

APPROXIMATION RESULTS TOWARD NEAREST NEIGHBOR HEURISTIC

Jérôme MONNOT

*LAMSADE, Université Paris-Dauphine, Paris, France
monnot@lamsade.dauphine.fr*

Abstract: In this paper, we revisit the famous heuristic called nearest neighbor (*NN*) for the traveling salesman problem under maximization and minimization goal. We deal with variants where the edge costs belong to interval $[a; ta]$ for $a > 0$ and $t > 1$, which certainly corresponds to practical cases of these problems. We prove that *NN* is a $(t+1)/2t$ -approximation for $\max TSP[a; ta]$ and a $2/(t+1)$ -approximation for $\min TSP[a; ta]$ under the standard performance ratio. Moreover, we show that these ratios are tight for some instances.

Keywords: Approximate algorithms, performance ratio, analysis of algorithms, traveling salesman problem.

1. INTRODUCTION

The classical traveling salesman problem can be formulated as follows: given K_n , a complete graph on n vertices with non-negative integer costs on its edges, the traveling salesman problem under minimization version, called $\min TSP$ (resp. maximization, called $\max TSP$) consists of minimizing (resp. maximizing) the cost of a Hamiltonian cycle, the cost of such cycle is the sum of its edge's costs. Moreover, when the edge-weights are in the set $\{a, a+1, \dots, b-1, b\}$, we will call of $TSP[a; b]$ problem. Several restrictions of this problem have often been studied in the literature, like Euclidean, metric or 1, 2 cases and very elegant positive or negative approximation results have being produced by Arora [1], Christofides [2], Papadimitriou and Yannakakis [7], Engebretsen and Karpinski [3], Papadimitriou and Vempala [6]. There are no special studies about this heuristic when edge-weights are in the set $\{a, a+1, \dots, b-1, b\}$.

In this paper, we revisit some approximation results for *Nearest Neighbor* algorithm (noted *NN*) described the first time by Karg and Thompson [5], also called the next best method in some sequencing jobs to a single production facility. This very simple heuristic has already been mainly studied by Fisher et al. [4] for *max TSP* and by Rosenkrantz et al. [8] for *min metric - TSP* and consist in starting from any vertex and visiting constantly the nearest vertex that has not been visited. In [4], the authors present several polynomial-time approximation algorithms, among which Nearest Neighbor achieving approximation ration $1/2$ for the maximization version whereas in [8], the results are less optimistic since they produce a $\theta(1/\log n)$ -approximation for minimization metric version, by using an approximation measure, called performance ratio, defined as:

$$\rho[\pi]_A(I) = \min \left\{ \frac{A(I)}{OPT(I)}, \frac{OPT(I)}{A(I)} \right\}$$

where $A(I)$ is the value of algorithm A and $OPT(I)$ is the value of an optimal solution on the instance I of a combinatorial problem π .

The performance ratio is a number less than or equal to 1, and is equal to 1 when $A(I) = OPT(I)$. Note that, compared to some definitions, we have inverted the performance ratio in the case of the minimization problems. Hence, we will always consider the ratio value as being between 0 and 1. We say that A is an r -approximation if for any instance I , we have $\rho_A(I) \geq r$.

A case that seems to be very common in practical situations appears when d_{\max}/d_{\min} is upper bounded by a constant. We prove that, when edge-costs belong to the interval $[a;ta]$, Nearest Neighbor is a $(t+1)/2t$ -approximation for the maximization problem and yields a $2/(t+1)$ -approximation for the minimization version.

The previous guaranteed performance on theses heuristics are strengthened by our results in both versions. Moreover, we show that ratios are tight.

2. THE NEAREST NEIGHBOR ALGORITHM

This algorithm depends on the goal of the traveling salesman problem, so when we study the maximization case, we replace goal by *max* else goal by *min*.

