ИЗСЛЕДВАНЕ ВЛИЯНИЕТО НА ПОСЛЕДОВАТЕЛНОСТТА НА ИЗПЪЛНЕНИЕ НА ОПЕРАТОРИТЕ НА ГЕНЕТИЧНИТЕ АЛГОРИТМИ ЗА ПАРАМЕТРИЧНА ИДЕНТИФИКАЦИЯ НА ФЕРМЕНТАЦИОНЕН ПРОЦЕС

Мария АНГЕЛОВА, Таня ПЕНЧЕВА

E-mails: **maria.angelova@clbme.bas.bg, tania.pencheva@clbme.bas.bg**

РЕЗЮМЕ

Изследвани са четири вида стандартни генетични алгоритми с различна последователност на изпълнение на основните генетични оператори мутация, кръстосване и селекция. Влиянието на последователността на генетичните оператори върху възможността за намиране на оптимално решение, както и върху времето за достигането му, е изследвана за целите на параметрична идентификация на ферментационен процес на дрождите Saccharomyces cerevisiae. Различните видове генетични алгоритми водят до сходни стойности на оптимизационния критерий, но генетичният алгоритъм с последователност на операторите мутация, кръстосване, селекция е чувствително по-бърз от останалите.

Ключови думи*: Генетични алгоритми, Кръстосване, Мутация, Селекция, Параметрична идентификация, Ферментационен процес, Saccharomyces cerevisiae.*

Въведение

Ферментационните процеси (ФП) като обекти за моделиране и висококачествено управление се характеризират с взаимозависими и променящи се във времето процесни променливи. Поради тези и редица други специфични особености ФП се описват с нелинейни модели с много комплексна структура. Важна стъпка за адекватното моделиране на ФП е изборът на подходяща оптимизационна процедура за параметрична идентификация на модела. Конвенционалните оптимизационни методи не могат да преодолеят спецификата на ФП и да достигнат до желано решение. Ето защо като алтернатива се използват генетичните алгоритми, които са стохастичен глобален оптимизационен метод. Тяхното приложение е предпоставка за робастност и достигане до глобален минимум, което ги прави много подходящи за параметрична идентификация на модели на ФП.

Генетичните алгоритми (ГА) като метод за оптимизация са директна, случайна търсеща техника за намиране на глобално решение в комплексното многомерно пространство на търсене [1]. От биологична гледна точка ГА са базирани на механиката на естествената селекция и генетика, съобразно еволюционната теория на Дарвин. Те успешно се използват в редица области на науката за решаване на инженерни и оптимизационни задачи и са доказани като много подходящи при оптимизация на силно нелинейни проблеми. Именно поради тази причина ГА се прилагат като особено надеждни за параметрична идентификация на модели на ФП [2, 3, 4, 5].

Стандартният генетичен алгоритъм (СГА) [1] е с последователност на изпълнение на генетичните операторите *селекция*, *кръстосване*, *мутация* и в настоящата разработка ще бъде означен като СГА-СКМ. Известни са и редица варианти на СГА [4, 6, 7], като например представеният в [4] модифициран генетичен алгоритъм, при който операторът *селекция* се изпълнява след операторите *кръстосване* и *мутация* и който в настоящата разработка ще бъде означен като СГА-КМС. Този алгоритъм предотвратява загубата на намерено добро решение от *кръстосването*, *мутацията* или и от двата оператора. СГА-КМС, приложен за параметрична идентификация на модел на полупериодичен ферментационен процес за бактериите *Escherichia coli*, намалява стойността на оптимизационния критерий, както и

времето за достигане до решение. Постигнатите обнадеждаващи резултати са предпоставка за търсене на други подобрения в СГА.

Обикновено в генетичните алгоритми операторът *мутацията* се изпълнява след оператора *кръстосване*. Правейки аналогия с протичащите в природата процеси може да се каже, че вероятността да се осъществи първо мутация в индивидите, след което те да бъдат кръстосани е съпоставима с тази двата процеса да се случват в обратна последователност. Цел на настоящата работа е да се изследва влиянието на последователността на изпълнение на генетичните оператори *кръстосване* и *мутация* при СГА. Споменатите вече СГА-СКМ и СГА-КМС са сравнени с новопредложени техни варианти с разменена последователност на операторите *кръстосване* и *мутация*. Получените общо четири варианта на СГА са сравнени по отношение на точност и бързодействие за целите на параметрична идентификация на модел на полупериодична култивация на дрождите *S. cerevisiae.*

