

УДК 681.3

**ПРИМЕНЕНИЕ ГЕНЕТИЧЕСКИХ АЛГОРИТМОВ В ЗАДАЧАХ АППРОКСИМАЦИИ
СЛОЖНЫХ МНОГОМЕРНЫХ ЗАВИСИМОСТЕЙ И ИДЕНТИФИКАЦИИ
ПАРАМЕТРОВ ТЕОРЕТИЧЕСКИХ МОДЕЛЕЙ**

*канд. техн. наук, доц. Д.О. ГЛУХОВ, Т.М. ГЛУХОВА, А.О. ЛУКЬЯНОВ
(Полоцкий государственный университет)*

Предложен метод построения аналитической аппроксимации n -мерных данных, основанный на применении генетического алгоритма. Особенностью метода является то, что кодирование пространства поиска выполняется в виде дерева разбора алгебраического выражения синтаксическим анализатором контекстно-свободной грамматики класса LR(1). Кроме этого, в ходе эволюционного процесса, помимо применения мутаций структуры (при условии их положительного влияния), выполняется этап мутации коэффициентов, позволяющих избежать попадания в локальные экстремумы целевой функции. А также на каждом шаге эволюционного процесса предусмотрен этап поиска экстремума в пространстве коэффициентов и этап упрощения аналитической модели.

Ключевые слова: генетические алгоритмы, популяция, мутация, аппроксимация, контекстно-свободная грамматика, нечеткая логика.

Под сложностью многомерной зависимости мы будем понимать неопределенность относительно присутствующих в данных функциональных зависимостей между входными переменными и выходным значением функции. Сложные многомерные зависимости встречаются, в частности, в следующих задачах:

- идентификация параметров аналитических моделей аппроксимации многомерных зависимостей;
- идентификация параметров теоретических вероятностных распределений;
- идентификация параметров моделей трендов при прогнозировании.

Аппроксимация данных в условиях отсутствия априорной информации о реальной модели требует применения универсальных аппроксиматоров, таких как искусственные нейронные сети, нечеткие логические аппроксиматоры. Однако аналитическое представление модели является наиболее компактным и наиболее точным. Методы аппроксимации, основанные на идентификации параметров некоторой теоретической модели, решают задачу, используя неизменяемую модель, принимаемую из каких-либо предположений разумности.

Нами делается попытка применения генетического алгоритма для поиска наиболее точной аналитической модели аппроксиматора. Генетический алгоритм выполняет не только идентификацию параметров теоретической модели, но и поиск самой модели в пространстве возможных модификаций структуры. Целенаправленное движение популяции моделей к моделям, обеспечивающим качественную аппроксимацию осуществляется благодаря закону эволюционного отбора по критерию минимизации среднеквадратического отклонения (СКО).

Генетический алгоритм (ГА) – это эвристический алгоритм поиска, используемый для решения задач оптимизации и моделирования путём последовательного подбора, комбинирования и вариации искоемых параметров с использованием механизмов, напоминающих биологическую эволюцию.

Ранее нами были предприняты попытки решения данной задачи на многоагентных системах, которые для аналитических моделей в пространстве 3-х переменных показали результаты, превосходящие результаты искусственных нейронных сетей, но на ограниченном разнообразии моделей [1, 2, 3]. Основным недостатком таких аппроксиматоров является ограниченный набор преобразований структуры, сформулированный в виде локальных эвристик. Эвристики являются очевидными для человека-эксперта способами изменения структуры. Соответственно неочевидные для эксперта модификации становятся невозможными. На наш взгляд, применение ГА снимает данное ограничение.

Кроме этого нами были построены аппроксиматоры на нечеткой логике (НЛА). Нечеткие логические системы, предложенные Лофти Заде [8], получили развитие в работах [7, 9]. Если рассматривать неизвестный параметр как непрерывный, то в этом случае можно провести параллель между выводом о значении неизвестного параметра и приближением функции и говорить о свойстве нечеткой системы выступать в роли лингвистического аппроксиматора. Преимуществом таких аппроксиматоров является их быстродействие, а основным недостатком – ограниченные возможности по повышению точности аппроксимации.

