An Experimental Evaluation of Genetic Process Mining

Chris J. Turner
Cranfield University
Cranfield
Bedfordshire, UK.
c.j.turner@cranfield.ac.uk

Ashutosh Tiwari
Cranfield University
Cranfield
Bedfordshire, UK.
a.tiwari@cranfield.ac.uk

ABSTRACT
This paper aims to ascertain the optimum values for two fitness function parameters within a process mining genetic algorithm; the κ parameter, which reduces the likelihood of process models with extra behaviour being selected and the γ parameter, which restricts the selection of models containing duplicate tasks. The experiments conducted in this research also include the use of a decaying rate for the mutation operator in order to promote greater accuracy in the mined process models. The paper concludes that the optimum setting of the fitness function parameters will in fact vary depending on the constructs found in each process model. This paper finds that a higher value for one of the fitness function parameters allows for simple process constructs to be mined with greater accuracy. The use of a decaying rate of mutation is also found to be beneficial in the correct mining of simple processes.

Categories and Subject Descriptors
J.1. [**Administrative Data Processing**] – Business

General Terms
Algorithms

Keywords
Business process mining, GA, Mutation rate, Chromosome fitness

1. INTRODUCTION
Data mining practice has been developed and adapted to create the business process mining techniques that are now being used to reconstruct actual business processes by mining event logs containing process execution data. These event logs are typically hosted within Business Process Management (BPM) and workflow systems, though they may also be accessible through other process related systems installed within a company. The need for companies to learn more about how their processes operate in the real world is a major driver behind the development and increasing use of process mining techniques. Currently many approaches to process mining make use of heuristic algorithms ('rules of thumb' based on assumptions about business process patterns). Process mining is uniquely challenging as there is often just one correct representation of a process which must be reconstructed from process execution traces. The mining algorithm must also able to mine more complex process constructs such as loops and cope with the presence of duplicate tasks [1]. The work of Alves de Medeiros [1], on GA use in process mining, is used as a base for the experiments in this paper.

GECCO'07, July 7–11, 2007, London, England, United Kingdom.
ACM 978-1-59593-697-4/07/0007.

2. EXPERIMENTAL EVALUATION
The key stages of the process mining algorithm are detailed in Alves de Medeiros et al. [1]: 1) Read the Event Log, 2) Build the Initial Population, 3) Calculate Individuals' Fitness, 4) Stop and Return the Fittest Individuals and 5) Create Next Population.

This paper studies the effect of applying a decaying rate of mutation to the mutation function employed by Alves de Medeiros [1]. Authors on evolutionary techniques, such as Cox [2], point to the use of a decaying rate of mutation in genetic algorithms. This paper also experiments with the κ and γ parameters in the fitness measure employed by Alves de Medeiros [3]. The κ parameter is used in the weighting of the extra behaviour punishment measure, and the γ parameter is used in the weighting of the punishment for duplicate tasks that share input and/or output tasks.

3. CONCLUSIONS
From the experiments conducted as part of this research it is clear that an optimum general setting for the κ and γ parameters of the fitness function for use in all mining activities is not realistic. The use of a higher value for the γ setting of the fitness function was shown to bring benefits for the mining of relatively simple processes. It is likely that a wider range of variables, such as crossover, population size and number of generations, need to be considered in the mining of more complex processes. However it may be possible to derive recommended settings for different classes of process. This conclusion can also be drawn for the mutation setting. The use of the decaying rate of mutation was shown to be beneficial in the mining of simple processes with a short sequence of tasks and few loops.

4. REFERENCES
[1] Alves de Medeiros, A. K.; Weijters, A. J. M. M., and van der Aalst, W. M. P. *Using Genetic Algorithms to Mine Process Models: Representation, Operators and Results*. WP 124. , Eindhoven Technical University, Eindhoven, 2004.

[2] Cox, E. *Fuzzy Modelling and Genetic Algorithms for Data Mining and Exploration*, Morgan Kaufmann, London, 2005.

[3] Alves de Medeiros, A. K, *Genetic Process Mining*. Ph. D Thesis, Eindhoven Technical University, Eindhoven, The Netherlands, 2006.