REAL WORLD APPLICATIONS
Poster Papers

Rajkumar Roy and David Davis, chairs

Optimized Interest Metric of Rules and One-to-One Marketing Using Connection Networks

Sung-Soon Choi and Byung-Ro Moon
School of Computer Science and Engineering,
Seoul National University, Seoul, 151-742 Korea
{irranum,moon}@soar.snu.ac.kr

With the explosive growth of the electronic commerce data, rule finding became a crucial part in marketing. A number of different metrics to quantify "interestingness" or "goodness" of rules, including *support* and *confidence* [1], have been proposed [3]. We suspect that the most suitable interest metics for different data sets are different one another. In this context, we find the optimized interest metric for the given data set.

We found that most of the proposed metrics can be expressed with respect to supports. Based on this observation, we express the degree of the connection from an item X to an item Y, $f(X \to Y)$, as the function of support values of X, Y, and $X \cup Y$. We replace those support values with independent variables x, y, and z, respectively, and then we set a model of $f(x, y, z)$ $(=f(X \to Y))$ as follows:

$$f(x, y, z) = (a_x x^{e_x} + b_x)(a_y y^{e_y} + b_y)(a_z z^{e_z} + b_z)\,.$$

We use a genetic algorithm to search the optimal coefficients and exponents of $f(x, y, z)$ for the data set.

Now we construct a connection network based on the optimized metric. We set a vertex for each item. We put an arc from the vertex X to the vertex Y with the weight $f(X \to Y)$. We have a directed graph $G = (V, A)$, where V is the vertex set and A is the arc set.

We perform one-to-one marketing using the connection network. Suppose that a customer purchased the products $X_1, X_2, \ldots, X_k$ so far. Let $N(X_i)$ be the set of neighbor vertices of X_i in the connection network $(1 \le i \le k)$. We define a score function for recommendation, $s : V \longmapsto R$, as follows (R is the set of real numbers.): if $Y \in \bigcup_i N(X_i)$, $s(Y) = \{\sum_{1 \le i \le k} f(X_i \to Y)\}/(a_k k^{e_k} + b_k)$, otherwise, $s(Y) = 0$.

We recommend the products of high scores to the customer. The value of the score function $s(Y)$ for a product Y is proportional to the weight sum of the arcs from the previously purchased products to the product Y. Figure 1 shows an example connection network. We divide the sum of weights by a function of k (the number of previously purchased products). This prevents the recommendations from flowing into a few customers that purchased excessively many products before.

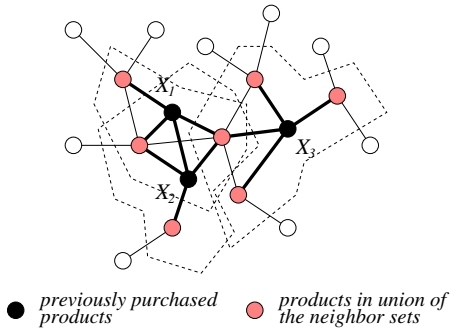

Figure 1: A connection network

Such a recommendation strategy is different from the existing ones based on the customers' profiles (for example, collaborative filtering [2]) in that it performs recommendations just by the relationships between products regardless of the customers' profiles. Experimental results with field data showed that the connection network model is fairly good not only in the computational cost but also in the recommendation quality.

This work was supported by KOSEF through the Statistical Research Center for Complex Systems at Seoul National University (SNU) and Brain Korea 21 Project. The RIACT at SNU provided research facilities for this study.

References

[1] R. Agrawal, T. Imilienski, and A. Swami. Database mining: A performance perspective. *IEEE Trans. on Knowledge and Data Engineering*, 5(6):914–925, 1993.

[2] D. Goldberg, D. Nichols, B. M. Oki, and D. Terry. Using collaborative filtering to weave an information tapestry. *Comm. of the ACM*, 35(12):61–70, 1992.

[3] R. J. Bayardo Jr. and R. Agrawal. Mining the most interesting rules. In *Proc. of the fifth ACM SIGKDD Int'l Conf. on Knowledge Discovery and Data Mining*, pages 145–154, 1999.

A Genetic Algorithm for Joint Optimization of Spare Capacity and delay in Self-Healing Network

H W Chong
Department of Computer Science
City University of Hong Kong
83 Tatchee Avenue., Hong Kong

Sam Kwong
Department of Computer science
City University of Hong Kong
83 Tatchee Avenue., Hong Kong

Abstract

This paper presents the use of multi-objective Genetic Algorithms (mGA) to solve the capacity and routing assignment problem arising in the design of self-healing networks using the Virtual Path (VP) concept. The aims to minimize the sum of working and backup capacity usage and transmission delay often compete and contradict with each other. Multi-objective Genetic algorithm is a powerful method for this kind of multi-objective problems. In this paper, a multi-objective GA approach is proposed to achieve the above two objectives while a set of customer traffic demands can still be satisfied and the traffic is 100% restorable under a single point of failure. We carried out a few experiments and the results illustrate the trade-off between objectives.

1 INTRODUCTION

Reducing network protection costs while maintaining an acceptable level of survivability is one of the main objectives of the network planners. The transmission delay from a source node to a destination node depends very much on the paths chosen. Every customer likes to have a route with minimum delay, the TELCO operator cannot promise every customer that the route is the shortest. It becomes an interesting problem for one to choose routes that will compromise the interests of the TELCO operator and different customers.

The above problem can be regarded as the combination of two sub problems:
1. The objective to minimize capacity subject to a constraint on delay with given a network topology and assigned traffic requirements.
2. The objective to minimize delay subject to a constraint on the total capacity with a given network topology and assigned traffic requirements.

Obviously, these two sub-problems are dependent and cannot be easily solved without considering each other's existence. That is an issue that makes the above problem a relatively hard one, which is not easy to be solved using classical optimization methods.

To tackle the above problems, a GA-based multi-objective optimization approach is presented in this paper. We will see how the method obtains a Pareto set of solutions in that any single set of solution can be freely chosen according to the fulfillment of the system requirements. In this paper, we will also see that spare capacity and delay are jointed optimized and to provide a highly available service.

2 EXPERIMENTAL RESULTS

To validate the effectiveness of the proposed approach, we have performed experiment in network shown in figure 1. The experiment is based on customer traffic demands of 100 working virtual paths and the highest ranked chromosomes in the final generation are depicted in figure 2. A Pareto optimal set can be clearly obtained by applying the multi-objective GA-based approach.

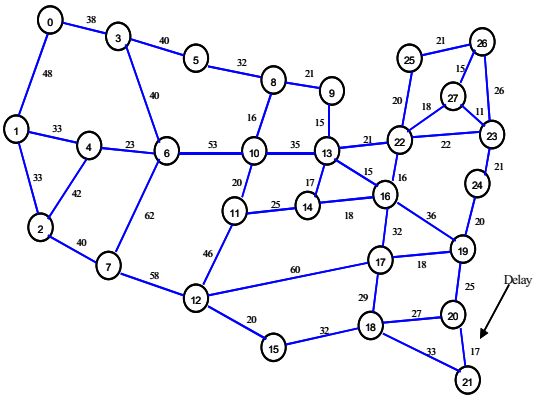

Figure 1. Network 1

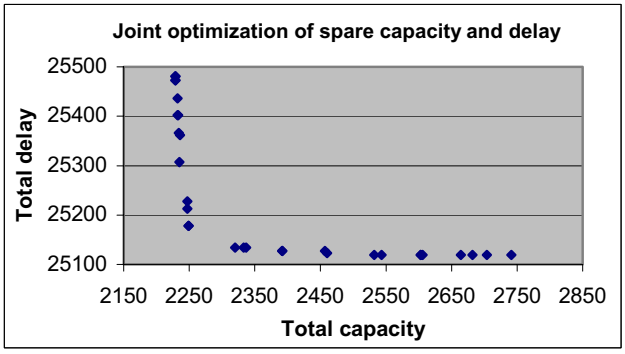

Figure 2. Final pareto front of Network 1

A real coded genetic algorithm for the optimisation of reaction rate parameters for chemical kinetic modelling in a perfectly stirred reactor