Контекстно-свободная грамматика G_{exp} класса LR(1) алгебраических выражений, предложенная нами, представлена ниже в форме Бекуса-Наура:

```

formula:      exp          {ANode* o = (ANode*)$1; createdNodes.push_back(o);}
|             formula ';' exp {ANode* o = (ANode*)$3; createdNodes.push_back(o);}
;

exp :         NUM          {ANodeNUM* o = new ANodeNUM($1); $$ = (void*) o;}
|           '(' exp ')' {$$ = $2;}
|           exp '+' exp   {ANode* o = (ANode*) new ANodeOp((ANode*)$1, (ANode*)$3, PLUS); $$ = (void*) o;}
|           exp '-' exp   {ANode* o = (ANode*) new ANodeOp((ANode*)$1, (ANode*)$3, MINUS); $$ = (void*) o;}
|           exp '/' exp   {ANode* o = (ANode*) new ANodeOp((ANode*)$1, (ANode*)$3, DIV); $$ = (void*) o;}
|           exp '*' exp   {ANode* o = (ANode*) new ANodeOp((ANode*)$1, (ANode*)$3, MUL); $$ = (void*) o;}
|           'p' 'o' 'w' '(' exp ',' exp ')' {ANode* o = (ANode*) new ANodeOp((ANode*)$5, (ANode*)$7, POW);
|                                           $$ = (void*) o;}
|
|           's' 'i' 'n' '(' exp ')'         {ANode* o = (ANode*) new ANodeFunc((ANode*)$5, SIN); $$ = (void*) o;}
|           'c' 'o' 's' '(' exp ')'         {ANode* o = (ANode*) new ANodeFunc((ANode*)$5, COS); $$ = (void*) o;}
|           'l' 'o' 'g' '(' exp ')'         {ANode* o = (ANode*) new ANodeFunc((ANode*)$5, LOG); $$ = (void*) o;}
|           'e' 'x' 'p' '(' exp ')'         {ANode* o = (ANode*) new ANodeFunc((ANode*)$5, EXP); $$ = (void*) o;}
|           't' 'a' 'n' '(' exp ')'         {ANode* o = (ANode*) new ANodeFunc((ANode*)$5, TAN); $$ = (void*) o;}
|           'c' 't' 'a' 'n' '(' exp ')'     {ANode* o = (ANode*) new ANodeFunc((ANode*)$6, CTAN); $$ = (void*) o;}
|           NAME {ANode* o = (ANode*) new ANodeNAME(string((char*)$1)); $$ = (void*) o; delete $1;}
;
    
```

Как мы видим, в процессе синтаксического разбора строится дерево правостороннего вывода в виде объектно-ориентированной структуры. Для получения возможности компиляции алгебраического выражения в дерево вложенных друг в друга объектов нами разработана система классов, представленная на диаграмме классов (рисунок 1).

