Traveling salesman should not be greedy: domination analysis of greedy-type heuristics for the TSP

Gregory Gutin
Department of Computer Science
Royal Holloway, University of London
Egham, Surrey TW20 0EX, UK
G.Gutin@rhul.ac.uk

Anders Yeo
Department of Computer Science
Aarhus University, Aarhus, Denmark
yeo@daimi.au.dk

Alexey Zverovich
Department of Computer Science
Royal Holloway, University of London
Egham, Surrey TW20 0EX, UK
A.Zverovich@rhul.ac.uk

Abstract

Computational experiments show that the greedy algorithm (GR) and the nearest neighbor algorithm (NN), popular choices for tour construction heuristics, work at acceptable level for the Euclidean TSP, but produce very poor results for the general Symmetric and Asymmetric TSP (STSP and ATSP). We prove that for every $n \geq 2$ there is an instance of ATSP (STSP) on n vertices for which GR finds the worst tour. The same result holds for NN. We also analyze the repetitive NN (RNN) that starts NN from every vertex and chooses the best tour obtained. We prove that, for the ATSP, RNN always produces a tour, which is not worse than at least $n/2-1$ other tours, but for some instance it finds a tour, which is not worse than at most $n-2$ other tours, $n \geq 4$. We also show that, for some instance of the STSP on $n \geq 4$ vertices, RNN produces a tour not worse than at most 2^{n-3} tours. These results are in sharp contrast to earlier results by G. Gutin and A. Yeo, and A. Punnen and S. Kabadi, who proved that, for the ATSP, there are tour construction heuristics, including some popular ones, that always build a tour not worse than at least $(n-2)!$ tours.

Keywords: TSP, domination analysis, greedy algorithm, nearest neighbor algorithm

1 Introduction

In this note we consider the *Asymmetric Traveling Salesman Problem (ATSP)*: given a weighted complete directed graph, $(\overset{\leftrightarrow}{K}_n, c)$, where n is the number of vertices and c is the weight function from the arc set of $\overset{\leftrightarrow}{K}_n$ to the set of reals, one seeks a hamiltonian cycle of minimum total weight. Below we call a hamiltonian cycle a *tour* and $c(a)$ the *cost* of a for an arc a of $\overset{\leftrightarrow}{K}_n$. For a tour T, its cost $c(T)$ is the sum of the costs of its arcs. The *Symmetric TSP (STSP)* is defined similarly to the ATSP apart from the fact that $\overset{\leftrightarrow}{K}_n$ is replaced by the complete undirected graph K_n. Since an instance of the STSP can be transformed into an "equivalent" instance of the ATSP by replacing every edge $\{x, y\}$ of K_n by the pair $(x, y), (y, x)$ of arcs of the costs equal to the cost of $\{x, y\}$, every heuristic for the ATSP can be used for the STSP. We well use the term TSP when it is not important whether the ATSP or STSP is under consideration.

It is well-known that for the majority of combinatorial optimization problems (including the TSP) even the problem to find an approximate solution (within a guaranteed constant factor from the optimum) is NP-hard. As a result, heuristics for such problems are usually compared using computational experiments. Glover and Punnen [3] suggested a new approach for evaluation of heuristics that compares heuristics according to their so-called domination number. We define this notion only for the TSP since its extension to other problems is obvious. The *domination number* of a heuristic $\mathcal{A}$ for the TSP is the maximum integer $d(n)$ such that, for every instance $\mathcal{I}$ of the TSP on n vertices, $\mathcal{A}$ produces a tour T which is not worse than at least $d(n)$ tours in $\mathcal{I}$ including T itself. Observe that an exact algorithm for the ATSP (STSP) has domination number $(n-1)!$ $((n-1)!/2)$.

Clearly, the domination number is well defined for every heuristic, and a heuristic with higher domination number may be considered a better choice than a heuristic with lower domination number. (This kind of comparison is somewhat similar to the standard comparison of approximation algorithms, which continues to be the most popular choice of theoretical performance analysis.)

