https://doi.org/10.26160/2474-5901-2024-44-37-45

ИДЕНТИФИКАЦИЯ ДИНАМИЧЕСКИХ СИСТЕМ НА ОСНОВЕ КОМБИНАЦИИ МЕТОДОВ ГЕНЕТИЧЕСКОГО ПРОГРАММИРОВАНИЯ И РАЗРЕЖЕННОЙ РЕГРЕССИИ

Чжан Лэ́э[1], Филимонов Н.Б.[1,2]

[1]Московский государственный технический университет им. Н.Э. Баумана, Москва, Россия;

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

Ключевые слова: динамическая система, экспериментальные данные, структурно-параметрическая идентификация, метод генетического программирования (GP), метод разреженной идентификации (SINDy), гибридный метод идентификации GP-SINDy.

Аннотация. Статья посвящена разработке и исследованию нового гибридного метода идентификации нелинейных динамических систем, основанного на комбинации метода генетического программирования (GP) и метода разреженной регрессии SINDy. Здесь метод GP обеспечивает структурную, а метод SINDy - параметрическую идентификацию системы. Эффективность гибридного метода идентификации апробирована на задаче построения математической модели динамической системы Лоренца по результатам экспериментальных наблюдений.

IDENTIFICATION OF DYNAMIC SYSTEMS BASED ON A COMBINATION OF METHODS GENETIC PROGRAMMING AND SPARSE REGRESSION

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

[1]Bauman Moscow State Technical University, Moscow, Russia;

[2]Lomonosov Moscow State University, Moscow, Russia

Keywords: dynamic system, experimental data, structural-parametric identification, genetic programming method (GP), sparse identification method (SINDy), hybrid GP-SINDy identification method.

Abstract. The article is devoted to the development and research of a new hybrid method for the identification of nonlinear dynamical systems based on a combination of the genetic programming method (GP) and the sparse regression method SINDy. Here, the GP method provides structural identification, and the SINDy method provides parametric identification of the system. The effectiveness of the hybrid identification method is tested on the task of constructing a mathematical model of the Lorenz dynamical system based on the results of experimental observations.

В современной теории и практике управления одним из основных, актуальных и активно развивающихся разделов является *идентификация систем*, связанная с построением математических моделей описания систем, адекватно отражающих основные их свойства [1]. Построение моделей систем (объектов, явлений, процессов) и их исследование является ключевой проблемой во всех областях науки и техники. При этом не всегда возможно построение теоретических моделей систем путем изучения их свойств и протекающих в них процессов. Чаще целесообразно построение не теоретических, а *эмпирических моделей систем* путем обработки результатов экспериментальных наблюдений внешних проявлений поведения системы. Данные модели систем лишены физического смысла, однако они удовлетворительно описывают экспериментальные данные, характеризующие свойства реальных систем, и являются единственно возможной альтернативой строгим теоретическим моделям.

Все большую популярность в задачах идентификации систем по экспериментальным данным находят методы и технологии эволюционных вычислений и, в частности, метод генетического программирования (Genetic Programming, GP) [2]. В настоящей работе развивается рассмотренный в работах авторов [3-5] подход к построению эмпирических математических моделей динамических систем методом GP.

Задача идентификации динамических систем по данным экспериментальных наблюдений

Считаем, что идентифицируемая динамическая система представляется «черным ящиком» (Black Box, BB), т.е. некоторой системой, о структуре и внутренних свойствах которой ничего неизвестно. Известными считаются лишь доступные для измерений в любой момент входные воздействия и реакции на них. Полагаем, что динамика системы BB описывается некоторым нелинейным дифференциальным уравнением вида

$$\dot{\boldsymbol{x}}(t) = \boldsymbol{f}(\boldsymbol{x}(t)),\ \boldsymbol{x}(t_0),$$

где $\boldsymbol{x}(t)$ – вектор состояния системы; $t > t_0$ – время, t_0 – начальный момент времени; $\boldsymbol{f}(\boldsymbol{x})$ – неизвестная нелинейная вектор-функция, определяющая динамику (уравнения движения и ограничения) системы; $\boldsymbol{x}(t_0)$ – вектор начального состояния системы.