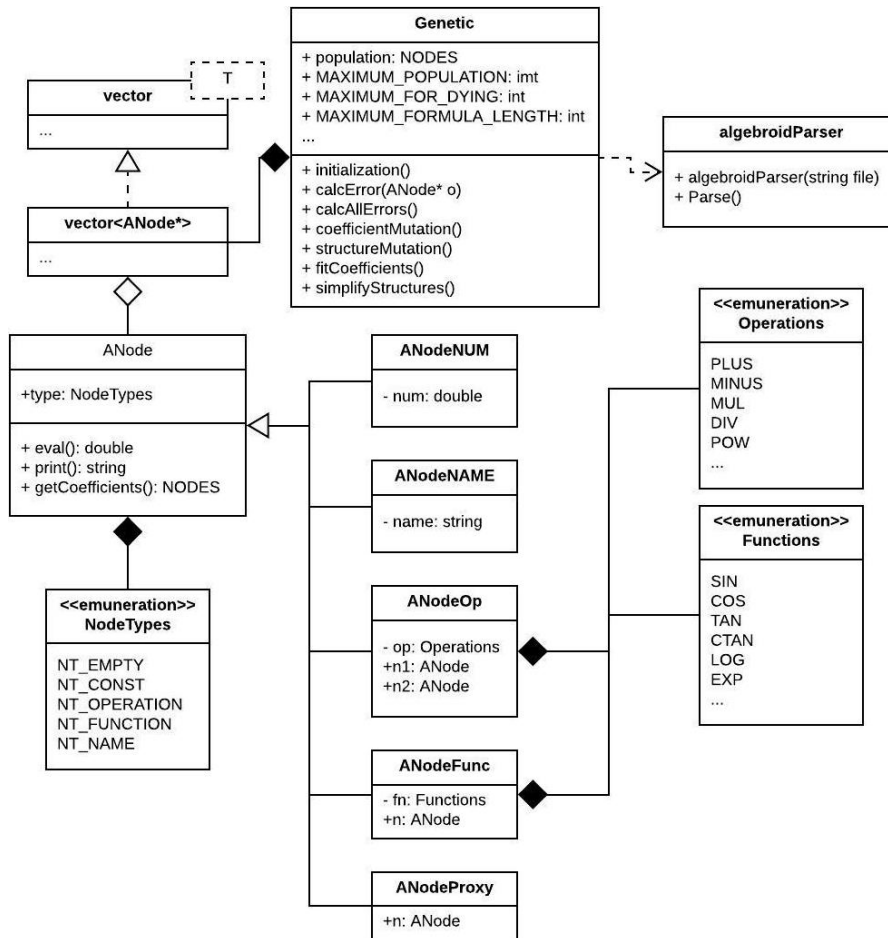


Рисунок 1. – Диаграмма классов программного обеспечения ГА поиска модели аналитического n-мерного аппроксиматора

В результате синтаксического разбора файла инициализации создается начальная популяция аналитических n -мерных моделей. Начальные варианты моделей формулируются экспертом на формальном языке грамматики G_{exp} .

Объекты базового класса ANode и производных классов ANodeNUM, ANodeNAME, ANodeOp, ANodeFunc, ANodeProхy атрибутируют узлы семантического стека и формируют дерево правостороннего разбора алгебраического выражения.

Класс Genetic реализует основные этапы ГА:

- 1) initialization() – инициализация популяции экспертно-заданными для случая n -переменных алгебраическими выражениями;
- 2) calcAllErrors() – вычисление ошибки аппроксимации для каждой особи в ее текущем состоянии;
- 3) coefficientMutation() – процесс случайного клонирования особей с мутацией коэффициентов для обеспечения возможности выхода из локальных экстремумов;
- 4) structureMutation() – процесс клонирования с мутацией структуры, при условии положительного эффекта от мутации;
- 5) fitCoefficients() – процесс поиска экстремума целевой функции в пространстве коэффициентов.

В реализации данного процесса нами выбран алгоритм градиентного спуска, но с применением координатной релаксации. Данный алгоритм не является быстрым, но достаточно надежен;

- б) simplifyStructure() – упрощает структуру выражения в специфических случаях.

Упрощение структуры приводит к удалению и замене узлов семантического дерева (именно для процедуры упрощения предусмотрен класс ANodeProхy) и предусмотрено в следующих случаях:

- $x * 0 \mid 0 * x \rightarrow 0;$
- $x^0 \rightarrow 1; \quad x^1 \rightarrow x;$
- $x * 1 \mid 1 * x \rightarrow x;$

Упрощение срабатывает, когда алгоритм обнаруживает сходимость коэффициента к 0 или 1. Блок-схема алгоритма эволюционного поиска представлена на рисунке 2.

