

USING GENETIC PROGRAMMING TECHNIQUES FOR INERTIA-FREE SYSTEM IDENTIFICATION TASKS

E. D. Mihov

Siberian Federal University
79, Svobodny Av., Krasnoyarsk, 660041, Russian Federation
E-mail: edmihov@mail.ru

The problem of identification of inertia-free objects is being investigated. The overall pattern of the process investigated is being described. As a research object, a stochastic inertia-free process of modeling has been chosen. A feature of the process under consideration is the fact that unmanaged but controlled variables influence the process under investigation. In addition, the process under investigation is affected by unmanaged and uncontrolled variables.

The levels of priori information has been briefly considered and characterized. For each level of priori information the identification method has been described. Particular attention is paid to the levels of priori information under which the identification task in a “broad” sense needs to be addressed.

As a method of identification, genetic programming is considered. Method of genetic programming has been chosen as a research object since this method is more commonly used in the identification problem. Despite the frequency with which this method is used, it is interesting to look at the results of this method under different conditions. As changing conditions, the object’s complexity and the change in the volume of the training sample were used.

For the identification process, objects with different structures were selected. The dependence of the time of finding the structure of the object on the size of the training sample was investigated. As shown by studies, there is no clear correlation between the time of finding the structure of the object and the size of the training sample.

In addition, relationship between the time of the structure and the “complexity” of the object was investigated. As a criterion of “complexity” of the object, the number of input variables was taken. The study showed correlation between some values; with the increase in the number of input variables, the time of finding the structure of the process also increased.

Keywords: inertia-free object, genetic programming, identification.

ИСПОЛЬЗОВАНИЕ МЕТОДОВ ГЕНЕТИЧЕСКОГО ПРОГРАММИРОВАНИЯ ДЛЯ ЗАДАЧ ИДЕНТИФИКАЦИИ БЕЗЫНЕРЦИОННЫХ СИСТЕМ

Е. Д. Михов

Сибирский федеральный университет
Российская Федерация, 660041, г. Красноярск, просп. Свободный, 79
E-mail: edmihov@mail.ru

Исследуется проблема идентификации безынерционных объектов. Описана общая схема исследуемого процесса. В качестве объекта исследования был взят процесс моделирования стохастическим безынерционным процессом. Особенностью рассматриваемого процесса является тот факт, что на исследуемый процесс имеют влияние неуправляемые, но контролируемые переменные. Также на исследуемый процесс влияют неуправляемые и неконтролируемые переменные.

Кратко были рассмотрены и охарактеризованы уровни априорной информации. Для каждого уровня априорной информации описан способ идентификации. Особое внимание уделено уровням априорной информации, при которых необходимо решать задачу идентификации в широком смысле.

В качестве метода идентификации рассмотрено генетическое программирование. Метод генетического программирования был выбран в качестве объекта исследования по причине того, что данный метод все чаще используется в задаче идентификации. Несмотря на частоту использования данного метода, интересно посмотреть на результаты работы данного метода в различных условиях. В качестве меняющихся условий было усложнение объекта и изменение объема обучающей выборки.

Для процесса идентификации подбирались объекты с различной структурой. Исследовалась зависимость времени нахождения структуры объекта от размера обучающей выборки. Как показали исследования, явной зависимости между временем нахождения структуры объекта и размером обучающей выборки не наблюдается.

Помимо этого, была исследована зависимость между временем нахождения структуры от сложности объекта. В качестве критерия сложности объекта было взято количество входных переменных. Исследование показало, что между данными величинами есть связь: при увеличении количества входных переменных увеличивалось и время нахождения структуры исследуемого процесса.

Ключевые слова: безынерционный объект, генетическое программирование, идентификация.

Introduction. The identification of many stochastic objects is often limited to the identification of static systems. The most common pattern of the discrete-continuous process is shown in fig. 1 [1–4].

A is unknown object's operator, $\overline{x(t)}$ is the vector of the object's output variables; $\overline{u(t)}$ is the vector of the object's input variables; $\overline{\mu(t)}$ is the vector of the object's controlled but unmeasurable input variables; $\overline{\lambda(t)}$ is the vector of the object's and controlled but unmeasurable input variables; $\xi(t)$ is the random action, $\omega^i(t): i = 1, 2, \dots, k$ are the variables in the process; (t) is the continuous time; $H^\mu, H^u, H^x, H^z, H^q, H^\omega$ measuring devices; $\mu_t, u_t, x_t, \omega_t$ are measures of $\mu(t), u(t), x(t), \omega(t); h^\mu(t), h^u(t), h^x(t), h^\omega(t)$ are the random effects.

Levels of a priori information. Let's consider the different levels of priori information [5–7]:

1. Systems with parametric uncertainty. Parametric level of priori information assumes that there is a parametric structure with unknown parameters. Iterative probabilistic procedures are used for parameter estimation. The problem of identification in the narrow sense is solved in this case;

2. Systems with nonparametric uncertainty. A nonparametric level of priori information does not have the parametric structure, but it has some qualitative information about the process. For example it has uniqueness, or ambiguity of its characteristics, linearity for dynamic processes or the nature of its nonlinearity. Nonparametric statistics methods are used to solve problems of identification at this level (it is identification in a "broad" sense);

3. Systems with parametric and nonparametric uncertainty. System that does not correspond to any types described above is very important from the practical point of view. For example, for individual characteristics of a multiply connected process, on the basis of physico-chemical laws, energy laws, the law of mass conservation, balance relationships, parametric patterns can be derived. However, for others it is not possible.

