Theoretical Computer Science ••• (••••) •••–•••

Contents lists available at ScienceDirect

Theoretical Computer Science

www.elsevier.com/locate/tcs

Mathematical analysis of the cumulative effect of novel ternary crossover operator and mutation on probability of survival of a schema

Apoorva Mishra *, Anupam Shukla

Soft Computing and Expert System Laboratory, ABV – Indian Institute of Information Technology & Management, Gwalior, Madhya Pradesh, 474010, India

ARTICLE INFO

Article history:
Received 23 October 2015
Received in revised form 29 February 2016
Accepted 29 July 2016
Available online xxxx
Communicated by W. Banzhaf

Keywords:
Genetic algorithms
Ternary crossover
Mutation
Schema
Schema survival

ABSTRACT

The concept of schema plays a vital role in the study of genetic algorithms. The effect of selection, simple crossover and mutation on schemata has already been studied rigorously by several researchers. In this paper a novel ternary crossover operator is introduced and its effects on the probability of survival of a schema are meticulously analyzed. A theorem regarding the effect of novel crossover operator on survival of schemata is proved and based on that the combined effect of novel crossover and mutation on probability of survival of a schema is established mathematically.

1. Introduction

Genetic algorithms are ubiquitous. There are innumerable areas where they have been successfully used [16]. Some of the most common operators used in genetic algorithms are: selection, crossover and mutation. Various variants of these basic operators have also been proposed to improve the performance of the traditional genetic algorithm [2,11] and have been successfully applied in the optimization of various problems [17].

Apart from genetic algorithms several other evolutionary [9,19] and nature inspired strategies [1,4–6,15] have also been proposed to solve optimization problems involving multiple objectives. As exhaustive analysis of the performance of evolutionary algorithms is difficult, various theoretical ways of analyzing them have been proposed [3]. Out of the various approaches of optimization which have been proposed, one proves to be better than the others under certain circumstances, while another approach might outperform it under different circumstances [20].

The schema theory [7] plays a vital role in the field of genetic algorithms as it enables us to estimate the number of instances of a schema in a particular generation based on the number of its instances in the previous generation. A lot of research has been done in this field and a generalized expression for schemata count has been proposed [13]. Schema theory has also been extended from genetic algorithms to genetic programming [21].

A schema represents a general class of chromosomes. A binary schema consists of a string over the alphabet $\{0, 1, *\}$; where '*' represents "don't care" bits and could be either '0' or '1'. Multiple strings in the search space belong to one

* Corresponding author.
E-mail addresses: apoorvamish1989@gmail.com (A. Mishra), dranupamiiitm@gmail.com (A. Shukla).

http://dx.doi.org/10.1016/j.tcs.2016.07.035
0304-3975/

Please cite this article in press as: A. Mishra, A. Shukla, Mathematical analysis of the cumulative effect of novel ternary crossover operator and mutation on probability of survival of a schema, Theoret. Comput. Sci. (2016), http://dx.doi.org/10.1016/j.tcs.2016.07.035

schema. Positions in the string where '0' or '1' occurs are termed as fixed positions. Schema theory deals with the study of the existence of a schema in the population at various generations and with the analysis of the effect of selection [8], crossover [12,14,18] and mutation [10] on the chance of survival of a schema.

The rest of the paper is organized in the following manner: Section 2 discusses the basic concept of schema as available in the literature and introduces the meaning of jargons like: defining length, order, fragility etc. with suitable examples.

In Section 3, the traditional crossover operator is explained with suitable examples and the effect of simple crossover on the probability of survival of schema is derived step by step for proper understanding of the concept [7]. Then, the combined effect of simple crossover and mutation on the probability of survival of a schema is derived mathematically. These concepts are also available in literature [7].

In Section 4.1, a novel ternary crossover operator is introduced and the different possible cases regarding its working are thoroughly investigated and represented using suitable examples. A theorem is proved on the basis of the observations regarding the effects of a novel ternary crossover operator.

Section 4.2 discusses the effect of the novel ternary crossover operator on the probability of survival of a schema from a generation to the next one. Different cases are considered and analyzed to gain better insight about the process and another theorem is proved (which is an extension of the previous theorem). Based on these theorems the effects of the novel ternary crossover operator on schemata are meticulously analyzed and the combined effect of novel ternary crossover operator and mutation on probability of survival of schema is represented mathematically.

In Section 5, the implications of the survival of a schema from a generation to the next one are discussed.

Section 6 concludes the various findings of the mathematical analysis.

2. Concept of schema

The basic concept of schema has been described in Section 1. In order to understand it thoroughly in the context of genetic algorithms, we give some basic definitions.

Definition 1 *(Order of a schema).* The order of a schema $o(H)$ is the number of fixed positions in the string.

Definition 2 *(Defining length of a schema).* The defining length of a schema $d(H)$ is the distance between the first and the last fixed positions in the string, i.e. the number of places where crossover can disrupt the schema.

Definition 3 *(Fragility of a schema).* The fragility of a schema is the number $\frac{d(H)}{(l-1)}$, (where, $d(H)$ is the defining length and l is the length of the schema) i.e. the proportion of places where the schema can be disrupted.

Definition 4 *(Gene).* A gene is a bit string of arbitrary length that represents a part of an individual.

Definition 5 *(Chromosome).* A chromosome is a collection of different genes arranged in a sequential order.

Definition 6 *(Fitness).* The fitness of an individual is a measure of its ability to solve a problem.

Definition 7 *(Crossover).* Crossover is the genetic operator that performs the task of recombination of two individuals (parents) to generate individuals of the new generation (offspring).

Definition 8 *(Selection).* Selection is the process by which the pairs of individuals that will participate in crossover operation are selected.

Definition 9 *(Mutation).* Mutation is the process by which new characteristics are introduced into the population.

Let us try to understand the concept of schema with the help of an example:

Suppose, we have a schema $H_1 = 10^{**}0^{*}1^{***}$.
For this schema,

Length $l(H_1) = 10$ (total number of bits in the string representing it).

Order of schema $o(H_1) = 4$ (Number of fixed bits)

Defining length $d(H_1) = 7 - 1 = 6$ (Difference between the last fixed bit and first fixed bit)

Vulnerability or fragility $\dfrac{d(H_1)}{(l-1)} = 6/9 = 0.66667$

Vulnerability represents the portion of the schema where crossover will destroy it.

Please cite this article in press as: A. Mishra, A. Shukla, Mathematical analysis of the cumulative effect of novel ternary crossover operator and mutation on probability of survival of a schema, Theoret. Comput. Sci. (2016), http://dx.doi.org/10.1016/j.tcs.2016.07.035