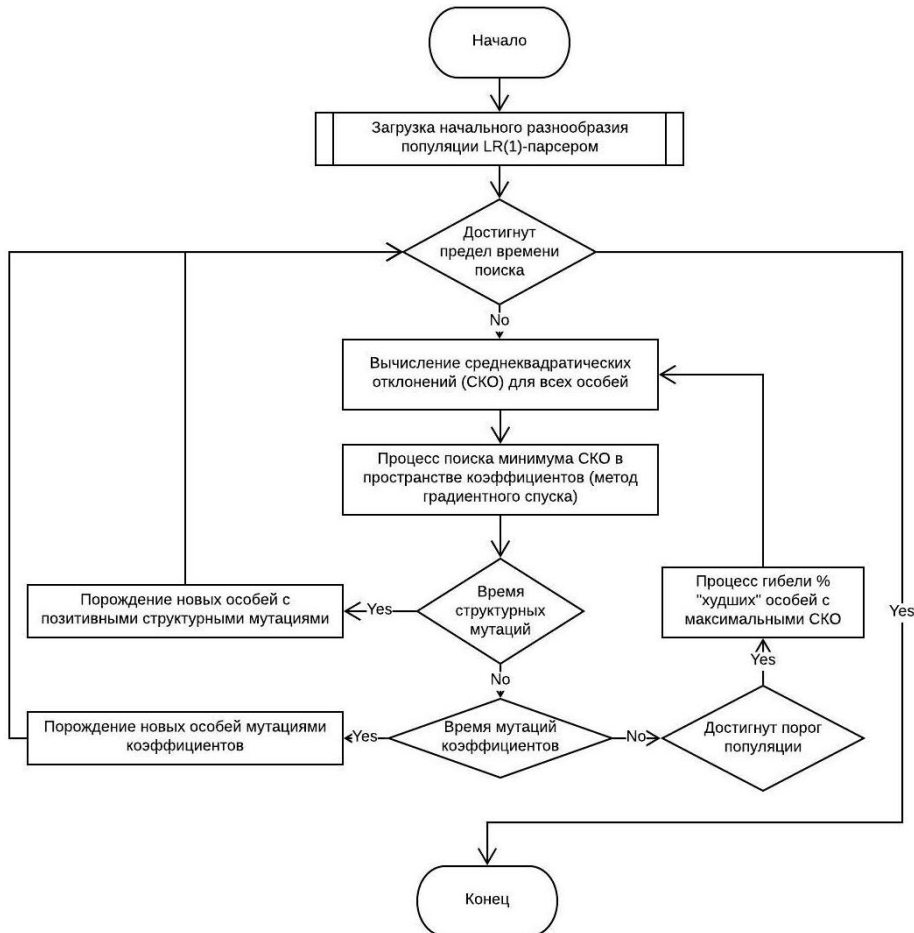
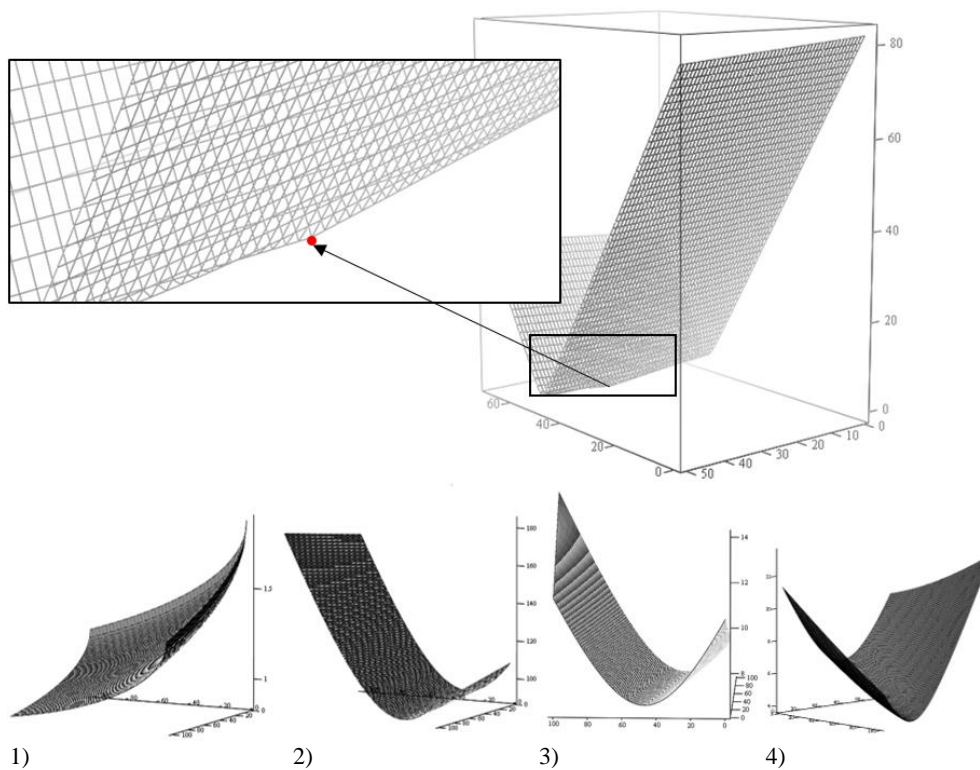


Рисунок 2. – Блок-схема алгоритма эволюционного поиска разработанного генетического алгоритма