Стандартни генетични алгоритми с размяна на последователността на изпълнение на операторите *мутация* и *кръстосване*

В настоящата разработка е представена модификация на стандартния генетичен алгоритъм [1], която се състои в размяна на последователността на изпълнение на операторите *мутация* и *кръстосване*. В резултат се получава нов алгоритъм СГА-СМК, чието действие е представено по-долу накратко. В началото се създава първична популация от *n* хромозоми и се изчислява обектната стойност на всеки индивид в началната популация. Следва подреждане и селектиране на най-подходящите индивиди в поколението, след което избраните индивиди се репродуцират. По време на репродукцията в представения тук СГА-СМК първо се изпълнява *мутацията*, а след това *кръстосването*. В термините на ГА *мутация* означава произволна промяна на стойността на ген в поколението. При *кръстосването* гените от родителите се комбинират с цел да се формира изцяло нова хромозома. По този начин се получават нови индивиди, като стремежът е те да наследят възможно най-добрата комбинация от характеристики на родителите. Накрая се изчислява обектната стойност на индивидите и се извършва замяна на част от индивидите с новополучените в това поколение. Алгоритъмът спира когато се удовлетвори някой от критериите за прекратяването му, като в разглеждания случай това е изпълнението на определен брой генерации. Фиг. 1 представя Matlab кода за гореописания СГА-СМК. Предложената размяна на последователността на изпълнение на операторите *мутация* и *кръстосване* е приложена и към СГА-КМС, в резултат на което е получена и още една модификация на СГА, а именно СГА-МКС. Matlab кода на СГА-МКС не е включен в настоящата разработка, но лесно може да бъде генериран на базата на представения на Фиг. 1 код.

Параметрична идентификация на полупериодична култивация на *S. cerevisiae* чрез приложение на четирите варианти на СГА

Дрождите са важни микроорганизми, с широк спектър на приложение в хлебната и бирената индустрия, генното инженерство и др. Следователно култивацията на дрожди, както и последващата обработка на данните от измерванията по време на и след самия процес е от изключително значение. В настоящото изследване са използвани експериментални данни от полупериодична култивация на *S. cerevisiae*, проведена в Института по техническа химия в Хановер, Германия. Култивацията на дрождите е осъществена в 2 l реактор, използвайки среда на Шацман [8]. Концентрацията на глюкозата в подхранващия разтвор е 35 $g.l^{-1}$. Температурата се поддържа на 30°C, pH – на 5.5. Скоростта на разбъркване е 1200 rpm, а скоростта на аериране се поддържа на 300 $l.h^{-1}$. Субстратът (глюкоза) и разтвореният кислород се измерват он-лайн, докато биомасата и етанола – офлайн.

```
% Initialise population
Chrom = crtbp(NIND, NVAR*PRECI);
% Counter
gen = 0;
% Evaluate initial population
ObjV = objfun1(bs2rv(Chrom, FieldDR));

% Generational loop
while gen < MAXGEN,
        % Assign fitness values to entire population
        FitnV = ranking(ObjV);
        % Select individuals for breeding
        SelCh = select('rws', Chrom, FitnV, GGAP);
        % Apply mutation
        SelCh = mut(SelCh);
        % Recombine individuals (crossover)
        SelCh = recombin('xovdp', SelCh, 0.7);
        % Evaluate offspring, call objective function
        ObjVSel = objfun1(bs2rv(SelCh, FieldDR));
        % Reinsert offspring into population
        [Chrom ObjV] = reins(Chrom, SelCh, 1, 1, ObjV, ObjVSel');
        % Increment counter
        gen = gen + 1;
end
```

