

ИНФОРМАЦИОННЫЕ
ТЕХНОЛОГИИ
И ТЕЛЕКОММУНИКАЦИИ

INFORMATION
TECHNOLOGIES
AND TELECOMMUNICATIONS

УДК 004.94

DOI: 10.17212/2782-2001-2025-2-111-122

Гибридный метод структурно-параметрической идентификации динамических систем^{*}

Л. ЧЖАН

105005, РФ, г. Москва, ул. 2-я Бауманская, 5, Московский государственный техни-
ческий университет

chzhan12@student.bmstu.ru

В различных сферах научно-исследовательской и производственной деятельности весьма актуальна задача структурно-параметрической идентификации динамических систем. Эта задача заключается в построении математической модели системы по экспериментальным данным наблюдений за ее поведением. Полученная модель используется как для исследования свойств системы, так и для управления системой. Эволюционные алгоритмы (особенно генетическое программирование) находят большую популярность при решении задачи структурно-параметрической идентификации. Результат идентификации методом генетического программирования представляется в символьном виде, что облегчает последующий анализ и управление. Однако по своей сути эволюционные алгоритмы являются стохастическими алгоритмами. Это означает, что они опираются на случайные процессы, что часто приводит к получению субоптимального решения. Повышение эффективности идентификации динамических систем методом генетического программирования необходимо. В настоящей статье представлен метод GP – SINDy (генетическое программирование с разреженной идентификацией, genetic programming with sparse identification) – новый гибридный метод идентификации динамических систем с помощью генетического программирования и разреженной идентификации. В предложенном методе процесс идентификации разделен на два этапа: сначала генетическое программирование применяется для определения структуры модели, а затем разреженная идентификация используется для определения соответствующих оптимальных параметров. Модель в форме дифференциальных уравнений построена на основе обработки наблюдений. Эффективность предложенного метода продемонстрирована на идентификации трех динамических систем. Результаты показывают, что метод GP – SINDy позволяет находить модели с высокой точностью и низкой сложностью (более высокой интерпретируемостью), что обуславливает его перспективность для структурно-параметрической идентификации динамических систем.

Ключевые слова: задача идентификации, построение математической модели, символьная регрессия, структурно-параметрическая идентификация, повышение эффективности идентификации, генетическое программирование, разреженная идентификация нелинейных динамик, гибридный метод идентификации GP – SINDy

^{*} Статья получена 13 января 2025 г.

ВВЕДЕНИЕ

Потребность в точных и интерпретируемых моделях в науке и технике стремительно растет в связи с очень широким классом динамики, которая исследуется в области современного управления [1–3]. Для решения подобных проблем может применяться структурно-параметрическая идентификация. Сутью структурно-параметрической идентификации является получение модели по результатам наблюдений поведения системы с целью изучения ее свойств и протекающих в ней процессов [4].

При решении задачи структурно-параметрической идентификации большую популярность находит метод генетического программирования (GP, Genetic Programming), основанный на дарвинской теории биологической эволюции [5]. Этот метод позволяет одновременно находить оптимальную структуру и параметры модели путем эволюционных вычислений. Результаты идентификации представляются в виде уравнений в символьном виде, удобном для дальнейшего использования [6].

Эффективность применения GP определяется случайным механизмом эволюционных вычислений, который часто приводит к неоптимальным параметрам модели системы [7]. Кроме того, при решении задачи идентификации динамических систем, при сложной их модели алгоритм GP часто заканчивается в окрестности найденного локального оптимума, трудно приближается к глобальному оптимуму, что не позволяет получить точную модель системы [8]. Иными словами, для сложных задач идентификации алгоритм GP позволяет находить решение субоптимальное, но нахождение точного решения может оказаться трудной задачей.

В связи с указанными недостатками применения GP весьма актуальным является разработка подхода к повышению его эффективности при решении задач идентификации математической модели системы по результатам наблюдений ее поведения. Такая возможность может быть обеспечена путем использования алгоритма GP совместно с другими алгоритмами оптимизации [9, 10].

В настоящей работе предлагается гибридный метод структурно-параметрической идентификации математических моделей системы по результатам наблюдений, который основан на комбинации метода GP и метода разреженной идентификации нелинейных динамик (SINDy, sparse identification of nonlinear dynamics). Следует отметить, что идентифицированная модель системы играет ключевую роль в раскрытии закономерностей ее поведения. Статья организована следующим образом: в первом разделе ставится и формулируется постановка задач идентификации, во втором разделе описываются алгоритм и схема предложенного гибридного метода идентификации, в третьем разделе представлены результаты компьютерного моделирования гибридного метода идентификации.

