СРАВНЕНИЕ НА НЯКОЛКО МОДИФИКАЦИИ НА МНОГОПОПУЛАЦИОНЕН ГЕНЕТИЧЕН АЛГОРИТЪМ

Калин КОСЕВ

Централна лаборатория по биомедицинско инженерство "Проф. Иван Даскалов"
Българска Академия на Науките
1113 София, ул. "Акад. Г. Бончев" № 105
E-mail: **kalinkosev@yahoo.com**

ABSTRACT
Four genetic algorithms (GA) are compared in this work – classical multi-population GA and three of its modifications. The proposed GA are tested for parameter identification of an E. coli non-linear fed-batch cultivation model. The obtained results show that the highest accuracy for parameter identification of the considered model is achieved with the multi-population GA with Modification 1. The optimization criterion value is 3.4939, obtained for time of 297.6343 s. The best calculation time is shown by the multi-population GA without mutation (Modification 2). The achieved minimal time of 195.36 s. is with slightly higher optimization criterion value – 4.0749.
Keywords*: Multi-population Genetic Algorithm, Modifications, Parameter Identification, E. coli Cultivation Process.*

УВОД

Най-популярният стохастичен метод за решаване на оптимизационни задачи е метода на еволюционното смятане, известен още като популационно базиран стохастичен метод. Това е клас от методи базирани на идеите на биологичната еволюция, която е водена от механизмите на репродукция, мутация и принципа на оцеляване на най-приспособения индивид. Едни от широко разпространените методи от групата на еволюционните методи са генетичните алгоритми (ГА). ГА представляват особен интерес поради тяхната универсалност и редицата предимства, които притежават [1]: лесно адаптиране към всеки проблем и големи възможности за настройка на алгоритмите посредством генетичните оператори и параметри.

С цел да се подобрят характеристиките на алгоритъма и качеството на полученото решение, както и за по-добра адаптация към конкретния оптимизационен проблем редица автори предлагат различни модификации и хибридизации на ГА. Известни са много ГА засягащи модификации на операторите на алгоритъма [2], допълнителни стъпки за филтриране на резултатите [3] и много др. [4, 5, 6, 7, 8]. В [9] за целите на параметрична идентификация на модели на ферментационни процеси е предложен модифициран многопопулационен генетичен алгоритъм без изпълнение на оператора *мутация*. Модификацията е направена, за да се подобри сходимостта и да се повиши ефикасността на идентификационната процедура при решаването на подобна задача с повишена сложност, произтичаща от спецификите на разглеждания проблем.

Цел на тази работа е изследване на характеристиките на клас ГА и избор на алгоритъм подходящ за приложение в реално време при решаване на задачи свързани с оптимизация и управление на биопроцеси. Сравнени са многопопулационен генетичен алгоритъм и три негови модификации. Разглежданите ГА са тествани при задача за параметрична идентификация на нелинеен модел на полупериодичен култивационен процес на бактерията *E. coli*. Изследвани са точността на полученото решението и сходимостта на алгоритмите.

ГЕНЕТИЧНИ АЛГОРИТМИ

Сравнени са четири генетични алгоритми:

- Класически многопопулационен генетичен алгоритъм (КМПГА);
- *Модификация 1*: многопопулационен генетичен алгоритъм с приложена модификацията предложена в [8];
- *Модификация 2*: многопопулационен генетичен алгоритъм с приложена модификацията предложена в [9];
- *Модификация 3*: многопопулационен генетичен алгоритъм с приложени модификации предложени в [8] и [9].

Класически многопопулационен генетичен алгоритъм

КМПГА имитира еволюцията на видовете по начин по-близък до естествената еволюция, отколкото еднопопулационния ГА. Субпопулациите при *КМПГА*, подобно на еднопопулационния, се развиват независимо една от друга за определен брой генерации. След изтичане на времето на изолация определен брой индивиди се разпределят между субпопулациите (*миграция*). Степента на генетично разнообразие и размяната на информация между субпопулациите се определят от процента на миграция, метода за избор на индивиди и схемата на миграция. Изборът на индивидите, които ще мигрират може да бъде постоянен, детерминиран или случаен. За всяка субпопулация се съставя набор от потенциални имигранти. От този набор, по случаен или точно определен начин, се определят индивидите, които ще мигрират. Както при еднопопулационния ГА и тук се изчисляват целева функция и функция на изява, извършват се операторите *кръстосване* и *мутация*. Циклите се изпълняват до достигане на поставеното условие за край на алгоритъма [1].

