

Экспериментальное исследование алгоритма генетического программирования с оптимизированными операторами на булевых данных

Захаров Алексей

Институт математики им. С.Л. Соболева СО РАН (Омский филиал)

Семинар "Модели и алгоритмы для задач составления
расписаний"

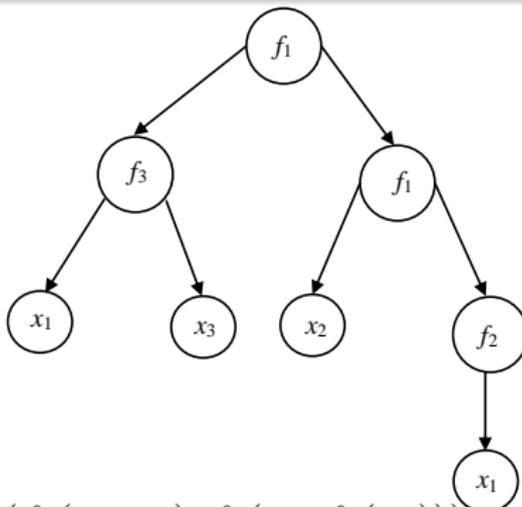
Исследование выполнено за счет гранта Российского научного фонда № 22-71-10015.

20.04.2024

Входные данные $\{(\bar{x}^i, y^i)\}$, $\bar{x}^i = (\bar{x}_1^i, \dots, \bar{x}_m^i)$, $i = 1, \dots, n$
(обучающая выборка).

Целевая функция $g(T) = \sum_{i=1}^n (y_i - T(\bar{x}_m^i))^2$,

$T(\bar{x}^i)$ – значение функционального дерева T на \bar{x}^i .



$$T(x_1, x_2, x_3) = f_1(f_3(x_1, x_3), f_1(x_2, f_2(x_1)))$$

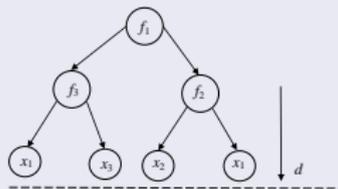
- 1 Koza J. R. (1992) Genetic Programming: On the Programming of Computers by Natural Selection. MIT Press, Cambridge, MA, USA.
- 2 Koza J.R., Poli R. Genetic Programming. In: Burke E.K., Kendall G. (eds) Search Methodologies. Springer, Boston, MA (2005).
- 3 Poli, R., Page J. Solving High-Order Boolean Parity Problems with Smooth Uniform Crossover, Sub-Machine Code GP and Demes. Genetic Programming and Evolvable Machines 1, 37-56 (2000).
- 4 Langdon W.B. Size Fair and Homologous Tree Crossovers for Tree Genetic Programming. Genetic Programming and Evolvable Machines 1, 95-119 (2000).
- 5 Moraglio A., Krawiec K., Johnson C.G. Geometric Semantic Genetic Programming. PPSN 2012. Lecture Notes in Computer Science, vol 7491. Springer, Berlin, Heidelberg (2012).
- 6 Santoso L.W., Singh B., Rajest S.S., Regin R., Kadhim K.H. A Genetic Programming Approach to Binary Classification Problem. EAI Endorsed Transactions on Energy Web. V. 8 (31) (2020).
- 7 Doerr B., Lissovoi A., Oliveto P. S. (1+ 1) genetic programming with functionally complete instruction sets can evolve Boolean conjunctions and disjunctions with arbitrarily small error. Artificial Intelligence, 319, 103906 (2023).

- 1: Построить начальную популяцию.
- 2: Повторить шаги 3–7 пока не будет выполнен критерий остановки.
- 3: Выбрать две особи T_1, T_2 из текущей популяции.
- 4: Применить оператор мутации к обоим особям T_1, T_2 с некоторой вероятностью и получить особи T'_1, T'_2 , соответственно.
- 5: Построить потомка T' , применив оператор кроссинговера к особям T'_1, T'_2 .
- 6: Выбрать лучшую особь T_b среди особей T', T_1 и T_2 .
- 7: Заменить худшую особь в популяции особью T_b .
- 8: Вернуть лучшую особь по отношению к значению целевой функции на протяжении работы алгоритма.