Thus, we are in a situation where the task of identification is formulated in terms of both parametric and nonparametric priori information. Then the models are an interconnected system of parametric and non-parametric ratios.

Genetic programming. Genetic programming algorithm is based on the same optimization algorithm [8–13].

This algorithm is applied when parametric structure of process is not known, but there is a sample where $(\overline{u_i}, x_i), i = \overline{1...n}, \overline{u_i}$ is the vector of the object's input

variables, x_i is the object's output variables, n is number of sample units.

Genetic programming reconstructs parametric structure, describing the process based on the available sample.

Since genetic programming is based on the optimization method (genetic algorithm), the process of finding the most optimal structure is iterative. The criterions of optimality are mean error, mean-square error, mean relative error and other characteristics that reflect structure accurateness.

The process of genetic programming is divided into the following stages [14; 15]:

- 1) the first genetic trees are created randomly;
- 2) receiving descendants from trees, obtaining descendants, can be performed using the following actions:
 - Replication;
 - Reproduction;
 - Mutation;
- 3) the trees that least satisfy the conditions of the optimum are destroyed;
- 4) verification of compliance with the conditions.

It is noticeable that the genetic programming process is very similar to genetic algorithms.

Let's describe central concepts of the genetic programming:

The concept of "genetic tree" is used as a description of the structure of an object in genetic programming. A genetic tree is a connected graph, where nodes represent actions, and finite elements (leaves) are variables (fig. 2).

We suggest the genetic tree of the following expression, as an example:

$$f(x_1, x_2) = x_1 * (x_2 - 7) * (2/\sin(x_1)).$$

The restoration of the structure from the represented tree begins with the leaves.

It is also important to describe the terms replication, reproduction and mutation.

Replication is the action towards the tree that results in an exact copy of the parent tree (fig. 3).

Reproduction is receiving descendants by replacing the branches of parent trees with each other (fig. 4).

Mutation is receiving a descendant by changing the parent tree branch to randomly created one (fig. 5).

The last step in genetic programming is verification of result.

If the verification confirms to the conditions, then the researcher uses the resulting tree for object modeling.

If the conditions have not been achieved, the genetic programming process is repeated, beginning with step 2.

Computational experiment. Object modeling process is demonstrated in different conditions. In the process of computational experiments, the relationship between the dimension of the process under investigation and the speed of finding the parametric structure using genetic programming will be demonstrated.

For genetic programming, the following settings have been introduced:

- the number of trees = 10;
- prediction: average forecast error < 0.1;
- the maximum recovery time = 240 second.

The maximum time given for recovery is necessary to accelerate the experiment.

The structure will be restored for the following functions: $f(x) = x_1 + x_2 + 5$, $x_1 \in (0,3)$, $x_2 \in (0,3)$. We will check the dependence of the time of finding the structure from the value of the training sample for this example.

The dependence of the time of finding the structure from the value of the training sample example is not observed (fig. 6). This can be explained by the fact that

this algorithm more depends on “successfulness” of the original tree choice.

It is worth mentioning that during experiments the genetic algorithm was completely restored, resulting in: $f(x) = (x_2 + 5) + x_1$, where the restored structure does not copy the original, for example: $f(x) = x_1 + x_1 + \cos(x_1) + x_2 + 3 + \sin(x_2)$, however, when $x_1 \in (0,3)$, $x_2 \in (0,3)$, structure is close to the original.

The structure will be restored for the following functions: $f(x) = \sin(x_1)*x_1 + 6*\cos(x_2)$, $x_1 \in (0,3)$, $x_2 \in (0,3)$. We will check the dependence of the time of finding the structure on the value of the training sample for this example.