1. ПОСТАНОВКА ЗАДАЧИ ИДЕНТИФИКАЦИИ

Задача идентификации связана с проблемой построения математических моделей на основе наблюдений за поведением динамических систем [11]. Следует выделить четыре основных компонента процесса идентификации, который ведет от результатов наблюдений к искомой идентифицированной

модели: данные наблюдений, набор моделей-кандидатов, критерий соответствия и проверка результата идентификации [12]. В связи с этим задача идентификации динамических систем на основе этих наблюдений состоит из следующих этапов: сбор данных, построение модели системы и оценка ее адекватности. Следует отметить, в настоящей работе задача сбора данных не рассматривается.

Как упоминалось ранее, задачу идентификации динамических систем можно сформулировать следующим образом. Предполагается, что динамика системы в переменных состояния описывается следующим дифференциальным уравнением:

$$\frac{d}{dt}\mathbf{x}(t) = \mathbf{f}(\mathbf{x}(t)), \quad (1)$$

где $\mathbf{x}(t) = [x_1(t), x_2(t), \dots, x_n(t)]$ – вектор состояния системы; t – время; $\mathbf{f}(\mathbf{x}(t))$ – неизвестная функция системы.

Известны результаты наблюдений поведения системы, состоящие из вектора состояния $\mathbf{x}(t)$ в моменты времени $t = t_1, t_2, \dots, t_m$ в виде матрицы $\mathbf{X} = (\mathbf{x}(t_1), \mathbf{x}(t_2), \dots, \mathbf{x}(t_m))^T$. Требуется по результатам наблюдений $\mathbf{X}$ идентифицировать математическую модель системы в виде уравнения (1), т. е. найти неизвестную функцию системы $\mathbf{f}(\mathbf{x}(t))$. При этом для оценки степени адекватности идентифицированной модели исходной системе можно применять различные подходы (например, среднеквадратическую ошибку, среднюю абсолютную ошибку и др.).

В настоящей работе структурно-параметрическая идентификация применяется для построения модели исследуемой системы. Она позволяет автоматически определять как структуру модели, так и ее параметры. Полученная модель представлена в виде дифференциальных уравнений в символьной форме. Такая характеристика делает ее особенно эффективной в случаях, когда соотношения, описывающие систему, неизвестны или сложно формализуемы.

2. МЕТОДОЛОГИЯ GP – SINDy-ИДЕНТИФИКАЦИИ

Как отмечено во введении, несмотря на эффективность GP при решении задачи идентификации, подход имеет два ограничения: 1) неоптимальные параметры и 2) сходимость к локальному оптимуму [7, 8]. В настоящей работе разработан двухэтапный гибридный метод GP – SINDy для структурно-параметрической идентификации математической модели, представленной в виде дифференциальных уравнений, по данным измерений исследуемой системы. Метод GP применяется для извлечения структурных кандидатов модели на начальном этапе, затем используется метод SINDy для определения параметров модели. Таким образом, задачу можно рассматривать как структурную часть, решаемую с помощью GP, и часть, связанную с оптимизацией параметров, решаемую с помощью SINDy.

Метод GP представляет собой метод символьной регрессии с помощью эволюционных вычислений при решении задачи идентификации и основан

на дарвинском естественном отборе и выживании наиболее приспособленных [13, 14]. Он работает с популяцией особей (часто представленных в виде деревьев), где каждая является одним возможным решением задачи. Все возможные решения (особи) составляют набор / пространство решений. Этот метод можно рассматривать как метод параллельного перебора огромного количества возможностей для поиска лучших решений, только в данном случае генетические операции обеспечивают эволюцию в направлении оптимума.

На рис. 1 представлена общая схема GP. В основном алгоритм состоит из следующих шагов: генерация начальной популяции особей (дерево 1, ..., N , N – размер популяции) случайным образом, оценка их приспособленности по заданной функции, проверка условий завершения, выбор наилучших решений и создание нового поколения особей генетическими операциями из них. Поскольку в настоящей статье сам метод GP не является центром исследования, подобную информацию о GP можно найти в литературе [13, 14], настройка параметров описана в нашей предыдущей работе [6].

