Solving the Probabilistic Travelling Salesman Problem Based on Genetic Algorithm with Queen Selection Scheme

Yu-Hsin Liu
Department of Civil Engineering, National Chi Nan University
Taiwan

1. Introduction

The probabilistic travelling salesman problem (PTSP) is an extension of the well-known travelling salesman problem (TSP), which has been extensively studied in the field of combinatorial optimization. The goal of the TSP is to find the minimum length of a tour to all customers, given the distances between all pairs of customers whereas the objective of the PTSP is to minimize the expected length of the *a priori* tour where each customer requires a visit only with a given probability (Bertsimas, 1988; Bertsimas et al., 1990; Jaillet, 1985). The main difference between the PTSP and the TSP is that in the PTSP the probability of each node being visited is between 0.0 and 1.0 while in TSP the probability of each node being visited is 1.0. Due to the fact that the element of uncertainty not only exists, but also significantly affects the system performance in many real-world transportation and logistics applications, the results from the PTSP can provide insights into research in other probabilistic combinatorial optimization problems. Moreover, the PTSP can also be used to model many real-world applications in logistical and transportation planning, such as daily pickup-delivery services with stochastic demand, job sequencing involving changeover cost, design of retrieval sequences in a warehouse or in a cargo terminal operations, meals on wheels in senior citizen services, trip-chaining activities, vehicle routing problem with stochastic demand, and home delivery service under e-commerce (Bartholdi et al., 1983; Bertsimas et al., 1995; Campbell, 2006; Jaillet, 1988; Tang & Miller-Hooks, 2004).

Early PTSP computational studies, dating from 1985, adopted heuristic approaches that were modified from the TSP (e.g., nearest neighbor, savings approach, spacefilling curve, radial sorting, 1-shift, and 2-opt exchanges) (Bartholdi & Platzman, 1988; Bertsimas, 1988; Bertsimas & Howell, 1993; Jaillet, 1985, 1987; Rossi & Gavioli, 1987). With its less than satisfactory performance in yielding solution quality, researchers in the recent years switch to metaheuristic methods, such as ant colony optimization (Bianchi, 2006; Branke & Guntsch, 2004), evolutionary algorithm (Liu et al., 2007), simulated annealing (Bowler et al., 2003), threshold accepting (Tang & Miller-Hooks, 2004) and scatter search (Liu, 2006, 2007, 2008). Because the genetic algorithm (GA), a conceptual framework of the population-based metaheuristic method, has been shown to yield promising outcomes for solving various complicated optimization problems in the past three decades (Bäck et al., 1997; Davis, 1991;

Source: Travelling Salesman Problem, Book edited by: Federico Greco, ISBN 978-953-7619-10-7, pp. 202, September 2008, I-Tech, Vienna, Austria

Goldberg, 1989; Holland, 1992; Liu & Mahmassani, 2000), this study will propose an optimization procedure based on GA framework for solving the PTSP.

Mainly, the author of this chapter proposes and tests a new search procedure for solving the PTSP by incorporating the nearest neighbor algorithm, 1-shift and/or 2-opt exchanges for local search, selection scheme, and edge recombination crossover (ERX) operator into genetic algorithm (GA) framework. Specifically, the queen GA, a selection approach which was proposed recently and yielded promising results (Balakrishnan et al., 2006; Stern et al., 2006), will be tested against the traditional selection mechanisms (i.e., fitness-proportional, tournament, rank-based and elitist selections) for its comparative effectiveness and efficiency in solving the PTSP. Unlike traditional selection mechanisms used in GA which selects both parents from the entire population based on their fitness values, the queen GA creates a subgroup of better solutions (the queen cohort), and uses at least one of its members in each performed crossover. To validate the effectiveness and efficiency of the proposed algorithmic procedure, a set of heterogeneous (90 instances) and homogeneous (270 instances) PTSP test instances as used in the previous studies (Liu, 2006, 2007, 2008; Tang & Miller-Hooks, 2004) will be used as the base for comparison purpose.

The remainder of this chapter is organized as follows. In the next section, expressions for exactly and approximately evaluating the *a priori* tour for the PTSP are introduced. The details of the proposed algorithmic procedure for the PTSP are then described. The results of the numerical experiments are presented and discussed in the next section, followed by concluding comments.

2. Definition and evaluation of the PTSP

The PTSP is defined on a directed graph $G := (V, E)$, where $V := \{0, v_1, v_2, ..., v_n\}$ is the set of nodes or vertices, $E \subseteq V \times V$ is the set of directed edges. Node 0 represents the depot with the presence probability of 1.0. Each non-depot node v_i is associated with a presence probability p_i that represents the possibility that node v_i will be present in a given realization. Given a directed graph G, the PTSP is to find an *a priori* Hamiltonian tour with minimal expected length in G.

2.1 Exact evaluation for the *a priori* tour

Solving the PTSP mainly relies on computing the expected length of an *a priori* tour. The computation of the expected length of a specific *a priori* PTSP tour τ, denoted as $E[\tau]$, depends on the relative location of nodes on that tour and the presence probability of each node in a given instance. By explicitly considering all realizations based on the presence of each individual node, the expected length of tour τ can be calculated. For an n-node PTSP instance, a tour τ has 2^n possible realizations. The probability of realization r_j, $p(r_j)$, can be calculated based on the presence probability of each individual node. Let $L[r_j(\tau)]$ describe the tour length of τ for realization r_j under the assumption that nodes not in r_j are simply skipped in the tour. The expected tour length can then be formally described as

$$E[\tau] = \sum_{j=1}^{2^n} p(r_j) L[r_j(\tau)] \tag{1}$$

The computation of expected length based on Equation (1) is inefficient, because the computational complexity increases exponentially with an increasing number of nodes.

Therefore, Jaillet & Odoni (1988) proposed an approach to exactly calculate $E[\tau]$ in the complexity of $O(n^3)$ for the PTSP.

$$E[\tau] = \sum_{i=0}^{n} \sum_{j=i+1}^{n+1} \{d_{\tau(i)\tau(j)} p_{\tau(i)} p_{\tau(j)} \prod_{k=i+1}^{j-1} (1 - p_{\tau(k)})\} \tag{2}$$

d_{ij} represents the distance between nodes i and j; $\tau(i)$ denotes the node that has been assigned the i^{th} stop in tour τ and $p_{\tau(i)}$ is the probability of node $\tau(i)$. $\tau(0)$ and $\tau(n+1)$ represent node 0, which is the depot.