Пусть результаты проведения экспериментального наблюдения входных переменных (начальное состояние $\boldsymbol{x}(t_0)$) и выходных переменных (текущее состояние $\boldsymbol{x}(t)$) системы BB в дискретные фиксированные моменты времени $t = t_i$, $i = \overline{1, m}$, сосредоточены в матрице $\boldsymbol{X}$:

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

Требуется по результатам наблюдения $\boldsymbol{X}$ решить задачу структурно-параметрической идентификации системы BB т.е. построить математическую модель системы в виде функции $\boldsymbol{f}(\boldsymbol{x})$, ставящей в соответствие входным переменным $\boldsymbol{x}(t_0)$ выходные переменные $\boldsymbol{x}(t)$. При этом для оценки степени адекватности модели системы ее оригиналу, т.е. близости расчетных данных $\boldsymbol{x}(t_i)$ и экспериментальных данных $\tilde{\boldsymbol{x}}(t_i)$, введем в рассмотрение ошибки идентификации

$$\boldsymbol{\varepsilon}(t) = (\varepsilon(t_1), \varepsilon(t_2), ..., \varepsilon(t_m))^{\mathrm{T}},\ \varepsilon(t_i) = \boldsymbol{x}(t_i) - \tilde{\boldsymbol{x}}(t_i),\ i = \overline{1, m},$$

и критерий точности идентификации в виде средней абсолютной ошибки:

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

Таким образом, поставленная задача идентификации системы ВВ является задачей нахождения эмпирической математической модели системы, обеспечивающей экстремум критерию точности идентификации:

$$J(\boldsymbol{\varepsilon}(t)) \to \min .$$

GP-алгоритм идентификации динамических систем

При решении задачи идентификации системы методом GP в трактовке Джона Козы (J.R. Koza) [6, 7] в качестве особей, т.е. потенциальных решений выступают компьютерные программы – иерархические композиции процедур и входных данных, отражающих состояние системы. Здесь программы представляются в виде синтаксических бинарных деревьев решений и составляются из переменных, констант и функций, причем листья дерева соответствуют терминалам, а внутренние узлы – функциям.

Метод GP позволяет решать задачу идентификации системы в символьном виде путем ее сведения к задаче символьной регрессии [3, 5], т.е. к поиску оптимальной структуры и параметров математической модели системы путем перебора различных произвольных суперпозиций функций-кандидатов (программ) из некоторого заданного набора. На рисунке 1 представлен общий вид GP-алгоритма идентификации системы «ВВ».

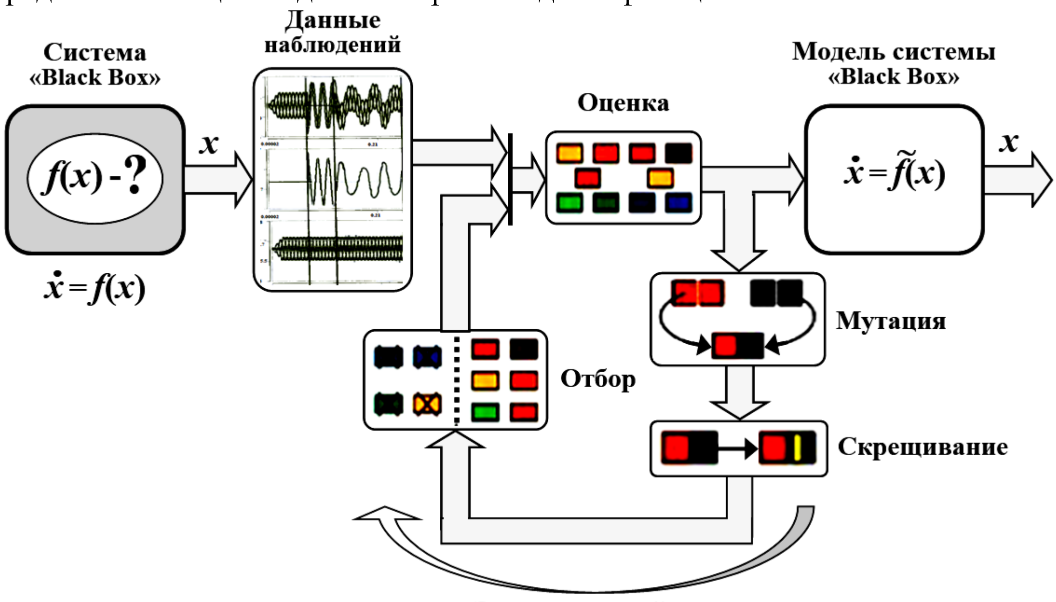

Рис. 1. Схема GP-алгоритма идентификации системы «ВВ»