Фиг. 1 Matlab код на СГА-СМК

Математическият модел за култивация на *S. cerevisiae*, получен на базата на баланс на масите, се представя в следния вид [9]:

$$\frac{dX}{dt} = \mu X - \frac{F}{V} X \tag{1}$$

$$\frac{dS}{dt} = -q_S X + \frac{F}{V}\left(S_{in} - S\right) \tag{2}$$

$$\frac{dE}{dt} = q_E X - \frac{F}{V} E \tag{3}$$

$$\frac{dO_2}{dt} = -q_{O_2} X + k_L^{O_2} a \left(O_2^* - O_2\right) \tag{4}$$

$$\frac{dV}{dt} = F \tag{5}$$

където S, X, E и O_2 са непрекъснати и диференцируеми функции, описващи съответно концентрациите на субстрат [g.l^{-1}], биомаса [g.l^{-1}], етанол [g.l^{-1}] и разтворен кислород [%]; O_2^* – концентрация на насищане на разтворен кислород [%]; $k_L^{O_2} a$ – масообменен коефициент [h^{-1}]; S_{in} – концентрация на субстрат в подхранващия разтвор [g.l^{-1}]; μ, q_S, q_E, q_{O_2} – специфични скорости на растеж/консумация съответно на биомаса, субстрат, етанол и разтворен кислород
[h^{-1}]; F – скорост на подхранване [l.h^{-1}]; V – обем на биореактора [l].

За разглежданата култивация специфичните скорости на растеж/консумация в модела (1)-(5) имат следния вид [9]:

$$\mu = \mu_{2S} \frac{S}{S + k_S} + \mu_{2E} \frac{E}{E + k_E},\ q_S = \frac{\mu_{2S}}{Y_{SX}} \frac{S}{S + k_S},\ q_E = -\frac{\mu_{2E}}{Y_{EX}} \frac{E}{E + k_E},\ q_{O_2} = q_E\ Y_{OE} + q_S Y_{OS} \tag{6}$$

където μ_{2S}, μ_{2E} са максималните стойности на скоростите на растеж [h^{-1}]; k_S, k_E – константи на насищане [g.l^{-1}]; Y_{ij} – икономични коефициенти [g.g^{-1}].

Специфичните скорости, отразени чрез зависимости (6), отговарят на *състояние на смесено оксидиране* [9]. Процесът на култивация на дрожди навлиза в това състояние, когато концентрацията на субстрат спадне под критичното ниво. Специфичната скорост на растеж на биомасата се представя като сума от две събираеми, описващи дрождевия растеж

съответно за сметка на консумацията на субстрат и на етанол. И двете събираеми имат структурата на кинетика на Моно, която се използва и за специфичните скорости на консумацията на етанол и на субстрат. Специфичната скорост на консумация на кислород се представя като сума от пропорционални отношения спрямо специфичните скорости на консумацията на етанол и на субстрат.

Като оптимизационен критерий в настоящото изследване е използвано средното квадратичното отклонение между изхода на модела и експерименталните данни, получени по време на култивацията.

$$J_Y = \sum \left(Y - Y^* \right)^2 \rightarrow \min , \qquad (7)$$

където Y са експерименталните и Y^* - моделно предсказаните данни, $Y = [X, S, E, O_2]$.

Параметричната идентификация на модела (1)-(5) е проведена в среда *Matlab* с използването на *Genetic Algorithm Toolbox* [10, 11, 12]. Генетичните параметри и операторите на СГА са настроени съгласно предходно изследване на авторите [13].

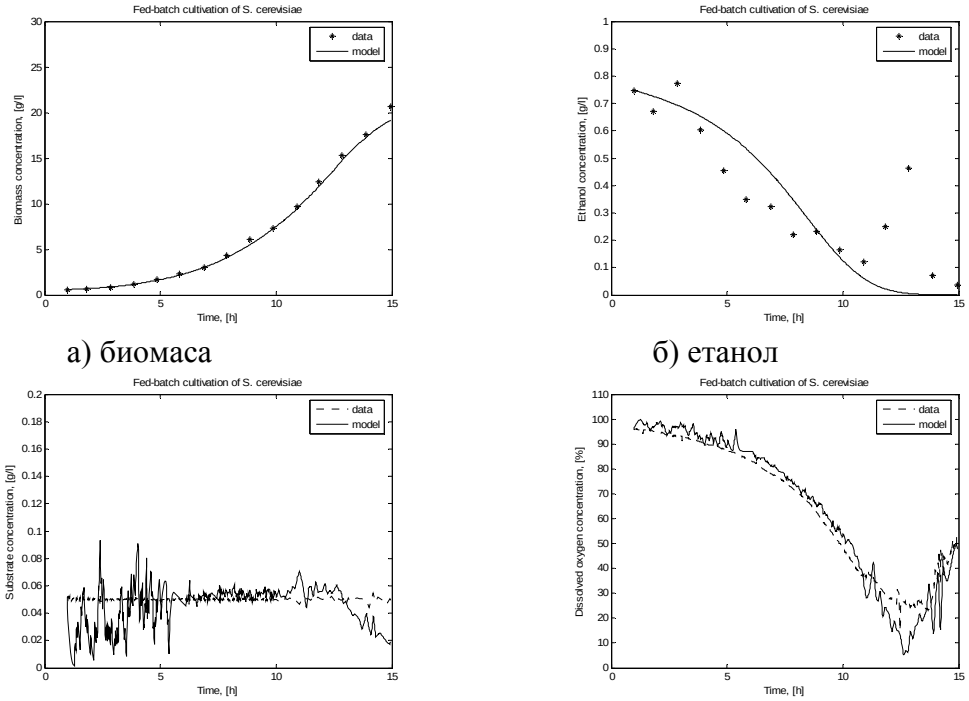
Параметричната идентификация е проведена чрез приложение на четирите варианта на СГА, обект на настоящото изследване: СГА-СКМ [1], предложеният тук СГА-СМК с изпълнение на оператора *мутация* преди оператора *кръстосване*, СГА-КМС [4] и предложената тук негова модификация СГА-МКС отново с разменена последователност на операторите *мутация* и *кръстосване*. Сравнението между четирите варианта на СГА е представено в Таблица 1.

