Evolutionary Algorithms for the Minimum Tardy Task Problem

Enrique Alba¹, Guillermo Leguizamón², and Guillermo Ordóñez²

¹ Universidad de Málaga, Complejo Tecnológico, Campus de Teatinos, 29071 Málaga, Spain. eat@lcc.uma.es
² LIDIC Universidad Nacional de San Luis

² LIDIC - Universidad Nacional de San Luis, Ejército de Iso Andes 950 (5700) San Luis, Argentina legui@unsl.edu.ar

Abstract. This paper compares the performance of several evolutionary algorithms for solving the combinatorial optimization problem known as the "minimum tardy task problem". A generational, a steadystate, and a cellular genetic algorithm, plus a new ant colony optimization approach to this problem are evaluated on different instances of the problem. The algorithms have all reported high quality results although no problem-specific changes of the three first evolutionary algorithms were made other than in the fitness function. Despite the high quality results of the genetic algorithms, the ant algorithm worked out the faster, more stable, and better scalable results. We show in this paper that optimum results can be got with basic techniques that are only sampling a tiny fraction of the search space.

keywords: evolutionary algorithms, ant systems, combinatorial optimization, decentralized selection models.

1 Introduction

Complex problem solving usually means dealing with unknown or non-existent computationally tractable solutions ([4] and [10]). In the past few years, several researchers used algorithms based on the model of organic evolution as an attempt to solve hard optimization and adaptation problems [6]. Due to their representation scheme for search points, Genetic Algorithms (GA) [5] are the most promising and easily applicable representatives of evolutionary algorithms for the problems discussed in this paper. In addition, a relatively recent search paradigm known as ant colony approach [15] is producing excellent results in different areas, thus we are interested in considering it in our study.

Our aim in this paper is to present algorithms representing different branches of the evolutionary computation domain to test their relative efficiency and efficacy on a scheduling class of problems: the minimum tardy task problem ("mttp"). The contribution of this work is two-fold. First, a performance comparison between well known families of evolutionary algorithms for solving a combinatorial optimization problem is made. These algorithms are: the generational genetic algorithm (genGA), the steadystate genetic algorithm (ssGA) [11], the cellular genetic algorithm (cGA) [9], and the ant colony algorithm for subset problems (ACAsp) [16, 17]. Second, the paper reports the improvement achieved on already known results for similar problem instances, and enlarges considerably the size of instances that can be solved to the optimum by consuming really modest computational resources.

The outline of the paper is as follows: Section 2 presents an overview of the working principles of genetic algorithms and ant colony algorithms. Section 3 presents the minimum tardy task problem. The encoding, the fitness function, and other specific problem-solving information is explained in this section. The experimental results for each problem instance are described in Section 4. The paper ends by summarizing our findings in Section 5.

2 The Evolutionary Algorithms

Genetic Algorithms (GAs) initially developed by Holland in the sixties, are guided random search algorithms based on the model of biological evolution (see e.g. [5,6]). Consequently, the field of Evolutionary Computation (EC), of which genetic algorithms is part, has borrowed much of its terminology from biology. These algorithms rely on the collective learning process within a population of individuals, each of which represents a search point in the space of potential solutions for a given optimization problem (objective function). The population evolves to-

wards increasingly better regions of the search space by means of randomized processes of selection, mutation, and recombination. The selection mechanism favors individuals of better objective function value to reproduce more often than worse ones when a new population is formed. Recombination allows for the mixing of parental information when this is passed to their descendants, and mutation introduces innovation into the population. Usually, the initial population is randomly initialized and the evolution process is stopped after a predefined number of iterations.

The ant colony optimization technique (Dorigo et al. [15]) as a new meta-heuristic for hard combinatorial optimization problems. Ant algorithms, that is, instances of the ant colony optimization metaheuristic, are basically a multi-agent system where low level interactions between single agents (called artificial ants) result in a complex behavior of the whole system. Ant algorithms have been inspired by colonies of real ants, which deposit a chemical substance (called *pheromone*) on the ground. This substance influences the choices they make: the larger the amount of pheromone on a particular path, the larger the probability that an ant selects the path. Artificial ants, in ant algorithms, are stochastic construction procedures that probabilistically build a solution by iteratively adding solutions components to partial solutions by taking into account (i) heuristic information on the problem instances being solved, if available, and (ii) pheromone trails which change dynamically at run-time to reflect the acquired search experience.