[NN_{goal}]

input: $I = (K_n, d)$ instance of *goal TSP*;

output: An acyclic permutation p of I ;

Take arbitrarily $x_1 \in V$;

Set $S = \{x_1\}$ and $z = x_1$;

While $S \neq V$ do

Take $y \notin S$ such that $d(z, y) = \text{goal}\{d(z, w) \mid w \notin S\}$ (line a);
 Set $y = p(z)$ and $z = y$;
 End while;
 $p(y) = x_1$;
 return p ;

We assume that when there are ties in different steps of algorithm, they can be broken by taking the vertex with minimum index, so in particular we always start with vertex x_1 . This algorithm yields an Hamiltonian cycle since an acyclic permutation describes a feasible solution by the set $\{(x, p(x)) \mid x \in V\}$ (where p points out to the successor of x in the cycle) and its complexity-time is $O(n^2)$.

The authors of [4] have proved by linear programming method that $\max TSP$ is $1/2$ -approximable, whereas we prove by a combinatorial technique that more generally $\max TSP[a; ta]$ is $(t+1)/2t$ -approximable for all $t > 1$.

Theorem 1.1. *The algorithm $[NN_{\max}]$ is a $\frac{t+1}{2t}$ -approximation for $\max TSP[a; ta]$ and this ratio is tight.*

Proof: Let $I = (K_n, d)$ be an instance on n vertices, such that $a \leq d(e) \leq ta$ for all edge e and let p^* (resp. p) be an acyclic permutation describing an optimal solution of I (resp. the solution returned by NN). We split V into $V_1 = \{x \in V \mid d(x, p(x)) < d(x, p^*(x))\}$ and $V_2 = \{x \in V \mid d(x, p(x)) \geq d(x, p^*(x))\}$. Note that $V_2 \neq \emptyset$ since by construction $x_1 \in V_2$. Moreover if $V_1 \neq \emptyset$ then the nearest neighbor heuristic is optimal and we have the main key following result:

$$\forall x \in V_1, d(p^*(x), p \circ p^*(x)) \geq d(x, p^*(x)). \quad (1.1)$$

Indeed, let $x \in V_1$; by construction $p^*(x)$ correspond to a previous step of algorithm than x (else $x \in V_2$) and then at the step $p^*(x)$, we have $x \notin S$ and the expected result.

Finally, we have

$$\begin{aligned} 2NN_{\max}(I) &= \sum_{x \in V} d(x, p(x)) + \sum_{x \in V} d(p^*(x), p \circ p^*(x)) \geq \\ &\geq \sum_{x \in V_2} d(x, p(x)) + \sum_{x \in V_1} d(p^*(x), p \circ p^*(x)) + a|V_1| + a|V_2| \\ &\geq \sum_{x \in V_2} d(x, p^*(x)) + \sum_{x \in V_1} d(x, p^*(x)) + an \geq \\ &\geq OPT_{\max}(I) + \frac{1}{t}OPT_{\max}(I). \end{aligned}$$

We now show that this ratio is tight. Let $J_n = (K_n, d)$ be an instance defined by: $V = \{x_i \mid 1 \leq i \leq 2n\}$ and for all i, j such that $1 \leq i \leq n < j \leq 2n$, we have $d(x_i, x_{j-n}) = d(x_i, x_j) = ta$ and $d(x_i, x_j) = a$. The nearest neighbor solution is described by $\forall i \leq 2n-1, p(x_i) = x_{i+1}$ and $p(x_{2n}) = x_1$ and an optimal solution by $\forall i \leq n-1, p^*(x_i) = x_{n+i}, p^*(x_{n+i}) = x_{i+1}$ and $p^*(x_n) = x_{2n}, p^*(x_{2n}) = x_1$. Finally, we have:

$$\rho_{NN_{\max}}(J_n) = \frac{\alpha(n+1)(t+1) - 2a}{2atn} \rightarrow \frac{t+1}{2t}. \quad \blacklozenge$$

In order to study the behavior NN_{\min} , we will establish a mathematical relation between respective solutions returned by algorithm on two instances linked by reduction. Moreover, we show that this relation remains true for OPT_{\max} and OPT_{\min} .