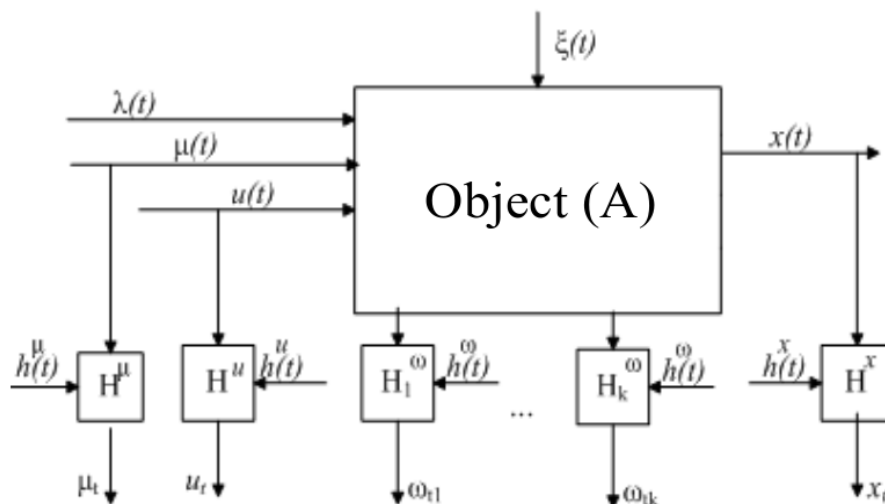


Fig. 1. Common pattern of the discrete-continuous process

Рис. 1. Общая схема исследуемого процесса

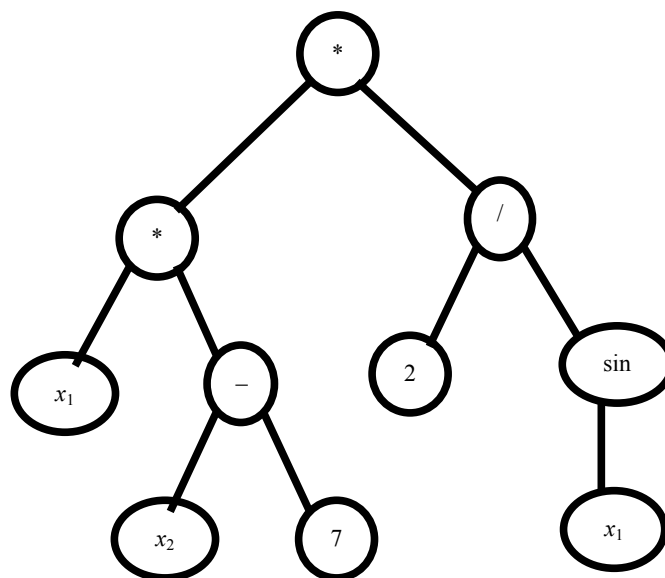


Fig. 2. Genetic tree example

Рис. 2. Пример генетического дерева

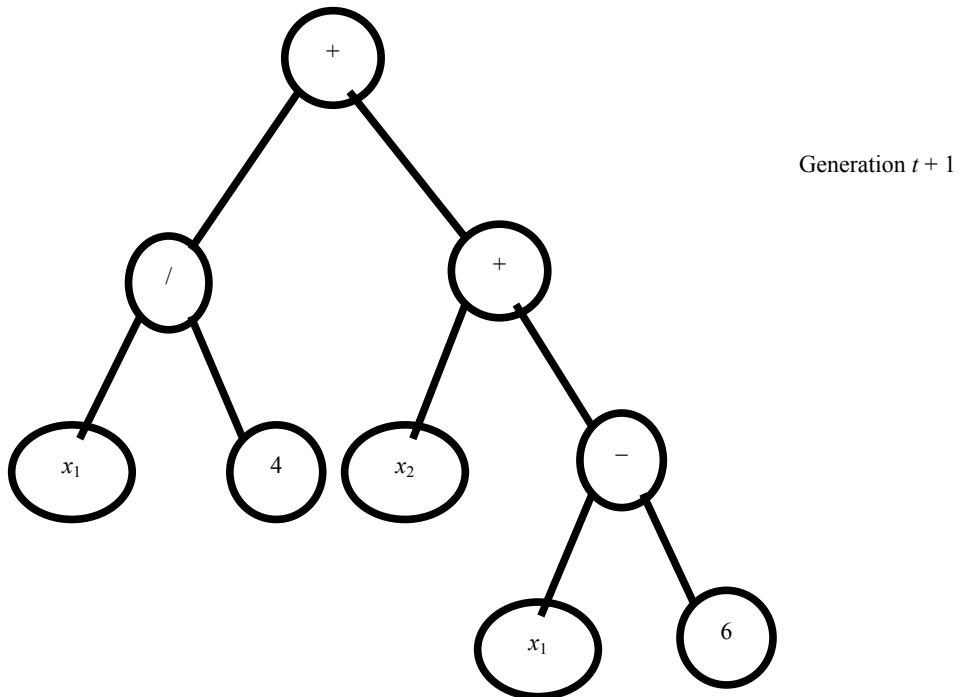
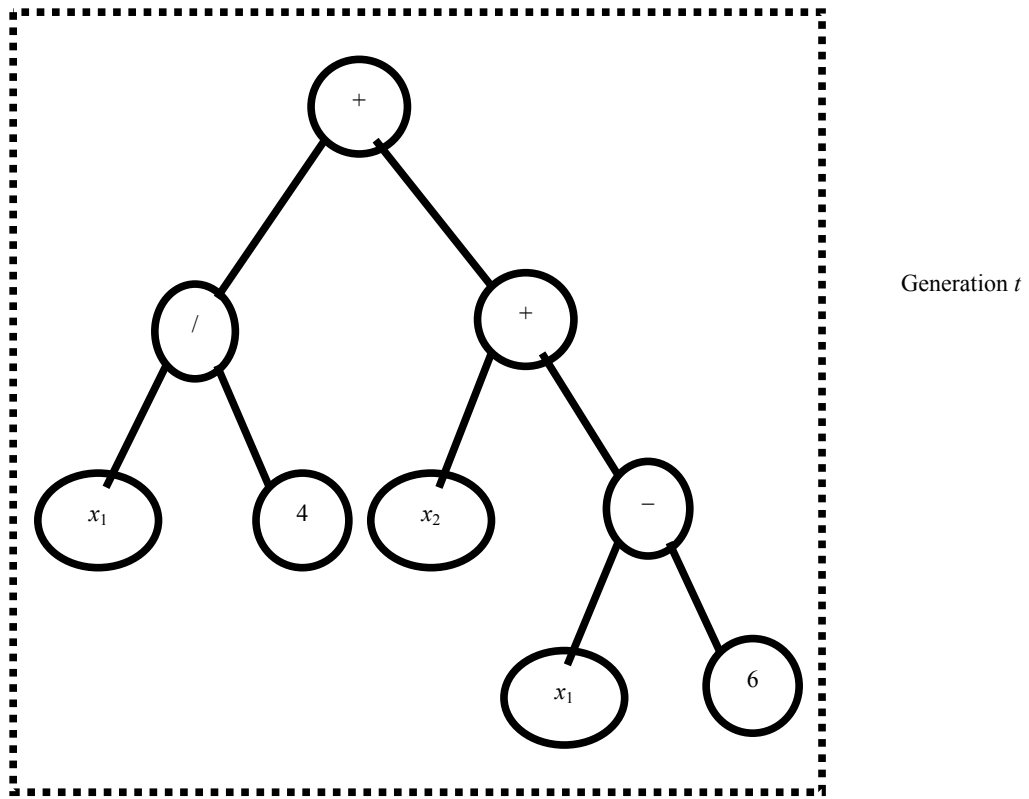