Методы генерации начальной популяции

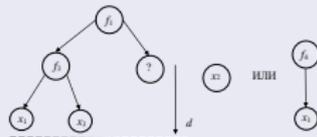
Full method

Полное дерево заданной глубины



Grow method

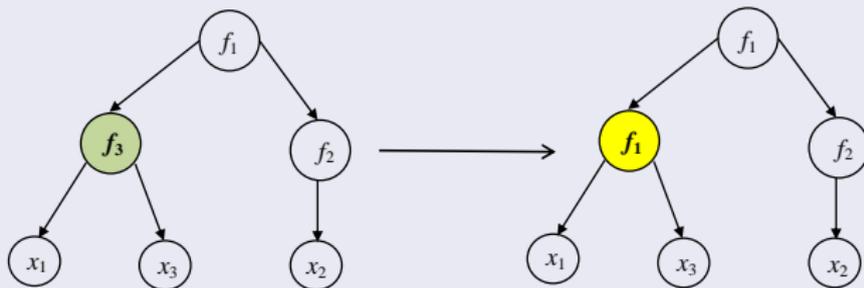
Каждый узел с некоторой вероятностью: внутренний или терминальный (лист)
Задана верхняя граница на глубину дерева



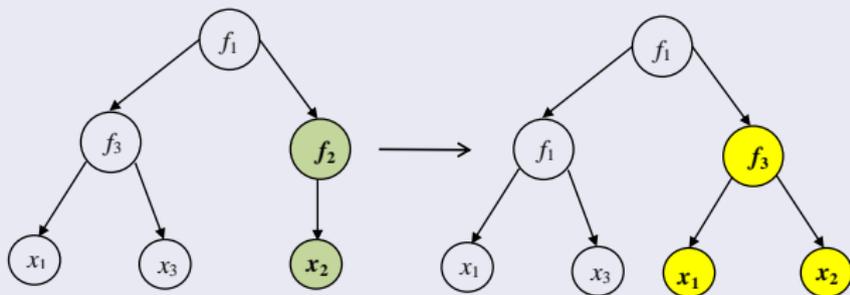
Ramped half-and-half method

Группы деревьев с заданной величиной i : от нижней до верхней границы. Каждая группа: 50% полных деревьев с глубиной i , 50% деревьев – методом grow с верхней границей i . В каждой группе равное кол-во элементов (почти).

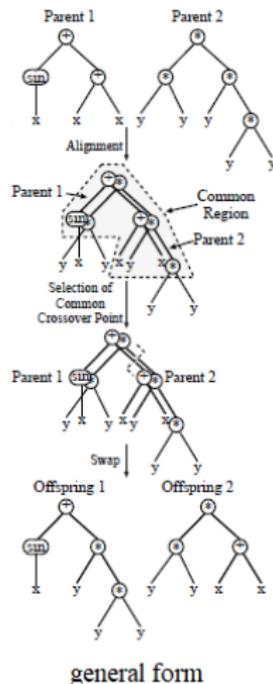
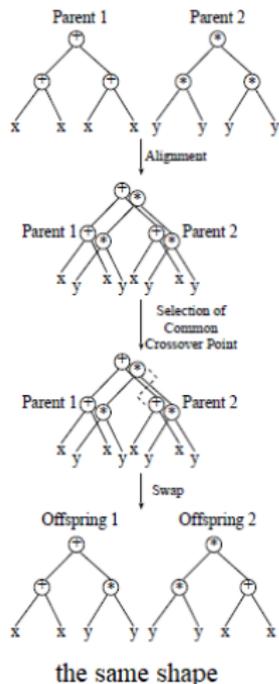
Point mutation (GP-PM)



Subtree mutation (GP-SM)