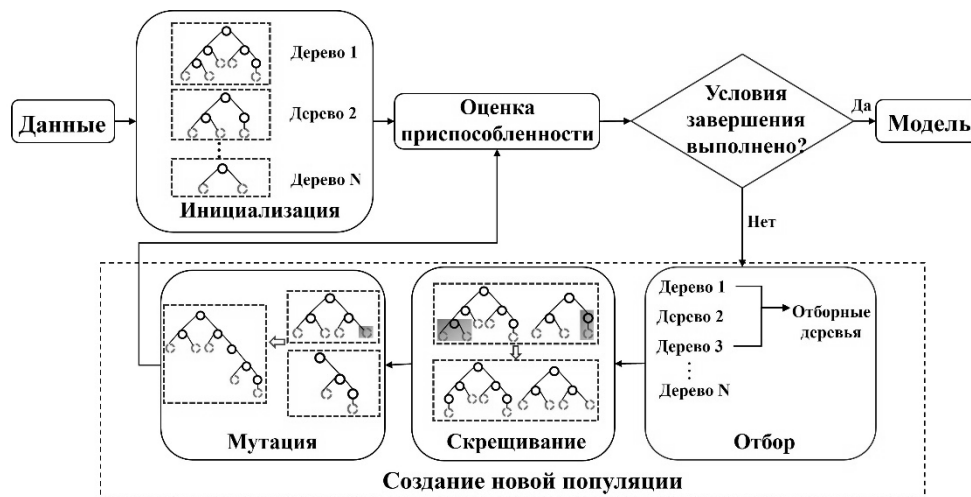


Рис. 1. Схема алгоритма GP. Каждое дерево состоит из функций (части схемы, выделенные сплошной линией) и терминалов (части схемы, выделенные пунктирной линией)

Fig. 1. The GP algorithm diagram. Each tree consists of functions (solid parts) and terminals (dashed parts)

При использовании метода GP для решения задачи идентификации наблюдается следующая тенденция: на начальном этапе с увеличением числа эволюций получаемая модель становится всё более точной и одновременно более сложной. Однако после достижения определенного порога сложности дальнейшее усложнение модели не приводит к значимому повышению точности [15]. Это связано с тем, что при достижении высокой точности алгоритм предпочитает добавлять множество более мелких членов для получения более высокой точности. А причина неточности в значительной степени зависит от случайно сгенерированных констант в процессе эволюционных вычислений [7]. В связи с этим в настоящей работе GP используется только для поиска релевантных компонентов из большой библиотеки кандидатов, т. е. структуры модели.

Метод SINDy – это регрессионный метод для извлечения «скупой» динамики из данных временных рядов, модель которых является разреженной [16, 17]. Использование метода SINDy в задаче идентификации основано на предположении, что многие системы имеют относительно небольшое количество активных членов в управляющих уравнениях.

Для того чтобы определить неизвестную модель динамической системы, данные состояний системы $\mathbf{x} = [x_1 \ x_2 \ \dots \ x_n]^T$ и их производных $\dot{\mathbf{x}}$ собираются в матрицу данных по времени:

$$\mathbf{X} = \begin{bmatrix} \mathbf{x}^T(t_1) \\ \mathbf{x}^T(t_2) \\ \vdots \\ \mathbf{x}^T(t_m) \end{bmatrix} = \begin{bmatrix} x_1(t_1) & x_2(t_1) & \dots & x_n(t_1) \\ x_1(t_2) & x_2(t_2) & \dots & x_n(t_2) \\ \vdots & \vdots & \ddots & \vdots \\ x_1(t_m) & x_2(t_m) & \dots & x_n(t_m) \end{bmatrix},$$

$$\dot{\mathbf{X}} = \begin{bmatrix} \dot{\mathbf{x}}^T(t_1) \\ \dot{\mathbf{x}}^T(t_2) \\ \vdots \\ \dot{\mathbf{x}}^T(t_m) \end{bmatrix} = \begin{bmatrix} \dot{x}_1(t_1) & \dot{x}_2(t_1) & \dots & \dot{x}_n(t_1) \\ \dot{x}_1(t_2) & \dot{x}_2(t_2) & \dots & \dot{x}_n(t_2) \\ \vdots & \vdots & \ddots & \vdots \\ \dot{x}_1(t_m) & \dot{x}_2(t_m) & \dots & \dot{x}_n(t_m) \end{bmatrix},$$

где $\mathbf{X}, \dot{\mathbf{X}} \in \mathbb{R}^{m \times n}$.