Здесь реализуются традиционные шаги Дарвиновского эволюционного цикла: генерация начальной популяции особей – случайных потенциальных решений; оценка приспособленности решений по значению функции приспособленности (критерия $J(\boldsymbol{\varepsilon}(t))$; формирование новой популяции, более приспособленной, чем предыдущая, путем выбора наилучших решений и применения генетических операторов селекции, скрещивания и мутации; проверка условий останова эволюционного поиска и принятие решения об окончании, либо о повторении итерации поиска. Итогом данного эволюционного поиска является математическая формула в виде оптимальной оценки $\tilde{f}(x)$ искомой функции $f(x)$, т.е. эмпирическая математическая модель идентифицированной системы «ВВ».

Несмотря на общепризнанную высокую результативность метода GP в задачах идентификации динамических систем, он часто оказывается эффективным при нахождении лишь структуры, но не параметров искомой модели системы. Причина данного недостатка метода GP заложена, по-видимому, в самой природе алгоритма эволюционного поиска параметров модели системы, основанного на простой комбинации случайно сгенерированных констант. Возникает естественный вопрос о возможности модификации метода GP для повышения его эффективности при параметрической идентификации систем. Для этого предлагается комбинация метода GP с известным методом разреженной идентификации SINDy (Sparse Identification of Nonlinear Dynamics), предложенным в работе [8].

SINDy-алгоритм идентификации динамических систем

В последние годы для моделирования динамических систем по результатам экспериментальных наблюдений все большую популярность приобретает метод SINDy [9, 10, п.7.3], основанный на использовании разреженной линейной регрессии на разреженной динамике моделируемой системы. Данный метод был предложен с целью идентификации нелинейных динамических систем в предположении, что уравнения системы содержат сравнительно небольшое число важных, доминирующих членов, которые и определяют ее динамику. Действительно, для многих нелинейных систем искомая нелинейная функция $f(x)$, определяющая ее динамику, содержит всего несколько ненулевых членов в пространстве возможных функций. Так, например, в уравнениях системы Лоренца фигурируют лишь линейные и квадратичные члены.

Метод SINDy в задачах идентификации динамических систем заключается в решении оптимизационной задачи разреженной регрессии на библиотеке возможных базовых нелинейных функций-кандидатов путем исключения терминов, коэффициенты которых меньше порогового значения, определяющего качество и точность идентификации модели системы. Основное достоинство метода SINDy – способность идентифицировать экономные модели динамических систем, содержащие только самые важные члены, необходимые для объяснения наблюдаемого поведения системы, что позволяет избежать практически нереализуемый комбинаторный перебор всевозможных вариантов моделей-кандидатов системы. При этом выбор набора потенциальных функций-кандидатов является произвольным со стремлением к разреженной динамике системы и тестированию на тестовой выборке данных с учетом особенностей уже известных моделей динамических систем в исследуемой области.

Общий вид SIBDy-алгоритма идентификации системы «ВВ» на примере хаотической системы Лоренца представлен на рисунке 2.

