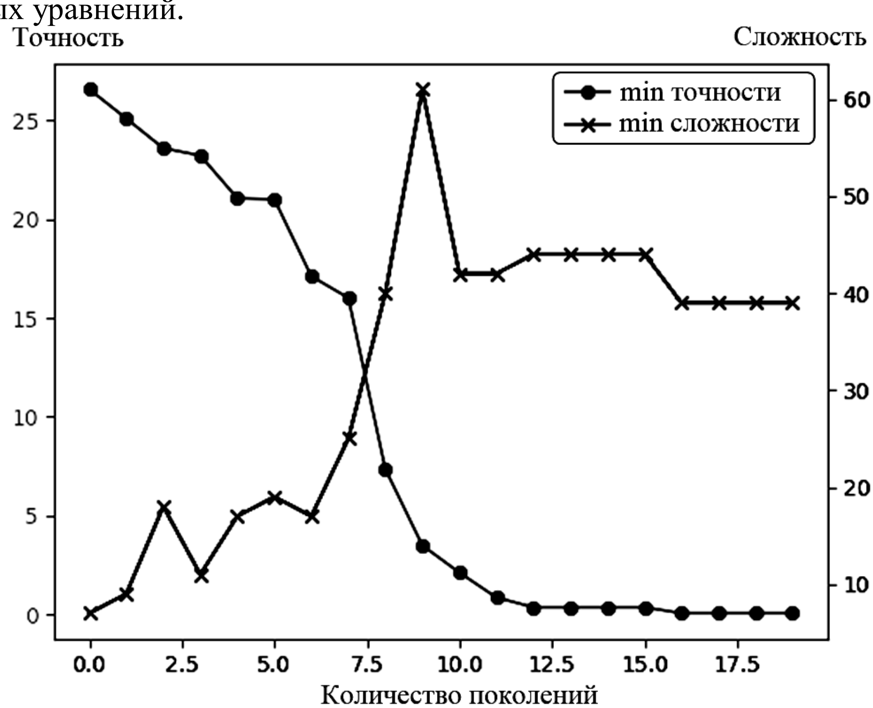

Рис. 2. Точность и сложность модели системы Лоренца, полученной во время эволюционного поиска при размере популяции, равной 1000

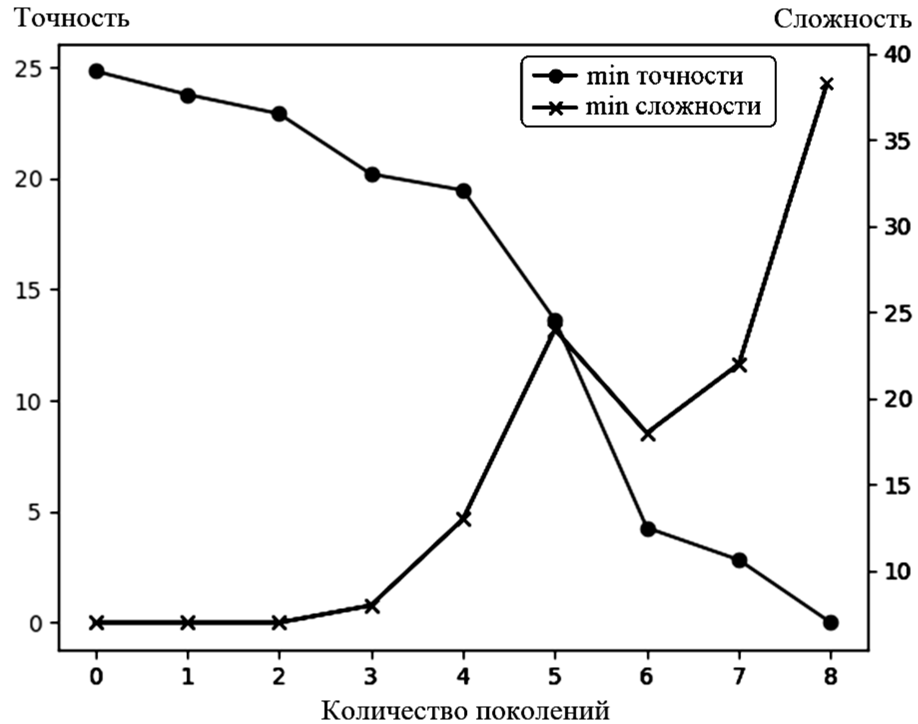

Рис. 3. Точность и сложность модели системы Лоренца, полученной во время эволюционного поиска при размере популяции, равной 5000