Дальше библиотека p функций-кандидатов построена из данных $\mathbf{X}$:

$$\boldsymbol{\Theta}(\mathbf{X}) = [\boldsymbol{\theta}_1(\mathbf{X}) \ \dots \ \boldsymbol{\theta}_p(\mathbf{X})], \quad (2)$$

где $\boldsymbol{\Theta}(\mathbf{X}) \in \mathbb{R}^{m \times p}$, каждый $\boldsymbol{\theta}_j(\mathbf{X})$ есть функция-кандидат, $\boldsymbol{\Theta}(\mathbf{X})$ может состоять из постоянных, полиномов, тригонометрических функций и т. д.

В результате идентифицированная динамическая система может быть представлена в виде

$$\dot{\mathbf{X}} = \boldsymbol{\Theta}(\mathbf{X})\boldsymbol{\Xi},$$

где $\boldsymbol{\Xi} = [\xi_1 \ \xi_2 \ \dots \ \xi_n] \in \mathbb{R}^{p \times n}$ – матрица коэффициентов.

Затем модель построена путем нахождения матрицы коэффициентов $\boldsymbol{\Xi}$ с использованием разреженной регрессии [15]. Схема алгоритма SINDy представлена на рис. 2.

Модель, полученная с помощью алгоритма SINDy, является разреженной, универсальной и символьной, что делает ее пригодной для дальнейшего использования в задачах управления. Однако построение таких моделей во многом определяется заранее заданной библиотекой функций-кандидатов (2),

причем при небольшом размере библиотеки метод SINDy не сможет обеспечить получение эффективной модели для систем.

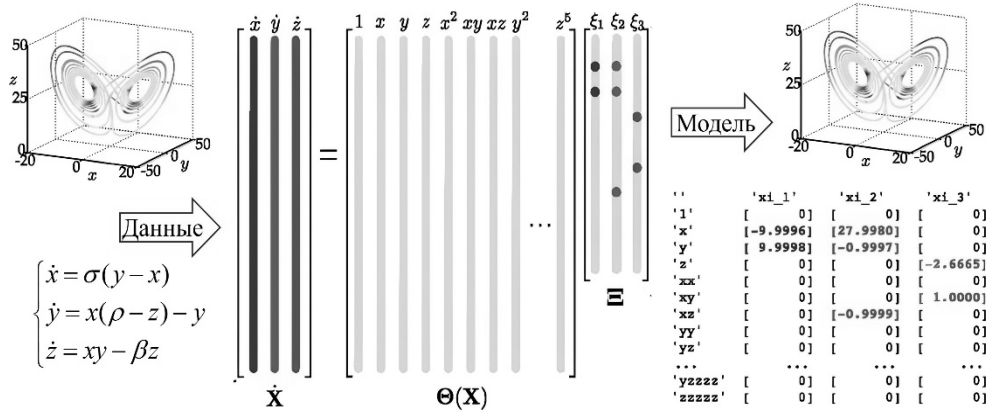


Рис. 2. Схема алгоритма SINDy

Fig. 2. The SINDy algorithm diagram

Метод GP-SINDy. В статье предлагается двухэтапный метод для структурно-параметрической идентификации математических моделей в виде дифференциальных уравнений по наблюдаемым данным для динамической системы. В предлагаемом методе GP применяется для определения структуры модели, а для параметров используется SINDy, соответственно этот метод называется GP – SINDy. Алгоритм процедуры GP – SINDy показан на рис. 3.



Рис. 3. Алгоритм гибридного метода GP – SINDy

Fig. 3. The hybrid GP – SINDy method algorithm

Сначала проводится процесс идентификации с помощью GP, в результате получается предварительная модель. В настоящей работе структура модели получается из частей предварительной модели, в будущем будем обсуждать другой подход. Затем с помощью SINDy проводится разреженная регрессия полученной структуры и получаются соответствующие параметры модели. Структура и параметры вместе образуют полную модель.

Важно подчеркнуть, что модели, полученные из наблюдений, отражают лишь взаимосвязи между переменными, но при определенных условиях (модель несложная и с четкой структурой) они помогут исследователям выявить свойство системы и процессов.

При использовании предложенного подхода GP – SINDy для решения задачи идентификации полученные модели, структурные уравнения которых

линейны по параметрам и содержат нелинейные функции. Таким образом, уравнение (1) переформировано:

$$\frac{d}{dt} \mathbf{x}(t) = \mathbf{f}(\mathbf{x}(t)) = \sum_{i=1}^r c_i F_i(\mathbf{x}(t)),$$

где $F_i(\mathbf{x}(t))$ – функции, включающие основные компоненты (линейные и нелинейные функции-кандидаты) структуры модели, сгенерированные методом GP; c_i – соответствующие параметры функций, полученные из матрицы коэффициентов методом SINDy; r – количество функциональных компонентов модели.