L. Elliott
Department of Applied Mathematics, University of Leeds, Leeds, LS2 9JT, UK
lionel@amsta.leeds.ac.uk

D.B. Ingham
Department of Applied Mathematics, University of Leeds, Leeds, LS2 9JT, UK
amt6dbi@amsta.leeds.ac.uk

A.G. Kyne
Centre for Computational Fluid Dynamics, University of Leeds, Leeds, LS2 9JT, UK
fueagk@sun.leeds.ac.uk

N.S. Mera
Centre for Computational Fluid Dynamics, University of Leeds, Leeds, LS2 9JT, UK
fuensm@sun.leeds.ac.uk

M. Pourkashanian
Department of Fuel and Energy, University of Leeds, Leeds, LS2 9JT, UK
fue6lib@sun.leeds.ac.uk

C.W. Wilson
Centre for Aerospace Technology, QinetiQ group plc, Cody Technology Park, Ively Road, Pyestock, Farnborough, Hants, GU14 0LS, UK
C.W.Wilson@qinetiq.co.uk

Abstract

In this study we develop a real coded genetic algorithm for the determination of the chemical reaction rates parameters (*A*'s, *b*'s and *E*'s in the Arrhenius expression) for the hydrogen combustion in a perfectly stirred reactor (PSR). The algorithm is tested on a hydrogen/air mixture but it can be applied for other, more complex hydrocarbon fuels.

1 GENERAL FORMATTING INSTRUCTIONS

The chemistry of combustion may be modelled using a system of chemical reactions for which the rates of each reaction are known. The net chemical production rate of each species results from a competition between all the chemical reactions involving that species. We assume that each reaction proceeds according to the law of mass action and the forward rate coefficients are in modified Arrhenius form rate:

$$k_{fi} = A_i T^{b_i} \exp\left(-\frac{E_i}{RT}\right) \qquad (1)$$

for $i = 1,...,N_R$, where T is the temperature, $R = 1.9860$ cal mol^{-1} K^{-1} is the universal gas constant and there are N_R competing reactions occurring simultaneously. The rate equation (1) contains the three parameters A_i, b_i and E_i for the i^{th} reaction. Databases that give measurements of reaction rate parameters, (*A*'s, *b*'s and *E*'s in the Arrhenius expression) for various reactions are commercially available, but a large variation in the reaction rates is generally observed. It is the possibility of the determination of these parameters for each reaction, based upon the outlet species mole fractions alone, which is investigated in this study, using a real coded genetic algorithm.

An genetic algorithm optimisation procedure is set up in an attempt to recover the species profiles (to within any experimental uncertainty) resulting from numerous sets of operating conditions by calculating new reaction rate parameters that lie between predefined boundaries. The inversion process aims to determine the unknown reaction rate parameters $((A_i, b_i, E_i), i=1,\ldots,N_R)$ by searching for the set of reaction rates parameters that gives the best fit to a set of given data. If data is measured for N_S different sets of reactor conditions, and all the K species are measured for every reactor condition, then the data will consist of a set of KN_S species concentration measurements and we look for the set of reaction rate parameters that gives the best fit to these measurements. This is done by looking for the maximum of the function

$$f((A_i, b_i, E_i)_{i=1,N_R}) = \left\{10^{-8} + \sum_{j=1}^{N_S}\sum_{k=1}^{K}\left(Y_{jk}^{calc} - Y_{jk}^{meas}\right)^2\right\}^{-1}$$

where Y_{jk}^{calc} and Y_{jk}^{meas} are the calculated and the measured mole concentrations of the k^{th} species in the j^{th} set of reactor conditions. The real coded genetic algorithm employed is found to be very efficient in constructing new reaction mechanisms.

Congressional Redistricting Using a TSP-based Genetic Algorithm

Sean L. Forman

Department of Mathematics and Computer Science
Saint Joseph's University, Philadelphia, PA 19131
sforman@sju.edu

1 OVERVIEW

A redistricting plan partitions a state's tracts (often counties or census tracts) into a set of congressional districts. Special interests attempt to mold districts to their political advantage, often inviting judicial review in the process. The quality of a plan has been judged by the courts using a fairly standard set of criteria: (1) the districts must be contiguous, (2) the districts must be of essentially equal population, (3) the districts should be of a pleasing (compact) shape, and (4) some consideration should be given to regions of common political interest. We will introduce a GA that attempts to satisfy the first three criteria.

2 ENCODING AND OPERATORS

This GA's chromosome is a string representing a path visiting all of a state's tracts exactly once. This encoding is taken from common GA solutions to the Traveling Salesman Problem. To generate a districting plan from a chromosome, an optimal district population is computed using the state's population and its allotted number of representatives. As we travel along our path, each tract's population is added to a cumulative population count. Once the necessary population threshold is reached, the set of traversed tracts forms a new district. The process repeats with the remaining districts and the remaining length of chromosome. This technique forces district populations to be relatively close to the desired size.

The objective function minimized considers a combination of the compactness (as measured by the district's perimeter squared over its area) and the number of discontiguous parts formed, both of which are to be minimized.

Several common TSP-based genetic operators are used to combine chromosomes for subsequent generations. Maximal preservative crossover and exchange mutation have produced the best results in *ad hoc* tests of possible operators.

A heuristic, called discontiguity patch, is applied following the generation of a new population. This heuristic searches for small discontiguous sets of tracts within our encoding and then transfers them within the string to a geographically nearby tract.

3 AN EXAMPLE: IOWA

As a proof of concept, preliminary results have been produced using Iowa as a test case. Census Bureau data was used to generate tract size, boundary and population information. Iowa is a medium size state with five congressional districts. Work is continuing on parameter selection, operator choice, and comparison with other existing techniques.

Acknowledgments

I would like to thank Alberto Segre and Saint Joseph's University for support of this work, the U.S. Census Bureau Geography Division, and the reviewers for their comments.

References

[1] P. G. di Cortona, C. Manzi, A. Pennisi, F. Ricca, and B. Simeone. *Evaluation and Optimization of Electoral Systems*. Monographs on Discrete Mathematics and Applications. SIAM, 1999.

[2] P. Larra naga, C. Kuijpers, R. Murga, I. Inza, and S. Dizdarevic. Genetic algorithms for the traveling salesman problem: A review of representations and operators. *Artificial Intelligence Review*, 13:129–170, 1999.

Efficient Affine 2D-Image Registration using Evolutionary Strategies

Héctor Fernando Gómez García
Arturo González Vega
Arturo Hernández Aguirre
CIMAT, Depto. de Computación
Guanajuato, Gto. 36240, MEXICO
hector,gonzart,artha@cimat.mx

Carlos A. Coello Coello
CINVESTAV-IPN Sección de Computación
México, D.F. 07300
ccoello@cs.cinvestav.mx

Abstract

We propose a new methodology to the general affine image registration problem that in two steps achieves near optimal results with high percentage of success using evolutionary strategies.

Principles

The image registration problem is *to find the mapping between two images* I_1 and I_2 that gives the best correspondence. Equation 1

$$I_2(x, y) = I_1(f(x, y)) \tag{1}$$

Approximations to f can be constructed by some transformations: affine and projective amongst several.

An affine transform is a linear transform composed of the following geometric transformations: translation, rotation, scaling, stretching, and shearing [2]. The general $2D$ affine transformation is expressed as shown in Equation 2.

$$\begin{bmatrix} x_2 \\ y_2 \end{bmatrix} = \begin{bmatrix} a_{1,3} \\ a_{2,3} \end{bmatrix} + \begin{bmatrix} a_{1,1} & a_{1,2} \\ a_{2,1} & a_{2,2} \end{bmatrix} \begin{bmatrix} x_1 \\ y_1 \end{bmatrix} \tag{2}$$

Our approach to image registration uses a $(\mu + \lambda)$ Evolutionary Strategy for searching the six real variables described in Eq. 2. Crossover operation for control and object variables is generalized intermediate, mutation follows the standard formulation as explained in Bäck's book [1]. Our method uses only sample points equally spaced and distributed over the image. Thus, about 0.4% of image pixels are used for registration in our experiments. Fitness function is based on the similarity measure "absolute difference of intensities", as follows:

$$fitness = \frac{1}{1 + \sum_{\{x,y\}\in\Omega} |I_1(f(x, y)) - I_2(x, y)|} \tag{3}$$

Where Ω is the set of sample points. The fitness function takes values in $[0, 1]$ to represent $[nomatch \rightarrow perfectmatch]$.