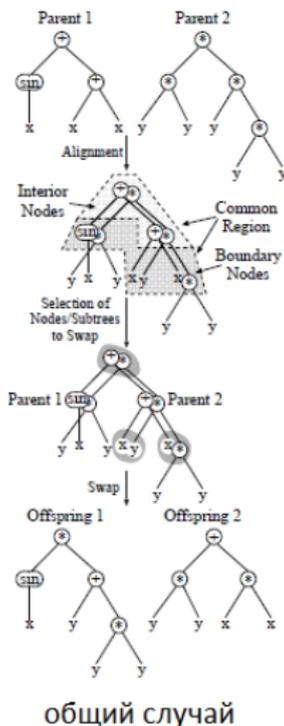
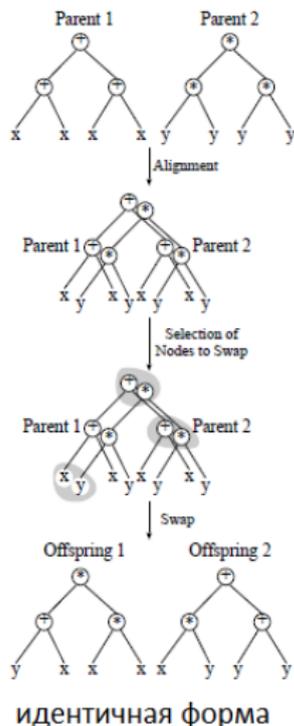


One-point crossover (GP-OPX)



Poli R., Langdon W.B. On the search properties of different crossover operators in genetic programming (1998)

Uniform prosrossover (GP-UX)



Poli R., Page J. Solving high-order Boolean parity problems with smooth uniform crossover, sub-machine code GP and demes (2000)

Optimal recombination problem (ORP)

The definition is based on genes transmission.¹

Optimal recombination problem

^a Given an instance I of combinatorial optimization problem with the set of feasible solutions Sol and two parents $\mathbf{p}^1 = (p_1^1, \dots, p_l^1), \mathbf{p}^2 = (p_1^2, \dots, p_l^2)$ from Sol .

The goal is to find the offspring $\mathbf{p}' \in \text{Sol}$ such that

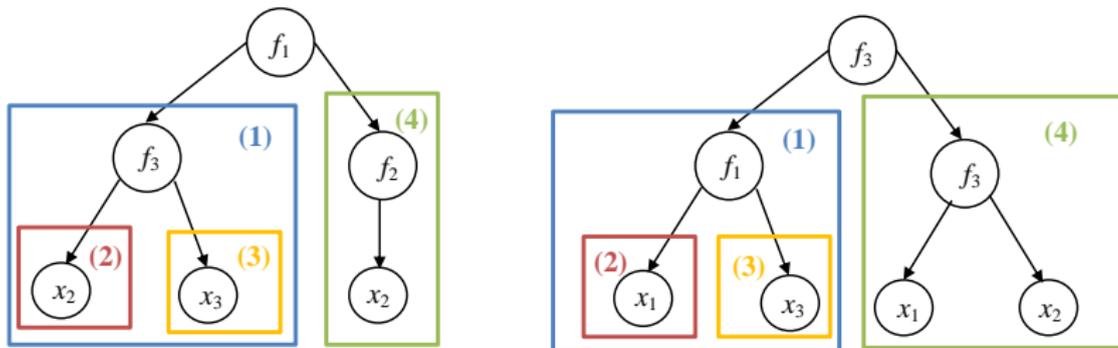
- 1 $p'_j = p_j^1$ or $p'_j = p_j^2 \quad \forall j = 1, \dots, l,$
- 2 for each $\bar{\mathbf{p}} \in \text{Sol}$ such that $\bar{p}_j = p_j^1$ or $\bar{p}_j = p_j^2 \quad \forall j$ the inequality holds

$$f(\mathbf{p}') \leq f(\bar{\mathbf{p}})$$

(in case of minimization problem).

^aA.V. Eremeev, J.V. Kovalenko. Optimal recombination in genetic algorithms for combinatorial optimization problems (2014)

¹Radcliffe, N.J.: The algebra of genetic algorithms (1994)



Оптимизированные операторы рекомбинации

Optimized one-point (O-OPX): выбор наилучшего среди 4 возможных потомков (1), (2), (3), (4).

Optimized uniform (O-UX): выбор наилучшего среди 2^3 возможных потомков (все возможные комбинации (2), (3), (4)).