3. РЕЗУЛЬТАТ МОДЕЛИРОВАНИЯ

Эффективность предложенного метода продемонстрирована на примерах трех динамических систем: системы Лоренца, математического маятника и системы Ван дер Поля. Результаты тестирования $\mathbf{X}$ и $\hat{\mathbf{X}}$ были получены путем численного моделирования методом Рунге – Кутты в Matlab с использованием уравнений из таблицы (система-оригинал). Данные разделены на две части: первая половина используется для идентификации модели, вторая – для верификации с целью проверки ее эффективности (пример представлен на рис. 4). В качестве критерия оценки эффективности идентифицированной модели была использована средняя абсолютная ошибка: чем меньше ее значение, тем эффективнее модель. Алгоритм метода реализован в среде Python с помощью системы эволюционных вычислений DEAP. В целом, в работе используется алгоритм GP, предложенный в работе [14], и настройка параметров, как в [6].

Результаты тестирования подходов GP и GP – SINDy на трех динамических системах (система Лоренца, маятник и система Ван дер Поля)

Results of testing GP and GP – SINDy approaches on three dynamic systems (Lorenz, pendulum and Van der Pol)

Система Лоренца	Маятник	Система Ван дер Поля
Системы-оригиналы:		
$\begin{cases} \dot{x} = \sigma(y - x) \\ \dot{y} = x(\rho - z) - y \\ \dot{z} = xy - \beta z \end{cases}$ $\sigma = 10, \rho = 28, \beta = 8/3$	$\begin{cases} \dot{\theta} = w \\ \dot{w} = -\frac{g}{l} \sin(\theta) \end{cases}$ $g = 9.8, l = 1$	$\begin{cases} \dot{x} = y \\ \dot{y} = -ay(x^2 - 1) - x \end{cases}$ $a = 5$
Идентифицированные GP:		
$\begin{cases} \dot{x} = 10.129(y - x) \\ \dot{y} = 27.988x - xz - y \\ \dot{z} = xy - 3z + 8.151 \end{cases}$	$\begin{cases} \dot{\theta} = w \\ \dot{w} = -9.735 \sin(\theta) \end{cases}$	$\begin{cases} \dot{x} = y \\ \dot{y} = -x + 4y - 13/3 yx^2 \end{cases}$

Окончание таблицы

End of the Table

Система Лоренца	Маятник	Система Ван дер Поля
Идентифицированные GP-SINDy:		
$\begin{cases} \dot{x} = 10(y - x) \\ \dot{y} = 28x - y - xz \\ \dot{z} = xy - 2.6667z \end{cases}$	$\begin{cases} \dot{\theta} = w \\ \dot{w} = -9.8000 \sin(\theta) \end{cases}$	$\begin{cases} \dot{x} = y \\ \dot{y} = -x + 5y - 5yx^2 \end{cases}$
Точности:		
GP	0.4211/0.0798/2.2990	0/0.0469
GP – SINDy	0/0.7948	0/0

Важно подчеркнуть, что при идентификации динамика данной системы полностью неизвестна, доступны только ее результаты наблюдений. В итоге получается модель системы в виде дифференциальных уравнений. В таблице представлены результаты идентификации трех систем методами GP и GP – SINDy. Результаты представлены в виде уравнений модели и соответствующих значений ошибок.