Fig. 3. Replication's example

Рис. 3. Пример репликации

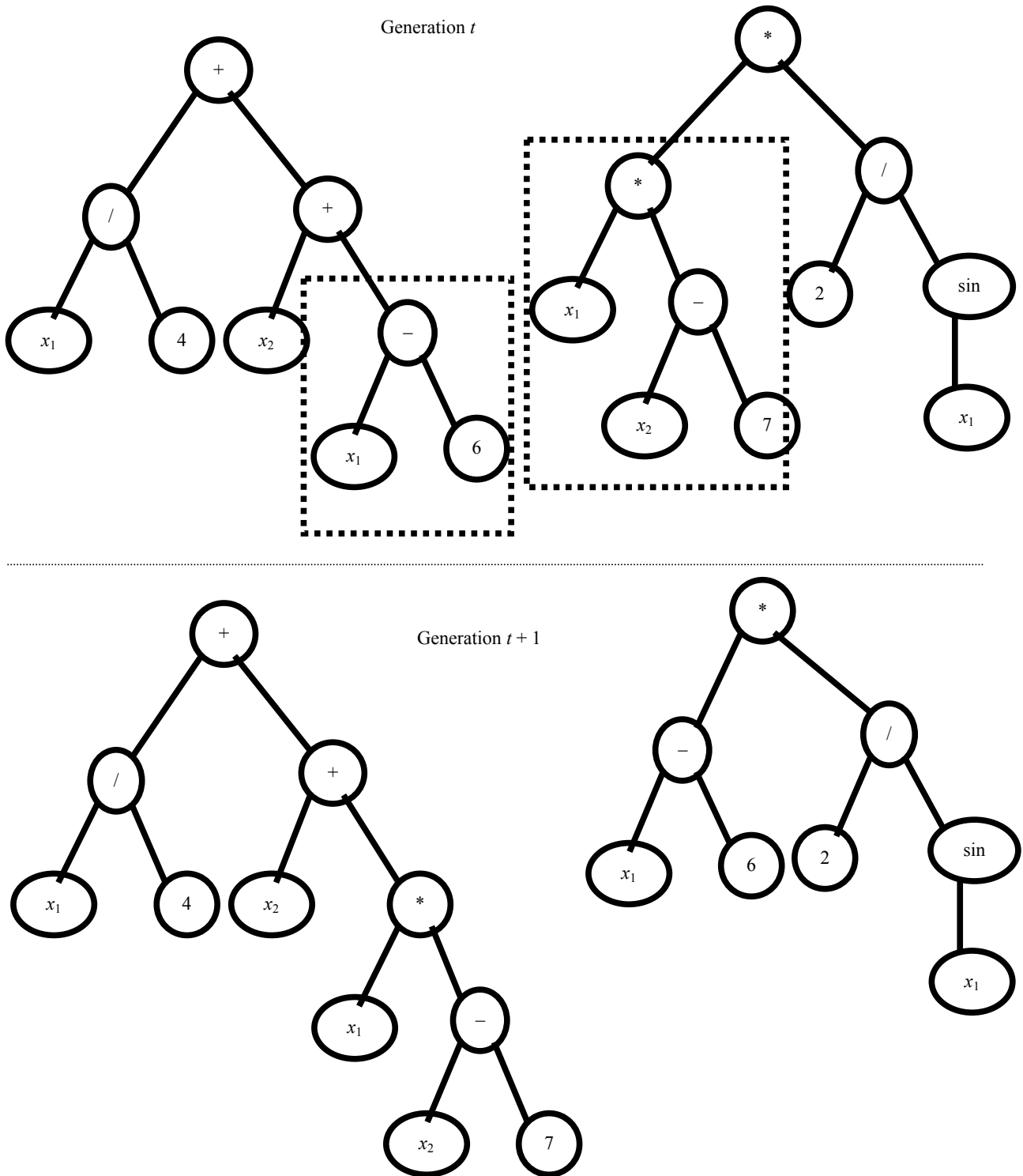


Fig. 4. Reproduction's example

Рис. 4. Пример размножения

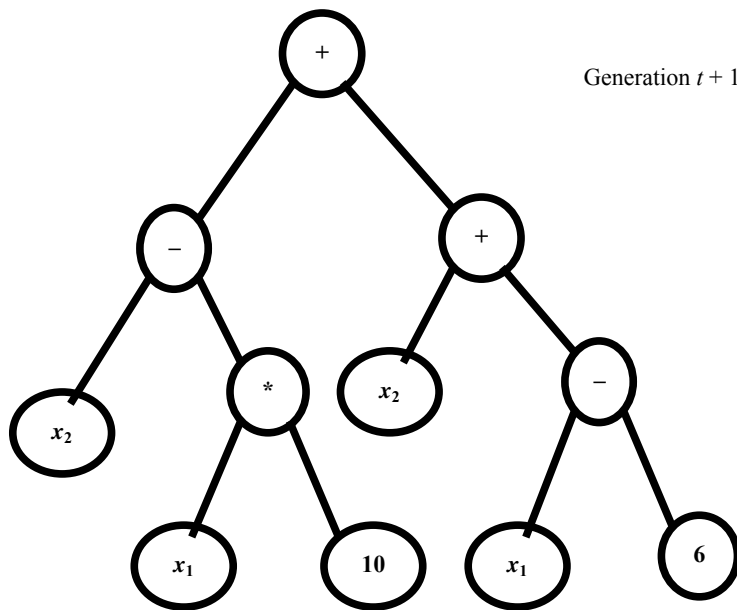
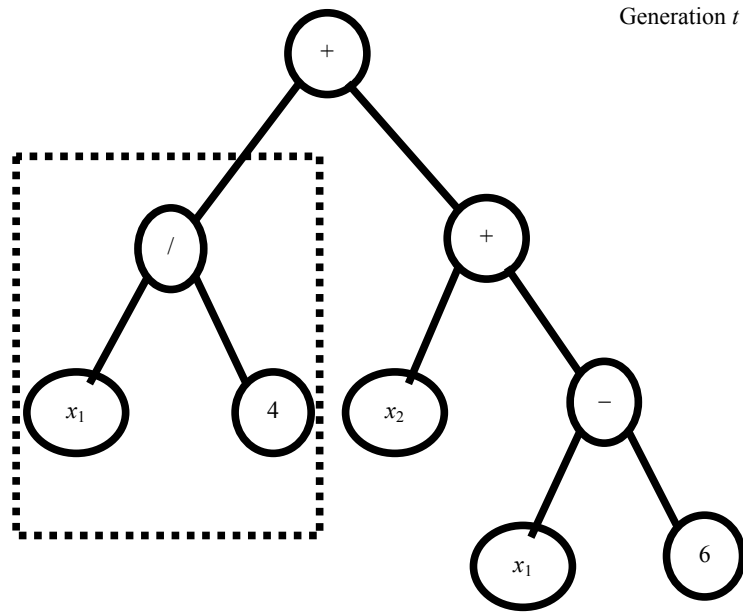


Fig. 5. Mutation's example

Рис. 5. Пример мутации

The dependence of the time of finding the structure on the value of the training sample example is not observed either (fig. 7).

The genetic algorithm restore the function: $f(x) = \cos(x_2) + \cos(x_2) + \cos(x_2) + \cos(x_2) + \cos(x_2) + \cos(x_2) + (x_1 * \sin(x_1))$, but the restored structure does not copy the original, for example: $f(x) = ((3 * x_1) / 4) + \cos(x_2) * 6$, but, when $x_1 \in (0, 3)$, $x_2 \in (0, 3)$, this structure is close to the original.

When comparing fig. 6 and fig. 7, we can see that the average time to restore the structure is greater in example 2. It's the reason to suppose that the time of modeling depends on the "complexity" of the modeling object.

Then more units (nodes and leaves) the genetic tree requires to restore the parametric form of the original function, the more "complicated" this function is. The value of input parameters also has dependence on the "complexity".

In the next experiment, we restore the function with 5 input variables: $f(x) = x_1 + x_2 + x_3 + x_4 + x_5$, $x_1 \in (0, 3)$, $x_2 \in (0, 3)$. Let's check the dependence of the modeling time on the value of the training sample.

As can be seen from fig. 8, there is no dependence of the time of finding the structure on the size of the training sample.

The genetic algorithm restores the function: $f(x) = (x_5 + x_3) + (x_1 + (x_4 + x_2))$, but the restored structure does not copy the original, for example: $(x_3 + (5 + ((x_5/4) * x_3) + ((x_1/(5 - x_2)) * (x_4 - \sin(x_1 + x_5))))))$, but, when $x_1 \in (0,3)$, $x_2 \in (0,3)$, this structure is close to the original.

The modeling time depends on the “complexity” of the modeling function shown in the fig. 9. You can see that in some cases the modeling process is fast. This happens in those cases when the algorithm choses the close to the desired function original tree.

As we can see in fig. 9, there is strong correlation between the modeling’s time and the value of input variables.