Methodology

The method consists of the following steps:
Step 1: Smooth both original images to be registered. Smooth procedure uses a convolution with a Gaussian kernel of $\sigma = 2$. Create the sample set of 256 pixels from smoothed image. Run the registration algorithm and stop it when fitness $f \geq 0.1$.
Step 2: Seed a new population with all individuals of the last generation of previous step.
Step 3: Load the original images and create a new sample set of 256 pixels. Run the registration algorithm and stop when condition is reached (for instance $f = 0.99$).

The goal is to apply our proposed method *stepwise refining affine image registration* to an affine transformation with gross distortion.

Results

We repeated 50 times the experiment using the same 256 x 256 pixels images and applying the same transformation. The initial population was created around the identity transformation. Values for μ and λ were 250 and 50. The correct transformation was found 92% of the times . In average 1468 generations were necessary in order to converge to the solution.

In the overall we have proposed a highly promising methodology. To the best of our knowledge no registration method based on affine transforms and evolutionary computation had accomplished this level of consistency and accuracy.

References

[1] T. Back. *Evolutionary Algorithms in Theory and Practice*. Oxford University Press, New York, 1996.
[2] B. Jahne. *Digital Image Processing: Concepts, Algorithms, and Scientific Applications*. Springer Verlag, Berlin, 1997.

Piston Pump Mobile Unity Tour Problem: An Evolutionary View

Marco César Goldbarg **Elizabeth Ferreira Gouvêa** **Francisco Dantas de M. Neto**

DIMAp, UFRN
Campus Universitário, Lagoa Nova
Natal, Brazil, 59 072-084

Abstract

This work presents a version of the Piston Pump Unity Tour Problem. The problem is solved by three evolutionary algorithms designed according to genetic and transgenetic approaches.

1 INTRODUCTION

A Piston Pump Mobile Unity, PPMU, is a vehicle used in the exploitation of onshore petroleum fields to recover the oil of some wells that have not enough energy to lift the oil to the surface. The times needed for the PPMU to move from a well to another one and to install and uninstall its equipment depend on factors such as conditions of the roads, well characteristics and the PPMU operator's experience. These variables can be empirically determined by means of the history of operations. The objective is to determine a sequence of visits of the PPMU to the wells such that the oil volume recovered is maximized, given a certain working period. In this version of the Piston Pump Mobile Unity Tour Problem, PPP, a well can not be considered more than once in the same working period. The PPMU working period can contain more than one tour. Three evolutinary algorithms were implemented to solve the Piston Pump Mobile Unity Tour Problem (PPP): a genetic and two transgenetic ones.

2 ALGORITHMS

For the PPP a chromosome is a sequence of wells that will be visited by the PPMU. The contribution of a well to the fitness of a chromosome is its volume divided by the distance between the well and the last position of the PPMU. The fitness of a chromosome is the summation of the contributions of all the wells in that chromosome. The Genetic Algorithm used the roullete wheel method to select the parents of the next generation. Given two parents, two offspring are generated by fixing one parent and trying to insert one well of the other parent in the sequence of visits of the former. Mutation is based on the wells with low occurrence in the population. One of those wells is chosen at random and inserted into a chromosome.

Computational Transgenetics, CT, is a multiagent proposal that aims at managing the process of information insertion in an evolutionary context. CT considers the epigenetic, the sexual reproduction and the prokaryotic recombination paradigms of evolution. CT provides two classes of evolutionary algorithms: the Extra-Intracellular Transgenetic Algorithms, EITAs (Goldbarg and Gouvêa, 2001), and Proto-Gene (Gouvêa and Goldbarg, 2001). CT infiltrates information in an evolutionary process by means of transgenetic agents. A transgenetic agent may be composed by one or more memes and an operating method. A meme is a proposal to organize a given DNA segment. The memes used on the transgenetic algorithms were pairs of wells. There were generated $n/2$ pairs of wells, where n is the number of wells of an instance. The wells were sorted by oil volume in non crescent order and, to generate a pair, two wells were chosen at random from the first half of the wells, that is the half with highest volumes.

3 COMPUTATIONAL EXPERIMENTS

The approach used in the transgenetic algorithms showed a better performance than the Genetic Algorithm. The ProtoG algorithm presented the better performance for all the instances generated. Furthermore, real PPP problems of a brazilian petroleum company were also solved by the three algorithms presented and, again, ProtoG found the best solutions.

Acknowledgments

This research was funded by Finep/CTPetro and ANP/RH program.

References

M. C. Goldbarg and E. F. Gouvêa (2001). Extra-Intracellular Transgenetic Algorithm, *GECCO 2001 Proceedings of the Genetic and Evolutionary Computation Conference*, 762. San Francisco, CA: Morgan Kaufmann.

E. F. Gouvêa and M. C. Goldbarg (2001). ProtoG: a Computational Transgenetic Algorithm, In *Proceedings of MIC 2001 – 4th Metaheuristics International Conference*, Porto, Portugal, 625-631.

Using Genetic Algorithms To Optimise Guillotine Cutting Operations

Alberto Gómez **David de la Fuente** **Jose Parreño** **Javier Puente**

Dept. de Admon. Empresas y Contabilidad
Universidad de Oviedo
33204 Gijón, Asturias (Spain)

Abstract

This study presents an application to optimise the use of an L- cut guillotine machine. The application has two distinct parts to it; first, a number of rectangular shapes are placed on as few metal sheets as possible by using Genetic Algorithms. Secondly, the sequence for cutting these pieces has to be generated. The guillotine's numeric control then uses this sequence to make the cuts.

1 INTRODUCTION

The application that was designed focuses in particular on optimising the use of the L-type guillotine that the company owns. In contrast to traditional guillotines, which can only cut vertically or horizontally, the L-type guillotine can cut in both these directions simultaneously, which lends greater flexibility to cutting and also leads to more efficient use of the steel sheets.

The programme to be implemented must situate the different pieces that the client has ordered on the metal sheet, bearing in mind that orders are generally made for more than one unit of a particularly dimensioned piece, and that more than one metal sheet is generally required to complete any given order. Positioning should be such as make the maximum use of the sheet material, which is tantamount to minimising the number of metal sheets that are used.

Once positioned, the programme provides the guillotine's numeric control with the cutting sequence of the sheets (the order the pieces are to be cut in). The guillotine begins cutting in the top, right-hand corner of the metal sheet, and successive cuts leave the pieces and the leftover material. To do this, an algorithm was designed based on positioning the pieces within the base surface and generating the corresponding cutting sequence for numeric control.

2 THE SOLUTION PROPOSED FOR CUTTING

The genetic algorithm designed to minimise the number of steel sheets that were used. The steps that were followed to do this will next be described. The codification used in the algorithm is based on integer numbers; each piece to be placed on the metal sheet is assigned a number, and that individual is formed with a string of numbers (a string of parts), The order in which the numbers appear in the string represents the positions on the metal sheet of the pieces.

3 CUTTING THE PIECES

A second algorithm had to be applied in order to instruct the guillotine's numeric control on how to cut the metal sheets in order to obtain the pieces. The algorithm responsible for carrying out this task acquires the distribution of the pieces in the metal sheet from the Genetic Algorithm, and generates an output file that tells the guillotine's numeric control what order to cut the pieces in.

4 RESULTS

To check the quality of the solutions provided by the algorithm, a set of examples have been applied, and it has been shown that the quality of the solutions provided by our algorithm is better in most cases, and only obtains similar results when the number of rectangles to be placed is small, in which cases exact methods can be used anyway.

References

A. Gómez, and D. De la Fuente (2000). Resolution of strip-packing problems with genetic algorithms. Journal of the Operational Research Society, 51:1289-1295.

S. Jakobs (1996). On genetic algorithms for the packing of polygons. *European Journal of Operational Research* 88:165-181.