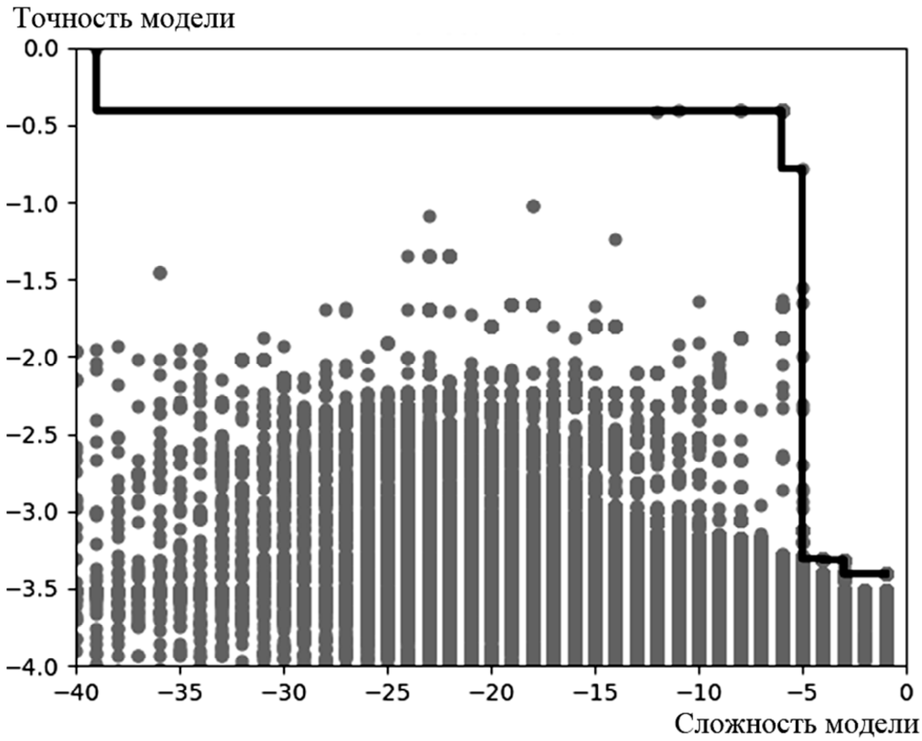

Рис. 4. Парето-фронт точности и сложности модели системы Лоренца по переменной *x*

Очевидно, что желание получить простую и одновременно точную модель, в принципе, неосуществимо, поскольку в силу векторной связи улучшение одной характеристики модели неизбежно приводит к ухудшению другой. В связи с этим все большую актуальность приобретает проблема достижения разумного компромисса, балансировки точности и сложности модели, которая является важным этапом при построении математических моделей динамических систем.

Заключение

В работе рассмотрена задача структурно-параметрической идентификации динамической системы, т.е. построения ее эмпирической математической модели методом ГП. Приведен пример построения эмпирической модели динамической системы Лоренца, иллюстрирующий эффективность метода ГП в задачах структурно-параметрической идентификации системы на основе обработки экспериментальных данных. Затронут вопрос компромисса между точностью и сложностью идентифицированной модели.