В качестве целевой функции нами выбрано среднеквадратичное отклонение между теоретическими и расчетными значениями функции. В m -мерном пространстве коэффициентов аналитических моделей поверхность СКО может иметь локальные экстремумы в связи с ограниченностью области данных и сложностью аппроксимируемых зависимостей. Однако характер поверхности позволяет выдвинуть предположение о возможности применения алгоритма градиентного спуска для поиска локального или глобального экстремума. Характер поверхности СКО в пространстве двух коэффициентов a, b функции одной переменной проиллюстрирован на рисунке 3.



$$1) y(x) = \sqrt{\frac{a}{x}} + \sqrt{\frac{b}{x^2}}; \quad 2) y(x) = ax^2 + bx; \quad 3) y(x) = a \log(x) + bx; \quad 4) y(x) = 5e^{1/x} + 3x$$

Рисунок 3. – Поверхность СКО в пространстве коэффициентов a, b моделей

В ряде работ, посвященных данной тематике, предлагается подбирать минимизирующие СКО коэффициенты в результате мутации коэффициентов и применять алгоритм Нелдера – Мида для поиска локального экстремума [5, 6].

С целью минимизации среднеквадратической ошибки отклонения некоторой аналитической функции от дискретного набора данных мы провели сравнительный анализ применимости различных численных методов оптимизации и предоставляем пользователю системы выбор конкретного метода. Наш подход основан на двух положениях:

1. Мутация коэффициентов необходима, для скачкообразного перехода в новую точку поиска экстремума, что позволяет алгоритму запустить несколько стратегий поиска экстремума из различных областей n -мерного пространства поиска, а, следовательно, избежать попадания в локальные экстремумы.

2. Применение наиболее качественных методов поиска экстремума в зависимости от специфики задачи (метод градиентного спуска, симплекс метод, и др.) для идентификации коэффициентов с заданной точностью.

В численных экспериментах наилучшую скорость сходимости показал метод покоординатного спуска, при котором шаг выполняется по координате, дающей максимальный выигрыш. Однако поиск и реализация более эффективных алгоритмов оптимизации остаются предметом дальнейшего исследования.

Важно отметить, что нами предложен новый метод отсева эквивалентных формул. Поскольку одна и та же формула в силу коммутативности операций (+, -, /, *) и дистрибутивности соответствующих пар операций, а также возможности произвольной расстановки скобок, может быть записана множеством различных вариантов, мы предложили формировать числовые хеши формул, используя специальные коды переменных, коэффициентов и функций и зеркально измененную алгебру (+ вместо *, - вместо /, * вместо +, / вместо -).

Пример работы данной хеш-функции выглядит следующим образом:

Формула	Хеш
$2 * \log(x1) * x2$	16994
$\log(x1) * 2 * x2$	16994
$x2 * 2 * \log(x1)$	16994
$\log(x1) * x2 * 2$	16994
$2 * (\log(x1) + x2)$	2.89382e+007
$2 * \log(x1) + x2 * 2$	2.89722e+007

Пример работы метода покоординатного спуска для семи особей популяции с точностью до 0.001 приведен в таблице 1. Из примера также видно, что нами разработан механизм обработки результатов в виде нечисел и бесконечностей, как больших значений СКО. Такой подход позволяет продолжать поиск без возникновения исключительных ситуаций переполнения и деления на ноль.

Таблица 1. – Пример результата идентификации коэффициентов модели методом покоординатного спуска

Формула	Среднеквадратичное отклонение
До идентификации коэффициентов	
1) $5 * x1 * x1 + 3$	0.00000
2) $2 * x1 + 3 * x2 + 4$	56.92715
3) $2 * \text{pow}(x2, 2) + 2$	154.52993
4) $15 * \text{pow}(x1, 2) + 13$	166.70333
5) $2 * x1 * x1 + 3 * x2 * x2 + 2 * x1 * x2 + 3 * x1 + 3 * x2 + 3$	359.15568
6) $11 * \text{pow}(x1, 3) + 3$	+INF
7) $\text{pow}(x1, x2) * 5$	+INF
После идентификации коэффициентов	
1) $5 * x1 * x1 + 3$	0.00000
2) $27.9921 * x1 + -2.02022 * x2 + 3.12347$	17.37498
3) $1.8 * \text{pow}(x2, 1.5) + 1.579$	57.33454
4) $5.00068 * \text{pow}(x1, 1.99993) + 2.99844$	0.00100
5) $4.99973 * x1 * x1 - 0.48889 * x2 * x2 + 0.33344 * x1 * x2 - 3.33275 * x1 + 4.86652 * x2 + 3.2222$	0.00099
6) $5.00068 * \text{pow}(x1, 1.99993) + 2.99844$	0.00100
7) $\text{pow}(x1, x2) * 3.3e-006$	50.52150