Computational experiments show that the greedy algorithm (GR) and the nearest neighbor algorithm (NN), popular choices for tour construction heuristics, work at acceptable level for the Euclidean TSP (see e.g. [7, 9]), but produce very poor results for the general Symmetric and Asymmetric TSP (see, e.g., [1, 2, 6, 7]). For the ATSP, GR builds a tour by repeatedly choosing the cheapest eligible arc of $(\overset{\leftrightarrow}{K}_n, c)$ until the chosen arcs form a tour; an arc $a = (u, v)$ is *eligible* if the out-degree of u in D and the in-degree of v in D equal zero, where D is the digraph induced by the set S of chosen arcs, and a can be added to S without creating a non-hamiltonian cycle. NN starts its tour from a fixed vertex i_1, goes to the nearest vertex i_2 (i.e., $c(i_1, i_2) = \min\{c(i_1, j) : \ j \neq i_1\}$), then to the nearest vertex i_3 (from i_2) distinct from i_1 and i_2, etc. The repetitive NN (RNN) starts NN from every vertex and chooses the best tour obtained.

We analyze GR, NN and RNN using the domination number approach. We prove that for every $n \geq 2$ there is an instance of ATSP (STSP) on n vertices for which GR finds the worst tour, i.e., the domination number of GR for the ATSP (STSP) is 1. The same result holds for NN. We show that, for the ATSP, RNN always produces a tour, which is not worse than at least $n/2 - 1$ other tours, but for some instance on n vertices it finds a tour, which is not worse than at most $n - 2$ other tours, i.e., the domination number of RNN is between $n/2$ and $n - 1$. We also prove that, for the STSP, the domination number of RNN is at most 2^{n-2}. These results are in sharp contrast to earlier results by G. Gutin and A. Yeo [4, 5], and A. Punnen and S. Kabadi [8], who proved that, for the ATSP, there are tour construction heuristics, including some popular ones (such as the Karp-Steele patching algorithm, which is a good choice for the ATSP [2]) that always build a tour not worse than at least $(n-2)!$ tours. (It follows from the simple construction mentioned in the last sentence of the first paragraph of this section that those heuristics have domination number at least $(n-2)!/2$ for the STSP.) This provides some theoretical explanation why "being greedy" is not so good for solving the TSP.

2 Results

In the following theorems we use the notions of forward and backward arcs in $\overset{\leftrightarrow}{K}_n$, $V(\overset{\leftrightarrow}{K}_n) = \{1, 2, ..., n\}$. We call an arc (i, j) *forward* (*backward*) if $i < j$ ($j < i$).

Theorem 2.1 *The domination number of GR for the TSP is 1.*

Proof: We show this theorem only for the ATSP; the proof for the STSP is omitted. We construct an instance of the ATSP for which GR produces the worst tour. Let the cost of every arc (i,j) be $n\min\{i,j\}+1$ with the following exceptions: $c(i,i+1)=in$ for $i=1,2,\ldots,n-1$, $c(i,1)=n^2-1$ for $i=3,4,\ldots,n-1$, and $c(n,1)=n^3$.

Since the cheapest arc is $(1,2)$, GR constructs the tour $T=(1,2,\ldots,n,1)$. The cost of T is

$$\sum_{i=1}^{n-1} in + c(n,1).$$

Suppose that there is a tour H in $(\overleftrightarrow{K}_n,c)$ such that $c(H)\geq c(T)$. The tour H must contain the arc $(n,1)$ since

$$c(n,1) > n\cdot\max\{c(i,j):\ 1\leq i\neq j\leq n, (i,j)\neq(n,1)\}.$$

This implies that H contains a hamiltonian path P from 1 to n of cost at least $\sum_{i=1}^{n-1} in$. Let e_i be an arc of P whose tail is i. Observe that $c(e_i)\leq in+1$ and P must have a backward arc, say e_k. Since $c(e_k)\leq (k-1)n+1$, we have $c(P)\leq(\sum_{i=1}^{n-1} in)+(n-1)-n$, a contradiction. □

The proof of this theorem implies that the domination number of NN for TSP is also 1. Certainly, this is the case if one always starts from the vertex 1. More often, NN is initiated from a random vertex. In this case, on at least one of the n instances obtained from the instance in the theorem by exchanging vertices 1 and i, $i=1,2,\ldots,n$, NN will produce the worst tour. However, the following two theorems show that the situation is slightly better for RNN.