2.2 Approximate evaluation for the *a priori* tour

Even though (2) yields a polynomial evaluation time for the PTSP, the resulting $O(n^3)$ time for calculating $E[\tau]$ is still very long, especially for metaheuristic methods which need to repeatedly evaluate the objective function value $E[\tau]$. In this study, the proposed GA needs to repeatedly compare two solutions (i.e., the new solution before and after local search procedure, which is described in the next section) based on their values of $E[\tau]$. Therefore, the depth approximation originally proposed by Branke & Guntsch (2004) was adopted. The depth approximate evaluation of $E[\tau]$ shown in (3) have been used to significantly increase the computation efficiency under the scatter search framework (Liu, 2006).

$$E_{\lambda}^{AP}[\tau] = \sum_{i=1}^{n} \sum_{j=i+1}^{\min\{n+1, i+\lambda\}} \{d_{\tau(i)\tau(j)} p_{\tau(i)} p_{\tau(j)} \prod_{k=i+1}^{j-1} (1 - p_{\tau(k)})\} \tag{3}$$

The only difference between (2) and (3) is the choice of truncation position λ in (3). Equation (3) will have the computational complexity of $O(n\lambda^2)$, instead of $O(n^3)$ in (2). It is easy to see that (3) becomes more accurate when λ increases. A larger value of λ, however, requires more computation efforts for the computation of (3). Equation (3) can perform a very good approximation of $E[\tau]$ with a smaller value of λ when the value of $p_{\tau(k)}$ gets larger, because $\prod_{k=i+1}^{j-1}(1 - p_{\tau(k)})$ will yield a very small value and can be omitted. Nevertheless, Equation (3) will need a larger value of λ to perform a good approximation when the value of $p_{\tau(k)}$ is small. The approximation usually yields some errors in comparison to the exact evaluation. To overcome that, the two-stage comparison proposed by Liu (2008) intends to exactly evaluate the $E[\tau]$ value by using the depth approximation evaluation (Equation 3) in the first stage and the exact evaluation (Equation 2) in the second stage. The detailed use of the depth approximation evaluation shown in Equation (3) to accelerate the proposed algorithm is referred to Liu (2008).

3. Solution algorithm

The proposed GA consists of four components as shown in Fig. 1. They are the initialization, local search, selection scheme, and crossover. When starting to solve the PTSP (Generation 0, g = 0), initial solutions are generated based on the nearest neighbor algorithm, which are then improved by the local search. Then, a specific selection mechanism is called into place to further select solutions to be mated based on their solution quality (objective function value). Pairs of solutions are used to generate the new solutions via edge recombination crossover (ERX). The newly generated solutions are then improved using the local search. The solutions are allowed to evolve through successive generations until a termination criterion is met. The detailed description of the embedded components is illustrated in the following sections.

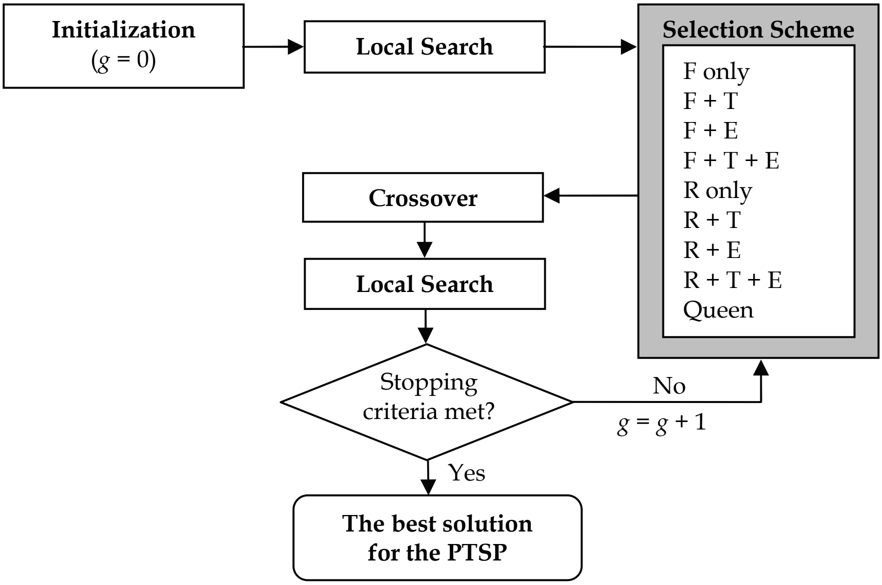
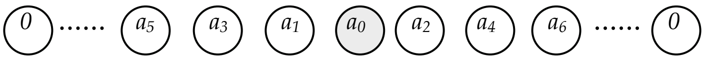


Fig. 1. The general procedure of the genetic algorithm for the PTSP.

3.1 Initialization

This procedure is designed to generate m initial solutions (m = 15 in this study). Considering a PTSP with n nodes (excluding the depot, node *0*), the farthest node, a_0, from node *0* is selected first and randomly inserted into a location between $(\lfloor (n+1)/2 \rfloor - 4)$ and $(\lfloor (n+1)/2 \rfloor + 4)$. The nearest neighbor algorithm is then used to build up the sequence of the tour. After selecting node a_0, the nearest node (a_1) from a_0 is selected and inserted in front of a_0. The second nearest node (a_2) from a_0 is selected and inserted behind a_0. Then, among the remaining nodes, the nearest node (a_3) from a_1 is selected and inserted in front of a_1, while the nearest node (a_4) from a_2 is selected and inserted behind a_2. The 1st initial solution (tour) is thus built by following the above rule and expressed as follows.



To create diverse solutions, the remaining initial solutions are generated using the above rule with slight modifications. The only difference lies in whenever l = 6, 12, 18, ..., instead of using the nearest node from a_{l-2}, a_l is randomly chosen from the first or second nearest node from a_{l-2}.

3.2 Local search

This component is used in an attempt to further enhance the solution generated via a local search procedure. As the previous study has investigated the performance of diversified local search strategy by stochastically selecting two different local search methods (i.e., 1-shift and 2-opt exchanges) and found that combining 1-shift and 2-opt (1-shift/2opt) is the most effective local search for the PTSP (Liu, 2008). Therefore, the 1-shift/2-opt is then adopted to improve the solution generated in the proposed GA algorithm.

