A Further Result on the Schema Theorem
of Partheno-Genetic Algorithm *

LI Maojun

(College of Electrical & Information Engineering, Hunan University · Changsha, 410082,P,R,China;

Department of Electrical Engineering, Changsha University of Electrical Power · Changsha, 410077,P.R.China)

TONG Tiaosheng

(College of Electrical & Information Engineering, Hunan University · Changsha, 410082,P.R.China)

Abstract: Schema theorem is a fundamental theory of genetic algorithms (GAs). This paper thoroughly analyses and studies the schema theorem of partheno-genetic algorithm (PGA) and gives out its formula, and then approximately estimates the probability that genetic operators of PGA destroy schema.

Key words: pantheno-genetic algorithm; genetic operator; schema theorem

Document code: A

单亲遗传算法图式定理的进一步分析

李茂军

(湖南大学电气与信息工程学院·长沙,410082;长沙电力学院电力工程系·长沙,410077)

童调生

(湖南大学电气与信息工程学院·长沙,410082)

摘要: 图式定理是遗传算法的重要理论基础之一.本文对单亲遗传算法的图式定理进行了全面的分析研究,给出了单亲遗传算法图式定理的表述式,并对各种遗传算子破坏图式的概率作了估算.

关键词: 单亲遗传算法;遗传算子;图式定理

1 Introduction

The parthene-genetic algorithm (PGA)[1,2] is a genetic algorithm (GA) using ordinal character strings as chromosomes, and its genetic operation is achieved by the genetic operators operating in one chromosome only such as gene exchange operator, gene shift operator and gene inversion operator instead of crossover operators operating between two chromosomes such as PMX, OX and CX[3]. Compared with the traditional genetic algorithms (TGAs), the genetic operation in PGA is simpler and more efficient, the initial population need not be varied and there is no "immature convergence", while PGA has the basic features of TGAs. When solving combinatorial optimization problems such as train line holding problem[1], travelling salesman problem[4] and pattern clustering problem[5], PGA is distinctly more ef-

ficient than TGAs.

Schema theorem is a fundamental theory of GAs. On the basis of Reference [6], this paper further analyses and studies the schema theorem of PGA and gives out its formula, and then approximately estimates the probability that genetic operators of PGA destroy schemas.

2 The genetic operators of PGA

Definition 1 Gene exchange operation of PGA is the procedure in which two or several genes (characters) in a chromosome (character string) are exchanged by a certain probability p_e. The exchanged genes are randomly selected.

Gene exchange operation can be divided into one-point gene exchange operation and multi-point gene exchange operation. The former exchanges only two genes in a chromosome; the latter takes a definite positive in-

teger u_e at first, and then randomly selects a positive integer $j \in \{1,2,\cdots,u_e\}$, finally exchanges j pair genes in a chromosome.

Definition 2 Gene shift operation of PGA is to shift the genes (characters) in one or several sub-strings in a chromosome (character string) backward in turn and to shift the last gene (character) in this (these) sub-string(s) to the first location, by a certain probability p_s. In the gene shift operation, the shifted sub-string(s) and its (their) length are randomly determined.

Gene shift operation can also be divided into one-point gene shift operation and multi-point gene shift operation. The former only shifts the genes in one sub-string in a chromosome; the latter takes a definite positive integer u_s at first, and then randomly selects a positive integer $j \in \{1,2,\cdots,u_s\}$, finally shifts the genes in j sub-strings in a chromosome.

Definition 3 Gene inversion operation of PGA is to invert the genes (characters) in one or several sub-strings in a chromosome (character string) by a certain probability p_i. In the gene inversion operation, the inverted sub-string(s) and its (their) length are randomly determined.

Gene inversion operation can also be divided into one-point gene inversion operation and multi-point gene inversion operation. The former only inverts the genes in one sub-string in a chromosome; the latter takes a definite positive integer u_i at first, and then randomly selects a positive integer $j \in \{1,2,\cdots,u_i\}$, finally inverts the genes in j sub-strings in a chromosome.

Definition 4 Gene leap operation of PGA is the procedure in which one or several genes (characters) in a chromosome (character string) leap to other values in the same order gene set[2], by a certain probability p_m. In the gene leap operation, the changed genes are randomly selected.

Gene leap operation consists of one-point gene leap operation and multi-point gene leap operation. The former takes only one gene in a chromosome to leap to other values in the same order gene set; the latter takes a definite positive integer u_m at first, and then randomly selects a positive integer $j \in \{1,2,\cdots,u_m\}$, finally takes j genes in a chromosome to leap to other values in the same order gene set.

Definition 5 Gene recombination operation of PGA is all the genetic operations that adjust the location of the ordinal genes (characters) in a chromosome such as gene exchange operation, gene shift operation and gene inversion operation.

3 Formula of the schema theorem of PGA

The formula of the schema theorem has relation with the operative process of PGA. This paper analyses and studies the schema theorem of PGA based on the model operative process of PGA that has given out in Reference [1].

