МОДИФИЦИРАН МНОГОПОПУЛАЦИОНЕН ГЕНЕТИЧЕН АЛГОРИТЪМ С ПОДОБРЕНА СХОДИМОСТ

Калин КОСЕВ, Олимпия РОЕВА

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

ABSTRACT
In this work a Multi-population genetic algorithm (MPGA) without the performing of the operator mutation is proposed. The idea is to reduce the convergence time and therefore to increase the identification procedure effectiveness. Considered MPGA without mutation is compared with classical MPGA for a parameter identification of an E. coli fed-batch cultivation model. It is attended to the optimization criterion value and to the time necessary for carrying out the calculations. The results show better work of the MPGA without mutation. The proposed algorithm finds the solution for time 55.7% less than MPGA with mutation. In the same time the accuracy of the obtained mathematical model is comparable with the one obtained from MPGA with mutation.
Keywords*: Multi-population Genetic Algorithm, Mutation, Identification, Non-linear model, Cultivation Process.*

УВОД

Широкото приложение на различните видове генетични алгоритми (ГА) е продиктувано от техните основни предимства пред конвенционалните методи за оптимизация, а именно [5]:

- извършват оптимизация на базата на предварително дефинирана потребителска функция;
- работят в кодирано параметрично пространство и извършват търсенето в множество от точки;
- работят на базата на вероятностни преходни правила.

Класическият ГА (еднопопулационен ГА) е универсална вероятностна техника на търсене, ефикасна и работоспособна за широк кръг от проблеми. Една възможност за подобряване на качеството на решението е въвеждането на множество от популации – т. нар. многопопулационен ГА (МПГА). МПГА имитира еволюцията на видовете по начин по-близък до естествената еволюция, отколкото еднопопулационния ГА.

Приложението на ГА за оптимизация на биопроцеси е представено още през 1996 [1]. Оттогава ГА навлизат широко в областта на оптимизацията и управлението на биопроцеси [2, 3, 4].

Мотивирани от желанието да подобрят характеристиките на алгоритъма, както и адаптацията му за ефективно решаване на поставената задача редица автори предлагат различни модификации и хибридизации [6, 7, 8]. В [9] са предложени ГА с хибридна мутация. Представеният ГА прилага стохастична процедура „hill climbing" в процеса на мутация. В [10] е реализиран ГА без мутация. Авторите предлагат техника наречена „директно заместване", без включване на мутация, с цел да се предотврати загубата на алели, за разглеждания от тях проблем.

Цел на тази работа е да се намери модификация на ГА, която води до подобряване на сходимостта на алгоритъма за конкретно решаваната задача. Сходимостта на даден

алгоритъм е от съществени значение при приложението му за решаване на проблеми в реално време, каквито са и задачите за управление на индустриални биопроцеси.

В тази работа е представен модифицирам многопопулационен ГА (ММПГА), при който се изпълнява само оператора *кръстосване*. За подобряване на сходимостта на алгоритъма е изключен оператора *мутация*. С цел да се покаже, че не се губи точността на получените решения с ММПГА алгоритъмът е сравнен с класически МПГА (КМПГА). Двата ГА са тествани при задача за параметрична идентификация на нелинеен модел на полупериодичен култивационен процес. Разгледан е реален процес на култивиране на бактерията *E. coli*, един от най-широко използваните микроорганизми за биологичен синтез.

МНОГОПОПУЛАЦИОНЕН ГЕНЕТИЧЕН АЛГОРИТЪМ БЕЗ МУТАЦИЯ

Значимостта на оператора *мутация* е все още въпрос на разискване. Някои автори смятат, че този оператор играе второстепенна роля в ГА. Други, определят мутацията като възможност за предотвратяване на попадане на решението в локален максимум. Влиянието на оператора *мутация*, както и на всеки един оператор на ГА, основно се определя от конкретно решаваната задача. За разглежданата в тази статия задача – параметрична идентификация на нелинеен модел на полупериодичен култивационен процес, оператора *мутация* може да бъде изключен, което води до подобрена сходимостта на МПГА, без загуба на точността на решението. По този начин е реализиран ММПГА без изпълнение на оператора *мутация*.

Предложеният ММПГА работи по-подобен на КМПГА начин. Субпопулациите при МПГА се развиват независимо една от друга за определен брой генерации. След изтичане на времето на изолация даден брой индивиди се разпределят между субпопулациите. Това е така наречената *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 [5]. Изборът на индивидите, които ще мигрират може да бъде постоянен, детерминиран (на базата на *функцията на изява*) или случаен. Съществуват много възможни структури на миграция на индивидите между субпопулациите. Най-общата стратегия за миграция е неограничената миграция. В този случай индивидите могат да мигрират от всяка една субпопулация в друга. За всяка субпопулация се съставя набор от потенциални имигранти. От този набор, по случаен или детерминиран начин, се определят индивидите, които ще мигрират. Тук се изчисляват *целевата функция* и *функцията на изява* и се извършва операцията *кръстосване*. Програмата се изпълняват до достигане на поставеното условие за край на алгоритъма.