Optimization of CDMA based Wireless System

Alex C H Ho
Department of Computer Science
City University of Hong Kong
83 Tatchee Avenue, Hong Kong

Sam Kwong
Department of Computer Science
City University of Hong Kong
83 Tatchee Avenue, Hong Kong

Abstract

This paper presents an optimization problem of spreading codes (sequences) in asynchronous Direct Sequence Code Division Multiple Access (DS/CDMA) systems. Bit Error Rate (BER) is one of the performance measurements and it is related to the spreading sequences mathematically. Therefore, the performance of CDMA systems highly depends on the assignment of the unique spreading sequences to the users in which different types and subsets of sequences of a selected sequence type are designed. If the type and the subset of the sequences are not carefully designed, then the Multiple Access Interference (MAI) can be very high. Thus, the channel capacity can be reduced significantly. Hence, proper selection of the type of sequences and its sequence subset are necessary and important to achieve good performance of the system. In this paper, we will use Gold. Kasami and Multiple Spreading [1] sequences as the spreading sequences and apply the genetic algorithm to optimize the sequences such that the MAI is minimized. Although the Multiple Spreading has many advantages over the other PN code, its set of sequences still is not the best if no Genetic Algorithm is applied. Therefore it can be concluded the GA based Multiple Spreading is the best solution to produce PN codes for DS-CDMA system.

1 INTRODUCTION

Direct Sequence (DS) is a well known Spread Spectrum Technique in the current communication system, thus, we are interested in this work. In DS-CDMA, the data signal is multiplied by a Pseudo Random Noise Code (PN code). The codes used for spreading have low cross-correlation values and are unique to every user. This is the reason that a receiver, which has knowledge about the code of the intended transmitter, is capable of selecting the desired signal but other receiver only refer the signal as noise. Multiple Access (MA) is another important consideration in wireless communication system. It means that multiple, simultaneous users can be supported by the wireless system. In other words, a large number of users share a common pool of radio channels and any user can gain access to any channel (each user is not necessary assigned to the same channel). A multiple access method is a definition of how the radio spectrum is divided into channels and how channels are allocated to the many users of the system. In order to maximize the multiple access capacity, direct sequence code division multiple access (DS-CDMA) system which based on the SS technique is one of the best multiple access method. However, its high multiple access capacity is limited by the problem of MAI (MA Interference). So, it will be the first problem to be overcome in order to optimize the performance of DS-CDMA system. This multiple-access interference problem exists as all users can use the same frequency band at any time and it can present a significant problem if the power level of the desired signal is significantly lower (due to distance) than the power level of the interfering user. However, this problem can be improved by generating the set of PN code sequences properly for multiple access.

The main objective of this work is to optimize the DS-CDMA based wireless system. All of the PN code generating techniques, have their own pros and cons. Walsh-Hadamard codes is orthogonal but as it uses only discrete carrier frequency, it does not have a single, narrow autocorrelation peak and also the spreading is not over the whole bandwidth. M-sequences, Gold-codes and Kasami-codes can spread over the whole bandwidth and Kasami-codes even have a large code set but all of them are non-orthogonal. So, the system using these codes is easier to suffer from the problem of MAI. Their disadvantages make them impossible to maximize the multiple access capacity and thus degrade the performance of DS-CDMA system. It seems that Multiple Spreading is the best method to generate PN codes as it can provide orthogonal codes within a cell while maintaining mutual randomness between users of different cells. Our work is aimed at finding an algorithm to generate a set of PN codes sequences which can suppress the MAI and so maximize the multiple access capacity of a DS-CDMA system in order to optimize the performance of a DS-CDMA based communication system. The performance of several PN codes generating algorithms will be compared, including Gold code, Kasami code and Multiple Spreading. Simulated Annealing and Genetic Algorithm will also be analyzed to find out the best algorithm to optimize the performance of a DS-CDMA system. The analysis has shown us that the GA based PN codes have better SNR and lower BER. When it is compared with another optimization algorithm - Simulated Annealing, the result shows that Genetic Algorithm is better.

References.

[1] Esmael H. Dinan and Bijan Jabbari "Spreading Codes for Direct Sequence CDMA and Wideband CDMA Cellular Networks", IEEE Communications Magazine, September 1998

Modeling Convection Coefficients with Genetic Algorithms

Zhou Ji **Dipankar Dasgupta**
Division of Computer Science
Department of Mathematical Sciences
The University of Memphis
Memphis, TN 38152

1 INTRODUCTION

Natural convection is an important mode of heat transfer in many fields. Convection coefficient, Nusselt number *Nu*, is usually modeled as a function of Rayleigh number *Ra*, which describes the conditions that affect convection. Inclination and spacing are important parameters of stacked cylinders. Of the latest models for two-cylinder array includes both inclination θ and spacing d as parameters (Ji, et al 1998). The coefficients were obtained from experimental data by an empirical combination of linear regression and exponential function fitting. This paper uses GA to seek for better coefficients.

2 APPLICATION OF GENETIC ALGORITHM

We assume the three coefficients *A*, *B*, and *n* in $Nu = (A + B \cdot Ra\cos\theta)D^n$ are within the value ranges of [0, 10], [0,1], and [0,1], respectively. A chromosome is encoded as a set of three coefficients {*A*, *B*, *n*}.

The objective function (f) is the mean-square-root of the deviations between the model and the experimental data at all the experimental points. Then individual's ranking of *F* is used to decide the selection probability.

Binary representation with ranking selection scheme is used. The initial population is generated randomly. Single point crossover is used. Crossover rate is chosen to be 60%. Mutation is applied to each of the three coefficients separately in the chromosomal representation. Elitism is used where the best individual is always passed to the next generation unchanged.

3 EXPERIMENTAL RESULTS ALGORITHM

The GA program was developed in Visual C++ with *chrom*, *popul*, *convecDoc* and *convecView* as major functional components.

An improved model for Nusselt number is obtained.

Top cylinder:
$Nu = (2.2032 + 0.307\times10^{-3} Ra\cos\theta)\, d^{0.2295}$.

Bottom cylinder:
$Nu = (2.5891 + 0.507\times10^{-3} Ra\cos\theta)\, d^{0.0444}$.

The mean-square-root deviations of them from the experimental data are 0.1218 and 0.1065 for the top and bottom cylinders, respectively. Those of the original model are 0.2854 and 0.3970. They are cut by 57% and 73% using GA.

To analyze GA performance, tests with some variations of parameters were compared. Average of 20 runs is taken for each choice.

Populations size 25, 50, 100 are compared with selection pressure 1.5 and mutation rate 0.1. Some tests with population 25 did not converge. Selection pressures 1, 1.25, 1.5, 1.75, and 2 are compared. Higher selection rate did not improve the performance significantly. Results from different mutation rate from 0 through 1are compared. For mutation rate 0.01, only 1 out of the 20 runs converged. On the other hand, large mutation rate up to close to 1 all resulted to the optimum. This result suggests that either the fitness landscape of this problem is very rugged, or the crossover operation does not capture the good schema effectively.

4 CONCLUSIONS

Modeling of Nusselt numbers of two vertically stacked inclined cylinders is improved significantly over the original method. On the other hand, it also confirmed that the relative dependence of *Nu* on the three coefficients *Ra*, θ and *d*. The method looks promising to obtain better model while there is much to improve in the implementation.

REFERENCES

Dasgupta, D and Z. Michalewicz,Editorsn (1997), *Evolutionary Algorithms in Engineering Applications*, Published by Springer Verlag, Inc.Berlin.

Shklover, G. C., and S. E. Gusev (1988). Natural Convection Heat Transfer in Horizontal Tubular Bundles, *Heat Transfer-Soviet Research*, V. 20, p.746.

Zhou Ji, Hua Zhu, and S. Gir (1998). An Experiment on the Effect of Spacing on Natural Convection from Vertical Stacked Inclined Cylinder with a Constant Heat Flux, Project Report (Mechanical), University of Memphis.