The procedures of 1-shift and 2-opt exchanges are briefly summarized as follows. Given an *a priori* tour τ, its 1-shift neighborhood is the set of tours obtained by moving a node at position i to position j with the intervening nodes being accordingly shifted backwards one space. The 2-opt exchange is the set of tours obtained by reversing a section of τ.

The depth approximate evaluation of expected length of the *a priori* tour shown in (3) is then used to increase the computational efficiency. For a specific tour τ, $E_{\lambda}^{AP}[\tau]$ is always less than the value of $E[\tau]$ because of the truncation in calculating $E_{\lambda}^{AP}[\tau]$. Let τ_b and τ_a denote the *a priori* tour before and after a specific local search method, respectively. It means that no improvement has been found after the local search if $E_{\lambda}^{AP}[\tau_a] \geqq E[\tau_b]$. Equation (2) is used to exactly evaluate the solution after the local search if $E_{\lambda}^{AP}[\tau_a] < E[\tau_b]$. If the local search yields a better $E[\tau]$ value than the one from the original solution (i.e., $E[\tau_a] < E[\tau_b]$), the new solution (τ_a) will replace the original solution (τ_b). If no improvement has been found after the local search, no replacement will be made. The above procedure is repeated N_{LS} times for each solution ($N_{LS} = 25$ in this study).

3.3 Selection scheme

Selection scheme is the process of choosing the mating pairs from the current population and to create the new solutions based on crossover operator. To investigate the performance of the queen GA, four popularly used selection mechanisms are used as a benchmark in this study: fitness-proportional, rank-based, tournament, and elitism selections.

3.3.1 Fitness-proportional selection (F)

Under the fitness-proportional selection method, the probability of selecting a particular solution for reproduction is proportional to its own fitness (i.e., $E[\tau]$) relative to the average fitness of the entire current generation. With this selection method, the best solution tends to produce the largest amount of offspring and hence survive to future generations. This procedure can be regarded as a “biased” roulette wheel where each string in the current population occupies a roulette wheel slot sized in proportion to its fitness (Goldberg, 1989). Selection can be done by simply spinning the weighted roulette wheel, and fitter strings will have higher chances of being selected. This process can be simulated by the following expression:

$$q_k = \frac{1/f_k}{\sum_{t=1}^{m} 1/f_t} \tag{4}$$

where q_k is the probability of selecting solution k to produce offspring, and m is the population size. The f_k is the fitness value of the k^{th} solution in the current generation.

Because the PTSP is a minimization problem, $1/f_k$ is used as the appropriate weight for the k^{th} solution.

3.3.2 Rank-based selection (R)

Under the rank-based selection, the probability of selecting a particular solution for reproduction is determined by the rank of its fitness. This process can be simulated by the following expression:

$$q_k = \frac{1/r_k}{\sum_{t=1}^{m} 1/r_t} \tag{5}$$

where r_k is the rank of the fitness value for the k^{th} solution.

3.3.3 Tournament selection (T)

Tournament selection, inspired by the competition in nature among individuals for the right to mate, picks two solutions using the proportional or rank-based selection from the population and the fittest one is selected for reproduction (Goldberg, 1989; Davis, 1991). Each solution can participate in an unlimited number of tournaments. The two winning solutions in the tournament are then subjected to the crossover operators.

3.3.4 Elitism (E)

Under the elitism selection strategy, the top N_e strings (N_e is determined by the analyst) of the current generation in terms of fitness value are kept and propagated to the next generation (Davis, 1991). The remaining solutions in the next generation are then generated based on the tournament selection method and the crossover operators. This procedure guarantees that the best solution in the next generation is not worse than the one in the current generation.

3.3.5 Queen GA

According to the concept of queen GA, the top N_{top} solutions in terms of its fitness value of the population are selected to be the members of queen. Then, one of the parents is chosen from the queen members and the other parent is randomly selected from the whole population excluding the already chosen member. These two selected parents are then mated based on the crossover operator. The queen members are dynamically updated based on the quality of the new solutions generated. A newly solution generated will become a queen member if the new solution has a better objective function value than the one with the worst objective value in the queen subset.

3.3.6 Experiment design of selection schemes

In addition to queen GA, eight schemes are designed by combining one or several selection methods from four popularly used selection mechanisms mentioned previously (i.e., fitness-proportional, rank-based, tournament, and elitism selection). Explicitly, since the tournament and elitism selections need to work with fitness-proportional (F) or rank-based (R) selection, eight selection schemes are designed and used in the numerical experiment in this study. They are fitness-proportional selection only (F), fitness-proportional and tournament selection (F+T), fitness-proportional and elitism selection (F+E), fitness-proportional, tournament and elitism selection (F+T+E), rank-based selection only (R), rank-based and tournament selection (R+T), rank-based and elitism selection (R+E), rank-based, tournament and elitism selection (R+T+E).

3.4 Edge recombination crossover (ERX)

The main purpose of this component is to create new solutions using a given pair of solutions generated by "selection". Based on the results from previous studies (Liu et al., 2007; Potvin, 1996), the edge recombination crossover (ERX) from genetic algorithms performed best when compared to other crossover strategies for both in TSP and PTSP. Therefore, ERX was adopted in this study.

ERX was proposed by Whitley et al. (1989) to solve the traditional TSP. A 5-node PTSP is used as an example to describe the procedure of ERX. Assuming that two solutions (tours) are chosen from the "selection"--(0, 4, 3, 1, 2, 0) and (0, 1, 2, 3, 4, 0), the edges connected to each node are as follows. For node 0, the first solution indicates that node 0 connects to nodes 2 and 4 and the second solution shows that node 0 connects to nodes 1 and 4. Therefore, node 0 connects to nodes 1, 2, and 4 by considering these two solutions. Similarly, node 1 connects to nodes 0, 2, 3; node 2 connects to nodes 0, 1, 3; node 3 connects to nodes 1, 2, 4; node 4 connects to nodes 0, 3. These are the initial edge lists for each node.