Эксперимент на булевых деревьях

Дерево $T = (V, E)$

Листья содержат элементы из множества

$X = \{x_1, x_2, \dots, x_m\}, x_i \in \{0, 1\}, i = 1, 2, \dots, m.$

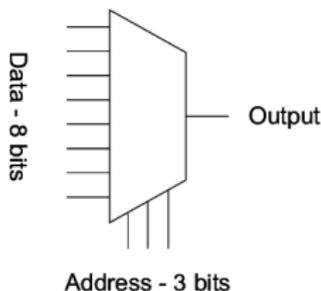
Базовые функции $\mathcal{F}_1 = \{\wedge, \vee, \neg\wedge, \neg\vee\},$

$\mathcal{F}_2 = \{\wedge, \vee, \neg\wedge, \neg\vee, \sim, XOR\},$

Тестовые примеры

Входные данные - вся таблица истинности

1. 6-multiplexor (6-mux).
2. even- n -parity (even- n).



6 - MUX

0 1 2 3
011011 → 0

11 - MUX

0 1 2 3 4 5 6 7
01111010001 → 1

Kalkreuth, R.: A comprehensive study on subgraph crossover in cartesian genetic programming. In: Proceedings of the 12th International Joint Conference on Computational Intelligence (IJCCI 2020). pp. 59–70 (2020).

Базовые функции AND, OR, NAND, NOR, XOR and EQ.
Среднее кол-во вычислений целевой функции.

Задача	LS + PM, O-OPX	(1+4)-CGP	(1+ λ)-CGP	Canonical-CGP	(μ + λ)-CGP
even-3	361	3177	2495	3107	1565
even-4	3547	15420	16523	54967	11135
even-5	7232	45542	34375	28413	43476

Doerr, B., Lissovoi, A., Oliveto, P.S.: (1+ 1) genetic programming with functionally complete instruction sets can evolve boolean conjunctions and disjunctions with arbitrarily small error. *Artificial Intelligence* 319, 103906 (2023).

Тестовые данные $AND_n = x_1 \wedge x_2 \wedge \dots \wedge x_n$ и
 $OR_n = x_1 \vee x_2 \vee \dots \vee x_n$.

Базовые функции AND, OR

Среднее кол-во вычислений целевой функции n_{obj} .

Задача	LS + PM, макс.глубина	LS + PM, n_{obj}	Doerr et al., n_{obj}
AND_4	2	100	46.3 – 51.2
AND_4	3	60	38.8 – 42.5
AND_4	5	32	38.9 – 39.1

Langdon, W.B.: Size fair and homologous tree crossovers for tree genetic programming. Genetic programming and evolvable machines 1, 95–119 (2000).

Базовые функции AND, OR, IF-THEN-ELSE, NOT.

Среднее кол-во вычислений целевой функции.

Задача	Init	LS + SM	Langdon
6-мих	RHH 2–6	16706	24000 – 38000
6-мих	RHH 5–8	20539	30000 – 44000

LS + PM, O-OPX.

Базовые функции $\{\wedge, \vee, \neg\wedge, \neg\vee\}$.

Задача	Потоки	Время запуска	Кол-во итераций	Кол-во вычисл. целевой ф-ции
even-4	1	14.77 мин	25 523	5 343 646
even-4	10	14.56 мин	21 398	4 499 204
even-4	50	10.88 мин	17 505	3 670 803
6-mux	1	4 мин	4 499	516 404
6-mux	10	4.6 мин	5 803	638 534
6-mux	50	3.55 мин	5 518	585 709

Базовые функции $\{\wedge, \vee, \neg\wedge, \neg\vee, \sim, XOR\}$.

Задача	Потоки	Время запуска	Кол-во итераций	Кол-во вычисл. целевой ф-ции
even-3	1	43 мсек	19.26	1091.6
even-3	5	44 мсек	15.7	774.67
even-3	10	51 сек	17.33	935.4
even-4	1	0.23 сек	46.87	3 967
even-4	5	0.2 сек	62.43	5 047
even-4	10	0.22 сек	53.6	4 745
even-5	1	0.9 сек	93	9 079
even-5	5	0.68 сек	113	10 290
even-5	10	0.65 сек	97	9 160

- 1 Параллелизация локального поиска
- 2 Параллелизация вычислений целевой функции
- 3 Оптимизированные операторы в Cartesian Genetic Programming
- 4 NP-трудность оптимальной рекомбинации для булевых функций

Спасибо за внимание!