For this paper, we developed and implemented three genetic algorithms and an ant colony algorithm. The GAs under study include generational, steady-state, and cellular genetic algorithms. The first two are sub-classes of panmictic algorithms, the latter one is a sub-class of structured EAs. Panmictic algorithms consider all the population as a mating pool for selecting individuals for reproduction, while structured EAs define some kind of neighborhood for each individual, and restrict reproduction to mates selected from its neighborhood. The reader is referred to [2] for more details on panmictic and structured genetic algorithms. The ACAsp presented here is basically an Ant System which combines some features taken from other ant algorithms.

2.1 Genetic Algorithms

In this section we describe briefly our three basic GAs. Our genGA, like most GAs described in the literature, is generational, i.e., at each generation, the new population consists entirely of offspring formed

by parents in the previous generation (although some of these offspring may be identical to their parents).

In steady-state selection [11], only a few individuals are replaced in each generation. With ssGA, the least fit individual is replaced by the offspring resulting from crossover and mutation of the fittest individuals. The ssGA selects two parents, recombines their contents to create one single offspring, applies mutation, and inserts the result back into the population for the next iteration step. Often, ssGA converges to an optimum faster than genGA (although in some cases it converges to a local optimum).

Finally, the cGA population is structured in a toroidal 2D grid and the neighborhood defined on it always contains 5 strings: the one under consideration and its north, east, west, and south neighboring strings. The grid used in our tests is a 7×7 square. Different parameters for the grid are discussed in [3]. Fitness proportional selection is used in the neighborhood along with the one–point crossover operator. The latter yields only one child: the one having the larger portion of the best parent. Since a string belongs to several neighborhoods, any change in its contents affects its neighbors in a smooth manner, representing an appropriate tradeoff between a convergence and a wide exploration of the search space.

2.2 Ant Colony Algorithms

Early experiments with ant algorithms were connected with ordering problems such as the Traveling Salesperson Problem, the Quadratic Assignment Problem, as well as the Job Shop Scheduling, Vehicle Routing, Graph Coloring and Telecommunication Network Problem [15]. The Ant System (AS) was the first example of an ant colony optimization algorithm to be proposed in the literature. However, AS was not competitive with state-of-the-art algorithms for TSP, the problem to which the original AS was applied. There exist several improvements to the original version of AS, much of them applied to the TSP. The more important improvements are: AS with an elitist strategy for updating the pheromone trail levels, AS_{rank} (a rank-based version of Ant System), \mathcal{MAX} - \mathcal{MIN} Ant System (\mathcal{MMAS}), and the Ant Colony System (ACS). A detailed description of these versions can be found in [18].

More recently, promising results were reported in [16,17] from the application of a new version of an Ant System to the Multiple Knapsack and Maximum Independent Set Problems, two examples of subset problems with constraints, for which an ant colony algorithm was first applied. This paper aims to go further in this direction in order to evaluate the robustness and feasibility of applying an ant algorithm to a different subset problem, the minimum tardy task problem, according to the general concept behind an ant colony optimization meta-heuristic. Our ant colony algorithm for subset problems, called ACAsp, works as follows: In each cycle of the algorithm m ants are released to build a new solution from the empty set and all problem components being feasible. Each new component is selected step by step from the set of allowed componments and incorporated to the solution under construction. At the same time, the set of allowed components is updated in order to eliminate all components that became infeasible after a new component, says i is added to the solution. Before finishing the cycle, the pheromone trail levels are updated accordingly.

3 The Minimum Tardy Task Problem

The minimum tardy task problem is a NP-Complete task-scheduling problem [4]. The following is a formal definition of the minimum tardy task problem [10]:

Problem instance:

Feasible solution: A one-to-one scheduling function g defined on $S \subseteq T$, $g: S \longrightarrow Z^+ \cup \{0\}$ that satisfies the following conditions for all $i, j \in S$:

- 1. If g(i) < g(j) then $g(i) + l_i \le g(j)$ which insures that a task is not scheduled before the completion of an earlier scheduled one.
- 2. $g(i) + l_i \leq d_i$ which ensures that a task is completed within its deadline.