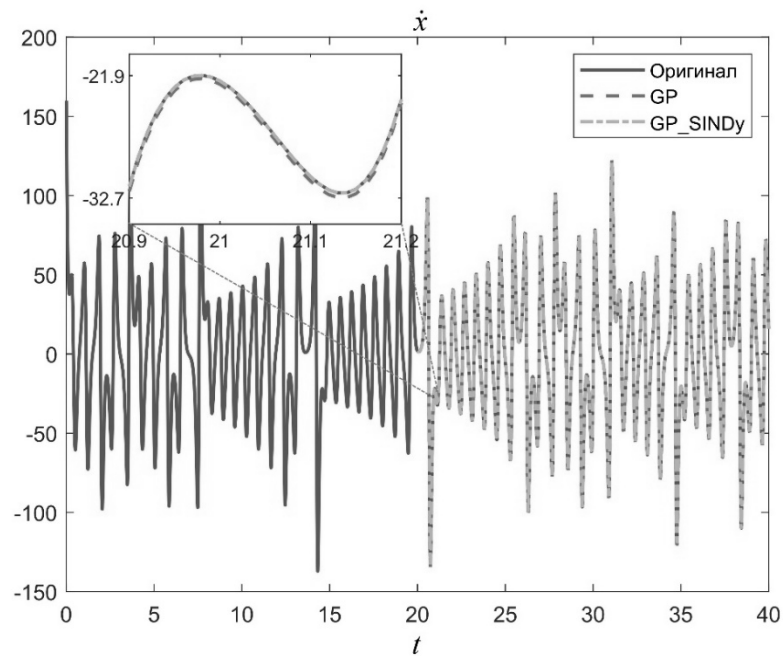


Рис. 4. Сравнение результатов идентификации переменной x системы Лоренца методами GP и GP – SINDy

Fig. 4. Comparison of the identification results of the variable x of the Lorenz system by GP and GP – SINDy methods

На основе полученных результатов можно сделать вывод, что модели, полученные методом GP – SINDy, действительно способны отражать динамику систем (модели обладают высокой точностью и правильной структурой). Кроме того, по сравнению с традиционным GP предложенный GP – SINDy более эффективен при решении задачи структурно-параметрической идентификации динамических систем.

ЗАКЛЮЧЕНИЕ

В настоящей статье предлагается новый метод GP – SINDy, предназначенный для идентификации математической модели по данным наблюдений. В предлагаемом методе процесс идентификации разделяется на две части: структурную и параметрическую, которые решаются с помощью методов GP и SINDy соответственно. В результате получается модель в дифференциальной форме. Эффективность предложенного метода подтверждена тестированием на ряде динамических систем. Результаты демонстрируют, что GP – SINDy является перспективным методом для идентификации интерпретируемых и точных моделей изучаемых систем.

Следует отметить, что, несмотря на тестирование метода исключительно на экспериментальных данных, теоретическая база и полученные результаты позволяют предположить его эффективность для реальных данных. Помимо простого построения модели с высокой точностью и низкой сложностью, долгосрочной целью настоящей работы является использование полученной информации для выявления ключевых закономерностей системы и управления ею.

СПИСОК ЛИТЕРАТУРЫ

1. Rudin C. Stop explaining black box machine learning models for high stakes decisions and use interpretable models instead // *Nature Machine Intelligence*. – 2019. – Vol. 1 (5). – P. 206–215.
2. Lipton Z.C. The mythos of model interpretability: in machine learning, the concept of interpretability is both important and slippery // *Queue*. – 2018. – Vol. 16 (3). – P. 31–57.
3. A survey of methods for explaining black box models / R. Guidotti, A. Monreale, S. Ruggieri, F. Turini, F. Giannotti, D. Pedreschi // *ACM Computing Surveys (CSUR)*. – 2018. – Vol. 51 (5). – Art. 93. – P. 1–42. – DOI: 10.1145/3236009.
4. Карасева Т.С. Эволюционные алгоритмы решения задач символьной регрессии для идентификации динамических систем: дис. ... канд. техн. наук: 2.3.1. – Красноярск, 2023. – 128 с.
5. Дивеев А.И., Шмалько Е.Ю. Классические методы символьной регрессии для поиска структур математических выражений (обзор) // *Вопросы теории безопасности и устойчивости систем*. – 2018. – № 20. – С. 100–132.
6. 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.
7. Iba H., Sakamoto E. Inference of differential equation models by genetic programming // *Information Sciences*. – 2008. – Vol. 178 (23). – P. 4453–4468.
8. The convergence mechanism and strategies for non-elitist genetic programming / H. Ni, F. Zeng, B. Yu, F. Sun // *Applied Mechanics and Materials*. – 2013. – Vol. 347–350. – P. 3850–3860. – DOI: 10.4028/www.scientific.net/amm.347-350.3850.
9. Qian L., Wang H., Dougherty E.R. Inference of noisy nonlinear differential equation models for gene regulatory networks using genetic programming and Kalman filtering // *IEEE Transactions on Signal Processing*. – 2008. – Vol. 56 (7). – P. 3327–3339.