Constructing X-of-N Attributes with a Genetic Algorithm

Otavio Larsen Alex A. Freitas Julio C. Nievola

Postgraduate Program in Applied Computer Science
Pontificia Universidade Catolica do Parana
Rua Imaculada Conceicao, 1155. Curitiba – PR. 80215-901. Brazil.
{larsen, alex, nievola}@ppgia.pucpr.br
http://www.ppgia.pucpr.br/~alex

Abstract

The predictive accuracy obtained by a classification algorithm is strongly dependent on the quality of the attributes of the data being mined. When the attributes are little relevant for predicting the class of a record, the predictive accuracy will tend to be low. To combat this problem, a natural approach consists of constructing new attributes out of the original attributes. Many attribute construction algorithms work by simply constructing conjunctions and/or disjunctions of attribute-value pairs. This kind of representation has a limited expressiveness power to represent attribute interactions. A more expressive representation is X-of-N [Zheng 1995]. An X-of-N condition consists of a set of N attribute-value pairs. The value of an X-of-N condition for a given example (record) is the number of attribute-value pairs of the example that match with the N attribute-value pairs of the condition. For instance, consider the following X-of-N condition: X-of-{"Sex = male", "Age < 21", "Salary = high"}. Suppose that a given example has the following attribute-value pairs: {"Sex = male", "Age = 51", "Salary = high"}. This example has 2 out of the 3 attribute-value pairs of the X-of-N condition, so that the value of the X-of-N condition for this example is 2.

In our GA an individual represents a X-of-N attribute, i.e. the set of N attribute-value pairs composing a X-of-N attribute. Each attribute-value pair is of the form $A_i = V_{ij}$, where A_i is the i-th attribute and V_{ij} is the j-th value belonging to the domain of the A_i. The current version of our GA can cope only with categorical attributes. (Continuous attributes are discretized in a preprocessing step.) The value of N is an integer number varying from 2 to 7. The fitness function is the information gain ratio of the constructed attribute.

In order to evaluate how good the new attributes constructed by the GA are, we have compared the performance of the C4.5 algorithm using only the original attributes with the performance C4.5 using both the original attributes and the new attributes constructed by the GA. Hereafter we refer to the former and to the latter as the original data set and the extended data set, respectively. The performance of C4.5 in both the original data set and the extended data set was measured with respect to the classification error rate. The experiments were done by using public-domain data sets available from http://www.ics.uci.edu/~mlearn/MLRepository.html.

The results are shown in Table 1. The results for the first four data sets of Table 1 were produced by a 10-fold cross-validation procedure. The results for the last three data sets (the monks data sets) were obtained by using the predefined partition of the data into training and test sets. The second and third columns of Table 1 show the error rate obtained by C4.5 in the original data set and the extended data set (with the new X-of-N attributes), respectively. The numbers after the symbol "?" denote standard deviations. For each data set, the difference in the error rates of the second and third columns is deemed to be significant when the two error rate intervals (taking into account the standard deviations) do not overlap. When the error rate of the "original + X-of-N" attributes is significantly better (worse) than the error rate of the "original" attributes, there is a "(+)" (("-")) sign in the third column. Note that the X-of-N attribute constructed by the GA significantly improved the performance of C4.5 in three data sets (tic-tac-toe, promoters and monks-2), and it significantly degraded the performance of C4.5 in just one data set (monks-3). In the other three data sets the difference in the error rates was not significant.

Table 1: Error rate obtained by C4.5 in seven data sets

Data Set	Error Rate (%)	
	Original attributes	original + X-of-N attrib.
hepatitis	22.84 ? 1.83	26.34 ? 4.99
Wisc. breast cancer	4.57 ? 0.64	4.83 ? 0.78
tic-tac-toe	14.31 ? 1.14	5.34 ? 0.47 (+)
promoters	18.88 ? 2.19	13.25 ? 1.98 (+)
monks-1	0.00 ? 0.00	0.00 ? 0.00
monks-2	29.60 ? 0,04	0.00 ? 0.00 (+)
monks-3	0.00 ? 0.00	2.80 ? 0,01 (–)

Acknowledgment

This research project is financially supported by Motorola, Jaguariuna –SP, Brazil.

Reference

[Zheng 1995] Z. Zheng. Constructing nominal X-of-N attributes. *Proc. 14th Int. Joint Conf. on Artificial Intelligence*, 1064-1070. Morgan Kaufmann, 1995.

An Efficient Genetic Algorithm for Fixed Channel Assignment Problem with Limited Bandwidth Constraint

Shouichi Matsui **Isamu Watanabe** **Ken-ichi Tokoro**
Communication & Information Research Laboratory (CIRL)
Central Research Institute of Electric Power Industry (CRIEPI)
2-11-1 Iwado-kita, Komae-shi, Tokyo 201-8511, JAPAN
E-mail: {matsui,isamu,tokoro}@criepi.denken.or.jp, Phone: +81-3-3480-2111

Abstract

This paper proposes a new efficient genetic algorithm for a fixed channel assignment problem with limited bandwidth constraint. Results are given which show that our GA produces far better solutions than the previously proposed GA the problem.

1 INTRODUCTION

The channel assignment problem (CAP), or frequency assignment problem (FAP) is a very important problem today, but is a difficult, NP-hard problem. To achieve the optimal solution of fixed channel assignment problems, most proposed algorithms try to minimize the amount of necessary channels under satisfying a set of given constraint (e.g., Matsui and Tokoro, 2001).

However, the total number of available channels, or bandwidth of frequencies, are given and fixed in many situations. Minimizing the bandwidth becomes meaningless for such applications. To address the problem, Horng et al. proposed a new cost model, where the objective is to minimize the total cost, the sum of the cost of blocked calls and the cost of interference. They also proposed a GA to solve the problem (Horng et al., 2001). But the performance of the GA is not good enough.

We propose a new genetic algorithm for the problem. The proposed GA is tested using the same problems by Horng et al., and the performance is far better than their GA.

2 THE PROPOSED ALGORITHM

The main idea of the proposed GA is to use a chromosome representation that encodes sequences of codes of a virtual machine that performs assignment. The GA searches a good sequence of codes that minimize the total cost.

The proposed steady-state GA uses a local search and a modified version of the adaptive mutation scheme proposed by Matsui and Tokoro (2001).

Table 1: Solutions of the problems.

	GA by Horng et al.		Proposed GA	
P	Best	Average	Best	Average
P1	203.4	302.6	3.7e-4	3.7e-4
P2	271.4	342.5	4.1	4.1
P3	1957.4	2864.1	5.8	7.4
P4	906.3	1002.4	243.8	249.7
P5	4302.3	4585.4	564.1	710.5
P6	4835.4	5076.2	594.7	746.0
P7	20854.3	21968.4	3322.0	3832.8
P8	53151.7	60715.4	28976.1	30815.6

3 EXPERIMENTS AND RESULTS

We tested the performance of the GA by running 100 times for 8 problems proposed by Hong et al.(2001). For all the experiments, the population size was 100, the maximal number of generation was 1000, the crossover ratio was 0.8, and the number of new offsprings was equal to the population size.

The comparison with the results by Horng et al.(2001) are shown in Table 1. The column *P* denotes the problem name, the column 'Best' shows the minimum cost found in the 100 runs, and the column 'Average' shows the average cost over the 100 runs. Table 1 shows that our GA performs very well, it outperforms the previous GA (Horng et al., 2001) for all cases, the cost obtained by our GA is very small compared to the ones by theirs.

References

Horng, J.T., M.H. Jin, and C.Y. Kao (2001). Solving fixed channel assignment problems by an evolutionary approach, *Proc. of GECCO-2001*, pp.351–358.

Matsui, S. and K. Tokoro (2001). Improving the performance of a genetic algorithm for minimum span frequency assignment problem with an adaptive mutation rate and a new initialization method, *Proc. of GECCO-2001*, pp.1359–1366.

Genetic Programming for Attribute Construction in Data Mining

Fernando E. B. Otero **Monique M. S. Silva** **Alex A. Freitas**

Pontificia Universidade Catolica do Parana (PUC-PR)
Postgraduate program in applied computer science
Rua Imaculada Conceicao, 1155
Curitiba – PR. 80215-901. Brazil
{fbo, mmonique, alex}@ppgia.pucpr.br
Tel./Fax: (55) (41) 330-1669 (c/o Alex Freitas)
Web page: www.ppgia.pucpr.br/~alex

This paper addresses the classification task of data mining. The goal of attribute construction is to construct new attributes out of the original ones, transforming the original data representation into a new one where regularities in the data are more easily detected.