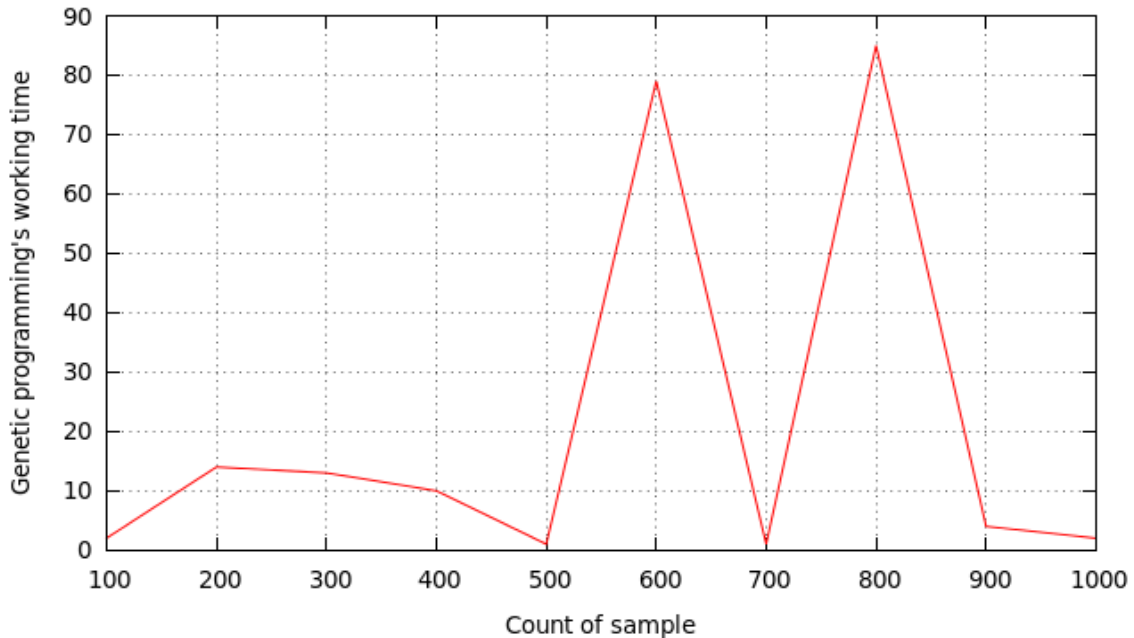


Fig. 6. Dependence of the time of finding the structure from the value of the training sample, example 1

Рис. 6. Зависимость времени нахождения структуры от величины обучающей выборки, пример 1

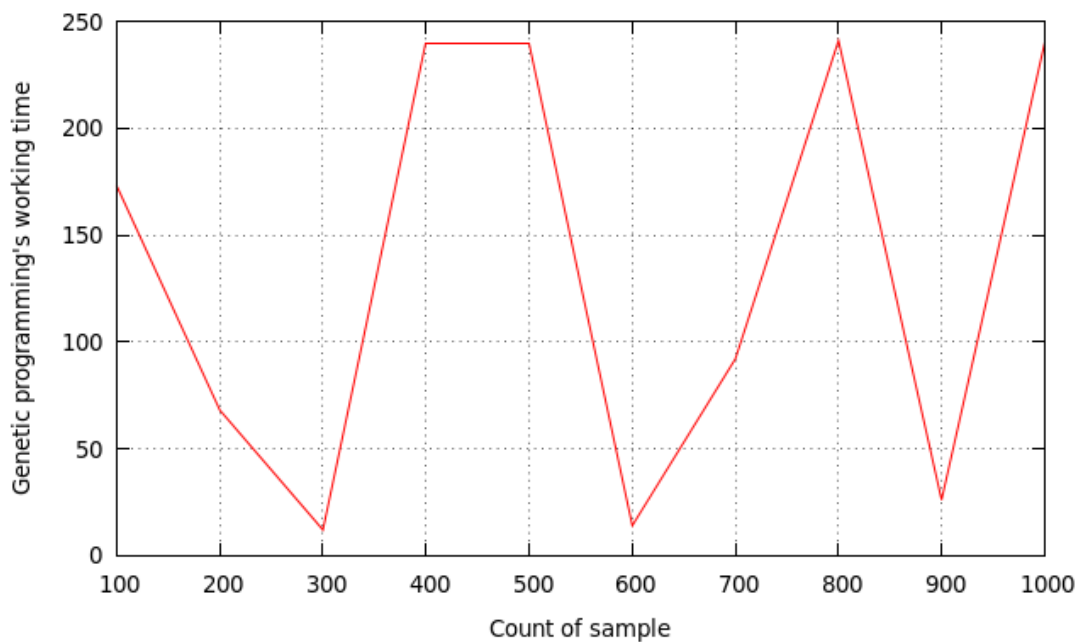


Fig. 7. Dependence of the time of finding the structure on the value of the training sample, example 2

Рис. 7. Зависимость времени нахождения структуры от величины обучающей выборки, пример 2