Objective function: The tardy task weight $W = \sum_{i \in T-S} w_i$, which is the sum of the weights of unscheduled tasks.

Optimal solution: The schedule S with the minimum tardy task weight W.

A subset S of T is feasible if and only if the tasks in S can be scheduled in increasing order by deadline without violating any deadline [10]. If the tasks are not in that order, one needs to perform a polynomially executable preprocessing step in which the tasks are ordered in increasing order of deadlines, and renamed such that $d_1 \leq d_2 \leq \cdots \leq d_n$.

The following example is from [8].

Example: Consider the following problem instance of the minimum tardy task problem:

3 Tasks: Lengths: 2 4 1 7 4 3 5 2 Deadlines: 20 3 6 8 10 15 165 15 20 16 19 10 25 17 18 Weights:

- a) $S = \{1, 3, 5, 6\}$ is a feasible solution and the schedule is given by: g(1) = 0, g(3) = 2, g(5) = 3 and g(6) = 7. The objective function value amounts to $\sum_{i \in T-S} w_i = w_2 + w_4 + w_7 + w_8$.
- b) $S' = \{2, 3, 4, 6, 8\}$ is infeasible. We define g(2) = 0, and task 2 finishes at time $0 + l_2 = 4$ which is within its deadline $d_2 = 5$. We schedule tasks 3 at 4, i.e. g(3) = 4, which finishes at $g(3) + l_3 = 5$ which is within its deadline $d_3 = 6$. But task 4 cannot be scheduled since $g(4) + l_4 = 5 + 7 = 12$ and will thus finish after its deadline $d_4 = 8$.

For our experiments, we use three problem instances: "mttp20" (20 tasks), "mttp100" (100 tasks) and "mttp200" (200 tasks). The first problem instance can be found in [8]. The second and third problem instances were generated as follows. First, the problem instance of very small size n5 is constructed.

Tasks: 1 2 3 4 5 Lengths: 3 6 9 12 15 Deadlines: 5 10 15 20 25 Weights: 60 40 7 3 50

This problem instance can be used to construct problem instances for any arbitrarily large number of tasks n where $n=5\times t$ ($t\geq 1$). We now describe how to construct a minimum tardy task problem instance of size $n=5\times t$ from the 5-task model. The first five tasks of the large problem are identical to the 5-model problem instance. The length l_j , deadline d_j , and weight w_j of the j^{th} task, for $j=1,2,\ldots,n$, is given by: $l_j=l_i,\ d_j=d_i+24\cdot m$ and

$$w_j = \left\{ \begin{array}{ll} w_i & \text{if } j \equiv 3 \, \mathrm{mod} \, 5 \text{ or } j \equiv 4 \, \mathrm{mod} \, 5 \\ (m+1) \cdot w_i & \text{otherwise} \end{array} \right.,$$

where $j \equiv i \mod 5$ for i=1,2,3,4,5 and $m=\lfloor (j-1)/5 \rfloor$. As can be seen, the 5-task model is repeated t times with no alterations in the task lengths. The weights and deadlines are scaled to force an optimum schedule represented by the bit string 11001 repeated t times. By leaving tasks j unscheduled, where $j\equiv 3 \mod 5$ or $j\equiv 4 \mod 5$, we get a fitness of $(7+3)\cdot t$. In other words, the tardy task weight for the globally optimal solution of this problem is $2\cdot n$.

The experiments on this scheduling problem are performed with different instances. The first problem instance is of moderate size, but nevertheless, is a challenging exercise for any heuristic. In the absence of test problems of significantly large sizes, we proceed by using the explained scalable instance generator for deriving instances of arbitrary sizes, and more importantly, whose optimal solution can be pre-computed. This allows us to compare our results to the optimum solution, as well as to the existing best solution (using genetic algorithms).