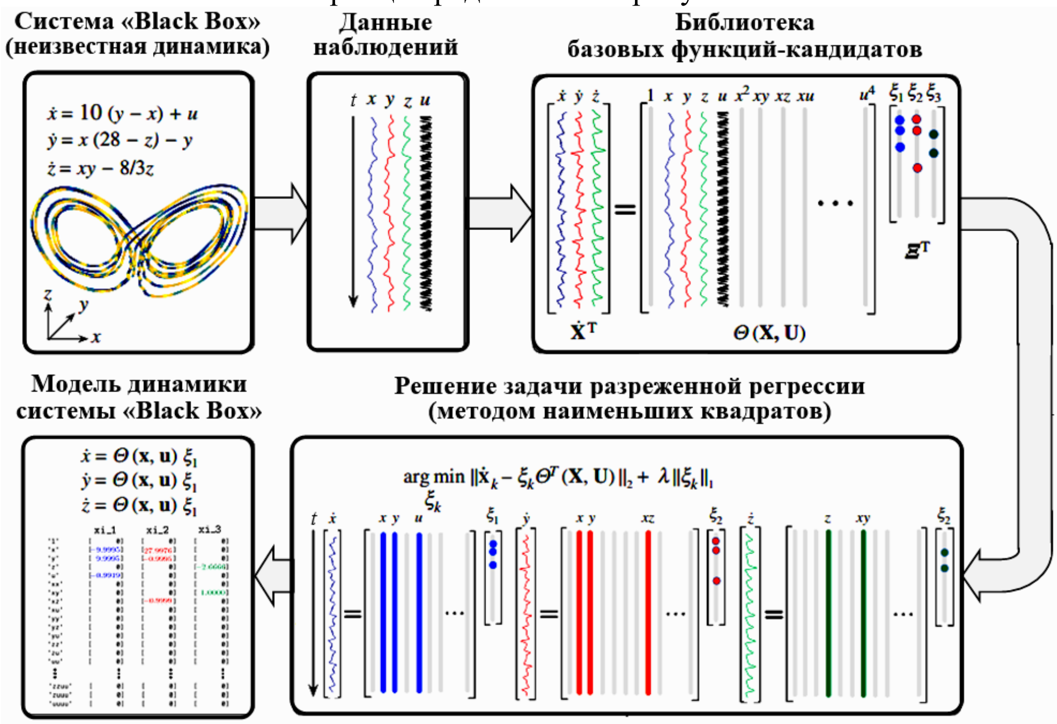

Рис. 2. Схема SINDy-алгоритма идентификации системы «ВВ»

В данном алгоритме ключевыми являются следующие алгоритмические модули:

– модуль «differentiate», выполняющий численное дифференцирование матрицы данных $\boldsymbol{X}$ для получения матрицы $\dot{\boldsymbol{X}}$;

– модуль «feature_library», формирующий библиотеку потенциальных функций-кандидатов $\boldsymbol{\Theta}(\boldsymbol{X})$;

– модуль «optimizers» предоставляющий набор разреженных оптимальных регрессионных решателей для определения вектора активных параметров-кандидатов $\boldsymbol{\Xi} = [\xi_1\ \xi_2 \ldots, \xi_p]$, соответствующих наименьшему количеству библиотечных функций.

Выходные данные SINDy-алгоритма представляют собой эмпирическую математическую модель системы, которая является аппроксимацией обобщенной линеаризованной матричной модели вида

$$\dot{\boldsymbol{X}} \approx \boldsymbol{\Theta}(\boldsymbol{X}))\boldsymbol{\Xi}\,,\ \dot{\boldsymbol{x}} = \boldsymbol{f}(\boldsymbol{x}) \approx \sum_{k=1}^{p} \theta_k(\boldsymbol{x})\xi_k\,,$$

наилучшим образом описывающей измеренные данные и отражающей динамическое поведение системы.

GP-SINDy-подход к идентификации динамических систем

Метод GP характеризуется высокой эффективностью при построении математической модели системы в широком смысле, т.е. в задачах структурной идентификации, а метод SINDy характеризуется высокой эффективностью при построении математической модели системы в узком смысле, т.е. в задачах параметрической идентификации. В связи с этим для повышения эффективности структурно-параметрической идентификации динамических систем по экспериментальным данным на основе метода GP предлагается его модификация путем интеграции с методом SINDy. Данная интеграция, комбинируя методы GP и SINDy, формирует новый гибридный подход к идентификации динамической системы «ВВ», именуемый методом GP-SINDy, схема которого представлена на рисунке 3.