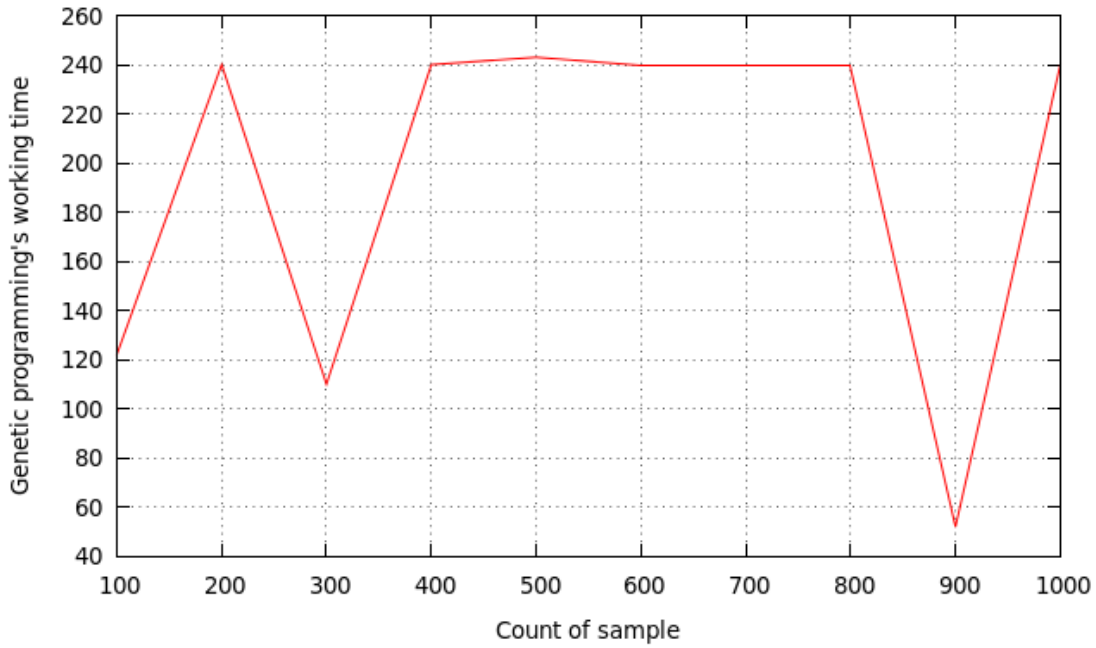


Fig. 8. Dependence of the time of finding the structure on the value of the training sample, example 3

Рис. 8. Зависимость времени нахождения структуры от величины обучающей выборки, пример 3

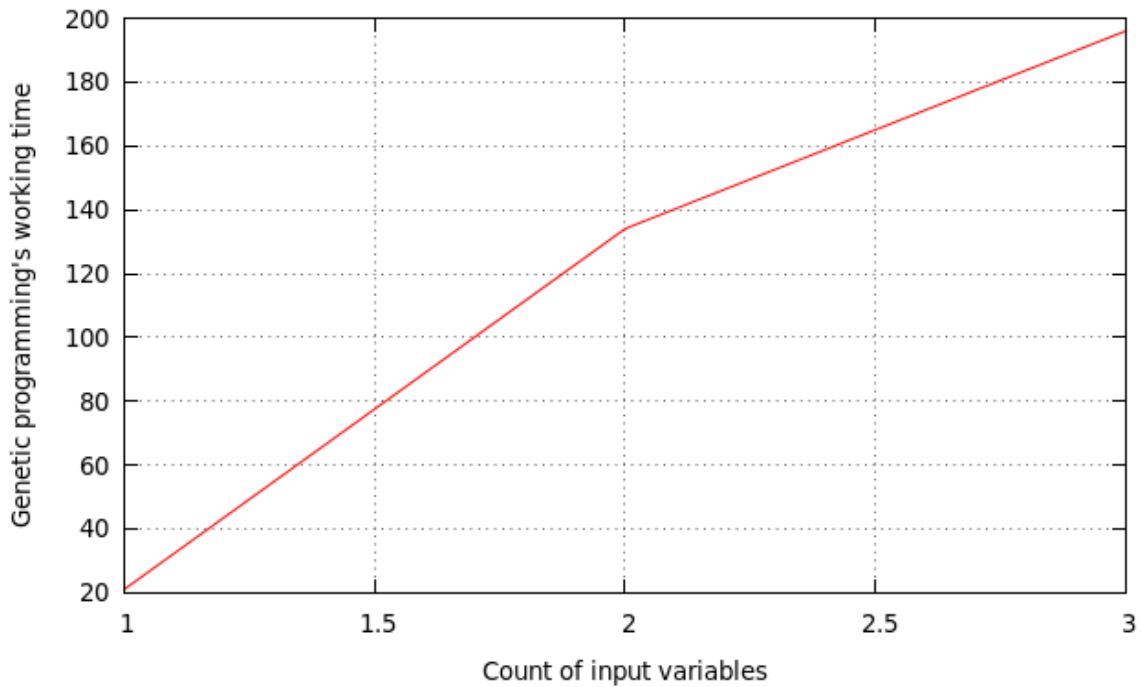


Fig. 9. Dependence of the time of finding the structure on the number of input variables, example 4

Рис. 9. Зависимость времени нахождения структуры от количества входных переменных, пример 4

Conclusion. The method of genetic programming for finding the parametric structure of the inertial-free object has been presented. It has been shown that the time of finding the parametric structure depends on train sample value. Also, the time of finding the parametric structure is dependent on the number of input variables.

Acknowledgments. The work was carried out with the financial support of the Ministry of Education and

Science of the Russian Federation during the implementation of the comprehensive project for the creation of high-tech production, contract No. 03.G25.31.0279.