Assume gene recombination operator and gene leap operator destroy schema by the probability p_{rd} and p_{md} respectively, then the number of schema H in the new individuals generated from population t by gene recombination operator and gene leap operator is

$$M_1(H,t) = M(H,t)(1 - p_d), \qquad (1)$$

Gene recombination operator and gene leap operator search subspace Λ_e and subspace Λ_m[2] respectively, that is, they process the schema in their respective subspace, so that p_d in formula (1) should take p_{rd} or p_{md} respectively, while they do not combine one with the other. The number of schema H in population $(t + 1)$ is

$$m(H,t+1) \geqslant [m(H,t) + m_1(H,t)]\frac{f(H)}{2\bar{f}}, \qquad (2)$$

where $\bar{f}$ is the average fitness of $2N$ individuals (where N is the number of all the individuals in a population) that contain N individuals in population t and N new individuals produced by the genetic operation, $f(H)$ is the average fitness of the individuals including schema H. Formula (2) isn't an equation because we ignore the number of schema H obtained from non-schema H. From formula (1) and formula (2), we have

$$m(H,t+1) \geqslant m(H,t)\frac{f(H)}{\bar{f}}\left(1 - \frac{p_d}{2}\right). \qquad (3)$$

That is the formula of the schema theorem of PGA, which gives out the floor lever of the number of the schema H in next population.

4 Estimation of the probability of the genetic operators destroying schema

To improve computing efficiency, the genetic operators, on one hand, should generate more new schema; but on the other hand, should maintain the excellent

schemata with high fitness, that is, the probability of genetic operators destroying schema should neither be too high nor too low. For the above reasons, it is extremely important to estimate probability of genetic operators destroying schemata for improving computing efficiency.

4.1 Estimation of the probability of the gene recombination operators destroying schema

4.1.1 Estimation of the probability of the gene exchange operator destroying schema

In Reference [6], the probability of one-point gene exchange operator destroying schema H has be given out, that is,

$$P_{ed} = \frac{P_e(H - O(H))O(H)}{l^2}. \tag{4}$$

Considering $u_e \ll l$, the average probability of multi-point gene exchange operator destroying schema H is

$$P_{ed} = \sum_{i=1}^{u_e} \frac{1}{u_e} P_e(1 - P_{el}) \approx$$

$$P_e - \frac{P_e}{u_e} \sum_{i=1}^{u_e} \left(\frac{l - O(H)}{l} \right)^{2i}, \tag{5}$$

where p_{el} is the probability that schema H can live.

4.1.2 Estimation of the probability of the gene shift operator destroying schema

The sum of character strings in length from 2 to l contained in schema H is

$$\beta = \frac{l(l-1)}{2}. \tag{6}$$

If the number of character strings that only consist of wildcards (*) in those character strings is r, then the probability of one-point gene shift operator destroying schema H is

$$P_{sd} = P_s(1 - \frac{\gamma}{\beta}). \tag{7}$$

Considering $u_s \ll l$, and having $u_s \ll \beta$, the average probability of multi-point gene shift operator destroying schema H is

$$P_{sd} = \frac{1}{u_s} \sum_{i=1}^{u_s} P_s(1 - P_{sl}) \approx$$

$$P_s - \frac{P_s}{u_s} \sum_{i=1}^{u_s} \frac{\gamma(\gamma - 1)\cdots(\gamma - i + 1)}{\beta^i}, \tag{8}$$

where p_{sl} is the probability that schema H can live.

4.1.3 Estimation of the probability of the gene inversion operator destroying schema

Definition 6   Assume that character string A con-

sists of ordinal genes and character string A is converted into character string B by inverting all the genes in character string A. If $A = B$, the character string A is called symmetrical character string.

The sum of character strings in length from 2 to l contained in schema H is β. Assume that the number of the symmetrical character strings in those character strings is θ. Then, probability of one-point gene inversion operator destroying schema H is

$$P_{id} = P_i(1 - \frac{\theta}{\beta}). \tag{9}$$

Considering $u_i \ll l$, the average probability of multi-point gene inversion operator destroying schema H is

$$P_{id} = \frac{1}{u_i} \sum_{i=1}^{u_i} P_i(1 - P_{il}) \approx$$

$$p_i - \frac{p_i}{u_i} \sum_{i=1}^{u_i} \frac{\theta(\theta - 1)\cdots(\theta - i + 1)}{\beta^i}, \tag{10}$$

where p_{il} is the probability that schema H can live.

The symmetrical character strings can be divided into two classes, one only consists of wildcards (*) and the other consists of wildcards (*) and fixed genes (characters). The number of the former is γ and the number of the latter is δ. From $\theta = \gamma + \delta$, we have $\theta \geqslant \gamma$, that is, the gene shift operator destroys schema more easily than gene inversion operator does.