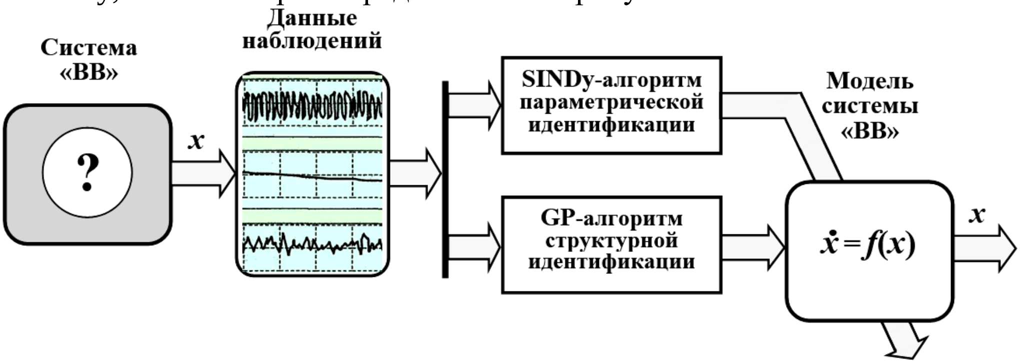

Рис. 3. Схема гибридного метода GP-SINDy идентификации динамической системы «ВВ»

Алгоритм предложенного гибридного метода GP-SINDy структурно-параметрической идентификации динамической системы «ВВ» содержит GP-алгоритм, обеспечивающий нахождение эволюционным генетическим поиском структуры системы и SINDy-алгоритм, обеспечивающий нахождение разреженной оптимальной регрессией параметров системы.

Для иллюстрации эффективности предлагаемого гибридного метода GP-SINDy идентификации динамических систем рассмотрим задачу построения по результатам экспериментальных наблюдений математической модели системы Лоренца (E. Lorenz), описываемой системой трех нелинейных дифференциальных уравнений первого порядка:

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

Рассматривая данную систему «черным ящиком», была решена задача ее идентификации методами GP и GP-SINDy по результатам наблюдения вектора состояния $\boldsymbol{X} = (x, y, z)^{\mathrm{T}}$ и его скорости $\dot{\boldsymbol{X}} = (\dot{x}, \dot{y}, \dot{z})^{\mathrm{T}}$. При этом модель системы Лоренца строилась последовательно, отдельно для каждой из трех ее подсистем первого порядка, описываемых дифференциальными уравнениями относительно трех переменных x, y, z.

Результаты сравнительного анализа точности идентификации системы обоими методами GP и GP-SINDy, характеризующей степень адекватности идентифицированной модели системы ее оригиналу, представлены на рисунке 4 и в таблице 1.

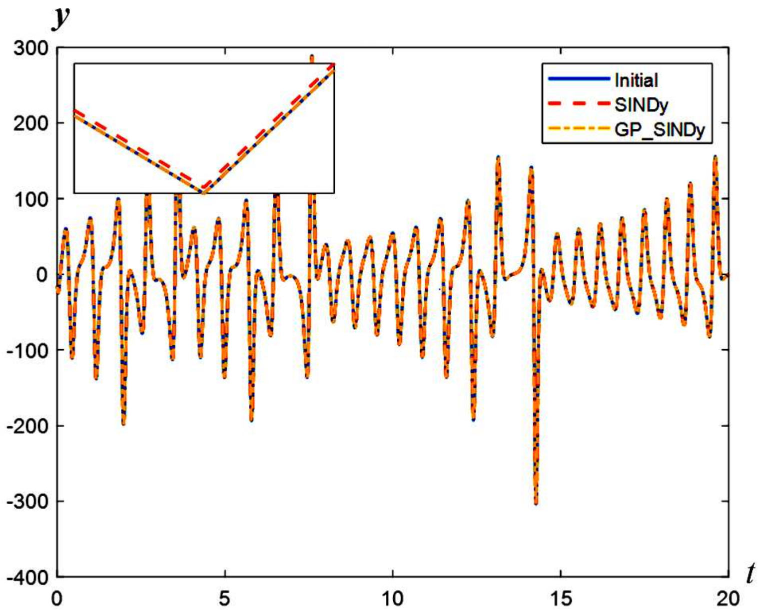

Рис. 4. Динамика второй подсистемы системы Лоренса относительно переменной y и ее модели, полученной методами GP и GP-SINDy

Табл. 1. Сравнительный анализ точности идентификации модели системы Лоренса методами GP и GP-SINDy

Идентификация подсистем системы Лоренса	Точность идентификации методом GP	Точность идентификации методом GP-SINDy
$\dot{x} = \ldots$	0,4211	0
$\dot{y} = \ldots$	0,07978	0
$\dot{z} = \ldots$	2,2990	0,7948

Сравнительный анализ результатов построения эмпирической математической модели системы Лоренса показал, что гибридный метод GP-SINDy идентификации значительно превосходит по точности метод GP идентификации.