3.1 Solving "mttp" with Genetic Algorithms

A schedule S can be represented by a vector x = (x_1, x_2, \ldots, x_n) where $x_i \in \{0, 1\}$. The presence of task i in S means that $x_i = 1$, while its absence is represented by a value of zero in the i^{th} component of x. In order to apply a genetic algorithm, we use the fitness function described in [8] which allows infeasible strings and uses a graded penalty term. As expected, significant portions of the search space of some of the problem instances we tackle are infeasible regions. Rather than ignoring the infeasible regions, and concentrating only on feasible ones, we do allow infeasibly bred strings to join the population, but for a certain price. A penalty term incorporated in the fitness function is activated, thus reducing the infeasible string's strength relative to the other strings in the population. The fitness function uses $\sum_{i=1}^{n} w_i$ as offset term, to make sure that no infeasible string has a better fitness value than a feasible one.

In essence, the fitness function to be minimized for the tardy task problem is given by:

$$f_3(\mathbf{x}) = \sum_{i=1}^{n} w_i \cdot (1 - x_i) + (1 - s) \cdot \sum_{i=1}^{n} w_i + \sum_{i=1}^{n} w_i x_i \cdot \mathbf{1}_{R+} \left(l_i + \sum_{\substack{j=1 \ x_j \text{ schedulable}}}^{i-1} l_j x_j - d_i \right)$$
(1)

The third term keeps checking the string to see whether a task could have been scheduled or not, as explained earlier. It makes use of the indicator function:

$$\mathbf{1}_{A}(t) = \begin{cases} 1 \text{ if } t \in A \\ 0 \text{ otherwise} \end{cases}$$
 (2)

Also note that s=1 when \boldsymbol{x} is feasible, and s=0 when \boldsymbol{x} is infeasible.

3.2 Solving "mttp" with Ant Colony Algorithms

The ACAsp for "mttp" is basically an Ant System including the main features of a \mathcal{MMAS} and an elitist strategy. This algorithm is very similar to that

ant algorithms presented in [16,17] for the Maximum Independent Set and Multiple Knapsack problems, respectively. The main difference is in the local heuristic which is formulated here as $\eta_i = w_i/l_i$, i.e., the rate between the weight associated to task idivided by its respective duration l_i . Therefore, the bigger the value of η_i , the more preference is assigned to task i to be part of the solution under construction. It is also important to note that the heuristic (η) and trail (τ) values are both involved in the probability distribution used for selecting the tasks to be included in the solution in each cycle of the ant algorithm. Each time that a new component is added to the solution under construction, the ACAsp eliminates the set of tasks that are not longer schedulable. Thus, ACAsp only builds feasible solutions.

4 Experimental Runs

We now proceed to discuss the parameters of our experiments. We performed a total of 100 experimental runs for each of the problem instances and each of the algorithms.

For the GAs, whenever no parameter setting is explicitly stated, all experiments reported here are performed with a standard parameter setting: population size $\mu=50$, one-point crossover, crossover rate $p_c=0.6$, bit-flip mutation, mutation rate $p_m=1/n$ (where n is the string length), and proportional selection. These were the settings used with the same problem instances reported in [8]. For the ACAsp, all experiments reported here are performed with the following parameter setting: colony size 5, $\alpha=1$, $\beta=3$, $\rho=0.5$, and 50 cycles.

What follows is the convention used to present the results of the experimental runs. For each problem instance, we present one table with the results of our EAs. The first column for each evolutionary algorithm gives the best fitness value encountered during the 100 runs. The second column for each evolutionary algorithm records the number of times each one of these values is attained during the 100 runs. The values given in the first row of the table are the average number of evaluations it took to obtain the maximum value. The first value recorded under f(x) is the globally optimal solution. For example, Table 1 reports that genGA obtained the global optimal (whose value is 41) 73 times out of the 100 runs. The table also indicates that the optimum value was obtained after 2174.7 evaluations when averaged over the 100 runs.