Модификация 1

Модифицираният мултипопулационен генетичен алгоритъм (*Модифицикация 1*) има структура, подобна на структурата на КМПГА [8]. Алгоритъмът генерира *n* на брой случайни решения, принадлежащи на зададената област на търсене. За генерираната популация се изчисляват целева функция и функция на изява. На следващите две стъпки се осъществяват две селекции на базата на функцията на изява. С първия избор се запазва дефиниран брой от “най-добрите” индивиди. С втория избор се определя броя на родителите, които ще сформират новата популация отново на базата на функцията на изява. На следващата стъпка операторите *кръстосване* и *мутация* се прилагат върху втория избор на индивиди, в резултат на което се получава ново поколение. Алгоритъмът изчислява функцията на изява на поколението. Популацията за последващо изпълнение се формира чрез заместване на старата популация със запазените “най-добри” хромозоми от първоначалната популация и хромозомите от новото поколение. Заместването става като индивидите с по-висока степен на изява изместват тези с по-малка степен на изява. По този начин се гарантира запазването на “най-добрите” решения без значение дали те са били в първоначалната или в новополучената популация. Сформираната по този начин популация се използва при следващия цикъл на алгоритъма. Алгоритъмът се изпълнява до достигане на зададения брой генерации.

Модификация 2

Разглежда се класически многопопулационен генетичен алгоритъм без *мутация* [9]. Алгоритъмът има структура, подобна на структурата на КМПГА. При реализирането му се изпълнява само оператора кръстосване. Изключването на оператора *мутация* води до значително подобряване на сходимостта на алгоритъма при конкретно разглежданата задача – параметрична идентификация на нелинеен модел на полупериодичен култивационен процес.

Модификация 3

Реализиран е многопопулационен ГА с разглежданите по-горе модификации – изключване на оператора *мутация* [9] и модификацията, предложена в [8].

ПОЛУПЕРИОДИЧЕН ПРОЦЕС НА КУЛТИВИРАНЕ НА *E. COLI*

Култивирането на рекомбинантните микроорганизми *E. coli* е често пъти най-икономичният начин за получаване на ценни фармацевтични продукти като инсулин, ваксини, хормони, ензими, растежни фактори и др. Математическият модел на процеса може да се представи със следната система нелинейни диференциални уравнения [10]:

$$\frac{dX}{dt} = \mu_{max} \frac{S}{k_S + S} X - \frac{F}{V} X \tag{1}$$

$$\frac{dS}{dt} = -\frac{1}{Y_{S/X}} \mu_{max} \frac{S}{k_S + S} X + \frac{F}{V}\left(S_{in} - S\right) \tag{2}$$

$$\frac{dA}{dt} = \frac{1}{Y_{A/X}} \mu_{max} \frac{S}{k_S + S} A - \frac{F}{V} X \tag{3}$$

$$\frac{dV}{dt} = F \tag{4}$$

,

където X е концентрация на биомаса, [g·l^{-1}]; S – концентрация на субстрат, [g·l^{-1}]; A – концентрация на ацетат, [g·l^{-1}]; S_{in} – концентрация на субстрат в подхранващия разтвор, [g·l^{-1}]; μ_{max} – максимална специфична скорост на растеж, [h^{-1}]; k_S – константа на насищане, [g·l^{-1}]; $Y_{S/X}$, $Y_{A/X}$ – икономични коефициенти, [g·g^{-1}]; F – скорост на подхранване, [l·h^{-1}]; V – работен обем на биореактора, [l].

За целите на параметричната идентификация са използвани реални експериментални данни от полупериодичен процес на култивиране на *E. coli MC4110*, представени в [11].

Критерият на оптимизационната задача има вида:

$$J = J_X + J_S + J_A \rightarrow min \tag{5}$$

$$J_X = \sum_{i=1}^{n}\left(X_{exp}(i) - X_{mod}(i)\right)^2;\ J_S = \sum_{i=1}^{n}\left(S_{exp}(i) - S_{mod}(i)\right)^2;\ J_A = \sum_{i=1}^{n}\left(A_{exp}(i) - A_{mod}(i)\right)^2.$$

където X_{exp}, S_{exp}, A_{exp} са експериментални данни за концентрацията на биомасата, субстрата и ацетата, X_{mod}, S_{mod}, A_{mod} – предсказаните от модела данни, n – брой данни.

Многопопулационния генетичен алгоритъм и предложените три модификации са тествани при задачата за параметрична идентификация на модел (1) – (4), при оптимизационен критерии (5).