Список литературы

1. Бахтадзе Н.Н., Гинсберг К.С., Боровских Л.П. Идентификация систем на пути создания общей теории идентификации объектов управления // Проблемы управления. – 2015. – № 3. – С. 79-83.
2. Дейч А.М. Методы идентификации динамических объектов. – М.: Энергия, 1979. – 240 с.
3. Дроздов А.Л. Алгоритм идентификации характеристик динамической системы по данным наблюдений // Автоматика и телемеханика. – 2000. – № 5. – С. 58-66.
4. Фатуев В.А., Каргин А.В., Понятский В.М. Структурно-параметрическая идентификация динамических систем. – Тула: Изд-во ТулГУ, 2003. – 156 с.
5. Филимонов А.Б., Филимонов Н.Б. Структурно-параметрическая идентификация линейных динамических объектов // Мехатроника, автоматизация, управление. – 2022. – Т. 23, № 5. – С. 227-235.
6. Dang T.P., Diveev A.I., Sofronova E.A. A Problem of Identification Control Synthesis for Mobile Robot by the Network Operator Method // Proc. of the 11th IEEE Conf. on Industrial Electronics and Applications (ICIEA 2016). Hefei, China, 2016, pp. 2413-2418.
7. Губайдуллин И.М., Дивеев А.И., Константинов С.В., Софронова Е.А. Разработка кинетических моделей сложных химических реакций методом сетевого оператора // Современные проблемы науки и образования. – 2014. – № 6. – С. 157.
8. Любимова Т.В., Горелова А.В. Идентификация нечеткой системы на основе метода генетического программирования // Наука и современность. – 2015. – С. 164-169.
9. Карасева Т.С. Идентификация дифференциальных уравнений первого порядка самонастраивающимся алгоритмом генетического программирования // Актуальные проблемы авиации и космонавтики. – 2019. – Т. 2. – С. 44-46.
10. Koza J.R. Genetic Programming: On the programming of computers by means of natural selection. – The MIT Press, 1992. – 836 p.

11. Poli R., Langdon W.B., McPhee N.F. A field guide to genetic programming (with contributions by JR Koza). – GPBiB, 2008. – 252 p.
12. Курейчик В.М, Родзин С.И. Эволюционные вычисления: генетическое и эволюционное программирование // Новости искусственного интеллекта. – 2003. – № 5. – С. 13-19.
13. Клименко И.С. Методология системного исследования. – Саратов: Вузовское образование, 2020. – 273 с.

References

1. Bakhtadze N.N., Ginzburg K.S., Borovskikh L.P. Identification of systems on the way to creating a general theory of identification of control objects // Control problems. 2015, no. 3, pp. 79-83.
2. Deich A.M. Methods of identification of dynamic objects. – M.: Energy, 1979. – 240 p.
3. Drozdov A.L. Algorithm of identification of the dynamic system characteristics by observation data // Automation and Telemechanic. 2000, no. 5, pp. 58-66.
4. Fatuev V.A., Kargin A.V., Ponyatskiy V.M. Structural-parametric identification of dynamic systems: tutorial. – Tula: Publ. house of TulSU, 2003. – 156 p.
5. Filimonov A.B., Filimonov N.B. Structural parametric identification of linear dynamic object // Mekhatronika, Avtomatizatsiya, Upravlenie. 2022, vol. 23, no. 5, pp. 227-235.
6. Dang T.P., Diveev A.I., Sofronova E.A. A Problem of Identification Control Synthesis for Mobile Robot by the Network Operator Method // Proc. of the 11th IEEE Conf. on Industrial Electronics and Applications (ICIEA 2016), Hefei, China. 2016, pp. 2413-2418.
7. Gubaidullin I.M., Diveev A.I., Konstantinov S.V., Safronova E.A. Development of kinetic models of complex chemical reactions by the network operator method // Modern problems of science and education. 2014, no. 6, p. 157.
8. Lyubimova T.V., Gorelova A.V. Identification of a fuzzy system based on the method of genetic programming // Science and modernity. 2015, pp. 164-169.
9. Karaseva T.S. Identification of first-order differential equations by the self-configuring genetic programming algorithm // Actual problems of aviation and cosmonautics. 2019, vol. 2, pp. 44-46.
10. Koza J.R. Genetic Programming: On the programming of computers by means of natural selection. – The MIT Press, 1992. – 836 p.
11. Poli R., Langdon W.B., McPhee N.F. A field guide to genetic programming (with contributions by JR Koza). – GPBiB, 2008. – 252 p.
12. Kureichik V.M., Rodzin S.I. Evolutionary computing: Genetic and evolutionary programming // Artifical Intelligence News. 2003, no. 5, pp. 13-19.
13. Klimenko I.S. Methodology of system research. – Saratov: University Education, 2020. – 273 p.

Чжан Лэлэ – аспирант	**Zhang Lele** – postgraduate student
Филимонов Николай Борисович – доктор технических наук, профессор	**Filimonov Nikolay Borisovich** - – doctor of technical sciences, professor
nbfilimonov@mail.ru	

Received 21.03.2023