В качестве примера применения, построенного нами ГА для решения прикладной задачи, рассмотрим проблему учета переходных процессов в стационарных неизотермических моделях транспортировки газа.

Формулировка проблемы: при включении компрессорной станции начинается подогрев газа и, несмотря на то, что нагретый газ еще не распространился по трубопроводу, стационарная модель описывает такое состояние трубопровода, которое возникнет после перехода в установившейся режим. Таким образом, существенные участки трубопровода в модели оказываются с завышенным значением температуры газа и, учитывая сильную зависимость плотности газа от температуры, с заниженным значением запаса газа.

Мы рассчитали коэффициент корреляции, который для рассматриваемых участков с ростом температуры показал высокую отрицательную корреляцию. Так, для приведенных на рисунке 4 в качестве примера графиках коэффициент корреляции составил $Correl(X, Y) = \{-0,84369, -0,88839, -0,88509, -0,85517\}$.

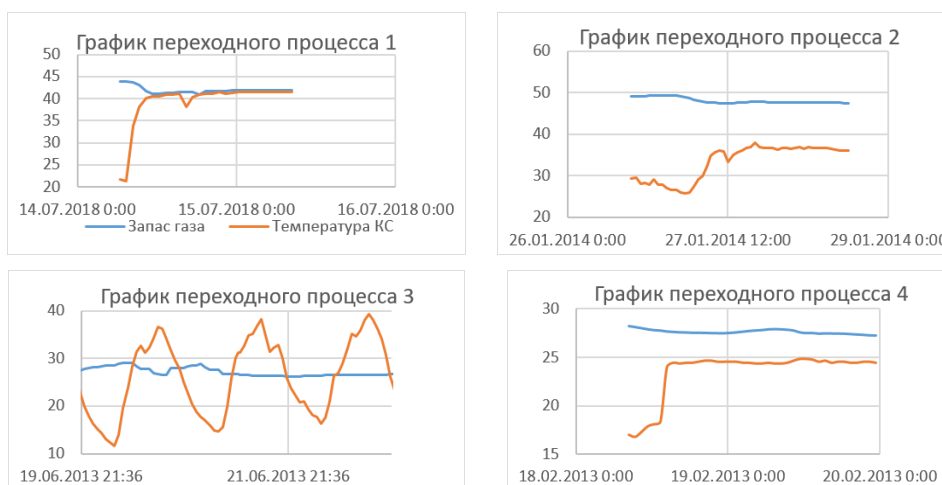


Рисунок 4. – Примеры графиков переходных процессов в рамках модели магистрального газопровода ОАО «Газпром трансгаз Беларусь»

Являясь авторами программного комплекса расчета запаса газа на магистральном газопроводе ОАО «Газпром трансгаз Беларусь», мы попытались устранить влияние таких переходных процессов путем введения инерционности изменения температуры на выходе компрессорной станции в течение времени перехода системы к установившемуся режиму.

Нами был применен ГА для поиска аналитической зависимости между значением времени переходного процесса y и текущего значения расхода газа Q , температуры грунта T_g и перепада температур на выходе компрессорной станции dT . Задачей аппроксиматора был поиск такой аналитической зависимости, которая обладает минимальным СКО для описания $y(Q, T_g, dT)$. Сравнительные результаты работы ГА приведены в таблице 2.