4.2 Estimation of the probability of the gene leap operator destroying schema

In Reference [6], probability of one-point gene leap operator destroying schema H has be given out, that is,

$$P_{md} = \frac{P_m O(H)}{l}. \tag{11}$$

Considering $u_m \ll l$, the average probability of multi-point gene leap operator destroying schema H is

$$P_{md} = \frac{P_m}{u_m} \sum_{i=1}^{u_m} (1 - p_{ml}) \approx$$

$$p_m - \frac{p_m}{u_m} \sum_{i=1}^{u_m} \left(\frac{l - O(H)}{l} \right)^i, \tag{12}$$

where p_{ml} is the probability that schema H can live.

5 Conclusion

This paper analyses and studies the schema theorem of PGA thoroughly. From the point of view of processing schema, the genetic operators of PGA have their respective function. During the genetic operation, gene recom-

bination operator and gene leap operator continually produce new schemata, while selection operator, on one hand, maintains the excellent schemata with high fitness, but on the other hand, falls into disuse bad schemata with low fitness. Similarly with TGA, through genetic operator processing schemata, the individuals in the population continually move towards the optimal individual in PGA, finally the optimal solution can be gained.

References

[1] Li Maojun, Tong Tiaosheng and Luo Longfu. A partheno-genetic algorithm and its application [J]. J. of Hunan University (Natural Sciences), 1998,25(6):56 – 59

[2] Li Maojun and Tong Tiaosheng. A partheno-genetic algorithm and analysis on its global convergence [J]. Chinese J. of Automation, 1999,11(2):119 – 123

[3] Pedro Larranaga, Cindy M H Kuijpers, Roberto H Murga, et al. Learning Bayesian network structures by searching for the best ordering with genetic algorithms [J]. IEEE Trans. on System. Mao, and Cybernetics, Part A: System and Humans, 1996,26(4):487 – 493

[4] Li Maojun and Tong Tiaosheng. A partheno-genetic algorithm solving serial combinatorial optimization [J]. Systems Engineering and Electronics, 1998,20(10):58 – 61

[5] Li Maojun, Fan Shaosheng and Tong Tiaosheng. The application of partheno-genetic algorithm in pattern clustering problem [J]. Pattern Recognition and Artificial Intelligence, 1999,12(1):32 – 37

[6] Li Maojun, Tong Guangyu and Tong Tiaosheng. Studies of schema theorem of partheno-genetic algorithm [A]. In: Proc. The Conference on Control and Decision of China (CDC'98) [C]. Dalian: Dalian University of Maritime Affairs Press, 1998,332 – 335

[7] Li Maojun, Qiu lifang and Tong Tiaosheng. The analysis on searching efficiency of partheno-genetic algorithm [J]. Journal of Changsha University of Electric Power (Natural Science),1999,14(1):48 – 50

[8] Sun Yanfeng and Wang zhongmo. Studies of schema theorem on genetic algorithm [J]. Control and Decision, 1996,11(suppl.1):221 – 224

本文作者简介

李茂军 1964 年生.分别于 1986 年和 1998 年在湖南大学获学士学位和硕士学位.现为湖南大学控制理论与控制工程学科在职博士研究生,长沙电力学院电力工程系副教授.主要研究领域:智能控制,进化计算等.

童调生 1936 年生.1961 年毕业于东北大学.现为湖南大学控制理论与控制工程学科教授,博士生导师.主要研究领域:最优控制,智能控制等.

---

[4] Kapila V and Haddad W M. Memoryless H_∞ controllers for discrete-time systems with time delay [J]. Automatica, 1998,34(9):1141-1144

[5] de Souza C E, Fu M and Xie L. H_∞ analysis and synthesis of discrete-time systems with time-varying uncertainty [J]. IEEE Trans. Automatic Control,1993,38(3):459 – 462

[6] Li Y C and Xu X M. The quadratic stability of a discrete-time system with structured uncertainties [J]. Int. J. Control, 1999,72(16):1427 – 1435

[7] Yu L. Robust stabilization of uncertain discrete-time linear systems [J]. Control and Decision,1999,14(2):169 – 172

[8] Song S and Kim J K. H_∞ control of discrete-time linear systems with norm-bounded uncertainties and time delay in state [J]. Automatica, 1998,34(1):137 – 139

[9] Madmoud M S and Xie L. Guaranteed cost control of uncertain discrete systems with delays [J]. Int. J. Control,2000,73(2):105 – 114

本文作者简介

郗耀军 1970 年生.现为南京理工大学控制理论与控制工程专业博士研究生.主要研究方向为时滞系统控制,鲁棒控制以及计算机控制理论与应用.

杨成梧 1936 年生.1961 年毕业于哈尔滨军事工程学院.现为南京理工大学教授,博士生导师.目前研究方向为 2D 系统,广义系统,H_∞ 控制和离散事件动态系统.