The operation of the ERX is described as follows. Assuming that node 0 is selected as the starting node for the new solution, all edges incident to node 0 must be deleted from the initial edge list. As described, from node 0 we can go to nodes 1, 2, or 4, while nodes 1 and 2 have two active edges and node 4 has only one active edge by deleting node 0 from the initial edge list. The node with the fewest active edge, node 4, is picked as the node next to node 0 in the new solution. Then, the edge list for the remaining nodes (nodes 1, 2, and 3) is further updated by deleting node 4. The updated edge list is node 1 (2, 3), node 2 (1, 3), and node 3 (1, 2). From node 4, we can only go to node 3 (as node 0 is already deleted from the list). Therefore, node 3 is chosen to be the node next to node 4 in the new solution. The new solution generated is further improved by the local search.

3.5 The procedure after the first generation

The newly generated solutions from the ERX and local search are used to update the population in terms of the objective function value. The above procedure is repeated until a termination criterion is met. However, if there are no solutions to be updated for the population in the current generation, the initialization is used to generate (m - m_1) new solutions in the next generation, but keeping m_1 high quality solutions ($m_1 = 2$, in this study). In addition, if the previous three generations converge to the same best solution, the local search is used to improve that "converged" solution by repeating N_{LS2} times to exhaustively search the neighborhood of that "converged" solution ($N_{LS2} = 300$, in this study).

4. Numerical experiments and results

There are two types of data sets, heterogeneous and homogeneous PTSP, used as numerical experiments in this study to examine the performance of different selection schemes under GA framework for the PTSP. First, 90 heterogeneous PTSP instances were generated by Tang & Miller-Hooks (2004) with size n = 50, 75, and 100. Three groups of problem sets categorized by different intervals of customer presence probabilities were created for each problem size (n = 50, 75, and 100). Presence probabilities of customer nodes were randomly generated from a uniform distribution on intervals (0.0, 0.2], (0.0, 0.5], (0.0, 1.0], one for each problem size. Second, there were 270 homogeneous PTSP instances generated by the author and used in the previous study of Liu (2008) with size n = 50, 75, and 100 associated with nine probability values (p = 0.1, 0.2, …, 0.9). For both homogeneous and heterogeneous PTSP, the presence probability of the depot (node 0) was assigned as 1.0. Ten different problem instances were randomly generated for each presence probability of customer nodes. For each instance, the coordinates of one depot and n customer nodes (x_i, y_i) were generated based on a uniform distribution from $[0, 100]^2$. The Euclidean distance for each pair of nodes was calculated by using $d_{ij} = \sqrt{(x_i - x_j)^2 + (y_i - y_j)^2}$.

To compare the effectiveness among nine different selection schemes under GA framework, the preset maximum number of generations (G_{max}) was used as the termination criterion (G_{max} is set to be two times the number of nodes, i.e., $G_{max} = 2n$, in this study) for both heterogeneous and homogeneous PTSP. The average solution quality is examined and compared among nine different selection schemes. In this study, the proposed methods were used to solve each problem instance 30 times to enhance the robustness of the results. That is, the average statistics for the methods proposed in this study are based on a 300-run average. The numerical results of heterogeneous and homogeneous PTSP are discussed in Section 4.1 and 4.2, respectively.

4.1 Results of heterogeneous PTSP

4.1.1 Descriptive statistics of average $E[\tau]$ values obtained by the heterogeneous PTSP

Average $E[\tau]$ values found from nine different selection schemes for the heterogeneous PTSP are reported in Table 1. Definitions of terms used in the column headings are given as follows. n denotes problem size, which is the number of customer nodes. p represents the customer presence probability interval (0.0, p].

The best average value of $E[\tau]$ among the nine selection schemes (i.e., F, F+T, F+E, F+T+E, R, R+T, R+E, R+T+E, Queen) for each problem size with different presence probability interval is shown in shaded. As shown in Table 1, the average $E[\tau]$ values obtained by only using fitness-proportional (F) or rank-based (R) selection strategy are consistently worse than the ones obtained by the other seven selection strategies. The solution quality becomes much better when adding tournament (T) and/or elitism strategies to fitness-proportional (F) or rank-based (R) selection. It indicates that fitness-proportional (F) or rank-based (R) selection should combine tournament (T) and/or elitism strategies to obtain acceptable outcomes.

Moreover, except for p = 0.5 when n = 50, the average $E[\tau]$ values obtained by adding elitism to fitness-proportional (F) selection strategy (F+E) performs better than the ones obtained by adding tournament to fitness-proportional (F) selection strategy (F+T). Furthermore, except for p = 0.5, 1.0 when n = 50, the average $E[\tau]$ values obtained by adding elitism to rank-based (R) selection strategy (R+E) performs better than the ones obtained by adding tournament to rank-based (R) selection strategy (R+T). It reveals that the average $E[\tau]$ values obtained by keeping the best solution(s) to the successive generations can generally perform better than the ones obtained by only applying tournament selection to fitness-proportional (F) or rank-based (R) selection.

Finally, as shown in Table 1, the average $E[\tau]$ values obtained by adding elitism to fitness-proportional (F) or rank-based (R) selection strategy are similar to the ones obtained by combining both elitism and tournament to fitness-proportional (F) or rank-based (R) selection strategy. Overall, the queen, F+E, F+T+E, R+E, and R+T+E are better selection strategies and yielded similar average $E[\tau]$ value for the heterogeneous PTSP than the other four selection strategies.

n	p	F	F+T	F+E	F+T+E	R	R+T	R+E	R+T+E	Queen
50	0.2	225.110	224.854	224.839	224.832	224.868	224.838	224.835	224.834	224.831
	0.5	343.901	341.585	341.675	341.426	341.935	341.347	341.504	341.331	341.499
	1.0	459.504	450.583	450.235	450.964	452.853	449.539	450.916	451.383	451.272
75	0.2	267.731	266.071	265.943	265.958	266.239	265.970	265.929	265.959	265.958
	0.5	415.129	404.257	403.526	403.879	406.728	403.782	403.485	403.748	403.705
	1.0	555.256	534.013	527.832	527.421	540.306	529.276	527.300	527.295	526.765
100	0.2	304.779	301.318	300.859	300.873	301.791	301.084	300.830	300.825	300.837
	0.5	480.752	466.813	463.747	462.578	469.663	464.671	462.661	463.381	461.556
	1.0	684.758	649.544	626.749	625.105	660.210	641.668	625.056	624.490	624.144

Table 1. Computational Results for the Heterogeneous PTSP

4.1.2 Inferential statistics analysis of nine selection schemes for heterogeneous PTSP

Since the assumption of normal distribution is hardly met in minimization problems, the permutation test (Basso et al., 2007), instead of parametric tests, is adopted for statistical testing in the study. A Monte Carlo method with 10,000 permutations is used to obtain the approximate p-value of the permutation test. A set of two-sample permutation tests is conducted to investigate if any statistically significant differences exist between the best average $E[\tau]$ value obtained and the ones obtained by the other eight selection schemes. Table 2 shows the p-values of the permutation tests, where $\alpha = 0.05$ is considered statistically significant in this study.