Таблица 2. – Сравнительные результаты работы ГА

Метод аппроксимации	Модель	СКО
ГА	$1.66597 * \text{pow}(1.7573, 1.1225 * \log(3.21177 * \text{pow}(\Delta T * 4.23434, 1.1878))) / \text{pow}(\log(0.0858575 * Q), 0.0742341) - 4.15915 * \Delta T$	1,078879
1) ГА	$52 - 14.48 * \text{pow}(\Delta T, 0.08961) / \text{pow}(Q, -0.1755)$;	1,330954
2) ГА	$102.607 * \text{pow}(\Delta T, -0.369501) / \log(0.265488 * Q)$;	1,200513
3) НЛА	Модель описана в работе [10]	1,080616

Найденное ГА решение превзошло по точности нечеткий логический аппроксиматор с подобранными оптимальными настройками сглаживания, дав при этом наиболее компактное представление аппроксиматора.

Динамику улучшения качества аппроксимации можно представить графиком позитивных мутаций (рисунок 5).

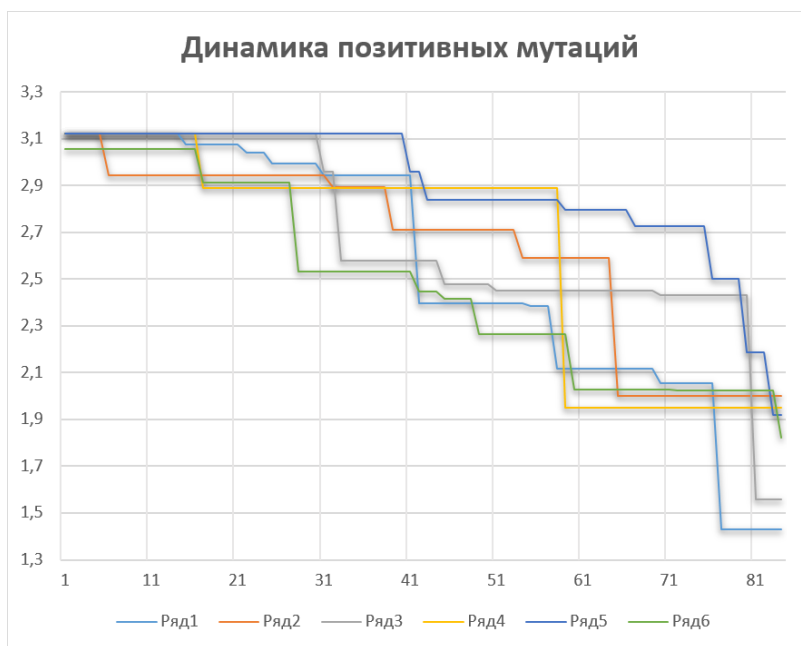


Рисунок 5. – График изменения СКО в результате шагов мутации

Одним из параметров предложенного ГА является предельный размер формулы, что позволяет получать большое разнообразие компактных вариантов аппроксиматоров.

Разработанный ГА имеет возможность гибкой настройки предельных порогов численности популяции, процента и периода гибели неперспективных особей, вероятности периода и вида мутаций структуры и коэффициентов, формулировки инструментов упрощения моделей, алгоритмов поиска экстремума. В частности, для решения задачи учета переходных процессов в стационарных неизотермических моделях транспортировки газа, нас интересует не только время, но и характер инерции температуры, что требует сбора и обработки дополнительной информации.

Выводы. В данной статье предложен новый метод построения аналитической аппроксимации сложных n -мерных зависимостей, основанный на применении генетического алгоритма.

Разработанный ГА имеет следующие особенности:

1) кодирование пространства поиска выполняется в виде дерева разбора алгебраического выражения синтаксическим анализатором контекстно-свободной грамматики класса LR(1);

2) в ходе эволюционного процесса, помимо применения мутаций структуры (при условии их положительного влияния), выполняется этап мутации коэффициентов, позволяющие избежать попадания в локальные экстремумы целевой функции;

3) на каждом шаге эволюционного процесса предусмотрен этап поиска экстремума в пространстве коэффициентов и этап упрощения аналитической модели.

Разработанный ГА был применен для решения прикладной задачи учета переходных процессов в стационарных неизотермических моделях транспортировки газа. Полученные модели были внедрены в рамках программного комплекса расчета запаса газа на магистральном газопроводе ОАО «Газпром трансгаз Беларусь», что обеспечило снижение влияния переходных процессов на дисбаланс газа.