Let us now proceed to analyze the results for the minimum tardy task problem starting with the ge-

Table 1. Overall best results of all experimental runs performed for "mttp20".

ssGA		genGA		cGA	
avg = 871.4		avg = 2174.7		avg = 7064.2	
f(x)	N	f(x)	N	f(x)	N
41	86	41	73	41	23
46	10	46	11	46	7
51	4	49	8	49	9
		51	3	51	9
		56	1	53	6
		57	1	54	1
		61	1	56	12
		65	2	≥ 57	33

Table 2. Overall best results of all experimental runs performed for "mttp100".

ssGA		genGA		cGA	
avg = 43442		avg = 45426		avg = 15390	
$f(oldsymbol{x})$	N	f(x)	N	f(x)	N
200	78	200	98	200	18
243	4	243	2	243	18
326	1			276	2
329	17			293	6
				316	1
				326	1
				329	37
				379	9
				≥ 429	8

Table 3. Overall best results of all experimental runs performed for "mttp200".

ssGA		genGA		cGA	
avg = 288261.7		avg = 83812.2		avg = 282507.3	
$f(oldsymbol{x})$	N	$f(oldsymbol{x})$	N	f(x)	N
400	18	400	82	400	6
443	8	443	9	443	7
476	2	476	2	493	1
493	1	493	2	529	34
516	1	496	1	543	1
529	42	529	3	579	10
579	3	629	1	602	1
602	1			629	8
665	23			665	17
715	1			≥ 679	15

netic algorithms. While ssGA outperforms the two other heuristics for the 20-task problem instance, genGA gives much better results for the 100-task and 200-task problems. For "mttp20", the local optimum of 46 differs from the global one by a Hamming distance of three. Compared to the results of [8] for "mttp20", ssGA performs much better, genGA is comparable, while cGA's performance is worse. For

"mttp200", genGA is a clear winner among the three genetic algorithms. This problem instance was not attempted by [8].

Similarly, Table 4 shows the results obtained from the application of the ACAsp to several instances of "mttp". The convention used in this table is as follows: the first column represents the size of the instance tested, the second column shows the objective value of the globally optimal solution, the third column represents the number of times each value is achieved out of 100 runs, and the last column shows the average number of evaluations to obtain the best solution. This number is actually an upper bound on the number of evaluations which is obtained as $m \times \overline{bc} \times n$, where m is the number of ants in the colony (population size), \overline{bc} is the average number of cycles used the get the best value, and n is the size of the problem. The number n is also included in the upper bound since each ant proceeds step by step in order to build a feasible solution. In each step of the construction process a new component is added, thus at most n components can be included in a solution in each cycle by each ant. Then, for each component added to the solution, an evaluation takes place in the ant algorithm. For example, the table shows for the instance of size 200 that the optimum value was obtained after 1000.00 evaluation when averaged over the 100 runs.

Clearly the ACAsp found the optimal value for all instances tested in each run. Is worth remarking that the ant algorithm was tested on larger instances of "mttp" when compared against the sizes of the instances tested by applying the GAs. Thus, instances of size 400, 800, 1200, 2000, 5000, and 10000 were solved optimally showing the scalability and good performance of the ant algorithm.

The ACAsp outperformed the GAs on the instances of size 20, 100, and 200. For the instances of larger sizes, ACAsp also showed a similar performance as for the smaller instances in regards of the quality of the results and the low number of evaluations needed to get the best values. For the GAs it is difficult and costly to solve the larger instances of the problem; we are performing more research on their scalability, and even plan their extension in the form of a parallel distributed fashion (multi-populations). However, it is very unlikely that their canonical behavior could compete with ACAsp in computational effort (number of evaluations); we should need to include some problem-specific operators to increase their efficiency on the "mttp" problem class.

ACAsp				
Size	f_{opt}	N	avg	
20	41	100	100.00	
100	200	100	510.00	
200	400	100	1000,00	
400	800	100	2100.00	
800	1600	100	4840.00	
1200	2400	100	8040.00	
2000	4000	100	16100.00	
5000	10000	100	132750.00	
10000	20000	100	364000.00	

5 Concluding Remarks and Future Work

This paper explores the applications of evolutionary algorithms for a combinatorial problem in the domain of task scheduling. Three genetic algorithms and an ant colony algorithm were faced to instances of increasing difficulty. Overall, our findings confirm the strong potentiality of genetic algorithms for finding globally optimal solutions with high probability in reasonable time, even in the case of hard multimodal optimization tasks when a number of independent runs is performed. However, it is mandatory noting that the ant colony algorithm offered very efficient results, clearly above the efficiency of the best of the genetic algorithms.