Several important findings are obtained. First, according to the results of the permutation tests, the average $E[\tau]$ values obtained by fitness-proportional (F) or rank-based (R) selection strategy are significantly higher than the best ones obtained by the other seven selection schemes for all of the tested cases. Second, the average $E[\tau]$ values obtained by Queen GA performs best in four out of the nine tested cases, and where they are not the best performing scheme, the average $E[\tau]$ values are not statistically significant different to the best ones obtained by the other eight selection schemes, except for n = 50 and p = 1.0. Third, for most of the test cases (21 out of 27 cases), the average $E[\tau]$ values obtained by F+T+E, R+E and R+T+E are not statistically significant different to the best ones obtained by these nine selection schemes. Finally, generally speaking, the average $E[\tau]$ values obtained by F+T, F+E and R+T performs statistically worse than the best ones obtained by the nine selection schemes for most of the test cases (20 out of 27 cases), except for n = 50 and p = 1.0, where the average $E[\tau]$ value obtained by R+T performs statistically better than the other eight selection schemes.

n	p	F	F+T	F+E	F+T+E	R	R+T	R+E	R+T+E	Queen
50	0.2	0.0000	0.0000	0.0040	1.0000	0.0000	0.0000	0.1044	0.7157	—
	0.5	0.0000	0.0056	0.0016	0.2814	0.0000	0.8413	0.0742	—	0.0574
	1.0	0.0000	0.0001	0.0301	0.0003	0.0000	—	0.0037	0.0000	0.0009
75	0.2	0.0000	0.0000	0.2865	0.1025	0.0000	0.0000	—	0.1026	0.1526
	0.5	0.0000	0.0762	0.9371	0.4485	0.0000	0.4828	—	0.6295	0.6664
	1.0	0.0000	0.0000	0.2261	0.3782	0.0000	0.0003	0.4642	0.4745	—
100	0.2	0.0000	0.0000	0.0046	0.0896	0.0000	0.0000	0.6137	—	0.3041
	0.5	0.0000	0.0000	0.0000	0.0376	0.0000	0.0000	0.0259	0.0052	—
	1.0	0.0000	0.0000	0.0036	0.1991	0.0000	0.0000	0.2004	0.6788	—

Table 2. *p*-value of Permutation test for the Heterogeneous PTSP

4.1.3 Comparison among the best performing scheme obtained in the study, the Queen GA and previous studies

As indicated in the previous section, in eight out of the nine tested cases (except for n = 50 and p = 1.0), the Queen GA either performs best or its performance not statistically significant different from the best ones obtained by the other eight selection schemes. The Queen as well as the the best performing scheme obtained in the study are compared against the previous studies in this section. The heterogeneous PTSP data generated by Tang & Miller-Hooks (2004) has been investigated in several studies (Tang & Miller-Hooks, 2004; Liu, 2006, 2007, 2008). The best average $E[\tau]$ values as well as the corresponding average CPU time in these studies (Previous Best) are listed in Table 3. In Table 3, the definitions of n and p are the same as in Table 1. $E[\tau]$ denotes the average value of the expected length of the *a priori* PTSP tour. CPU is the average CPU running time in seconds. The "Previous Best" results for the heterogeneous PTSP data were obtained by Liu (2006, 2007, 2008), except for n = 50 and p = 0.5, which were obtained by Tang & Miller-Hooks (2004). In Liu's studies (as well as the results of this study), all implementations were performed on an Intel Pentium IV 2.8 GHz CPU personal computer with 512 MB memory (3479 MFlops), while TMH's study was based on a 10-run average and was conducted on a DEC AlphaServer 1200/533 computer with 1 GB memory (1277 MFlops). The best average value of $E[\tau]$ among the three compared sets for each problem size with different presence probability interval is shown in shaded.

n	p	Best in this study		Queen		Previous Best	
		E[τ]	CPU (s)	E[τ]	CPU (s)	E[τ]	CPU (s)
50	0.2	224.8313	28.7	224.8313	28.7	224.8314	45.4
	0.5	341.3313	16.8	341.4989	16.2	341.3000*	72.4*
	1.0	449.5391	6.5	451.2717	8.4	450.2215	12.4
75	0.2	265.9293	108.9	265.9581	118.5	265.9315	240.6
	0.5	403.4846	46.3	403.7050	50.1	403.2347	51.8
	1.0	526.7646	28.6	526.7646	28.6	527.1907	41.5
100	0.2	300.8245	288.1	300.8370	269.5	300.8495	689.9
	0.5	461.5559	115.6	461.5559	115.6	462.2678	121.2
	1.0	624.1439	68.8	624.1439	68.8	624.6369	96.7

*Running on DEC AlphaServer 1200/533 computer with 1 GB memory (1277 MFlops)

Table 3. Computational Results for the Heterogeneous PTSP

The results in Table 3 show that the best of the average $E[\tau]$ values obtained in this study are better than the ones obtained by the "Previous Best." The only exception is when p = 0.5 and n = 75. The best average $E[\tau]$ value yielded performs 0.06% worse than the one obtained by the previous study (Liu, 2008), when p = 0.5 and n = 75. Moreover, the computation efforts used to yield the best results in this study are all less than the one used in "Previous Best." It suggests that the GA solution framework proposed in this study is a promising method for solving the heterogeneous PTSP. As for the Queen GA, the results show that it performs better than the "Previous Best" in terms of average $E[\tau]$ value and computational effort when n = 100. It suggests that the Queen GA is capable of effectively and efficiently solving relatively large-sized heterogeneous PTSP.