Theorem 1.2. *The algorithm $[NN_{\min}]$ is $\frac{2}{t+1}$ -approximation for $\min-TSP[a;ta]$ and this ratio is tight.*

Proof: Let $I = (K_n, d)$ be an instance on n vertices of $\min-TSP[a;ta]$, set $d_{\max} = \max_{e \in E} d(e)$ and $d_{\min} = \min_{e \in E} d(e)$. We transform instance I into instance $\alpha(I) = (K_n, d')$ just by changing the weight of edges by $d'(e) = d_{\max} + d_{\min} - d(e)$. It is clear that $\alpha(I)$ is still an instance verifying $a \leq d'(e) \leq ta$, so we can apply nearest neighbor algorithm on $\alpha(I)$ and we have:

$$NN_{\min}(I) = n(d_{\max} + d_{\min}) - NN_{\max}(\alpha(I)). \quad (1.2)$$

We show this equality by an inductive proof. Note p_{\min} (resp. p_{\max}) the solution produces by NN_{\min} (resp. NN_{\max}) on the instance I (resp. $\alpha(I)$). For an arbitrate step x (we identify current step with last vertex visited) if we have $y = p_{\min}(x)$ then $\forall z \in S, d(x, y) \leq d(x, z)$ and $\forall z \in S, d'(x, y) = d_{\max} + d_{\min} - d(x, y) \geq d_{\max} + d_{\min} - d(x, z) = d'(x, z)$, thus we have $y = p_{\max}(x)$ and more generally for any vertex $x, p_{\min}(x) = p_{\max}(x)$.

Moreover, this equality also holds for the respective optimal solution of I and $\alpha(I)$:

$$OPT_{\max}(\alpha(I)) = n(d_{\max} + d_{\min}) - OPT_{\min}(I). \quad (1.3)$$

Let p_{\min}^* be an optimal solution of I , it is a feasible solution of $\alpha(I)$, thus we have $OPT_{\min}(I) \geq n(d_{\max} + d_{\min}) - OPT_{\max}(\alpha(I))$. Conversely, since $\alpha \circ \alpha(I) = I$, we also have $OPT_{\min}(I) \leq n(d_{\max} + d_{\min}) - OPT_{\max}(\alpha(I))$.

Thanks to the equality (1.3) and since $OPT_{\min}(I) \geq d_{\min}n$, we also obtain:

$$OPT_{\max}(\alpha(I)) \leq n(d_{\min}t + d_{\min}) - OPT_{\min}(I) \leq tOPT_{\min}(I). \quad (1.4)$$

Finally, add equality (1.2) to (1.3) and thanks to previous theorem and inequality (1.4), we have:

$$\begin{aligned} NN_{\min}(I) - OPT_{\min}(I) &= OPT_{\max}(\alpha(I)) - NN_{\max}(\alpha(I)) \leq \\ &\leq \frac{t-1}{2t} OPT_{\max}(\alpha(I)) \leq \\ &\leq \frac{t-1}{2} OPT_{\min}(I) \end{aligned}$$

and the expected result holds.

We show that this ratio is tight by considering the instances $\alpha(J_n) = (K_n, d')$ where $J_n = (K_n, d)$ is defined as in the previous theorem. Thus, we obtain:

$$\rho_{NN_{\min}}(\alpha(J_n)) = \frac{2an}{a(n+1) + at(n-1)} \rightarrow \frac{2}{t+1}.$$