Because of the promising results of the ant colony algorithm in this kind of problems, we plan to study its fundamentals to find whether it is possible to include them inside other evolutionary algorithms. Besides, we will conduct further research on other similar problems to assess whether the efficient behavior of the ant colony meta-heuristic holds.

Acknowledgments

The first author acknowledges the funding of the Ministry of Science-Technology and FEDER under contract TIC2002-04498-C05-02 (the TRACER project), http://tracer.lcc.uma.es. The second and third authors acknowledge the funding of the Universidad Nacional de San Luis and the ANPCyT.

References

1. Alba, E., Khuri, S. (2001) Applying Evolutionary Algorithms to Combinatorial Optimization Problems.

- Lecture Notes in Computer Science, vol. 2074, Part II, Springer-Verlag, Berlin, Heidelberg, 689–700
- Alba, E., Troya, J. M. (1999) A Survey of Parallel Distributed Genetic Algorithms. Complexity, vol. 4, number 4, 31–52
- Alba, E., Troya, J. M. (2000) Cellular Evolutionary Algorithms: Evaluating the Influence of Ratio. Lecture Notes in Computer Science, vol. 1917, Springer-Verlag, Berlin, Heidelberg, 29–38
- Garey, M. R., Johnson, D. S. (1979) Computers and Intractability: A Guide to the Theory of NP-Completeness. W. H. Freeman and Co., San Francisco, CA
- Goldberg, D. E. (1989) Genetic Algorithms in Search, Optimization and Machine Learning. Addison Wesley, Reading, MA
- Holland, J. H. (1975) Adaptation in Natural and Artificial Systems. The University of Michigan Press, Ann Arbor, MI
- Brucker, P. (1998) Scheduling Algorithms. Springer-Verlag, 2nd edition
- Khuri, S., Bäck, T., Heitkötter, J. (1994) An Evolutionary Approach to Combinatorial Optimization Problems. Proceedings of the 22nd Annual ACM Computer Science Conference, ACM Press, NY, 66–73
- 9. Manderick, B., Spiessens, P. (1989) Fine-Grained Parallel Genetic Algorithms. *Proceedings of the 3rd ICGA*, Morgan Kaufmann, 428–433
- Stinson, D. R. (1987) An Introduction to the Design and Analysis of Algorithms. The Charles Babbage Research Center, Winnipeg, Manitoba, Canada, 2nd edition
- Syswerda, G. (1981) A Study of Reproduction in Generational and Steady-State Genetic Algorithms. Proceedings of FOGA, Morgan Kaufmann, 94–101
- Whitley, D., Rana, S., Dzubera, J., Mathias, K. E. (1996) Evaluating Evolutionary Algorithms. Artificial Intelligence, vol. 85, 245-276
- Wolpert, D. H., Macready, W. G. (1997) No Free Lunch Theorems for Optimization. *IEEE Transactions on Evolutionary Computation*, vol. 1, number 1, 67–82
- De Jong, K. A., Potter, M. A., Spears, W. M. (1997) Using Problem Generators to Explore the Effects of Epistasis. Proceedings of the 7th International Conference of Genetic Algorithms, Morgan Kaufman, 338-345
- Corne, D., Dorigo, M., Glover, F. (1999) New Ideas in Optimization. McGrawHill.
- Leguizamón, G., Michalewicz, Z. (1999) New Version of Ant System for Subset Problems. Proceedings of the 1999 Congress on Evolutionary Computation, IEEE Press, Piscataway, NJ., 1459–1464
- 17. Leguizamón, G., Michalewicz, Z., Schütz, M. (2001) An Ant System for the Maximum Independent Set Problem *Proceedings of the 2001 Argentinian* Congress on Computer Science. El Calafate, Argentina, 1027–1040.
- 18. Dorigo, M., Stützle, T. (2000) The Ant Colony Optimization Metaheuristic: Algorithms, Applications, and Advances Technical Report IRIDIA/2000-32. IRIDIA Universite Libre de Bruxells, Belgium