Благодарности. Работа выполнена при финансовой поддержке Министерства образования и науки РФ в ходе реализации комплексного проекта по созданию высокотехнологичного производства, договор № 03.G25.31.0279.

References

1. Tweedle V., Smith R. A mathematical model of Bieber Fever. *Transworld Research Network*. 2012, Vol. 37/661, No. 2, P. 157–177.
2. Soviets B. Ya., Yakovlev S. A. *Modelirovanie sistem* [Modeling of systems]. Moscow, Vysshaya shkola Publ., 2001, 343 p.
3. Antonov A. V. *Sistemnyy analiz* [The system analysis], Moscow, Vysshaya shkola Publ., 2004, 454 p.
4. Trusov P. V. *Vvedenie v matematicheskoe modelirovanie* [Introduction to mathematical modeling], Moscow, Logos Publ., 2005, 440 p.
5. Medvedev A. V. [Analysis of data in the identification problem]. *Sbornik nauchnykh statey Mezhdunarodnoy konferentsii komp'yuternyy analiz dannykh i modelirovanie*. 1995, Vol. 2, P. 201–207 (In Russ.)
6. Soviets B. Ya., Yakovlev S. A. *Modelirovanie sistem* [Modeling of systems]. Moscow, Vysshaya shkola Publ., 2001, 343 p.
7. Tirsina A. N. *Teoriya sistem i sistemnyy analiz* [Theory of systems and systems analysis]. Chelyabinsk, Znaniya Publ., 2002, 128 p.
8. Skobtsov Yu. A. *Osnovy evolyutsionnykh vychisleniy* [Fundamentals of evolutionary calculations]. Donetsk, DonNTU Publ., 2008, 326 p.
9. Emelyanov V. V., Kureichik V. V., Kureichik V. M. *Teoriya i praktika evolyutsionnogo modelirovaniya* [Theory and practice of evolutionary modeling]. Moscow, Fizmatlit Publ., 2003, 432 p. (In Russ.)
10. Kureichik V. M., Lebedev B. K., Lebedev O. K. *Poiskovaya adaptatsiya: teoriya i praktika* [Search adaptation: theory and practice]. Moscow, Fizmatlit Publ., 2006, 272 p.
11. Gladkov L. A., Kureichik V. V., Kureichik V. M. *Geneticheskie algoritmy* [Genetic Algorithms], Moscow, Fizmatlit Publ., 2006, 320 p.
12. Gladkov L. A., Kureichik V. V., Kureichik V. M. *Bioinspirirovannye metody v optimizatsii* [Bioinspired methods in optimization]. Moscow, Fizmatlit Publ., 2009, 384 p.
13. Bukatova I. L. *Evolutsionnoe modelirovanie i ego prilozheniya* [Evolutionary modeling and its applications]. Moscow, Nauka Publ., 1994, 232 p.
14. Lugger J. *Iskustvennyy intellekt. Strategiya i metody resheniya slozhnykh problem* [Artificial Intelligence. Strategy and methods for solving complex problems], Moscow, Williams Publ., 2003, 864 p.
15. Koza J. R. *Genetic Programming*. Cambridge / MA: MIT press, 1994, 836 p.

Библиографические ссылки

1. Tweedle V., Smith R. A mathematical model of Bieber Fever // *Transworld Research Network*. 2012. Vol. 37/661, № 2. P. 157–177.
2. Советов Б. Я., Яковлев С. А. Моделирование систем. М. : Высш. шк., 2001. 343 с.
3. Антонов А. В. Системный анализ. М. : Высш. шк., 2004. 454 с.
4. Введение в математическое моделирование / под ред. П. В. Трусова. М. : Логос, 2005. 440 с.
5. Медведев А. В. Анализ данных в задаче идентификации // *Компьютерный анализ данных и моделирование* : сб. науч. ст. Междунар. конф. 1995. Т. 2. С. 201–207.
6. Советов Б. Я., Яковлев С. А. Моделирование систем. М. : Высш. шк., 2001. 343 с.
7. Теория систем и системный анализ / под ред. А. Н. Тырсина. Челябинск : Знания, 2002. 128 с.
8. Скобцов Ю. А. Основы эволюционных вычислений. Донецк : ДонНТУ, 2008. 326 с.
9. Емельянов В. В., Курейчик В. В., Курейчик В. М. Теория и практика эволюционного моделирования. М. : Физматлит, 2003. 432 с.
10. Курейчик В. М., Лебедев Б. К., Лебедев О. К. Поисковая адаптация: теория и практика. М. : Физматлит, 2006. 272 с.
11. Гладков Л. А., Курейчик В. В., Курейчик В. М. Генетические алгоритмы. М. : Физматлит, 2006. 320 с.
12. Гладков Л. А., Курейчик В. В., Курейчик В. М. Биоинспирированные методы в оптимизации. М. : Физматлит, 2009. 384 с.
13. Букатова И. Л. Эволюционное моделирование и его приложения. М. : Наука, 1994. 232 с.
14. Люггер Дж. Искусственный интеллект. Стратегия и методы решения сложных проблем. М. : Вильямс, 2003. 864 с.
15. Koza J. R. *Genetic Programming*. Cambridge ; MA : MIT press, 1994. 836 с.