Vrije Universiteit Amsterdam
Exam Evolutionary Computing
15.06.2004

Note 1 Your name must be written on each sheet in CAPITALS.

Note 2 You can answer the questions in English or in Dutch.

Points to be collected: 100

Grade: points collected divided by 10, rounded up to the first decimal.

1. The Fraud Detection Office organises site visits of companies to scrutinize financial records and attempt to uncover fraudulent tax returns. They also maintain a database which holds the following information for each company: number of employees, annual turnover and declared profits for current and last six years, mean growth in turnover for all companies in that sector, and a label indicating whether the company had a case of fraud over the past six years. In order to improve their detection rate, and avoid wasting the time of honest companies, they wish to implement a system for predicting which companies are likely to act fraudulently, and have approached you as a consultant to design a GP-based system using their database and historical records of fraudulent behaviour.

   NB. If your solution requires, you may freely make assumptions on the problem, the data, etc., as long as they are realistic.

   (a) **(6p)** Describe a suitable GP representation for this problem, including likely function and terminal sets.

   (b) **(6p)** Specify a suitable fitness (evaluation) function.

   (c) **(6p)** Describe what crossover and mutation operators you would implement.

   (d) **(4p)** Describe how you would initialise the population

   (e) **(4p)** Explain what is bloat in GP, and how can it be avoided?

   (f) **(4p)** A commonly used technique with large population sizes is so-called "over-selection". Describe the basic principle of this technique.

2. (a) **(4p)** Explain what the term "operator adaptation" means with respect to evolutionary algorithms.

   (b) **(4p)** Explain the rationale behind the use of operator adaptation.

   (c) **(18p)** There are taxonomies for classifying adaptive evolutionary algorithms based on the concept of three different levels of adaptation. Describe these different levels of adaptation, giving an example of each. (NB. Adaptation in this question is not restricted to operators, your answer may consider other adaptive components.)

3. Many difficult combinatorial optimization problems have as their basis the task of finding the "best" order in which some sequence of events should occur (with each happening exactly once). Given n such events to be ordered, one representation is clearly as a permutation of the integers 1 to n.

   (a) **(4p)** Given such a permutation-based representation, why are normal position-wise independent mutation operators not generally used? Illustrate your answer with a simple example.

   (b) **(4p)** Given such a permutation-based representation, why are normal recombination operators such as 1-point or uniform crossover not generally used? Illustrate your answer with a simple example.

   (c) **(8p)** Describe an alternative way to permutation representation. That is, assume that the solution to a problem is a permutation and you may not use permutation representation in your EA. Show how (by what features, e.g., representation, fitness function, mutation and crossover operators, selection, etc.) can you achieve that your solution will be a permutation. (NB. It is sufficient if your solution *could* achieve this property – a guarantee is of course not required.)

4. (a) **(6p)** Describe the characteristic features of a modern Evolution Strategy (ES), in terms of representation, principal search operator and population models.

   (b) **(4p)** How does the way that selection pressure is applied differ between a modern ES and a generational Genetic Algorithm?

   (c) **(8p)** Typically modern ESs use intermediate recombination for strategy parameters and discrete recombination for problem variables. Explain these operators work, and how their effects on population diversity (i.e., on the spread of alleles/values in the population) differ.

   (d) **(6p)** Describe the process of mutation for an ES with n variables and a single mutation step size.

   (e) **(4p)** Why is the order in which the steps occur in this process so important for its successful functioning?