We use a standard tree-structure representation for each individual. The GP constructs new attributes out of the continuous (real-valued) attributes of the data set being mined. Each individual corresponds to a candidate new attribute. The terminal set consists of all the continuous attributes in the data being mined. The function set consists of four arithmetic operators (+, -, *, /) and two relational operators, namely "≤", "≥". We used tournament selection, standard tree crossover and point mutation. The fitness function was information gain ratio.

The experiments were performed with four public-domain data sets from the UCI data set repository, available at: `http://www.ics.uci.edu/~mlearn/MLRepository.html`. We compared the classification error rate of C4.5 using only the original attributes with the error rate of C4.5 using both the original attributes and the new attribute constructed by the GP. We did experiments with 3 values of the tournament size *k* and 3 values of maximum tree size (number of nodes). For each of the 9 combinations of these parameters values we ran a 10-fold cross-validation.

The results are reported in Tables 1, 2, 3, and 4. In the title of each table we report the data set and the error rate (in %) obtained by C4.5 using only the original attributes. In the tables themselves, each cell contains the error rate obtained by C4.5 using the attribute constructed by GP. The value of a cell is in bold if the error rate obtained using the attribute constructed by the GP is smaller than the error rate obtained using only the original attributes. The numbers after "±" denote standard deviations. In general, the results can be summarized as follows. In the Abalone data set the attribute constructed by GP led to a slight increase in error rate, but this increase was not significant. In the Wine data set the attribute constructed by GP lead to some reduction in error rate, but again this reduction was not significant. In the other two data sets (Balance-Scale and Waveform) the attribute constructed by GP led to a reduction in error rate which was significant – the corresponding error rate intervals (considering the standard deviations) do not overlap. In general, in the four data sets the differences in error rates associated with different combinations of parameter values was not significant, showing that GP was quite robust to variations in these two parameters.

Table 1: (Abalone) error rate of original attr.: 79.2 ± 0.37

	Maximum tree size (number of nodes)		
k	31	63	127
2	79.31 ± 0.36	**79.18 ± 0.35**	79.21 ± 0.41
4	79.21 ± 0.33	79.21 ± 0.33	**79.16 ± 0.36**
8	79.21 ± 0.33	79.21 ± 0.33	**79.16 ± 0.36**

Table 2: (Balance-scale) error of orig. attr.: 22.42 ± 1.34

	Maximum tree size (number of nodes)		
k	31	63	127
2	**11.47 ± 2.13**	**7.78 ± 0.66**	**8.58 ± 0.58**
4	**9.06 ± 0.42**	**8.26 ± 0.65**	**8.26 ± 0.44**
8	**8.58 ± 0.67**	**8.74 ± 0.68**	**8.10 ± 0.63**

Table 3: (Waveform) error rate of orig. attr.: 25.06 ± 0.66

	Maximum tree size (number of nodes)		
k	31	63	127
2	**23.04 ± 0.40**	**22.48 ± 0.45**	**22.68 ± 0.57**
4	**22.44 ± 0.59**	**22.86 ± 0.50**	**22.56 ± 0.57**
8	**22.22 ± 0.49**	**22.60 ± 0.42**	28.86 ± 2.74

Table 4: (Wine) error rate of original attr.: 6.48 ± 2.05

	Maximum tree size (number of nodes)		
k	31	63	127
2	**5.31 ± 1.38**	**4.72 ± 1.47**	**4.72 ± 1.47**
4	**3.54 ± 1.30**	**5.31 ± 1.63**	**3.54 ± 1.30**
8	**3.54 ± 1.30**	**4.72 ± 1.47**	**5.30 ± 1.62**

Evolving Good Recommendations

Supiya Ujjin **Peter J. Bentley**

Department of Computer Science, University College London
Gower Street, London S.Ujjin@cs.ucl.ac.uk
+44 (0)20 7679 2878

1 SYSTEM OVERVIEW

Recommender systems are new types of internet-based software tools, designed to help users find their way through today's complex on-line shops and entertainment websites. This paper focuses on the use of evolutionary search to fine-tune a profile-matching algorithm within a recommender system, to find profiles similar to the current user (or *active user*, *A*). Selected data from those profiles are then used to build recommendations. By evolving profile-matching, we tailor it to the preferences of individual users. This enables the recommender system to make more accurate predictions of users' likes and dislikes, and hence better recommendations to users.

In this research, the MovieLens dataset was used for initial experiments. The evolutionary recommender system uses 22 features from this data set: movie rating, age, gender, occupation and 18 movie genre frequencies.

Before recommendations can be made, the movie data is processed into separate profiles, one for each person, defining that person's movie preferences. We define *profile(j,i)* to mean the profile for user *j* on movie item *i*, see fig. 1. The profile of *j*, *profile(j)* is therefore a collection of *profile(j,i)* for all the items *i* that *j* has seen.

1	2	3	4	..22
Rating	Age	Gender	Occupation	18 Genre frequencies
5	23	0	45	000000100010000000

Figure 1: *profile*(*j*,*i*) - profile for user *j* with rating on movie item *i*, if *i* has a rating of 5.

Once profiles are built, the process of recommendation can begin. Given an active user *A*, a set or neighbourhood of profiles similar to *profile(A)* must be found.

The neighbourhood selection algorithm consists of three main tasks: *profile selection*, *profile matching* and *best profile collection*. It is not always feasible to use the entire database of profiles to select the best possible profiles. As a result, most systems opt for random sampling – performed by *profile selection*. Next, the profile matching process computes the distance or similarity between the selected profiles and the active user's profile using a modified Euclidean distance function (employing multiple features such as user's age, gender and movie genres). Every user places a different importance or priority on each feature. Our approach shows how weights defining user's priorities can be evolved by a genetic algorithm. Once all the Euclidean distances have been found, the *best profile collection* picks users most similar to *A* to form the neighbourhood of *A*. Because the neighbourhood set contains those users who are most similar to *A*, movies that these users like have a reasonable probability of being liked by *A*.

To calculate a fitness measure for an evolved set of weights for the active user, *w*(*A*), the recommender system finds a set of neighbourhood profiles for *A*. Three movie items that *A* has seen are then selected, where items with more ratings have a higher probability of being picked. The ratings of these users are then employed to compute the predicted rating for *A* on each movie item. The predicted vote, *predict_vote*(*A*,*i*), for *A* on item *i*, can be defined as:

$$predict\_vote(A,i) = mean_A + k\sum_{j=1}^{n} euclidean(A,j)(vote(j,i) - mean_j)$$

where: $mean_j$ is the mean vote for user *j, k* is a normalising factor such that the sum of the euclidean distances is equal to 1, *vote*(*j*,*i*) is the actual vote that user *j* has given on item *i, n* is the size of the neighbourhood.

Because *A* has already rated the movie items, it is possible to compare the actual rating with the predicted rating. So, the average of the differences between the three actual and predicted votes are used as fitness score to guide future generations of weight evolution.

2 EXPERIMENTS

Four sets of experiments were designed to observe the difference in performance between the evolutionary recommender system and a standard, non-adaptive recommender system based on the Pearson algorithm.

In all experiments, the GA recommender performed equally well (or better) than the Pearson algorithm. The results also suggested that random sampling is a good choice for the *profile selection* task of retrieving profiles from the database.

3 CONCLUSIONS

In conclusion, experiments demonstrated that, compared to a non-adaptive approach, the evolutionary recommender system was able to successfully fine-tune the profile matching algorithm. This enabled the recommender system to make more accurate predictions, and hence better recommendations to users.[1]

[1] Full details of this work can be found on-line: http://www.cs.ucl.ac.uk/staff/S.Ujjin/

Evolving Finite Automata with Two-Dimensional Output for DNA Recognition and Visualization

Edgar E. Vallejo
Computer Science Department
Tecnológico de Monterrey
Campus Estado de México, México
evallejo@campus.cem.itesm.mx

Fernando Ramos
Computer Science Department
Tecnológico de Monterrey
Campus Morelos, México
framos@campus.mor.itesm.mx

Abstract

This article presents a computational model for biosequence recognition and visualization. Finite automata with 2-D output recognize and create maps of DNA sequences. Populations of such machines evolve discrimination capabilities using genetic algorithms. Experimental results indicate that evolved machines are capable of recognizing HIV sequences in a collection of training and validation sets. In addition, we found that these sequences map to similar structures on the cartesian space.