Таблица 1. Резултати от параметричната идентификация на полупериодична култивация на *S. cerevisiae*

Параметър	***СГА-СКМ***	***СГА-СМК***	***СГА-КМС***	***СГА-МКС***
J_Y	*0.0223*	*0.0221*	*0.0225*	*0.0223*
CPU time, s	*73.8281*	*73.4688*	*64.8281*	*59.5156*
μ_{2S}, h^{-1}	*0.9616*	*0.9038*	*0.9211*	*0.9119*
μ_{2E}, h^{-1}	*0.0971*	*0.1320*	*0.0872*	*0.0966*
k_S, $g.l^{-1}$	*0.1154*	*0.1119*	*0.1176*	*0.1109*
k_E, $g.l^{-1}$	*0.7963*	*0.7990*	*0.7620*	*0.7987*
Y_{SX}, $g.g^{-1}$	*0.4279*	*0.4072*	*0.4279*	*0.4316*
Y_{EX}, $g.g^{-1}$	*1.2898*	*1.7699*	*1.2898*	*1.3170*
k_{la}^{o2}, h^{-1}	*38.5895*	*116.4160*	*127.2898*	*141.1076*
Y_{OS}, $g.g^{-1}$	*313.8285*	*898.6292*	*989.8014*	*993.2537*
Y_{OE}, $g.g^{-1}$	*234.7797*	*281.1797*	*62.6547*	*166.6377*

Както се вижда от Таблица 1, стойностите на оптимизационния критерий, получени при приложението на четирите вида стандартни генетични алгоритми, са твърде сходни, което гарантира, че от гледна точка на точност изпълнението на оператора *мутация* преди оператора *кръстосване* не води до намаляване степента на адекватност на модела. От гледна точка на бързодействие обаче, изпълнението на оператора *мутация* преди оператора *кръстосване* води до намаляване времето за достигане до глобален минимум. Докато приложението на СГА-СМК не намалява чувствително (<1%) времето за достигане до решение в сравнение със СГА-СКМ, то приложението на СГА-МКС води до намаляване на времето спрямо СГА-КМС с 9%. И при дрождите *S. cerevisiae* се потвърди резултатът, представен в [4] за бактериите *E. coli*, а именно, че изпълнението на оператора *селекция* след операторите *кръстосване* и *мутация* чувствително намалява времето за намиране на оптимално решение. Най-високо бързодействие се постига при СГА-МКС, което е с 24% по-малко от времето, за което СГА-СКМ достига до глобален минимум, при това без загуба на точност на решението. Следователно, в резултат на представеното сравнение, последователността на изпълнение на операторите *мутация*, *кръстосване*, *селекция* е най-оптимална по отношение на бързодействие със запазена висока степен на адекватност на решението.

Поради сходство на резултатите тук са представени само тези, получени с новопредложения и доказан като най-оптимален по отношение на бързодействие и точност стандартен генетичен алгоритъм, с изпълнение на операторите в последователност *мутация*, *кръстосване*, *селекция*. Фиг. 2 представя резултатите от експерименталните и моделно предсказаните данни, съответно за биомаса, етанол, субстрат и разтворен кислород.