Theorem 2.2 *Let $n\geq 4$. The domination number of RNN for the ATSP is at least $n/2$ and at most $n-1$.*

Proof: We first consider the following instance of the ATSP, which proves that the RNN has domination number at most $n-1$. Let $N>2n$. Let all arcs $(i,i+1)$, $1\leq i<n$, have cost iN, all arcs $(i,i+2)$, $1\leq i\leq n-2$, cost $iN+1$, and all remaining forward arcs (i,j) cost $iN+2$. Let a backward arc (i,j) have cost $(j-1)N$.

When NN tour T starts at $i\notin\{1,n\}$, it has the form $(i,1,2,...,i-1,i+1,i+2,...,n,i)$ and cost

$$\ell=\sum_{k=1}^{n-1} kN - N + 1.$$

When T starts at 1 or n, we simply have $T=(1,2,...,n,1)$ of cost $\sum_{k=1}^{n-1} kN>\ell$. Let $\mathcal{F}$ denote the set of all tours T described above (note that $|\mathcal{F}|=n-1$). Observe that any tour in $\mathcal{F}$ has cost at least ℓ. Let C be any tour not in $\mathcal{F}$. Let B denote the set of backward arcs in C, and define the length of a backward arc (i,j) by $i-j$. Let q denote the sum of the lengths of the arcs in B. Since C is a tour (and therefore there is a path from n to 1) we have $q\geq n-1$. The cost of C is at most $\sum_{i=1}^{n}(iN+2)-qN-|B|N$, since if (i,j) is an arc in B, then the corresponding term $iN+2$ in the sum can be replaced by the real cost $(j-1)N=iN+2-(i-j+1)N-2$ of the arc. We have

$$\begin{aligned}\textstyle\sum_{i=1}^{n}(iN+2)-qN-|B|N &\leq \ell+N-1+2n+nN-qN-|B|N\\ &=\ell+2n+N(n+1-q-|B|)-1.\end{aligned}$$

Since C is not in $\mathcal{F}$ we have $|B| \geq 2$, implying that $2n + N(n+1-q-|B|) - 1$ is negative except for the case of $q = n-1$ and $|B| = 2$. We may conclude that the cost of C is less than ℓ, as $q = n-1$ and $|B| = 2$ would imply that C belongs to $\mathcal{F}$. Therefore all cycles not in $\mathcal{F}$ have cost less than those in $\mathcal{F}$.

In order to prove that RNN has domination number at least $n/2$, assume that this is false, and proceed as follows. RNN constructs n tours, but several of them may coincide. By the assumption, there exist at least three tours that coincide. Let $F = x_1 x_2 ... x_n x_1$ be a tour such that $F = F_i = F_j = F_k$, where F_s is the tour obtained by starting NN at x_s and x_i, x_j and x_k are distinct. Without loss of generality, we may assume that $i = 1$ and $2 < j \leq 1 + (n/2)$. For every m, with $j < m \leq n$, let C_m be the tour obtained by deleting the arcs $(x_i, x_{i+1}), (x_j, x_{j+1}), (x_m, x_{m+1})$ and adding the arcs $(x_i, x_{j+1}), (x_m, x_{i+1}), (x_j, x_{m+1})$. Note that $c(C_m) \geq c(F)$, since $c(x_i, x_{i+1}) \leq c(x_i, x_{j+1})$ (because we used NN from x_i to construct F_i), $c(x_j, x_{j+1}) \leq c(x_j, x_{m+1})$ (since we used NN from x_j to construct F_j) and $c(x_m, x_{m+1}) \leq c(x_m, x_{i+1})$ (since NN chose the arc $x_m x_{m+1}$ on F_j, when the arc $x_m x_{i+1}$ was available). Therefore the cost of F is at most that of $F, C_{j+1}, C_{j+2}, \ldots, C_n$, implying that the domination number is at least $n - j + 1 \geq n/2$, a contradiction. □

We call a tour $x_1 x_2 \ldots x_n x_1$, $x_1 = 1$, of the STSP *pyramidal* if $x_1 < x_2 < \ldots < x_k > x_{k+1} > \ldots > x_n$ for some index k. Since every pyramidal tour $x_1 x_2 \ldots x_n x_1$, $x_1 = 1$, is determined by the set $\{x_2, x_3, \ldots, x_{k-1}\}$ or the set $\{x_{k+1}, x_{k+2}, \ldots, x_n\}$ (clearly, $x_k = n$), we obtain that the number of pyramidal tours of the STSP is 2^{n-3}.