ПОЛУЧЕНИ РЕЗУЛТАТИ

Разглежданите генетични алгоритми са реализирани в среда *Matlab 7.5*. Всички числени експериментите са извършени на платформа *Windows Vista*, с процесор Intel Core2Duo, 2.20GHz, 3.00GB DDR III RAM. Използван е *Genetic Algorithm Toolbox* [12] като са направени необходимите модификации в предложените там функции и алгоритми. Основните генетични оператори и параметри на реализираните алгоритми са представени в Таблица 1 [13].

Таблица 1. Генетични оператори и параметри

Оператор	**Вид**	**Параметър**	**Стойност**
кодиране	двоично	степен на кръстосване	0.70
кръстосване	двуточково	степен на мутация	0.1
мутация	битова инверсия	степен на миграция	0.20
селекция	принцип на рулетката	брой индивиди	100
функция на изява	линеен ранкинг	брой генерации	100
заместване	според функция на изява	брой субпопулации	5
		генерации за мигриране	20

Задачата за параметрична идентификация на модел (1) – (4) е решена на базата на реални експериментални данни [11]. С цел да се получи представителна извадка, поради стохастичния характер на ГА, всяка една идентификационна задача е повторена 10 пъти.

Получените резултати от четирите ГА – *КМПГА*, *Модификация 1*, *Модификация 2* и *Модификация 3* са обобщени в Таблица 2. Представени са числените стойности на оценяваните параметри на модела – μ_{max}, k_S, $Y_{S/X}$ и $Y_{A/X}$; стойността на оптимизационния критерий (*J*), както и на J_X, J_S и J_A. Следено е и изчислителното време необходимо за намиране на решение (*T*). Показани са усреднените стойности от 10-те процедури за всички следени параметри, резултатът получен за минимално изчислително време, както и резултатът получен за максимално изчислително време.

От представените числени стойности се вижда ясно, че *Модификация 1* намира решението за по-малко изчислително време в сравнение с *КМПГА* и по трите показатели (средна стойност, мин. време и макс. време). Тук е постигната и най-голяма точност на намереното решение. Получените резултати потвърждават, че предложената в [8] модификация предотвратява загуба на „най-добрата хромозома” и води до увеличаване на точността на решението.

Най-добро изчислително време дава *Модификация 2*, т.е. *КМПГА* без *мутация*, отново относно трите показатели (Таблица 2). Изключването на оператора *мутация* намалява значително изчислителното време. Постигнато е мин. време за намиране на решението от 195.36 s. Грешката в този случай е пренебрежимо по-голяма. Стойността на оптимизационният критерий при решението за мин. време е 4.0749 сравнена с тази получена с *Модификация 1* – 3.9090.

Таблица 2. Резултати от параметричната идентификация

ГА	Показател	μ_{max}	k_S	$1/Y_{S/X}$	$Y_{A/X}$	J_X	J_S	J_A	J	T, s
КМПГА	средно	0.5716	0.0422	2.0197	0.0136	2.3831	1.2583	0.0004	3.6418	316.0752
	мин. време	0.5760	0.0434	2.0192	0.0152	2.4876	1.2554	0.0004	3.7434	288.1806
	макс. време	0.5601	0.0388	2.0200	0.0158	2.4363	1.1442	0.0005	3.5810	359.8007
Модиф. 1	средно	0.5843	0.0459	2.0201	0.0129	2.4198	1.3892	0.0012	3.8102	277.8347
	мин. време	0.6000	0.0514	2.0197	0.0110	2.3131	1.5951	0.0008	3.9090	259.2737
	макс. време	0.5752	0.0430	2.0206	0.0193	2.2126	1.2794	0.0019	3.4939	297.6343
Модиф. 2	средно	0.5780	0.0445	2.0204	0.0170	2.4615	1.3631	0.0021	3.8267	203.0400
	мин. време	0.6102	0.0531	2.0199	0.0137	2.4631	1.6115	0.0003	4.0749	195.3601
	макс. време	0.6082	0.0535	2.0211	0.0127	2.3878	1.6648	0.0004	4.0530	214.7198
Модиф. 3	средно	0.5833	0.0458	2.0207	0.0170	2.4276	1.3830	0.0023	3.8130	240.6050
	мин. време	0.5958	0.0501	2.0202	0.0223	2.5629	1.5171	0.0043	4.0844	208.8073
	макс. време	0.5483	0.0352	2.0211	0.0169	2.4597	1.0401	0.0008	3.5005	264.4685

Изключването на оператора *мутация* и при предложената в [8] модификация на ГА - *Модификация 3*, намалява сходимостта на алгоритъма в сравнение с тази на *Модификация 1*

– от 259.27 s на 208.80 s. Предложените *Модификации 2* и *3* значително увеличават сходимостта на ГА, като в същото време нарастването на стойността на грешката е пренебрежимо малко.