10. Icke I., Bongard J.C. Improving genetic programming based symbolic regression using deterministic machine learning // IEEE Congress on Evolutionary Computation (CEC), Cancun, Mexico. – IEEE, 2013. – P. 1763–1770. – DOI: 10.1109/CEC.2013.6557774.
11. Филимонов А.Б. Филимонов Н.Б. Структурно-параметрическая идентификация линейных динамических объектов // Мехатроника, автоматизация, управление. – 2022. – Т. 23, № 5. – С. 227–235. – DOI: 10.17587/mau.23.227-235.
12. Ljung L. System identification: theory for the user. – Upper Saddle River, NJ: Prentice Hall PTR, 1999. – 540 p.
13. Koza J.R. Genetic programming: on the programming of computers by means of natural selection. – The MIT Press, 1992. – 836 p.
14. Poli R., Langdon W.B., McPhee N.F. A field guide to genetic programming / with contributions by J.R. Koza. – GPBiB, 2008. – 252 p.
15. Чжан Л. Анализ компромисса между точностью и сложностью идентифицированных моделей динамических систем // Системы анализа и обработки данных. – 2024. – № 2 (94). – С. 85–93. – DOI: 10.17212/2782-2001-2024-2-85-93.
16. Brunton S., Proctor J., Kutz J. Discovering governing equations from data by sparse identification of nonlinear dynamical systems // Proceedings of the National Academy of Sciences. – 2016. – Vol. 113 (15). – P. 3932–3937.
17. Brunton S., Kutz J. Data-driven science and engineering: machine learning, dynamical systems, and control. – Cambridge: Cambridge University Press, 2019. – 495 p.

Чжан Лэлэ, аспирант кафедры системы автоматического управления МГТУ им. Баумана. Получила степень магистра в области системы автоматического управления в МГТУ им. Баумана в 2021 г. Научный интерес – математическое моделирование, идентификация динамических систем методом эволюционных алгоритмов, компромисс между точностью и сложностью идентифицированных моделей. E-mail: chzhanl2@student.bmstu.ru

Zhang Lele, a PhD student in the Bauman Moscow State Technical University (BMSTU), the department of automatic control systems. She graduated from BMSTU and obtained her master degree in 2021. Her research interests include mathematical modeling, identification of dynamical systems by evolutionary algorithms, and trade-off between accuracy and complexity of identified models. E-mail: chzhanl2@student.bmstu.ru

DOI: 10.17212/2782-2001-2025-2-111-122

A hybrid method for structural-parametric identification of dynamic systems*

L. ZHANG

Bauman Moscow State Technical University, 5, 2-ya Baumanskaya Street, Moscow, 105005, Russian Federation

chzhanl2@student.bmstu.ru

Abstract

In various fields of research and production activities the problem of structural-parametric identification of dynamic systems is very important. This task consists in constructing a mathematical model of the system based on the experimental data of observation of its behavior. The obtained model can be used both for investigation of system properties and for system control. Evolutionary algorithms, especially genetic programming, have found great popularity

* Received 13 January 2025.

in building such models. The result of identification by genetic programming is represented in a symbolic form, which facilitates subsequent analysis and control. However, evolutionary algorithm are inherently stochastic algorithms, which means that they rely on random processes, which often leads to a suboptimal solution. Improving the efficiency of identification of dynamic systems by the genetic programming method is necessary. This paper presents GP – SINDy (genetic programming with sparse identification), a new hybrid method for identification of dynamic systems using genetic programming and sparse identification. In the proposed method, the identification process is divided into two stages: first, genetic programming is applied to determine the model structure, and then sparse identification is used to determine the corresponding optimal parameters. A model in the form of differential equations is constructed on the basis of observations. The effectiveness of the proposed method is demonstrated on the identification of three dynamic systems. The results show that the GP – SINDy allows finding models with a high accuracy and a low complexity (higher interpretability), which makes it promising for structural-parametric identification of dynamical systems.

Keywords: identification problem, mathematical model construction, symbolic regression, structural-parametric identification, identification efficiency improvement, genetic programming, sparse identification of nonlinear dynamics, hybrid method of identification of GP – SINDy