1 FINITE AUTOMATA WITH TWO-DIMENSIONAL OUTPUT

A finite automaton with two-dimensional output is a special type of finite automata (Hopcroft and Ullman, 1979). It has a read-only input tape. The input tape head can read one symbol at a time, and in one move the input tape head shifts one square to the right. The output of the machine is produced on the cartesian space. The output tape head writes one symbol at the current position and in one move the output tape head advances one coordinate (up, down, left or right). The machine either accept or reject each sequence.

2 EXPERIMENTS

We applied a genetic algorithm to a population of finite automata with two-dimensional output to evolve the ability of discriminating a collection of DNA sequences from a collection of training and validation sets. These sets consist of positive and negative examples of HIV sequences. We used a generational genetic algorithm with tournament selection and elitism (Mitchell, 1996). Genome representation was derived from the concatenation of values of the finite automata state transition function (Vallejo and Ramos, 2001). Fitness was defined as the number of sequences correctly classified.

3 RESULTS

We performed several runs with this model. In most runs, the evolutionary process yielded finite automata capable of classifying all sequences correctly. In addition, we found that HIV sequences map to very similar structures on the cartesian space.

4 CONCLUSIONS

In this work, we extended the finite automata model to perform biosequence recognition and two-dimensional output. We evolved a population of such machines to perform sequence recognition. The evolutionary process yielded machines with recognition and generalization capabilities. The evolved machines were capable of producing a convenient biosequence representation.

References

J.E. Hopcroft, J.D. Ullman (1979). *Introduction to Automata Theory, Languages and Computation.* Addison Wesley Publishing Company.

M. Mitchell (1996). *An Introduction to Genetic Algorithms.* The MIT Press.

D.W. Mount (2001). *Bioinformatics: Sequence and Genome Analysis.* Cold Sring Harbor Laboratory.

E.E Vallejo, F. Ramos (2001). Evolving Turing Machines for Biosequence Recognition and Analysis. In J. Miller, M. Tomassini, P.L. Lanzi, C. Ryan, A.G.B. Tettamanzi W.B. Langdon (eds.) *4th European Conference on Genetic Programming, EuroGP2001* LNCS 2038. pp. 192-203. Springer-Verlag.

Real-World Shop Floor Scheduling by Ant Colony Optimization

Andre Vogel*
Dept. of Economics
Technical University
D-09107 Chemnitz
Germany

Marco Fischer‡
Dept. of Economics
Technical University
D-09107 Chemnitz
Germany

Tobias Teich¶
Dept. of Economics‖
Technical University
D-09107 Chemnitz
Germany

Summary

Manufacturing Control Problems are still often solved by manual scheduling, that means only by out of the experience of the workers. Modern algorithms, such as Ant Colony Optimization, have proved their capacity to solve this kind of problems in theory.

The objective of this work was to compare the developed Ant Algorithm to the results achieved by a Genetic Algorithm tested on the same real-world data.

The algorithm was coupled to the ERP-System (Daamgard Axapta was used) in the way, displayed in figure 1.

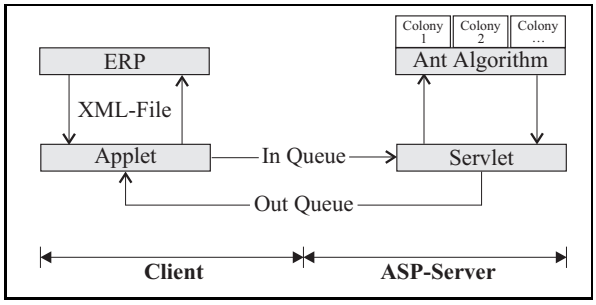

Figure 1: Communication structure

The shop floor scheduling problem was tested on real world data provided by a German engineering company. In order to be able to dispose of real world problem sizes and data, a set of numbers was recorded within a period of two months (e.g. release date, the finishing date and the due date of each job, ...).

The used algorithm based on an already applied one [1], which was successfully tested on the widely known Job Shop Scheduling.

*Email: andre.vogel@wirtschaft.tu-chemnitz.de
‡Email: marco.fischer@wirtschaft.tu-chemnitz.de
¶Email: t.teich@wirtschaft.tu-chemnitz.de
‖phone: ++49-371-531 42 52

In order to compare the reached solution quality, we used the mean flow time and the mean lateness of jobs. For the comparison of the results achieved by the means of the Ant Algorithm and other methods we used a priority rule scheduler (PRIORULE) and GACOPA [2]. The priority rule scheduler was allowed to calculate a number of single, standard priority rules (e.g.: shortest processing time, earliest due dates etc.) and combinations of these rules. The best out of the achieved results was used. Manual scheduling achieved a mean lateness of 2.4 days and a mean flow time of 13.2 days.

Table 1: Relative results (manual scheduling = 100%)

scheduler	*mean lateness*	*mean flowtime*
MANUAL	100.0 %	100.0 %
PRIORULE	95.3 %	97.9 %
GACOPA	55.3 %	85.0 %
ANTS	61.8 %	87.7 %

The following parameter settings were used: $\alpha = 1$; $\beta = 1$; $\rho = 0,01$; $c = 0,8$; $\gamma = 0,8$; $q = 2$. The results presented in table 1 show that ants perform very well, but not as good as the GACOPA the results are compared with.

This work has been supported by the Collaborative Research Center 457 and the CBS GmbH.

References

[1] Fischer, M., Vogel, A., Teich, T., Fischer, J.: A new Ant Colony Algorithm for the Job Shop Scheduling Problem. in: Proceedings of the Genetic and Evolutionary Computation Conference, San Francisco, 2001.

[2] Kaeschel, J., Meier, B., Fischer, M., Teich, T.: Real-World Applications: Evolutionary Real World Shop Floor Scheduling using Parallelization and Parameter Coevolution, in: Proceedings of the Genetic and Evolutionary Computation Conference, Las Vegas, 2000.

A Comparison of Genetic Algorithm Methods in Aerial Spray Deposition Management

L. Wu
AI Center
University of Georgia
Athens, GA 30602
wulei@arches.uga.edu
706-559-9553

W.D. Potter
AI Center
University of Georgia
Athens, GA 30602
potter@cs.uga.edu
706-542-0361

K. Rasheed
AI Center
University of Georgia
Athens, GA 30602
khaled@cs.uga.edu
706-542-3444

H. Thistle, J. Ghent, D. Twardus
USDA Forest Service

M. Teske
Continuum Dynamics
Princeton, NJ

Abstract

In this paper, we describe several genetic algorithm methods to deal with the spray parameter optimization problem, and compare them with the original heuristic method SAGA.

1 SUMMARY

SAGA (Spray Advisor using Genetic Algorithm) was developed to heuristically search for an optimal or near-optimal set of input parameters needed to achieve a certain aerial spray deposition (Potter et al., 2000). It is a typical parametric design problem, in which the genetic algorithm evolves the spray parameters in a continuous domain, and the simulation engine evaluates the solution. The search space is rather huge, so the simple GA has difficulty converging to a global optimum.

First, we developed SAGA2 from SAGA. The original generational genetic algorithm is replaced by a steady-state genetic algorithm. The original roulette wheel selection is replaced by tournament selection. Instead of using a single type of crossover and mutation operator, in SAGA2 we combine several kinds of crossover and mutation operators and apply them with different probabilities. Second, we developed SAGA2NN from SAGA2. SAGA2NN generates the initial population from a large pool of individuals. In order to save computation time, it uses a neural network to approximate the fitness. The learning rule of the neural network is backpropagation with momentum. SAGA2NN also applies this method in the process of crossover and mutation. Then, we developed SAGADO by applying GADO (Genetic Algorithm for Design Optimization) in this problem. GADO is a general-purpose approach to solving the parametric design problem (Rasheed, 1998). Simulated annealing is also a widely used global stochastic optimization technique. In order to compare its performance with the genetic algorithm, we developed SASA based on simulated annealing. The cooling schedule adopted is geometric cooling. The temperature is updated using the formula: $T_{i+1} = \alpha T_i$ ($\alpha \in (0, 1)$).

