Chapter 16

Efficiency Analysis of Genetic Algorithm and Genetic Programming in Data Mining and Image Processing

Ayan Chatterjee
Sarboday Public Academy, India

Mahendra Rong
Bangabasi Evening College, India

ABSTRACT

Today, in the age of artificial intelligence and machine learning, Data mining and Image processing are two important platforms. GA and GP are value based and program based randomized searching tools respectively and these two are very much useful in the fields’ data mining and image processing for handling different issues. In this chapter, a review is made on ability of GA and GP in some applications of these two fields. Here, the selected subfields of data mining are market analysis, fraud detection, risk management, sports analysis, protein interaction, classification of data, drug discovery and feature construction. The similar in image processing are enhancement and segmentation of images, face recognition, photo mosaic generation, data embedding, image pattern classification, object detection and Graphics Processor Unit (GPU) development. The efficiencies of GA and GP in these particular applications are analyzed with corresponding parameters, comparing with other non-GA and non-GP approaches of the corresponding subfields.

INTRODUCTION

In present decade, optimization is an essential part in our everyday life. In any type of efficiency analysis, this is an important in each platform. Optimization procedures are majorly divided into two parts- hard computing and soft computing. Soft computing is more realistic than hard computing in every situation. Genetic Algorithm (GA) is a useful optimization approach of soft computing.

DOI: 10.4018/978-1-5225-2440-3.ch016
Efficiency Analysis of Genetic Algorithm and Genetic Programming

Genetic Algorithm (GA) was invented by John Holland in mid-sixties in university of Michigan. But idea of natural selection and natural genetics were developed by David Goldbarg. This is probabilistic and global search depending on principles of evolution. It was built up on the basis of three different hereditary properties of animals- selection, cross-over and mutation. Actually, the concept of creation of children from previous generation is used to obtain global optimum value in this useful as well as interesting soft computing approach.

According to differentiation of the operators’ selection, cross-over and mutation, different GA schemes are developed. According to number of objectives, there are two type genetic algorithms are made up like each type optimization approach- single objective genetic algorithms and multi objective genetic algorithms. Binary coded GA, Real coded GA, Messy GA etc are some single objective GA approaches and weighted sum, NSGA and NSGA-II etc are some multi objective GA approaches. In different platforms, different GA shows their optimization efficiencies in different ways. Among them, computer science, biology, economics, mathematics, pharmacology etc. are major application fields of genetic algorithm. In computer science, some efficient fields of GA are artificial intelligence, network security, data mining, image processing, pattern recognition etc.

Genetic Programming (GP) is an extended version of GA. In GA, population is created with a set of values and in GP; population is created with a set of sub programs, more specifically with a set of instructions. The concept of GP was first proposed by Lawrence J. Fogel in 1964. But at that time, he only introduced about this idea and simply implemented it in finite state automata. It was broadly used in 1990 with machine learning and artificial intelligence. Some specific application fields of GP are expert system, horn clause, networking, automation theory, software engineering, data mining, image processing etc.

In this chapter, we mainly focus on the application of GA and GP in data mining and image processing. We choose these two fields of application, because- these two are in a broad range of particular research application as well as GA and GP both are used in these two particular cases. So, we can analyze the efficiency of GA and GP with proper experiments.

BACKGROUND

In this section, we will describe some typical natures of GA and GP, observed in different situations as well as we mention about some comparative analysis of GA and GP in some application fields. In 2014, Mujahid Tabassum and Kuruvilla Mathew proposed a genetic algorithm analysis towards optimization solutions. In that paper, they discussed about some non linear, large and discrete type attributes containing problems with high degree complexity solution. Actually they mainly focused on different parts of artificial intelligence which are solved using GA. In this paper, the particular application fields are evolvable hardware applications, robotics, engineering design, data encryption and computer gaming. The comparative analysis is developed among GA and traditional methods with the mentioned application fields in that paper. In 1997 and 2006, two papers were published on data mining using GA and GP respectively. In the paper, application of GA in Data mining (published in 1997), Robert E. Marmaelestein proposed a comparative analysis between KNN classifier and genetic algorithm with Mushroom data set clustering. In that particular clustering, it is observed that, GA takes 15 generations to obtain 0% training error, whereas KNN classifier takes 3 generations to get the same result with same population size, probability of mutation and probability of cross-over. In the paper, application of GP in data