Схематично ММПГА може да се представи както следва:

1. [***Старт***] Генериране на *к* на брой случайни субпопулации, всяка от които с *n* хромозоми.
2. [***Целева функция***] Изчисляване на *целевата функция* за всяка хромозома *n* в субпопулациите.
3. [***Функция на изява***] Изчисляване на *функцията на изява* за всяка хромозома *n* в субпопулациите.
4. [***Нова популация***] Създаване на нови популации чрез:
   4.1. [***Селекция***] Избиране на родителски хромозоми от субпопулациите съобразно тяхната *функция на изява*.
   4.2. [***Кръстосване***] Кръстосване на избраните родители и формиране на ново поколение.
5. [***Заместване***] Поставяне на полученото поколение в старата субпопулация и използване на новополучената субпопулация за последващо изпълнение на алгоритъма.
6. [***Мигриране***] Мигриране на индивидите между субпопулациите, след изтичане на времето на изолация.

7. [***Проверка***] Спиране на алгоритъмът и връщане на “най-доброто” решение сред популациите ако крайното условие е удовлетворено. В противен случай алгоритъмът продължава.
8. [***Цикъл***] Повтаряне на **Стъпка 2**.

Операторите и параметрите, които се дефинират за представеният МПГА без мутация са обобщени в Таблица 1.

Таблица 1. Оператори и параметри на ММПГА

Оператор	**Вид**	**Параметър**	**Стойност**
кодиране	двоично	степен на кръстосване	0.70
кръстосване	двуточково	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

При КМПГА са използвани операторите и параметрите представени в Таблица 1, като степента на мутация е 0.1.

Като подходящи оператори и параметри различни автори предлагат различни решения, в зависимост от поставения проблем. Дефинираните в тази работа оператори и стойности на параметрите на ГА се базират на предишни изследвания за решаванана тук задача за параметрична идентификация [11].

ПАРАМЕТРИЧНА ИДЕНТИФИКАЦИЯ НА НЕЛИНЕЕН МОДЕЛ НА ПОЛУПЕРИОДИЧНА КУЛТИВАЦИЯ 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 полупериодичен процес на култивиране на *E. coli* може да се представи със следната система нелинейни диференциални уравнения [12]:

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

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

$$\frac{dV}{dt} = F\,, \tag{3}$$

където X е концентрация на биомаса, [g/l]; S – концентрация на субстрат, [g/l]; S_{in} – концентрация на субстрат в подхранващия разтвор, [g/l]; μ_{max} – максимална специфична скорост на растеж, [h^{-1}]; k_S – константа на насищане, [g/l], $k_S > 0$; $Y_{S/X}$ – икономичен коефициент, [-], $Y_{S/X} > 0$; F – скорост на подхранване, [l/h]; V – работен обем, [l], $V > 0$.

За целите на параметричната идентификация са използвани реални експериментални данни от полупериодичен процес на култивиране на *E. coli MC4110*, представени в [13].

Критерият на оптимизационната задача има вида:

$$J = J_X + J_S + J_A \to min \tag{4}$$

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

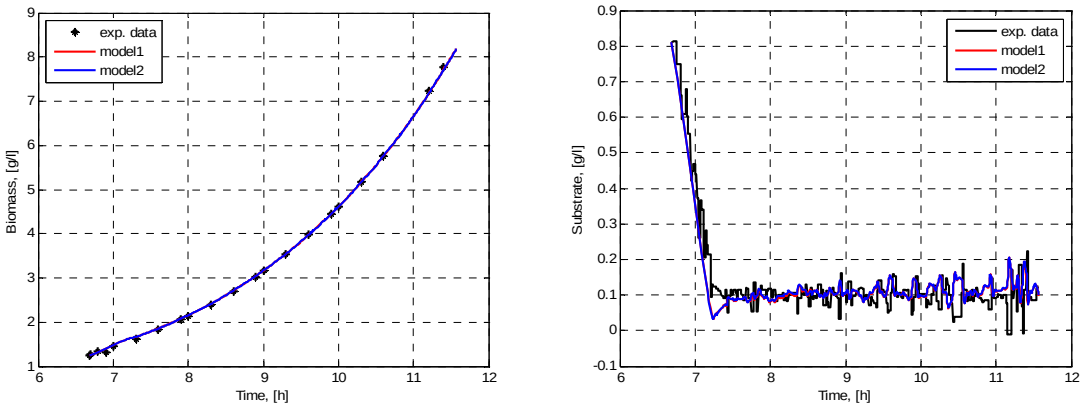
където X_{exp}, S_{exp} са експериментални данни за концентрацията на биомасата, субстрата и ацетата, X_{mod}, S_{mod} – предсказаните от модела данни, n – брой данни.

За решаване на поставената задача бяха изпълнени голям брой експерименти с двата ГА – КМПГА и ММПГА. Всички числени експериментите бяха извършени в среда *Matlab 7.5*, с процесор Intel Core2Duo, 2.20GHz, 3.00GB DDR III RAM. С цел да се получи представителна извадка, поради стохастичния характер на ГА, всяка една идентификационна задача е повторена 10 пъти. Представените в статията резултати са усреднените стойности от 10-те процедури. В Таблица 2 са показани получените стойности на параметрите на модела (μ_{max}, k_S и $Y_{S/X}$), на критерия (J), включително за J_X и J_S, както и времето за намиране на решението (T). Представени са усреднените стойности и резултатите получени при минимално и максимално време T – от 10-те процедури.