2 RESULTS AND ANALYSIS

We ran these methods on three practical spray parameter specifications provided by Forest Service managers. Each method ran five times with different random seeds. Evolution process of one of these parameter settings is shown in Fig. 1. SAGA2 and SAGADO outperform SAGA in all experiments. It shows that exquisite choice of type of GA, selection, crossover and mutation operator can boost GA performance. SAGA2NN converges very fast, which is useful in real aerial spray applications. We hoped it could achieve even better results, but our experiments don't show that. We think the reason is that the advantage of the neural network is counteracted by premature convergence of the genetic algorithm. SASA performs much poorer than the genetic algorithm methods. It shows that genetic algorithms perform better than simulated annealing in an extremely complicated domain such as aerial spray deposition management.

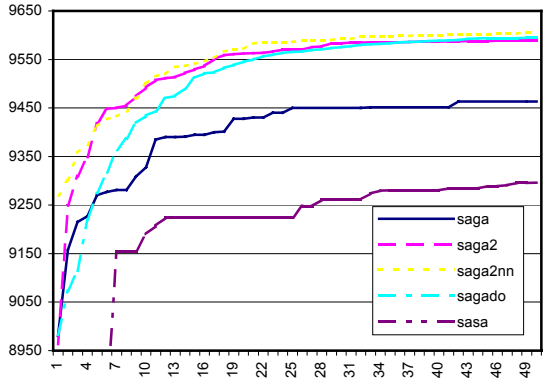

Fig. 1 Evolution Process of Parameter Setting III

References

W.D. Potter, W. Bi, D. Twardus, H. Thistle, M.J. Twery, J. Ghent, M. Teske (2000). Aerial Spray Deposition Management Using the Genetic Algorithm. IEA-EIA Conference 2000.

K. Rasheed (1998). GADO: A Genetic Algorithm for Continuous Design Optimization. Ph.D. Thesis, Department of Computer Science, Rutgers University, New Brunswick, NJ, 1998.

Functional Test Generation for Digital Integrated Circuits Using a Genetic Algorithm

Xiaoming Yu† Alessandro Fin‡ Franco Fummi‡ Elizabeth M. Rudnick†

† Center for Reliable & High-Performance Computing, University of Illinois, Urbana, IL
‡ Dipartimento di Informatica, Verona, ITALY

In order to manage the complexity and ensure the correctness of modern integrated circuit (IC) designs, such as microprocessors, designers usually adopt a top-down approach. Given the functional specification of a design, designers translate it into a description written in some hardware description language, such as VHDL, Verilog, or SystemC. This initial design description is typically entered at the register transfer level (RTL). However, a higher level behavioral description may be used instead before the design is further refined to the RTL. The RTL is subsequently synthesized into a network of transistors or logic gates, known as the gate-level description. Using this design process, designers can focus on the most important part of a design. Since a design may go through behavioral, register transfer, and gate levels, it is important to verify the initial description and also the functional equivalence of descriptions at different levels of abstraction. Insufficient verification can lead to product recalls when bugs are finally detected by customers, which can be extremely costly. Simulation-based approaches, which involve applying level-dependent test generation techniques, are typically used for verifying a design. Tests are generated after each level of the design is completed and then simulated at that level and the higher level. Results are compared across different levels to check if there is a mismatch. However, test generation is a difficult and time-consuming process.

Genetic algorithms (GA's) have been shown to be effective when solving state space search problems. In the area of test generation for manufacturing test of digital ICs, GA's have been effectively used to solve the gate-level test generation problem. In this work, GA's are proposed to generate functional tests for higher levels that can be reused at lower levels in a homogeneous design and test generation environment. The functional tests can be used for design verification and also for development of tests for manufactured parts. In the proposed method, a GA is first used to generate tests at the behavioral level for descriptions written in SystemC. As the design is translated into RTL, the generated tests are simulated first and another GA is used to generate additional tests to obtain better coverage using coverage metrics targeted at the more detailed design. Then at the gate level, vectors generated at the behavioral and register transfer levels are reused before a gate-level test generator is applied. By reusing functional tests from higher levels, the overall test generation time can be decreased, thus reducing the development period of the design. Furthermore, the functional tests generated at behavioral and register transfer levels are useful in identifying design errors during the design verification process.

A simple GA was used to repeatedly generate test sequences to target statement coverage, i.e., coverage of the statements in a SystemC or RTL description, and path coverage, i.e., coverage of all paths in the SystemC or RTL description. The fitness value of each individual of the GA is determined by the number of covered statements or paths. The GA tries to cover each statement or path several times so that each bit error, which reproduces the single stuck-at fault at behavioral and register transfer levels, gets a chance to be detected. Each sequence generated by the GA is fault simulated to remove the detected bit errors. The GA stops when no more bit errors can be detected. By targeting statement coverage or path coverage, only logic simulation, which is much faster than fault simulation, is involved in the GA, and it is performed only once for each candidate sequence. This is the first reported approach for targeting path coverage automatically.

Experiments were carried out on a sample of the ITC99 benchmarks to demonstrate the effectiveness of the proposed functional test generation method. The GA was used to generate test sequences at the RTL targeting 100% path coverage. The sequences were then simulated at the gate level to remove detected faults before a gate-level test generator was applied. Experimental results show that the fault coverages achieved by the proposed method are greater than or equal to those of gate level test generation alone for five out of six circuits in less time. For the two large circuits b07 and b11, significantly higher fault coverages are achieved in much less test generation time.

Multiobjective Evolutionary Algorithm Approach for Solving Integer Based Optimization Problems

Jesse B. Zydallis, Todd A. Sriver and Gary B. Lamont*
Department of Electrical and Computer Engineering
Graduate School of Engineering and Management
Air Force Institute of Technology WPAFB OH 45433
{Jesse.Zydallis, Todd.Sriver, Gary.Lamont}@afit.edu

Abstract

Multiobjective Evolutionary algorithms (MOEAs) are often well-suited for complex combinatorial Multiobjective optimization problems (MOPs). Integer based MOPs are prevalent in real world applications where there exist a discrete amount of a component or quantity of an item. Presented here is the application of a building block based MOEA, the MOMGA-II, to a NP Complete problem and real-world application. Excellent results are obtained.

1 INTRODUCTION

Over the last few years we have shown the efficiency and effectiveness of our Multiobjective Messy Genetic Algorithm (MOMGA) and MOMGA-II as applied to various unconstrained MOPs [1]. We expand our MOP test suite to include integer based, constrained MOPs. The MOMGA-II is applied to the NP Complete (NPC) knapsack problem and a real-world logistics problem. This algorithm is unique in its explicit manipulation of Building Blocks (BBs) and variable string lengths to find the solution to complex MOPs as compared to other MOEAs.

2 INTEGER OPTIMIZATION

Integer optimization problems are a class of real-world problems encountered in industry, science, engineering, and government. In the first MOP utilized, the 0/1 knapsack problem, one has a knapsack to place items of value in, where each item has a particular weight and value associated with it. There is a constraint that the knapsack has a maximum capacity and that only whole items may be placed within the knapsack. The multiobjective formulation of this problem includes an arbitrary number of knapsacks that the overall profit is to be maximized.

The second MOP analyzed here is the advanced logistics problem (MOP-ALP) which is concerned with allocating appropriate resources for use in various tasks. It is a real-world optimization problem with strict equality constraints. This MOP formulation consists of three objective functions to minimize.

3 CONCLUSIONS

The results of the MOMGA-II when applied to a NPC and a real-world highly constrained optimization problem illustrate that the MOMGA-II is capable of solving integer based problems, constrained integer based problems and highly constrained integer based problems with strict equality constraints. The MOMGA-II obtains results that are favorable when compared with other well-known MOEAs for the knapsack problem. Having success with the NPC multiple knapsack problem, and the real world MOP-ALP, the MOMGA-II can be applied to other multiobjective versions of NPC problems and real-world problems that map to them.

References

[1] Jesse B. Zydallis and Gary Lamont. Solving of Discrete Multiobjective Problems using an Evolutionary Algorithm with a Repair Mechanism. In *Proceedings of the 44th IEEE 2001 Midwest Symposium on Circuits and Systems (MWSCAS 2001)*, volume 1, pages 470–473, Dayton, OH, August 2001.

*The views expressed in this article are those of the authors and do not reflect the official policy or position of the United States Air Force, Department of Defense, or the U.S. Government.