We give another proof of this theorem by a straightforward analysis of this heuristic in the special case where the edge-costs are only a and ta . We split V into $V_1 = \{x \in V \mid d(x, p(x)) = a\}$ and $V_2 = \{x \in V \mid d(x, p(x)) = ta\}$ and we have that V_1 (resp. V_2) is isomorphic to the edge set of cost a (resp. at) taken by the heuristic, so we have:

$$NN_{\min}(I) = a|V_1| + at|V_2| = an + a(t-1)|V_2|. \quad (1.5)$$

We do the same partition for an optimal solution p^* ; so we split V into $V_1^* = \{x \in V \mid d(x, p^*(x)) = a\}$ and $V_2^* = \{x \in V \mid d(x, p^*(x)) = ta\}$. We also have the following result:

$$OPT_{\min}(I) = an + a(t-1)|V_2^*|. \quad (1.6)$$

Moreover, the key following result establishes one relationship between sets V_i , $i = 1, 2$ thanks to optimal acyclic permutation p^* :

$$p^*(V_2 \cap V_1^*) \subseteq V_1. \quad (1.7)$$

Indeed, this mathematical relation shows that for each mistake of algorithm (i.e. $x \in V_2 \cap V_1^*$), we can find a step for which the heuristic works well (i.e. $y \in V_1$). The proof is not presented here. Finally, since p^* is a permutation, we have:

$$\begin{aligned} 2a(t-1)|V_2| &= a(t-1)|V_2 \cap V_2^*| + a(t-1)|V_2 \cap V_1^*| + a(t-1)|V_2| = \\ &= a(t-1)|V_2 \cap V_2^*| + a(t-1)|p^*(V_2 \cap V_1^*)| + a(t-1)|V_2| \leq \\ &\leq a(t-1)|V_2 \cap V_2^*| + a(t-1)(|V_1| + |V_2|) \leq \\ &\leq (an + a(t-1)|V_2 \cap V_2^*|) + atn - 2an \leq \\ &\leq OPT_{\min}(I) + atn - 2an. \end{aligned}$$

Thus, we have:

$$\begin{aligned} NN_{\min}(I) &= an + a(t-1) |V_2| \leq \\ &\leq \frac{1}{2} OPT_{\min}(I) + \frac{atn}{2} \leq \\ &\leq \frac{t+1}{2} OPT_{\min}(I) \end{aligned}$$

and the expected result holds. \blacklozenge

Finally let us notice we could show that this algorithm gives the same performance ratio for the two versions of Hamiltonian path problem (with or without a specified endpoint) through a slight modification of line a of algorithm. Nevertheless for the maximization version where the two endpoints are specified, this heuristic yields no constant approximation ratio when d_{\max}/d_{\min} is not upper bounded by a constant.

REFERENCES

- [1] Arora, S., "Polynomial time approximation scheme for Euclidian TSP and other geometric problems", *Proc. F.O.C.S.*, 1996, 2-11.
- [2] Christofides, N., "Worst-case analysis of a new heuristic for the traveling salesman problem", Technical report 338, Grad. School of Industrial Administration, CMU, 1976.
- [3] Engebretsen, L., and Karpinski, M., "Approximation hardness of TSP with bounded matrices", E.C.C.C. Report TR00-089, 2000.
- [4] Fisher, M.I., Nemhauser, G.L., and Wolsey, L.A., "An analysis of approximations for finding a maximum weight Hamiltonian circuit", *Op. Res.*, 27 (4) (1979) 799-809.
- [5] Karg, L.L., and Thompson, G.L., "A heuristic approach to traveling salesman problems", *Management Sci.*, 10 (1964) 225-248.
- [6] Papadimitriou, C.H., and Vempala, S., "On the approximability of the traveling salesman problem", *Proc. S.T.O.C.* 2000, 2000.
- [7] Papadimitriou, C.H., and Yannakakis, M., "The traveling salesman problem with distance one and two", *Math. Oper. Res.*, 18 (1993) 1-11.
- [8] Rosenkrantz, D.J., Stearns, R.E., and Lewis, P.M., "An analysis of several heuristics for the traveling salesman problem", *S.I.A.M. J. Comp.*, 6 (1977) 563-581.