Таблица 2. Резултати от параметричната идентификация

	Параметри			**Критерии**			**Време**
	μ_{max}	k_S	$1/Y_{S/X}$	J_X	J_S	J	T, s
средно КМПГА	0.5716	0.0422	2.0197	2.3831	1.2583	3.6418	316.0752
средно ММПГА	0.5780	0.0445	2.0204	2.4615	1.3631	3.8267	203.0400
min КМПГА	0.5760	0.0434	2.0192	2.4876	1.2554	3.7434	288.1806
min ММПГА	0.6102	0.0531	2.0199	2.4631	1.6115	4.0749	195.3601
max КМПГА	0.5601	0.0388	2.0200	2.4363	1.1442	3.5810	359.8007
max ММПГА	0.6082	0.0535	2.0211	2.3878	1.6648	4.0530	214.7198

На Фигура 1 графично е представено сравнението между експерименталните данни и предсказанията на моделите, получени с ММПГА и КМПГА. Показани са основните променливи на процеса – X и S.



Фигура 1. Сравнение между експериментални данни и модел
Модел 1 получен с КМПГА и *Модел 2* получен с ММПГА

Представените резултати ясно показват, че получените математически модели (*Модел 1* получен с МПГА с мутация и *Модел 2* получен с МПГА без мутация) са идентични относно моделирането на динамиката на X и много близки относно динамиката на S. Тези резултати се потвърждават и от подобните оценки на параметрите, показани в Таблица 2.

Единствената съществена разлика в резултатите за двата модела е времето за намиране на решението. КМПГА намира оценките на параметрите за време 316.08 s., докато ММПГА за 203.04 s. С предложената модификация се подобрява с 55.7% сходимостта на МПГА, като в същото време се запазва точността на решението.

ЗАКЛЮЧЕНИЕ

В тази работа е предложена модификация на МПГА без изпълнение на оператора *мутация*. Идеята е да се подобри сходимостта на алгоритъмът и да се повиши ефикасността на идентификационната процедура. Разгледан е МПГА като по-добра алтернатива на еднопопулационния. МПГА представя по-близка интерпретация на естествената еволюция на видовете в сравнение с еднопопулационния. С цел намаляване на времето за е изпуснат оператора *мутация*. Предложената модификация е сравнена с класически МПГА при задача за параметрична идентификация на модел на полупериодичен процес на култивиране на *E. coli*. Проследена е промяната на грешката и времето необходимо за изчисляване на решенията. Получените резултати показват значително подобряване на сходимостта на ММПГА. Решението е намерено за време с 55.7% по-малко от времето необходимо на КМПГА. Едновременно с това полученият с ММПГА математически модел е съпоставим по точност с този, получен с КМПГА.

Благодарности: *Тази работа е финансирана от Фонд "Научни изследвания" по проекти: ДМУ 02/4 "Висококачествено управление на биотехнологични процеси с приложение на модифицирани конвенционални и метаевристични методи" и ДИД-02-29 "Моделиране на процеси с фиксирани правила за развитие".*

ЛИТЕРАТУРА

[1] Morriyama H., K. Shimizu, On-line optimization of culture temperature for ethanol fermentation using genetic algorithm, *J. Chem. Tech. Biotechnol*, 66, 1996, 217-222.

[2] Al-Duwaish H. N., A Genetic Approach to the Identification of Linear Dynamical Systems with Static Nonlinearities, *Int. J. of Systems Science*, 31(3), 2000, 307-313.

[3] Na J.-G., Y. K. Chang, B. H. Chung, H. C. Lim, Adaptive Optimization of Fed-batch Culture of Yeast by using Genetic Algorithms, *Bioprocess and Biosystems Engineering*, 24, 2002, 299-308.

[4] Roeva O., St. Tzonkov, A Genetic Algorithm for Feeding Trajectory Optimization of Fed-batch Fermentation Processes, *Int. J. Bioautomation*, 12, 2009, 1-12.

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

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

[8] Roeva O., A Modified Genetic Algorithm for a Parameter Identification of Fermentation Processes, *Biotechnology and Biotechnological Equipment*, 20(1), 2006, 202-209.

[9] Katayama K., H. Sakamoto, The Efficiency of Hybrid Mutation Genetic Algorithm for the Travelling Salesman Problem, *Math. and Comp. Modelling*, 31, 2000, 197-203.

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

[11] Roeva O., Tuning of genetic algorithm for parameter identification of fed-batch cultivation process, *Fifth National Young Scientific-Practical Session*, Bulgaria, May 21-22, 2007, 94-98.

[12] Crueger, W., D. Grueger. *Biotechnology. A textbook of industrial microbiology*. Sci. Tech. Inc., 1984.

[13] 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.