ЛИТЕРАТУРА

1. Глухов, А.О. Многоагентные структуры для решения задачи коммивояжера / А.О. Глухов, В.В. Трофимов // Проблемы менеджмента : сб. науч. тр. – СПб. : Изд-во СПбГУЭФ, 2000. – Вып. 3. – С. 71–76.
2. Трофимов, В.В. Аппроксимация на многоагентных структурах / В.В. Трофимов, А.О. Глухов // Экономическая кибернетика: системный анализ в экономике и управлении : сб. науч. тр. – СПб. : Изд-во СПбГУЭФ, 2000. – Вып. 1. – С. 40–53.
3. Modified hybrid genetic algorithm of discreet optimization problems / Glukhov A.O. [et al.] / Proceedings of 2017 20th IEEE International Conference on Soft Computing and Measurements, SCM 2017 6 July 2017, Scopus 7970603. – St. Petersburg ; Russian Federation, 2017. – P. 417–419.
4. Пожарский, Д.А. Генетический алгоритм для нахождения коэффициентов аппроксимации функции в контактных задачах для цилиндра [Электронный ресурс] / Д.А. Пожарский, Н.Б. Золотов, И.Е. Семенов // Молодой ученый. – 2017. – № 24. – С. 122–125. – Режим доступа: <https://moluch.ru/archive/158/44625/>. – Дата доступа: 09.10.2019.
5. Кильдюшов, М.С. Программа для восстановления аппроксимированных алгебраических функций от нескольких переменных по набору дискретных значений функции [Электронный ресурс] / М.С. Кильдюшов // Интернет-журнал «Науковедение». – 2015. – Т. 7, № 5. – Режим доступа: <http://naukovedenie.ru/PDF/136TVN515.pdf>. – Дата доступа: 09.10.2019.
6. Кильдюшов, М.С. Применение генетических алгоритмов для восстановления аппроксимированных алгебраических функций с определенной точностью / М.С. Кильдюшов // Наука и бизнес: пути развития ; Фонд развития науки и культуры, Тамбов. – 2016. – № 1 (55), – С. 25–31.
7. Driankov, D. An introduction to fuzzy control / D. Driankov, H. Hellendoorn, M. Reinfrank. – Springer-Verlag, 1993. 5. Zadeh, Lotfi A. Fuzzy Sets / Lotfi A. Zadeh // Information & Control. – 1965. – Vol. 8. – P. 338–353.
8. Zadeh, Lotfi A. Fuzzy Sets / Lotfi A. Zadeh // Information & Control. – 1965. – Vol. 8. – P. 338–353.
9. Glukhov, D.O. Dynamic expert system by fuzzy inference rules to automations an examination of complex objects / D.O. Glukhov // Budownictwo i Inzynieria Srodowiska. – Zielonogorsk : Politechnika Zielonogorska, 1998. – P. 105–109.
10. Глухов, Д.О. Мягкие вычисления для организации компьютерного представления номограмм на примере вычисления предельного коэффициента ползучести / Д.О. Глухов, Т.М. Глухова, С.П. Кундас // Вестн. Полоц. гос. ун-та. Сер С, Фундам. науки. – 2010. – № 3. – С. 2–6.
11. Звонков, В.Б. Сравнительное исследование классических методов оптимизации и генетических алгоритмов / В.Б. Звонков, А.М. Попов // Вестн. Сибир. гос. аэрокосмич. ун-та им. М.Ф. Решетнева. – 2013. – С. 23–27.

Поступила 20.09.2019

APPLICATION OF GENETIC ALGORITHMS IN THE PROBLEMS OF APPROXIMATION OF COMPLEX MULTIDIMENSIONAL DEPENDENCES AND IDENTIFICATION OF THE PARAMETERS OF THEORETICAL MODELS

D. GLUKHOV, T. GLUKHOVA, A. LUKYANOV

Proposes a method for constructing an analytical approximation of n-dimensional data, based on the use of a genetic algorithm. A feature of the method is that the encoding of the search space is performed in the form of a parsing tree for an algebraic expression by the parser of the context-free grammar of the class LR(1). In addition, during the evolutionary process, in addition to the use of structure mutations (subject to their positive influence), the stage of mutation of the coefficients is performed, which allows avoiding the target function falling into local extremum. And also at each step of the evolutionary process, there is a stage for searching for an extremum in the space of coefficients and a stage for simplifying the analytical model.

Keywords: genetic algorithms, population, mutation, approximation, context-free grammar, fuzzy logic.