а) биомаса б) етанол

в) субстрат г) разтворен кислород

Фиг. 2 Експериментални и моделни данни за концентрациите на основните биохимични променливи

Резултатите от сравнението на новопредложените СГА-СМК и СГА-МКС със СГА-СКМ и СГА-КМС за целите на параметрична идентификация на полупериодична култивация на *S. cerevisiae* показват ефективността на размяната на последователността на изпълнение на операторите *мутация* и *кръстосване* по отношение на бързодействие.

Заключение

В настоящото изследване са предложени две модификации на стандартни генетични алгоритми с размяна на последователността на изпълнение на операторите *мутация* и *кръстосване*. Новопредложените СГА-СМК и СГА-МКС са сравнени със СГА-СКМ и СГА-КМС за целите на параметрична идентификация на полупериодична култивация на *S. cerevisiae*. Получените резултати показват, че приложението на операторите в последователност *мутация*, *кръстосване*, *селекия* води до чувствително подобряване на бързодействието на алгоритъма, без това да се отразява на степента на адекватност на решението. Приложението на СГА-МКС води до намаляване на времето за намиране на решение с 9% спрямо СГА-КМС и 24% спрямо времето на СГА-СКМ.

Благодарности

Изследванията са частично финансирани по проект BG051PO001-3.3.04/40 по ОП "Развитие на човешки ресурси" на ЕСФ и МОМН, и по договор ДИД 02-29 "Моделиране на процеси с фиксирани правила за развитие", финансиран от ФНИ.

Литература

1. Goldberg D., Genetic Algorithms in Search, Optimization and Machine Learning, Addison-Wiley Publishing Company, Massachusetts, 1989.

2. Carrillo-Ureta G. E., P. D. Roberts, V. M. Becerra, Genetic Algorithms for Optimal Control of Beer Fermentation, Proceeding of the 2001 IEEE International Symposium on Intelligent Control, Mexico City, Mexico, 2001, 391-396.
3. Ranganath M., S. Renganathan, C. Gokulnath, Identification of Bioprocesses Using Genetic Algorithm, Bioprocess Engineering, 1999, 21, 123-127.
4. Roeva O., A Modified Genetic Algorithm for a Parameter Identification of Fermentation Processes, Biotechnol. & Biotechnol., 2006, 20, 202-209.
5. Roeva O., Genetic Algorithm for a Parameter Estimation of a Fermentation Process Model: A Comparison, Bioautomation, 2005, 3, 19-28.
6. Cordon O., F. Herrera, Hybridizing Genetic Algorithms with Sharing Scheme and Evolution Strategies for Designing Approximate Fuzzy Rule-based Systems, Fuzzy Sets and Systems, 2001, 118, 235-255.
7. Kuo R. J., C. H. Chen, Y. C. Hwang, An Intelligent Stock Trading Decision Support System through Integration of Genetic Algorithm based Fuzzy Neural Network and Artificial Neural Network, Fuzzy Sets and Systems, 2001, 118, 21-45.
8. Pencheva T., O. Roeva, I. Hristozov, Functional State Approach to Fermentation Processes Modelling, Tzonkov St., B. Hitzmann (Eds.), Prof. Marin Drinov Academic Publishing House, Sofia, 2006.
9. Zhang X.-C., A. Visala, A. Halme, P. Linco, Functional State Modelling Approach for Bioprosesses: Local Models for Aerobic Yeast Growth Processes, J. Proc. Contr., 1994, 4(3), 127-134.
10. Chipperfield A. J., P. Fleming, H. Pohlheim, C. M. Fonseca, Genetic Algorithm Toolbox for Use with MATLAB, User's Guide, Version 1.2. Dept. of Automatic Control and System Engineering, University of Sheffield, UK, 1994.
11. Chipperfield A. J., P. Fleming, The Matlab Genetic Algorithm Toolbox, 1995.
12. MatWorks Inc. Genetic Algorithm Toolbox, User's Guide, 1999.
13. Angelova M., St. Tzonkov, T. Pencheva, Parameter Identification of a Fed-batch Cultivation of *S. cerevisiae* using Genetic Algorithms, Serdica Journal of Computing, 2010, 4(1), 11-18.