The next theorem gives an upper bound for the domination number of RNN for the STSP. Even though the theorem leaves a possibility that this domination number is exponential, it is still much smaller than $\Theta((n-2)!)$.

Theorem 2.3 *Let $n \geq 4$. The domination number of RNN for the STSP is at most 2^{n-3}.*

Proof: We consider the following instance of the STSP, which proves that RNN for the STSP has domination number at most 2^{n-3}. Let $N > 2n$. Let all edges $\{i, i+1\}$, $1 \leq i < n$, have cost iN, all edges $\{i, i+2\}$, $1 \leq i \leq n-2$, cost $iN + 1$, and all remaining edges $\{i, j\}$, $i < j$, cost $iN + 2$.

Let c_{RNN} be the cost of the cheapest tour constructed by RNN. It is straightforward to verify that

$$c_{\text{RNN}} = c(12 \ldots n1) = \sum_{i=1}^{n-1} iN + N + 2. \quad (1)$$

Let $T = x_1 x_2 \ldots x_n x_1$ be a tour in K_n, $x_1 = 1$; we orient all edges of T such that T becomes a directed cycle T'. Some of arcs in T' are forward, others are backward. For a backward arc $e = (j, i)$, we define its length as $q(e) = j - i$. We denote the sum of the lengths of backward arcs in T' by $q(T')$. (By the definition of a backward arc the length of every backward arc is positive.) Let $c_{\max}$ be the cost of the most expensive non-pyramidal tour T. Since the number of pyramidal tours is 2^{n-3}, to prove this theorem it suffices to show that $c_{\max} < c_{\text{RNN}}$.

Observe that $q(T') \geq n$ for every T' corresponding to a non-pyramidal tour T. Let H be a non-pyramidal tour of cost $c_{\max}$, and let $e_i = (i, j)$ be an arc of H'. If e_i is forward, then $c(e_i) \leq iN + 2$, and if e_i is backward, then $c(e_i) \leq jN + 2 = iN + 2 - q(e_i)N$. Thus,

$$c_{\max} \leq \sum_{i=1}^{n} (iN + 2) - q(H')N \leq \sum_{i=1}^{n-1} iN + 2n$$

as $q(H') \geq n$. Since $N > 2n$ and by (1), we conclude that indeed $c_{\max} < c_{\mathrm{RNN}}$. □

By the construction mentioned in the last sentence of the first paragraph of Section 1 and the lower bound in Theorem 2.2, the domination number of RNN for the STSP is at least $n/4$. It would be interesting to find the exact values of the domination number of RNN for the ATSP and STSP. It would be of certain interest to compute the domination numbers of several more heuristics and to analyze how the behavior of heuristics in computational experiments depends on their domination numbers.

References

[1] J. Cirasella, D.S. Johnson, L.A. McGeoch and W. Zhang, The asymmetric traveling salesman problem: algorithms, instance generators and tests, To appear in Proc. 3rd Workshop on Algorithm Engineering and Experiments (ALENEX 01), Washington DC, 2001

[2] F. Glover, G. Gutin, A. Yeo and A. Zverovich, Construction heuristics for the asymmetric traveling salesman problem, to appear in *Europ. J. Oper. Res.*

[3] F. Glover and A. Punnen, The traveling salesman problem: New solvable cases and linkages with the development of approximation algorithms, *J. Oper. Res. Soc.* **48** (1997) 502-510

[4] G. Gutin and A. Yeo, Polynomial approximation algorithms for the TSP and the QAP with factorial domination number, to appear in *Discrete Applied Math.*

[5] G. Gutin and A. Yeo, TSP tour domination and Hamilton cycle decomposition of regular digraphs, to appear in *Oper. Res. Letters*

[6] G. Gutin and A. Zverovich, Evaluation of the Contract-or-Patch Heuristic for the Asymmetric TSP, submitted

[7] D.S. Johnson and L.A. McGeoch, The traveling salesman problem: a case study in local optimization. *Local Search in Combinatorial Optimization,* E.H.L. Aarts and J.K. Lenstra (eds.), Wiley, N.Y., 215-310 (1997)

[8] A.P. Punnen and S. Kabadi, Domination analysis of some heuristics for the asymmetric traveling salesman problem, submitted

[9] G. Reinelt, *The traveling salesman problem: Computational Solutions for TSP Applications. Springer Lecture Notes in Computer Sci.* 840 (1994), Springer-Verlag, Berlin