4.2 Results of homogeneous PTSP

4.2.1 Descriptive statistics of average *E* [τ] values obtained by the homogeneous PTSP

Average $E[\tau]$ values found from nine different selection schemes for the homogeneous PTSP are reported in Table 4. In Table 4, the definitions of n and p are the same as in Table1. The best average value of $E[\tau]$ among the nine selection schemes (i.e., F, F+T, F+E, F+T+E, R, R+T, R+E, R+T+E, Queen) for each problem size with different presence probability is shown in shaded. As the similar results obtained in the heterogeneous PTSP, the average $E[\tau]$ values obtained by only using fitness-proportional (F) or rank-based (R) selection strategy are consistently worse than the ones obtained by the other seven selection strategies. The solution quality becomes much better when adding tournament (T) and/or elitism (E) strategies to fitness-proportional (F) or rank-based (R) selection. Moreover, except for p = 0.3 when n = 50, the average $E[\tau]$ values obtained by adding elitism to fitness-proportional (F) selection strategy (i.e., F+E) performs better than the ones obtained by adding tournament to fitness-proportional (F) selection strategy (i.e., F+T). Furthermore, except for p = 0.3, 0.4 when n = 50, the average $E[\tau]$ values obtained by adding elitism to rank-based (R) selection strategy (i.e., R+E) performs better than the ones obtained by adding tournament to rank-based (R) selection strategy (i.e., R+T). Finally, the average $E[\tau]$ values obtained by adding elitism to rank-based (R) selection strategy are similar to the ones obtained by combining both elitism and tournament to rank-based (R) selection strategy. Overall the queen, F+T+E, R+E, and R+T+E are better selection strategies and yielded similar average $E[\tau]$ value for the homogeneous PTSP than the other five selection strategies.

4.2.2 Inferential statistics analysis of nine selection schemes for homogeneous PTSP

A set of two-sample permutation tests is conducted to investigate if any statistically significant differences exist between the best average $E[\tau]$ value obtained and the ones obtained by the other eight selection schemes. Table 5 shows the p-values of the permutation tests, where α = 0.05 is considered statistically significant in this study.

Several important findings are obtained. First, according to the results of the permutation tests, the average $E[\tau]$ values obtained by F only, R only and F+T are significantly higher than the best ones obtained by the other six selection schemes for all of the tested cases. Second, the average $E[\tau]$ values obtained by Queen GA performs best in 8 out of 27 tested cases, and where they are not the best performing scheme, the average $E[\tau]$ values are not statistically significant different to the best ones obtained by the other eight selection schemes, except for n = 75 and p = 0.6. Third, for most of the test cases (70 out of 81 cases), the average $E[\tau]$ values obtained by F+T+E, R+E and R+T+E are not statistically significant different to the best ones obtained by these nine selection schemes. Finally, the average $E[\tau]$ values obtained by F+E and R+T performs statistically worse than the best ones obtained by the nine selection schemes for most of the test cases (40 out of 54 cases).

n	p	F	F+T	F+E	F+T+E	R	R+T	R+E	R+T+E	Queen
50	0.1	233.907	233.550	233.497	233.493	233.584	233.513	233.492	233.492	233.492
	0.2	312.887	311.251	311.079	311.033	311.488	311.034	310.998	311.006	310.995
	0.3	371.020	366.525	366.788	366.170	367.575	366.097	366.424	366.632	366.492
	0.4	413.906	406.654	405.985	405.792	408.614	405.010	405.656	405.699	405.466
	0.5	467.415	456.167	453.551	453.791	459.147	454.205	453.581	453.486	453.204
	0.6	515.228	498.553	494.441	493.196	503.028	496.461	492.888	492.565	492.738
	0.7	537.288	519.762	510.409	509.883	525.096	516.295	509.516	509.762	509.492
	0.8	580.616	562.011	551.838	552.437	568.825	557.246	550.880	551.649	551.506
	0.9	586.400	565.562	562.089	561.712	572.469	561.706	560.520	561.090	561.496
75	0.1	277.591	276.112	275.827	275.822	276.302	275.976	275.824	275.819	275.820
	0.2	369.227	363.290	362.206	361.628	364.299	362.419	361.878	361.895	361.623
	0.3	460.647	448.300	444.228	444.268	451.166	446.191	444.101	444.083	444.365
	0.4	514.566	500.111	493.371	493.100	503.418	497.185	493.801	493.083	492.856
	0.5	563.640	537.817	526.367	525.293	546.373	532.653	525.790	525.704	525.308
	0.6	623.310	597.093	578.021	577.570	602.857	589.736	577.194	574.769	576.791
	0.7	666.105	638.798	621.849	620.450	648.911	632.238	619.659	618.957	619.248
	0.8	712.283	688.327	659.604	658.339	693.720	677.008	658.942	656.115	656.658
	0.9	757.030	722.544	690.629	690.952	733.558	711.425	690.537	690.196	690.150
100	0.1	310.330	306.549	305.727	305.682	307.103	306.172	305.685	305.676	305.682
	0.2	435.561	422.562	418.959	418.552	424.865	420.063	418.046	418.428	418.515
	0.3	526.932	507.731	497.024	496.876	512.953	502.780	496.402	497.076	497.298
	0.4	619.191	593.193	575.482	574.381	600.909	586.779	574.386	574.636	574.569
	0.5	679.219	648.563	618.385	616.023	657.506	637.732	617.572	616.625	616.519
	0.6	733.975	703.389	662.915	660.517	711.493	689.266	660.917	659.644	659.688
	0.7	809.507	775.264	730.042	726.416	786.035	761.061	726.758	727.200	726.707
	0.8	857.957	811.857	751.417	749.322	827.972	795.440	750.532	748.208	749.040
	0.9	880.283	844.058	791.853	790.753	856.049	830.113	791.278	789.900	788.850