ЗАКЛЮЧЕНИЕ

В тази работа са сравнени четири генетични алгоритъма – класически многопопулационен генетичен алгоритъм и три негови модификации. Предложените генетични алгоритми са тествани при задача за параметрична идентификация на нелинеен модел на полупериодичен култивационен процес на бактерията *E. coli*. На основата на извършените числени експерименти могат да се направят следните заключения за характеристиките на изследваните генетични алгоритми:

1. С най-голяма точност са намерени оценките на параметрите в разглеждания модел с използването на класически многопопулационен генетичен алгоритъм с модификация, предложена в [8] (*Модификация 1*). Стойността на оптимизационният критерий е 3.4939, получена за време 297.6343 s.

2. Най-добро изчислително време дава класически многопопулационен генетичен алгоритъм без *мутация* (*Модификация 2*). Постигнато е мин. време за намиране на решението от 195.36 s. със стойност на оптимизационния критерий – 4.0749.

В резултат от проведените експерименти и анализ на получените резултати многопопулационният генетичен алгоритъм без *мутация* (*Модификация 2*) е определен като подходящ за приложение в реално време при решаване на задачи свързани с оптимизация и управление на биопроцеси. Това е алгоритъмът с най-добра сходимост като едновременно с това, точността на полученият модел на полупериодичен култивационен процес на *E. coli* с съпоставима с точността на останалите решения.

Благодарности: Тази работа е финансирана от Фонд "Научни изследвания" по договори: ДМУ 02/4 и ДИД-02-29.

ЛИТЕРАТУРА

[1]. Goldberg, D. *Genetic algorithms in search, optimization and machine learning*. Addison-Weslcy Publishing Company, Reading, Massachusetts, 1989.

[2]. Heng Li, Peter E. D. Love, S. Ogunlana, Genetic algorithm compared to nonlinear optimization for labour and equipment assignment, *Building Research & Information*, 26(6), 1998, 322-329

[3]. Liu J., Tang L., Production planning, scheduling and control, A modified genetic algorithm for single machine scheduling, *Computers & Industrial Engineering*, 37(1-2), 1999, 43-46

[4]. Andre J., P. Siarry, T. Dognon, An improvement of the standard genetic algorithm fighting premature convergence in continuous optimization, *Advances in Engineering Software*, 32(1), 2001, 49-60.

[5]. Kim D. H., A. Abraham, J. H. Cho, A hybrid genetic algorithm and bacterial foraging approach for global optimization, *Information Sci.*, 177, 2007, 3918-3937.

[6]. Alsumait J. S., J. K. Sykulski, A. K. Al-Othman, A hybrid GA–PS–SQP method to solve power system valve-point economic dispatch problems, *Applied Energy*, 87, 2010, 1773-1781.

[7]. Rowe J., I. East, Direct replacement: A Genetic Algorithm without Mutation which Avoids Deception, *Lecture notes in Artificial Intel.*, Springer, 956, 1995, 41-48.

[8]. Roeva O., A Modified Genetic Algorithm for a Parameter Identification of Fermentation Processes, *Biotech. & Biotechnol. Equipm.* 20(1), 2006, 202-209.

[9]. Косев К., О. Роева, Многопопулационен генетичен алгоритъм с подобрена сходимост, *VIII Национална младежка научно-практическа сесия'10*, 10-11 май 2010, под печат.

[10]. Crueger, W., D. Grueger. *Biotechnology. A textbook of industrial microbiology*. Sci. Tech. Inc., 1984.
[11]. Arndt M., S. Kleist, G. Miksch, K. Friehs, E. Flaschel, J. Trierweiler, B. Hitzmann, A Feedforward-feedback Substrate Controller Based on a Kalman Filter for a Fed-batch Cultivation of *Escherichia coli* Producing Phytase, *Computers and Chemical Engineering*, 2004, 29(5), 1113-1120.
[12]. Chipperfield A. J., P. J. Fleming, *The Matlab Genetic Algorithm Toolbox*, 1995.
[13]. Roeva O., Multipopulation genetic algorithm: A tool for parameter optimization of cultivation processes models, *LNCS*, Springer, 4310, 2007, 255-262.