REFERENCES

1. Rudin C. Stop explaining black box machine learning models for high stakes decisions and use interpretable models instead. *Nature Machine Intelligence*, 2019, vol. 1 (5), pp. 206–215.
2. Lipton Z.C. The mythos of model interpretability: in machine learning, the concept of interpretability is both important and slippery. *Queue*, 2018, vol. 16 (3), pp. 31–57.
3. Guidotti R., Monreale A., Ruggieri S., Turini F., Giannotti F., Pedreschi D. A survey of methods for explaining black box models. *ACM Computing Surveys (CSUR)*, 2018, vol. 51 (5), art. 93, pp. 1–42. DOI: 10.1145/3236009.
4. Karaseva T.S. *Evolutsionnye algoritmy resheniya zadach simvol'noi regressii dlya identifikatsii dinamicheskikh system*. Diss. kand. tekhn. nauk [Evolutionary algorithm for solving symbolic regression problems for identification of dynamic systems. PhD eng. sci. diss.]. Krasnoyarsk, 2023. 128 p.
5. Diveev A.I., Shmalko E.Yu. Klassicheskie metody simvol'noi regressii dlya poiska struktur matematicheskikh vyrazhenii (obzor) [Modern methods of symbolic regression and their modifications (review)]. *Voprosy teorii bezopasnosti i ustoichivosti system*, 2018, no. 20, pp. 133–158. (In Russian).
6. Zhang L., Filimonov N.B. Razrabotka i issledovanie algoritma geneticheskogo programmirovaniya dlya strukturno-parametricheskoi identifikatsii dinamicheskikh sistem [Development and research of genetic programming algorithm for structural-parametric identification of dynamic system]. *Vysokoproizvoditel'nye vychislitel'nye sistemy i tekhnologii = High-Performance Computing Systems and Technologies*, 2024, vol. 8, no. 1, pp. 196–201.
7. Iba H., Sakamoto E. Inference of differential equation models by genetic programming. *Information Sciences*, 2008, vol. 178 (23), pp. 4453–4468.
8. Ni H., Zeng F., Yu B., Sun F. The convergence mechanism and strategies for non-elitist genetic programming. *Applied Mechanics and Materials*, 2013, vol. 347–350, pp. 3850–3860. DOI: 10.4028/www.scientific.net/amm.347-350.3850.
9. Qian L., Wang H., Dougherty E.R. Inference of noisy nonlinear differential equation models for gene regulatory networks using genetic programming and Kalman filtering. *IEEE Transactions on Signal Processing*, 2008, vol. 56 (7), pp. 3327–3339.
10. Icke I., Bongard J.C. Improving genetic programming based symbolic regression using deterministic machine learning. *IEEE Congress on Evolutionary Computation (CEC)*, Cancun, Mexico, 2013, pp. 1763–1770. DOI: 10.1109/CEC.2013.6557774.

11. Filimonov A.B. Filimonov N.B. Strukturno-parametricheskaya identifikatsiya lineinykh dinamicheskikh ob"ektov [Structural-parametric identification of linear dynamic objects]. *Mekhatronika, avtomatizatsiya, upravlenie* = *Mechatronics, Automation, Control*, 2022, vol. 23, no. 5, pp. 227–235. DOI: 10.17587/mau.23.227-235.
12. Ljung L. *System identification: theory for the user*. Upper Saddle River, NJ, Prentice Hall PTR, 1999. 540 p.
13. Koza J.R. *Genetic programming: on the programming of computers by means of natural selection*. The MIT Press, 1992. 836 p.
14. Poli R., Langdon W.B., McPhee N.F. *A field guide to genetic programming*. With contributions by J.R. Koza. GPBiB, 2008. 252 p.
15. Zhang L. Analiz kompromissa mezhdu tochnost'yu i slozhnost'yu identifikatsionnykh modelей dinamicheskikh sistem [Analysis of the trade-off between accuracy and complexity of identified models of dynamic systems]. *Sistemy analiza i obrabotki dannykh* = *Analysis and Data Processing Systems*, 2024, no. 2 (94), pp. 85–93. DOI: 10.17212/2782-2001-2024-2-85-93.
16. Brunton S., Proctor J., Kutz J. Discovering governing equations from data by sparse identification of nonlinear dynamical systems. *Proceedings of the National Academy of Sciences*, 2016, vol. 113 (15), pp. 3932–3937.
17. Brunton S., Kutz J. *Data-driven science and engineering: machine learning, dynamical systems, and control*. Cambridge, Cambridge University Press, 2019. 495 p.

Для цитирования:

Чжан Л. Гибридный метод структурно-параметрической идентификации динамических систем // Системы анализа и обработки данных. – 2025. – № 2 (98). – С. 111–122. – DOI: 10.17212/2782-2001-2025-2-111-122.

For citation:

Zhang L. Gibrinyi metod strukturno-parametricheskoi identifikatsii dinamicheskikh sistem [A hybrid method for structural-parametric identification of dynamic systems]. *Sistemy analiza i obrabotki dannykh* = *Analysis and Data Processing Systems*, 2025, no. 2 (98), pp. 111–122. DOI: 10.17212/2782-2001-2025-2-111-122.