Table 4. Computational Results for the Homogeneous PTSP

n	p	F	F+T	F+E	F+T+E	R	R+T	R+E	R+T+E	Queen
50	0.1	0.0000	0.0000	0.0075	0.7634	0.0000	0.0000	1.0000	1.0000	—
	0.2	0.0000	0.0000	0.0021	0.2200	0.0000	0.0945	0.9051	0.7110	—
	0.3	0.0000	0.0099	0.0048	0.6845	0.0000	—	0.1424	0.0095	0.0581
	0.4	0.0000	0.0000	0.0070	0.0409	0.0000	—	0.0904	0.0404	0.1295
	0.5	0.0000	0.0000	0.4667	0.1794	0.0000	0.0211	0.3636	0.5619	—
	0.6	0.0000	0.0000	0.0108	0.4328	0.0000	0.0000	0.6682	—	0.7860
	0.7	0.0000	0.0000	0.2819	0.5964	0.0000	0.0000	0.9782	0.7682	—
	0.8	0.0000	0.0000	0.1894	0.0571	0.0000	0.0000	—	0.3047	0.4021
	0.9	0.0000	0.0000	0.1670	0.1866	0.0000	0.2785	—	0.4873	0.1266
75	0.1	0.0000	0.0000	0.0000	0.0045	0.0000	0.0000	0.0027	—	0.3458
	0.2	0.0000	0.0000	0.0030	0.9778	0.0000	0.0000	0.2195	0.1902	—
	0.3	0.0000	0.0000	0.7352	0.6279	0.0000	0.0000	0.9672	—	0.5809
	0.4	0.0000	0.0000	0.4570	0.7199	0.0000	0.0000	0.1626	0.7544	—
	0.5	0.0000	0.0000	0.1553	—	0.0000	0.0000	0.5710	0.5843	0.9881
	0.6	0.0000	0.0000	0.0000	0.0020	0.0000	0.0000	0.0009	—	0.0112
	0.7	0.0000	0.0000	0.0068	0.1305	0.0000	0.0000	0.4880	—	0.7738
	0.8	0.0000	0.0000	0.0037	0.0406	0.0000	0.0000	0.0171	—	0.5933
	0.9	0.0000	0.0000	0.6335	0.4054	0.0000	0.0000	0.7243	0.9652	—
100	0.1	0.0000	0.0000	0.0000	0.3462	0.0000	0.0000	0.1212	—	0.2154
	0.2	0.0000	0.0000	0.0194	0.1442	0.0000	0.0000	—	0.2872	0.2267
	0.3	0.0000	0.0000	0.3428	0.4848	0.0000	0.0000	—	0.2666	0.2004
	0.4	0.0000	0.0000	0.1009	—	0.0000	0.0000	0.9924	0.7035	0.7599
	0.5	0.0000	0.0000	0.0045	—	0.0000	0.0000	0.0663	0.4873	0.4821
	0.6	0.0000	0.0000	0.0100	0.4620	0.0000	0.0000	0.2514	—	0.9728
	0.7	0.0000	0.0000	0.0012	—	0.0000	0.0000	0.7460	0.4101	0.7955
	0.8	0.0000	0.0000	0.0038	0.3420	0.0000	0.0000	0.0453	—	0.4636
	0.9	0.0000	0.0000	0.0149	0.1190	0.0000	0.0000	0.0362	0.3671	—

Table 5. p-value of Permutation test for the Homogeneous PTSP

5. Concluding comments

In this chapter, a genetic algorithm is developed to solve the PTSP. The effectiveness and efficiency of nine different selection schemes were investigated for both the heterogeneous and homogeneous PTSP. Extensive computational tests were performed and the permutation test was adopted to test the statistical significance of the nine selection schemes. Several important findings are obtained. First, fitness-proportional (F) or rank-based (R) selection should combine tournament (T) and/or elitism strategies to obtain acceptable outcomes for both the heterogeneous and homogeneous PTSP. Second, the average $E[\tau]$ values obtained by keeping the best solution(s) to the successive generations can generally perform better than the ones obtained by only applying tournament selection to fitness-proportional (F) or rank-based (R) selection for both the heterogeneous and homogeneous PTSP. Third, the queen, F+T+E, R+E, and R+T+E are better selection strategies and yielded similar average $E[\tau]$ value for the heterogeneous and homogeneous PTSP than the other five selection strategies. Finally, the numerical results showed that the proposed solution procedure can further enhance the performance of the method proposed by previous studies in most of the tested cases for the heterogeneous PTSP in terms of objective function value and computation time. These findings showed the potential of the proposed GA in effectively and efficiently solving the large-scale PTSP.

6. Acknowledgement

This work was supported primarily by the National Science Council of Taiwan under Grant NSC 96-2416-H-260-008. The author is indebted to Dr. Elise Miller-Hooks for providing me with the test instances to be used in this paper.

7. References

Bäck , T.; Fogel, D. B. & Michalewicz, Z. (1997). *Handbook of Evolutionary Computation*, Oxford University Press, 0-75030392-1, Bristol, UK.

Balakrishnan, P. V.; Gupta, R. & Jacob, V. S. (2006). An investigation of mating and population maintenance strategies in hybrid genetic heuristics for product line designs. *Computers & Operations Research,* Vol. 33, No. 3, 639-659, ISSN: 0305-0548.

Bartholdi, J. J. & Platzman, L. K. (1988). Heuristics based on spacefilling curves for combinatorial problems in Euclidean space. *Management Science*, Vol. 34, No. 3, 291-305, ISSN: 0025-1909.

Bartholdi, J. J.; Platzman, L. K.; Collins, R. L. & Warden, W. H. (1983). A minimal technology routing system for meals on wheels. *Interfaces*, Vol. 13, No. 3, 1-8, ISSN: 0092-2102.

Basso, D.; Chiarandini, M. & Salmaso L. (2007). Synchronized permutation tests in replicated I × J designs. *Journal of Statistical Planning and Inference*, Vol. 137, No. 8, 2564-2578, ISSN: 0378-3758.

Bertsimas, D. (1988). *Probabilistic combinatorial optimization problems*, Ph.D. dissertation, Massachusetts Institute of Technology, MA, USA.

Bertsimas, D.; Chervi, P. & Peterson, M. (1995). Computational approaches to stochastic vehicle routing problems. *Transportation Science*, Vol. 29, No. 4, 342-352, ISSN: 0041-1655.

Bertsimas, D. & Howell, L. (1993). Further results on the probabilistic traveling salesman problem. *European Journal of Operational Research,* Vol. 65, No. 1, 68-95, ISSN: 0377-2217.

Bertsimas, D.; Jaillet, P. & Odoni, A. R. (1990). A priori optimization. *Operations Research*, Vol. 38, No. 6, 1019-1033, ISSN: 0030-364X.

Bianchi, L. (2006). *Ant colony optimization and local search for the probabilistic traveling salesman problem: a case study in stochastic combinatorial optimization*. Ph.D. dissertation, Université Libre de Bruxelles, Brussles, Belgium.

Bowler, N. E.; Fink, T. M. A. & Ball, R. C. (2003). Characterization of the probabilistic traveling salesman problem. *Physical Review E*, Vol. 68, 036703, ISSN: 1539-3755.