Список литературы

1. Isermann R., Munchhof M. Identification of Dynamic Systems: An Introduction with Applications. – Springer-Verlag Berlin, 2011. – 705 p.

2. Sette S., Boullart L. Genetic Programming: Principles and Applications // Engineering Applications of Artificial Intelligence. 2001, vol. 14, iss. 6, pp. 727-736.
3. Чжан Л., Филимонов Н.Б. Идентификация динамических систем на основе обработки экспериментальных данных методом генетического программирования // Journal of Advanced Research in Natural Science. – 2023. – № 18. – С. 4-12.
4. Чжан Л., Филимонов Н.Б. Разработка и исследование алгоритма генетического программирования для структурно-параметрической идентификации динамических систем // Высокопроизводительные вычислительные системы и технологии. – 2024. – Т. 8, № 1. – С. 196-201.
5. Чжан Л. Анализ компромисса между точностью и сложностью идентифицированных моделей динамических систем // Системы анализа и обработки данных. – 2024. – №2(94). – С. 85-93.
6. Patelli A. Genetic Programming Techniques for Nonlinear System Identification. – Editura Politehnium, 2011. – 243 p.
7. Szeifer F., Abonyi J. Genetic Programming for the Identification of Nonlinear Input-Output Models // Industrial & Engineering Chemistry Research. 2005, vol. 44, no. 9, pp. 3178-3186.
8. Brunton S.L., Proctor J.L., Kutz J.N. Discovering Governing Equations from Data by Sparse Identification of Nonlinear Dynamical Systems // Proceedings of the national academy of sciences. 2016, vol. 113, no.15, pp. 3932-3937.
9. de Silva B.M., Champion K., Quade M., Jean-Christophe Loiseau J-Ch., Kutz J.N., Brunton S.L. PySINDy: A Python Package for the Sparse Identification of Nonlinear Dynamics from Data // The Journal of Open Source Software. 2020, no. 5(49), pp. 1-4.
10. Брантон С.Л., Куц Дж.Н. Анализ данных в науке и технике. – М.: ДМК Пресс, 2021. – 574 с.

References

1. Isermann R., Munchhof M. Identification of Dynamic Systems: An Introduction with Applications. – Springer-Verlag Berlin, 2011. – 705 p.
2. Sette S., Boullart L. Genetic Programming: Principles and Applications // Engineering Applications of Artificial Intelligence. 2001, vol. 14, iss. 6, pp. 727-736.
3. Zhang L., Filimonov N.B. Identification of Dynamic Systems based on the Processing of Experimental Data by the Method of Genetic Programming // Journal of Advanced Research in Natural Science. 2023, iss. 18, pp. 4-12.
4. Zhang L., Filimonov N.B. Development and Research of Genetic Programming Algorithm for Structural-Parametric Identification of Dynamic Systems // High-performance computing systems and technologies. 2024, vol. 8, no. 1, pp. 196-201.
5. Zhang L. Analysis of the Trade-Off between Accuracy and Complexity of Identified Models of Dynamic Systems // Analysis and Data Processing Systems. 2024, no 2(94), pp. 85-93.
6. Patelli A. Genetic Programming Techniques for Nonlinear System Identification. – Editura Politehnium, 2011. – 243 p.
7. Szeifer F., Abonyi J. Genetic Programming for the Identification of Nonlinear Input-Output Models // Industrial & Engineering Chemistry Research. 2005, vol. 44, no. 9, pp. 3178-3186.

8. Brunton S.L., Proctor J.L., Kutz J.N. Discovering Governing Equations from Data by Sparse Identification of Nonlinear Dynamical Systems // Proceedings of the national academy of sciences. 2016, vol. 113, no.15, pp. 3932-3937.
9. de Silva B.M., Champion K., Quade M., Jean-Christophe Loiseau J-Ch., Kutz J.N., Brunton S.L. PySINDy: A Python Package for the Sparse Identification of Nonlinear Dynamics from Data // The Journal of Open Source Software. 2020, no. 5(49), pp. 1-4.
10. Brunton S.L., Kutz J.N. Data-Driven Science and Engineering. – Cambridge: Cambridge University Press, 2022. – 614p.

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 31.10.2024