Branke, J. & Guntsch, M. (2004). Solving the probabilistic TSP with ant colony optimization. *Journal of Mathematical Modelling and Algorithms*, Vol. 3, No. 4, 403-425, ISSN: 1570-1166.

Campbell, A. M. (2006). Aggregation for the probabilistic traveling salesman problem. *Computers & Operations Research,* Vol. 33, No. 9, 2703-2724, ISSN: 0305-0548.

Davis, L. (1991). *Handbook of Genetic Algorithms*. Van Nostrand Reinhold, ISBN: 0442001738, New York.

Goldberg, D. E. (1989). *Genetic Algorithms in Search, Optimization, and Machine Learning*, Addison-Wesley, ISBN: 0201157675, Reading, PA.

Holland, J. H. (1992). *Adaptation in Natural and Artificial Systems: An Introductory Analysis with Applications to Biology, Control, and Artificial Intelligence*, MIT Press, ISBN: 0262581116, Boston, MA.

Jaillet, P. (1985). *Probabilistic traveling salesman problems*, Ph.D. dissertation, Massachusetts Institute of Technology, MA, USA.

Jaillet, P. (1987). Stochastic routing problems, In: *Advanced school on stochastics in combinatorial optimization*, Andreatta, G., Mason, F. & Serafini, P. (Ed.), 192-213, World Scientific Publisher, ISBN: 9971504561, Singapore.

Jaillet, P. (1988). A priori solution of a traveling salesman problem in which a random subset of the customers are visited. *Operations Research*, Vol. 36, No. 6, 929-936, ISSN: 0030-364X.

Jaillet, P. & Odoni, A. R. (1988). The probabilistic vehicle routing problem, In: *Vehicle routing: methods and studies*, Golden, B. L. & Assad, A. A. (Ed.), 293-318, North-Holland, ISBN: 0444704078, Amsterdam.

Liu, Y-H. (2006). A scatter search based approach with approximation evaluation for the heterogeneous probabilistic traveling salesman problem, *Proceedings of 2006 IEEE Congress on Evolutionary Computation (CEC 2006)*, pp. 1603-1609, Vancouver, Canada.

Liu, Y.-H. (2007). A hybrid scatter search for the probabilistic traveling salesman problem. *Computers & Operations Research,* Vol. 34, No. 10, 2949-2963, ISSN: 0305-0548.

Liu, Y.-H. (2008). Diversified local search strategy under scatter search framework for the probabilistic traveling salesman problem. *European Journal of Operational Research,* Vol. 191, No. 2, 332-346, ISSN: 0377-2217.

Liu, Y.-H.; Jou, R.-C., Wang, C.-C. & Chiu, C.-S. (2007). An evolutionary algorithm with diversified crossover operator for the heterogeneous probabilistic TSP. *Lecture Notes in Artificial Intelligence*, 4617, 351-360, Springer, ISBN: 3540737286, Berlin.

Liu, Y.-H. & Mahmassani, H. S. (2000). Global maximum likelihood estimation procedure for multinomial probit model parameters. *Transportation Research, Part B*, Vol. 34B, No. 5, 419-449, ISSN: 0191-2615.

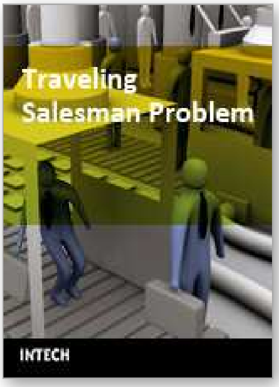
Potvin, J. Y. (1996). Genetic algorithms for the traveling salesman problem. *Annals of Operations Research*, Vol. 63, 339-370, ISSN: 0254-5330.

Rossi, F. & Gavioli, I. (1987). Aspects of heuristic method in the probabilistic traveling salesman problem, In: *Advanced school on stochastics in combinatorial optimization*, Andreatta, G., Mason, F. & Serafini, P. (Ed.), 214-227, World Scientific Publisher, ISBN: 9971504561, Singapore.

Stern, H.; Chassidim, Y. & Zofi, M. (2006). Multiagent visual area coverage using a new genetic algorithm selection scheme. *European Journal of Operational Research,* Vol. 175, No. 3, 1890-1907, ISSN: 0377-2217.

Tang, H. & Miller-Hooks, E. (2004). Approximate procedures for the probabilistic traveling salesperson problem. *Transportation Resesearch Record*, Vol. 1882, 27-36, ISSN: 0361-1981.

Whitley, D.; Starkweather, T. & Fuquay, D. (1989). Scheduling problems and traveling salesmen: the genetic edge recombination operator. *Proceedings of the Third International Conference on Genetic Algorithms (ICGA '89)*, pp. 133-140, ISBN: 1558600663, Fairfax, Virginia, June, 1989, Morgan Kaufmann, Palo Alto, CA.



Traveling Salesman Problem
Edited by Federico Greco

ISBN 978-953-7619-10-7
Hard cover, 202 pages
Publisher InTech
Published online 01, September, 2008
Published in print edition September, 2008

The idea behind TSP was conceived by Austrian mathematician Karl Menger in mid 1930s who invited the research community to consider a problem from the everyday life from a mathematical point of view. A traveling salesman has to visit exactly once each one of a list of m cities and then return to the home city. He knows the cost of traveling from any city i to any other city j. Thus, which is the tour of least possible cost the salesman can take? In this book the problem of finding algorithmic technique leading to good/optimal solutions for TSP (or for some other strictly related problems) is considered. TSP is a very attractive problem for the research community because it arises as a natural subproblem in many applications concerning the every day life. Indeed, each application, in which an optimal ordering of a number of items has to be chosen in a way that the total cost of a solution is determined by adding up the costs arising from two successively items, can be modelled as a TSP instance. Thus, studying TSP can never be considered as an abstract research with no real importance.

InTech Europe
University Campus STeP Ri
Slavka Krautzeka 83/A
51000 Rijeka, Croatia
Phone: +385 (51) 770 447
Fax: +385 (51) 686 166
www.intechopen.com

InTech China
Unit 405, Office Block, Hotel Equatorial Shanghai
No.65, Yan An Road (West), Shanghai, 200040, China
中国上海市延安西路65号上海国际贵都大饭店办公楼405单元
Phone: +86-21-62489820
Fax: +86-21-62489821